Basic Electronics

An Introduction to Electronics for Science Students

Curtis A. Meyer
Carnegie Mellon University

16 August, 2007

Email: `cmeyer@curtismeyer.com`
Web: `http://www.curtismeyer.com`

Preface

This book has been written to support a one-semester laboratory course in electronics that is taught at Carnegie Mellon University (CMU). Physics majors typically take this course during the second semester of their sophomore year. The course as taught at CMU consists of two one-hour lectures and two three-hour labs per week. Students taking this course are expected to have completed three semesters of introductory physics courses. This includes an introduction to modern physics as well as a first course on mathematical methods in physics. While in principle this course should be accessible to students after completing one year of study, the additional math from the sophomore level courses helps with understanding of the math used in electronics—in particular complex numbers. However, I have tried to write the book as a self contained package with sufficient review of this math that it should be accessible. However, some of the optional topics may not be appropriate.

The material in this book has been developed from lecture notes that were used throughout the course. It also includes material that is not normally covered in the course, but is felt to show some of the diversity in the topic. The purpose of this book is not to train expert electronic designers, but rather to expose science students to basic electronics concepts in conjunction with hands-on laboratory experience that closely matches the material in the book. When possible, I have tried to present the material in context with other physics courses, as well as trying to analyze equations the way that scientists view them. Students completing such a course should be ready to use a more-sophisticated text in designing circuits for use in other laboratory environments.

In writing this book, I have drawn on handouts from and discussion with my colleagues at CMU. In particular, I would like to acknowledge course handouts written by Bob Suter. I would also like to thank Brian Quinn for his careful reading of the original sets of lecture notes and Roy Briere for many discussions on the presentation of material. I also deeply appreciate the careful reading the manuscript by Ted LaPage. His comments have made the book substantially clearer. However, even with all this help, there will certainly still remain errors in the text. For those, I must take personal responsibility. Finally, I would like to thank the several years of CMU sophomores who finally managed to make me believe that a more appropriate text needed to be developed for the course and my wife Annette and daughter Allison for convincing me to take the plunge and actually write this book.

Contents

Chapter 1

Direct-Current Circuits

1.1 Introduction

This chapter is divided into two parts. The first is a review of basic electricity and magnetism concepts. We then review the definitions of voltage and current and move on to resistance, capacitance and then inductance. The second half of this chapter then looks in detail at direct-current circuits and the concept of equivalent circuits based on the current-versus-voltage curves of various devices. These latter concepts will be important throughout the remainder of the text.

1.2 Electricity and Magnetism

1.2.1 Coulomb's Law

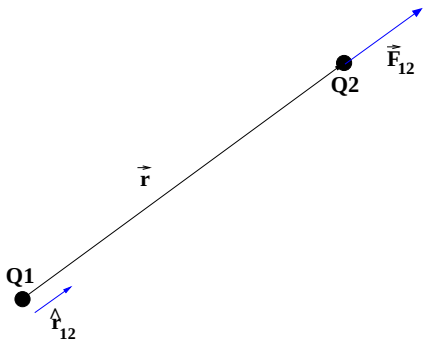

Figure 1.1: Two charges, Q_1 and Q_2 are separated by some radius, $\vec{r}_{12}$. The force exerted on charge 2 by charge 1 is given by Coulomb's law.

Any study of electricity and magnetism begins with Coulomb's law. This states that two electric charges separated by a distance r exert forces on each other given in equation 1.1, (see Figure 1.1).

$$\vec{F}_{1on2} = -\vec{F}_{2on1} = \frac{1}{4\pi\epsilon_o}\frac{Q_1 Q_2}{r^2}\hat{r}_{12}\,. \tag{1.1}$$

The quantity $\hat{r}_{12}$ is a unit vector in the direction of $\vec{r}_{12}$. As one can see, the force falls off like one over the distance squared, and is proportional to the two electric charges. In the MKS system, charge is measured in *Coulombs*, C. The basic unit of charge is the magnitude of the charge on the electron, $e = 1.6\times10^{-19}\,C$. The quantity ϵ_o is known as the *vacuum permittivity* and has a value of $\epsilon_o = 8.8541878176\times10^{-12}\,C^2/Nm^2$. From this, we find that the constant of proportionality in Coulomb's law is $\frac{1}{4\pi\epsilon_o} = 9\times10^9\,Nm^2/C^2$.

Coulomb's law can be generalized to the electric field of a point charge, $\vec{E}$, where for a point charge Q, the electric field is given as

$$\vec{E} = \frac{1}{4\pi\epsilon_o}\frac{Q}{r^2}\hat{r}.$$

This allows us to rewrite Coulomb's law for some electric charge, Q_2 placed in some electric field $\vec{E}$ as:

$$\vec{F} = Q_2\vec{E}$$

If we consider a magnetic field, $\vec{B}$, and moving charges in addition to our electric fields, then we can generalize the force law to

$$\vec{F} \quad = \quad Q_2 \left(\vec{E} + \vec{v} \times \vec{B} \right) .$$

In MKS units, $\vec{B}$ is measured in *Tesla*, T, such that $T = (N \cdot s)/(C \cdot m)$ or $N/(A \cdot m)$, in which the unit of current, *Ampere*, $A = C/s$ has been introduced.

1.2.2 Voltage

The potential energy of a charge in an electric field is measured relative to the potential energy at some reference point:

$$U(\vec{r}) - U(\vec{r}_o) \quad = \quad -Q \int_{\vec{r}_o}^{\vec{r}} \vec{E}(\vec{r}) \cdot d\vec{r}.$$

This is just the work required to move the charge Q from $\vec{r}_0$ to $\vec{r}$ through the electric field $\vec{E}$. Recall that the differential amount of work is $dW = \vec{F} \cdot d\vec{r}$ and that the potential energy is just the integral of this quantity. The potential energy, $U(\vec{r})$, is well-defined because the integral is independent of the path taken from $\vec{r}_0$ to $\vec{r}$. Potential energy has the units of energy or *Joules*, J.

The electric potential, or just *potential*, at point $\vec{r}$ is defined as the potential energy per unit of charge. The potential difference between the points $\vec{r}$ and $\vec{r}_o$ is given in equation 1.2. We refer to the potential *difference* between two points as the *voltage* between those points.

$$V(\vec{r}) - V(\vec{r}_o) \quad = [U(\vec{r}) - U(\vec{r}_o)] / Q \quad = \int_{\vec{r}_o}^{\vec{r}} \vec{E}(\vec{r}) \cdot d\vec{r}. \tag{1.2}$$

The units of potential difference, or voltage, are *Volts*, $V = J/C$. For positive Q, a point of higher potential energy is said to be at a higher potential. In fact, the charge carriers we usually deal with are electrons; we will finesse this point by pretending the charges are positive, but flowing in the opposite direction. This is discussed below. To summarize: a point of high potential is a point of high potential energy for positive charges; a point of negative potential is a point of negative potential energy for positive charges. Positive charges tend to move toward low potential points whereas electrons tend to move toward points of high potential. All potentials are defined relative to a chosen reference position and its potential, but in any physical problem, it is only the difference in potential that matters, not its absolute value at some given point. This difference is better referred to as the voltage between the points.

In electronics, we *apply a voltage to a circuit* by connecting the circuit to a *voltage source* which may be a battery, an electronic constant-voltage source (DC or "direct current" source) or a source whose voltage varies with time (AC or "alternating current" source in the case of a sinusoidal time dependence). Such a source always has (at least) two terminals (or connection points) and the "output voltage" of the source is the potential difference or voltage between these terminals.

> Physiological response to an applied voltage is due to the fact that our bodies contain charged particles which can move in response to an electric. We sense a shock because it interferes with normal body function. Our nerves operate by sending electrical signals to our spinal cord and brain through electrically conducting nerve fibers. A few volts on the skin generally does not produce a discernible result—that is, noticeable current does not reach nerve fibers. Response to larger voltage differences depends on the distance $|\vec{r} - \vec{r}_o|$ over which the voltage is applied and the physiological position of application. A voltage difference of thousands of volts applied over a distance of about 5 mm on the tip of a finger might only result in a surprise and a small burned spot.The same voltage applied from a finger on the left hand to a finger on the right hand might have dramatically different results! Here, an electrical disturbance might well pass near the heart which is controlled by electrical impulses.

A standard convention is to measure voltages relative to that of the *Earth* (i.e., $\vec{r}_0$ is some point deep in the ground). What this usually means is that we measure voltage relative to the power-company *ground* or relative to a cold water pipe which is literally attached to the earth (when things are working right, these are the same). As a practical matter, the boxes containing most electronic devices are grounded so that you do not shock yourself when you touch them.

An exception to this convention occurs when you use a battery or a *floating* power supply: these devices maintain (more or less) a fixed potential difference between their terminals. The potential *floats* relative to ground unless you explicitly connect one side to a conductor which is grounded. This is sometimes required and sometimes just a good idea. On rare occasions it cannot be done.

Since $V(\vec{r})$ is independent of the path from the reference point to $\vec{r}$, we can follow *any* path from one of the source's terminals to the other and we will see the same potential difference. An extension of this logic is known as (one of) **Kirchhoff's laws**:

$$\sum_{loop} V_i = 0, \tag{1.3}$$

where the V_i are voltage rises (or drops—you choose one or the other and use appropriate signs) across elements in a loop. Note that in going around a loop, we may have some voltage rises and some drops: once we go around an entire loop, we must return to the potential at which we started.

In electronics, we generally approximate wires as perfect conductors. Wires therefore have NO potential drop. The voltage at one end of a wire is the same as that at the other end. In circuit diagrams, we represent wires as lines. All points connected to the same line have to be at the same potential. In a circuit, the standard symbol for a DC voltage source is shown in Figure 1.2. The longer line is the terminal that is at the higher potential.

Figure 1.2: The standard electrical symbol for a DC voltage source. The terminal on the left is at a higher potential than that on the right.

1.2.3 Current

An electrical current corresponds to a flow of electrical charges. The current, I, is the charge per second passing through a cross-sectional area of interest. If there is a local density of carriers, n [m^{-3}] or, more conventionally, [cm^{-3}], of charges Q, moving with average velocity $\vec{v}$, then the amount of charge per second per area is

$$\vec{J} \;=\; nQ\vec{v}. \tag{1.4}$$

$\vec{J}$ is the current *density* with units of $C/(s \cdot m^2)$ or A/m^2. If A is the area, perpendicular to the average velocity, $\vec{v}$, over which this charge is flowing (think of a wire), then the magnitude of the total current is

$$I \;=\; JA\,. \tag{1.5}$$

Expanding this, we find that

$$I \;=\; nQvA\,. \tag{1.6}$$

In electrical circuits, the conducting objects are generally electrons flowing through various materials. For electrons, $Q = -e$, where $e = 1.602 \times 10^{-19}\,C$. Since the charge is negative (this is a convention that dates back to Benjamin Franklin—blame him!), the current flows in the opposite direction to the velocity or flow of electrons. However, I is unchanged if we replace Q by $-Q$ and $\vec{v}$ by $-\vec{v}$. Therefore,

we can just as well think of positively charged particles going in the opposite direction to the electrons. In electronics, we only concern ourselves with the direction of current flow and ignore the question of the sign of carrier.

So, in what direction does the charge flow? First of all, the electrons are bound in materials, so it would take a large amount of energy to make them flow out of the material. Thus, by configuring materials (metal wires and semiconductors) into specific shapes, we determine where current flows. In a metal, charges are free to distribute themselves in response to a potential field. If one applies a significant potential difference (voltage) between the ends of a wire, a very large current can flow. That is, if $\vec{E}$ is finite, a force will quickly generate a significant $\vec{v}$ and, thus, a significant current. On the other hand, even in a metal, collisions limit the velocity and this gives rise to *resistance*. In a *passive* circuit element, positive charge always flows from high voltage to low (that is, electrons flow in the opposite direction). Only where chemical (batteries) or other outside electromagnetic forces (DC or AC supplies) are active can charges flow in the opposite direction.

1.2.4 Conductivity, Resistivity and Resistance

In the previous section, we related the current in a material to the flow of charge carriers through the material (equation 1.4). In fact, it is an electric field, $\vec{E}$, in the material that causes the charge carriers to move. As such, it makes sense to define a proportionality constant between the vector current density, $\vec{J}$, and the electric field in a material, $\vec{E}$. This constant is known as the *conductivity* of the material and is given the symbol g. Equation 1.7 expresses this relationship.

$$\vec{J} = g\vec{E} \tag{1.7}$$

If we consider current flowing through a wire, equation 1.5 gives us the current I in terms of J and the area of the wire. Assuming that the electric field is uniform throughout the wire, the potential difference along the length, l, of the wire is just $V = E \cdot l$. If we put all of this together, we arrive at an expression that relates the current flowing through a wire to the potential difference between the ends of the wire.

$$I = \frac{gA}{l} \cdot V$$

From this, we can define the *conductance*, G, of the wire to be

$$G = \frac{gA}{l}$$

and the inverse, known as the resistance, R, of the wire to be

$$R = \frac{l}{gA}\,.$$

We can rewrite this as the familiar Ohm's Law,

$$V = I \cdot R \tag{1.8}$$

which can also be written in terms of conductance as

$$I = V \cdot G\,.$$

From equation 1.8, the dimensions of resistance are volts per ampere, or ohms, Ω. The standard symbol for a resistor is shown in Figure 1.3.

While the conductivity is a convenient number to report, one normally finds the inverse of the conductivity listed. This is defined as the resistivity of the material:

$$\rho = \frac{1}{g}\,.$$

Using the resistivity, the resistance of some material is given as $R = \rho l/A$. The dimensions of resistivity are Ωm. Table 1.1 gives resistivity values for several materials. Good conductors such as metals have a very small resistivity, while good insulators have a very large resistivity.

The resistivity of a material also has a temperature dependence. Equation 1.9 shows the temperature dependence of ρ. Resistivity is typically reported at some reference temperature, T_0, which is usually taken to be room temperature.

$$\rho(T) \quad = \quad \rho(T_0)\left[1 + \alpha\left(T - T_0\right)\right] \tag{1.9}$$

The temperature coefficient, α can then be used to compute the resistivity at some other temperature. The exact definition of the temperature coefficient, α, is

$$\alpha \quad = \quad \frac{1}{\rho}\frac{d\rho}{dT}\,. \tag{1.10}$$

From this latter definition, it is clear that equation 1.9 is only an approximation for temperature dependence of ρ. It is valid if the difference between the temperature and the reference temperature are *small enough*. It is valid for nearly everything that we might do in an electronics course with the exception of burning out a resistor!

Figure 1.3: The standard electrical symbol for a resistor.

Values for α are listed in Table 1.1. For metals, these numbers are positive, which means that the resistivity, and hence the resistance of the material goes up as the temperature is increased. Note that for semi-conductors, we find that α is negative. The resistivity goes down as the temperature is increased. We will see when we discuss semiconductors that the conductivity is in fact increasing with increasing temperature.

Material	ρ, $\Omega \cdot m$	α, K^{-1}
	Metals	
Copper	17.2×10^{-9}	0.00393
Silver	15.9×10^{-9}	0.0038
Gold	24.4×10^{-9}	0.0034
Aluminum	28.2×10^{-9}	0.0039
Brass	70×10^{-9}	0.0038
Iron	100×10^{-9}	0.005
Mercury	957.8×10^{-9}	0.00089
Nickel	78×10^{-9}	0.006
Tantalum	155×10^{-9}	0.0031
Tin	115×10^{-9}	0.0042
Zinc	58×10^{-9}	0.0037
	Semiconductors	
Carbon	35×10^{-6}	-0.0005
Germanium	0.46	-0.048
Silicon	0.64×10^{3}	-0.075
	Insulators	
Glass	10^{10} to 10^{14}	
Quartz	7.5×10^{17}	

Table 1.1: The resistivity, ρ and the thermal coefficient α for several different materials at 20 °C. Data for metals are taken from the *Handbook of Chemistry and Physics*, 56'th edition, CRC Press, Inc., Cleveland, OH. Data for the semiconductors and insulators are taken from Wikipedia.

1.2.5 Power

As with any physical system, we know that in electronics energy is conserved. However, rather than talking about energy, we usually discuss the time rate of change of energy, or the *power*, where power, P, is defined as

$$P = \frac{dE}{dt}.$$

Since energy is measured in Joules, the unit of power is a Joule per second, or a *Watt*.

$$1\,W = 1\,J/s$$

In electronics, we supply power to a circuit, some power is dissipated by the circuit (as heat), and some power can be delivered to the output of the circuit. For a current I flowing through a resistor R, the power dissipated in the resistor is just

$$P_R = I^2R.$$

Using Ohm's law, we can write the power dissipated in a resistor in two other equivalent forms:

$$\begin{aligned} P_R &= VI \\ P_R &= \frac{V^2}{R}. \end{aligned}$$

In addition to the dissipated power, we also have power supplied to a circuit. If there is a voltage source, V, which drives a current I through the circuit, then the power supplied to the circuit is

$$P_s = VI.$$

By conservation of energy, we can easily show that the power a circuit delivered to its output is equal to the power supplied minus the power dissipated.

Example: A 5-Volt DC power supply has a $1000\,\Omega$ resistor connected between its leads. How much power does the power supply provide and what is the power dissipated in the resistor? From Ohm's law, we know that the current through the resistor must be $I = V/R$, or $5\,mA$. This means that the power delivered by the supply is just $P = VI$, or $25\,mW$. The power dissipated in the resistor is $P_R = I^2R$, or again $25\,mW$.

1.2.6 Capacitance and Capacitors

If a uniform electric field, $\vec{E}$, exists between two parallel conducting plates which are not electrically connected together, then we will induce surface charges on the surfaces of the two plates, $+Q$ on the left-hand plate and $-Q$ on the right-hand plate. This is shown schematically in Figure 1.4. The two plates are each of area A and are separated by a distance l. The induced charges will be spread over the surfaces with surface densities $+\sigma$ and $-\sigma$ respectively, where $Q = \sigma \cdot A$. Continuing, we also know that the electric field of an infinite plate with uniform surface charge density σ is given as

$$E = \frac{\sigma}{\epsilon_0}.$$

The electric field creates a potential difference between the two plates of $V = El$, with the left-hand plate at a higher potential. So putting all of this together, we can arrive at a relationship between the charge on the capacitor plates, Q and the voltage between the two plates. This is given by:

$$Q = \frac{\epsilon_0 A}{l} \cdot V.$$

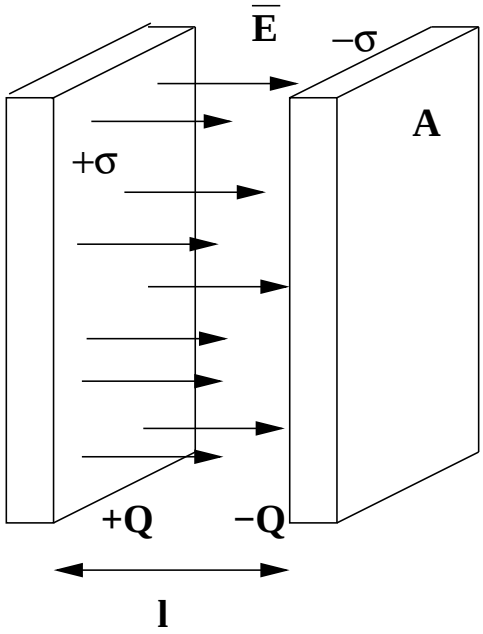

Figure 1.4: Two parallel conducting plates of area A separated by a distance l have an electric field $\vec{E}$ that points from the left plate to the right plate. A positive charge, $+Q$ is induced on the left-hand plate, while a negative charge, $-Q$ is induced on the right-hand plate. The surface charge densities are $+\sigma$ and $-\sigma$ respectively.

We define the constant of proportionality as the *capacitance*, C, of the parallel plates.

$$C = \frac{\epsilon_0 A}{l} \tag{1.11}$$

Figure 1.5: The standard electrical symbol for a capacitor.

The MKS unit of capacitance is the Farad, F, where a Farad is a Coulomb per Volt, $1F = 1C/V$. If an insulating material is inserted into the gap, it will have a permittivity ϵ, and in equation 1.11 will just replace ϵ_0 with ϵ. Normally, permittivity is written as a dielectric constant K times the vacuum permittivity. The standard symbol for a capacitor is shown in Figure 1.5.

$$\epsilon = K\epsilon_0$$

Typical values for permittivity are given in Table 1.2.

If we apply a potential difference, V, across a capacitor of capacitance C, then there will be a charges $\pm Q$ on the two capacitor plates such that

$$Q = CV\,. \tag{1.12}$$

We refer to the charge Q as the *stored* charge in the capacitor. If we place two capacitors in parallel to each other, and apply the same potential difference across both of them, then each will have a stored charge given by equation 1.12. The total stored charge will be the sum of the two individual charges. This would allow us to replace the two capacitors in parallel with a single capacitor with capacitance equal to the sum of the two individual capacitances. The *equivalent* capacitance of two capacitors in

Material	K
Vacuum	1.00
Air	1.0005
Glass	$5-10$
Mica	$3-6$
Silicon	11.7
Water	80
Barium titanate	1200

Table 1.2: Dielectric constants for various materials. The numbers are taken from Wikipedia.

parallel is just the sum of the two individual capacitances. For n capacitors in parallel, it is easy to show that the equivalent capacitance is just the sum of the individual capacitances.

$$C_{||} = C_1 + \cdots + C_n \tag{1.13}$$

If two capacitors are placed in series and a total voltage V is placed across them, then it is easy to see that each of the capacitors must have exactly the same charge on it, Q, and this Q is the charge that would be on the equivalent capacitance. We would also see that the potential difference across an individual capacitor is some fraction of the total V, in fact $V_i = Q/C_i$. The sum of the individual V_is must add up to the total voltage, V. This leads us to the formula for n capacitors in series.

$$\frac{1}{C_{series}} = \frac{1}{C_1} + \cdots + \frac{1}{C_n} \tag{1.14}$$

Electrolytic Capacitors

Figure 1.6: Three typical circuit symbols for an electrolytic capacitor. The anode is shown to the right in each of the symbols.

A particular type of capacitor is the *electrolytic* capacitor. This is made with two foil layers around a thin paper layer. The paper has been soaked in a a liquid electrolyte and one of the foils has an insulating oxide layer on the side next to the paper. The foil with the oxide layer is known as the *anode*, while the foil in electrical contact with the paper is known as the *cathode* layer. These capacitors have a voltage polarity requirement in that the anode be more positive than the cathode. If a reverse voltage is applied, it can destroy the central layer. The heat produced during the chemical reaction can cause the liquid layer to boil, and possibly explode the capacitor. In Figure 1.6 are shown three commonly used symbols for an electrolytic capacitor.

1.2.7 Inductance and Inductors

If we consider a solenoidal coil of N tightly wound turns and cross-sectional area A, the magnitude of the magnetic field in the center of the coil is given as

$$B = \mu_0 NI$$

where μ_0 is the vacuum permeability defined to be $4\pi \times 10^{-7} T/A$. The magnetic flux, Φ_B, is given as the product of B times A, so in terms of the magnetic flux, we have that

$$\Phi_B = \mu_0 NIA\,.$$

For any coil, we can relate the magnetic flux through the coil to the current flowing in the coil. This allows us to define the *self-inductance*, or just *inductance* of the coil, L, using the ratio in equation 1.15. The inductance of a coil is a constant that depends on geometry of the coil.

$$L = \frac{N\Phi_B}{I} \tag{1.15}$$

In the case where we can a explicitly compute Φ_B, it is possible to calculate L. In most cases, we would need to measure L. For the solenoid which we initially discussed, it is easy to show that the inductance is given as:

$$L = \mu_0 N^2 A\,.$$

The MKS unit for inductance is the *Henry*, H, where

$$1\,H = 1 Tm^2/A\,.$$

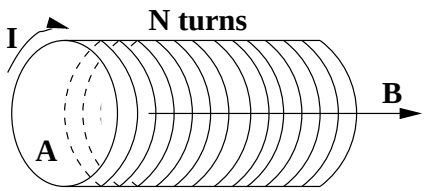

Figure 1.7: A solenoid of N turns and cross sectional area A has a current I flowing through it. This creates a magnetic field $\vec{B}$ in the center of the coil.

If we have some other material other than vacuum in the middle of the coil, we replace μ_0 with the magnetic permeability of the new material, μ. This is usually expressed in terms of relative permeability, K_m, where $\mu = K_m\mu_0$. Probably the most common core material is soft iron, which can have $K_m \sim 5000$, or even more.

We also recall from elementary electricity and magnetism that if we have a changing magnetic field passing through a coil, this induces an electromotive force, ε across the coil. If we have a current flowing through the coil, this generates a magnetic field. If the current changes, then the magnetic field changes, which in turn induces an EMF in the coil that opposes the change in the current. The consequence is that if we have some changing current in the coil, this will set up a potential difference, ε, (or V_L), across the coil. The standard electrical symbol for an inductor is shown in Figure 1.8.

$$V_L = L\frac{dI}{dt} \tag{1.16}$$

Figure 1.8: The standard electrical symbol for a inductor.

1.2.8 Transformers

A transformer consists of two separate coils of wire, both wrapped around the same core material. Because the core material has a very large μ, we generally assume that the same magnetic flux passes through each of the two coils. Figure 1.9 shows a schematic diagram of such a device. As with inductors, transformers are used with time varying voltages. An input voltage, $v_1(t)$ is connected to the two terminals on the left-hand side of the transformer. This voltage drives a current, $i_1(t)$ through the N_1 turns on the left-hand side of the transformer. This current produces a magnetic flux in the core material, given by:

$$\Phi_B(t) \approx \mu N_1 i_1(t) A\,.$$

This same magnetic flux passes through the N_2 turns of the coil on the right-hand side of the transformer. If we assume that there is no internal resistance, then the input side appears as an inductor in which

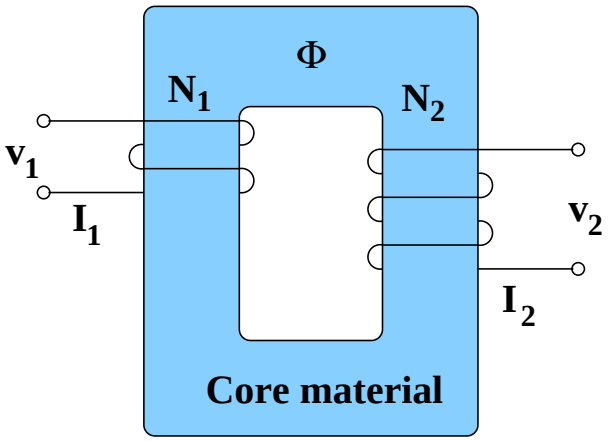

Figure 1.9: A transformer with N_1 windings on the input side and N_2 windings on the output side. In such a transformer, $v_2 = \frac{N_2}{N_1} \cdot v_1$.

the current and the voltage out of phase by 90°. We will come back to this point in chapter 2, where we will also find that in such a case, no power is dissipated. The power given as the product of v time i averages to zero.

If we initially have nothing connected to the right-hand terminals, then the ratio of the induced electromotive forces is just equal to the ratio of the number of windings in the two coils. These EMFs must be numerically equal to the voltages at the two sets of terminals. The magnitude of the voltage at the right-hand terminals is related to that at the left-hand terminals by equation 1.17.

$$v_2 = v_1 \cdot \frac{N_2}{N_1} \tag{1.17}$$

If N_2 is larger than N_1, then v_2 will be larger than v_1 and we have a so-called step-up transformer. If N_2 is smaller than N_1, then we have a step-down transformer. If we connect a load, R_L to the right-hand

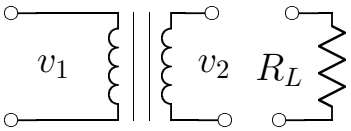

Figure 1.10: A transformer with a load resistance, R_L that can be connected to its output.

side of the transformer as shown in Figure 1.10, then by conservation of energy, the power delivered to the left-side of the transformer must be the same as that dissipated on right-hand side. The exact details of this are no relevant at this point, only that in order for energy to be conserved, we must have the following.

$$v_1 i_1 = v_2 i_2 \,.$$

We also have that the current on the right-hand side must be

$$i_2 = \frac{v_2}{R_L} \,,$$

so we find that the current on the input side is

$$i_1 = \frac{v_1}{(N_2/N_1)^2 \, R_L} \,.$$

To the input voltage, the transformer appears to be an equivalent load of:

$$R_{eq} \quad = \quad (N_2/N_1)^2 \, R_L \, .$$

The transformer not only changes voltages and currents, it also transforms resistances.

1.3 Simple Circuits

1.3.1 Time scales

In this course, we usually deal with circuits in which the rates of change of voltages and currents are small enough that we do not have to worry about propagation times. As long as the speed of light, c, divided by the fractional rate of change of a quantity gives a distance larger than the circuit dimensions, we can make the following approximations:

- The current is the same everywhere within an unbranched piece of a circuit.
- The potential at any point along a wire or combination of directly connected wires is the same (as long as the current isn't too large).
- Charge does not accumulate at any point in a circuit.

We can think of a set of directly connected wires as a single *electrical point* or *node*. The sum of currents coming into any node has to be zero (i.e., whatever comes in through one leg has to go out through another). This is the second of **Kirchhoff's laws**.

$$\sum_{node} I_i \quad = \quad 0 \tag{1.18}$$

We will now consider simple circuits consisting of voltage sources and resistors. If we apply some potential difference, V, across a resistance, R, then a current given by $I = V/R$ will flow through the resistor. More often we talk about this in terms of a current I flowing through some resistance R and producing a potential difference, $V = IR$, across the resistor. We can also use conservation of electric charge to show that if we have a set of elements connected in series (one after the other), then the same current, I, must flow through all of them. If we also assume that we have perfect wires (no voltage drop along the wire), then any two points in a circuit connected directly together by a wire must be at the same potential.

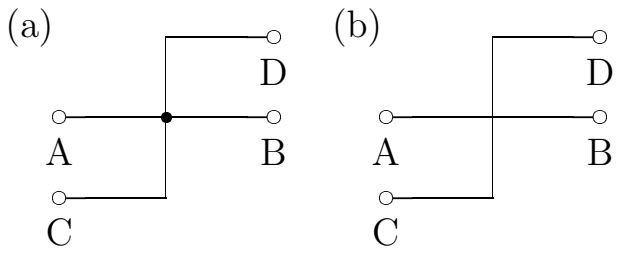

Figure 1.11: Circuit (a) shows a junction dot at the crossing of the two wires. This means that they are electrically connected. Circuit (b) does not have a junction dot at the intersection of the two wires which means that the wires are not connected at the crossing.

In drawing these circuits, it is sometimes necessary for the wires to cross over each other. We establish the convention shown in Figure 1.11 to determine whether two crossing wires are connected to each other. If we place a junction dot at the intersection as in 1.11(a), then the wires are connected together. Thus in figure 1.11(a), points A, B, C and D are all at the same potential. If there is

no junction dot, then no connection exists between the wires. In 1.11(b), points A and B are at the same potential, and points C and D are at the same potential. However, points A and C do not have to be at the same potential. In order for this convention to work, we have to assume that wires go straight through these junctions. They do not change directions at the junction. This means that in Figure 1.11(b), the points A and B are connected together, and the points C and D are connected together.

1.3.2 Resistors in Series and Parallel

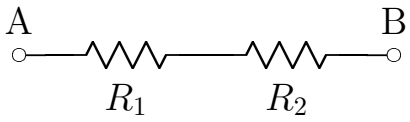

Figure 1.12: Two resistors connected in series.

Let us use these simple relations to examine the behavior of resistors in series and parallel. Consider a battery of voltage V connected to two resistors in series, R_1 and R_2, as shown in Figure 1.12. The same current, I, flows through both resistors. This tells us that the voltage drops across each resistor must add up to V.

$$\begin{aligned} V &= I \cdot R_1 + I \cdot R_2 \\ V &= I \cdot (R_1 + R_2) \end{aligned}$$

It is also true that the voltage divided by the total resistance of the two resistors must be I as well. If we define R_{eq} as the *equivalent* resistance of the two resistors, we find the following.

$$\begin{aligned} I \cdot R_{eq} &= V \\ I \cdot R_{eq} &= I \cdot (R_1 + R_2) \end{aligned}$$

In particular, we have that

$$R_{eq} = R_1 + R_2 \tag{1.19}$$

or two resistors in series are equivalent to a single resistor whose value is the sum of the two.

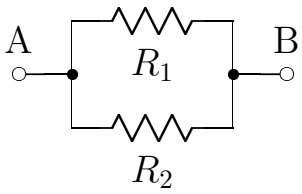

Figure 1.13: Two resistors connected in parallel.

If we now consider two resistors connected in parallel as shown in Figure 1.13, then we know that the voltage across each of the individual resistors must be the total voltage, V, but that a different current will flow through each resistor.

$$\begin{aligned} I_1 &= \frac{V}{R_1} \\ I_2 &= \frac{V}{R_2} \end{aligned}$$

Similarly, the current through the entire circuit, $I = I_1 + I_2$ is just:

$$I = \frac{V}{R_{eq}}$$

Combining all of these, we find that:

$$\frac{V}{R_{eq}} = \frac{V}{R_1} + \frac{V}{R_2}$$

or that the equivalent resistance of two resistors in parallel is given as:

$$\frac{1}{R_{eq}} = \frac{1}{R_1} + \frac{1}{R_2}\,. \tag{1.20}$$

1.3.3 Voltage and Current Meters

In the laboratory, we use ammeters to measure current and voltmeters to measure voltage. Both of these devices are based on a third device which produces a displacement that is proportional to the current flowing through it. In older equipment, this was a galvanometer with a needle that moved to indicate either positive or negative current. In modern digital devices, a circuit is used to produce the same effect. In both cases, we will refer to this as a galvanometer. The properties of a galvanometer are as follows:

1. The displacement is linearly proportional to the current flowing through the device for both positive and negative currents.
2. There is some maximum current, I_{max}, that can flow through the device. This current produces the largest displacement. Typically, I_{max} is a small number such as $1\,mA$.
3. The galvanometer coil has some internal resistance, R_{coil}, that is typically on the order of $10\,\Omega$.

Knowing these properties, we can turn a galvanometer into either an ammeter or a voltmeter using the two circuits shown in Figure 1.14.

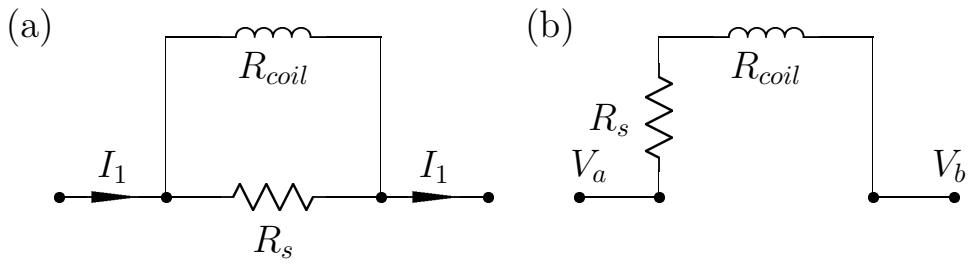

Figure 1.14: Circuit (a) shows the internal connections of a typical ammeter. The current is split to flow through both the galvanometer coil with resistance R_{coil} and the shunt resistor, R_s. Circuit (b) shows the internal wiring of a typical voltmeter. The current follows a single path through both the shunt resistor, R_S and the galvanometer coil with resistance R_{coil}.

The Ammeter

An ammeter measures the current flowing through the it—in order to do this, it has to be physically put in the path of the current that you want to measure. In order to protect the galvanometer coil, we need to physically split the current into two paths inside the meter. One path goes through the coil,

and the second path goes through a shunt resistor, R_s. If we want to measure currents up to $1\,A$, then our circuit has to split the current such that when $1\,A$ flows into the meter, only I_{max} is allowed to flow through the coil, and the remainder flows through the shunt resistor. This means that the shunt resistor is typically much smaller than the coil resistance, R_{coil}. In order to protect the meter, most meters are manufactured with a fuse in the same path as the coil. The fuse is chosen such that if you try to put more than a few times I_{max} through the coil, the fuse will blow.

As an example, let us assume that $R_{coil} = 10\,\Omega$, $I_{max} = 10\,mA$ and the maximum current we want to measure is $I = 10\,A$. The parallel combination of R_{coil} and R_s gives us an equivalent meter resistance of

$$R_{meter} = (R_s R_{coil}) / (R_s + R_{coil}) .$$

For a $10\,A$ current, the voltage drop across the meter is $V_m = I \cdot R_{meter}$.

$$\begin{aligned} V_m &= \frac{10\,A \cdot 10\Omega \cdot R_s}{10\Omega + R_s} \\ V_m &= \frac{100\,V R_s}{10\Omega + R_s} \end{aligned}$$

The current flowing through the coil is then given as $I_{coil} = V_m / R_{coil}$.

$$\begin{aligned} I_{coil} &= \frac{10\,A R_s}{10\Omega + R_s} \\ 10\,mA &= \frac{10\,A R_s}{10\Omega + R_s} \end{aligned}$$

This can be solved for the shunt resistance, where we find that

$$R_s = 0.101\,\Omega.$$

From this, the resistance of the meter is:

$$R_m = 0.10\,\Omega.$$

In order to both maintain precision and maximize the range of currents one can measure, a meter typically has several different shunt resistors which are switched in depending on the maximum current that we want to measure. While the resistance of an ammeter is quite small, it is not zero. As such, in any real circuit, there will be a small voltage drop across a current meter. This may or may not affect the results of your measurements.

The Voltmeter

A voltmeter measures the voltage between two points in a circuit. In order to do this, it needs to be connected in parallel to the circuit between the two points that it is measuring. Anytime we connect something into a circuit between two points with a voltage difference, current will flow through the device. To prevent the voltmeter from distorting the circuit, we want the current that flows through it to be very small. To accomplish this, we place a large resistor, R_s in series with the coil, R_{coil}. The resistance of the meter is then $R_m = R_{coil} + R_s$. An important consideration in the design of a voltmeter is that the maximum current should be small compared to the typical currents in a circuit.

As an example, let's assume that we want to measure a voltage up to $10\,V$ using a meter with $R_{coil} = 10\,\Omega$ and $I_{max} = 1\,\mu A$. Since the $1\,mA$ flows through the entire circuit, we want the $10\,V$ voltage drop to be across the entire meter. This then gives:

$$\begin{aligned} 10\,V &= (R_s + 10\,\Omega) \cdot 1\,\mu A \\ R_s &= 9999990\,\Omega \\ R_s &\approx 10 M\Omega. \end{aligned}$$

From this, we find that $R_m = 10000000\,\Omega$.

In general, the resistance of a voltmeter is very large, and for most practical purposes can be assumed to be infinite. However, if one is measuring voltages across very large load resistances, R_L, it is crucial that the meter resistance, R_m is very large compared to R_L. If not, it will be necessary to make corrections.

Measuring Voltages and Currents

Because physical ammeters do not have zero internal resistance and physical voltmeters do not have infinite internal resistance, using these devices may sometimes provide a distorted measurement. If you suspect that the meters are actually affecting your measurement, one way to account for this is to make measurements with both circuits shown in Figure 1.15. The voltages and currents that are measured in the two cases will be different, but the measurements can be combined to solve for what the current and voltage would be if the meters were not in the circuit. This exercise is left for the reader in problem 12.

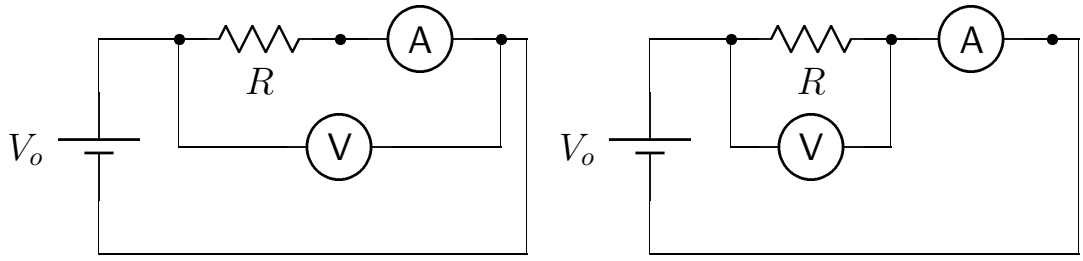

Figure 1.15: Two possible circuits for measuring the voltage across some load, R, and the current through the same load. In the limit of perfect voltmeters (infinite internal resistance) and ammeters (zero internal resistance), both of these circuits would yield the same results.

Measuring Resistances

Modern multi-meters that can measure voltage and current are usually also capable of measuring the resistance between their probes. They do this by setting up a known potential difference between the probes, and then measuring the current that flows through them. If we simply place a resistor between the probes, we will very likely get an accurate reading of its resistance. However, if we try to make a measurement of a resistor that is part of a circuit, we may encounter problems. It is likely that there is more than one path through the circuit between the two probes–not just the resistor we are probing. We would measure the equivalent resistance of all of these paths. If, as part of the functioning of the circuit, there is a potential difference across the resistor, then our reading can be even more distorted. It is even possible to measure an apparent negative resistance. If we want to measure the resistance of some component, we need to make the measurement when it is not part of a circuit.

1.3.4 Circuit Analysis using Kirchhoff's Rules

In many circuits, we can work our way through the circuit using the rules for series and parallel components to solve for voltage and current in any part of the circuit. However, some circuits get very messy when we apply these rules, and others are such that it is not possible to make progress with only these rules. In these cases, we need more powerful techniques to analyze circuits. One such technique is to use the so-called Kirchhoff's laws, as defined earlier in this chapter. These are:

- The sum of currents coming into any node has to be zero (equation 1.18).
- The sum of voltage drops around a closed loop has to be zero (equation 1.3).

We will use these rules to analyze the circuit shown in Figure 1.16.

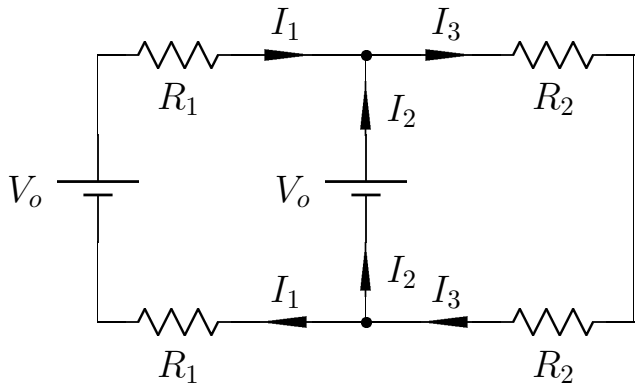

Figure 1.16: Example circuit for using Kirchhoff's rules.

There are two nodes in this circuit: the points in the center of the circuit where three wires come together. Let us start by examining the node at the center top of the circuit. As shown in the figure, we can define the current flowing out of R_1 into the node as I_1. The current that flows up from the voltage source into the node is I_2 and the current that flows out of this node is I_3. By conservation of charge, we know that the same currents flow through the wires that go into the lower node as well. If we apply the first Kirchhoff rule to the upper node, we have that I_1 and I_2 flow into the node and I_3 flows out. This gives us that:

$$I_1 + I_2 - I_3 = 0. \tag{1.21}$$

The lower node would yield the exact same equation. Now let us look at the loop around the outside of the circuit. If we follow the circuit around, every time we pass through a voltage source from negative to positive, we add the voltage. If we go in the reverse direction, we subtract the voltage. When a current goes through a resistor in the direction that we are going, we subtract $I \cdot R$. When the directions are opposite, we add $I \cdot R$. If we start from the battery at the left and write the equation as we go clockwise around the outside of the circuit, then we have:

$$0 = V_o - I_1 \cdot R_1 - I_3 \cdot R_2 - I_3 \cdot R_2 - I_1 \cdot R_1$$

$$0 = V_o - 2 \cdot (I_1 \cdot R_1 + I_3 \cdot R_2) \tag{1.22}$$

If we start at the same point, but move clockwise around the left-most loop, we have:

$$\begin{aligned} 0 &= V_o - I_1 \cdot R_1 - V_o - I_1 \cdot R_1 \\ 0 &= -2 \cdot I_1 \cdot R \end{aligned}$$

which tells us simply that $I_1 = 0$. Finally, we can look at the right-most loop in the circuit. Again, moving clockwise around this loop, we can write that

$$0 = V_o - I_3 \cdot R_2 - I_3 \cdot R_2$$

$$0 = V_o - 2 \cdot I_3 \cdot R_2 \tag{1.23}$$

Subtracting equation 1.23 from equation 1.22, we obtain the that:

$$0 = -2 \cdot I_1 \cdot R_1$$

or, we again find that $I_1 = 0$. We can then put this into equation 1.22 to obtain:

$$I_3 = \frac{V_o}{2R_2}.$$

Finally, from equation 1.21, we find that $I_3 = I_2$. If we write down all possible equations, there will be redundant information. However, some combinations of these equations may be easier to solve than others.

1.4 Power dissipation

Above, we discussed the fact that a charge moving through a potential difference in vacuum will acquire kinetic energy. We are not concerned here with vacuum electronics. In a material, charges experience collisions and lose kinetic energy. All the energy they would have gained shows up as heat or stored energy of some kind. Electrical resistance is analogous to a frictional force and gives rise to heat.

If a current, I, passes through a potential difference, V, the rate of energy dissipation or storage is given as:

$$P = VI. \tag{1.24}$$

If the current is passing through a resistive element of resistance R, then we also know that $V = I \cdot R$. We can rewrite the above equation to yield the power dissipated in the resistor as

$$P = I^2R = V^2/R. \tag{1.25}$$

1.5 Equivalent Circuits

As we work through this course, we will find that any linear two-terminal network is equivalent to a single resistor, R, in series with a single voltage source, V. The idea of equivalent circuits is something that should be familiar from introductory electricity and magnetism. We want to replace something complicated by something simpler that has the same behavior. Thus, we can more easily understand the global behavior of the circuit while not having to worry about the details of what is happening in the circuit. At this point, we effectively assume that any two-terminal network of resistors and voltage sources is equivalent to a single resistor R in series with a single voltage source V. This is a remarkable assumption about which we would be well advised to ask "Why is it true?" Before we answer this, let us explore what we mean by equivalent.

1.5.1 Simple Equivalents

We will start with two equivalents with which we are already quite familiar. These are the cases of resistors in series and parallel. Figure 1.17 shows two resistors in series while Figure 1.18 show two resistors in parallel. Each of these cases can be replaced by an equivalent resistance as given in the following equations.

$$\begin{aligned} R_{eq}^{\text{series}} &= R_1 + R_2 \\ \frac{1}{R_{eq}^{\text{parallel}}} &= \frac{1}{R_1} + \frac{1}{R_2} \end{aligned}$$

While we call this an equivalent resistance, what do we mean when we say they have the *same behavior*? The answer has to do with the relationship between the voltage, V, that we place across the device and the current , I, that flows through it. If we apply some known voltage across two equivalent elements,

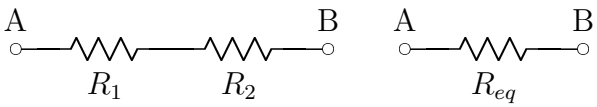

Figure 1.17: Two resistors in series, R_1, and R_2 can be replaced by a single equivalent resistance whose value is $R_{eq} = R_1 + R_2$.

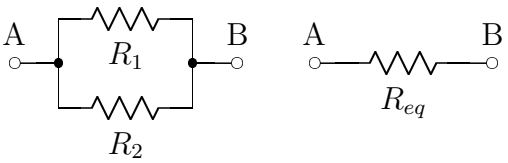

Figure 1.18: Two resistors in parallel, R_1, and R_2 can be replaced by a single equivalent resistance whose value is $\frac{1}{R_{eq}} = \frac{1}{R_1} + \frac{1}{R_2}$.

the same current will flow through each one. This is shown in figure 1.19 where a voltage V_o is applied across two resistors in series, R_1 and R_2. The current that flows through the circuit is

$$I_o = \frac{V_o}{R_1 + R_2}.$$

If the same voltage is applied to the equivalent resistance, R_{eq}, then the current is given as follows.

$$\begin{aligned} I_{eq} &= \frac{V_o}{R_{eq}} \\ &= \frac{V_o}{R_1 + R_2} \\ &= I_o \end{aligned}$$

The same current flows through both circuits.

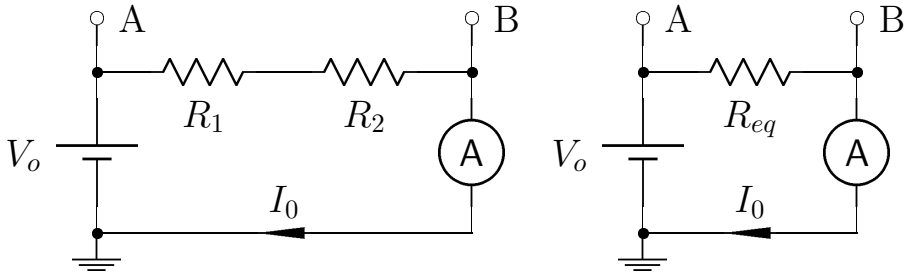

Figure 1.19: When a voltage V_o is applied across two resistors in series, a current I_o flows through them. When the same voltage, V_o, is applied across the equivalent resistance, the same current, I_o, flows.

1.5.2 I-V Curves

We can characterize the relationship between current and voltage with an I–V curve—a plot of the current through the device as a function of the voltage across the device. For a resistor, the I–V curve is a straight line with a slope of $\frac{1}{R}$ that passes through the point $(0, 0)$ as shown in Figure 1.20. In fact, this is the I–V curve for any network of resistors.

$$I = \frac{V}{R} \tag{1.26}$$

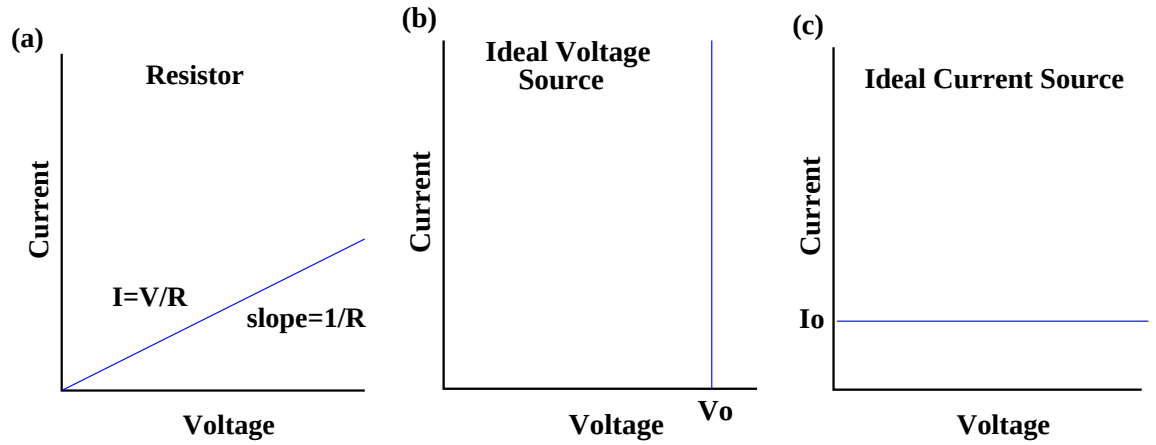

Figure 1.20: Plot (a) is the I–V curve of a resistor, R. The slope of the line is $1/R$. The line has an equation of $I = V/R$. Plot (b) is the I–V curve of an ideal voltage source. Plot (c) is the I–V curve of an ideal current source.

I–V curves apply to devices other than resistors as well. Let us consider an ideal battery. Such a device has a fixed voltage, V_o, between its two terminals and is known as a *voltage source*. We can characterize its behavior by an I–V curve. Figure 1.21 shows the voltage source and the circuit necessary to measure its I–V curve. The circuit consists of a variable resistor, R_{var}, and voltage meter, V, and a current meter, A. The I-V curve is produced by making measurements of I and V for different values of the resistance R. In making these measurements for an ideal voltage source, the voltage is always going to be V_o, while the measured current will be given by $I = V_o/R$. Such a curve is shown in Figure 1.20(b). **An ideal voltage source delivers the same voltage across its output terminals independent of what is attached to it.**

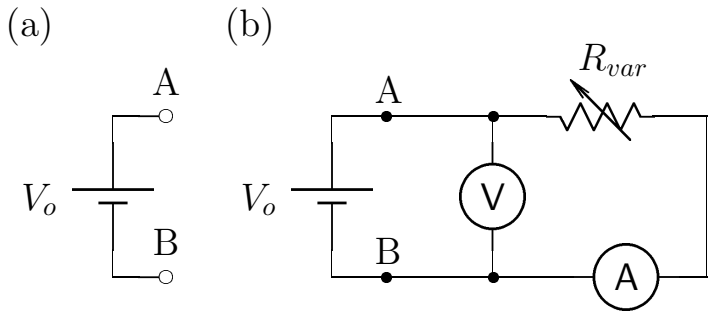

Figure 1.21: Figure (a) is an ideal voltage source of voltage V_o. Figure (b) is the circuit necessary to measure the I–V curve of the voltage source. It contains a variable resistor, R_{var}, a voltage meter, V, and a current meter, A.

Related to an ideal voltage source is an ideal current source. **An ideal current source delivers the same current through its output terminals independent of what is attached to it.** The I–V curve for such a device is shown in Figure 1.20(c). As we might well expect, perfect current or voltage sources do not exist, but it is possible to build devices that behave in this fashion over some finite range.

We can extend the concept of equivalent circuits and the I–V curve to any linear circuit that has two terminals (the points A and B in our resistor examples).

Example: Determine the I–V curve for the circuit shown in Figure 1.22(a).
In order to do this experimentally, we would need a variable resistor, R_{var}, a voltage meter, V, and a current meter, A. This now gives us the circuit on the right-hand side of Figure 1.22(b). To determine

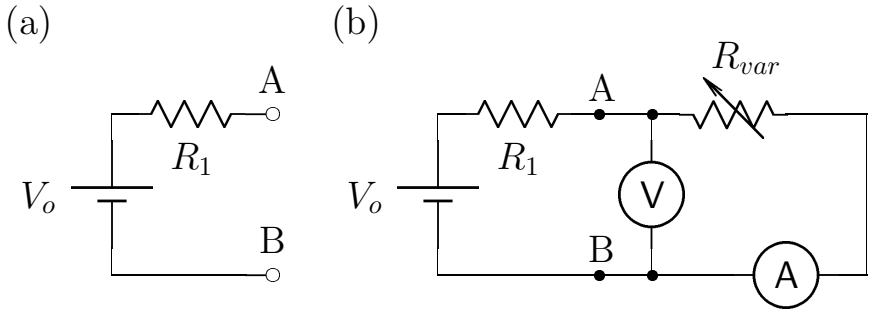

Figure 1.22: We want to determine the I–V curve for the circuit on the left-hand side. To do so, we build the circuit shown on the right-hand side of the plot.

the I–V curve, we now want to solve for both the current and voltage as a function of the value of the variable resistor. The current through the variable resistor is given as:

$$I = \frac{V_o}{R_1 + R_{var}}$$

Using this, we can get the voltage across the variable resistor as

$$\begin{aligned} V_{AB}(R_{var}) &= I \cdot R_{var} \\ &= V_o \cdot \frac{R_{var}}{R_1 + R_{var}} \end{aligned}$$

Let us now evaluate this for several values of the variable resistor. The results are given in Table 1.3, while Figure 1.23 is a plot of the I–V curve. The slope of the line is $-1/R_1$ and its equation is given as

$$I = \left(V_o - V_{AB}\right)/R_1$$

R_{var}	V_{AB}	I
0	0	V_o/R_1
$R_1/2$	$V_o/3R_1$	$2V_o/3R_1$
R_1	$V_o/2$	$V_o/2R_1$
$2R_1$	$2V_o/3$	$V_o/3R_1$
∞	V_o	0

Table 1.3: Example voltage and current data.

1.5.3 Equivalent Circuits

We can now return to our earlier question about equivalent circuits and discuss when it is possible to replace a two-terminal network with a resistor in series with a voltage source. First, it is restricted to circuits containing only *linear* elements (resistors obeying Ohm's law, and voltage and current sources). When this is the case, we can use Kirchhoff's laws to construct a set of linear algebraic equations involving voltages across and currents through the elements. Properly done, these equations can be

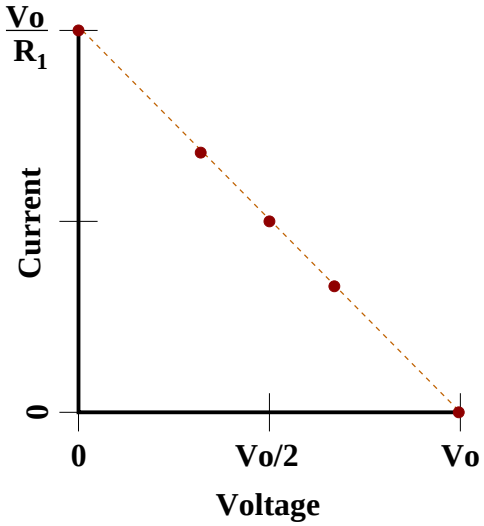

Figure 1.23: The I–V curve of the circuit in Figure 1.22.

reduced to N equations in N unknown currents. Each current will contain a contribution linearly proportional to each voltage source in the circuit divided by a coefficient with units of resistance.

Now a very important point: *The concept of an equivalent circuit applies to only one pair of terminals at a time.* Any pair of terminals you choose will have an equivalent circuit, but that equivalent will be different from one set of terminals to another. You can open a circuit up at any point and construct an I-V curve relating the current through the terminals to the voltage across them. This will have to be a linear relation because of the linearity of the governing equations. There are only two cases to worry about:

- The circuit is composed entirely of passive, linear elements - i.e., resistors. The open-circuit voltage will necessarily be zero. If we apply a positive voltage, we will put a positive current into the circuit. Hence, the I-V plot is a straight line through the origin with a positive slope (as long as we choose our signs right). The inverse slope is the equivalent resistance of the circuit.
- The circuit contains resistors and sources. The open circuit voltage (the voltage when no current is drawn) will be finite, giving a point on the horizontal or V-axis. With a short circuit across the terminals (i.e., a wire), there will be a finite current and zero voltage. Thus, we have a straight line that does not go through the origin. Any such line can also be generated by a single source and a single resistor: a voltage source in series with a resistor (the Thèvenin equivalent) or a current source in parallel with a resistor (the Norton equivalent).

We will find that these simple ideas of equivalent circuits also hold for alternating-current circuits as well as direct-current circuits. Because equivalent circuits are so useful in understanding circuit behavior, understanding how to use them is crucial to the mastery of the material in the remainder of this book.

1.5.4 Thèvenin and Norton Equivalents

Two very useful equivalent circuits are the Thèvenin equivalent circuit and the Norton equivalent circuit. Any two-terminal circuit can be replaced by either of these. In this course, the Thèvenin equivalent is more useful, but it is important to understand both of them. The Thèvenin equivalent circuit is an ideal voltage source, V_{th}, in series with a resistance, R_{th} (see Figure 1.24(a)). The Norton equivalent Circuit is an ideal current source, I_N, in parallel with a resistance, R_N (see Figure 1.24(b)). In the previous

example, we have already determined the I-V curve of the Thèvenin circuit. We now claim that the Norton circuit has an identical I–V curve to the Thèvenin circuit.

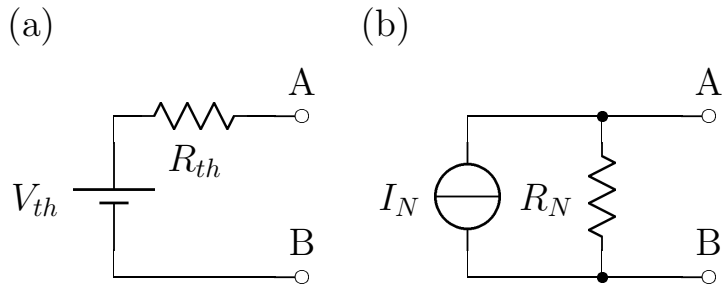

Figure 1.24: The Thèvenin equivalent circuit (a) and the Norton equivalent circuit (b).

We can show this by again examining the behavior of the circuit when a variable resistor is attached between the terminals A and B. It is easy to show that we obtain the following values of I and V as a function of R_{var}. The Norton and Thèvenin Circuits are equivalent to each other if we have $R_{th} = R_N$

R_{var}	V_{AB}	I
0	0	I_N
R_N	$I_N \cdot R_N/2$	$I_N/2$
∞	$I_N \cdot R_N$	0

and $I_N = V_{th}/R_{th}$.

If we are now given some circuit with two outputs, A and B, how do we go about determining the values of R_{th} and V_{th} for the equivalent circuits? The answer is that we determine the I–V curve of the circuit. In the lab, we do this by making a pair of measurements. The two obvious measurements to make are the *open-circuit voltage* and the *short-circuit current*. If we simply measure the voltage between A and B with nothing else hooked up, we measure the so-called *open-circuit voltage*. This voltage is V_{th}, (or I_N/R_N). If we then connect A to B with a wire, the current through the wire is known as the *short-circuit current*. This current is I_N (or V_{th}/R_{th}). The Thèvenin resistance can then be obtained as:

$$R_{th} = V_{th}/I_N. \tag{1.27}$$

However, unless one clearly understands the behavior of the circuit, shorting the terminals together is generally a very bad idea. It might be possible for a very large current to flow which could burn out either the measuring devices or even the circuit in question. A much better approach is to put some finite resistance, R_L, between A and B and measure the voltage across this resistor, V_L. In such a measurement, it can be shown that

$$R_{th} = R_L \cdot \left(\frac{V_{th} - V_L}{V_L} \right) \tag{1.28}$$

The Thèvenin and Norton theorems also hold for AC circuits. We simply replace the term "resistance" with "impedance" and the same rules apply. The "equivalent impedance" may require more than one circuit element, but conceptually, it is still just a number (a complex number, as we will see).

Example: A $1\,k\Omega$ resistor is placed between the terminals of some circuit and a voltage drop of $1\,V$ is measured. A $5\,k\Omega$ resistor is placed across the terminals and a voltage drop of $2\,V$ is measured. What are V_{th} and R_{th} of the Thèvenin equivalent circuit?

From the data provided, we can determine the currents flowing through the two load resistors: $I_{1k} = 1\,mA$ and $I_{5k} = 0.4\,mA$. This now gives us two points on our I–V curve: $(1\,V,1\,mA)$ and $(2\,V,0.4\,mA)$. The slope of the line between the points is $-1/R_{th}$, so we have:

$$\begin{aligned} -1/R_{th} &= (1\,mA - 0.4\,mA)\,/\,(1\,V - 2\,V) \\ &= 0.6\,mA/-1\,V \\ &= -.0006\,\Omega^{-1} \\ R_{th} &= 1.67\,k\Omega \end{aligned}$$

The voltage drop across the load resistor is now given as

$$\begin{aligned} V_L &= V_{th} \cdot R_L/\,(R_{th} + R_L) \\ 1\,V &= V_{th} \cdot 1\,k\Omega/2.67\,k\Omega \\ V_{th} &= 2.67\,V \end{aligned}$$

Example: Determine the parameters of the Thèvenin equivalent for the circuit shown in Figure 1.25. To compute the parameters of the Thèvenin equivalent circuit, we need to mathematically determine

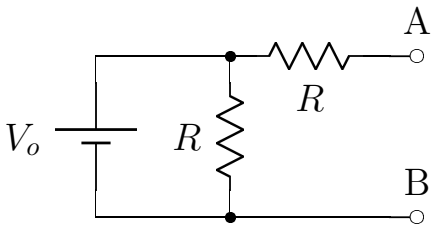

Figure 1.25: Circuit for which the Thèvenin equivalent is computed.

the open-circuit voltage and the short circuit current. The open circuit voltage just gives V_{th}.

$$V_{th} = V_o$$

If we short the two terminals together, the short-circuit current will be I_N. In this case, The current coming out of the voltage source will be $I = 2V_o/R$. This current will split equally over both branches of the circuit, so we will find that

$$I_N = V_o/R$$

This then yields that the Thèvenin resistance is:

$$\begin{aligned} R_{th} &= \frac{V_o}{I_N} \\ R_{th} &= \frac{V_o}{V_o/R} \\ R_{th} &= R \end{aligned}$$

Example: Consider the circuit shown in Figure 1.26. Let us start by looking left into the circuit across points A and B. The Thèvenin equivalent of the circuit is shown as the circuit in Figure 1.27. By using the techniques we have discussed for simplifying circuits, it is easy to show that the open-circuit voltage, $V_{AB} = \frac{V_o}{5}$. This is just the Thèvenin voltage of the circuit.

$$V_{th} = \frac{V_o}{5}.$$

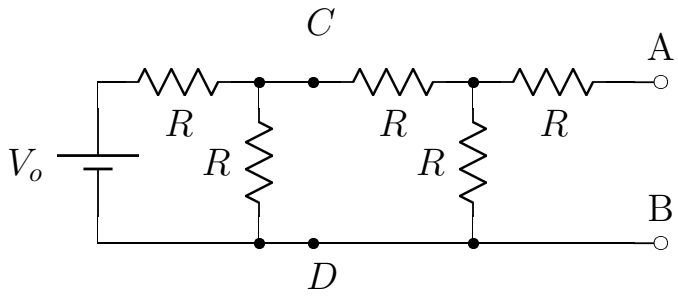

Figure 1.26: A sample circuit with two output terminals, A and B.

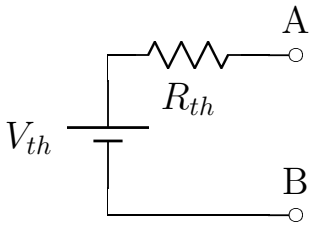

Figure 1.27: The Thèvenin equivalent circuit of the the circuit in Figure 1.26 when one looks to the left across the terminals A and B.

In order to determine R_{th} of the circuit, we need to determine the current that would flow through a short circuit between A and B. Again, using the techniques developed earlier, we find that

$$I_N = \frac{V_o}{8R}.$$

This then leads to

$$R_{th} = \frac{V_{th}}{I_N}$$
$$R_{th} = \frac{8}{5}R.$$

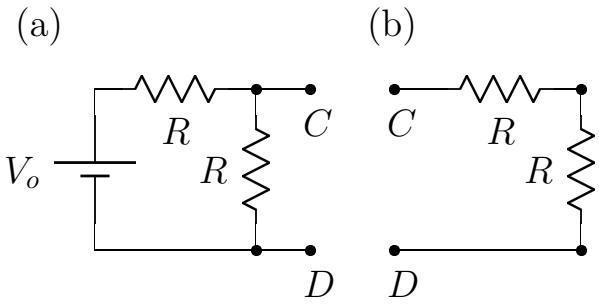

Figure 1.28: If we cut the circuit from Figure 1.26 and cut between the points C and D, we produce two new circuits. If we look left across the points C and D, we see the circuit shown in (a). If we look to the right across the points, we see the circuit shown in (b). Note that we have removed the resistor on the branch of the original circuit going to the point A as it will not affect the behavior of the new circuits.

It is also possible to cut the circuit at some other point and examine the Thèvenin equivalents. To do this, let us cut the circuit at the points C and D. If we do so, we can replace the two parts of the

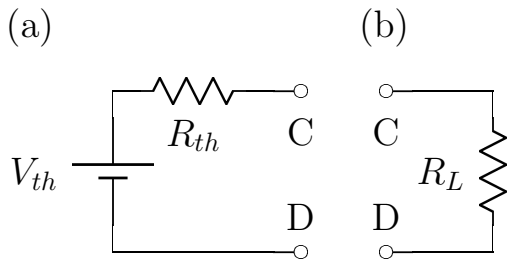

Figure 1.29: The equivalent circuits for the two circuits in Figure 1.28.

circuit with the equivalents shown in Figure 1.28. Note that in the circuit 1.28(b), we have dropped the resistor that went to the point A. No current will be able to flow through this resistor, so it cannot affect the behavior of the circuit. Figure 1.29 shows the equivalents for the two new circuits.

Let us now determine the parameters of the two equivalent circuits. The easy case is for Figure 1.28(b). Its equivalent is just a single resistance that is equivalent to R and R in series. This gives us that

$$R_L = 2R\,.$$

For part (a) of the circuit, the voltage across C and D is just $V_o/2$. This gives

$$V_{th} = \frac{V_o}{2}\,.$$

If we now short the circuit, we find that the Norton current is

$$I_N = \frac{V_o}{R}\,.$$

This then yields the Thèvenin resistance

$$\begin{aligned} R_{th} &= \frac{V_{th}}{I_N} \\ R_{th} &= \frac{1}{2}R\,. \end{aligned}$$

1.6 The Voltage Divider

Much of what we do in electronics depends on understanding the behavior of a very simple circuit known as a *voltage divider*. This circuit is shown in Figure 1.30. A voltage divider has an input voltage, V_{in}, and an output voltage, V_{out}. We can determine the relationship between them as follows. The current flowing through the two resistors is given as:

$$I = \frac{V_{in}}{R_1 + R_2}.$$

The voltage across the second resistor is $I \cdot R_2$, which yields that:

$$V_{out} = V_{in} \cdot \frac{R_2}{R_1 + R_2}, \tag{1.29}$$

or in a slightly different form:

$$\frac{V_{out}}{V_{in}} = \frac{R_2}{R_1 + R_2}. \tag{1.30}$$

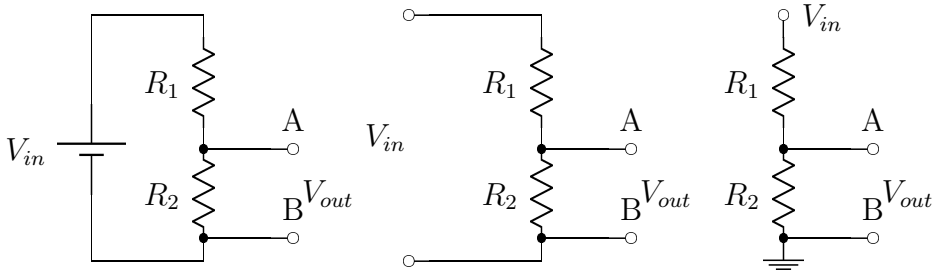

Figure 1.30: The basic voltage divider circuit. All three circuits are identical, they just represent different ways of drawing the same thing and all will be used.

In using a voltage divider, we would like to be able to consider it an ideal voltage source of voltage V_{out}.

What is shown in Figure 1.30 is an open-circuit voltage divider, or alternatively a voltage divider with an infinite load resistance. More typically, we place some finite load resistance R_L across the output of a voltage divider as shown in Figure 1.31.

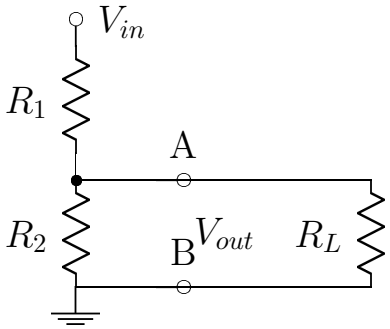

Figure 1.31: A voltage divider with a load resistance.

We now want to ask what the impact of this load resistance is on our assumption of an ideal voltage source. It is relatively straightforward to determine what the new value of V_{out} is. The circuit now behaves as if it is a voltage divider with R_2 replaced by the parallel combination of R_2 and R_L; let us call this R_2'.

$$R_2' = \frac{R_2 R_L}{R_2 + R_L} \tag{1.31}$$

In fact, V_{out} will no longer be the same as before the load was connected. However, if R_L is much larger than R_2, then we find that:

$$\begin{aligned} R_2' &\approx \frac{R_2 R_L}{R_L} \\ R_2' &\approx R_2. \end{aligned}$$

In such a limit, the output voltage will not be significantly changed. However, we should be a bit more careful in analyzing the behavior of our circuit. If we put equation 1.31 into equation 1.30, we get the following:

$$\begin{aligned} \frac{V_L}{V_{in}} &= \frac{R_2'}{R_1 + R_2'} \\ \frac{V_L}{V_{in}} &= \frac{R_2 \cdot R_L}{R_1 \cdot R_2 + R_1 \cdot R_L + R_2 \cdot R_L} \end{aligned}$$

$$\frac{V_L}{V_{in}} = \frac{R_2}{R_1+R_2}\cdot\frac{1}{1+\frac{R_1R_2}{R_1+R_2}\frac{1}{R_L}}$$

This can be simplified to yield equation 1.32, where $R_1 \parallel R_2$ is the equivalent resistance of R_1 in parallel with R_2. If R_L is much larger than this parallel combination, then the second factor becomes approximately one.

$$\frac{V_L}{V_{in}} = \frac{R_2}{R_1+R_2}\cdot\frac{1}{1+\frac{R_1\parallel R_2}{R_L}} \tag{1.32}$$

In the limit where the load resistance, R_L, is much larger than $R_1 \parallel R_2$, a voltage divider will behave as an ideal voltage source.

1.6.1 The Thèvenin Equivalent of the Voltage Divider

Let us now examine the Thèvenin equivalent of our voltage divider. To determine V_{th}, we find the open-circuit voltage. This is just V_{out} from equation 1.30.

$$V_{th} = V_{in}\frac{R_2}{R_1+R_2}$$

We can find the Norton current by determining the short-circuit current. This yields that

$$I_N = \frac{V_{in}}{R_1},$$

and putting these two together, we find that

$$\begin{aligned} R_{th} &= \frac{V_{th}}{I_N} \\ R_{th} &= \left(V_{in}\frac{R_2}{R_1+R_2}\right)/\left(\frac{V_{in}}{R_1}\right) \\ R_{th} &= \frac{R_1\cdot R_2}{R_1+R_2} \end{aligned}$$

$$R_{th} = R_1 \parallel R_2. \tag{1.33}$$

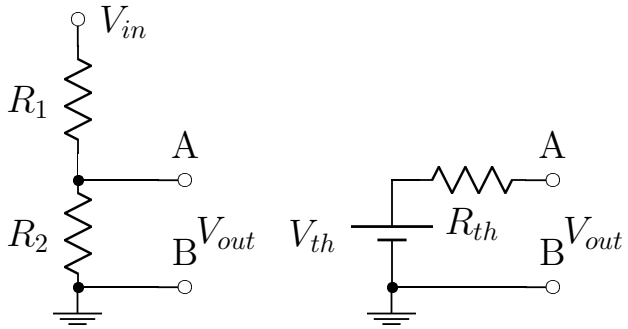

Figure 1.32: The Thèvenin equivalent circuit (right) of the voltage divider circuit (left). As described in the text, $V_{th} = V_{in}\frac{R_2}{R_1+R_2}$ and $R_{th} = R_1 \parallel R_2$.

1.6.2 Power Dissipation in a Voltage Divider

While a voltage divider can be very convenient in allowing us to set voltages, a useful question to ask is how much power is dissipated by the device. The power dissipated in a resistor is $P = I^2R$. In our voltage divider with no load attached to it, the current through both resistors is $I = V_{in}/(R_1 + R_2)$. This means that the power dissipated is:

$$\begin{aligned} P &= R_1 I^2 + R_2 I^2 \\ P &= (R_1 + R_2) \cdot \left(\frac{V_{in}}{R_1 + R_2} \right)^2 \\ P &= \frac{V_{in}^2}{R_1 + R_2}. \end{aligned} \tag{1.34}$$

For a given input voltage, the larger the resistors in the divider are, the smaller the power dissipated is. However, the normal operating state of a voltage divider is to have a load resistor connected to it, where

$$R_L \gg R_1 \,||\, R_2 \,.$$

This latter condition tends to push down the values of R_1 and R_2, and thereby increase the power dissipated in the divider.

We now want to look at the power delivered to the load. If the circuit is well built, very little current is diverted into the load, so $I_L \ll I$. We can be a bit more precise than this. We know that $R_L \gg R_1 \,||\, R_2$. We can write $R_L = \alpha R_1 R_2/(R_1 + R_2)$, where $\alpha \gg 1$. From this we can find that the current through the load is:

$$\begin{aligned} I_L &\approx V_{in} \frac{R_2}{R_1 + R_2} \frac{1}{R_L} \\ I_L &\approx \frac{V_{in}}{\alpha R_1}. \end{aligned}$$

Thus we see that the power dissipated in the load is:

$$\begin{aligned} P_{load} &= R_L I_L^2 \\ P_{load} &= \left(\alpha \frac{R_1 R_2}{R_1 + R_2} \right) \left(\frac{V_{in}}{\alpha R_1} \right)^2 \\ P_{load} &= \left(\frac{V_{in}^2}{R_1 + R_2} \right) \left(\frac{1}{\alpha} \right) \left(\frac{R_2}{R_1} \right) \\ P_{load} &= P_\circ \cdot \left(\frac{1}{\alpha} \right) \cdot \left(\frac{R_2}{R_1} \right) \end{aligned}$$

where $P_\circ$ is the nominal power in the divider as in equation 1.34. The second term in the latter equation gives the ratio of the power dissipated in the load to that in the circuit. For most normal divider operations, R_1 and R_2 are similar in size. This means that the power dissipated in the load is much smaller than what is dissipated in the circuit.

In the limit where $R_2 \gg R_1$, it is possible that the above statement is not true. In this particular case, almost all of the voltage drop is across R_2, and we can approximate $R_1 = 0$. In this case, R_L can have any value and the voltage drop across it will still be V_{in}. The power dissipated in R_2 and R_L can be written as:

$$\begin{aligned} P_2 &= \frac{V_{in}^2}{R_2} \\ P_L &= \frac{V_{in}^2}{R_L} \end{aligned}$$

These can have any ratio we want, just by choosing the appropriate R_2 and R_L. However, this is a very *strange* case. With $R_1 = 0$, there is really no voltage divider. This is far from the normal operating condition for a divider, but it goes to show that one has to be careful about making general statements.

Under normal operating conditions for a voltage divider, we find that the majority of the power is dissipated in the circuit and not delivered to the load. While voltage dividers may deliver a stable voltage, they are not a very good power supply.

Problems

1. Justify to yourself: an *electron volt* (eV) is the potential energy difference per volt per electron rather than per Coulomb: eV's have energy units whereas V is called a *potential*, not a *potential energy*; when an electron is accelerated through a potential difference of 1 Volt (in vacuum), it picks up an energy of $1\text{eV} = e \times 1\,\text{V} = 1.602 \times 10^{-19}\text{J}$.

2. In typical metals, you should know that the density of conduction electrons is $n \sim 10^{23}\text{cm}^{-3}$ (close to Avogadro's number). For $I = 1$ Amp in a wire of diameter 1 mm, what is v, the average speed of the electrons?

3. Consider the circuit shown in Figure 1.33 below. In terms of R_1, R_2 and V_o, what is the voltage between A and B?

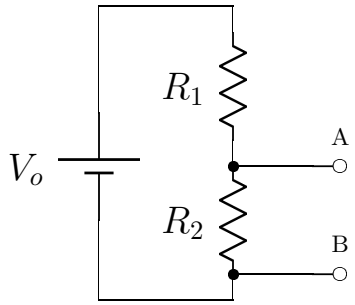

Figure 1.33: The circuit for problem 3.

4. Explicitly demontrate that equation 1.13 is true for two capacitors in parallel.

5. Explicitly demontrate that equation 1.14 is true for two capacitors in series.

6. For the circuit in Figure 1.33 above, what is the ratio of $R_2 : R_1$ such that the voltage across A and B is $\frac{1}{2}V_o$? What is the ratio of $R_2 : R_1$ such that the voltage across A and B is $\frac{1}{3}V_o$? $\frac{1}{10}V_o$?

7. Consider the circuit shown in Figure 1.34 which is built using six identical resistors. Use Kirchoff's rules to solve for the current flowing through each resistor in the circuit. What is the voltage drop across each resistor in the circuit? (To facilitate labeling, use a notation such as I_{pq} and V_{pq} where these represent the current flowing from point p to point q, and the voltage drop in going from point p to point q.)

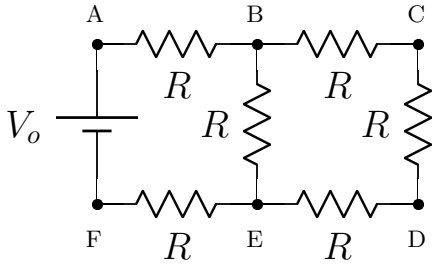

Figure 1.34: The circuit for problem 7.

8. We now attach two output terminals to the circuit from problem 7. The resulting circuit is shown in in Figure 1.35. **(a)** What is the voltage between the terminals G and H? **(b)** What current flows from G to H? **(c)** If we connect a wire from G to H, what current flows through the wire and what is the voltage between G and H?

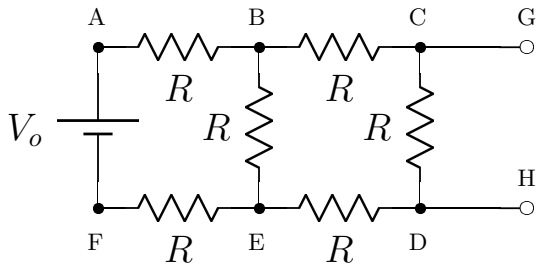

Figure 1.35: The circuit for problem 8.

9. Replace the circuit from problem 8 with the simpler one shown in Figure 1.36. V_{th} is a voltage source and R_{th} is a new resistance. What are the values of V_{th} and R_{th} such that you get the same answer to the current and voltage questions as in problem 8?

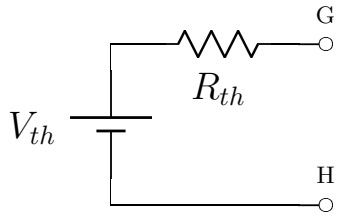

Figure 1.36: The circuit for problem 9.

10. Replace the circuit in Figure 1.35 with the one shown in Figure 1.37 where I_N is a current source that always delivers I_N amps of current and R_N is a new resistance. What are the values of I_N and R_N such that you get the same answer to the current and voltage between G and H as in problem 8?

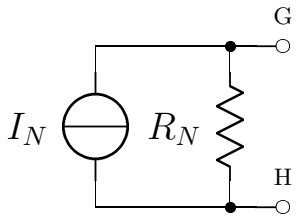

Figure 1.37: The circuit for problem 10.

11. We can generate I-V curves for these circuits by placing a variable resistor, R_V, between G and H. As we vary the value for R_V from 0 to ∞, we map a set of (V, I) points. As stated previously, this curve should be linear. Pick 4-5 values of R_V and evaluate I and V between the G and H terminals. You can do this for either the Thèvenin or the Norton equivalent circuit. Does it matter which one you choose? Plot these values as a graph of I versus V and show that it is indeed linear. What is the slope of the line?

12. Consider the two circuits shown in Figure 1.15 for measuring the voltage across R and the current through R. Assume that the internal resistance of the voltmeter is $R_v = 100\,R$ and that the internal resistance of the ammeter is $R_a = 0.01\,R$. In terms of V_o and R, what are the measured voltages and currents in each of the two circuits?

13. Show that the open-circuit voltage of a circuit is the Thèvenin voltage of the circuit.

14. Show that the short-circuit current is the Norton current of the circuit.

15. Show that equation 1.28 is correct.

16. Determine the parameters of the Thèvenin equivalent for the circuit shown in Figure 1.38.

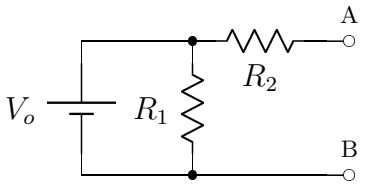

Figure 1.38: The circuit for problem 16.

17. You are given a *black-box* device with two output terminals. You are asked to characterize the behavior of this device, so you proceed to measure an $I-V$ curve for the circuit. You measure the following two (I,V) points: $(2.0I_b, V_b)$ and $(I_b, 5.0V_b)$. (**a**) Sketch the Thèvenin and Norton equivalent circuits for the black box. (**b**) Accurately sketch the $I-V$ curve for the black box. Be sure to carefully label your plot. (**c**) From your graph, determine the Thèvenin voltage, V_{th}, the Thèvenin resistance, R_{th}, the Norton current, I_N and the Norton resistance, R_N for your black box. Express your answer in terms of V_b and I_b. (**d**) A load resistance of value $R_L = R_{th}$ (equal to the Thèvenin resistance) is connected across the output terminals of your black box. In terms of V_{th} and I_N, what is the voltage across R_L and the current through R_L?

18. You have an *R2R ladder* with two output terminals as shown in Figure 1.39. (**a**) Sketch the Thèvenin equivalent of the R2R ladder as seen when looking into the output terminals to the right of the circuit. (**b**) Sketch the IV curve for your R2R ladder as seen from the output terminals. Label your axes in terms of V_{th} and R_{th}. (**c**) What is the Thèvenin equivalent voltage, V_{th}, and the Thèvenin resistance, R_{th}, for the R2R ladder? (Hint: You may find it useful to calculate the voltage at the node between the $5\,k\Omega$ resistors first.)

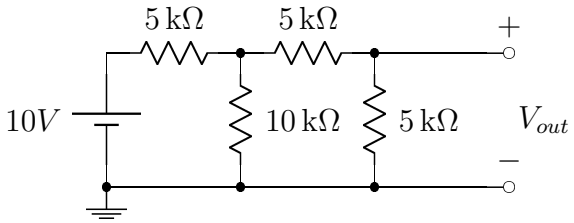

Figure 1.39: The circuit for problem 18.

19. In Figure 1.31, let $R_1 = R_2 = R$. If $R_L = 2 \cdot R$, by what fraction does V_{out} differ from its open-circuit value? If $R_L = 10 \cdot R$? $R_L = 100 \cdot R$?

20. You are asked to build a voltage divider that will take a $12\,V$ input voltage and deliver $4\,V$ into an $8\,k\Omega$ load. Pick good values of R_1 and R_2 for this circuit.

21. You are asked to more carefully design the circuit above such that the power dissipated in the voltage divider is a minimum and the output voltage sags by no more than $0.2\,V$ when the load is connected. What values of R_1 and R_2 should you choose?

22. You are given the circuit shown in Figure 1.40 with unknown values for R_1, R_2, V_1 and V_2. (**a**) Sketch the form of the Thèvenin equivalent circuit as viewed across A and B for the circuit. (**b**) Sketch the form of the Norton equivalent circuit as viewed across A and B for the circuit. (**c**) Sketch the Thèvenin equivalent circuit as viewed across D and E when looking toward the right.

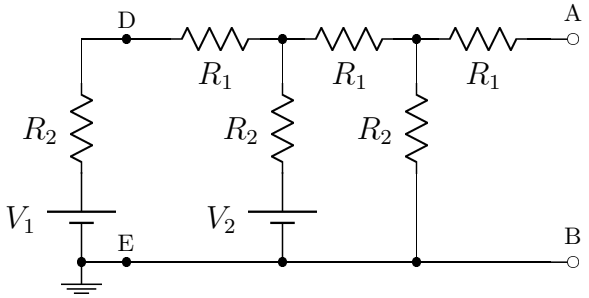

Figure 1.40: The circuit for problem 22.

23. A black-box voltage source has a Thèvenin voltage of $V_{th} = 10.0\,V$ and a Thèvenin resistance of $R_{th} = 50.0\,\Omega$. To the output of this box, you will attach two resistors in series, R_1,R_2.
You have two design goals for your circuit:
• The voltage across R_2 should be very close to $\frac{1}{3}V_{th}$.
• The power dissipated in the external resistors should be kept low (under $50\,mW$).
(**a**) Accurately plot the I-V curve for the black box voltage source alone (R_1 and R_2 are not attached). (**b**)Choose reasonable values for R_1 and R_2 such that the circuit will operate as desired. (Justify your choice based on physics reasons and the design goals.) (**c**) The complete circuit that you built (with output taken across R_2) can be replaced by a Thèvenin equivalent circuit with a new Thèvenin voltage of V'_{th} and a new Thèvenin resistance of R'_{th}. What are the numerical values of V'_{th} and R'_{th}, based on the circuit that you designed in part **b**? (**d**) Your lab partner accidentally attaches a load with very small R_L ($< 50.\,\Omega$) to your completed circuit. Will your circuit satisfy the design goals? Briefly explain.

24. Consider the circuit shown in Figure 1.41, built from two ideal voltage sources, V_o, and four identical resistors, R. The output from the circuit is measured between the connectors at A and B. Answer the following questions in terms of V_o and R. (**a**) What is the open-circuit voltage measured between the connectors A and B, V_{AB}? (**b**) A wire is now connected from A to B and some current, I_{AB}, flows through the wire. What are I_{AB} and the voltage, V_{AB}? (**c**) Draw the Thèvenin equivalent circuit as seen from the connectors A and B for the original circuit. Determine both V_{th} and R_{th}. (**d**) Sketch an I–V curve for the original circuit as seen from the connectors A and B. Carefully label your axes and indicate what the voltage and current are at the intercepts. (**e**)Some unknown load is connected between A and B and you measure a voltage of $V_{AB} = \frac{1}{2}V_{th}$ across the load. What is the resistance of the load?

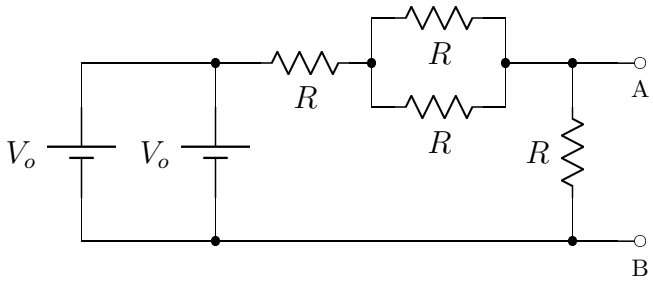

Figure 1.41: The circuit for problem 24.

25. Consider the circuit shown in Figure —reffig:prob:1:25 which is built out of an ideal voltage source, V_0, and six identical resistors of resistance R. They are labeled as R_1, R_2, R_3, R_4, R_5 and R_6 in the circuit below so that they can be distinguished. They all have the same resistance R. Answer the following questions in terms of V_0 and R. (**a**) What voltage difference will be measured between

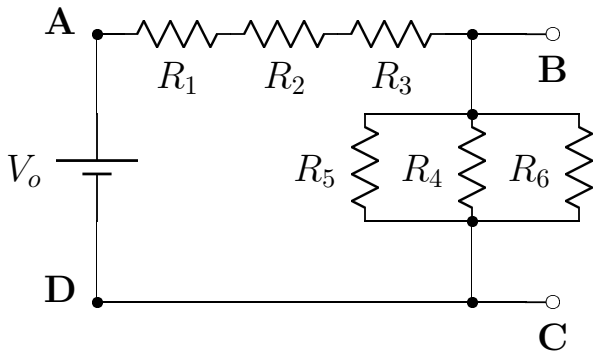

Figure 1.42: The circuit for problem 25.

the points **A** and **D** ($V_{AD} = V_A - V_D$)? (**a**) What is the voltage difference between the points **C** and **A** ($V_{CA} = V_C - V_A$)? (**c**) What voltage difference will be measured between the points **B** and **C** ($V_{BC} = V_B - V_C$)? (**d**) What current will flow through the resistor R_5? (**e**) Sketch the Thèvenin equivalent circuit as seen looking into **B** and **C** towards the left. What are the Thèvenin voltage and Thèvenin resistance in terms of V_0 and R? (**f**) If we remove the voltage source and look into the circuit through **A** and **D** towards the right, what are the Thèvenin voltage and Thèvenin resistance that we see?

Chapter 2

Alternating-Current Circuits

2.1 Introduction

We are now going to look at circuits in which the voltage and currents are time-dependent, rather than constant. We will use a notation in which V and I represent voltages and currents that are constant in time, while $v(t)$ and $i(t)$ will represent time-dependent voltages and currents. We will start by examining the case where we have:

$$v(t) = v_o \cos(\omega t + \phi_v) \tag{2.1}$$
$$i(t) = i_o \cos(\omega t + \phi_i)\,. \tag{2.2}$$

We will only be considering circuits in which the frequency, $f = \omega/2\pi$, is not changed by the circuit. Both the voltage and current will have the same frequency, but as indicated above, they may not have the same phase (ϕ_v may not be equal to ϕ_i). We will also start by looking at only a single frequency, but will note that it is possible to build up any time-dependent voltage as a sum of cosines of different frequencies (Fourier Analysis). We also note at this point that in doing the mathematical circuit analysis, we typically use the angular frequency, ω. In doing lab work where one measures frequencies, the quantity f is relevant. It is important to remember the factor of 2π that connects these two quantities. The quantity f is measured in units of cycles per second, or *Hertz* (Hz). The angular frequency, ω, is measured in units of radians per second. The radians are typically supressed and we simply measure ω in inverse seconds, s^{-1}.

The schematic in Figure 2.1 is meant to represent essentially any electronic circuit. We consider a signal, $v_i(t)$, (which may come from a strain gage, an NMR pick-up coil, a photon or particle detector device, the light detector in a CD player, ...) as the input to the circuit. The circuit, represented by the box, performs some function on $v_i(t)$ and sends out a voltage, $v_o(t)$, which is appropriate for driving some other device (the input card on a computer, a meter, audio speakers, ...). The output voltage may be an amplified or attenuated form of the input or may measure some other property of the input signal, but as mentioned above, the output signal and the input signal both have the same frequency.

The circuits that we study are referred to as *linear systems*. We can think of this in simple terms relative to the black-box circuit in Figure 2.1. Let us assume that an input voltage of v_i^a produces an output voltage of v_o^a and that an input voltage of v_i^b produces an output voltage of v_o^b. If the black-box circuit is linear, then an input voltage that is some linear combination of the two previous input voltages,

$$v_{in}^c = \alpha v_i^a + \beta v_i^b,$$

where α and β are constants, it will produce an output voltage that is the same linear combination of the two output voltages:

$$v_o^c = \alpha v_o^a + \beta v_o^b,$$

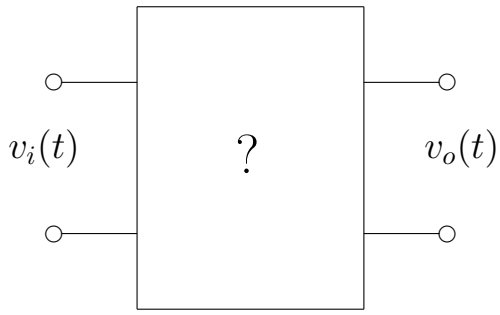

Figure 2.1: A black-box circuit with two input terminals and two output terminals.

A more mathematical definition of this is that a linear system (mechanical, electrical, or other) is one whose response to external excitation can be described by a *linear integral-differential equation*. In other words,

$$\begin{aligned} v_o &= a_0 v_i + a_1 \dot{v} + i + a_2 \ddot{v} + i + ... \\ &+ b_0 + b_1 \int dt' v_i + b_2 \int dt'' \int dt' v_i + ... \end{aligned} \tag{2.3}$$

Many systems in nature are approximately linear and, fortunately, the analysis of their response is relatively straightforward. In dealing with linear electrical circuits, we will learn analysis tools that can be applied to many other circumstances. Electrical circuits are a particularly nice context in which to learn this analysis because one can (almost) always build a real circuit and demonstrate the results of the analysis.

If we look at what equation 2.3 can do to a sinusoidal input signal of frequency ω, it is straightforward to show that all of the terms on the right-hand side will be either sines or cosines containing the **same** frequency, ω. The sum itself can shift the timing of the peaks and valleys of the output relative to the input, (known as shifting the phase), or it can change the amplitude of the signal. It cannot change the frequency.

We will study several types of circuits which perform functions on input signals and produce appropriate outputs. AC circuit analysis provides the means to analyze and characterize a wide variety of types of circuits (but not all types – notably not digital circuits). As we proceed through the book, detailed examples will be worked out which illustrates the considerations necessary in building circuits and using electrical instruments.

2.2 Complex Notation

It is possible to describe time-dependent electronic circuits in terms of oscillatory functions with magnitudes and phases. However, it is much simpler to do do this using *complex* notation. We treat quantities as complex, with real and imaginary parts. In such a notation, operations that might involve multiple trigonometric identities and many lines of not particularly clear algebra can be replaced with simple multiplications and divisions of complex numbers. Before proceeding with time-varying voltages, we therefore pause for a brief review of complex numbers.

A complex number, z, can be expressed as a combination of real and imaginary parts, where the imaginary part is multiplied by $\sqrt{-1}$. In mathematics and physics, $\sqrt{-1}$ is represented by the symbol i. Unfortunately, in electronics, the symbol i is reserved for a time-varying current. As such, we need to introduce an alternative notation for $\sqrt{-1}$. In electronics, we define $j = \sqrt{-1}$. So rather than writing $z = a + ib$ as we normally would, we write $z = a + jb$.

If we have a complex number, $z = a + jb$, we can represent it as a point in the complex plane as shown in Figure 2.2. We can also express this complex number in polar coordinates as a magnitude,

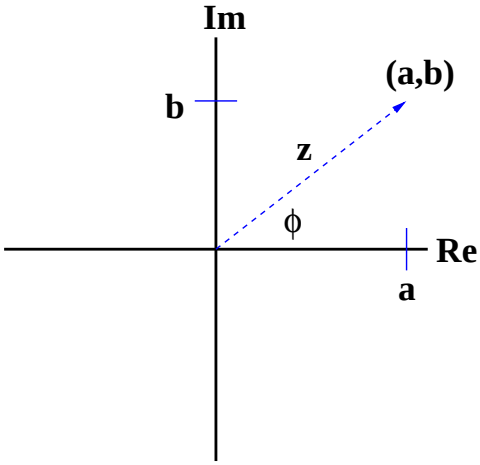

Figure 2.2: A point (a, b) in the complex plane.

$| z |$, and phase, ϕ, where

$$| z | = \sqrt{a^2 + b^2} \tag{2.4}$$

$$\tan(\phi) = \frac{b}{a}. \tag{2.5}$$

We can express the real and imaginary parts of z as follows:

$$Re(z) = | z | \cos\phi \tag{2.6}$$

$$Im(z) = | z | \sin\phi \tag{2.7}$$

An important operation with complex numbers is the complex conjugate. If $z = a + jb$, then the complex conjugate of z is $z^* = a - jb$. The complex conjugate replaces j with $-j$ in a complex number. The product of z and z^* is then

$$\begin{aligned} z \cdot z^* &= (a + jb) \cdot (a - jb) \\ z \cdot z^* &= \left(a^2 + b^2\right) + (-abj + abj)) \\ z \cdot z^* &= \left(a^2 + b^2\right) \\ z \cdot z^* &= | z |^2 \end{aligned} \tag{2.8}$$

If we return to equations 2.6 and 2.7, then we can write the complex number z as:

$$z = | z | (\cos\phi + j \sin\phi) . \tag{2.9}$$

At this point, we will make a small aside to consider the power series expansion of the function $e^{j \cdot x}$.

$$e^{j \cdot x} = 1 + jx + \frac{(jx)^2}{2!} + \frac{(jx)^3}{3!} + \frac{(jx)^4}{4!} + \frac{(jx)^5}{5!} + \cdots$$

We can slightly rearrange the terms in this equation such that we collect all the real parts together, and all the imaginary parts together. This gives is the expression as follows:

$$e^{j \cdot x} = \left(1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots\right) + j\left(x - \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots\right) .$$

Looking carefully at this, we note that the real part is just the expansion for the cosine function, while the imaginary part is the expansion for the sin function. This allows us to rewrite the above expression as:

$$e^{j \cdot x} = \cos(x) + j \sin(x) \; . \tag{2.10}$$

From this, we can rewrite equation 2.9 in a more compact form:

$$z = \mid z \mid e^{j\phi} \tag{2.11}$$

which is known as the *polar form* of a complex number. The complex conjugate of this is just

$$z^* = \mid z \mid e^{-j\phi}$$

where as before, we have simply replaced j with $-j$. We can also get the magnitude of the polar form as we did with the Cartesian form.

$$\begin{aligned} zz^* &= \mid z \mid^2 e^{j\phi} e^{-j\phi} \\ &= \mid z \mid^2 \end{aligned}$$

The polar form makes multiplication and division much easier than the Cartesian form does. If we have two complex numbers, $z_1 = r_1 e^{j\phi_1}$ and $z_2 = r_2 e^{j\phi_2}$, their product is given as follows.

$$\begin{aligned} z_1 \cdot z_2 &= r_1 e^{j\phi_1} \cdot r_2 e^{j\phi_2} \\ &= r_1 \cdot r_2 e^{j(\phi_1 + \phi_2)} \end{aligned} \tag{2.12}$$

The product has a magnitude that is the product of the magnitudes of the individual complex numbers, and a phase that is the sum of the phases of the individual numbers. Similarly, the ratio of the two numbers is given as:

$$\frac{z_1}{z_2} = \frac{r_1}{r_2} e^{j(\theta_1 - \theta_2)} \tag{2.13}$$

where we take the ratio of the magnitudes and the difference of the phases.

If we now consider a pair of complex numbers whose magnitudes are both one, $z_1 = e^{j\phi_1}$ and $z_2 = e^{j\phi_2}$, then the the multiplication or division of these numbers results in simply a rotation in the complex plane.

While polar form of complex numbers greatly facilitates multiplication and division, addition and subtraction are clearly easier to carry out with the Cartesian form.

Another convenient features of the exponential representation of complex numbers is the way it factors. In electronics, we will very often need to take derivatives and integrals of time dependent voltages. As an example of how the exponential representation makes this easy, we consider a traveling wave. It can be written in either of the following forms:

$$e^{j(kx - \omega t)} = e^{jkx} e^{-j\omega t}. \tag{2.14}$$

In performing a spatial derivative, the time-dependent factor is just a multiplicative constant:

$$\begin{aligned} \frac{\partial}{\partial x} e^{j(kx-\omega t)} &= e^{-j\omega t} \frac{\partial}{\partial x} e^{jkx} \\ \frac{\partial}{\partial x} e^{j(kx-\omega t)} &= e^{-j\omega t} \left[jk e^{jkx} \right] \\ \frac{\partial}{\partial x} e^{j(kx-\omega t)} &= jk e^{j(kx-\omega t)}. \end{aligned} \tag{2.15}$$

In short, taking a spatial derivative is the same as multiplying by jk. You can show that a time derivative is the same as multiplying by $-j\omega$. Furthermore, integration corresponds to multiplying by the inverse of these quantities.

2.3 Characterizing Time-dependent Voltages

The language that we use for time-dependent circuits is very similar to that used in describing time-independent circuits. We have voltages and currents as before. However, we generalize resistance to something which we call *impedance*. While impedance has the same units as resistance, namely ohms, it also has the ability to change the phase between the voltage and the current. We typically use the symbol Z for impedance. In the next section we will see that impedances in series and parallel combine in the same way as resistors.

If we have a time-dependent voltage as given in equation 2.1, we can characterize it with several measures. The quantity v_o is referred to as the *amplitude* of the voltage. We also often consider the *peak-to-peak* voltage, $2v_o$ in our example. We might also consider the time average of the voltage, but this is zero. On the other hand, $v^2(t)$ averaged over a full cycle is both non-zero and a useful quantity, as it is related to the power dissipated. We go a bit beyond this and define the *root mean square* voltage to be the square root of this average:

$$V_{RMS} = \sqrt{< v^2(t) >} \tag{2.16}$$

For $v(t)$ given as in equation 2.1, we can find V_{RMS} by averaging over an integer number of cycles, $(T = N \cdot \frac{2\pi}{\omega})$.

$$\begin{aligned}
V_{RMS} &= \left[\frac{\int_0^T v_o^2 \cos^2(\omega t + \phi_v) dt}{T}\right]^{\frac{1}{2}} \\
V_{RMS} &= \left[\frac{v_o^2}{T}\left(\frac{1}{2\omega}\sin(\omega t + \phi_v)\cos(\omega t + \phi_v) + \frac{1}{2}t\right)\Big|_0^T\right]^{\frac{1}{2}} \\
V_{RMS} &= \left[\frac{T v_o^2}{2T}\right]^{\frac{1}{2}} \\
V_{RMS} &= \frac{v_o}{\sqrt{2}}
\end{aligned}$$

If we now return to a time-dependent voltage as in equation 2.1, we can write this slightly differently as:

$$v(t) = v_o Re\left(e^{j\omega t} e^{j\phi_v}\right). \tag{2.17}$$

For convenience, we will omit the Re from this expression, but note that when we measure a voltage, we look only at the real part at that particular instant in time. More generally, we write the voltage and current as complex functions where it is implied that when we make a measurement we will only see the real part of the expression. The voltage and current from equation 2.1 and 2.2 can now be written as

$$v(t) = v_o e^{j\omega t} e^{j\phi_v} = v_o e^{j(\omega t + \phi_v)} \tag{2.18}$$

$$i(t) = i_o e^{j\omega t} e^{j\phi_i} = i_o e^{j(\omega t + \phi_i)} \tag{2.19}$$

We can now generalize Ohm's law for $V = IR$ to $v(t) = i(t)Z$ where the complex impedance, Z, contains the relative phase as well as the ratio of the amplitudes of $v(t)$ and $i(t)$. The impedance, Z, can now be written as:

$$Z(\omega) = \frac{v(t)}{i(t)} \tag{2.20}$$

$$Z(\omega) = \frac{v_o}{i_o} e^{j(\phi_v - \phi_i)}. \tag{2.21}$$

2.3.1 Power Dissipation in AC Circuits

In determining the power in an AC circuit, we start with the expression for a DC circuit:

$$P = V \cdot I \,. \tag{2.22}$$

Because we have a time-varying signal, we need to average it over some time. We typically choose power per cycle, which implies an average over an integer number of periods.

$$P_{avg} = \frac{1}{T} \cdot \int_{t=0}^{T} v(t) \cdot i(t)\, dt \tag{2.23}$$

In order to proceed, we need to look closely at what we actually put into equation 2.23 for $v(t)$ and $i(t)$. To do this, we note that power is a measured quantity. As such , it should be computed using the measured voltage and current. Because of this, we want to put the measured, or real parts of the voltage and current into equation 2.23. If we are using the complex exponential form for current and voltage, then we can rewrite equation 2.23 as:

$$P_{avg} = \frac{1}{T} \int_{t=0}^{T} Re(v)\, Re(i)\, dt \,. \tag{2.24}$$

Example: If the voltage and current are in phase,

$$\begin{aligned} v(t) &= v_o e^{j\frac{2\pi}{T}t} \\ i(t) &= i_o e^{j\left(\frac{2\pi}{T}t\right)} \end{aligned}$$

then the average power is given by

$$P_{avg} = \frac{1}{T} \int_{t=0}^{T} Re\left[v_o e^{j\frac{2\pi}{T}t}\right] \, , Re\left[i_o e^{j\left(\frac{2\pi}{T}t\right)}\right] dt \,.$$

We can simplify this to give that

$$\begin{aligned} P_{avg} &= \frac{v_0 i_0}{T} \cdot \int_{t=0}^{T} \cos\left(2\pi t/T\right) \cos\left(2\pi t/T\right) dt \\ P_{avg} &= \frac{v_0 i_0}{T} \cdot \int_{t=0}^{T} \cos^2\left(2\pi t/T\right) dt \,. \end{aligned}$$

The average of $\cos^2$ over one period is just $\frac{1}{2}$, so we find that the average power $P_{avg} = \frac{1}{2} v_o i_o$.

We can now ask what would have happened if the power were $Re\left[v(t)i(t)\right]$ rather than equation 2.24? If we write this as P'_{avg}, this would give us the following expressions for the power.

$$\begin{aligned} P'_{avg} &= \frac{1}{T} \int_{t=0}^{T} Re\left[v_o e^{j\frac{2\pi}{T}t} i_o e^{j\left(\frac{2\pi}{T}t\right)}\right] dt \\ P'_{avg} &= \frac{v_0 i_0}{T} \int_{t=0}^{T} \left[\cos^2(\frac{2\pi}{T}t) - \sin^2(\frac{2\pi}{T}t)\right] dt \\ P'_{avg} &= 0 \end{aligned}$$

Clearly, the average power is not zero! The crucial point to remember here is that we have to take the real part of both the voltage and the current when evaluating power.

Example: Let us consider a case in which the current and voltage are out of phase by 90°:

$$\begin{aligned} v(t) &= v_o e^{j\frac{2\pi}{T}t} \\ i(t) &= i_o e^{j\left(\frac{2\pi}{T}t-\frac{\pi}{2}\right)} \end{aligned}$$

We can substitute these into equation 2.24 to give that

$$P_{avg} = \frac{1}{T}\int_{t=0}^{T} Re\left[v_o e^{j\frac{2\pi}{T}t}\right]\, Re\left[i_o e^{j\left(\frac{2\pi}{T}t-\frac{\pi}{2}\right)}\right]\, dt$$

and then taking the real parts of both the current an voltage, we find that:

$$P_{avg} = \frac{v_0 i_0}{T}\cdot\int_{t=0}^{T}\cos\left(2\pi t/T\right)\,\sin\left(2\pi t/T\right)\,dt\,.$$

or the average power $P_{avg}=0$. If the voltage and current are 90° out of phase, the average dissipated power per cycle is zero.

2.3.2 Time-dependent Circuits

As mentioned earlier, impedances in series and parallel add just like resistors in series and parallel. In showing this, we will carry out a derivation that only applies to sinusoidal signals. However, we will later show that any periodic signal can be built up out of sums of sines and cosines. This is all we need to be able to generalize the results to any signal.

For a series combination, the same instantaneous current runs through both elements, therefore the current $i(t)$ is the same. Thus,

$$\begin{aligned} v(t) &= i(t)\cdot Z_1 + i(t)\cdot Z_2 \\ v(t) &= i(t)\left(Z_1+Z_2\right) \\ v(t) &= i(t)Z_{series} \end{aligned}$$

Therefore we find that

$$Z_{series} = Z_1 + Z_2. \tag{2.25}$$

Note that addition of complex impedances is carried out by separately adding their real and imaginary parts. If the impedances are in polar form, we must express them as real and imaginary parts before adding. We cannot just add amplitudes or phase angles!

For a parallel combination, the instantaneous voltage across the elements is the same, therefore the total current i_o is

$$\begin{aligned} i(t) &= \frac{v(t)}{Z_1}+\frac{v(t)}{Z_2} \\ &= v(t)\left(\frac{1}{Z_1}+\frac{1}{Z_2}\right) \\ &= v(t)\frac{1}{Z_{parallel}} \end{aligned}$$

so we find that $Z_{parallel}=\left(\frac{1}{Z_1}+\frac{1}{Z_2}\right)^{-1}$. This can be rewritten as:

$$Z_{parallel} = \frac{Z_1 Z_2}{Z_1+Z_2} \tag{2.26}$$

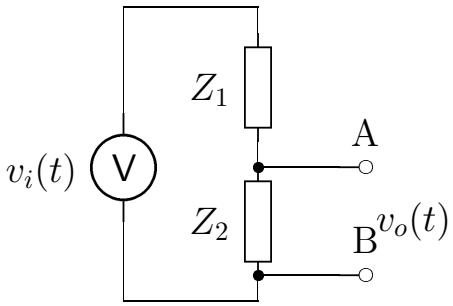

Figure 2.3: The basic voltage divider circuit with a time-dependent input voltage, $v_i(t)$ and a time-dependent output voltage, $v_o(t)$. The impedances are shown as Z_1 and Z_2 in the circuit.

Let us revisit the voltage divider discussed earlier. The time-dependent version is shown in Figure 2.3. The analysis of this circuit proceeds exactly as before. The total impedance is just

$$Z = Z_1 + Z_2,$$

for two impedances in series. From this, the current through the circuit is given as:

$$i(t) = v_i(t)/Z.$$

The voltage drop across Z_2 can be written as v_o and is given as

$$\begin{aligned} v_o(t) &= i(t) \cdot Z_2 \\ v_o(t) &= \frac{Z_2}{Z_1 + Z_2} \cdot v_i(t) \end{aligned} \tag{2.27}$$

The form of equation 2.27 is exactly the same as we found for a time-independent circuit.

2.4 The Gain of a Circuit

If we now focus on a time-dependent voltage having a single frequency, f, or $\omega = 2\pi f$, then we can include this in our analysis. The first point to note is that the impedances, Z, very often depends on ω, so that we have $Z(\omega)$ in general. We also noted earlier that the types of circuits at which we are looking do not change the frequency. If we have some input voltage,

$$v_i = v_i(\omega, t),$$

then the output voltage can be written as

$$v_o = v_o(\omega, t),$$

and the current can be written as

$$i = i(\omega, t).$$

Using this, we can rewrite our voltage divider relation as:

$$v_o(\omega, t) = \frac{Z_2(\omega)}{Z_1(\omega) + Z_2(\omega)} \cdot v_i(\omega, t).$$

The multiplicative factor that relates v_o to v_i is called the *gain*, and is given the symbol $G(\omega)$. In the case of the voltage divider, the gain is

$$G(\omega) = \frac{Z_2(\omega)}{Z_1(\omega) + Z_2(\omega)}. \tag{2.28}$$

This is one particular example of a gain, but in general we will be able to write that:

$$v_o(\omega, t) \quad = \quad G(\omega) \cdot v_i(\omega, t)\,, \tag{2.29}$$

where the gain depends on the structure of the circuit in question.

2.5 Bode Plots

The *gain* of a circuit, G, as discussed in equation 2.29, is a parameter of significant interest, and we will often find it useful to plot gain versus frequency. Note that in alternating-current circuits we may see frequencies from a few Hz up to tens of MHz (or more), so that the x-axis of our graph will usually be $\log_{10}(f)$. At the same time, since the gain itself may vary by orders of magnitude, we also plot the logarithm of the gain, resulting in a log-log plot of gain versus frequency.

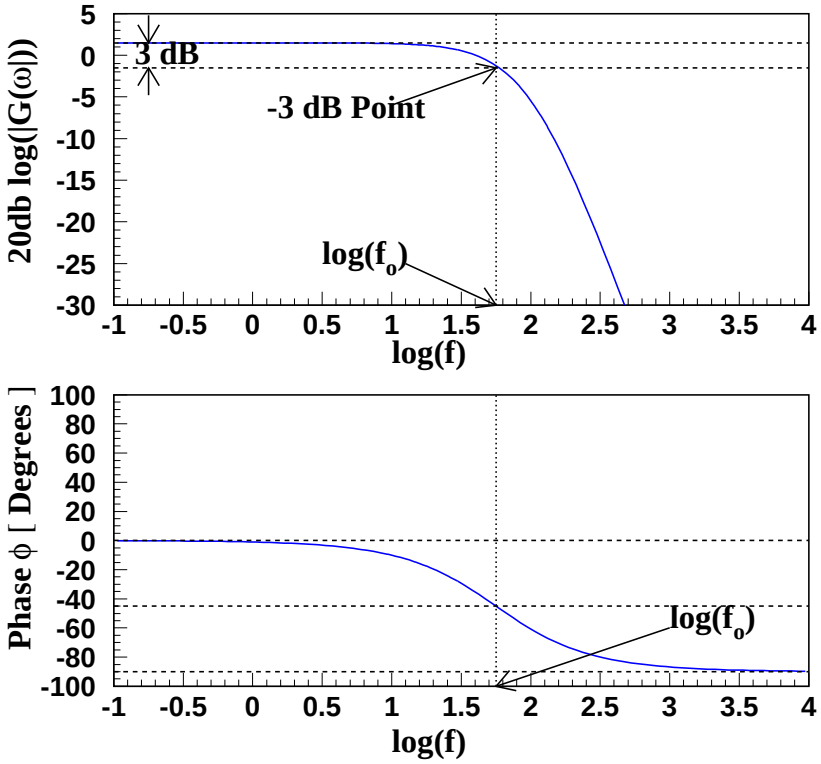

Figure 2.4: The upper plot is a typical Bode plot showing the $-3\,dB$ point. The gain is constant for low frequencies, and then starts to fall off rapidly for frequencies above the $-3\,dB$ point. The lower plot is the phase difference between the output and the input as a function of $\log f$. This changes by 90° as the frequency goes from 0 to ∞.

In these graphs, the point at which the power has fallen by a factor of 2 from some reference value is also of interest. As power is proportional to the voltage squared, in order for power to fall by a factor of 2, voltage (and hence gain) must fall by a factor of $\sqrt{2}$ (note that $\log_{10}(\sqrt{2}) = 0.15$). It is particularly

convenient to plot

$$20\,dB\,\log_{10}(\mid G(\omega)\mid) \quad \text{vs} \quad \log_{10} f$$

which is known as a *Bode Plot*. (The unit dB is defined below.) On such a plot, $20\,dB\log(\mid G(\omega)\mid)$ will drop by $3\,dB$ when the power falls by a factor of 2. We will refer to this as the $3\,dB$ point. The upper plot of Figure 2.4 shows a typical Bode plot with the gain falling off towards high frequencies. The $3\,dB$ point occurs when we are $3\,dB$ below the maximum value. In this plot, the maximum is at $1.5\,dB$ so the $3\,dB$ point is when the curve equals $-1.5\,dB$.

A *bel* (named for Alexander Graham Bell) is a factor of 10 change in power. This is commonly used to measure quantities such as gain and attenuation. For example, the attenuation, in bels, is

$$A_{bel} \;=\; \log_{10}(P_{out}/P_{in})\,.$$

Since $P \propto V^2$ for resistive loads, we have

$$\begin{aligned} A_{bel} &= \log_{10}((V_{out}/V_{in})^2) \\ A_{bel} &= 2\log_{10}(\mid V_{out}/V_{in}\mid). \end{aligned}$$

A decibel is a tenth of a bel, so the numerical value of the attenuation in decibels is ten times more than it is in bels, so

$$A_{\text{decibel}} \;=\; 20\log_{10}(\mid V_{\text{out}}/V_{\text{in}}\mid). \tag{2.30}$$

Rather than quoting a fall-off or rise in decibels per decade, one sometimes encounters decibels per *octave*. One octave in frequency is a factor of 2. It is easy to show that $20\,dB/decade$ is the same as $6\,dB/octave$.

In Figure 2.4, the gain is constant over a large frequency range, and then starts to fall off along what looks like a straight line. What is the physical meaning of this straight-line fall-off? We could fit it with an expression of the form

$$\begin{aligned} 20\,dB\log[\mid G(\omega)\mid] &= \alpha + \beta\log f \\ \log[\mid G(\omega)\mid] &= \frac{\alpha}{20\,dB} + \frac{\beta}{20\,dB}\log f \\ \mid G(\omega)\mid &= \text{constant}\cdot f^{\frac{\beta}{20\,dB}} \end{aligned} \tag{2.31}$$

Thus, the gain has the functional form of a *power law* of the frequency. If p is the power or exponent, then a power law has the form

$$\mid G\mid \;\propto\; f^p. \tag{2.32}$$

The slope parameter, β, in equation 2.31 is usually measured in units of dB/decade—by how many db does it change when the frequency changes by a factor of 10 (one decade)? For this particular plot, the exponent p in equation 2.32 is

$$p \;=\; \beta/(20\,db)\,.$$

If we measure a slope of $\beta = -20\,dB$/decade, then the exponent is just $p = -1$, and the gain falls off as $1/f$. If we measure twice this slope, $\beta = -40\,dB$/decade, then the gain falls off as $1/f^2$.

Circuits in which the gain rapidly falls off with frequency are referred to as filters. One goal of such circuits is to pass all frequencies on one side of a cut-off frequency, and to exclude all frequencies on the other side. How well the circuit does this is measured by the slope of the fall-off on a Bode plot.

Related to the Bode plot is one which looks at the phase difference between the input and output voltages as a function of $\log(f)$. The lower portion of Figure 2.4 shows such a plot corresponding to

the Bode plot above it. In this particular case, the phase difference is 0° for low frequencies, and then falls to −90° as the frequency becomes very large. At the $-3\,db$ point, the phase difference has fallen to −45°.

The gain is in general a complex number and can be expressed as

$$G(\omega) = \mid G(\omega) \mid e^{j\phi(\omega)}. \tag{2.33}$$

Both the Bode and phase plots taken together fully characterize the gain function of a circuit. As such, Bode plots are a very important tool in analyzing circuits involving time-varying voltages. If we return to upper plot in Figure 2.4, the fact that the horizontal line is at $1.5\,db$ tells us that for low frequencies, the magnitude of the gain is:

$$\begin{aligned} G &= 10^{1.5/20} \\ G &\approx 1.2. \end{aligned}$$

If the circuit has a gain of 1, then $20\,dB \log \mid G \mid = 0$, if the circuit has a gain of 10, then $20\,dB \log \mid G \mid = 20\,dB$. A gain of 100 yields $40\,dB$. Similarly, if the gain is only 0.1, then the curve will read $-20\,dB$. Second, a phase difference of 0 corresponds to a situation in which the input and output voltage are exactly in phase. In this sense, we could characterize a wire as having a gain of 1.0 and a phase of 0°. It just passes the input voltage through. As we proceed through this course, we will rely on the Bode plots and the phase versus $\log(f)$ to show the behavior of our circuits as a function of frequency. Understanding these is crucial to this course.

2.6 The Impedance of Circuit Elements

We are now in a position to determine the impedances of various components that we will use in circuits: resistors, capacitors and inductors.

In the case of a resistor, we have the simple expression that $v = iR$, or that $Z_R = R$. The impedance of a resistor is real, and just the value of the resistance. This means that in a resistor, the current and voltage are *in phase* (there is no phase difference between the two).

Next, let us consider a capacitor. We recall from basic electricity and magnetism that the charge stored in a capacitor, q, is the voltage across the capacitor plates, v, times the capacitance, C. We also know that the time derivative of the charge, dq/dt, is the current, i. From this, we get:

$$\begin{aligned} v(t) &= \frac{1}{C} q(t) \\ \frac{dv}{dt} &= \frac{1}{C}\frac{dq}{dt} \\ \frac{dv}{dt} &= \frac{i(t)}{C} \\ v(t) &= \frac{1}{C}\int i(t)dt \end{aligned}$$

If we consider a current $i(t) = i_o e^{j\omega t}$ flowing through the capacitor, then the voltage across the capacitor is given as:

$$\begin{aligned} v(t) &= \frac{i_o}{C}\int e^{j\omega t} \\ &= \frac{i_o}{j\omega C} e^{j\omega t} \end{aligned}$$

From this, we can compute the impedance of a capacitor to be:

$$Z_C(\omega) = \frac{v(t)}{i(t)}$$

$$\begin{aligned} Z_C(\omega) &= \left[\frac{i_o}{j\omega C}e^{j\omega t}\right] / \left[i_o e^{j\omega t}\right] \\ Z_C(\omega) &= \frac{1}{j\omega C}. \end{aligned} \tag{2.34}$$

We can rewrite this slightly if we note that $\frac{1}{j} = -j$, and that $-j = e^{-j\frac{\pi}{2}}$. These can be combined to give us that $Z_C = \frac{1}{\omega C}e^{-j\frac{\pi}{2}}$.

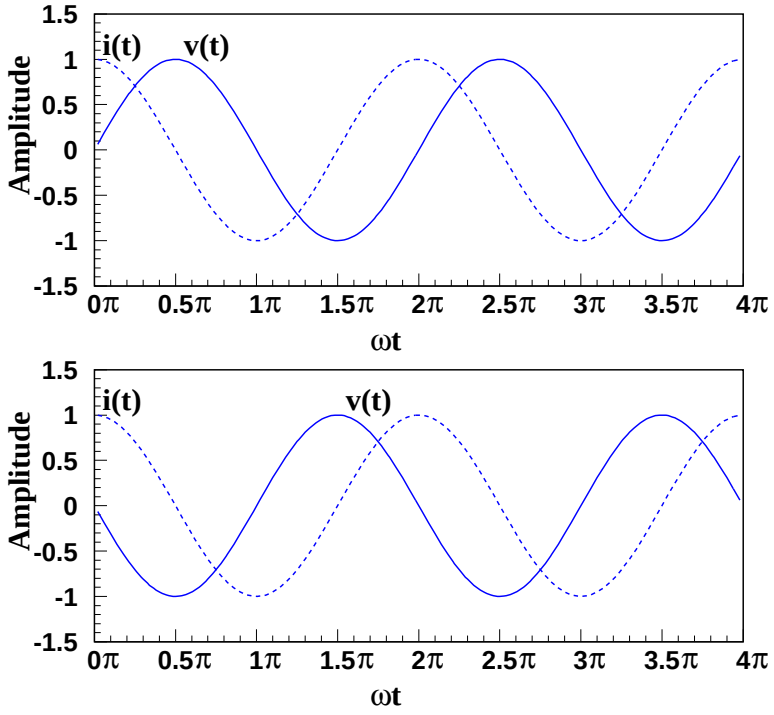

Figure 2.5: The upper plot shows the voltage and current as a function of $\omega \cdot$ time in a capacitor. The current is said to *lead* the voltage by $90°$. The lower plot shows the voltage and current as a function of $\omega \cdot$ time in an inductor. The current is said to *lag* the voltage by $90°$.

Finally, we examine the inductor. The voltage across an inductor is given as $v_L(t) = L\frac{di}{dt}$. If we have a current $i(t) = i_o e^{j\omega t}$ flowing through an inductor, then the voltage across the inductor is given as:

$$\begin{aligned} v_L(t) &= L\frac{d}{dt}i_o e^{j\omega t} \\ v_L(t) &= j\omega i_o L e^{j\omega t}. \end{aligned}$$

From this, we can compute the impedance of an inductor to be:

$$\begin{aligned} Z_L(\omega) &= \frac{v(t)}{i(t)} \\ Z_L(\omega) &= \left[j\omega L i_o e^{j\omega t}\right] / \left[i_o e^{j\omega t}\right] \\ Z_L(\omega) &= j\omega L. \end{aligned} \tag{2.35}$$

Similarly, if we note that $j = e^{j\frac{\pi}{2}}$, then we can write that the impedance is $Z_L(\omega) = \omega L e^{j\frac{\pi}{2}}$. The results for the resistor, capacitor and inductor are summarized in Table 2.1.

Component	Impedance	
R	R	R
C	$\frac{1}{j\omega C}$	$\frac{1}{\omega C}e^{-j\frac{\pi}{2}}$
L	$j\omega L$	$\omega L e^{j\frac{\pi}{2}}$

Table 2.1: Impedances of resistors, capacitors and inductors.

For a current given as $i(t) = i_o \cos\omega t$, we show the resulting voltage in both a capacitor and an inductor in Figure 2.5. The upper plot shows the voltage in the capacitor *lagging* the current in the capacitor by 90°. The lower plot shows the voltage across the inductor *leading* the current in the inductor by 90°.

$$\begin{aligned} v_C(t) &= i(t) \cdot Z_C \\ v_C(t) &= i_o e^{j\omega t} \frac{1}{\omega C} e^{-j\frac{\pi}{2}} \\ v_C(t) &= \frac{i_o}{\omega C} e^{j(\omega t - \frac{\pi}{2})} \end{aligned} \tag{2.36}$$

Example: From basic electricity and magnetism, we know that the equivalent capacitance of two capacitors in parallel is just the sum of the two capacitors.

$$C_{eq}^{parallel} = C_1 + C_2.$$

For two capacitors in series, the equivalent capacitance is given as

$$C_{eq}^{series} = \left(\frac{1}{C_1} + \frac{1}{C_2}\right)^{-1}.$$

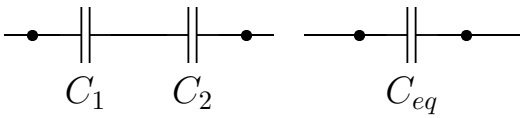

Figure 2.6: Two capacitors, C_1 and C_2, in series and the corresponding equivalent capacitance, C_{eq}.

However, equations 2.25 and 2.26 say that impedances add like resistors. How can these two statements be consistent? Let us consider two capacitors, C_1 and C_2 in series as shown in Figure 2.6. The two impedances are $Z_1 = \frac{1}{j\omega C_1}$ and $Z_2 = \frac{1}{j\omega C_2}$. We can write that the equivalent impedance must be

$$Z_{eq} = \frac{1}{j\omega C_{eq}}$$

where Z_{eq} is given as

$$Z_{eq} = Z_1 + Z_2\,.$$

We can rewrite this as follows:

$$\begin{aligned} \frac{1}{j\omega C_{eq}} &= \frac{1}{j\omega C_1} + \frac{1}{j\omega C_2} \\ \frac{1}{j\omega}\left(\frac{1}{C_{eq}}\right) &= \frac{1}{j\omega}\left(\frac{1}{C_1} + \frac{1}{C_2}\right) \\ \frac{1}{C_{eq}} &= \frac{1}{C_1} + \frac{1}{C_2}. \end{aligned}$$

This yields the result in equation 2.37, which is the expected form for two capacitors in series.

$$C_{eq} = \frac{C_1 C_2}{C_1 + C_2} \tag{2.37}$$

Example: Any real inductor not only has an inductance L, but also has some internal resistance, R_L. We can schematically treat this as an inductor L and a resistor R_L in series as shown in Figure 2.7. The equivalent inductance of such a combination is given as:

$$Z_{eq} = Z_L + Z_R \tag{2.38}$$
$$Z_{eq} = j\omega L + R_L. \tag{2.39}$$

The magnitude of the impedance is

$$\mid Z_{eq} \mid = \sqrt{(\omega L)^2 + (R_L)^2}, \tag{2.40}$$

and the phase is

$$\phi_L = \tan^{-1}\left(\frac{\omega L}{R_L}\right) \tag{2.41}$$

which allows us to write that

$$Z_{eq} = \sqrt{(\omega L)^2 + (R_L)^2}\ e^{j\phi_L}\,. \tag{2.42}$$

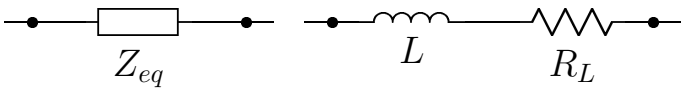

Figure 2.7: A real inductor.

We can now examine Z_{eq} as a function of the frequency, ω. As $\omega \to 0$, we find that $Z_{eq} \to R_L$, while as $\omega \to \infty$, we find that $Z_{eq} \to \omega L$. If we were to plot this, we would most likely be interested in values of frequency that spanned many orders of magnitude. Hence it would be convenient to plot Z_{eq} versus $\log(f)$ rather than f. In order to do this, we will rewrite the expression for $\mid Z_{eq} \mid$ as

$$\frac{\mid Z_{eq} \mid}{R_L} = \sqrt{1 + (\omega L/R_L)^2}\,.$$

In Figure 2.8(a), we plot the ratio of Z_{eq}/R_L versus ω on a log-log plot. The frequency, ω, is given in units of $\frac{R_L}{L}$. As with our Bode plots, to fully understand the behavior of Z, we should also make a plot of the phase. This is shown in Figure 2.8(b). For low frequencies, the phase is close to 0°, the impedance is nearly real and equal to the value of the internal resistance, R_L. As the frequency increases, the phase climbs steadily towards 90°. It passes through 45° when $\omega = \frac{R_L}{L}$.

While we often treat an inductor as a *pure* value, the reality is that it will always have an associated resistance. Only in the limit of large frequency, compared to R_L/L, can we ignore the resistance.

Example: Let us look at a real inductor, which has inductance L and resistance R_L, and apply some voltage $v(t) = V_{in}e^{j\omega t}$ to it. If we measure the voltage across the inductor, we need to take the real part of the expression for the voltage, which we can expand as:

$$v(t) = V_{in}\cos(\omega t) + jV_{in}\sin(\omega t)\,.$$

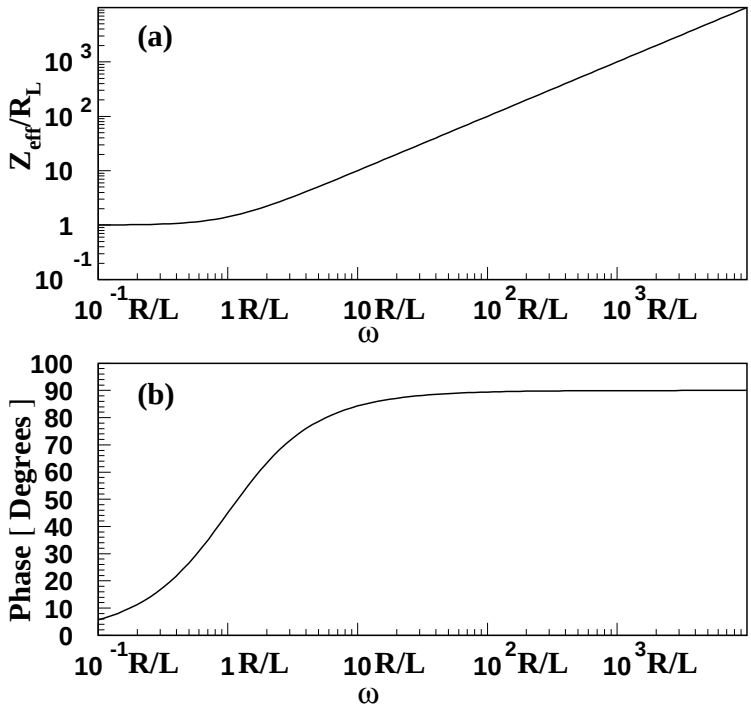

Figure 2.8: Plot (a)is a log–log plot of Z_{eff}/R_L versus ω, (in units of R_L/L). Plot (b) is the phase, (in degrees) versus ω, (in units of R_L/L), plotted on a linear-log plot.

Taking the real part yields just gives us the $\cos(\omega t)$ part of the expression; we measure:

$$v_{meas}(t) \quad = \quad V_{in}\cos(\omega t)\,.$$

Now let us examine the current through the inductor. The current is $i(t) = v(t)/Z$, where the impedance of the inductor is $Z_L = R_L + j\omega L$. In order to proceed, we will express the impedance in polar form:

$$Z_L \quad = \quad \mid Z_L \mid e^{j\phi_L}\,.$$

The magnitude of the impedance is

$$\mid Z_L \mid \quad = \quad \sqrt{R_L^2 + \omega^2 L^2}\,.$$

The phase angle can be computed as the arc-tangent of the imaginary part of Z_L divided by the real part. This gives

$$\phi_L \quad = \quad \tan^{-1}\left(\frac{\omega L}{R_L}\right)\,.$$

It is now possible to write the current in the inductor as

$$\begin{aligned} i(t) \quad &= \quad \frac{V_{in}e^{j\omega t}}{\mid Z_L \mid e^{j\phi_L}} \\ i(t) \quad &= \quad \frac{V_{in}}{\mid Z_L \mid}\cdot e^{j(\omega t - \phi_L)}\,. \end{aligned}$$

If we now measure the current, we need to take the real part of the expression. If we expand the complex exponential into sines and cosines, we get:

$$i(t) = \frac{V_{in}}{\mid Z_L \mid} \cdot \left[\cos\left(\omega t - \phi_L\right) + j \sin\left(\omega t - \phi_L\right)\right] ,$$

from which we easily see that the measured current is:

$$i_{meas} = \frac{V_{in}}{\mid Z_L \mid} \cdot \cos\left(\omega t - \phi_L\right) .$$

Finally, if we are interested in the power dissipated in the inductor, we can use $P = v(t) \cdot i(t)$, however, again we need to be careful. The current and voltage that we use here are the measured values, not the complex values. This gives that:

$$P(t) = \frac{V_{in}^2}{\mid Z_L \mid} \cdot \cos\left(\omega t\right) \cdot \cos\left(\omega t - \phi_L\right) .$$

In reality, we are probably not interested in the instantaneous power, but rather the average power over one full cycle of the voltage. This is then

$$P_{avg} = \frac{1}{T} \int_0^T P(t) dt$$

where T is the length of one full cycle, or the period ($T = \frac{2\pi}{\omega}$). We can expand this to give

$$\begin{aligned}
P_{avg} &= \frac{\omega}{2\pi} \frac{V_{in}^2}{\mid Z_L \mid} \int_{t=0}^{t=\frac{2\pi}{\omega}} \left[\cos\left(\omega t\right) \cdot \cos\left(\omega t - \phi_L\right)\right] dt \\
P_{avg} &= \frac{\omega}{2\pi} \frac{V_{in}^2}{\mid Z_L \mid} \int_{t=0}^{t=\frac{2\pi}{\omega}} \left[\cos^2\left(\omega t\right) \cos\phi_L + \cos\left(\omega t\right) \sin\left(\omega t\right) \sin\phi_L\right] dt \\
P_{avg} &= \frac{V_{in}^2}{\mid Z_L \mid} \cos\phi_L \frac{1}{2} \\
P_{avg} &= \frac{1}{2} I^2 R_L
\end{aligned}$$

where $I = V_{in}/Z_L$. For an inductor with no internal resistance, $R_L = 0$, then $\phi_L = 90°$ and $\cos\phi_L = 0$. The average power would be zero. If we use the trigonometric identity that

$$\cos\left[\tan^{-1}\left(x\right)\right] = \frac{1}{\sqrt{1+x^2}} ,$$

then we can write

$$\begin{aligned}
\cos\phi_L &= \frac{1}{\sqrt{1+\omega^2 L^2/R_L^2}} \\
\cos\phi_L &= \frac{R_L}{\mid Z_L \mid}
\end{aligned}$$

which then gives the average power as

$$P_{avg} = \frac{R_L V_{in}^2}{2 \mid Z_L \mid^2} .$$

As $R_L \to 0$, this goes to zero as we expected.

2.6.1 Power Dissipation

Let us consider an impedance $Z = R + jX$ which has a current $i(t) = i_0 e^{j\omega t+\phi}$ flowing through it. Let us define $\theta(t) = \omega t + \phi$ which then gives that

$$\begin{aligned} i(t) &= i_0 e^{j\theta(t)} \\ i(t) &= i_0 \left[\cos\theta(t) + j\sin\theta(t)\right] \end{aligned}$$

The voltage across the impedance will be $v(t) = i(t)Z$, which we can written as

$$\begin{aligned} v(t) &= i_0 \left[\cos\theta(t) + j\sin\theta(t)\right] \cdot \left[R + jX\right] \\ v(t) &= \left[i_0 R\cos\theta(t) - i_0 X\sin\theta(t)\right] + j\left[i_0 R\sin\theta(t) + i_0 X\cos\theta(t)\right] \end{aligned}$$

The average power dissipated is the average over one cycle of the real part of the current times the real part of the voltage. Thus

$$\begin{aligned} P_{avg} &= \frac{1}{T}\int_0^T Re\left[i(t)\right] Re\left[v(t)\right] \\ P_{avg} &= \frac{\omega}{2\pi}\int_0^{2\pi/\omega} i_0\cos\theta(t)\left[i_0 R\cos\theta(t) + i_0 X\sin\theta(t)\right] dt \\ P_{avg} &= \frac{\omega}{2\pi}\int_0^{2\pi/\omega} \left[i_0^2 R\cos^2\theta(t) + i_0^2 X\cos\theta(t)\sin\theta(t)\right] dt \end{aligned}$$

We note that the average of a sine times a cosine over a full cycle is zero, while that of a cosine squared over a full cycle is $\frac{1}{2}$. This gives us

$$P_{avg} = \frac{1}{2} i_0^2 R\,.$$

The average power dissipated in an impedance $Z = R + jX$ is independend of the imaginary part of the impedance. It only depends on the real part, or the resistance. Power is dissipated in resistors, but not in capacitors or pure inductors.

2.7 Time Domain Analysis

When analyzing time-dependent voltages, we can consider their behavior as functions of time, or as functions of frequency. In this section, we will focus on the time dependence, or the so-called *time domain*. For a simple oscillatory voltage given as $v(t) = v_o\cos\omega t$, the voltage is just a cosine function of the time. However, even with a voltage source that is a simple cosine function, opening or closing a switch in a circuit may introduce additional terms to the the voltage that may die out over some finite time.

2.7.1 The RC Circuit

Let us consider the circuit shown in Figure 2.9. Initially, the switch is open, there is no voltage across the capacitor ($V_C = 0$) and no current flows through the circuit. At time $t = 0$, the switch is closed. Current flows through the circuit and charges up the capacitor. Eventually, there is a voltage of $V_C = V_o$ on the capacitor and the current stops flowing. We want to examine the voltage difference between **A** and **B** as a function of time after the switch is closed.

We can use one of the Kirchhoff rules to write that the sum of all the voltage drops as we go around the circuit must be zero. This gives us the equation

$$0 = V_o - \frac{q(t)}{C} - i(t)R$$

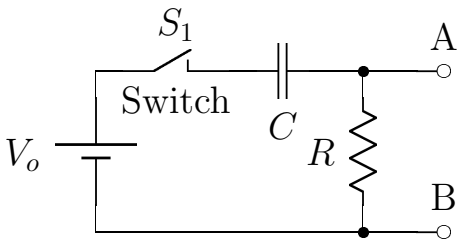

Figure 2.9: A RC circuit.

which, since $i(t) = dq/dt$, can be rewritten as

$$\frac{V_o}{R} = \frac{dq(t)}{dt} + \frac{1}{RC}q(t). \tag{2.43}$$

Because this is a first-order linear differential equation in q, we expect that the solution will be an exponential. We also anticipate that at time $t = 0$, the charge is $q(0) = 0$, and it then increases up to some maximum value. As such, we will guess a solution of the form given in equation 2.44 where Q_f is the final charge on the capacitor and α has dimensions of inverse time.

$$q(t) = Q_f\left(1 - e^{-\alpha t}\right) \tag{2.44}$$

$$\frac{dq(t)}{dt} = \alpha Q_f e^{-\alpha t}. \tag{2.45}$$

Putting equations 2.44 and 2.45 into equation 2.43, and then rearranging slightly, we arrive at the expression

$$\frac{V_o}{R} = \left(\alpha Q_f - \frac{Q_f}{RC}\right)e^{-\alpha t} + \frac{Q_f}{RC}.$$

In order to have a solution that is valid for all times, t, the term that multiplies the exponential must be zero and the remaining terms must be equal on the two sides of the equation. In other words

$$0 = \alpha Q_f - \frac{Q_f}{RC}$$
$$0 = \frac{V_o}{R} - \frac{Q_f}{RC}.$$

This then yields

$$\alpha = \frac{1}{RC} \text{ and}$$
$$Q_f = C \cdot V_o.$$

Using these constants, we can determine $q(t)$ from equation 2.44 and $i(t)$ from equation 2.45. The voltage across the capacitor is then $v_C(t) = q(t)/C$, while that across the resistor is $v_R(t) = i(t)R$. This gives

$$v_C(t) = V_o\left(1 - e^{-t/RC}\right) \tag{2.46}$$

$$v_R(t) = V_o e^{-t/RC}. \tag{2.47}$$

The two voltages are plotted in Figure 2.10 as a function of time. Note that for all times, the sum of v_C and v_R is equal to V_o—just our Kirchhoff loop equation. The product of RC is given a special name, the *characteristic time*, and is often written using the symbol τ. At a time of $t = \tau$, the exponential factor is just $e^{-1} \approx 0.37$. For each characteristic time interval, the exponential falls by another factor of ≈ 0.37.

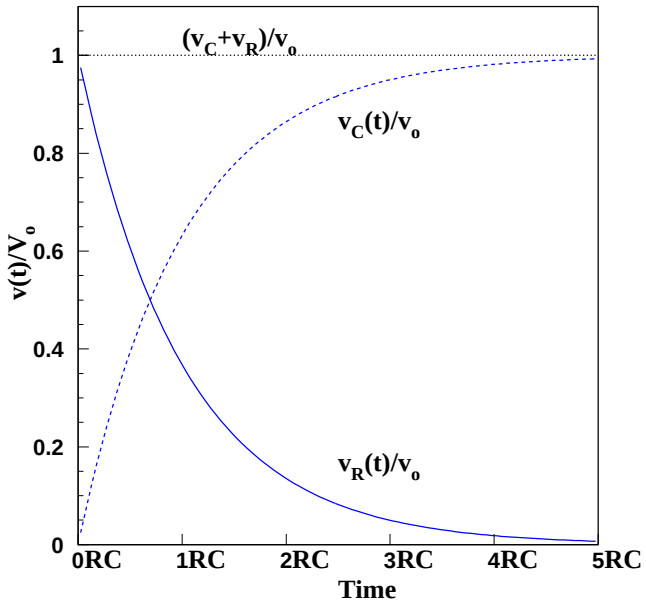

Figure 2.10: The voltage across the resistor and capacitor as a function of time for the circuit in Figure 2.9. The time is given in units of the characteristic time, $\tau = RC$, while the voltage is given as a fraction of the voltage of the source, v_o. The switch is closed at time $t = 0$.

2.7.2 The RL Circuit

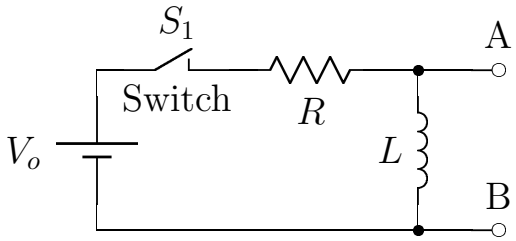

Figure 2.11: An RL circuit.

We can examine a similar circuit involving a resistor and an inductor, as shown in Figure 2.11. Initially, no current flows through the circuit. At time $t = 0$, we close the switch, and current starts to flow. As with the previous case, we can write down the Kirchhoff loop equation as we go around the loop:

$$0 \quad = \quad V_o - R\,i(t) - L\,\frac{d\,i}{dt}$$

which can be rewritten as:

$$\frac{d\,i}{dt} + \frac{R}{L}i(t) \quad = \quad \frac{V_o}{L}. \tag{2.48}$$

This equation is quite similar to equation 2.43, so we will guess a similar form for the solution, but for $i(t)$.

$$\begin{aligned} i(t) &= i_f\left(1-e^{-\alpha t}\right) \\ \frac{d\,i}{dt} &= \alpha i_f e^{-\alpha t}. \end{aligned}$$

Inserting these equations into equation 2.48, and then rearranging the parts, we find:

$$0 = i_f e^{-\alpha t}\left(\alpha - \frac{R}{L}\right) + \frac{1}{L}\left(R i_f - V_o\right)$$

In order for this equation to be true for all times, the terms in parentheses must be zero. This then yields that:

$$\begin{aligned} i_f &= \frac{V_o}{R} \text{ and} \\ \alpha &= \frac{R}{L}. \end{aligned}$$

We can use these to solve for the current as a function of time. This can in turn be used to determine the voltage as a function of time across the resistor and the inductor.

$$v_R(t) = V_o\left(1-e^{-\frac{R}{L}t}\right) \tag{2.49}$$

$$v_L(t) = V_o\, e^{-\frac{R}{L}t} \tag{2.50}$$

As with the RC circuit before, we can define a characteristic time of the RL circuit to be $\tau = L/R$. The voltage falls off $\frac{1}{e}$ for every characteristic time interval, τ.

2.7.3 Characteristic Times

We might note at this point that the various time-dependent circuits have characteristic times associated with them. It is often possible to deduce these times purely by dimensional analysis. We need only find a combination of components (more precisely, of units) which yields time units. For an RC circuit, $R\cdot C$ has dimensions of $\Omega\cdot F = s$. For the LR circuit, recall that the MKS unit of inductance is the Henry. H. Since $V_L = L\frac{di}{dt}$, $1\,H = 1V\cdot s/A$, so L/R will have dimensions of time.

In the two previous circuits, we see that characteristic times appear in decaying exponentials. If we look at the circuits on time scales much longer than the characteristic times, the solutions will be steady-state—the exponential factors $e^{-t/\tau}$ will be essentially $e^{-\infty}\approx 0$. We could also consider time scales much smaller than the characteristic times, where by much smaller we mean a very small fraction of τ (0.1τ, 0.01τ, $\cdots$). In these situations, we could expand the exponential and only keep the first and second term

$$e^{-t/\tau} \approx 1 - \frac{t}{\tau}. \tag{2.51}$$

In such a limit, the time dependencies are all linear.

As we proceed, we will find that, corresponding to the characteristic times we have seen, there will be characteristic frequencies given as $\omega_c = 1/\tau$. A circuit's response to frequencies either much less or much greater than the characteristic frequencies will be quite different. We will return to this in the next chapter.

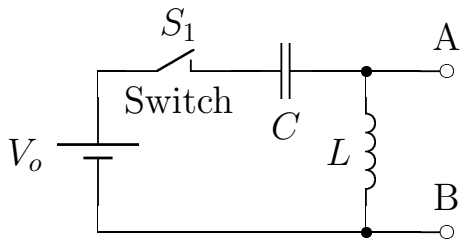

Figure 2.12: An LC circuit.

2.7.4 The LC Circuit

If we now consider an LC circuit as shown in Figure 2.12, we might start by guessing what the characteristic time should be. We are looking for some combination of Henrys and Farads that give time units. Recall that:

$$1\,\text{Farad} \quad = \quad (1\,\text{Coulomb})\,/\,(1\,\text{Volt})$$

Thus the product of $L \cdot C$ will have dimensions of time squared, so we might anticipate that the characteristic time will be

$$\tau \quad = \quad \sqrt{L\,C}\,.$$

At time $t = 0$ we close the switch, but we are told the the charge on the capacitor at this time is zero and the current through the inductor is also zero. As in the previous two cases, we can write the Kirchhoff loop equation for the circuit:

$$0 \quad = \quad V_o - \frac{q(t)}{C} - L\frac{di(t)}{dt}.$$

Since $di/dt = d^2q/dt^2$, we can rearrange the equation to yield:

$$\frac{d^2q(t)}{dt^2} + \frac{1}{LC}q(t) \quad = \quad \frac{V_o}{L}. \tag{2.52}$$

Second-order differential equations of this form have oscillatory solutions. We will try one of the form:

$$\begin{aligned} q(t) \quad &= \quad q_o + A\cos\omega_o t + B\sin\omega_o t \\ \frac{d^2q(t)}{dt^2} \quad &= \quad -\omega_o^2\,(A\cos\omega_o t + B\sin\omega_o t) \end{aligned}$$

which when put into equation 2.52 gives

$$\frac{V_o}{L} \quad = \quad \left[\frac{1}{LC} - \omega_o^2\right][A\cos\omega_o t + B\sin\omega_o t] + \frac{q_o}{LC}$$

The only way that this equation can be true for all possible times is for the part that multiplies the trigonometric functions to be zero and the remaining parts to be equal.

$$\omega_o \quad = \quad \frac{1}{\sqrt{LC}} \tag{2.53}$$

$$q_o \quad = \quad C \cdot V_o \tag{2.54}$$

If we now put all of this together, we get that

$$\begin{aligned} q(t) &= V_o C + [A\cos\omega_o t + B\sin\omega_o t] \\ i(t) &= \omega_o\left[B\cos\omega_o t - A\sin\omega_o t\right]. \end{aligned}$$

Applying the condition that at time $t = 0$, $i(0) = 0$ gives that $B = 0$. If we then apply the condition that at $t = 0$, $q(0) = 0$, we find that $V_o C + A = 0$ or $A = -V_o C$. This then gives

$$\begin{aligned} q(t) &= V_o C\left[1 - \cos\omega_o t\right] \\ i(t) &= V_o C\omega_o \sin\omega_o t \\ \frac{di(t)}{dt} &= \frac{V_o}{L}\cos\omega_o t \end{aligned}$$

(where we have made use of $\omega_o^2 = 1/LC$ in the last equation). We can then use these to evaluate the voltages across the capacitor and the inductor as a function of time. Doing so, we find

$$v_C(t) = V_o\left[1 - \cos\omega_o t\right] \tag{2.55}$$

$$v_L(t) = V_o\cos\omega_o t \tag{2.56}$$

These voltages are plotted in Figure 2.13. It is interesting to note that the voltage oscillates with frequency $\omega_o = 1/\sqrt{LC}$, and that the voltage across the capacitor gets larger (by a factor of two) than the voltage from the source. We also see that the characteristic time, $\tau = \sqrt{LC}$, that we predicted before doing this example corresponds to $1/\omega_o$.

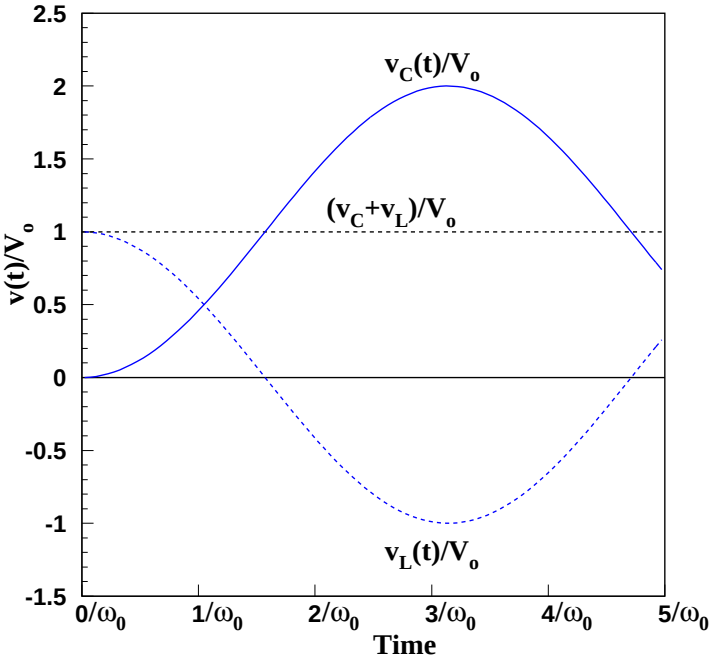

Figure 2.13: The voltage as a function of time across the capacitor and inductor in Figure 2.12.

2.8 Fourier Analysis

We will consider some function, $f(t)$, where $f(t)$ is bounded and is periodic with period T, ($f(t+T) = f(t)$). If, over a single period, T, $f(t)$ is continuous except for possibly a finite number of jump discontinuities and there are a finite number of maxima and minima, then it is possible to represent $f(t)$ as a *Fourier series* as given in equation 2.57:

$$f(t) = \frac{a_o}{2} + \sum_{n=1}^{\infty} \left[a_n \cos\left(\frac{2n\pi t}{T}\right) + b_n \sin\left(\frac{2n\pi t}{T}\right) \right] . \tag{2.57}$$

This series will converge to $f(t)$ at every point where $f(t)$ is continuous, and to $\frac{1}{2}\left(f(t^+) + f(t^-)\right)$ (the average of the left-hand and right-hand limits across the discontinuity) for every point where $f(t)$ is discontinuous. The coefficients in the sum are determined as follows:

$$a_n = \frac{2}{T} \int_{-T/2}^{T/2} f(t) \cos\left(\frac{2n\pi t}{T}\right) dt \quad n = 0, 1, 2, 3, \cdots \tag{2.58}$$

$$b_n = \frac{2}{T} \int_{-T/2}^{T/2} f(t) \sin\left(\frac{2n\pi t}{T}\right) dt \quad n = 1, 2, 3, 4, \cdots \tag{2.59}$$

An alternate formulation is in terms of phase-shifted-cosine functions. This is obtained by rewriting equation 2.57 as in equation 2.60.

$$f(t) = \frac{a_o}{2} + \sum_{n=1}^{\infty} \left[c_n \cos\left(\frac{2n\pi t}{T} + \phi_n\right) \right] \tag{2.60}$$

The coefficient c_n and the phase ϕ_n can be obtained from a_n and b_n

$$c_n = \sqrt{a_n^2 + b_n^2} \tag{2.61}$$

$$\phi_n = \tan^{-1}\left(\frac{a_n}{b_n}\right) \tag{2.62}$$

The original coefficients can be obtained from the latter two as

$$a_n = c_n \cos\phi_n \tag{2.63}$$

$$b_n = c_n \sin\phi_n. \tag{2.64}$$

The expansion in equation 2.60 can be obtained from the *Exponential Fourier Series* as in equation 2.65.

$$f(t) = \frac{1}{2} \sum_{n=-\infty}^{n=\infty} c_n e^{j\omega_n t} \tag{2.65}$$

in which

$$c_n = \frac{2}{T} \int_{-T/2}^{T/2} f(t) e^{-j\omega_n t} dt \tag{2.66}$$

where $\omega_n = n \cdot (2\pi/T)$.

Figure 2.14 are shows several waveforms that are commonly encountered in electronics as well as their Fourier transforms written in the form of equation 2.57. Let us initially look at the first three wave forms: the square wave, the triangle wave and the the saw-tooth wave. All three of these functions are *odd*, $f(-t) = -f(t)$. The Fourier transforms of these functions **only** involve sines, not cosines. This makes sense in that $\sin(-x) = -\sin(x)$; the sine function is odd. This is a general property, namely if $f(t)$ is odd, then all the a_n terms in equation 2.58 will be zero.

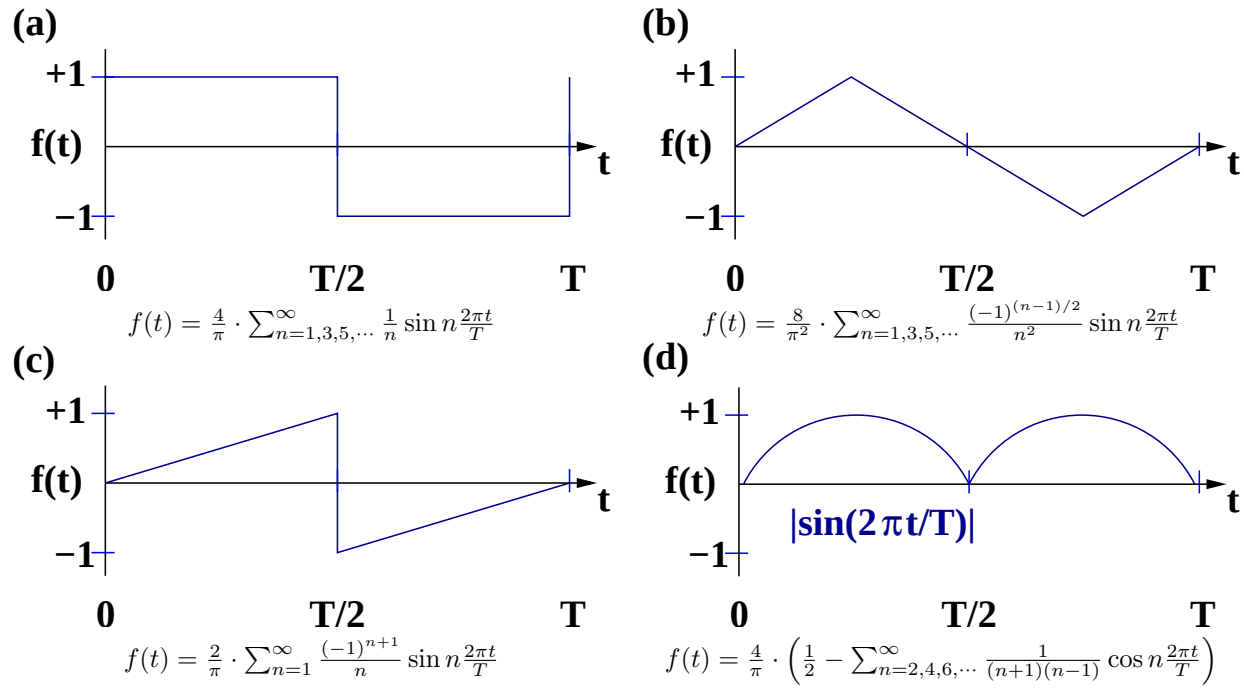

Figure 2.14: Fourier transforms of several functions typically seen in electronics. (a) is a square-wave function, (b) is a triangle wave, (c) is a saw-tooth wave and the (d) is the absolute value of a sine wave.

If we now look at the fourth example, $f(t) =\mid \sin(2\pi t/T) \mid$, this function is even: $f(-t) = f(t)$. Its Fourier transform only involves cosines. In the case where we have an even function, all the b_n terms in equation 2.59 will be zero.

Let us now ask how quickly the sum in equation 2.57 converges. We begin by looking at the $f(t) =\mid \sin(2\pi t/T) \mid$ example in Figure 2.14. For this waveform,

$$\mid \sin(2\pi t/T) \mid \;=\; \frac{4}{\pi} \cdot \left(\frac{1}{2} - \sum_{n=2,4,6,\cdots}^{\infty} \frac{\cos n\frac{2\pi t}{T}}{(n+1)(n-1)} \right) . \tag{2.67}$$

The first six terms in the sum are

$$\begin{aligned}
n = 0 \quad & \frac{2}{\pi} \\
n = 2 \quad &= -\frac{4}{\pi}\frac{1}{3} \cos(4\pi t/T) \\
n = 4 \quad &= -\frac{4}{\pi}\frac{1}{15} \cos(8\pi t/T) \\
n = 6 \quad &= -\frac{4}{\pi}\frac{1}{35} \cos(12\pi t/T) \\
n = 8 \quad &= -\frac{4}{\pi}\frac{1}{63} \cos(16\pi t/T) \\
n = 10 \quad &= -\frac{4}{\pi}\frac{1}{99} \cos(20\pi t/T) \; .
\end{aligned}$$

Figure 2.15 shows six plots corresponding to the inclusion these terms compared to the actual function. These go from (a) which shows the constant term to (f) which shows inclusion of terms up through $n = 10$. It is interesting to note how quickly this converges; after six terms there is little difference between the truncated series and the function.

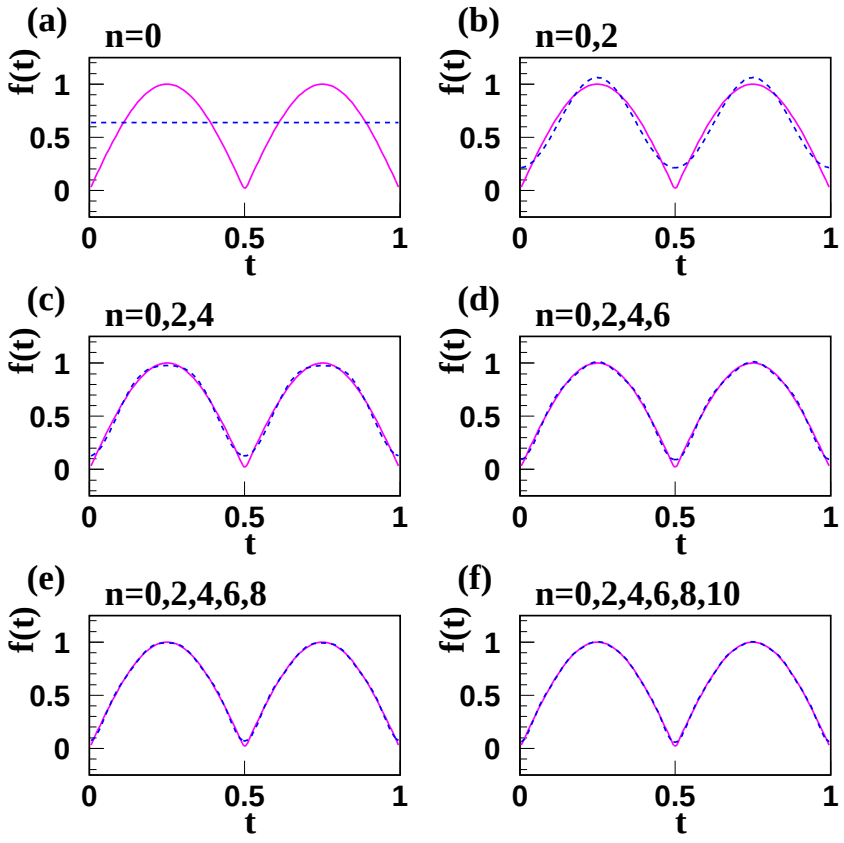

Figure 2.15: The function $|\sin(2\pi t/T)|$ and progressively more accurate approximations to the function. The Fourier expansion is given in Figure 2.14 as $f(t) = \frac{4}{\pi} \cdot \left(\frac{1}{2} - \sum_{n=2,4,6,\cdots}^{\infty} \frac{1}{(n+1)(n-1)} \cos n \frac{2\pi t}{T} \right)$. The solid magenta line is the function while the blue dashed curve is the approximation for progressively more terms in the expansion. (a) has the $n = 0$ term, (b) has $n = 0, 2$, (c) has $n = 0, 2, 4$, (d) has $n = 0, 2, 4, 6$, (e) has $n = 0, 2, 4, 6, 8$ and (f) has $n = 0, 2, 4, 6, 8, 10$. In this particular example, there is little difference between the approximation in (f) and the actual function.

Now let us consider the square-wave function given as the first example in Figure 2.14. Here the series is

$$f(t) \quad = \quad \frac{4}{\pi} \cdot \sum_{n=1,3,5,\cdots}^{\infty} \frac{1}{n} \sin n \frac{2\pi t}{T}.$$

Again, we compute the first six terms in this series.

$$\begin{aligned} n = 1 \qquad & \frac{4}{\pi} \sin(2\pi t/T) \\ n = 3 \qquad & \frac{4}{\pi} \frac{1}{3} \sin(6\pi t/T) \\ n = 5 \qquad & \frac{4}{\pi} \frac{1}{5} \sin(10\pi t/T) \end{aligned}$$

$$
\begin{aligned}
n = 7 \qquad & \frac{4}{\pi}\frac{1}{7}\sin\left(14\pi t/T\right) \\
n = 9 \qquad & \frac{4}{\pi}\frac{1}{9}\sin\left(18\pi t/T\right) \\
n = 11 \qquad & \frac{4}{\pi}\frac{1}{11}\sin\left(22\pi t/T\right)
\end{aligned}
$$

First we note that the coefficient is falling off much more slowly in this case than in the previous. Here it goes as $\frac{1}{n}$ whereas before is was as $\frac{1}{n^2}$. We anticipate that the convergence of this series will be much slower. In Figure 2.16 we produce the same six plots that we had previously. As expected, the convergence is not nearly as good. The very sharp edges require a large number of terms. The other effect that we see is the small oscillations (ringing) of the approximation around the true function.

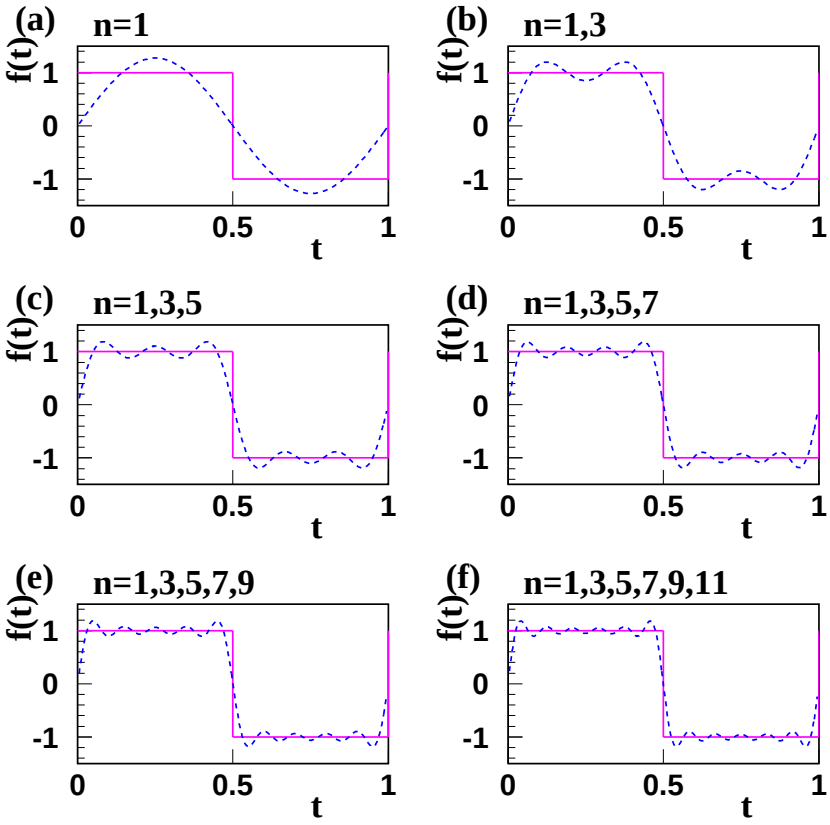

Figure 2.16: The square wave function and progressively more accurate approximations to the function. The Fourier expansion is given in Figure 2.14 as $f(t) = \frac{4}{\pi}\cdot\left(\sum_{n=1,3,5,\cdots}^{\infty}\frac{1}{n}\sin n\frac{2\pi t}{T}\right)$. The solid magenta line is the function while the blue dashed curve is the approximation for progressively more terms in the expansion. (a) has the $n = 1$ term, (b) has $n = 1, 3$, (c) has $n = 1, 3, 5$, (d) has $n = 1, 3, 5, 7$, (e) has $n = 1, 3, 5, 7, 9$ and (f) has $n = 1, 3, 5, 7, 9, 11$. In this particular example, even going out to $n = 11$, there are still observables deviations between the approximation and the actual function.

If we look at the other two examples, the triangle wave and the saw-tooth function, we find the former converges as $\frac{1}{n^2}$ while the latter converges as $\frac{1}{n}$. The slower converging saw-tooth function has

the sharp edge at $T/2$, while the triangle is a more-smoothly varying function. The bottom line from this is that very sharp edges require a lot more terms in the series.

2.8.1 Filtering Signals

If we now consider some circuit that filters out specific frequencies, we can ask about the consequences of such filtering on various periodic functions. We imagine a filter that lets high-frequency pieces through, but cuts out low-frequency components. In Figure 2.17 are shown what would happen to equation 2.67 as we remove the low-frequency parts of the wave. In this particular case, not much of the wave remains after we start removing the $n = 2$ and $n = 4$ terms in the sum.

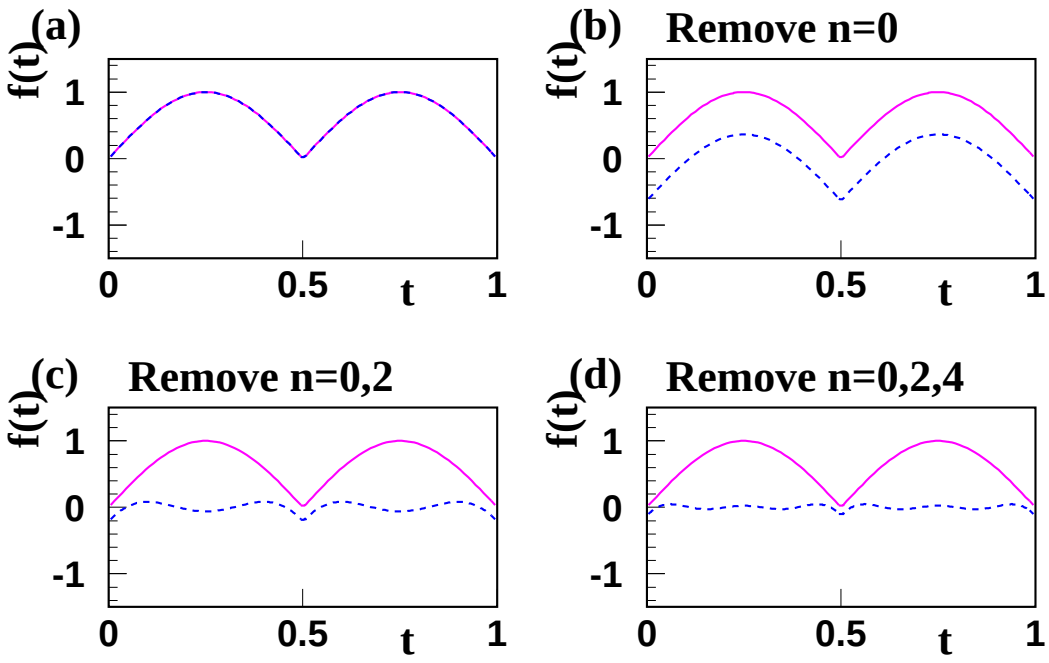

Figure 2.17: These plots show the Fourier expansions for $|\sin(2\pi t/T)|$ from which we have removed the low-frequency components. The solid-magenta curve is the actual wave form. The dashed-blue curve is what remains after we have filtered out low frequencies. (a) has no filtering, (b) has the $n = 0$ term removed, (c) has the $n = 0, 2$ terms removed and (d) has the $n = 0.2.4$ terms removed.

Problems

1. What is the RMS voltage for a square wave voltage defined by the following equations?

$$\begin{aligned} v(t) &= v_o \quad \cos(t) > 0 \\ v(t) &= -v_o \quad \cos(t) < 0 \\ v(t) &= 0 \quad \cos(t) = 0 \end{aligned}$$

2. What is the RMS voltage for a saw-tooth wave voltage defined by the following equations?

$$\begin{aligned} v(t) &= v_o(1-t) \quad 0 < t \leq 2 \\ v(t) &= v_o(3-t) \quad 2 < t \leq 4 \\ v(t) &= v_o(5-t) \quad 4 < t \leq 6 \\ v(t) &= \cdots \end{aligned}$$

3. On a set of axes representing the complex plane, draw the complex number $z = 5 + 3j$. Then, compute z^*, z^2, $z \times z^*$, $z + z^*$, and $z - z^*$. Do each calculation using symbols (i.e., $z = x + jy$), then substitute the numerical values.

4. What is the magnitude of the number $z = e^{j\theta}$?

5. Plot the point $e^{j\pi/3}$ in the complex plane. Show and calculate the real and imaginary parts.

6. Calculate the polar coordinates of the number $z = 5 + 3j$. Write this number as a complex exponential, i.e., in polar form.

7. Given $z_1 = a + jb$ and $z_2 = c + jb$, express the following in Cartesian form: (**a**) $z_1 \cdot z_2$, (**b**) $z_1 \cdot z_2^*$, (**c**) z_1/z_2 , (**d**) $\frac{1}{z_1} + \frac{1}{z_2}$, (**e**) $z_1 \cdot z_1$ and (**f**) $z_1 \cdot z_1^*$.

8. Given the complex number $z = \frac{1}{\sqrt{2}} - j\frac{1}{\sqrt{2}}$, determine the following in Cartesian coordinates: (**a**) z^2 , (**b**) z^3 and (**c**) z^4. (**d**) Express z in polar form and repeat **a**, **b** and **c**.

9. Consider the complex number $z = -2 + 2j$. What is $\sqrt{z}$ in Cartesian coordinates? **Hint:** Express z in polar coordinates in order to take the square root.

10. Compute the following derivative using both complex notation and the explicit cosine function:

$$\frac{\partial^2}{\partial x^2} \mathrm{Re}\left\{ e^{j(kx - \omega t)} \right\} \tag{2.68}$$

    Show that you also get the same answer if you factor the exponential as suggested by equation. 2.14.

11. What is $|Ae^{j(kx-\omega t)}|$? How does this compare to $|A\cos(kx - \omega t)|$? How do you explain this contrast?

12. An impedance $Z = 1000\,(1 + j)\,\Omega$ is connected to an AC voltage source of amplitude $10\,V$ and frequency $f = 60\,Hz$ as shown in Figure 2.18. You may assume that at $t = 0$, the AC voltage is at a maximum. (**a**) What is the current, $i(t)$, in the impedance? (**b**) What is the power dissipated during one AC cycle in the impedance?

13. An impedance Z is built from a resistor and capacitor connected in parallel. When connected to an AC voltage source with a frequency of $f = 60\,Hz$, the impedance has a numerical value of $Z = 1000\,(1 - j)\,\Omega$. The impedance is connected as shown in Figure 2.19, where the voltage source has an amplitude of $10\,V$ and at $t = 0$, the AC voltage is at a maximum. (**a**) What are the values of R and C? (**b**) What is the power dissipated during one AC cycle in the impedance? (**c**) What is the current in the resistor, $i_R(t)$? (**d**) What is the current in the capacitor, $i_C(t)$? (**e**) What fraction of the power from part **b** is dissipated in the resistor and the capacitor?

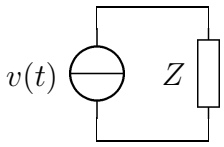

Figure 2.18: The circuit for problem 12.

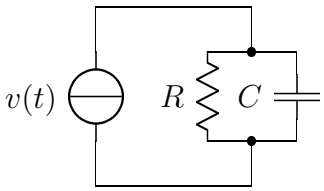

Figure 2.19: The circuit for problem 13.

14. A capacitor C and inductor L are connected in series as shown in Figure 2.20. What is the impedance, Z, of the combination? At what frequency, ω, is the magnitude of the impedance zero?

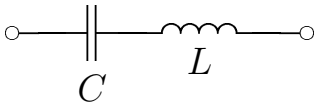

Figure 2.20: The circuit for problem 14.

15. A capacitor C, inductor L and resistor R are connected in series as shown in Figure 2.21. What is the impedance, Z, of the combination? At what frequency, ω, is the magnitude of the impedance a minimum?

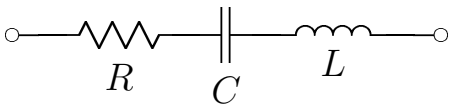

Figure 2.21: The circuit for problem 15.

16. A capacitor C and inductor L are connected in parallel as shown in Figure 2.22. What is the equivalent impedance, Z, of the circuit? Express your answer as $Z = R + jY$ where R and Y are real numbers.

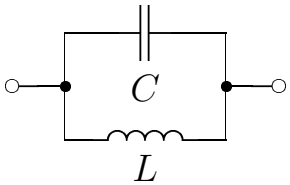

Figure 2.22: The circuit for problem 16.

17. A resistor R, capacitor C and inductor L are connected in parallel as shown in Figure 2.23. What is the equivalent impedance, Z, of the circuit? Express your answer as $Z = R + jY$ where R and Y are real numbers. At what frequency, ω, is the phase of the impedance equal to $0°$? What is the magnitude of the impedance at this frequency?

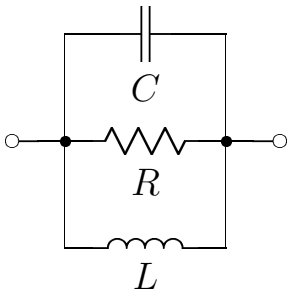

Figure 2.23: The circuit for problem 17.

18. A resistor R, capacitor C and inductor L are connected as shown in Figure 2.24. What is the equivalent impedance, Z, of the circuit? Express your answer as $Z = R + jY$ where R and Y are real numbers. At what frequency does the phase of the impedance equal $0°$?

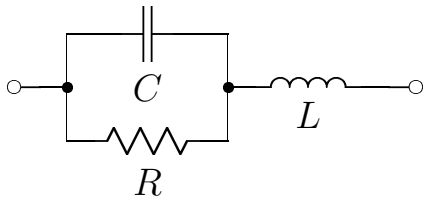

Figure 2.24: The circuit for problem 18.

19. Two resistors, R, and capacitor C and inductor L are connected as shown in Figure 2.25. What is the equivalent impedance, Z, of the circuit? Express your answer as $Z = R + jY$ where R and Y are real numbers. At what frequency does the phase of the impedance equal $0°$?

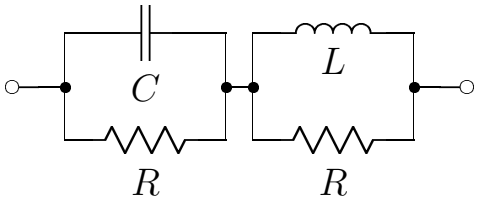

Figure 2.25: The circuit for problem 19.

20. The circuit shown in Figure 2.26 consists of a battery of constant voltage, $(2.0\,V)$, two unknown components and a switch. The output voltage is measured across the nodes **a** and **b**. Figure 2.26 below shows a scope trace that is measured across the outputs of the circuit above. The trace is generated by closing the switch at time $t = 0$ and measuring the output voltage as a function of time. The time is measured in seconds and the output voltage is measured in Volts. (**a**) Sketch the voltage across circuit element **A** as a function of time. (**b**) What is the characteristic time (roughly) of the circuit? You can read this off the plot. (**c**) If **A** and **B** are single components that we have studied in lab, there are at least two circuits that will produce the above behavior. Sketch both of them. (**d**) You measure the equivalent resistance of the the two elements as combined in series and find a value of roughly 100Ω. Which of your two circuits is correct, and what are the two components, including their values?

21. You plan to build the AC source circuit shown on the left side of Figure 2.27. The voltage $v_1(\omega)$ is assumed to be an ideal voltage source. (**a**) Draw and label the Thèvenin equivalent circuit as seen

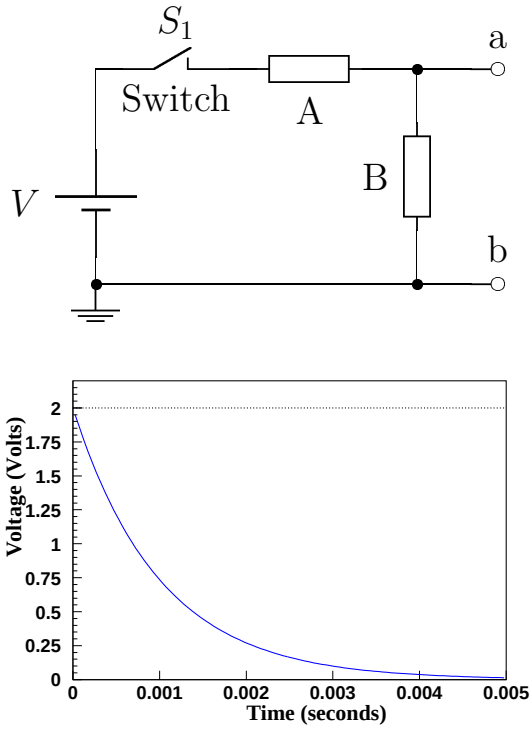

Figure 2.26: The circuit and scope trace for problem 20.

looking left into the output terminals. (**b**) Suppose you now want to measure the output of the above circuit with your oscilloscope. The equivalent circuit of a Tektronix TDS 3032 oscilloscope input is shown in the right side of Figure 2.27. The resistor has a value of $R_s = 10M\Omega$ and the capacitor has a value of $C_s = 8pF$. Write and simplify an expression for the input impedance of the oscilloscope; write this as a resistance times a dimensionless quantity. What is the characteristic frequency of this impedance? Make a log-log sketch for the magnitude of the impedance, $\mid Z_s \mid$, versus frequency, ω. (**c**) At what frequency will $\mid Z_{th} \mid$ of the source equal $\mid Z_s \mid$ of the scope? (**d**) Take $v_s(\omega)$ to be the voltage across the scope probe. Sketch the Bode plot of $\mid v_s \mid / \mid v_{th} \mid$. What is the $-3dB$ frequency in Hz? Your sketches should show the low- and high-frequency asymptotic behavior and the correct characteristic frequencies.

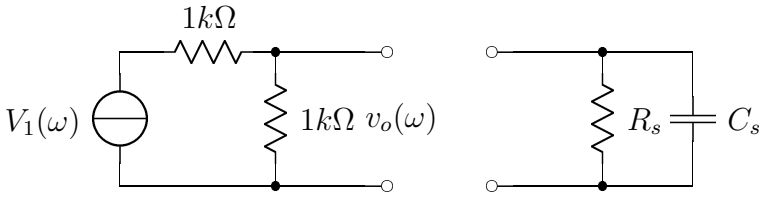

Figure 2.27: The circuits for problem 21.

22. Show that a fall-off of $6\,dB$/octave is equal to $20\,dB$/decade. (Recall that an octave is a factor of two in frequency, while a decade is a factor of ten.)

23. If the curve on a Bode plot is falling off at $60\,dB$/decade, what is the frequency dependence of the gain?

24. A circuit has a gain that goes as

$$G(\omega) \quad = \quad \frac{A}{\sqrt{\omega}}\,.$$

What are the dimensions of the constant A? What is the slope of the Bode plot in dB/decade?

25. Determine the Fourier transform of the function $f(t) = \sin^2(2\pi t/T)$. How many terms are needed to produce a good approximation?

26. Determine the Fourier transform of the function $f(t) = \sin^3(2\pi t/T)$. How many terms are needed to produce a good approximation?

27. Determine the Fourier transform of the function $f(t) = \sin^4(2\pi t/T)$. How many terms are needed to produce a good approximation?

Chapter 3

Filtering Circuits

3.1 Circuit Analysis in the Frequency Domain

While analysis of circuits in the time domain, as discussed in chapter 2, does provide useful information, the more typical situation is to examine the response of a circuit in the *frequency domain*. At some level, we have already discussed this in terms of the gain of a circuit. For the black box circuit shown in Figure 3.1, we write down the relationship between v_i and v_o as in equation 3.1.

$$v_o(\omega, t) = G(\omega) v_i(\omega, t) \tag{3.1}$$

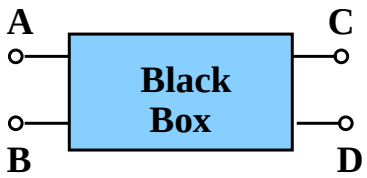

Figure 3.1: A black-box circuit with two inputs, A and B, and two outputs, C and D.

The gain, $G(\omega)$, is a function of the frequency that needs to be determined for each circuit. In addition to the gain of a circuit, two other characteristics are important: the *input impedance* and the *output impedance*. The two Thèvenin equivalent circuits shown in Figure 3.2 describe what these mean. Looking into the input of a black box, the circuit will appear as some load impedance, Z_{th}^{in}, or just the *input impedance*, Z_{in}. Looking back into the black box through the output, we will most likely see some voltage, v_{th}, and an *output impedance*, Z_{th}^{out}, or just Z_{out}.

We will generally characterize circuits by G, Z_{in} and Z_{out}. If we think back to our work with constant-voltage circuits, we can identify a few desirable characteristics of the input and output impedances. First, we would like the input impedance, Z_{in}, to be large, particularly when compared to the output impedance of any circuit connected to the input of this circuit. If this is true, then the output voltage will not be pulled down by this circuit. Second, we would like the output impedance to be small, particularly when compared to any loads that we connect the the circuit. We will see that it would be very desirable to design our circuits with almost infinite input impedance and zero output impedance. In reality, this is very difficult to do without more complicated circuit elements such as transistors and op-amps.

Now we are ready to calculate the gain functions of various circuits. We will start with low-pass and high-pass filters, and then move on to various RLC circuits. We will note that when we are determining the phase difference between the input and the output, we normally have an expression of the form

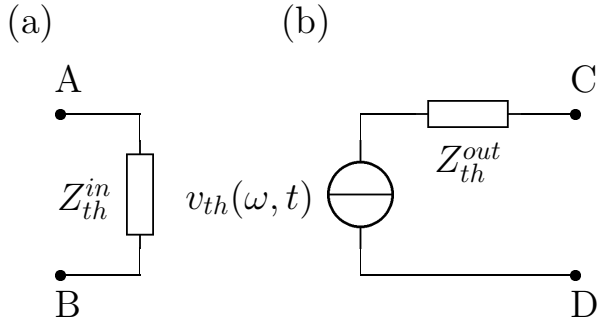

Figure 3.2: Circuit (a) shows the equivalent looking into the input side of a black-box circuit. This is characterized by some input impedance, Z_{th}^{in}. Circuit (b) shows the equivalent looking into the output side of a black-box circuit. This is characterized by a Thèvenin voltage, v_{th}, and an output impedance, Z_{th}^{out}.

$\tan\phi = a$. However, this may be ambiguous as to what ϕ we mean: $\tan\phi = \tan(\phi \pm \pi) = a$. Figure 3.3 shows two pairs of points, where the points in each pair have the same value of the tangent function. Unfortunately, in electronics we need to know which of the two possible values of ϕ we really want. In order to try to alleviate this in the text, we are going to write the expression for these phase angles as $\tan\phi = y/x$. In Figure 3.3, we would write the four expressions:

$$\begin{aligned} \tan(\phi_1) &= b/a \\ \tan(\phi_1 + \pi) &= -b/-a \\ \tan(\phi_2) &= -d/c \\ \tan(\phi_2 + \pi) &= d/-c\,. \end{aligned}$$

The ratio will be the value on the y-axis (with sign) over that along the x-axis (with sign). This notation uniquely identifies ϕ in the range of 0 to 2π.

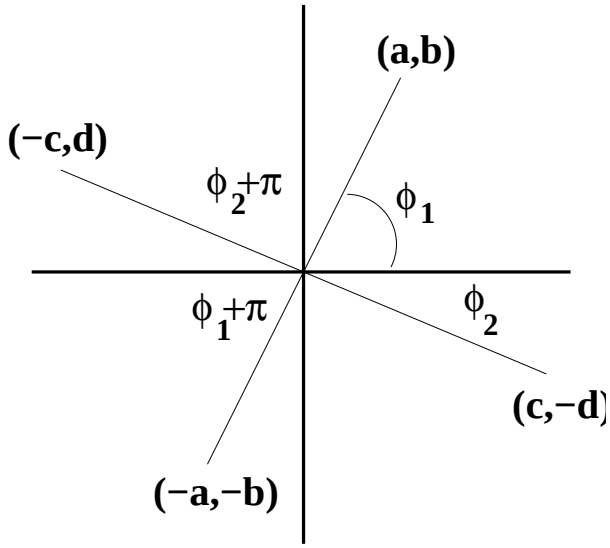

Figure 3.3: The ambiguity in $\tan(\phi)$.

3.2 The Low-pass Filter

A low-pass filter is a circuit that lets through low-frequency signals essentially without loss, but strongly damps out high-frequency signals. Figure 3.4 shows the low-pass filter configuration of an RC circuit. The values of R and C will allow us to characterize the circuit very simply. The gain, or more accurately $G \cdot v_i$, tells us the open-circuit output voltage, or Thèvenin equivalent source voltage.

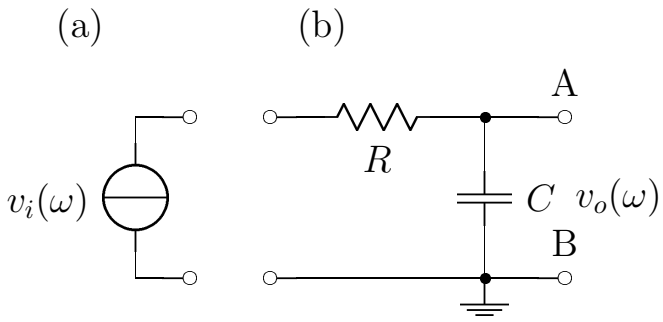

Figure 3.4: A low-pass RC filter circuit, (b), that can be connected to some AC voltage source, (a).

3.2.1 The Gain of the Low-pass Filter

The analysis of the voltage gain is conceptually straightforward. The circuit in Figure 3.4 is just a generalization of the voltage divider discussed earlier. As long as nothing is connected across the output, the same current flows through R and C. This current is just $v_i/(Z_R + Z_C)$, which means that the voltage across the capacitor is just Z_C times the current.

$$\begin{aligned} v_o(\omega) &= v_i(\omega)\frac{Z_C}{Z_R + Z_C} \\ v_o(\omega) &= v_i(\omega)\frac{\frac{1}{j\omega C}}{R + \frac{1}{j\omega C}} \end{aligned}$$

At this point, we note that the product ωRC is dimensionless. We can rewrite this as follows, which gives the gain for a low-pass filter as in equation 3.2.

$$\begin{aligned} v_o(\omega) &= v_i(\omega)\left[\frac{1}{1 + j\omega RC}\right] \\ G_{lp}(\omega) &= \frac{1}{1 + j\omega RC} \end{aligned} \tag{3.2}$$

The characteristic time $\tau = RC$ appears explicitly in these equations. However, in the frequency domain, we refer to a *characteristic frequency*, ω_{RC}, as

$$\omega_{RC} = \frac{1}{RC}\,.$$

We can now rewrite equation 3.2 in terms of ω_{RC} in such a way that it is explicitly dimensionless. This is usually a good practice as it makes it very easy to see that an equation is dimensionally correct. It also makes it quite clear that the behavior of the function depends on the relative sizes of the frequency and the characteristic frequency.

$$G_{lp}(\omega) = \frac{1}{1 + j \cdot (\omega/\omega_{RC})} \tag{3.3}$$

To see the asymptotic behavior of equation 3.3, we take the limits $\omega << \omega_{RC}$ and $\omega >> \omega_{RC}$. The region $\omega \sim \omega_{RC}$ is the cross-over region between asymptotic behaviors. It is worth noting that $(RC)^{-1}$, with units s^{-1}, is to be compared to ω and **not** to $f = \omega/(2\pi)$. This conclusion comes out of the above analysis. You might want to think of radians as being more *natural* or *physical* units than the simple counting of cycles.

Before doing more analysis, let's look at the circuit and make sure we can intuitively obtain the asymptotic behaviors. At low frequencies, $|Z_C|$ is large compared to R. Hence, most of the input voltage will appear across the capacitor. This will yield that $|G_{lp}| \to 1$ and the phase shift $\phi \to 0$. At high frequencies, $|Z_C|$ becomes small compared to R. In this case, most of the voltage will appear across the resistor and we will find that $|G_{lp}| \to 0$. Since the current will be determined by R (and therefore will be in-phase with the input voltage), $\phi_{lp} \to -\pi/2$ because of the $-j$ in Z_C. *The circuit will "pass" low frequencies (from input to output) but will attenuate high frequencies.*

Now we obtain analytic verification of the intuitive trends just described. It is a useful exercise to verify the algebra leading from equation 3.2 to the following two expressions.

$$|G_{lp}(\omega)| = \left[\frac{1}{1+(\omega/\omega_{RC})^2}\right]^{1/2} \tag{3.4}$$

$$\tan\left(\phi_{lp}(\omega)\right) = \frac{-\omega/\omega_{RC}}{1}. \tag{3.5}$$

We have written equation 3.5 explicitly as the imaginary part of the gain divided by the real part of the gain; thus the 1 in the denominator. This shows that ϕ_{lp} lies in the fourth quadrant of the complex plane since the imaginary part of G_{lp} is negative and the real part is positive.

In the limit $\omega << \omega_{RC}$, equation 3.4 goes to one and equation 3.5 goes to zero—as expected. The circuit reproduces the input at its output terminals. The magnitude of the gain becomes constant—this is a power law with $\alpha = 0$,

$$\begin{aligned} 20\,dB \log G &= \text{constant} \\ G &\sim f^0 \end{aligned}$$

which is a horizontal line on the Bode plot shown in Figure 3.5. On the linear phase angle scale, this quantity also becomes constant with a value of zero. In the limit $\omega >> \omega_{RC}$, equation 3.4 becomes ω_{RC}/ω, which is much smaller than 1. The phase in equation 3.5 becomes the inverse tangent of a large negative number—the phase angle goes to $-\frac{\pi}{2}$. While the phase angle plot becomes constant, the Bode plot follows another straight line, this time with a slope of $-20\,dB$/decade. This is a power law with $\alpha = -1$,

$$G \sim f^{-1}.$$

Notice that the overall Bode plot is simple. It is two straight lines connected by a curved piece in the neighborhood of $\omega = \omega_{RC}$. At this *characteristic frequency*, ω_{RC}, the magnitude of the gain is $1/\sqrt{2} \approx 0.707$; the base-10 log of this is -0.1505, so this point is below the $0\,dB$ level (gain of one) by just about $3\,dB$. The so-called $-3\,dB$ point. At the same frequency, the phase is $-\frac{\pi}{4}$, or $-45°$ as shown in Figure 3.5.

3.2.2 The Output Impedance of the Low-pass Filter

To compute the impedance seen from the output terminals, we assume that the source can be modeled as an ideal source in series with a resistance, r_s. This is shown in Figure 3.6. If the source impedance were complex rather than just a real resistance (r_s), then the *input signal* to the filter could contain distortions of the original source signal (frequency dependent amplitudes and phases). However, even just a resistance, r_s, can affect the nature of the output. We will see this again.

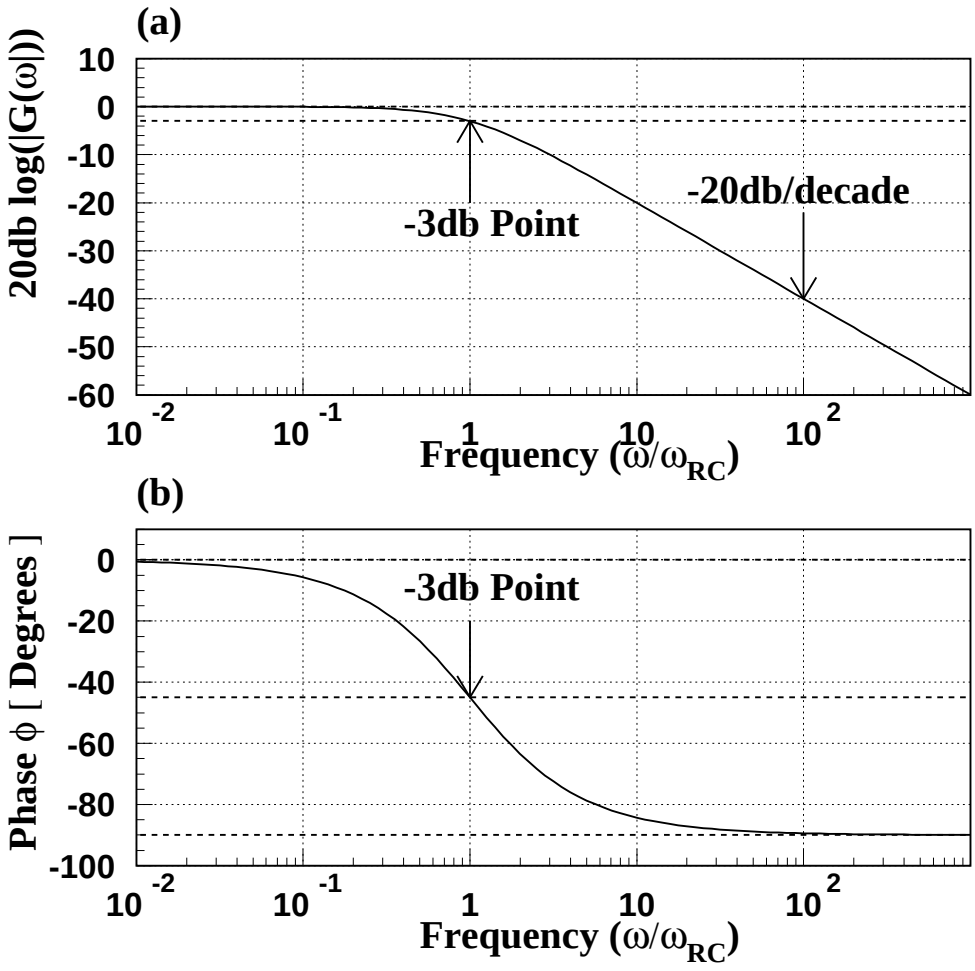

Figure 3.5: Frequency response of the low-pass filter. (a) shows the Bode plot for the response in decibels while (b) shows the phase shift in units of degrees. The frequency axis is scaled by $\omega_{RC} = (RC)^{-1}$; by re-scaling this axis, these curves describe any RC low-pass filter.

By comparing the short-circuit current to the open-circuit voltage, the Thèvenin equivalent impedance can be found just as we did for the resistive voltage divider. The Thèvenin, or *output*, impedance as seen across the A and B terminals in Figure 3.6 is just Z_C in parallel with $(R + r_s)$.

$$\begin{aligned} Z_{out} &= \frac{(R+r_s)\cdot\left(\frac{1}{j\omega C}\right)}{(R+r_s)+\left(\frac{1}{j\omega C}\right)} \\ Z_{out} &= \frac{(R+r_s)}{1+j\omega\,(R+r_s)\,C} \end{aligned} \tag{3.6}$$

The Thèvenin equivalent circuit for the output terminals is just a voltage source of amplitude $|v_i(\omega)\cdot G_{lp}(\omega)|$ and phase angle given by equation 3.5, in which R has been replaced by $R + r_s$. This is connected in series with the impedance given in 3.6. At low frequencies the capacitor's impedance becomes large and

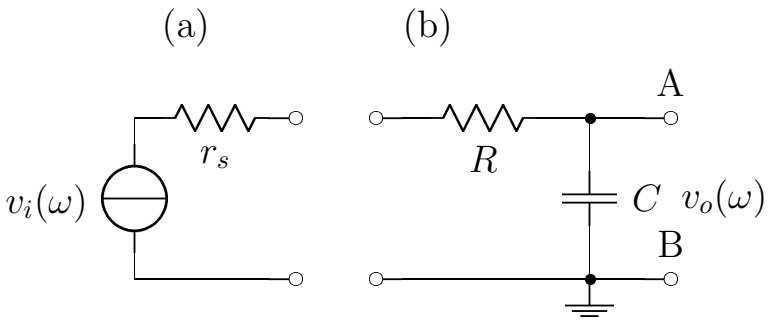

Figure 3.6: A low-pass circuit, (b), and a source, (a). The source includes some internal source resistance, r_s.

$Z_{out} \rightarrow (R + r_s)$, whereas at high frequencies, $Z_{out} \rightarrow 1/(j\omega C)$ (the capacitor impedance dominates the parallel combination). $|Z_{out}|$ is a maximum at low frequencies and decreases as ω increases. We can rewrite this slightly to emphasize the fact that if r_s is small relative to R, then it has little effect on the output impedance of the filter.

$$Z_{out} = R\left(1 + \frac{r_s}{R}\right) \cdot \frac{1}{1 + j\,(\omega/\omega_{RC}) \cdot (1 + r_s/R)} \tag{3.7}$$

3.2.3 The Input Impedance of the Low-pass Filter

From the input terminals, the filter (with no load connected) looks like a series combination of R and C:

$$\begin{aligned} Z_{in} &= R + \frac{1}{j\omega C} \\ Z_{in} &= R\left(1 - j\frac{\omega_{RC}}{\omega}\right). \end{aligned}$$

For frequencies much larger than ω_{RC}, $|Z_{in}| \rightarrow R$ and becomes larger as the frequency decreases. The Thèvenin equivalent circuit for the input terminals is just Z_{in}.

Note however, that *when we put a load, Z_L, on the output, the input impedance changes*: we have the parallel combination of C and Z_L in series with R. This is a serious drawback of the simple passive filter circuits we are considering here—it is a drawback that we will overcome later in the course by using transistors and/or operational amplifiers.

3.3 The High-pass Filter

A high-pass filter is a circuit that lets through high-frequency signals essentially without loss, but strongly attenuates low-frequency signals. Figure 3.7 shows the high-pass filter configuration of an RC circuit.

3.3.1 The Gain of the High-pass Filter

Again taking advantage of our understanding of a voltage divider, we can determine the voltage gain of the high-pass filter as follows:

$$\begin{aligned} v_o(\omega) &= v_i(\omega)\frac{Z_R}{Z_R + Z_C} \\ v_o(\omega) &= v_i(\omega)\frac{R}{R + \frac{1}{j\omega C}} \end{aligned}$$

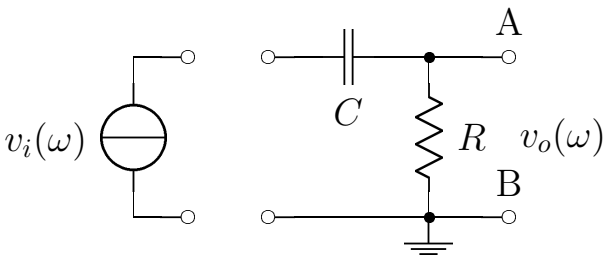

Figure 3.7: A high-pass RC filter circuit (right) connected to some input voltage source, (left).

$$v_o(\omega) = v_i(\omega)\left[\frac{j\omega RC}{1+j\omega RC}\right]$$

or

$$G_{hp}(\omega) = \frac{j\,(\omega/\omega_{RC})}{1+j\,(\omega/\omega_{RC})} \tag{3.8}$$

As before, we have arranged this expression so that all terms are dimensionless and taken advantage of the characteristic frequency, $\omega_{RC} = (RC)^{-1}$. The magnitude and phase of the gain are then:

$$|G_{hp}(\omega)| = \frac{\omega/\omega_{RC}}{[1+(\omega/\omega_{RC})^2]^{1/2}} \tag{3.9}$$

$$\tan\left(\phi_{hp}(\omega)\right) = \frac{1}{\omega/\omega_{RC}} \tag{3.10}$$

Figure 3.8(a) shows a Bode plot and (b) shows a phase versus $\log(f)$ for these. It is easy to show that equations 3.9 and 3.10 lead to the following asymptotic behaviors. For $\omega << \omega_{RC}$,

$$\begin{aligned} |G_{hp}| &\rightarrow \omega/\omega_{RC} << 1 \\ \phi_{hp} &\rightarrow \frac{\pi}{2}. \end{aligned}$$

For the limit where $\omega >> \omega_{RC}$,

$$\begin{aligned} |G_{hp}| &\rightarrow 1 \\ \phi_{hp} &\rightarrow 0 \end{aligned}$$

3.3.2 Input and Output Impedance

Figure 3.9 shows a high-pass filter connected to a voltage source with an internal resistance r_s. Looking at this as a voltage divider like we did for the low-pass filter, the output impedance can be calculated as:

$$\begin{aligned} Z_{out} &= (Z_s + Z_C) \,||\, Z_R \\ Z_{out} &= \frac{\left(r_s + \frac{1}{j\omega C}\right)\cdot R}{\left(r_s + \frac{1}{j\omega C}\right) + R} \end{aligned}$$

$$Z_{out} = R\cdot\frac{(r_s/R) - j\,(\omega_{RC}/\omega)}{(1+r_s/R) - j\,(\omega_{RC}/\omega)}$$

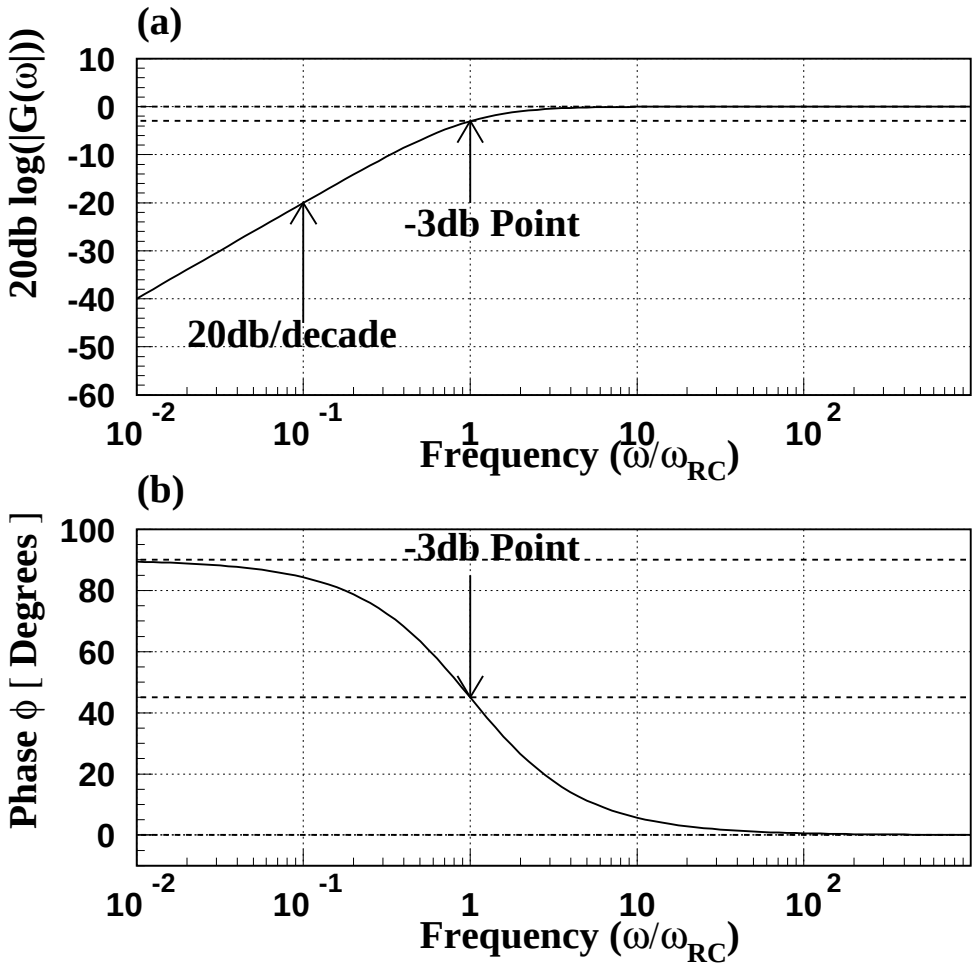

Figure 3.8: Frequency response of the high-pass filter. (a) shows the magnitude response in decibels while (b) shows the phase shift in units of degrees. The frequency axis is scaled by $\omega_{RC} = (RC)^{-1}$; by re-scaling this axis, these curves describe any RC high-pass filter.

$$Z_{out} \quad = \quad R \cdot \frac{1 + j\,(\omega/\omega_{RC})\,(r_s/R)}{1 + j\,(\omega/\omega_{RC}) \cdot (1 + r_s/R)}. \tag{3.11}$$

Again, (3.8) and (3.11) specify the Thèvenin equivalent circuit for the output terminals. Note that in the limit of $r_s << R$, equation 3.11 and equation 3.7 become the same.

Under no-load conditions, the input impedance is just the same as the low-pass filter. (See section 3.2.3).

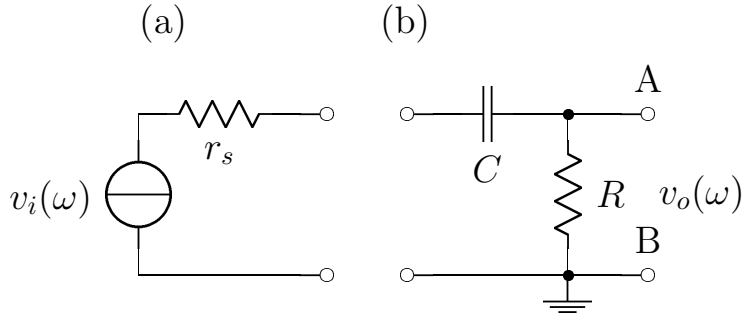

Figure 3.9: A high-pass RC filter circuit (b) connected to some voltage source with internal resistance r_s, (a).

3.4 Integrating and Differentiating Circuits

If we examine the output of either the high-pass or the low-pass filter, there are two regions defined by the high- and low-frequency limits: $\omega >> \omega_{RC}$ and $\omega << \omega_{RC}$. In one of these limits, the output voltage and the input voltage are nearly the same. In the other limit, the magnitude of the gain falls off with the frequency, and the magnitude of the phase goes to $\frac{\pi}{2}$. Let us look more carefully at this latter limit for the two filters. These are summarized in Table 3.1. If we now take a specific input voltage,

Filter	Limit	Gain	Phase
Low-pass	$\omega >> \omega_{RC}$	$\sim \omega^{-1}$	$-\frac{\pi}{2}$
High-pass	$\omega << \omega_{RC}$	$\sim \omega^{1}$	$\frac{\pi}{2}$

Table 3.1: Limits of the low-pass and high-pass filter.

$v_i = v_a \cos(\omega t)$, then the output for the two filters will be as in equation 3.12 for the low-pass filter and as in equation 3.13 for the high-pass filter.

$$\begin{aligned} v_o &\sim \frac{1}{\omega}\cos\left(\omega t - \frac{\pi}{2}\right) \\ v_o &\sim \frac{1}{\omega}\sin(\omega t) \\ v_o &\sim \int v_i dt \end{aligned} \tag{3.12}$$

$$\begin{aligned} v_o &\sim \omega\cos\left(\omega t + \frac{\pi}{2}\right) \\ v_o &\sim -\omega\sin(\omega t) \\ v_o &\sim \frac{dv_i}{dt} \end{aligned} \tag{3.13}$$

The two RC filters that we have looked at appear to behave as integrating and differentiating circuits. Lets us now be a bit more explicit. We consider an input voltage written in complex form as in equation 3.14. Recall that when we measure this, we are taking the real part of the equation.

$$v_i(t) = v_a e^{j(\omega t + \phi_a)} \tag{3.14}$$

If we now take the gain for a low-pass filter from equation 3.2, we can calculate the output voltage of our filter as:

$$\begin{aligned} v_o(t) &= G_{lp}(\omega)\cdot v_i(t) \\ v_o(t) &= \frac{1}{1+j\omega/\omega_{RC}}\cdot v_a e^{j\omega t+\phi_a} \end{aligned}$$

Now let us consider the limit in which $\omega >> \omega_{RC}$. We can approximate the previous equation as:

$$\begin{aligned} v_o(t) &= \frac{1}{j\omega/\omega_{RC}} \cdot v_a e^{j(\omega t+\phi_a)} \\ v_o(t) &= \frac{\omega_{RC}}{j\omega} \cdot v_a e^{j(\omega t+\phi_a)} \\ v_o(t) &= \omega_{RC} \cdot \int v_a e^{j(\omega t+\phi_a)} dt \end{aligned}$$

$$v_o(t) = \omega_{RC} \cdot \int v_i(t) dt \tag{3.15}$$

In the high-frequency limit, the output of the low-pass filter is the characteristic frequency times the integral of the input voltage.

A similar analysis can be carried out for the high-pass filter. We find that

$$v_o(t) = \frac{1}{\omega_{RC}} \cdot \frac{dv_i(t)}{dt} \tag{3.16}$$

We can now extend this to any periodic function by expanding the function as its Fourier series. If we represent an input signal in terms of a Fourier sum as

$$v(t) = \sum_{n=0}^{\infty} a_n \cos\left(n\omega_o t + \phi_n\right) ,$$

it is easy to take a derivative or integral of an arbitrarily complicated waveform. For the derivative, we have:

$$\begin{aligned} \dot{v}_i(t) &= \sum_{n=0}^{\infty} a_n(-n\omega_o) \sin(n\omega_o t + \delta_n) \\ \dot{v}_i(t) &= \sum_{n=0}^{\infty} a_n(-\omega_n) \cos(\omega_n t + \delta_n + \pi/2) \\ \dot{v}_i(t) &= \sum_{n=0}^{\infty} a_n \mathrm{Re}\left[(j\omega_n) e^{j(\omega_n t+\delta_n)}\right] , \end{aligned}$$

where $\omega_n = n\omega_o$. Thus, in the formula for the derivative, the frequency component at ω_n has just been multiplied by factor of $j\omega_n$. Inserting this factor is the same as taking the derivative!

If we have a circuit whose gain is $G(\omega) = j\omega/\omega_c$ (ω_c being some characteristic frequency—$(RC)^{-1}$ in our case), *over the range of frequencies in the input signal's Fourier representation*, then the output will be proportional to the derivative. The high-pass filter performs exactly this function in the region $\omega << \omega_c$.

A similar calculation leads to the conclusion that a gain function which scales with $1/(j\omega)$ leads to an output proportional to the integral of the input. The low-pass filter performs exactly this function in the region $\omega >> \omega_{RC}$.

Notice that in neither case can we build a circuit which will perform the appropriate operation on an arbitrary input signal. Eventually, the gain becomes constant as seen in Figs. 3.5 and 3.8. These mathematical functions work in the frequency region where the gain is small. This means that, with these circuits, the relevant signal may also be small and therefore subject to noise and measurement difficulties. We will be able to make better circuits when we include some amplification. On the other hand, the eventual *saturation* of the gain at some finite level will always be a limitation—we can't make circuits with infinite gain!

3.5 RLC Circuit Analysis

So-called RLC circuits play a large role in the modern world. An important use is in receivers where they select out a particular frequency, the *resonant frequency* of the circuit. This frequency is characterized by the inductance, L, and capacitance, C, of the circuit and is given as $1/\sqrt{LC}$. From earlier, we know that three characteristic frequencies can actually be defined for a circuit with R, L and C. We already have the LC, or resonant, frequency from above. The other two are the RC and the RL frequencies. We can define these three frequencies as follows.

$$\begin{aligned} \omega_{LC} &\equiv \frac{1}{\sqrt{LC}} \\ \omega_{RC} &\equiv \frac{1}{RC} \\ \omega_{RL} &\equiv \frac{R}{L} \end{aligned}$$

In this section, we will examine the response of various RLC circuits in terms of these three frequencies. In all cases, the resonant frequency is the crucial one in defining behavior. Choosing this frequency defines a relationship between L and C. We will see that the free parameter that controls the other two frequencies is the resistance of the circuit. In the time domain, the resistance controls how quickly a signal is damped out. In the frequency domain, we will see that the resistance controls how narrow the resonance is (around the resonant frequency). The smaller R is, the sharper the resonance and the bigger the gain. The limit of very small R for the two frequencies that depend on R is:

$$\begin{aligned} \omega_{RC} &\rightarrow \infty \\ \omega_{RL} &\rightarrow 0. \end{aligned}$$

3.5.1 The series RLC circuit.

Let us consider the series RLC circuit as shown in Figure 3.10. In such a circuit, we can nominally consider several possible output voltages. We see three of them in the figure. These are just the voltages across the three components in the circuit: v_o^R across the resistor, v_o^L across the inductor and v_o^C across the capacitor. Unfortunately, as discussed earlier, the resistor and the inductor are probably not separable. All inductors have a non-zero internal resistance. If we want to restrict ourselves to output voltages that we can measure, we are limited to v_o^C and v_o^{LR}. The latter is the voltage measured across the resistor and inductor, where we have assumed that the only resistance in the circuit is that of the inductor.

In analyzing the circuit, we can start by writing the total impedance seen by the source. This is just the series combination of the impedances of the three elements in the circuit.

$$\begin{aligned} Z_{tot} &= R + j\omega L + \frac{1}{j\omega C} \\ Z_{tot} &= R\left[1 - j\frac{\omega_{RC}}{\omega}\left(1 - \frac{\omega^2}{\omega_{LC}^2}\right)\right], \end{aligned} \tag{3.17}$$

where ω_{LC} and ω_{RC} are defined above. When the frequency is equal to the resonant frequency, $\omega = \omega_{LC}$, the imaginary part of Z_{tot} is zero and the magnitude of Z_{tot} is minimized: $\mid Z_{tot} \mid = R$. In addition, the current in the loop is in-phase with the source voltage.

We can now write the three output voltages in Figure 3.10 using the voltage divider formula. The voltage across any one of the elements, X, is

$$v_o^X(\omega) = \frac{Z_X}{Z_{tot}} \cdot v_i(\omega).$$

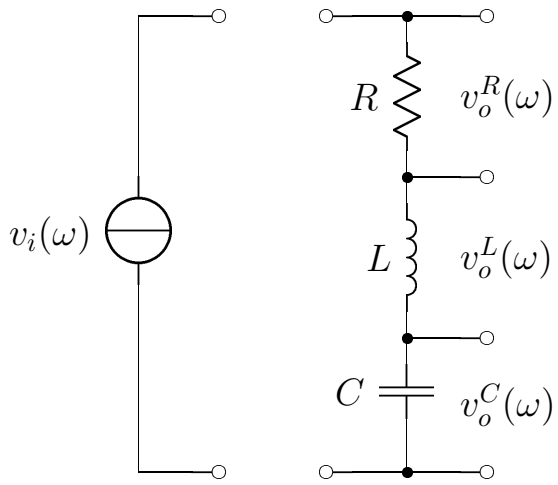

Figure 3.10: The series RLC circuit.

We can also write this as a gain across the specified component.

$$G_X(\omega) = \frac{Z_X}{Z_{tot}} \tag{3.18}$$

The component gain we can actually measure is that across the capacitor. In this case, we can write:

$$\begin{aligned} G_C(\omega) &= \frac{Z_C}{Z_{tot}} \\ G_C(\omega) &= \frac{1}{(1-\omega^2/\omega_{LC}^2) + j\,(\omega/\omega_{RC})}. \end{aligned} \tag{3.19}$$

From this, we can determine the magnitude and the phase of the capacitor gain.

$$\begin{aligned} \mid G_C(\omega) \mid &= \left[\frac{1}{(1-\omega^2/\omega_{LC}^2)^2 + (\omega/\omega_{RC})^2}\right]^{1/2} \\ \tan\left(\phi_C(\omega)\right) &= \frac{-\omega/\omega_{RC}}{1-\omega^2/\omega_{LC}^2} \end{aligned} \tag{3.20}$$

This suggests several limits we want to examine. First let us consider the case when the circuit is at the resonant frequency, $\omega = \omega_{LC}$. In this particular limit, we have that:

$$\begin{aligned} \mid G_C \mid &= \frac{\omega_{RC}}{\omega_{LC}} = \frac{\omega_{RL}}{\omega_{RC}} = \sqrt{\frac{L}{R^2 C}} \\ \phi_C &= -\frac{\pi}{2} \end{aligned}$$

As $R \to 0$, we see that $\mid G \mid \to \infty$. In the low-frequency limit, $\omega \to 0$, we have that

$$\begin{aligned} \mid G_C \mid &\to 1 \\ \phi_C &\to 0(-)\,, \end{aligned}$$

where we have written $0(-)$ to indicate that the phase approaches 0 from the negative side. Finally, we can consider the high-frequency limit, $\omega \to \infty$.

$$\begin{aligned} \mid G_C \mid &\to \left(\frac{\omega_{LC}}{\omega}\right)^2 \\ \phi_C &\to -\frac{\pi}{2} \end{aligned}$$

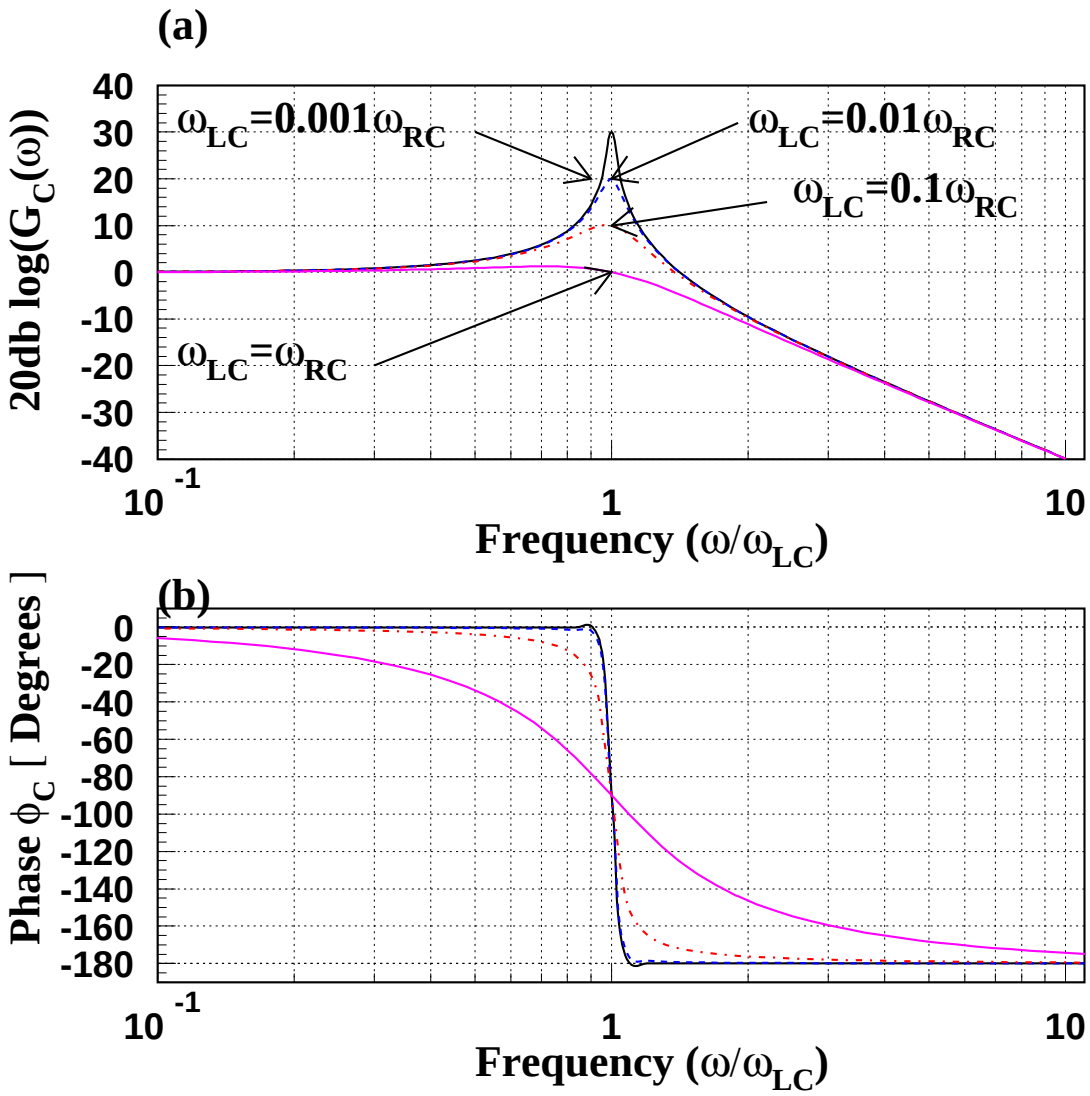

Figure 3.11: The Bode (a) and phase plot(b) for the gain measured across the capacitor in the series RLC circuit. The four curves are for progressively larger ω_{RC}, corresponding to progressively smaller R. We see that the resonance becomes sharper and the peak gain larger as R decreases.

Figure 3.11 shows the Bode and phase plot for the gain across the capacitor. In the Bode plot, note that the peak around the resonant frequency, ω_{LC}, gets sharper as ω_{RC} increases, or as mentioned earlier, as R decreases. If R becomes large enough, the resonance response vanishes.

In interpreting the tangent in equation 3.20, it is important to note the minus sign in the numerator. The denominator changes sign as we go through ω_{LC}, while the numerator remains negative. The gain moves from the fourth quadrant to the third quadrant of the complex plane as the frequency sweeps through the resonance.

We can now look at the gain across the resistor:

$$\begin{aligned} G_R(\omega) &= \frac{R}{Z_{tot}} \\ G_R(\omega) &= \frac{1}{1 - j\,(\omega_{RC}/\omega)\,(1 - \omega^2/\omega_{LC}^2)}. \end{aligned} \tag{3.21}$$

From this, we can write the magnitude and phase of the gain across the resistor as

$$\mid G_R(\omega) \mid \;= \frac{\omega/\omega_{RC}}{\left[(\omega/\omega_{RC})^2 + \left(1 - \omega^2/\omega_{LC}^2\right)^2\right]^{1/2}}$$

and

$$\tan\left[\phi_R(\omega)\right] = \left[\frac{\omega_{RC}}{\omega}\left(1 - \frac{\omega^2}{\omega_{LC}^2}\right)\right]/1. \tag{3.22}$$

In equation 3.22, the phase angle is positive when $\omega < \omega_{LC}$ and then changes sign at $\omega = \omega_{LC}$. As with the capacitor, we can consider several frequency limits for the gain and phase. We start with $\omega = \omega_{LC}$.

$$\begin{aligned} G_R(\omega) &= 1 \\ \phi_R &= 0 \end{aligned}$$

We can now examine the low-frequency limit, $\omega \ll \omega_{LC}$.

$$\begin{aligned} G_R(\omega) &\rightarrow \frac{\omega/\omega_{RC}}{\left[1 + (\omega/\omega_{RC})^2\right]^{1/2}} \\ G_R(\omega) &\rightarrow \omega/\omega_{RC} \rightarrow 0 \\ \tan(\phi_R) &\rightarrow \frac{\omega_{RC}/\omega}{1} \\ \phi_R &\rightarrow \frac{\pi}{2} \end{aligned}$$

The last limit is $\omega \gg \omega_{LC}$.

$$G_R(\omega) \rightarrow \frac{\omega/\omega_{RC}}{\omega^2/\omega_{LC}^2}$$

Noting that $\omega_{RC}/\omega_{LC}^2 = \omega_{RL}$, we can write

$$G_R(\omega) \rightarrow \frac{\omega_{RL}}{\omega} \rightarrow 0\,.$$

For the phase we have

$$\begin{aligned} \phi_R &\rightarrow \tan^{-1}\left(\frac{-\omega_{RC}\omega}{\omega_{LC}^2}\right) \\ \phi_R &\rightarrow -\frac{\pi}{2} \end{aligned}$$

Figure 3.12 shows the Bode and phase plots for the gain across the resistor. As with the capacitor, the resonance curve becomes sharper as R becomes smaller, but the maximum gain remains 1 for all values of R.

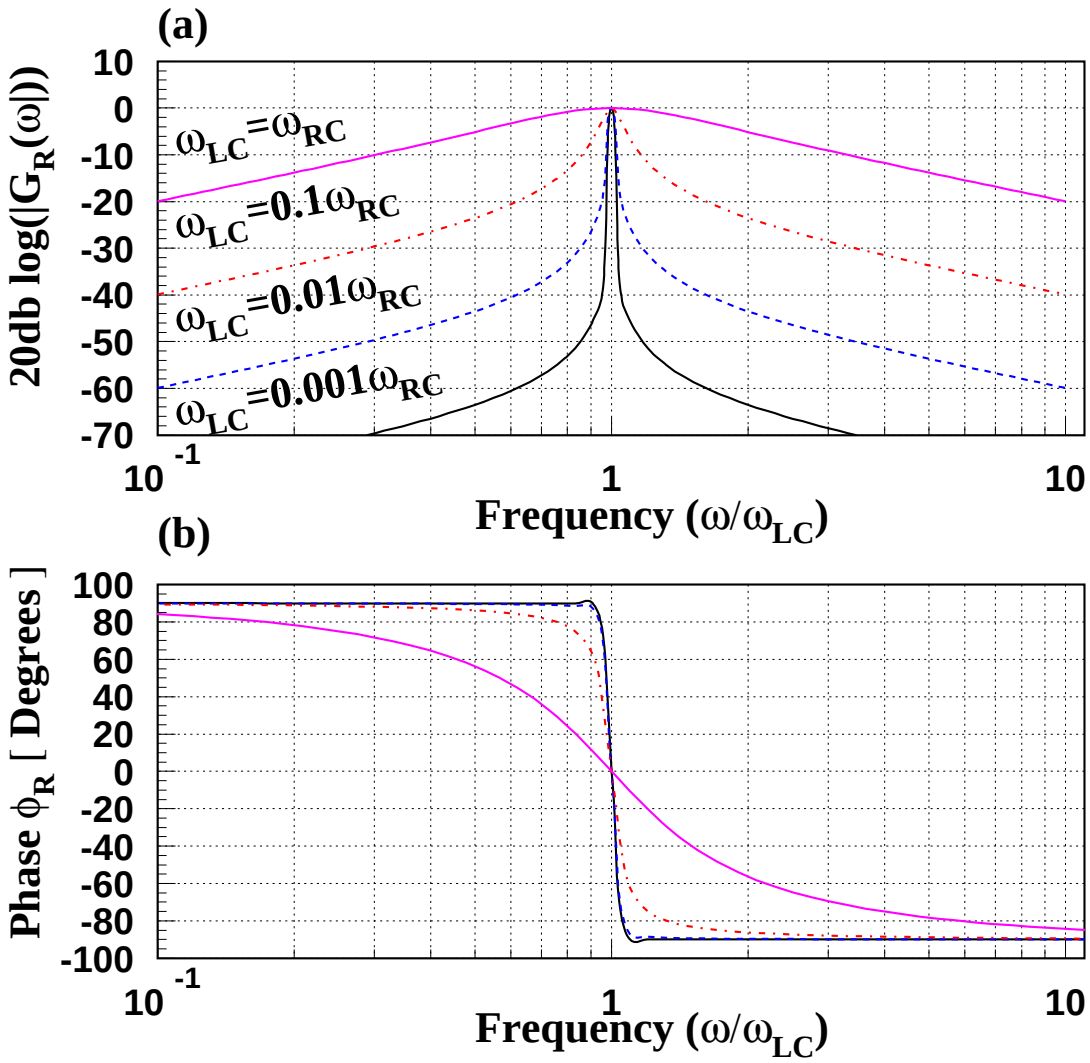

Figure 3.12: The Bode (a) and phase plot (b) for the gain measured across the resistor in the series RLC circuit. The four curves are for progressively larger ω_{RC} correspond to progressively smaller Rs. We see that the resonance gets sharper as R is decreased.

Finally, let us look at the gain across the inductor. While it is true that any real inductor will have some internal resistance associated with it, we can mathematically treat the case of a pure inductor. Under this assumption, we find:

$$\begin{aligned} G_L(\omega) &= \frac{j\omega L}{R - j\frac{1}{\omega C}(1-\omega^2/\omega_{LC}^2)} \\ G_L(\omega) &= \frac{1}{(1-\omega_o^2/\omega^2) - j\,(\omega_{RL}/\omega)} \end{aligned} \tag{3.23}$$

From this, we can get the magnitude and phase of the gain to be:

$$\mid G_L(\omega) \mid \;=\; \left[\frac{1}{\left(1-\omega_{LC}^2/\omega^2\right)^2+\left(\omega_{RL}/\omega\right)^2}\right]^{1/2}$$

and

$$\tan\left[\phi_L(\omega)\right] \;=\; \frac{\omega_{RL}/\omega}{1-\omega_{LC}^2/\omega^2}.$$

We can now examine the usual limits. The first is $\omega = \omega_{LC}$, where we find that:

$$\begin{aligned}\mid G_L(\omega_{LC}) \mid &= \frac{\omega_{LC}}{\omega_{RL}}\\ \mid G_L(\omega_{LC}) \mid &= \sqrt{\frac{L}{R^2C}}\\ \phi_L &= \frac{\pi}{2}.\end{aligned}$$

In the limit when $\omega \gg \omega_{LC}$, we find

$$\begin{aligned}\mid G_L \mid &\rightarrow 1\\ \phi_L &\rightarrow 0.\end{aligned}$$

Finally, in the limit where $\omega \ll \omega_{LC}$,

$$\begin{aligned}\mid G_L \mid &\rightarrow \left(\frac{\omega}{\omega_{LC}}\right)^2\\ \tan\left(\phi_L\right) &\rightarrow \frac{\omega\omega_{LR}}{-\omega_{LC}^2}\\ \phi_L &\rightarrow -\pi\end{aligned}$$

In Figure 3.13(a) we plot the gain and in (b) the phase as a function of $\log(f)$. We see that as R decreases, the resonance becomes sharper, and the gain at resonance increases.

Note that, while the gain of the circuit across individual elements may be larger than one, the sum of voltages across the three elements adds to the supply voltage, as it should. For example, at the resonant frequency,

$$\begin{aligned}v_R(t) &= V_s\cos(\omega_o t)\\ v_C(t) &= V_s\frac{\sqrt{L/C}}{R}\cos(\omega_o t-\pi/2),\end{aligned}$$

and

$$v_L(t) \;=\; V_s\frac{\sqrt{L/C}}{R}\cos(\omega_o t+\pi/2).$$

Due to the phase shifts between v_C and v_L, these two sum to zero and the sum of voltages has the same amplitude and phase as the source.

3.5.2 The LC-parallel RLC circuit.

Here, we have the parallel combination of L and C in series with R as shown in Figure 3.14. In this case, we can write the equivalent impedance of the circuit as Z_R in series with $Z_L \mid\mid Z_C$. The equivalent impedance of LC is

$$Z_{LC} \;=\; \frac{j\omega L}{1-\omega^2/\omega_{LC}^2}.$$

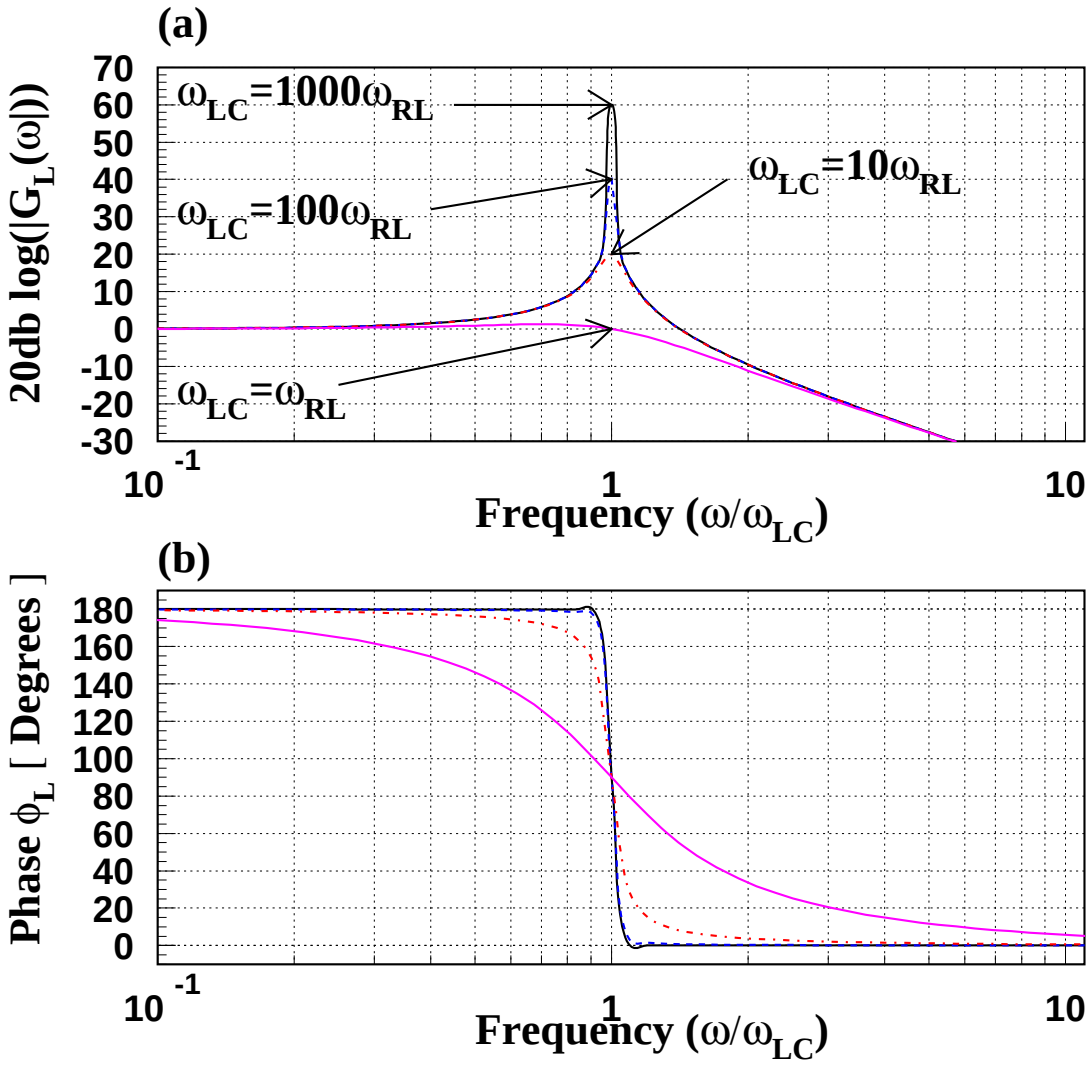

Figure 3.13: The Bode (a) and phase plot(b) for the gain measured across the inductor in the series RLC circuit. The four curves for progressively larger ω_{RC} correspond to progressively smaller R. We see that the resonance becomes sharper and the peak gain increases as R is decreased.

Using this, we write that the gain across the LC parallel combination is:

$$\begin{aligned} G_{LC}(\omega) &= \frac{Z_{LC}}{R + Z_{LC}} \\ G_{LC}(\omega) &= \frac{1}{1 - j\,(\omega_{RL}/\omega)\,(1 - \omega^2/\omega_{LC}^2)}. \end{aligned} \tag{3.24}$$

Thus,

$$| G_{LC}(\omega) | = \left[1 + (\omega_{RL}/\omega)^2 \left(1 - \omega^2/\omega_{LC}^2\right)^2\right]^{-1/2}$$

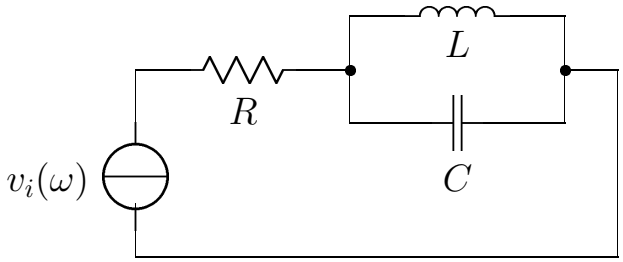

Figure 3.14: The LC-parallel RLC circuit.

and

$$\tan\left(\phi_{LC}(\omega)\right) = (\omega_{RL}/\omega)\,(1-\omega^2/\omega_{LC}^2).$$

Finally, we can look at the limits. When $\omega = \omega_{LC}$, we have:

$$\begin{aligned} |\,G_{LC}\,| &= 1 \\ \phi_{LC} &= 0 \end{aligned}$$

As we take $\omega \ll \omega_{LC}$, we find that:

$$\begin{aligned} G_{LC}(\omega) &\rightarrow \omega/\omega_{LR} \\ \phi_{LC} &\rightarrow \frac{\pi}{2}. \end{aligned}$$

In the limit where $\omega \gg \omega_{LC}$, we have:

$$\begin{aligned} |\,G_{LC}\,| &\rightarrow \omega_{RC}/\omega, \\ \phi_{LC} &\rightarrow -\frac{\pi}{2}. \end{aligned}$$

3.6 Driving Loads with Filters

If we build a high-pass or low-pass filter, we typically want to use the filtered output as input into some additional circuit. We can model this new circuit as a load impedance, Z_{LD}, connected between the output terminals of the filter. Figure 3.16 shows a low-pass filter driving such an impedance.

Because a filter can be considered a voltage divider, we can use its Thèvenin equivalent impedance to study the circuit's behavior when it is loaded. Recall that the Thèvenin equivalent impedance of a voltage divider is just the parallel combination of the two impedances in the divider. This will be the same for both high-pass and the low-pass filters, namely, Z_{out} is the parallel combination of Z_R and Z_C.

$$\begin{aligned} Z_{out} &= Z_C \,||\, Z_R \\ Z_{out} &= \frac{Z_C Z_R}{Z_C + Z_R} \\ Z_{out} &= \frac{R}{1 + j\omega/\omega_{RC}}. \end{aligned} \tag{3.25}$$

In order for the circuit not to be loaded down by Z_{LD}, we recall that $Z_{LD} \gg Z_{th}$, or in comparing to equation 3.25, we get the relation in equation 3.26 for the magnitude of the two impedances.

$$|\,Z_{LD}\,| \gg \frac{R}{\sqrt{1 + (\omega/\omega_{RC})^2}} \tag{3.26}$$

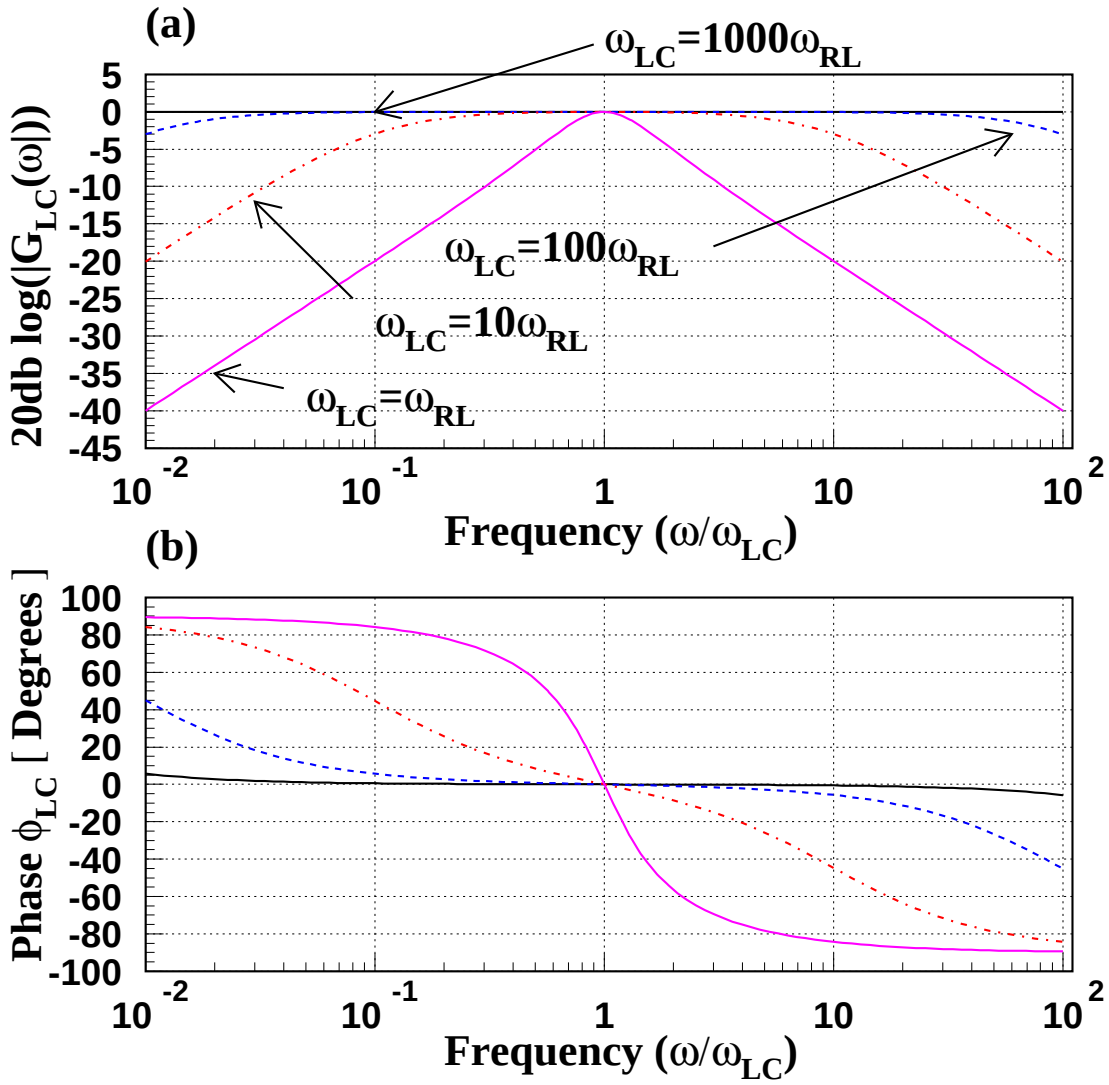

Figure 3.15: The Bode (a) and phase plot (b) for the gain across the parallel LC combination shown in Figure 3.10.

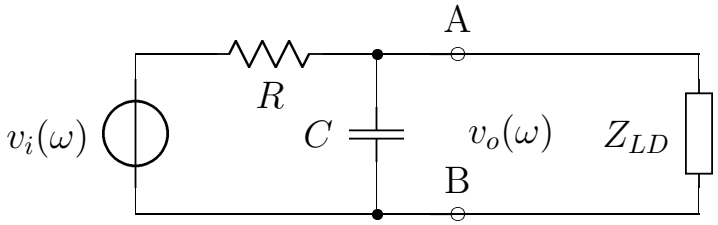

Figure 3.16: A low-pass filter driving a load impedance, Z_{LD}.

If we examine equation 3.26 in the case where $\omega = \omega_{RC}$, then we get that $\mid Z_{LD} \mid \gg R/\sqrt{2}$. Generally, we are operating in the limit of either $\omega \ll \omega_{RC}$ (low-pass) or $\omega \gg \omega_{RC}$ (high-pass). In the low-pass case, if $\omega \ll \omega_{RC}$, then the denominator of equation 3.26 becomes 1, and we get

$$\mid Z_{LD} \mid \quad \gg \quad R\,.$$

In the case of the high-pass filter, we look at the opposite limit. Here, the denominator of equation 3.26 becomes ω/ω_{RC}. Since $\omega_{RC} = 1/RC$, we see that the limit for Z_{LD} is

$$\mid Z_{LD} \mid \quad \gg \quad \frac{1}{\omega C} \qquad (\mathrm{high-pass})$$

We would like to use this to come up with the smallest load we can drive with this filter. In order to do this, we need to make some assumption about the frequency, ω. For the circuit to function as a high-pass filter, the natural choice is $\omega = \omega_{RC}$. With this, we can deduce that the load must satisfy

$$\mid Z_{LD} \mid \quad \gg \quad R/\sqrt{2}$$

an equation that can safely be applied for both the low-pass and the high-pass filters.

Example: We want to design a low-pass filter with a characteristic frequency of $f = 1590\,Hz$ which can drive a load of $\mid Z_{LD} \mid > 10\,k\Omega$. We are also told that the voltage supply that is used to drive the filter has an output resistance of $R_s = 50\,\Omega$. What are good values of R and C to choose for this circuit?

First recall that for the low-pass filter, the total resistance is $R + r_s$ (see equation 3.6). If start with our limit on the load impedance, we would like to choose $\mid Z_{LD} \mid \gg (R + r_s)/\sqrt{2}$. We can rearrange this to get $R + r_s \ll 14.1\,k\Omega$. We now need to decide how we want to treat the *much less*. If possible, a factor of 10 is a good choice. However, in some cases, we may be forced to choose a smaller number. Any choice that we make is going to have some consequence for the circuit. If we take a factor of 10, then we get that $R + r_s \approx 1.4\,k\Omega$ and $R \approx 1.35\,k\Omega$. We now want to choose C such that $1/RC$ will be ω_{RC}. This gives us

$$\begin{aligned} \frac{1}{(1350\,\Omega)C} &= 2\pi 1590\,Hz \\ C &= \frac{1}{1350\,\Omega} \cdot \frac{1}{10000\,s^{-1}} \end{aligned}$$

$$\begin{aligned} R &= 1.35k\Omega \\ C &= 74.nF \end{aligned}$$

3.7 Chaining Filters Together

Let us now continue to look at what happens when we try to connect sub-circuits together to perform a more complex function. We will use the filter circuits as an example, but the results will all turn out to hinge on Thèvenin equivalents, so the conclusions are general. These considerations are important when one wants to design circuits or sub-circuits to perform tasks in the lab.

3.7.1 The Band-pass Filter

Suppose we want to extract from our input signal the part that is located close to the characteristic frequency, ω_{RC}, while rejecting other frequency components. Based on what we have done so far, the natural thing to try is to hook a low-pass filter to a high-pass filter. Let us use the circuit shown Figure 3.17 and analyze the behavior of the combined circuits. In order to optimize the behavior, we

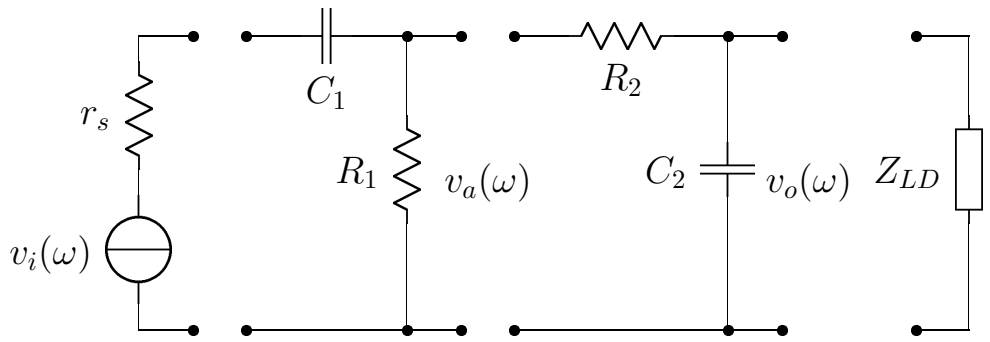

Figure 3.17: Connecting a low-pass filter to a high-pass filter to achieve a *band-pass* filter. We use the Thèvenin equivalent circuit for the source on the left. The next stage is the high-pass filter whose output is labeled v_a. This output is connected to the low-pass filter which, in turn, will be connected to some other circuitry represented by the complex load Z_{LD}. The voltage at the output of the low-pass filter is $v_o(\omega)$.

want for both filters to have the same characteristic frequency, ω_c. This gives that $R_1C_1 = R_2C_2$. A key element of this analysis will be using the Thèvenin equivalents of the various circuit elements. In this regard, the input and output impedances of the two filters are helpful.

To start, we would like to write that the voltage v_a at the point after the high-pass filter (the input to the low-pass filter) is:

$$v_a(\omega) = G_{hp}(\omega)v_i(\omega).$$

This is true under the condition that the effective load, Z_{eff}, is large compared to the output impedance of the low-pass filter, and, in turn, the input impedance of the low-pass filter is large compared to the output impedance of the high-pass filter. If we choose components such that this is true, then we can write that the voltage at the output of the low-pass filter will be:

$$\begin{aligned} v_o(\omega) &= G_{lp}(\omega)v_a(\omega) \\ v_o(\omega) &= G_{lp}(\omega)G_{hp}(\omega)v_i(\omega). \end{aligned} \tag{3.27}$$

If we now have the condition that Z_{LD} is large relative to the output impedance of the low-pass filter, then we may be able to achieve our goal of passing a limited frequency range. To summarize, we need to satisfy the following four conditions.

1. $R_1C_1 = R_2C_2$
2. $\mid Z_{in}^{hp} \mid \gg r_s$
3. $\mid Z_{in}^{lp} \mid \gg \mid Z_{out}^{hp} \mid$
4. $\mid Z_{LD} \mid \gg \mid Z_{out}^{lp} \mid$

Figure 3.18 shows the Bode plot and phase plot for the combined filter when the above conditions are satisfied. Figure 3.19 shows the gain plotted against $\log(f)$ on a linear scale.

We might note a few points about this combined filter. The maximum value of the gain is $\frac{1}{2}$ which occurs at the characteristic frequency, ω_{RC}. This comes from the fact that both the high-pass and the low-pass filters have a gain of $\frac{1}{\sqrt{2}}$ at ω_{RC}. Second, the phase is zero at ω_c, but for lower frequencies, it increases towards $\frac{\pi}{2}$, while for higher frequencies, it falls off towards $-\frac{\pi}{2}$. Finally, this band-pass filter lets through a range of frequencies. It essentially attenuates the gain at a rate of $20\,dB$/decade on either side of ω_{RC}.

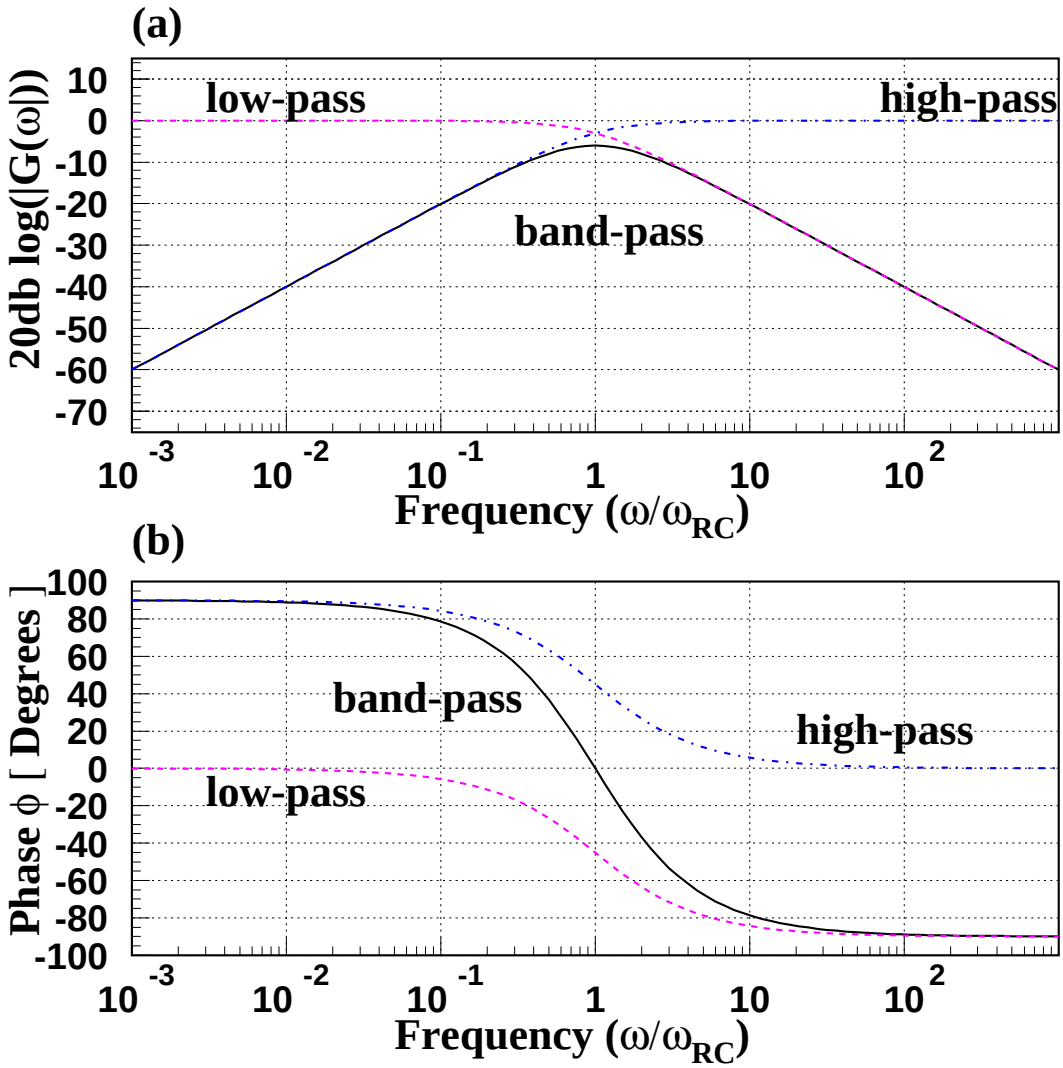

Figure 3.18: The Bode (a) and phase plot (b) of the *desired* response of the band-pass filter built from a combination of high-pass and low-pass filters. In the linear regions, the gain falls at 20dB per decade of frequency. The maximum value of the gain, at $\omega = \omega_{RC}$, is $\frac{1}{2}$, while the phase at that point is $0°$.

Example: Let us now look at a specific example. We take $\omega_{RC} = 2\pi(1000\,s^{-1})$, $(f = 1\,kHz)$. This means that we need $R_1C_1 = R_2C_2 = 1.6 \times 10^{-4}\,s$. Furthermore, we assume that the input source resistance is $r_s = 50\Omega$. The logic for determining actual values of the resistors and capacitors is as follows.

Since we want all of v_i to appear at the input to the high-pass filter, we want $r_s \ll \mid Z_{in}^{hp} \mid$. The input impedance of the filter is just Z_R in series with Z_C. This gives us that:

$$r_s \quad \ll \quad \mid R_1 \left(1 + \frac{1}{j\omega R_1 C_1}\right) \mid$$

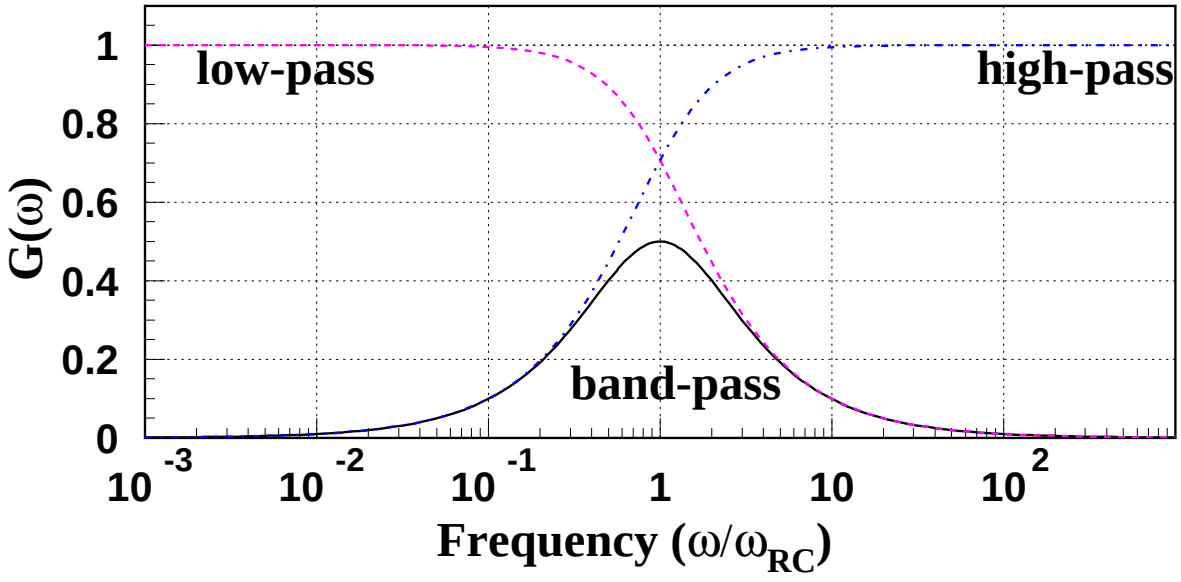

Figure 3.19: The same curves as in Figure 3.18, but with a linear vertical axis. This may improve your intuitive understanding of what is shown in the Bode plot (as well as in other Bode plots).

$$r_s \ll R_1\sqrt{\left(1+\frac{1}{(\omega/\omega_c)^2}\right)}$$

Evaluating this at $\omega = \omega_{RC}$, we find that:

$$\begin{aligned} R_1 &\gg \sqrt{2}\cdot r_s \\ R_1 &\gg 71\,\Omega \end{aligned}$$

Choosing $R_1 = 100 \cdot r_s = 5\,k\Omega$ should be quite safe. Having determined R_1, we can now calculate C_1 using the relation that $R_1C_1 = 1/\omega_{RC}$. This gives us

$$\begin{aligned} (5\,k\Omega)\cdot(C_1) &= \frac{1}{6280\,s^{-1}} \\ C_1 &= \frac{1}{5\,k\Omega}\cdot\frac{1}{6280\,s^{-1}} \end{aligned}$$

So that $C_1 \approx 0.03\,\mu F$. These values

$$\begin{aligned} R_1 &= 5\,k\Omega \\ C_1 &= 0.03\,\mu F\,, \end{aligned}$$

are convenient laboratory values.

We now have $v_a = G_{hp}v_i$, but *this is only true as long as there is no load on the high pass filter stage*. In other words, this is the *open-circuit* output voltage of the first stage.

To see what happens when we attach the low-pass circuit to the high-pass stage, we need the Thèvenin equivalent circuits for the output of the high-pass input of the low-pass, as shown in Figure 3.20. Again, we have a voltage divider in which we want all of v_a to appear across Z_{in}^{lp}. Therefore, we want $\mid Z_{out}^{hp} \mid \ll \mid Z_{in}^{lp} \mid$. The output impedance of the high-pass filter is just the parallel combination of Z_{R_1} and Z_{C_1}, while the input impedance of the low-pass filter is the series combination of Z_{R_2} and Z_{C_2}.

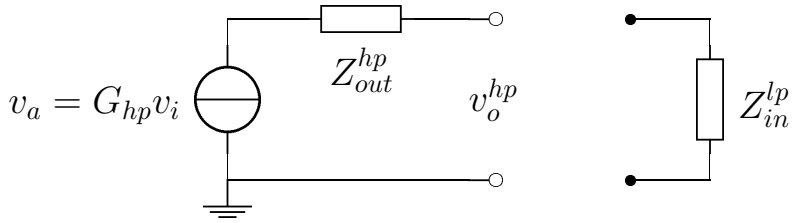

Figure 3.20: The equivalent circuits for the output of the high-pass and input to the low-pass filter. v_a is the open-circuit output of the high-pass stage. This stage has output impedance Z_{out}^{hp}. When the low-pass stage is connected, we are putting a load, Z_{in}^{lp} on the high-pass stage and, in general, the voltage at the input, v_o^{hp}, is no longer v_a.

Thus

$$\begin{aligned} |\frac{R_1}{1+j\omega R_1 C_1}| &\ll |R_2\left(1+\frac{1}{j\omega R_2 C_2}\right)| \\ |\frac{R_1}{1+j\omega/\omega_{RC}}| &\ll |R_2\left(1-j\frac{\omega_{RC}}{\omega}\right)| \\ R_1 &\ll R_2 |\left(1+j\frac{\omega}{\omega_{RC}}\right)\left(1-j\frac{\omega_{RC}}{\omega}\right)| \end{aligned}$$

If we take $\omega = \omega_{RC}$, then the previous equation simplifies to $R_1 \ll R_2 \cdot 2$. To satisfy this, we will choose $R_2 = 100 \cdot R_1$. Finally, by setting the characteristic frequency, we get the following values for R_2 and C_2.

$$\begin{aligned} R_2 &= 500\,k\Omega \\ C_2 &= 0.3\,nF. \end{aligned}$$

The circuit design is complete but we now we need to think about how well it will work. In particular, we have not yet looked at the load impedance, Z_{LD}. In order for the circuit to work as desired, we require that the load impedance is much larger than the output impedance of the low-pass filter: $|Z_{LD}| \gg |Z_{out}^{lp}|$. As we have seen, this means that $|Z_{LD}| \gg R_2$.

The value of $R_2 = 500\,k\Omega$ is uncomfortably large. Not only does the load we are driving have to have a very large input impedance, but with very large resistances, you have to worry about conduction through moisture, dust, and dirt. Furthermore, C_2 is uncomfortably small. Stray capacitance start to become important. If we had settled on a smaller factor than the 100 we used above, we would have had more comfortable component values, but we would have compromised signal quality. This kind of trade-off is found in essentially all circuit designs.

At the frequency we want to pass through the filter, $\omega = \omega_{RC}$, the gain is just $-3\,dB - 3\,dB = -6\,dB$ or $G = 0.5$. Unless the input signal is quite large, this could cause trouble—we'd like a circuit with an amplified output voltage rather than an attenuated one.

Finally, Z_{in} is dominated by the first stage (high-pass) and is reasonably large. However, if the input voltage were coming from a device with higher output impedance, we would quickly get into trouble with very large resistance values. It would be advantageous to have an *isolation stage* between the source voltage and our circuit. Such a device would have a gain of 1, a very large input impedance and a very small output impedance.

Example: Let us consider a slightly modified case for the band-pass filter shown in Figure 3.17. We will assume that both stages of the filter have the same characteristic frequency, ω_{RC}, and that the high-pass stage has $R_1 = R$ and $C_1 = C$. The low-pass stage will be assumed to have $R_2 = \alpha R$ and $C_2 = \frac{1}{\alpha}C$. We will now look at the behavior of the circuit at $\omega = \omega_{RC} = \frac{1}{RC}$. At the characteristic

frequency, we find that $Z_{C_1} = -jR$ and that $Z_{C_2} = -j\alpha R$. Using Kirchhoff's rules, or some other technique, we can find an exact expression for the gain of the two stages:

$$G_{hplp}(\omega_{RC}) = \frac{\alpha}{2\alpha + 1}.$$

Table 3.2 gives the gains for several values of α. When the low-pass stage has a large impedance, the circuit behaves as desired. If we reverse the order of the impedances, the gain will be dramatically reduced.

α	100	10	1	0.1	0.01
G_{lphp}	0.498	0.476	0.333	0.083	0.010

Table 3.2: The gain of the band-pass filter at ω_{RC} for several values of α, the ratio of R_2 to R_1.

We conclude that it is possible to connect circuit stages together to perform complex functions but that we have to be careful in such a design. The fundamental rule is that we want *low output impedances connected to high input impedances*. If we can arrange this and still have the circuit function, then we should sh the results we want.

As this example has shown, it's not easy to meet these requirements with the circuits we've studied so far. We will shortly develop an *impedance transformer* circuit in the form of a single transistor and also in the form of an integrated circuit called an operational amplifier. We will build amplifier circuits (transistor based and op-amp based) to boost signal amplitudes. And we will build filter circuits which have gain (that is, have $|G| > 1$).

3.8 Building Better Filters

In the last section, we examined chaining together a high-pass and a low-pass filter to achieve a simple band-pass filter. In doing so, we arrived at several impedance relations that would allow us to assume that the later stages of the filter did not load down the stages prior to them. We would no like to look at the effect of connecting several low-pass (or, equally, high-pass) filters together to achieve a sharper edge in the frequency fall-off. For simplicity, we will assume that we have managed to build the circuit shown in Figure 3.21, where the characteristic frequency of each stage is the same, ω_{RC}, and the component values are chosen such that the stages do not load each other down. Let us now examine the behavior of this circuit.

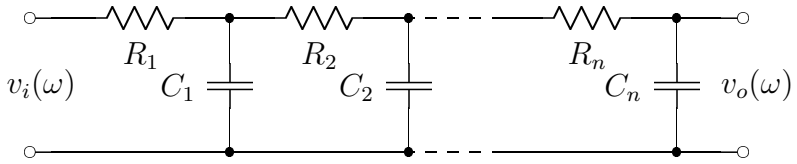

Figure 3.21: Connecting n low pass filters together to achieve a better filter.

We can start with the gain of the circuit. For a single low-pass filter, the gain is as given in equation 3.3. Additional stages will each have the same factor. This would give us that the gain of an n-stage filter is

$$G(\omega) = \frac{1}{(1 + j\omega/\omega_{RC})^n}. \tag{3.28}$$

We can get the magnitude of this gain by noting that every filter stage contributes a factor of

$$\frac{1}{\sqrt{1 + (\omega/\omega_{RC})^2}},$$

which gives that the magnitude of the gain of such a filter is

$$| G(\omega) | \quad = \quad \left[1 + (\omega/\omega_{RC})^2\right]^{-\frac{n}{2}} . \tag{3.29}$$

The Bode plots for several different values of n are shown in Figure 3.22. The gains are flat to a frequency near the characteristic frequency, after which they fall off at a rate of $n \times 20\,dB/$decade. In fact, at the characteristic frequency, it is easy to show that the curve is at the point $n \cdot -3\,dB$.

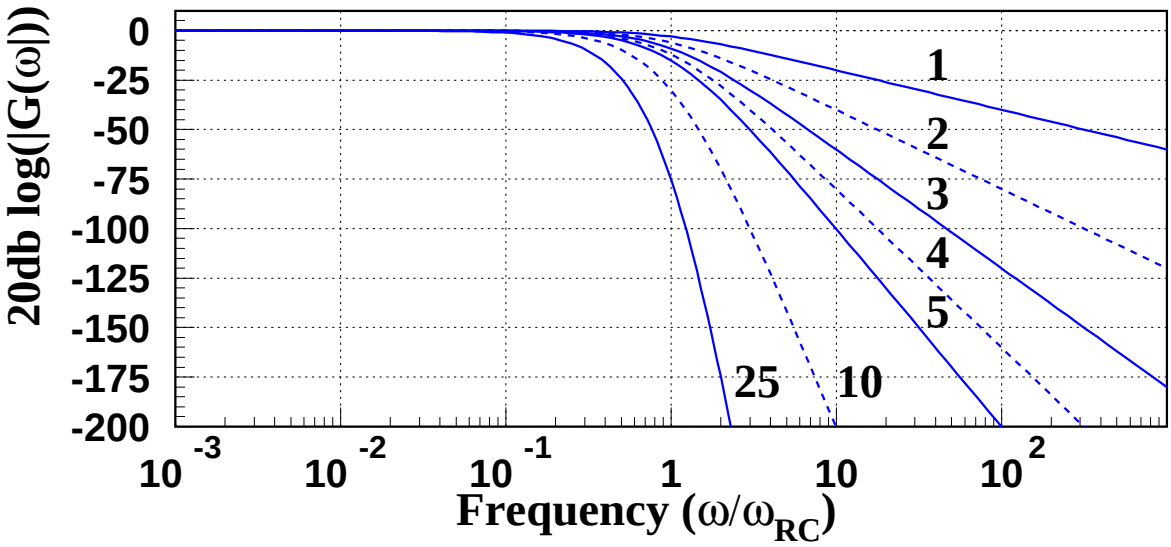

Figure 3.22: The Bode plots of several different multi-stage low-pass filters as per equation 3.29. The numbers next to the curves are the number of low-pass stages in the filter.

We can also compute the phase of the gain as a function of frequency. If the gain for $n = 1$ is

$$G(\omega) \quad = \quad | G | e^{j\phi} ,$$

then that for n stages can be expressed as

$$G(\omega) \quad = \quad \left(| G | e^{j\phi}\right)^n .$$

This can be simplified to yield

$$G(\omega) \quad = \quad | G |^n e^{jn\phi} .$$

Using the phase for $n = 1$ given in equation 3.5, we can plot the phase for several values of n. This is done in Figure 3.24 for up to $n = 4$.

Clearly, with multi-stage filters, we can make make the eventual slope of the frequency fall-off as steep as we like, but as shown in both Figure 3.23 and in the phase plot, Figure 3.24, the knee of the fall-off stays relatively smooth. We do not seem to be able to significantly sharpen this edge using only low-pass filters. If we refer back to some of the RLC circuits, there we saw extremely fast phase motion and an extremely sharp knee in some cases (e.g., Figure 3.11). The apparent conclusion here is that we need inductors and capacitors together to produce sharp edges in our filters. We will investigate this more carefully in the following section.

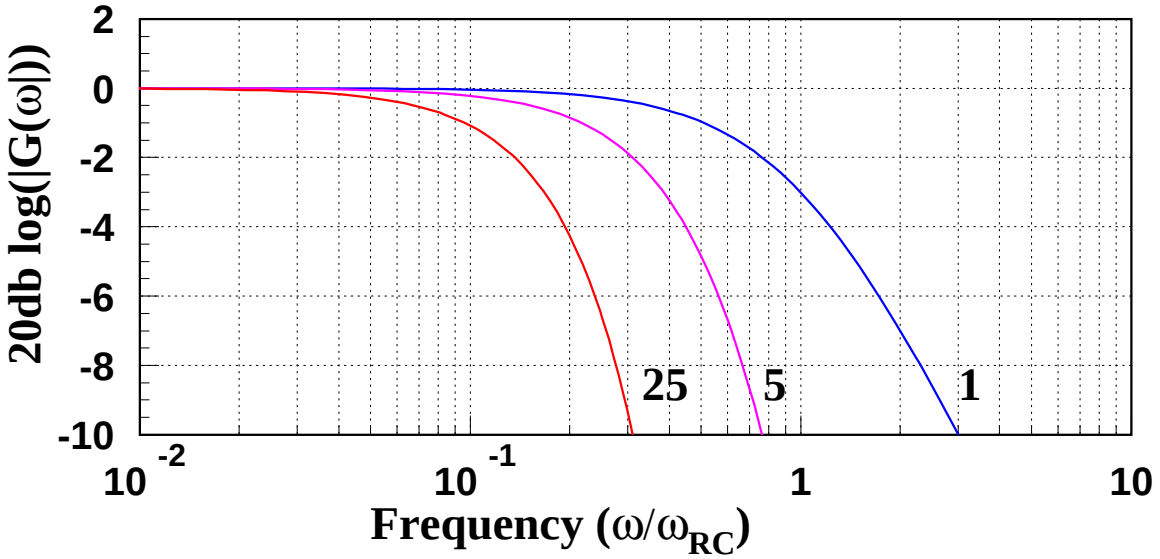

Figure 3.23: The Bode plots of several different multi-stage low-pass filters as per equation 3.29. This plot focuses on the knee in the frequency response. The numbers next to the curves indicate the number of low-pass stages in the filter.

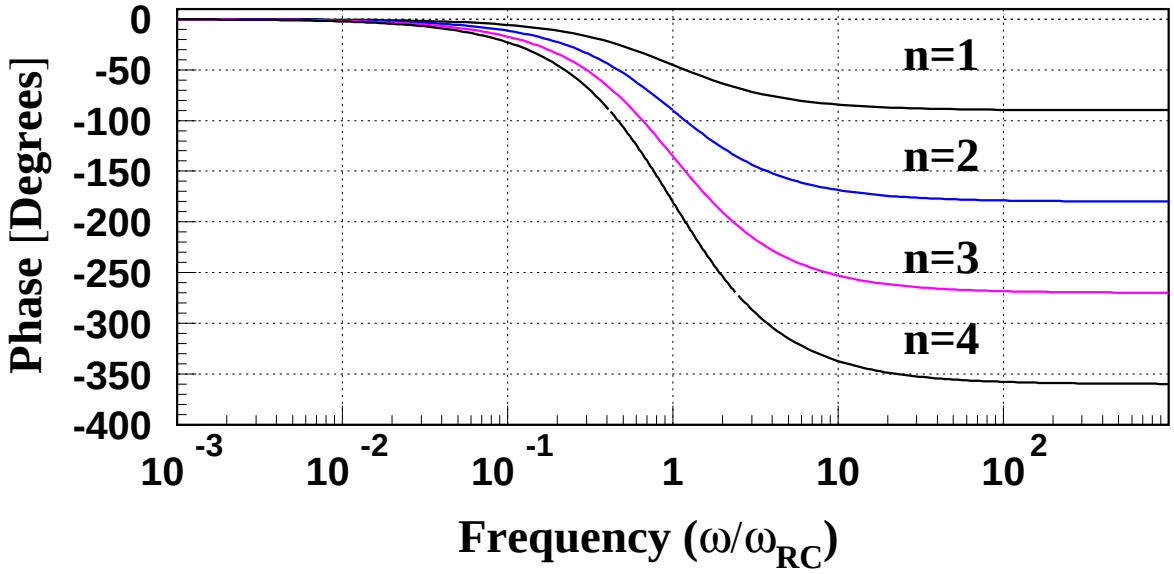

Figure 3.24: The phase plots of several different multi-stage low pass filters. The numbers on the plot indicate the number of low-pass filter stages, n.

3.9 Theoretical Considerations (Optional)

3.9.1 Complex Frequencies and Transfer Functions

Recall from section 2.7 that we could generate time dependencies for our voltages and current that were oscillatory (sines and cosines) and exponentially increasing or decreasing in time. Equation 3.30 shows

how all of this behavior can be combined to describe, say, a voltage. In this equation, both σ and ω are real numbers.

$$v(t) = V_0 e^{(\sigma + j\omega)t} \tag{3.30}$$

The exponential decay or rise comes from the σt, while the oscillatory behavior comes from the $j\omega t$ term. We can represent all of the voltages with which we have been working in this form.

Continuing along this path, we can define a *complex* frequency, s, such that $s = \sigma + j\omega$. This allows us to use a single parameter, s, to describe not only currents and voltages, but also such quantities as gain, $G(\omega)$.

In electronics theory, the gains we have studied are referred to as *transfer functions*. They can be generalized to couple voltage or current on one side to either voltage or current on the other, and generate a family of these functions. Note that gain is often referred to as $H(j\omega)$ in electronics:

$$H(j\omega) = G(\omega)$$

We will now generalize this to be a function of the complex frequency, s, so rather than $H(j\omega)$, we have $H(s)$. The next thing that we will note is that the transfer function can be written as the ratio of two polynomials, and these polynomials can be factorized as follows:

$$H(s) = A \cdot \frac{(s - s_a)(s - s_b)\cdots(s - s_m)}{(s - s_\alpha)(s - s_\beta)\cdots(s - s_\mu)} \tag{3.31}$$

Here, A is an overall scale factor while the $s_{a,\cdots,m}$ are the zeroes of the numerator and the $s_{\alpha,\cdots,\mu}$ are the zeroes of the denominator. Because this equation arises from one in which all of the coefficients are real, the roots of the polynomials must either be real or pairs of complex conjugates. This limits the cases that we need to examine in detail. As far as the transfer function goes, the zeroes of the numerator are zeroes of the transfer function, while the zeroes of the denominator cause the magnitude of the transfer function to go to infinity. These are referred to as *poles* of the transfer function. Any transfer function is fully described by its zeroes, its poles and the overall scale factor.

Example: Consider the low-pass filter with gain given in equation 3.3. Its transfer function can be written as:

$$H(s) = \omega_{RC} \frac{1}{s + \omega_{RC}}.$$

This transfer function has a single pole at $s = -\omega_{RC}$ and a scale factor of ω_{RC}.

Example: Similarly, the transfer function for the high-pass filter can be obtained from equation 3.8:

$$H(s) = \frac{s}{s + \omega_{RC}}.$$

This transfer function has a zero at $s = 0$ and a pole at $s = -\omega_{RC}$.

3.9.2 Graphical Representation

The poles and zeroes of a transfer function can be represented as points on the complex s plane. This is shown in Figure 3.25 for both the low-pass filter (a) and high-pass filter (b). We represent poles in the complex plane as solid circles and zeros of the transfer function as open circles.

We can use the complex plane to determine the magnitude and phase of a transfer function. In order to do this, let us consider the point on the negative real axis at $(-a, 0)$ as shown in Figure 3.26. This point can represent either a pole or a zero of a transfer function. To determine the magnitude an phase of the transfer function at a frequency ω, we draw a line from the point to ω on the positive imaginary

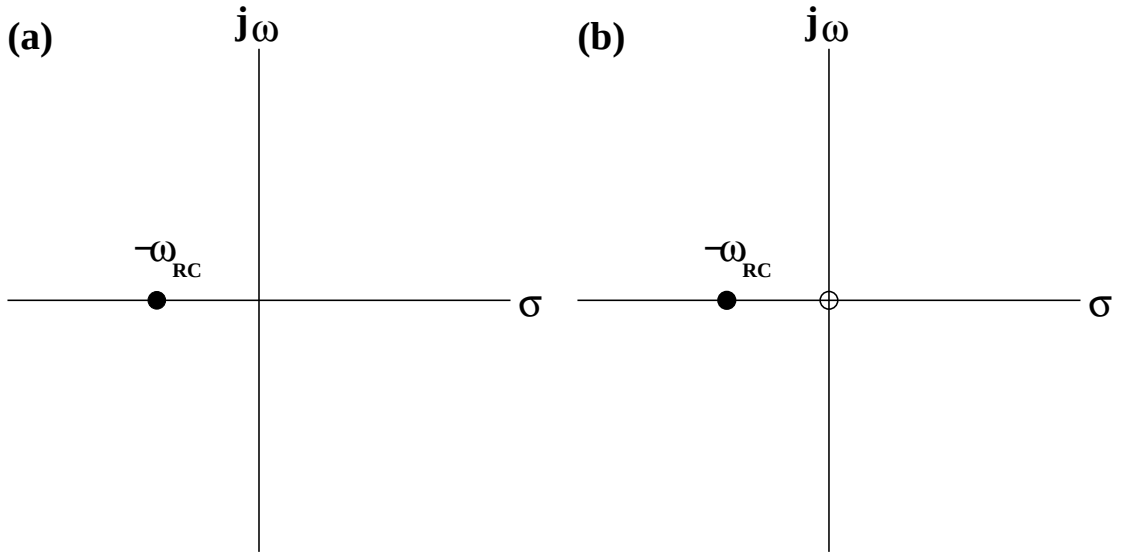

Figure 3.25: Representations of the low-pass (a) and high-pass (b) filters in the complex s-plane. Poles are represented by solid circles, while zeroes are represented as open circles. Both of these filters have a pole on the negative real axis at $-\omega_{RC}$, while the high pass filter also has a zero at $s = 0$.

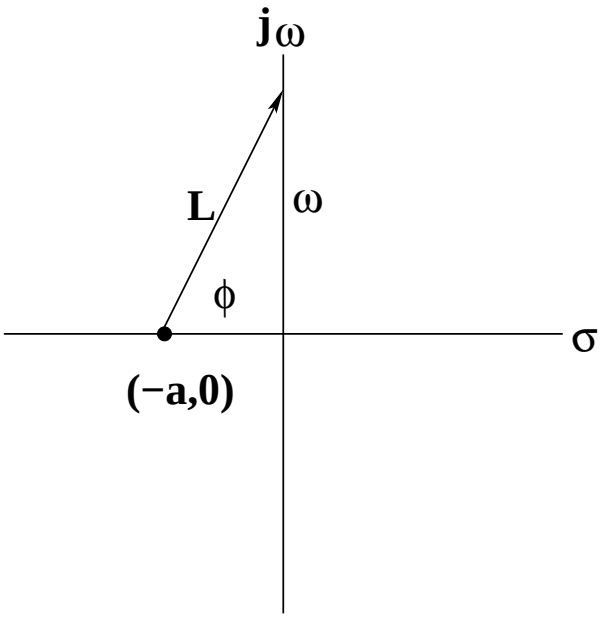

Figure 3.26: A graphical representation of the magnitude, L and angle, ϕ between two points in the complex s plane.

axis as shown. The angle ϕ is given by $\tan\phi = \frac{\omega}{a}$, while the length of the vector, $L = \sqrt{a^2 + \omega^2}$. If we had a zero of our transfer function at the point, then we have

$$G(j\omega) \quad = \quad A \cdot L \cdot e^{j\phi}$$

whereas if the point is a pole of the transfer function, then

$$G(j\omega) = A\frac{1}{L \cdot e^{j\phi}}.$$

If we have multiple poles and/or zeroes, then the total gain is just the product of the individual gains. This yields that the total phase is just the sum of all the phases of the zeroes minus the sum of all the pole phases, while the magnitude is the product of all the zero magnitudes divided by the product of all the pole magnitudes.

A Zero or a Pole at The Origin

Let us now consider the case where we have a single zero or pole at the origin. From our graphical representation, the angle from the origin to a point on the positive imaginary axis is always $\frac{\pi}{2}$—the phase angle a constant. Similarly, the length of the line is just ω. For a zero, we find that the gain is

$$\begin{aligned} G(\omega) &= \omega e^{j\frac{\pi}{2}} \\ G(\omega) &= j\omega, \end{aligned}$$

while for a pole, it is

$$\begin{aligned} G(\omega) &= \frac{1}{\omega}e^{-j\frac{\pi}{2}} \\ G(\omega) &= \frac{1}{j\omega}. \end{aligned}$$

These are plotted in Figure 3.27.

A Zero or a Pole on The Real Axis

If we have a single pole on the negative real axis at $-a$, then, from above, we have the following gains. For a zero, we have

$$\begin{aligned} G(\omega) &= \sqrt{a^2+\omega^2}\, e^{\tan^{-1}(\omega/a)} \\ G(\omega) &= a + j\omega \end{aligned}$$

while for a pole, we have

$$\begin{aligned} G(\omega) &= 1/\left(\sqrt{a^2+\omega^2}\right) e^{-\tan^{-1}(\omega/a)} \\ G(\omega) &= 1/\left(a+j\omega\right). \end{aligned}$$

These are plotted in Figure 3.28.

A Pair of Complex Poles or Zeroes

The remaining case of interest occurs when we have inductors, resistors and capacitors in our circuit. In such a case, the transfer functions factor to a pair of poles or zeros that are complex conjugates of each other. For zeroes, the transfer function is

$$H(s) = (s-s_a)(s-s_a^*),$$

and for poles,

$$H(s) = \frac{1}{(s-s_\alpha)(s-s_\alpha^*)}.$$

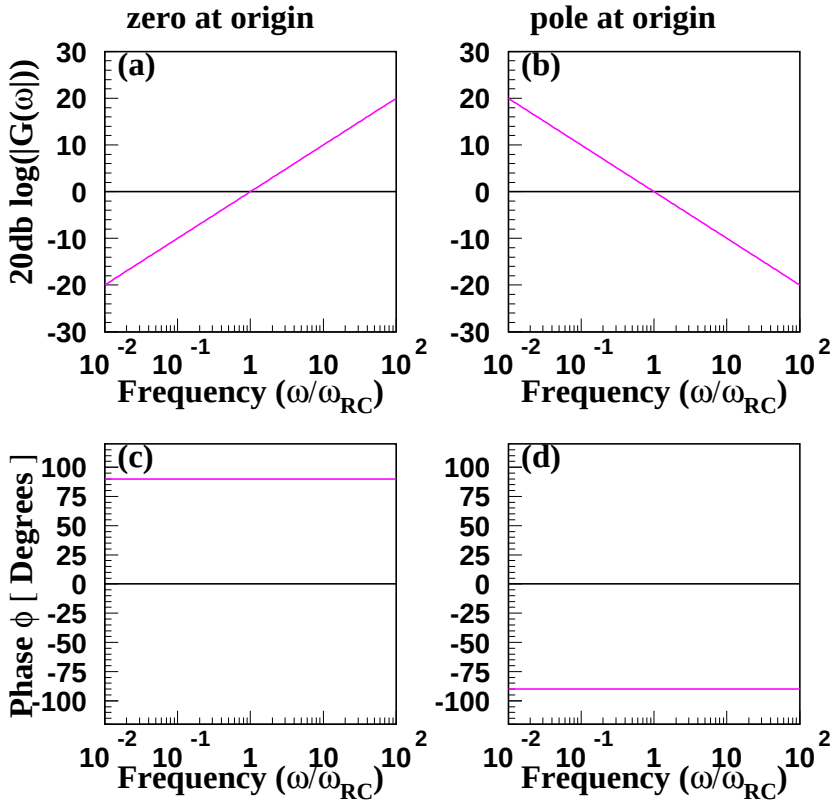

Figure 3.27: The Bode and phase plots for a zero (a),(c) and a pole (b),(d) at the origin.

Figure 3.29 shows s and s^* in the complex plane, and two vectors going to a point on the imaginary axis.

If s and s^* are zeroes, then the magnitude is just the product of the two lengths, while the total phase is the sum of the two angles. Note that at the point shown, the angle from the upper point is negative. At the origin, the total phase would be $0°$, as the two phases are opposite each other.

If s and s^* are poles, then the magnitude would be the inverse of the product of the amplitudes and the phase would be the negative of the sum of the phases. Figure 3.30 shows the Bode and phase plots for zeroes and poles, assuming that a is about 1% of ω_0.

This discussion applies to a circuit containing inductance and capacitance. The resonant frequency, ω_0, is $\frac{1}{\sqrt{LC}}$, while the distance away from the imaginary axis, a, is related to the resistance in the circuit. The smaller R is, the smaller a is, and the closer we are to having the two points on the imaginary axis. In such a limit, the resonance peak would be infinitely high. Beyond the resonance peak, the curve is increasing or decreasing by $40\,dB$ per decade.

As we have previously seen, filters containing only resistors and capacitors will not have sharp cutoffs. As the curves in Figure 3.30 show, the addition of inductors to our filters can give much sharper cutoffs (assuming our resistances are small).

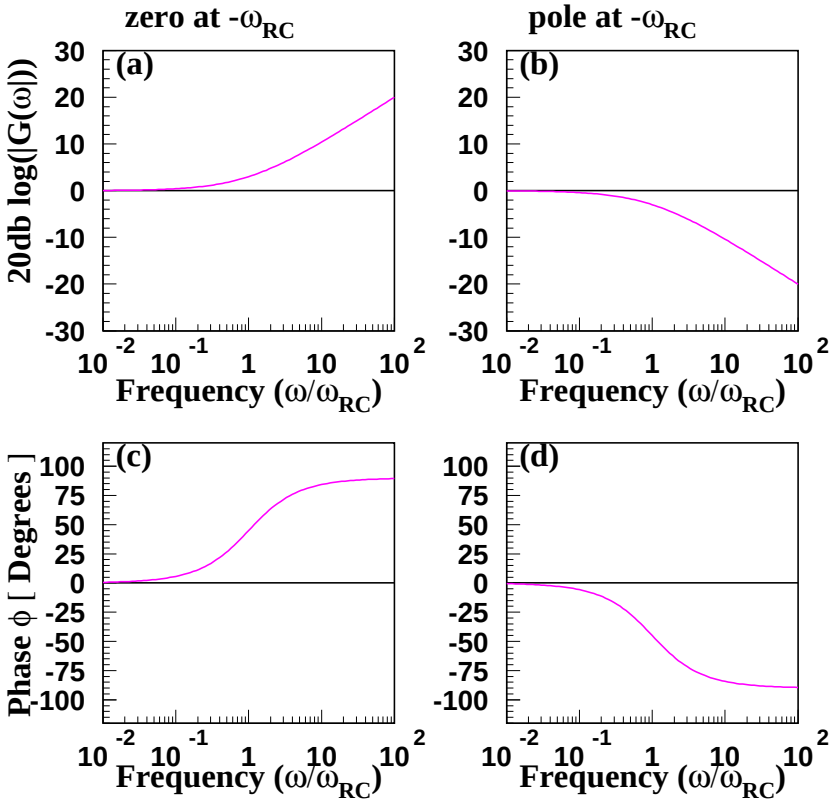

Figure 3.28: The Bode and phase plots for a zero (a),(c) and a pole (b),(d) on the negative real axis.

3.9.3 Pole-Zero Cancellation

Some systems have poles where we do not want them. A method to remove these is so-called pole-zero cancellation. We try to add a zero to the system at exactly the same place that we have an unwanted pole. In equation 3.31, this then causes the pole and the zero to cancel each other out. The details are beyond the scope of this book. However, qualitative understanding of the concept is useful.

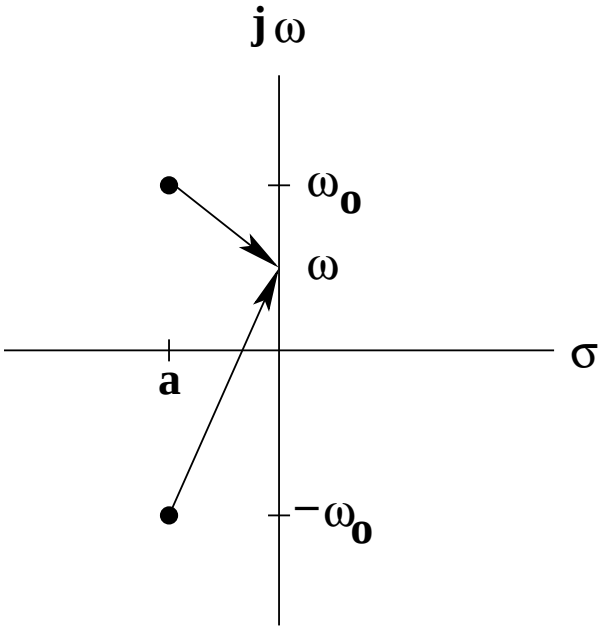

Figure 3.29: A pair of complex points, s and s^* in the s plane; one is at $-a + j\omega_o$, the other at $-a - j\omega_o$.

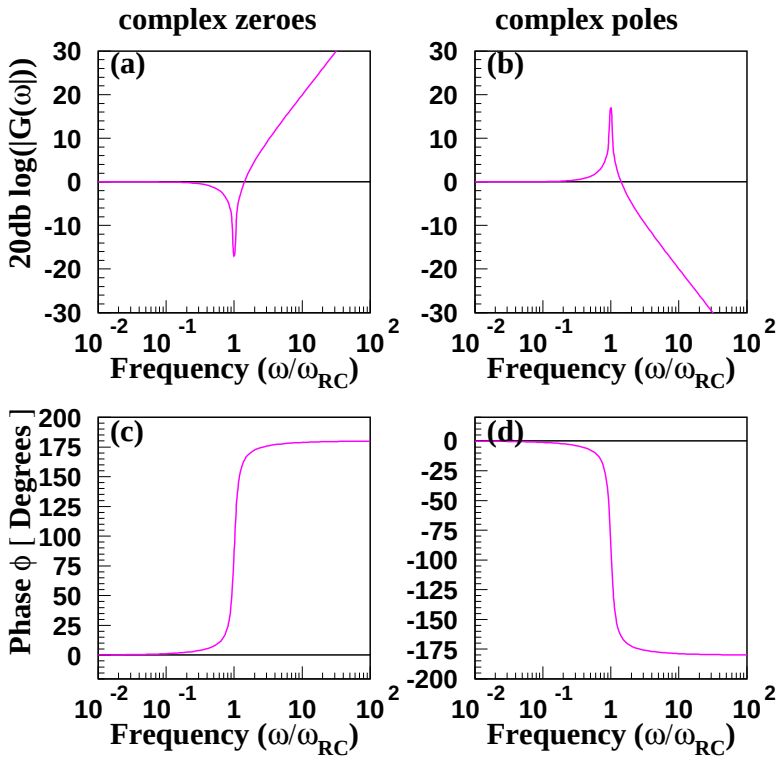

Figure 3.30: The Bode and phase plots for a pair of complex zeroes (a),(c) and poles (b),(d).

Problems

1. A low-pass filter is to be built using a resistor and a capacitor such that the circuit should attenuate an $\omega = 10000\,s^{-1}$ signal by a factor of 100. If the input impedance of the circuit should be $1\,k\Omega$ at $\omega = 100\,s^{-1}$, what values should be chosen for R and C?

2. A high-pass filter is to be built using a resistor and a capacitor such that the circuit should attenuate an $\omega = 100\,s^{-1}$ signal by a factor of 100. If the input impedance of the circuit should be $1\,k\Omega$ at $\omega = 10000\,s^{-1}$, what values should be chosen for R and C?

3. The circuit shown in Figure 3.31 is to be used as a filter. (**a**) Assuming that the inductor has zero resistance, what is the gain of the circuit as a function of ω? (**b**) Is this a high-pass or a low-pass filter? (**c**) What is the characteristic frequency of this filter? (**d**) What is the input impedance of the filter? (**e**) What is the output impedance of the filter?

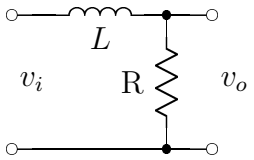

Figure 3.31: The circuit for problems 3 and 4.

4. The circuit shown in Figure 3.31 is to be used as a filter. (**a**) Assuming that the inductor has a non-zero internal resistance, R_L, what is the gain of the circuit as a function of ω? (**b**) Is this a high-pass or a low-pass filter? (**c**) What is the characteristic frequency of this filter? (**d**) What is the input impedance of the filter? (**e**) What is the output impedance of the filter?

5. The circuit shown in Figure 3.32 is to be used as a filter. (**a**) Assuming that the inductor has zero resistance, what is the gain of the circuit as a function of ω? (**b**) Is this a high-pass or a low-pass filter? (**c**) What is the characteristic frequency of this filter? (**d**) What is the input impedance of the filter? (**e**) What is the output impedance of the filter?

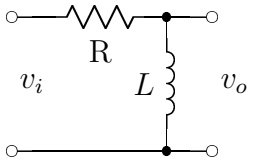

Figure 3.32: The circuit for problems 5 and 6.

6. The circuit shown in Figure 3.32 is to be used as a filter. (**a**) Assuming that the inductor has a non-zero internal resistance, R_L, what is the gain of the circuit as a function of ω? (**b**) Is this a high-pass or a low-pass filter? (**c**) What is the characteristic frequency of this filter? (**d**) What is the input impedance of the filter? (**e**) What is the output impedance of the filter?

7. Show that the filter shown in Figure 3.33 has a gain of

$$G = \frac{\omega L/R}{1 + (\omega L/R)^2}\left[(\omega L/R) + j\right] .$$

8. In problem 7 you found the gain of the RL filter shown in Figure 3.33. (**a**) What are the dimensions of the gain function? (**b**) Sketch the ratio $\frac{|V_o|}{|V_i|}$ as a function of $\log(\omega)$. Label relevant limits and, in particular, where the ratio is $\frac{1}{\sqrt{2}}$. (**c**) Sketch the phase difference, $\delta = \phi_o - \phi_i$, as a function of $\log(\omega)$ (ϕ_i is the phase of the input voltage, V_i, and ϕ_o is the output phase). Be careful to note if δ is positive or negative and to label relevant limits, in particular, at what frequency $\mid \delta \mid = \frac{\pi}{4}$. (**d**) You are asked to choose components for such a circuit so that the magnitude of the output voltage is 90% of the magnitude of the input voltage for a frequency of $f = 796\,Hz$. If you are forced to use an inductor of $L = 1\,mH$, what value of R should you choose?

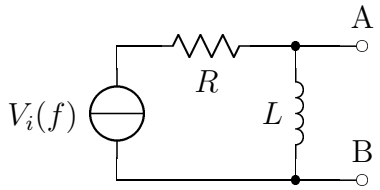

Figure 3.33: The figure for problems 7 and 8.

9. A simple circuit, as shown in Figure 3.34, is built from a voltage source with input voltage $v_{in}(t) = v_o e^{j\omega t}$ and two elements with impedance Z_1 and Z_2 connected in series. The output voltage, $v_{out}(t)$, is measured across Z_2 as shown in the figure. (**a**) What is the output voltage, v_{out}, in terms of v_{in}, Z_1 and Z_2? (**b**) Assume that Z_1 is an inductor with inductance L, and that Z_2 is a capacitor of capacitance C. Explain qualitatively why $v_{out} \to v_{in}$ for very low frequencies and $v_{out} \to 0$ for very large frequencies. (**c**) What is the gain of the circuit, $G(\omega)$, in terms of ω, L and C?
Most inductors have a small internal resistance r. If we were to account for this, we would arrive at the following gain function.

$$G(\omega) \quad = \quad \frac{(1-\omega^2 LC) - j\omega rC}{(1-\omega^2 LC)^2 + (\omega rC)^2}$$

What is the phase-shift, $\delta(\omega)$, in the following three cases: (**d**) $\omega = \omega_o = \frac{1}{\sqrt{LC}}$ (**e**) ω very small with respect to ω_o, and (**f**) ω very large with respect to ω_o?

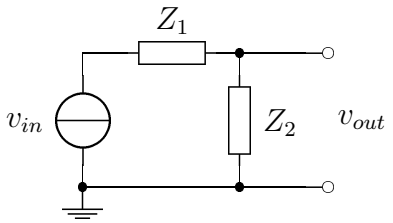

Figure 3.34: The circuit for problem 9 and 10.

10. Consider the LC filter in problem 9 (Figure 3.34). Show that if the inductor has an internal resistance r, that the gain is given as

$$G(\omega) \quad = \quad \frac{(1-\omega^2 LC) - j\omega rC}{(1-\omega^2 LC)^2 + (\omega rC)^2}\,.$$

11. You are given two RC filters as in Figure 3.35. One of these is a high-pass filter and the other is a low-pass filter, with gains given as follows:

$$G_{lp}(\omega) = \frac{1}{1+j\omega RC}$$
$$G_{hp}(\omega) = \frac{j\omega RC}{1+j\omega RC}.$$

(**a**) Explain qualitatively which of the filters is a high-pass filter and which is a low-pass filter. (**b**) For the low-pass filter, sketch $20dB \log \mid G(f) \mid$ as a function of $\log(f)$. Label relevant limits and, in particular, the $-3\,dB$ point. (**c**) For the low-pass filter, sketch the phase difference, $\delta = \phi_o - \phi_i$ as a function of $\log(f)$ (ϕ_i is the phase of the input voltage, V_i, and ϕ_o is the output phase). Be careful to note if δ is positive or negative and to label relevant limits, in particular, where $\mid \delta \mid = \frac{\pi}{4}$? (**d**) A friend from Electrical Engineering tells you that she has learned to build more complicated low-pass filters that have a high-frequency fall-off rate, in dB/decade, twice as large as the simple RC filter above. For her filter (at high frequencies), what is the fall-off (in dB/decade), and what is the f-dependence of her $\mid G(f) \mid$?

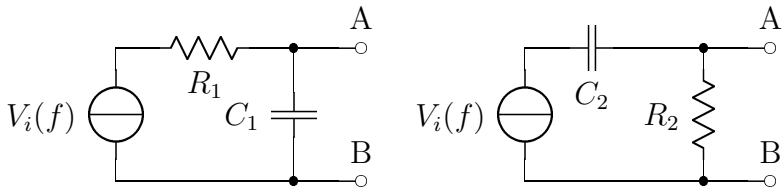

Figure 3.35: The circuits for problem 11.

12. The circuit shown in Figure 3.36 is constructed from two unknown linear components with impedances Z_a and Z_b. A time-varying voltage, $v_{in}(t) = v_o e^{j\omega t}$, is applied as shown. You measure an output voltage, $v_{out}(t)$, at the point shown in the figure. (**a**) In terms of Z_a and Z_b, what is the gain of the circuit? (**b**) You are told that Z_b is a resistor of value R, while Z_a is an inductor with inductance L and resistance r_L. In terms of R, r_L and L, what is the gain of the circuit, $G(\omega)$? (**c**) Is this circuit a high-pass or low-pass filter, (explain)?

13. Consider the circuit shown in Figure 3.36 where $Z_a = R + \frac{1}{j\omega C}$ and $Z_b = R$. (**a**) Show that the gain of the circuit can be expressed as:

$$G(\omega) = \frac{j\omega RC}{1+2j\omega RC}$$

(**b**) At what frequency, ω_o, does $20\,dB \log(\mid G \mid)$ go through the -3dB point? (Hint: What is the maximum value of the gain?) (**c**) Sketch $20 \log(\mid G \mid)$ versus $\log \omega$. (**d**) Sketch the phase of the gain as a function of $\log \omega$. (**e**) Is this circuit a high-pass or a low-pass filter?

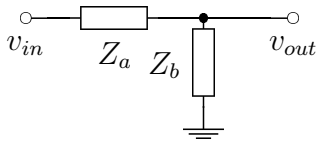

Figure 3.36: The circuit for problems 12 and 14.

14. Consider the modified RC low-pass filter shown in Figure 3.37. In the high-frequency limit, what is the gain of the circuit? Is this circuit a good low-pass filter?

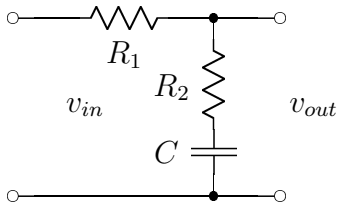

Figure 3.37: The circuit for problem 14.

15. Consider the modified RC high-pass filter shown in Figure 3.38. Show that the gain of the circuit is $R_2/(R_1 + R_2)$ times the gain of a normal high-pass filter with characteristic frequency $\omega_{RC} = \frac{1}{(R_1+R_2)C}$.

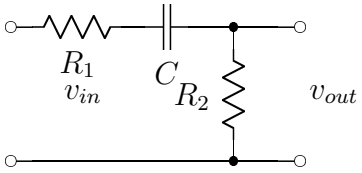

Figure 3.38: The circuit for problem 15.

16. Consider the RC filter in Figure 3.39, where the components are chosen such that $R_1C_1 = R_2C_2$. (**a**) Show that the gain of the circuit can be expressed as $G = \frac{R_2}{R_1+R_2}$. (**b**) Show that the gain of the circuit can be expressed as $G = \frac{C_1}{C_1+C_2}$.

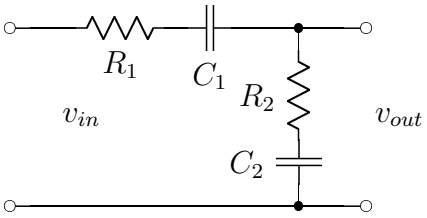

Figure 3.39: The circuit for problem 16.

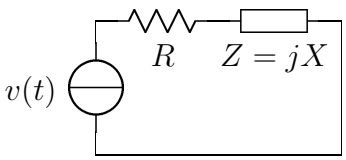

Figure 3.40: The circuit for problem 17.

17. Consider the circuit shown in Figure 3.40, in which an ideal voltage source provides a time-dependent input voltage, $v(t) = V_0 \cos(\omega t)$, to two elements in series. The first element is a resistor, R. The second element has a purely imaginary impedance $Z = jX$. (**a**) If Z were an inductor, L, what would be the value of X? (**b**) If Z were a capacitor, C, what would be the value of X? (**c**) Find the current, $i(t)$, flowing through the circuit. (You may give your answer as a complex wave or just the real part. But either way you should include the proper time dependence of $i(t)$.) (**d**) Find the phase difference between the current, $i(t)$, and the input voltage, $v(t)$. Under what

conditions will the current LEAD the voltage? (**e**) Find the potential difference, $v_Z(t)$, across the imaginary impedance $Z = jX$. Also find the magnitude of the gain, $|v_Z/v|$. (**f**) Find the **average** power dissipated in the resistor R. (**g**) Assume that Z is a capacitor of value C. At what frequency is the magnitude of the phase angle 45°? In the limit as the frequency gets very large, what is the magnitude of the phase angle?

Chapter 4

Introduction to Semiconductors

4.1 Introduction

In 1948, physicists John Bardeen, Walter H. Brattain and William Shockley of Bell Telephone Laboratories announced the invention of the transistor. It was also independently developed by Herbert Matarè and Heinrich Welker working at Westinghouse Laboratory in Paris. The semiconductors that underlie the functioning of the transistor have become nearly ubiquitous in the modern world.

In this chapter, we begin by discussing the physical principles underlying semiconductors. In particular we will examine the formation of energy bands and energy gaps. (This material is by no means complete and the interested reader should consult books on condensed matter physics for a detailed discussion of this topic.

We then move on to discuss the diode, one of the simpler semiconductor circuit elements. We also examine a small number of diode circuits. Again, the interested reader should consult a book such as Horowitz and Hill for a more complete listing of such circuits.

A discussion of diodes leads naturally to the bipolar transistor, which we will study in chapter 5.

4.2 Energy Bands in Materials

Let us look at a single atom consisting of a closed core of electrons around the nucleus and a single electron outside the core. We can talk about the electron moving in a potential about the positive charge at the center of the atom as shown in Figure 4.1. For the single electron, there are a set of solutions to the Schroedinger equation, $\psi_n(x)$. Each wave function, $\psi_n(x)$, has an energy eigenvalue of E_n, and for the electron in the nth state, the probability of finding the electron at some position x is $\psi^*(x)\psi(x)$. We can generalize this to more than one outer electron, though it is usually difficult or impossible to write down an analytic expression for ψ_n.

Now assuming that we have two such atoms with their closed cores centered at $\pm\frac{a}{2}$ as shown in Figure 4.2. We will assume that there is a single outer electron that can move between the two atoms. In the case of the two atoms very far apart, we have two possible solutions to the Schroedinger equation. If the electron is around the atom to the right, we would find $\psi_n(x-\frac{a}{2})$, while if the electron is around the left-hand atom, the solution would be $\psi_n(x+\frac{a}{2})$. In both cases, the electron would have energy E_n corresponding to the energy eigenvalue of the wave function ψ_n.

In reality, the above two solutions can only be approximate solutions for our atoms. In particular, our potential is symmetric about $x=0$ which means that the probability of finding the electron at x should be equal to that at $-x$. This means that our wave function must satisfy the relation that

$$\psi(-x) \quad = \quad \pm\psi(x)\,.$$

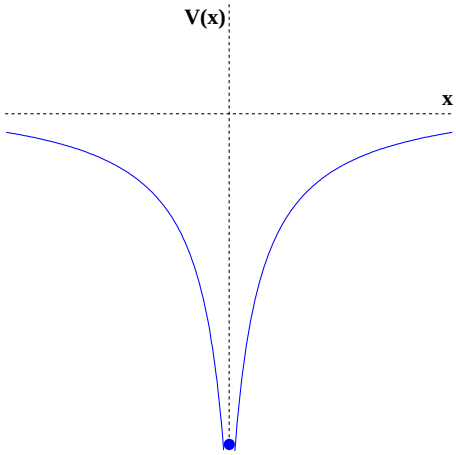

Figure 4.1: The electric potential around an atom.

In order to satisfy this condition, we can build two linear combinations of our above wave functions.

$$\psi_A = \frac{1}{\sqrt{2}}\left(\psi_n(x-\frac{a}{2}) + \psi_n(x+\frac{a}{2})\right) \tag{4.1}$$

$$\psi_B = \frac{1}{\sqrt{2}}\left(\psi_n(x-\frac{a}{2}) - \psi_n(x+\frac{a}{2})\right) \tag{4.2}$$

For the case when the two atoms are very far apart, the energy eigenvalues for these two solutions are equal to the energy of a single atom, E_n and are said to be *degenerate*.

$$E_A = E_B = E_n$$

If we now push the two atoms closer together by making a smaller, then we will eventually reach a point where the electron can easily hop back and forth between the two atoms. The wave functions above can still be used to approximate the the solutions, but the degeneracy between the two energies will be broken. The electric potential for such a configuration is depicted in Figure 4.2. Not only do the wave functions allow for the electron to be found near either of the two atoms, but the two degenerate energies split into distinct levels. One of these wave functions will have a higher energy than the other (Figure 4.3). The wave function derived from the sum (the symmetric wave function) has the lower energy, while the one arising from the difference (the antisymmetric solution) has the higher energy. In addition, each of these wave functions describes a situation in which the electrons are no longer associated with a single atom, but rather with both atoms.

Let us continue to add atoms, up to n, in an evenly-spaced one-dimensional array. We will assume n is quite large. As happened with two atoms, the n identical wave functions of the individual atoms mix. This results in n wave functions that are linear combinations of the products of the original ones. In addition, the n degenerate energy levels split into n new levels as shown in Figure 4.4. The spacing of the levels within the band scales like $1/n$ times the width of the band. For large numbers of atoms, this spacing can be quite small—$10^{-7} eV$—compared to the width of the band itself (which is on the order of $0.1\, eV$). This spacing is significantly smaller than the thermal energy, $k_B T$, of the atoms. Rather than distinct energy levels, we now have a band of allowed energy levels. Because of thermal motion, any energy within the band is allowed. We also have wave functions that are spread out over all of the atoms

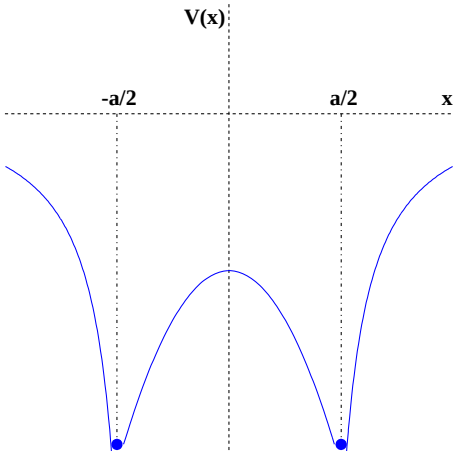

Figure 4.2: The potential well around two atoms that are close enough together for the electrons to hop back and forth between them.

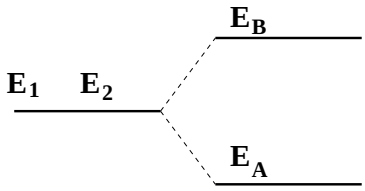

Figure 4.3: The two degenerate energies E_1 and E_2 split into two distinct energies. The higher energy, E_B, corresponds to the antisymmetric wave function, while E_A corresponds to the symmetric situation.

rather than being localized to individual atoms. The electron inside an energy band is in principle free to move within the band.

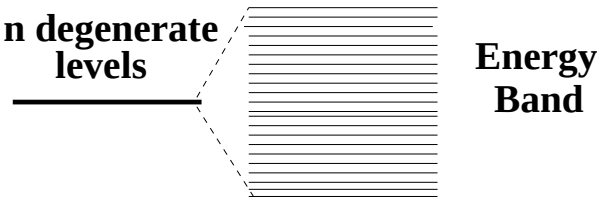

Figure 4.4: The n degenerate energy levels have split into n very closely spaced energy levels. This is referred to as an energy band. Typically the spacing between the individual levels is much smaller than the thermal energy, k_BT, of the atom.

The formation of energy bands leads to a shift from distinct levels to the formation of a band structure with gaps of forbidden energies. This is depicted in Figure 4.5. The nominal atomic levels are shown on the left. As the atoms come together in a solid, we eventually form the band and gap structure shown on the right. A key point here is that this happens to all the energy levels, even the ones that are not occupied by electrons. This leads to energy bands that are filled, bands that are partially filled,

and bands that are empty.

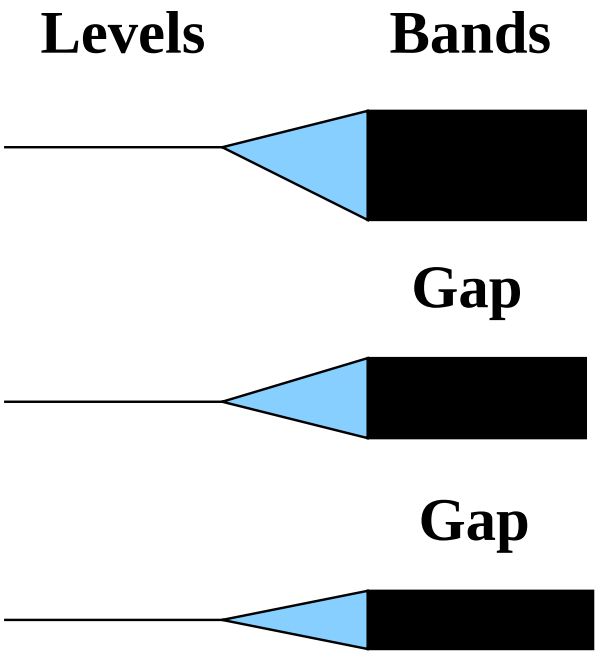

Figure 4.5: The splitting of distinct atomic energy levels into a band-gap structure.

4.3 Conduction in Materials

We recall from chapter 1 that the current density in a material is

$$\vec{J} = n \cdot q \cdot \vec{v}$$

where q is the charge of the current carrier, $\vec{v}$ is the average velocity of the carrier, and n is the number of charge carriers per unit volume. Implicit in this definition is that there are charge carriers, and that they are able to move through the material. Both of these conditions are satisfied if we have an energy band whose available states are not fully occupied with electrons. Such a band is known as a conduction band. The existence or nonexistence of such a band depends on the atomic, molecular and crystal structure of the material.

As we move through the elements, the number of electrons in the outer shells change. Materials like the noble gases have completely filled outer shells , while the alkali metals have single electrons in the outermost shell. However, for conduction we cannot talk about single atoms, but rather the solid form of the material. This means that not only are the atoms themselves important, but how the outermost electrons arrange themselves in the solid.

In metals, a partially filled band allows conduction. An example of this is shown in Figure 4.6, where the density of charge carriers (electrons) is plotted against the energy of the system. At zero temperature, the levels up to the Fermi Energy are filled, and those higher are empty, as indicated by the rectangular box extending to E_F in Figure 4.6. For a finite temperature, thermal excitation of atoms can provide energy for electrons to move into the empty levels. This is indicated by the curved line near E_F in Figure 4.6. Because there are empty, available levels in the band, electrons are able to move through the material. In particular, if we apply a potential difference across the material, electrons will move through the material with some non-zero average velocity, while the number of charge carriers corresponds roughly to the number of electrons in the band.

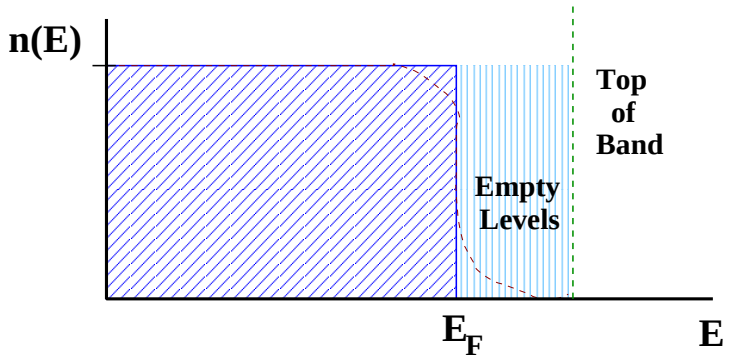

Figure 4.6: The filling of energy levels in some material. At zero temperature, all levels below the Fermi energy, E_F, are filled, and all those above it are empty. For finite temperature, thermal energy can excite electrons into the empty levels above E_F, thus depleting the levels just below E_F. This is indicated by the curved, dashed line near E_F.

For certain conducting materials, the conduction band may be nearly full. Figure 4.7 shows such a band. Plot (a) shows the occupancy of each level. If we apply a potential difference across the material, the charge carriers (electrons) will again flow through the material. However, we may well find that the conduction is not proportional to the number of electrons, n, but rather the number of empty levels—the *holes* into which the electrons can move. Figure 4.7(b) shows the energy levels in such a material. The conduction band is filled up to the Fermi energy and there is an energy gap between the top of the conduction band and the empty band above it.

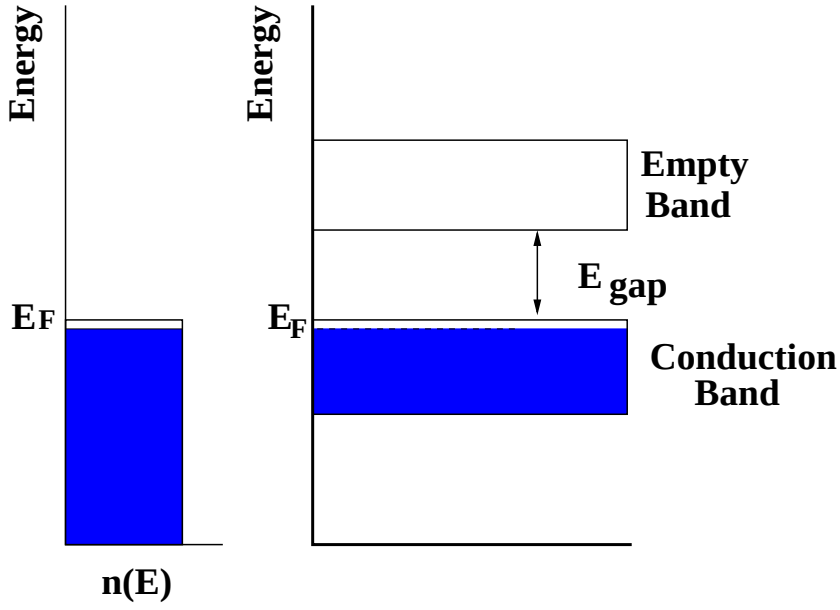

Figure 4.7: A nearly full conduction band in some material. Figure (a) shows the levels filled up to the Fermi energy, which is below the top of the band. (b) shows an energy diagram with the nearly filled conduction band, a energy gap, E_{gap} and then an empty band above that. The Fermi energy is inside the conduction band.

Figure 4.8 describes the two possible types of moving charge carriers. For the case where the conduction band is not nearly-full, we have the situation depicted in Figure 4.8(a). A potential difference

is applied to the material such that the bottom is more positive than the top. Electrons then move down through the material as shown in the time steps moving from left to right. In the case where the condition band is nearly full, we have the situation depicted in Figure 4.8(b). The holes into which the electrons can move are much more sparse, so as electrons flow through the material it actually appears as if the holes are moving in the opposite direction from the electrons. In this latter case, we can think of the conduction in two ways. The obvious is that there are some number of holes into which negative charge carriers (electrons) can move from top to bottom, or alternatively, that the holes behave like positive charges moving from bottom to top. Both of these views are equivalent, but we will find the latter more useful in the situations where the bands are nearly full. It appears that the conduction occurs with the holes moving from the more positive side to the more negative side of the material.

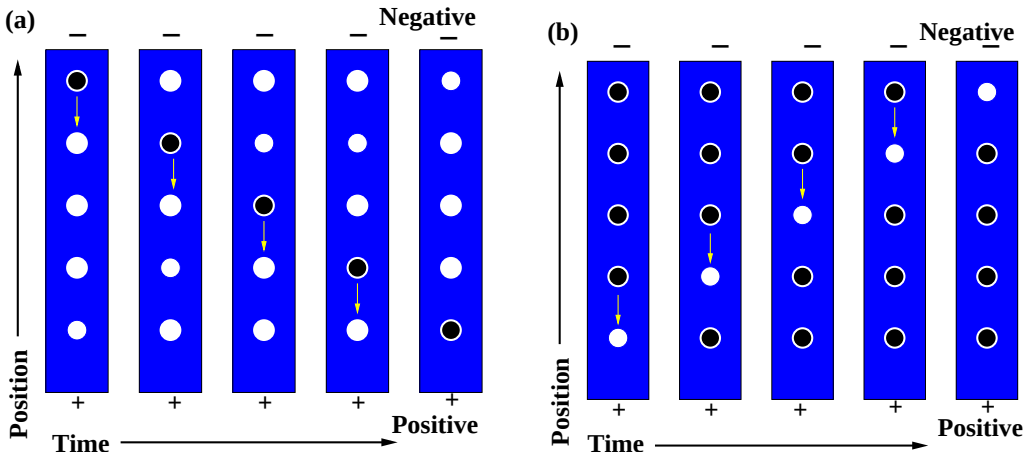

Figure 4.8: Conduction of electrons and holes through a solid. The electrons are shown as black circles filling the white holes. The bottom of the material is at a higher potential than the top. In (a), an electron moves from the top to the bottom of the material. In (b), the electrons move down to fill the available hole. The hole appears to move upward, so that positive charge flows from high potential to low.

We now consider a material in which all the levels in the conduction band are filled. In a solid, this can either be because the atom's shells are individually full, or because the bonds formed in making the solid have filled them. In such a case, there is no place within a band for an electron to move when an external voltage is applied. The material behaves as an insulator. However, it is possible for the material to conduct if we could either create holes in the filled bands (remove electrons), or cause some of the electrons to jump from the full band to the empty one above it. These bands are shown in Figure 4.9. In fact, if an electron were to somehow jump from the filled to the empty band, it would both leave a hole in the filled band, and provide an electron for conduction in the empty band.

4.4 Semiconductor Materials

Because the ability of a material to conduct depends on the density of charge carriers, n, we would like to know how often in an insulator we would find an electron-hole pair created. If the material is at some finite temperature T, then the relative probability of finding an electron in the empty band divided by the probability of finding it in the conduction band is:

$$\begin{aligned} P_E/P_F &= \frac{e^{-(E+E_{gap})/k_BT}}{e^{-(E)/k_BT}} \\ P_E/P_F &= e^{-E_{gap}/k_BT}. \end{aligned} \tag{4.3}$$

At room temperature, k_BT is approximately $1/40\,eV$. The band gaps, E_{gap}, for a few materials are given in Table 4.1. For crystal silicon, $E_{gap} = 1.1\,eV$. This yields a relative probability of about

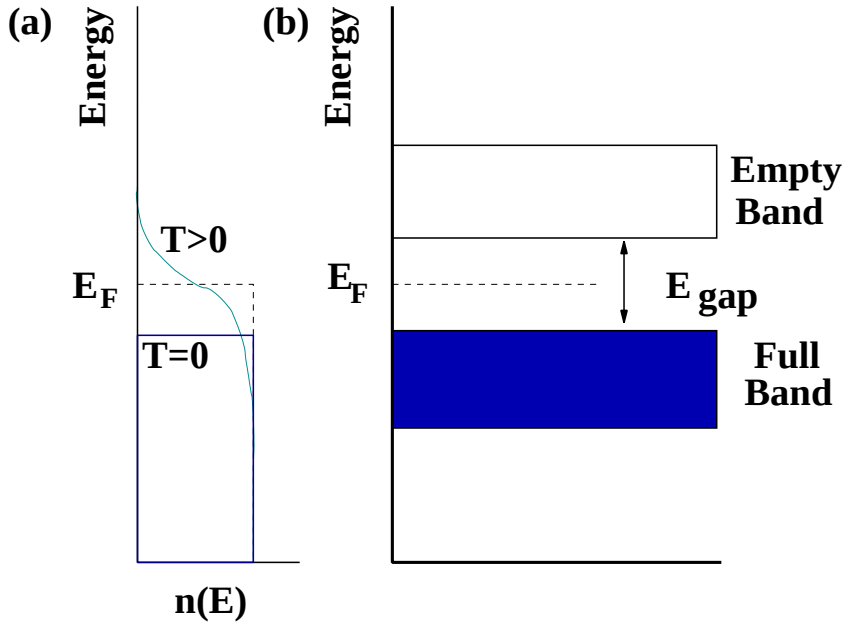

Figure 4.9: A material in which the lower energy band is completely full, and the upper one is empty with a gap between the two bands. The Fermi energy lies in the gap. At zero temperature, conduction will not be possible. At finite temperature, some electrons may be thermally excited into the empty band, which then allows conduction. This is indicated by the $T > 0$ occupation curve in the figure.

$e^{-40} \approx 4 \times 10^{-18}$, while a $2\,eV$ gap would give a relative probability of about 2×10^{-35}. In silicon, a very tiny number of electron-hole pairs are created at room temperature, while for materials with larger band gaps, there are nearly none. Materials that conduct because of this thermal excitation are referred to as semiconductors. In fact, the materials in Table 4.1 are all semiconductors. They are all materials

Material	Band Gap
CdS	$2.4\,eV$
Si	$1.1\,eV$
Ge	$0.7\,eV$
Te	$0.3\,eV$
InSb	$0.2\,eV$

Table 4.1: Band gaps of several semiconductor materials.

in which bonds formed between the atoms lead to a completely filled band. Effectively all possible bond sites are occupied. On a two-dimensional crystal lattice, one could visualize this using a square crystal structure. In such a picture, each atom has a bond to its nearest four neighbors, each of which is filled. This is shown schematically in Figure 4.10(a). The materials that make up typical semiconductors come from the part of the periodic table shown in Figure 4.10(b). The most common semiconductor material is silicon. It has four outermost electrons that can pair up with four adjacent atoms to created a filled energy band. Germanium is similar and is also a common semiconductor. Table 4.1 lists the band-gap energy of a few semiconductor materials including two materials that combine an element to with more electrons than silicon with another element fewer electrons than silicon.

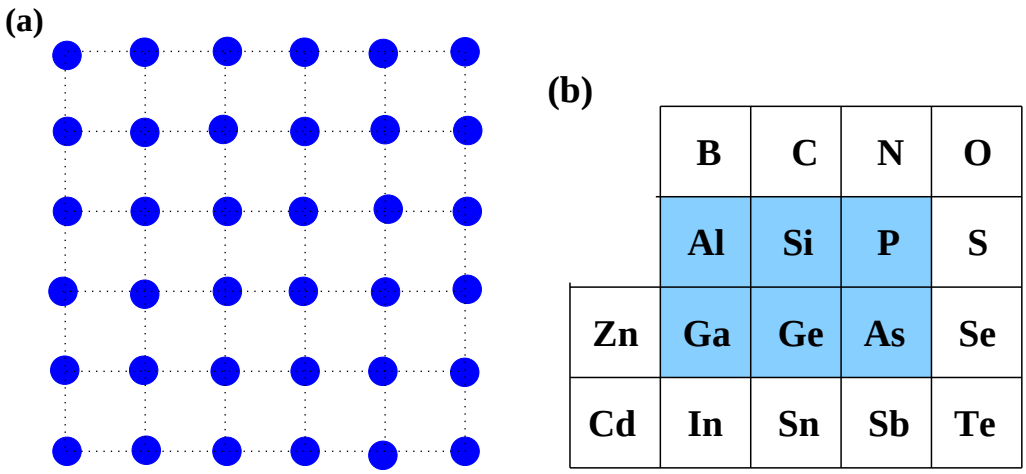

Figure 4.10: On the left is a schematic drawing showing the bonds between atoms in a crystal. The crystal has a completely filled energy band, which at zero temperature means that it is an insulator. However, at room temperature, sufficient electrons are thermally excited into the next band that the electrons in the excited band and the holes in the original band allow for finite conduction in the material. On the right is the part of the periodic table which the materials that make semiconductors are found.

4.4.1 Doped Semiconductors

The number of conduction of electrons in a semiconductor depends on the temperature and the band gap in the material. As we saw above, in most pure materials—even silicon–this number is too small to be useful. It becomes much more interesting when we no longer consider a pure semiconductor, but rather contaminate, or *dope*, the semiconductor with some other material. It is these doped semiconductors that are at the heart of all modern electronics.

Let us consider a semiconductor with a crystal structure as we saw in Figure 4.10. To this we mix in, or dope with, some small amount of an element with more electrons than the semiconductor (to the right of the material in the periodic table). For each of these doping atoms there will be an unpaired, and loosely bound electron associated with it. This is shown schematically in Figure 4.11(a) where the loosely bound electrons are shown as a circle around each of the scattered doping atoms.

The energy level diagram for this material is shown in Figure 4.11(b). The band gap is still there, as before, but we see the addition of the so-called *donor levels* to the diagram. It is in these levels that the loosely bound electrons are found. The key feature about these is that the energy gap between the donor levels and the empty conduction band is small compared to total band gap. Typical values are on the order of .02 to .03 eV, which are comparable to the thermal energy $k_B T$. Most of these donor electrons are actually thermally excited into the conduction band, and the material can conduct.

A doped semiconductor in which electrons are thermally excited into the conducting band is known as an *n-type* semiconductor. The negative charge carriers (electrons) are known as the *majority carrier* in an n-type semiconductor. The positive carriers (holes) are known as the *minority carrier*. While seemingly obvious, we emphasize that there are far more majority carriers than minority carriers in a doped semiconductor.

If instead of an element on the right-hand side of semiconductor material, we had chosen something from the left-hand side, the doped semiconductor would look like that shown in Figure 4.12(a) and be known as a *p-type* semiconductor. Here, there will be holes in the crystal structure into which an electron could go. These are indicated as the smaller atoms with circles around them. When we look at the energy level diagram, these holes appear as *acceptor levels* in the diagram. These are just slightly above the top of the full band, and again the energy gap to these levels is small. This means that most of these acceptor levels have an electron from the filled band thermally excited into them, leaving a

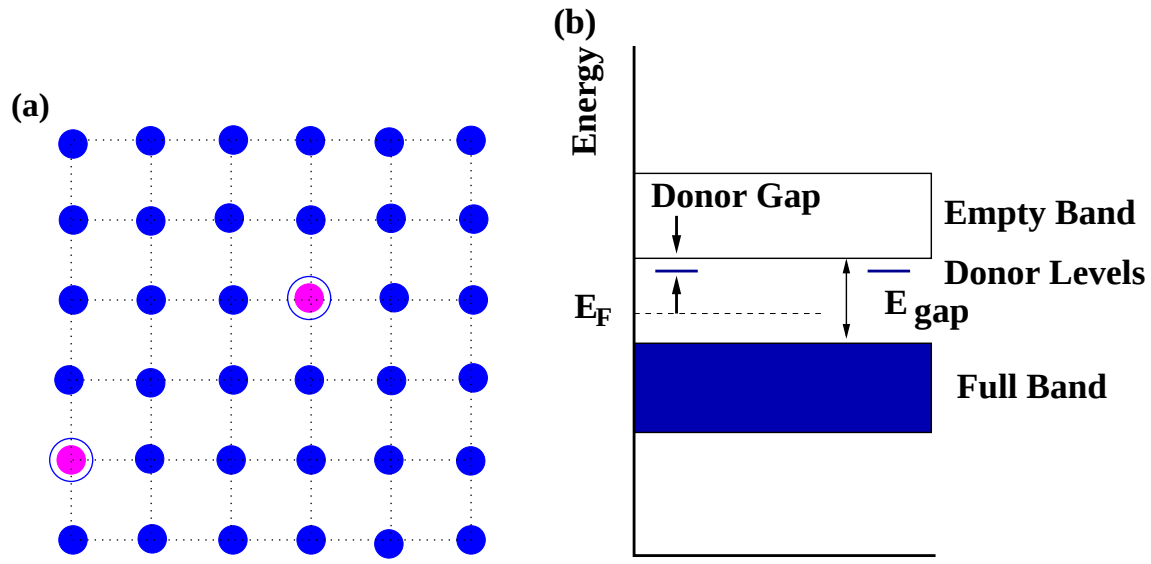

Figure 4.11: An n-type semiconductor. (a) shows a doped semiconductor in which the lighter colored atoms with circles around them represent an atom with an extra, or loosely bound electron. (b) shows the corresponding energy level diagram. The loosely bound electrons occupy the donor levels in the plot.

large number of holes in the formerly full band. This becomes a conduction band, with holes serving as the conductors. In a p-type semiconductor, the holes are the majority carrier and the electrons are the minority carrier.

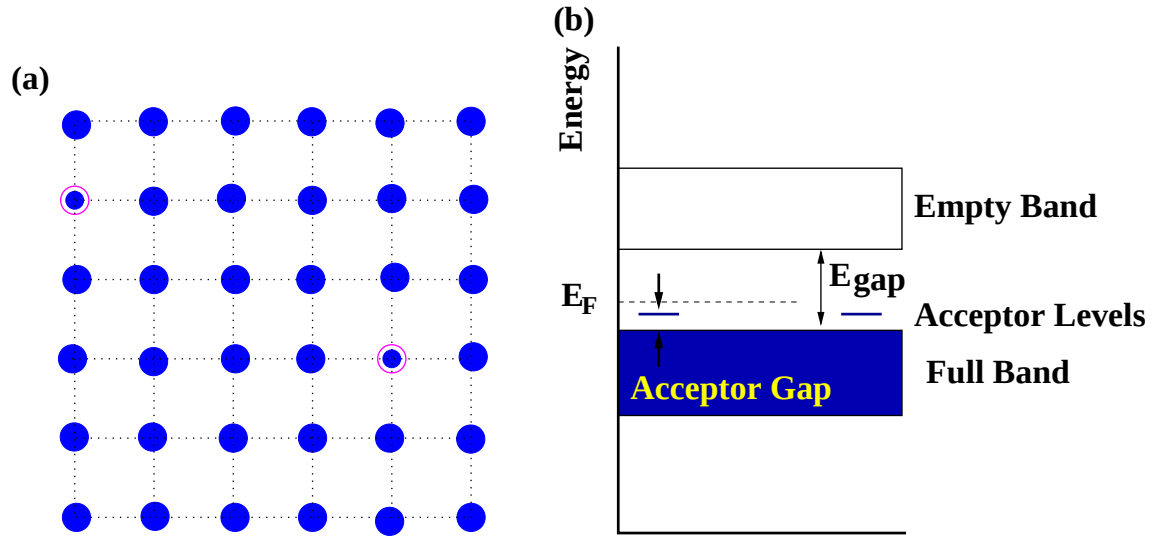

Figure 4.12: A p-type semiconductor. (a) shows a doped semiconductor in which the smaller dots with circles around them represent atoms missing an electron in their crystal structure. (b) shows the corresponding energy level diagram. The missing electrons correspond to the acceptor levels in the plot.

The conductance, $G = i/v$, in a semiconductor bar of material of cross-sectional area A and length

L is

$$G = \sigma\left(\frac{A}{L}\right) , \tag{4.4}$$

where σ is the conductivity of the material, which in a semiconductor, is

$$\sigma = q\left(\mu_e n + \mu_h p\right) . \tag{4.5}$$

The quantities n and p are the number of negative and positive charge carriers per unit volume. q is the electric charge of the charge carriers, and the μs are their mobility. This latter quantity effectively measures how easy it is to move the charges.

In doped semiconductors, conduction is dominated by one type of charge carrier. However, it is because both are present that these materials acquire some of their more interesting properties, as we will see later in this book.

4.4.2 The pn Junction

If we place an n-type material in contact with a p-type material, we create a *junction*, as shown in Figure 4.13(a). At such a junction, electrons will diffuse from the n-type material into the p-type as shown in Figure 4.13(b). As the diffusion occurs, the n-type material will acquire a net positive charge, while the p-type will acquire a net negative charge. This will set up an electric field from the n-type material towards the p-type material that will oppose additional diffusion, leaving the n-type material at a higher potential than the p-type, Figure 4.13(c). It is easy to see that this *pn-junction* can be used as a rectifier; indeed, it is the basis of the diode.

If we connect an external voltage source to the pn-junction as shown in Figure 4.14, then we can either increase or decrease the potential difference between the two sides of the junction. If we raise the potential of the n-type material relative to the p-type, we increase the potential difference between the two sides, making it even more difficult for current to flow. This is known as a *reverse-biased* junction. If we *forward-bias* the junction by raising the potential of the p-type material relative to the n-type, then we will lower the potential difference between the two sides. At some point current will start to flow.

Let us now try to be a bit more precise about this. If the external voltage is zero, some small number of electrons and holes still flow across the junction. However, the number of going from n-type to p-type equals the number going from p-type to n-type, so the net current flow is zero. If we consider the electrons, a few of them have sufficient thermal energy to move from the n-type to the p-type material. This current is known as the electron generation current, I_{g_e}. We also have electrons in the p-type material that diffuse to the boundary and are swept into the n-type material. This current is known as the electron recombination current, I_{r_e}. In order for the net electron current to be zero,

$$I_{g_e} + I_{r_e} = 0 .$$

We can make analogous arguments for the holes, leading to

$$I_{g_h} + I_{r_h} = 0 .$$

We first consider a reverse-biased junction. Assume that we apply some external potential V, raising the potential of the n-type material relative to the p-type material. The number of electrons that have sufficient energy to participate in the electron generation current is proportional to e^{-eV/k_BT}, which means that electron and hole-generation currents for some voltage $-V$ are as follows:

$$I_{g_e}(-V) = I_{g_e}(0)e^{-eV/k_BT} \tag{4.6}$$
$$I_{g_h}(-V) = I_{g_h}(0)e^{-eV/k_BT} \tag{4.7}$$

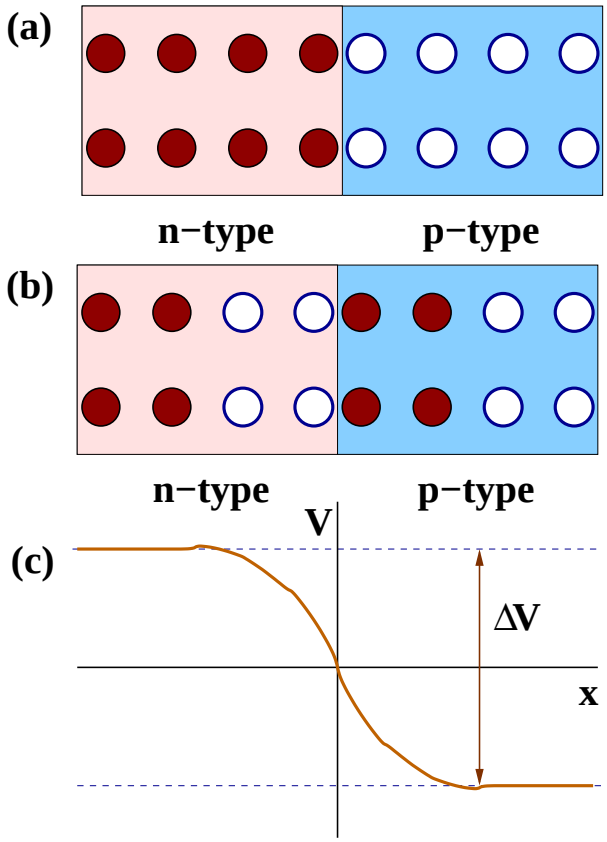

Figure 4.13: A pn junction (a) showing an n-type material in contact with a p-type material. In the n-type material, the charge carriers are electrons. In the p-type, they are holes. (b) shows the same junction in which some of the electrons from the n-type material have diffused into the p-type material. This sets up a potential difference, ΔV, across the junction, shown in (c).

Both of these quickly go to zero as $\mid -V \mid$ is made larger than $k_B T/e$. The regeneration currents arise from diffusion of charge, and do not depend on the voltage V. So, in reverse-biased operation, the current through the junction is

$$I_{reverse} \approx I_{r_e} + I_{r_h} .$$

If we now forward-bias the junction, then the generation currents in equations 4.6 and 4.7 become the following.

$$I_{g_e}(V) = I_{g_e}(0) e^{eV/k_B T} \tag{4.8}$$
$$I_{g_h}(V) = I_{g_h}(0) e^{eV/k_B T} \tag{4.9}$$

These quickly dominate the regeneration currents as V is made larger. If we define the saturation current, I_S, to be

$$I_S = I_{g_e}(0) + I_{g_h}(0) = -I_{r_e}(0) - I_{r_h}(0) ,$$

then we can write that the total current through the junction is

$$I(V) = I_S \left(e^{eV/k_B T} - 1 \right) . \tag{4.10}$$

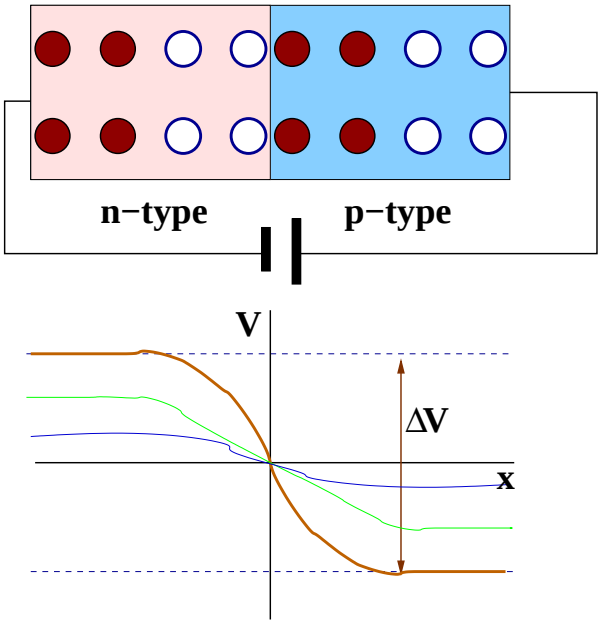

Figure 4.14: A pn junction to which an external voltage is used to bias the junction. The junction shown is forward-biased: the external voltage decreases the potential difference between the positive and negative sides. As this difference decreases, charge can flow through the junction. If the junction is reverse-biased, the potential difference between the n-type and p-type material increases.

The quantity $k_B T/e$ has dimensions of Volts. Recalling that the Boltzmann constant is $k_B = 1.38 \times 10^{-23}\, J/K$ and the electron charge $e = 1.6^{-19}\, C$, we can evaluate this factor as a function of temperature. For room temperature, $T = 293\, K$, we find that $k_B T/e = 24.3\, mV$. The exponential in equation 4.10 leads to large changes in current for small changes in voltage. Alternatively, if we consider some finite range of current, then the exponential gives rise to what appears to be a nearly constant voltage drop. We refer to this nearly constant voltage drop as the threshold voltage, V_{th}. Figure 4.15 shows some example I-V curves for currents in the range of a few hundred milliamperes,

If the forward bias voltage is less than V_{th}, then very little current can flow through the junction. For a larger forward bias across the junction, a large current can flow. The value of the V_{th} is controlled by both the temperature of the junction and the saturation current. Typical values of I_S for silicon-based junctions are in the range of 1 to $10\, pA$. As shown in Figure 4.15, this leads to a threshold voltage of 0.6 to $0.7\, V$ for diode currents on the order of a few hundred mA. Germanium-based semiconductors have threshold voltages on the order of 0.2 to $0.3\, V$, leading to saturation currents on the order of mA for similar diode currents.

4.5 Diodes

The pn-junctions which we have discussed are known as *diodes*, and, for the most part, Figure 4.15 is a very good representation of their I-V curves. The pin connected to the p-type semiconductor is known as the *anode*, while that connected to the n-type is the *cathode*. In circuits, we typically model diode behavior in a somewhat simpler fashion. The following two rules are very useful in analyzing diode circuits.

- If the forward bias applied to the diode is larger than V_{th}, then the current flowing through the diode is whatever it needs to be to make the voltage drop across the diode equal to V_{th}.

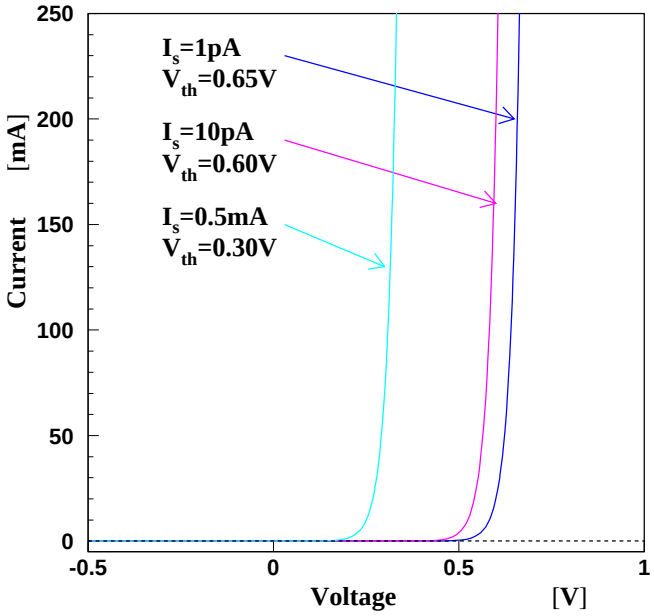

Figure 4.15: The I-V curve of a pn-junction as given by equation 4.10. We show the response for various values of the saturation current, I_S, for total currents in the range of a few hundred mA.

- If the forward bias applied is less than V_{th}, or the diode is reverse biased, no current flows through the diode and the voltage drop across the diode is the applied voltage.

Throughout the following, we will assume that we are using diodes with V_{th} of 0.6 to 0.7 V. We use the average value of $V_{th} = 0.65\,V$ for this, but it important to remember that this is only an approximation. The following example shows an application of this model to diode behavior.

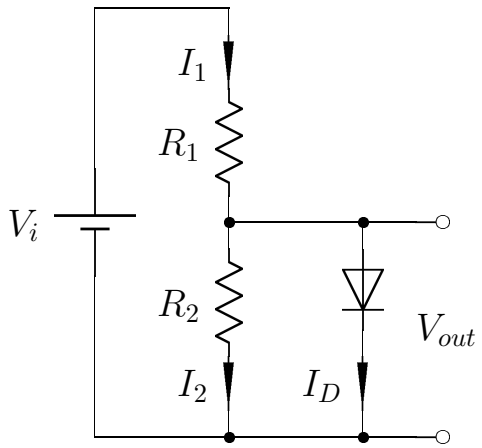

Figure 4.16: A diode attached across the output of a voltage divider.

Example: Consider the voltage divider shown in Figure 4.16. The input voltage is $V_i = 10\,V$ and the lower resistor has a value of $R_2 = 1\,k\Omega$. We examine two cases for the value of R_1: $19\,k\Omega$ and $9\,k\Omega$. We can determine the voltage, V_{out}, across R_2, and the current, I_2, through R_2 in each case. Let us start by ignoring the diode in the circuit. In this limit we have:

$$V_{out} = V_{in}\frac{R_2}{R_1 + R_2}$$

which leads to the following voltages for the two different values of R_1.

$$\begin{aligned} V_{out}^{(19)} &= 0.5\,V \\ V_{out}^{(9)} &= 1.0\,V \end{aligned}$$

We now put the diode back into the circuit. In the $19\,k\Omega$ case, the voltage across the diode will be $0.5\,V$. This voltage is smaller than the $0.65\,V$ needed to turn the diode on. As such, no current will flow through the diode. In this case, V_{out} will remain $0.5\,V$, and the current through R_2 will be $0.5\,mA$.

On the other hand, if R_1 is $9\,k\Omega$, then the initial voltage across the diode will be $1.0\,V$. This is larger than the $0.65\,V$ needed to turn the diode on. As such, the diode will turn on and current will start flowing through it $V_{out} = 0.65\,V$. If there is $0.65\,V$ across the diode, there will also be $0.65\,V$ across R_2. This means that $I_2 = 0.65\,V/1\,k\Omega = 0.65\,mA$. We know that there must be a $9.35\,V$ drop across R_1, which means that the current through R_1 is $I_1 = 9.35\,V/9\,k\Omega = 1.04\,mA$. The current through the diode is then $I_D = I_1 - I_2 = 0.39\,mA$.

4.5.1 Limitations on Diode Voltages

In a forward-biased diode, current flows to keep the voltage drop across the diode at the nominal diode drop, $\approx 0.65\,V$. Of course, there is a limit to how much power the diode can handle before it is damaged. In a reverse-biased diode, we do see a small reverse current, $I_{reverse}$. The number of charge carriers available is fixed. However, if the voltage is increased, the speed of these carriers can increase. At some point, they acquire sufficient energy to knock electrons out of the lattice. This produces both conduction electrons and holes, which in turn allows a much larger current to flow. The voltage at which this current starts to flow is known as the breakdown voltage and this effect is called avalanche multiplication.

A second reverse-bias effect can also occur in diodes. In this case the electric field in the junction layer can become so strong that it can dislodge electrons directly from their covalent bonds. This process is known as *Zener breakdown*. Both Zener breakdown and avalanche multiplication produce the same effect: a rapid increase in reverse current once the breakdown voltage is exceeded. Figure 4.17 shows the I-V curve of a diode including the breakdown voltage and large increase in the reverse current.

Diode breakdown at larger reverse voltages is typically due to avalanche multiplication, while lower-voltage breakdowns are usually Zener breakdown. If one is building a circuit that relies on the diode being reverse-biased, it is necessary to make sure that the breakdown voltage is not exceeded.

4.5.2 Zener Diodes

From the I-V curve in Figure 4.17, we can see that the voltage is nearly independent of current for diodes in the breakdown region. *Zener diodes* are designed to exploit this property, and are often used in circuits to provide a voltage reference. They are also referred to as *voltage reference diodes*, and are available in a large range of breakdown voltages from a few volts up to a several of hundred volts. Even though they are referred to as Zener diodes, they can break down by either the Zener effect or avalanche multiplication. Figure 4.18 shows the symbol for a $5.6\,V$ Zener diode. Note the symbol is slightly different than that for a normal diode. Also note that the breakdown voltage is indicated on the symbol.

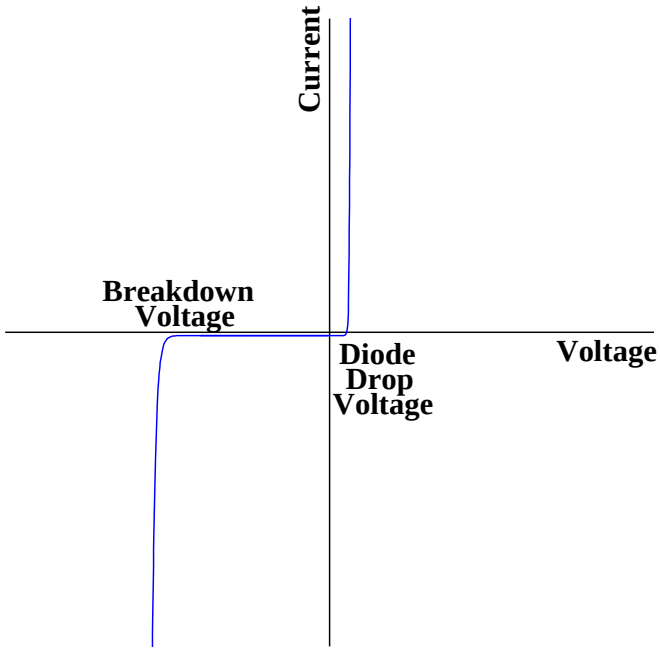

Figure 4.17: A diode I-V curve showing both the forward biased region with the normal diode drop and the reverse biased region with the breakdown voltage.

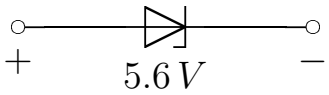

Figure 4.18: A $5.6\,V$ Zener diode with its breakdown voltage labeled in the circuit.

Example: Let us return to the voltage divider example we worked out earlier (see Figure 4.16). However, rather than a normal diode, we will use a $5.6\,V$ Zener diode, as shown in Figure 4.19. The input voltage is $V_i = 11.2\,V$ and the upper resistor has a value of $R_1 = 1\,k\Omega$. If R_2 is smaller than R_1, then the device functions as a voltage divider. However, if R_2 is larger than R_1, then the circuit wants to have more than $5.6\,V$ across R_2. This exceeds the breakdown voltage of the Zener, which allows current to flow. The voltage across R_2 will never be able to exceed $5.6\,V$.

4.6 Simple Diode Circuits

In this section, we present a few simple circuits involving diodes. These include the *rectifier*, which is used to convert from alternating to direct current, as well as circuits which use diodes to protect other circuit elements.

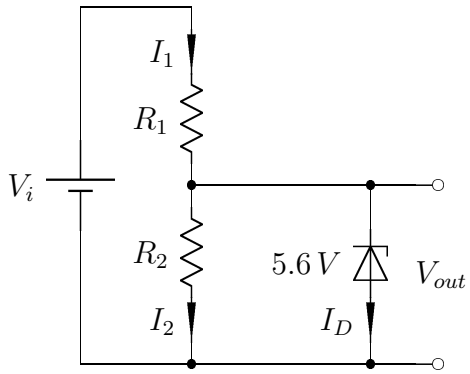

Figure 4.19: A Zener diode attached across the output of a voltage divider.

4.6.1 Conversion from AC to DC

Electricity is transmitted as alternating current, AC, but many of the appliances that are part of everyday life run on DC. The conversion from AC to DC is accomplished with a rectifier. This operation is important enough that will examine it in some detail, designing a series of progressively better circuits to carry it out.

> While AC is used to transmit power throughout the world, the exact specification of the AC depends on where one is. In North America, the AC has a frequency of $60\,Hz$ and an RMS voltage of $110\,V$. In Europe, the frequency is $50\,Hz$ and the RMS voltage is $220\,V$. Since what we often need is DC voltages, one might ask why electricity is not transmitted as such. One of the reasons that with AC, it is easy to change the voltage by using a transformer without loosing too much energy in heat. In such a conversion, the product of voltage times current stays constant. A second, and more important reason has to do with resistive power loss. Such a loss is given by I^2R. For a fixed power being delivered, a higher voltage leads to a smaller current, and hence a smaller power loss. While we could transmit DC at a high voltage, there is no simple device like the transformer to change the voltage. Unfortunately, most modern electrical appliances use DC at some level, and all have some circuit that converts AC to the needed DC circuit. Most homes today are swamped with AC to DC converters plugged into power strips.

The Half-wave Rectifier

Figure 4.20 is shows a half-wave rectifier. A transformer steps the AC line voltage down to a level more closely matched to the desired DC voltage. This is then passed through the circuit on the right, which consists of a diode and a resistor. Figure 4.21 shows the voltage across the resistor for a sinusoidal input voltage, $v_i(t) = V_0 \cos(\omega t)$. During the part of the AC cycle in which the voltage at the upper input terminal is more positive than that on the lower by at least one diode drop, the diode will allow current to flow. This will yield a voltage across the resistor that follows the positive cycle of the cosine function, minus the diode drop voltage.

Assuming that we do not exceed the breakdown voltage of the diode, in all other cases, the diode will not allow current to flow and the voltage across the output resistor will be zero.

The Full-wave Rectifier

The half-wave rectifier does not use the negative part of the cosine function and hence transmits only part of the power reaching it. By clever arrangements of diodes, we can avoid this problem. The circuit that does this is known as a full-wave rectifier and is shown in Figure 4.22. On the positive part of

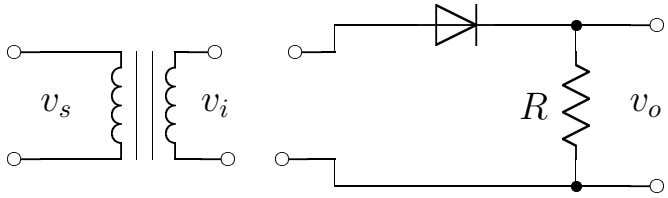

Figure 4.20: A half-wave rectifier circuit. An input source voltage, v_s, is shifted to v_i in the transformer. The voltage v_i then goes through the half-wave rectifier. The output as measured across the resistor, v_o, is shown in Figure 4.21.

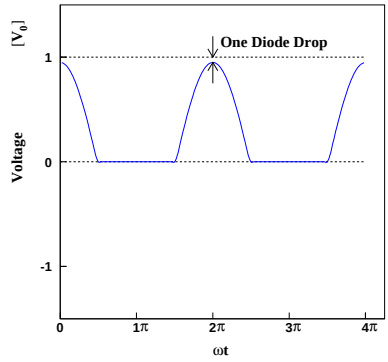

Figure 4.21: An input voltage is, $v_i(t) = V_o \cos(\omega t)$, is fed into the half-wave rectifier circuit shown in Figure 4.20, producing the indicated voltage across the resistor.

the cosine cycle, the upper-right and lower-left diodes function as a half-wave rectifier. However, on the negative part of the cycle, the lower-right and upper-left diodes also function as a half-wave rectifier, of the same polarity as before. This yields the output voltage shown in Figure 4.23. This is approximately the absolute value of the cosine input, but minus two diode voltage drops.

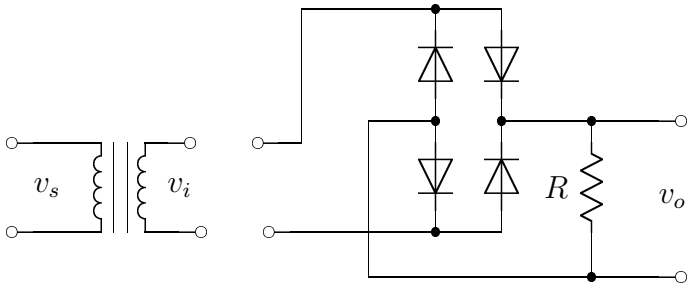

Figure 4.22: A full-wave rectifier circuit.

A Buffered Full-wave Rectifier

Assuming that the frequency of the input AC voltage is known, and constant, it is possible to use a capacitor to smooth out, or buffer, the voltage of the full-wave rectifier. This is certainly the case when converting the AC line voltages with frequencies of 50 to 60 Hz. In Figure 4.24, we have placed a capacitor across the output of the rectifier. This forms an RC circuit with the output resistor, with a

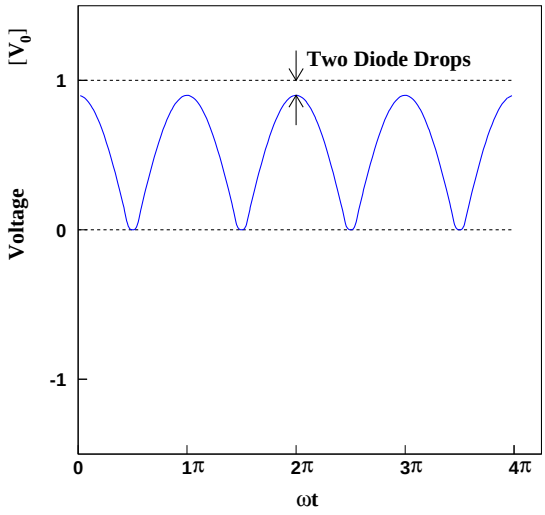

Figure 4.23: The output voltage from the full-wave rectifier.

characteristic time of $\tau = RC$. If the characteristic time is large compared to $\frac{1}{2}$ the period of the input wave, then the capacitor will charge up near the peaks of the cycles, but not have sufficient time to fully discharge during the troughs. The voltage across the resistor will resemble that shown in Figure 4.25. Here, we have achieved something very close to a DC voltage.

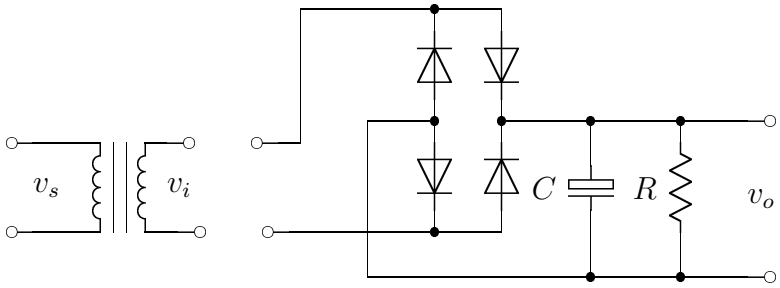

Figure 4.24: A full-wave rectifier circuit with a capacitor to filter the output voltage.

Voltage Regulators

For the sake of completeness, we will go the last step and add a solid-state voltage regulator chip to the output of our full-wave rectifier. Such a circuit is shown in Figure 4.26. The regulator is able to output a constant, very-low-ripple output voltage as long as the input voltage is within some finite range, typically, a fraction of the output voltage to several times the output voltage. As such, the buffering capacitor from the previous section is crucial, so the output voltage of the full-wave rectifier never falls below the minimum needed by the regulator.

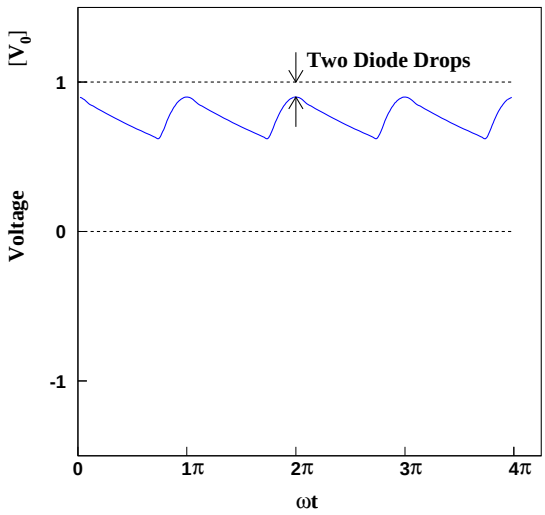

Figure 4.25: The output voltage of a full-wave rectifier with a buffering capacitor to smooth out the voltage.

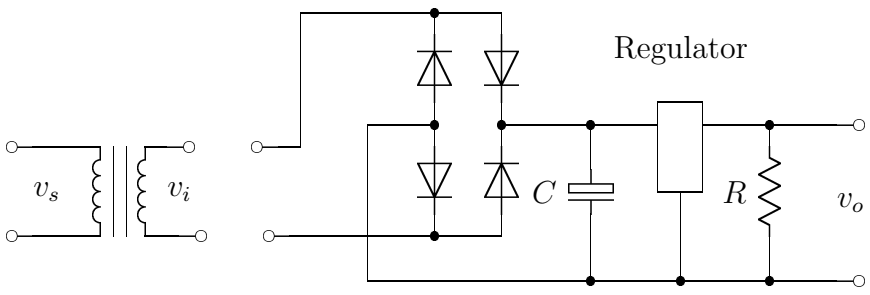

Figure 4.26: A full-wave rectifier circuit with a filter capacitor and a solid-state voltage regulator.

4.6.2 Diode-controlled Battery Backup

Related to the rectifier, which converts AC to DC, are battery backup circuits. These are used in devices that we do not want to shut off during brief power outages. While they are often associated with computers, they are far more ubiquitous than that. Many clocks, particularly alarm clocks, have battery backups.

A typical circuit is shown in Figure 4.27. The important point in this circuit is that the supply that has the higher voltage will have sufficient voltage drop across its diode to power the device. The other supply will not. Here, the DC supply runs at slightly higher voltage than the battery. When it is working, it supplies power; if it goes off, the battery becomes the higher voltage, and powers the device.

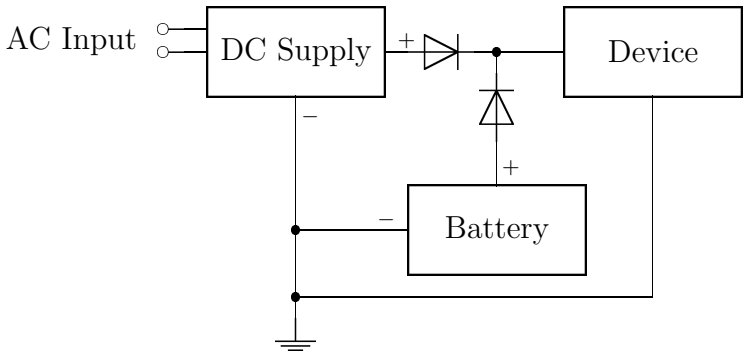

Figure 4.27: A battery backup for a DC power supply driving some device. The nominal voltage of the battery is slightly lower than that of the DC supply. The diodes only allow only the device with the higher voltage to provide current to the device. If the DC supply fails, its voltage goes to zero and the battery takes over.

4.6.3 Diode Voltage Clamp

A diode clamp maintains the voltage at some point at or below some specified value. The circuit shown in Figure 4.28 prevents the output voltage of the circuit from exceeding V_{CC} plus one diode voltage drop. If the voltage were to become larger than that, then sufficient current would flow through the diode to pull the voltage back to the maximum voltage. Such a circuit is useful in preventing voltage spikes from damaging circuits. The diode functions as a protection element.

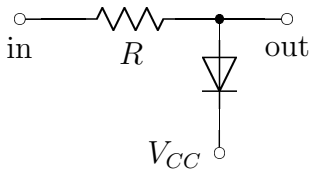

Figure 4.28: The voltage clamp prevents the output from exceeding V_{CC} plus one diode drop.

4.6.4 Diode Voltage Limiter

Related to the voltage clamp is the voltage limiter shown in Figure 4.29. In this circuit, if the output either exceeds or falls below the desired value by mores than one diode drop, then current will flow through the appropriate diode to keep the voltage within the specified range. Such a circuit can be modified by placing more than one diode in series in each limiter. We can also replace the ground with some other reference voltage.

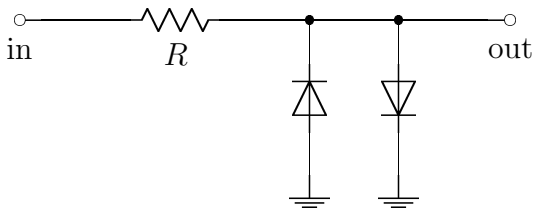

Figure 4.29: The voltage limiter maintains the signal within ±1 diode drop of ground.

4.6.5 Inductive Kick Blocker

Because the voltage across an inductor is proportional to how fast the current is changing in the inductor, if a switch is opened or closed when connected to an inductor, a large voltage, or *kick*, can result. A diode can be used to limit this kick by providing a path for the current to take to ground. This prevents a spark, or some other arcing, from occurring near the switch. Such a circuit is shown in Figure 4.30.

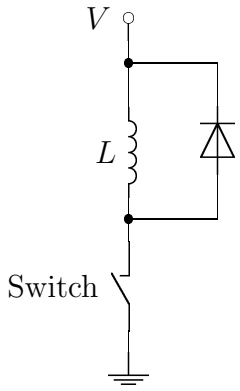

Figure 4.30: When the switch is opened or closed, the current changes rapidly in the inductor. This can cause a very large voltage drop across the inductor. The diode limits this by allowing current to flow through it.

Problems

1. If we have an n-type doped semiconductor in which the donor level is $0.05\,eV$ below the nominally empty conduction band, what fraction of the donor electrons are in the conduction band if the semiconductor is at $100\,K$? If it is at $200\,K$? If it is at room temperature? What are the consequences of this material in using semiconductors employed in space exploration?

2. Assume that the individual energy levels in a conduction band are actually spaced by about $10^{-5}\,eV$. At what temperature do these spacings become similar to the thermal energy?

3. In pure silicon, the band gap is $1.1\,eV$. At what temperature are 1% of the electrons in the conduction band?

4. Typically, we assume a diode to have a specific diode drop voltage. Consider a diode with a saturation current of $10\,pA$ and an operating current range between $5\,mA$ and $100\,mA$. By how much does the diode-drop voltage vary over this range of currents?

5. Consider the half-wave rectifier shown in Figure 4.20 where the transformer puts out a voltage with frequency $f = 60\,Hz$ and amplitude $10.0\,V$. Assuming that the diode drop is zero, what is the average output voltage over one full cycle? What is the RMS voltage over one full cycle?

6. Consider the half-wave rectifier shown in Figure 4.20 where the transformer puts out a voltage with frequency $f = 60\,Hz$ and amplitude $10.0\,V$. Assuming that the diode drop is $0\,V$, what is the average output voltage over one full cycle? What is the RMS voltage over one full cycle?

7. Consider the full-wave rectifier shown in Figure 4.22 where the transformer puts out a voltage with frequency $f = 60\,Hz$ and amplitude $10.0\,V$. Assuming that the diode drop is zero, what is the average output voltage over one full cycle? What is the RMS voltage over one full cycle?

8. In Figure 4.24, a filtering capacitor is used to hold up the voltage. If the output of the transformer has a frequency of $60\,Hz$ and the resistor has a minimum value of $R = 5000\,\Omega$, what value of C should be chosen to insure that the voltage does not drop below 50% of the maximum value?

9. A device needs an input voltage in the range of $4.35\,V$ to $5.65\,V$. Design a diode circuit that will do this.

10. A solid-state voltage regulator such as the LM7805 takes an input voltage between $7\,V$ and $30\,V$ and provides a DC output of $5\,V$. Explain what the output of the circuit in Figure 4.26 would be if we did not use the capacitor, C, in the circuit.

Chapter 5

Transistors

5.1 Introduction

Up until now, we have been dealing with passive circuit elements, none of which are able to increase the power of an input signal. In this chapter, we will start to look at an important class of active circuit elements known as *transistors* that can increase or *amplify*, the power of their input. Initially one might say that this violates the first law of thermodynamics. However, as we would expect, we use an external supply to provide the needed energy to the circuit. The active element simply feeds the supplied energy into the output signal. A black-box diagram of such an element is shown in Figure 5.1. The box has inputs and outputs, like our previous circuits; the new piece is the mechanism to provide external power. This is shown as V_{CC} and the ground in Figure 5.1. This external power will typically be provided as a DC voltage, V_{cc}, which is measured relative to ground. One consequence of this is that the maximum of $v_o(t)$ is V_{CC}. In order to be able to produce signals with large voltage outputs, we will need to provide a large V_{CC}. In some cases, we may replace the external ground with a second supply voltage, V_{EE}. As with V_{CC}, this also puts limits on the output of the circuit.

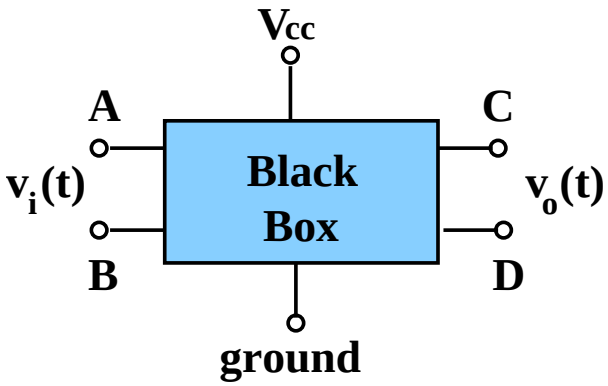

Figure 5.1: An *active* black-box circuit. The circuit has an input, $v_i(t)$, an output, $v_o(t)$, and a mechanism for supplying external power, V_{cc} and the *ground* connection.

In discussing transistors, we will start with a simple overview of bipolar junction transistors—often referred to as *bjt*s. We will then look at the details of the semiconductor junctions in the transistor, and then return to a more detailed discussion of transistor operation. This will then be followed by some specific examples of transistor circuits. Finally, we will briefly discuss field effect transistors.

5.2 Bipolar Junction Transistors

5.2.1 Overview of Operation

A bipolar junction transistor is effectively a sandwich of three different semiconductor layers. There are two types: a *negative-positive-negative* sandwich and a *positive-negative-positive* sandwich. These are referred to as *npn* and *pnp* transistors respectively. Figure 5.2 shows the symbols for each of these transistors.

Transistors are three-lead devices, with one lead coming from each of the three layers of the sandwich. The one connected to the middle layer is known as the *base*. The other two leads are known as the *collector* and the *emitter*. (When discussing transistors, we use the abbreviation B, C and E to refer to the *base*, *collector* and *emitter* respectively.)

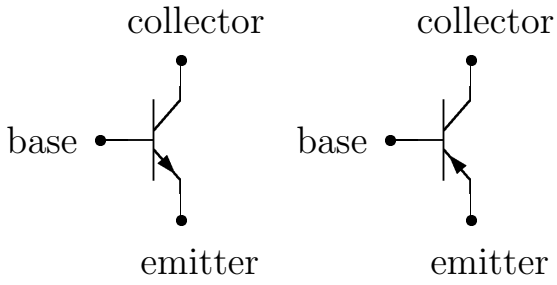

Figure 5.2: On the left is the symbol for an *npn* transistor; on the right is a *pnp* transistor. Both symbols show the three terminals of a transistor: base, collector and emitter.

In this section, we will examine the behavior of the *npn* transistor. However, the same principles hold true for the *pnp* transistor if one reverses the voltages. We will start with the behavior of an *npn* transistor in its normal operating mode. Figure 5.3 shows DC voltages (V_B, V_C and V_E) connected to the three terminals of the transistor. It also shows current flowing into the base and collector, I_B and I_C, and current flowing out of the emitter, I_E. Normal operation of the transistor will occur when the base-emitter junction is forward biased, $V_{BE} = V_B - V_E > 0$, and the base-collector junction is reverse biased, $V_{BC} = V_B - V_C < 0$. In fact, inside the transistor, the connection from B to E looks like a diode. Usually $V_{BE} \sim 0.6$ to $0.7\,V$. This is not something that we design, it is just the way that diodes work. We will assume $V_{BE} = 0.65\,V$ in the following discussion, but it is understood that the exact value of this drop depends on the transistor used. The connection from B to C also looks like a diode, but because this is reverse biased, no current flows from C to B. Finally, it is straightforward to see that $V_{CE} = V_C - V_E > 0$.

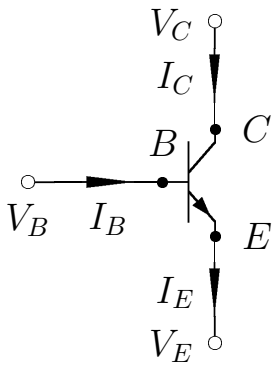

Figure 5.3: An *npn* transistor with DC voltages at each of its terminals, giving rise to DC currents as shown.

By conservation of electric charge, we have that

$$I_B + I_C = I_E. \tag{5.1}$$

In order for the transistor not to overheat and melt, these currents must remain below some maximum value. As I_E is the largest, we usually use it to determine the maximum current in the transistor.

The transistor has one final very important property: it functions like a *current-valve*. The current into the base, I_B, controls how much current flows into the collector, I_C. A small current into the base lets a large current flow into the collector. This relationship is described in equation 5.2, where β is the constant of proportionality. Typical values for β are from 50 to 300.

$$I_C = \beta I_B \tag{5.2}$$

Thus, a small base current gives rise to a much larger collector current, and, from equation 5.2, a large current flowing out of the emitter. This is what makes transistors useful. We can also define a parameter α as in equation 5.3. In normal transistor operation, α is slightly smaller than 1.

$$I_C = \alpha I_E \tag{5.3}$$

It is easy to see that α and β are related as follows:

$$\beta = \frac{\alpha}{1-\alpha} \tag{5.4}$$

$$\alpha = \frac{\beta}{1+\beta} \tag{5.5}$$

If α and β were constant, then we could design a circuit that would be able to provide a stable amplification of β. Unfortunately, this is not the case. For small emitter currents, the transistor can be *turning on*, and the values of β and α can depend strongly on the emitter current, I_E, with β doubling as I_E increases from a few tenths of a milliamp up to a few milliamps. For larger values of I_E, β will start to slowly fall off again, dropping by 10 to 20% as the current is increased to $\sim 100\,mA$. While this might have limited use, we would be much better served to design circuit whose performance does not depend on any particular value of β. Rather, they should only require that β is large. As we proceed, we will see how this comes about rather naturally.

Figure 5.4 shows an I–V curve for an *npn* transistor. We plot the collector current, I_C, against the potential difference between the collector and the emitter, V_{CE}. The four curves on the plot correspond to four different base currents, I_B. For each curve, there is some region where the transistor is turning on (the rapid rise of I_C on the left-hand side of the figure), after which the current I_C remains approximately constant up to the P_{max} line. The last feature on the graph is the maximum power, P_{max}. In order that the transistor not burn out [1], the power dissipated in the transistor,

$$P = I_C \cdot V_{CE}\,,$$

cannot exceed the maximum rated power for the transistor.

5.2.2 Bipolar Transistor Semiconductor Structure

As we have seen, a transistor is made from three layers of semiconductor material, either a p-layer between two n-layers, or an n-layer between two p-layers. We will discuss the former case, the so-called npn transistor. Figure 5.5 shows the structure of a typical bipolar npn transistor. The center layer is known as the base and the two outer layers are known as the collector and the emitter. Under normal operating conditions, the emitter emits electrons into the base. Most of these electrons diffuse through the base and are collected by the collector. In order for this to happen the base-emitter junction is forward-biased, while the base-collector junction is reverse-biased.

[1] There is some speculation that transistors actually function by magic. The small puff of smoke leaking out of the transistor when the power rating is exceeded may well be the magic leaking out.

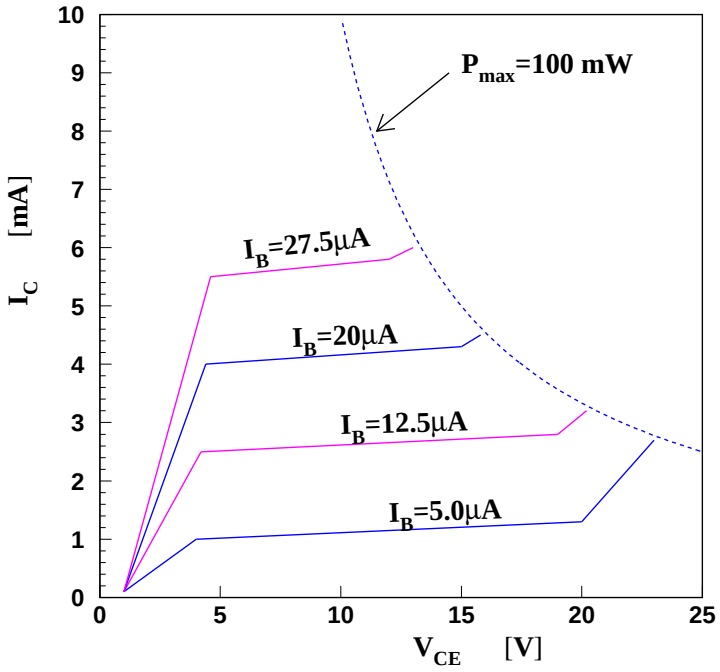

Figure 5.4: A representation of the I–V curve of an *npn* transistor with $\beta \sim 200$ and a maximum power of $100\,mW$.

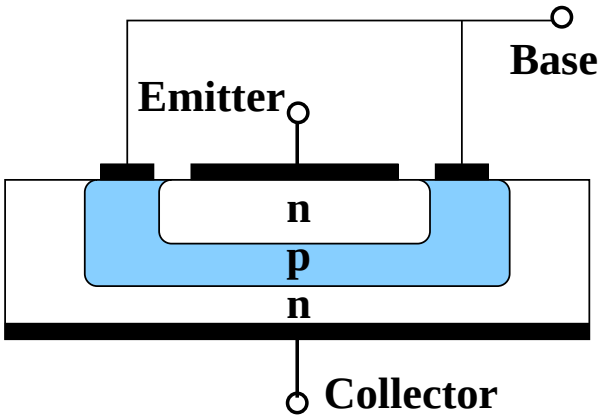

Figure 5.5: The semiconductor structure of a bipolar transistor. The *emitter* and *collector* are connected to n-type material while the *base* is connected to p-type material which is sandwiched between the n-type layers.

We start with the base-emitter pn-junction. If this is forward-biased, then the majority carriers from the emitter (electrons) move into the base, where they increase the concentration of minority carriers. Similarly, the majority carriers in the base (holes) flow into the emitter, where they increase the concentration of minority carriers. This results in a large current flow through the junction.

Next we examine the base-collector junction. This junction is reverse-biased, which from our un-

derstanding of diodes implies that there is a small reverse current due to diffusion of minority carriers in the base (electrons) into the collector. In the transistor, the injection of minority carriers into the base at the emitter junction and the extraction of minority carriers at the collector side leads to a large gradient in the concentration of minority carriers across the base. This causes these to diffuse from the emitter side to the collector side, and then into the collector. This reverse current, which is small in the diode, become large in the transistor. This is depicted in Figure 5.6.

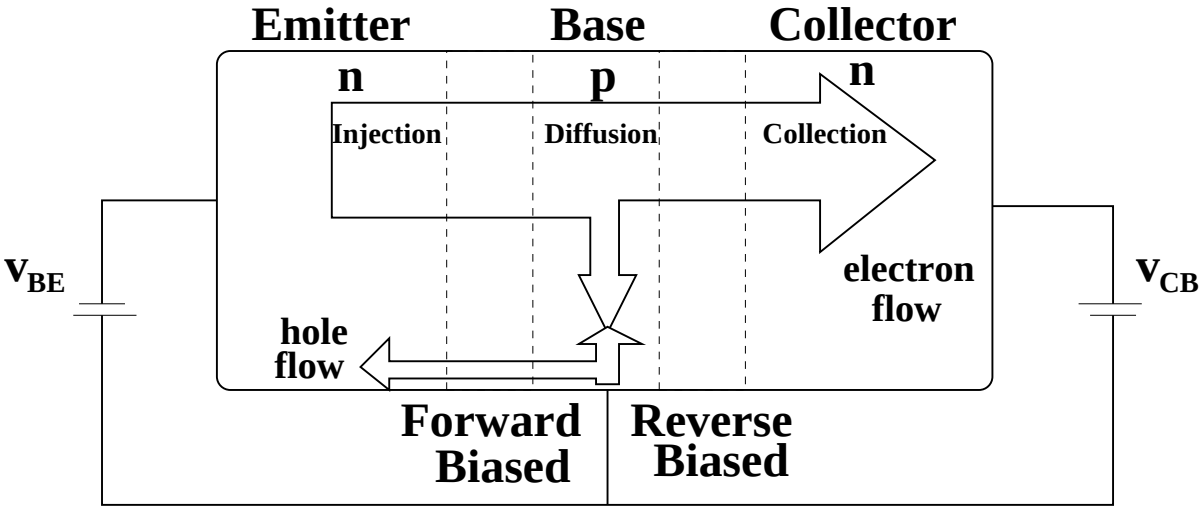

Figure 5.6: Carrier flow in a normally biased bipolar junction transistor.

A bipolar junction transistor appears quite symmetrical, meaning that it is in principle possible to reverse the roles of the collector and the emitter. Indeed, a transistor will operate in this reverse mode. However, the emitter-base and collector-base junctions are generally manufactured differently to make the forward-mode of operation work better than the reverse mode.

5.2.3 Transistor Operation

Consider an npn transistor that is biased in such a way to be in its normal operating range. The base is at a potential about $0.65\,V$ higher than the emitter, while the collector is at an even higher potential. The simple model which we have been using to describe the transistor states that the collector current, I_C, is β times the base current. Let us now take a closer look at the transistor behavior. In particular, for specific values of base current (I_B), supply voltage (V_{CC}), and loads, what are the voltage drops across the base-emitter and collector-emitter junctions, and what is the collector current (I_C). These are referred to as the *operating point* of the transistor.

We start with the input to the transistor, consisting of the current into the base, I_B, and the base-emitter voltage difference, V_{BE}. The base-emitter junction looks like a forward-biased diode, so we expect that the I-V curve should be that of a diode. This is shown in Figure 5.7. The nearly vertical rise of the the I-V curve occurs at about $0.65\,V$, but for small enough currents, B_{BE} shrinks, and below some voltage the diode turns off. This in turn shuts off the transistor. Let us now connect the base of the transistor to a voltage source with voltage V_S and some source impedance R_S as shown in Figure 5.8. If we look at the Kirchhoff voltage loop equation around the circuit, we have:

$$0 \quad = \quad V_s - I_B R_s - V_{BE}\,. \tag{5.6}$$

We can solve this for the current, I_B as a function of the base-emitter voltage, V_{BE}. This gives us:

$$I_B \quad = \quad \frac{V_s}{R_s} - \frac{1}{R_s} \cdot V_{BE}$$

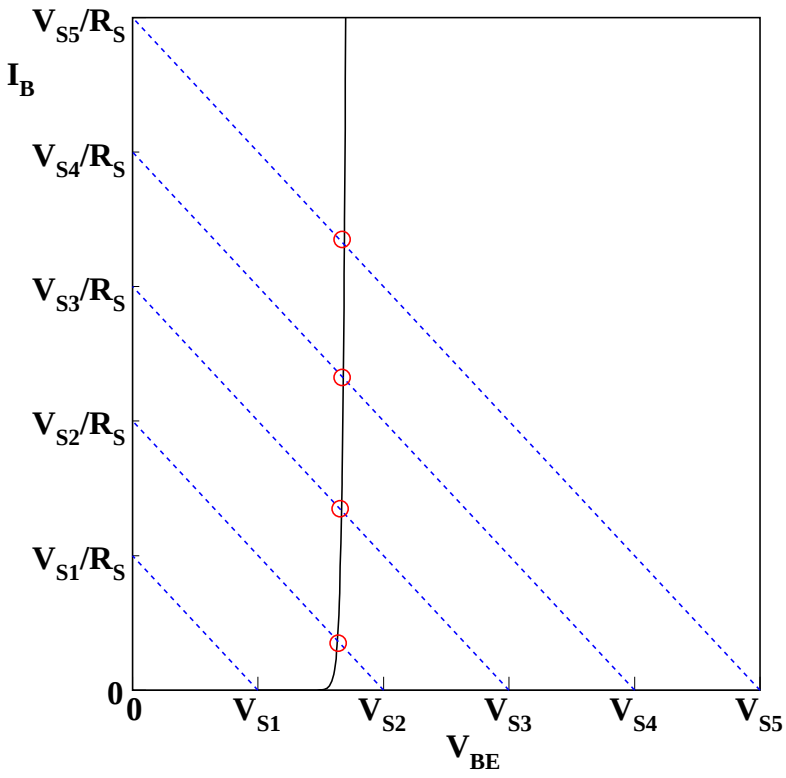

Figure 5.7: The I-V curve for the input to an npn transistor. The nearly vertical rising curve is the familiar diode curve. Overlaid on these are several I-V curves for a voltage source with internal resistance R_S. The intersections of these curves with the diode curve (circled) correspond to operating points of the transistor.

This I-V curve has been plotted in Figure 5.7 for a fixed R_s, but several values of the source voltage, V_s. If we choose a particular source voltage, say V_{s3} in the figure, then the intercept of the I-V line and the diode I-V curve determines the particular value of I_B and V_{BE}. We note that if we choose V_{s1}, then $I_B = 0$ and the diode is turned off. For a particular V_s between V_{s1} and V_{s2} the diode turns on, which turns the transistor on, and the voltage drop across the base-emitter junction is approximately constant.

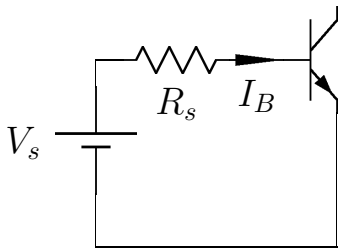

Figure 5.8: A voltage source V_s with source resistance R_s connected to the base of an npn transistor.

Now let us consider the output side of the transistor. Using the circuit shown in Figure 5.9. The

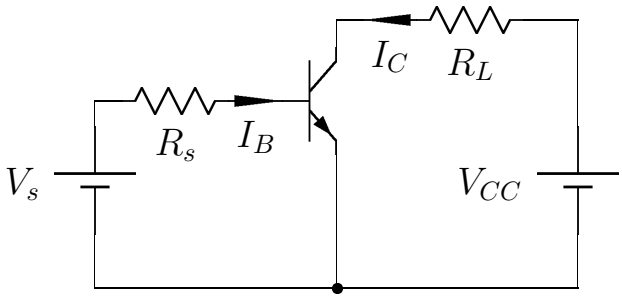

Figure 5.9: A voltage V_{CC} and load connected across the collector-emitter junction of our transistor. We have the same source and source resistance as earlier connected across the base-emitter junction of the transistor.

I-V curve for the output of a transistor is proportional to I_B in the active part of the operation of the transistor, as shown in Figure 5.10. This is represented by the series of flat lines, each of which corresponds to a different input current. On the right, these curves eventually run into the power limit, shown as the curved, dashed line. On the left-hand side, these curves all start from $V_{CE} = 0$ and $I_C = 0$. They then quickly rise along a nearly universal curve up to the operating current. We next consider the load lines as we did for the input circuit. If we take the Kirchhoff voltage loop around V_{CC}, R_L and V_{CE} in Figure 5.9,

$$0 \quad = \quad V_{CC} - I_C R_L - V_{CE} \,. \tag{5.7}$$

This can be solved for the current to yield

$$I_C \quad = \quad \frac{V_{CC}}{R_L} - \frac{1}{R_L} \cdot V_{CE} \tag{5.8}$$

This load line is drawn on Figure 5.10. As with the input, the intercept of the load line with the I-V curve of the transistor yields the operating point of the transistor, (V_{CE}, I_C). In the case where the load line intersects the flat part of the I-V curve, we are in the active region of the transistor. This is typically where one wants to operate the transistor. If we happen to choose a point that is to the right of the power curve, we are very likely to burn out the transistor. Finally, when the load line intersects the rising part of the I-V curve on the left-hand side of the plot, we are in *saturation*. This voltage is referred to as the *saturation voltage*, V_{CEsat}. Because all I-V curves with larger I_B also pass through the same load point, the values of V_{CE} and I_C no longer change as the base current is increased. Rather, they remain fixed at the saturation values, V_{CEsat} and I_{Csat}.

The above procedures provide a method of determining what the operating point is for a transistor circuit. I_B is determined on the input side, which selects a particular I-V curve on the output side. The intercept of the load line with the I-V curve then yields V_{CE}.

We can put much of this together by creating a transfer plot, a graph of V_{CE} versus the source voltage V_s. Such a plot is shown in Figure 5.11. For values of V_s smaller than the the nominal diode voltage drop, $0.65\,V$, the collector-emitter voltage is V_{CC}. The transistor is off, and the emitter is at ground. In the active region, the collector-emitter voltage falls from V_{CC} to V_{CEsat} over some finite range of source voltage, after which it remains at V_{CEsat}. The actual slope of the drop is

$$\text{slope} \quad = \quad -\frac{\beta R_L}{R_s} \,. \tag{5.9}$$

To show this, consider equations 5.6 and 5.7. From the latter, we can write that

$$I_B R_S \quad = \quad V_s - V_{BE} \,. \tag{5.10}$$

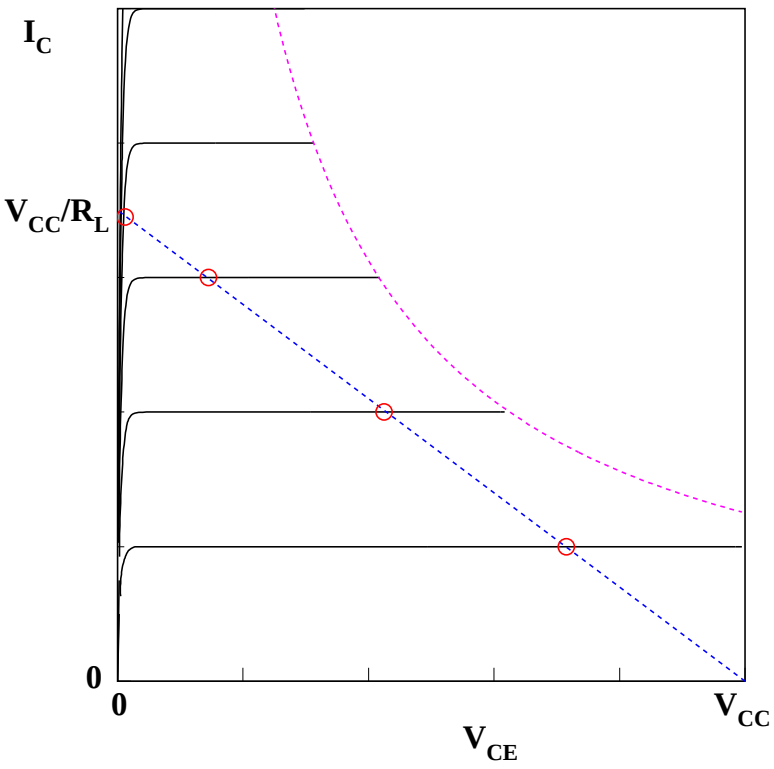

Figure 5.10: The I-V curve for the output to an npn transistor. The family of roughly horizontal lines correspond are the I-V of the output of the transistor for different values on the input voltage, I_B. The slanted line is the I-V curve for a voltage source V_{CC} and load resistor R_L. The circled intercepts indicate operating points of the transistor for a given input current and supply parameters. In particular, the collector-emitter voltage difference for the given output current.

However, in the active region, $I_B = I_C/\beta$, or

$$\frac{I_C}{\beta} R_S \quad = \quad V_s - V_{BE} \, .$$

From equation 5.8, we can substitute for I_C to obtain:

$$\frac{1}{\beta} \frac{R_S}{R_L} \left[V_{CC} - V_{CE} \right] \quad = \quad V_s - V_{BE} \, .$$

This then yields

$$V_{CE} \quad = \quad -\frac{\beta R_L}{R_S} V_s + \left[\frac{\beta R_L}{R_S} V_{BE} + V_{CC} \right] \tag{5.11}$$

which gives the correct slope. If we note that, in the active region, $V_{BE} \approx 0.65\,V$, we can also rewrite this as:

$$V_s \quad = \quad 0.65\,V + \frac{R_S}{\beta R_L} \left[V_{CC} - V_{CE} \right] \, .$$

When $V_{CE} = V_{CEsat}$, the source voltage at which the circuit saturates can be seen to be

$$V_{ssat} = 0.65\,V + \frac{R_S}{\beta R_L}\left[V_{CC} - V_{CEsat}\right] . \tag{5.12}$$

.

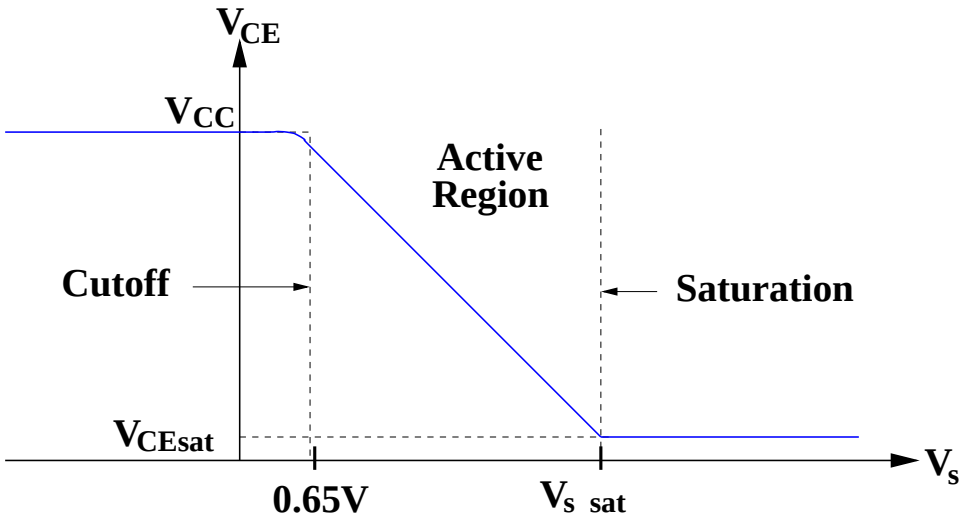

Figure 5.11: A transfer plot showing how V_{CE} is related to V_s. When V_s is below the cutoff voltage of the transistor, the transistor is off and $V_{CE} = V_{CC}$. The transistor then turns on and enters the active region, where the slope is given by equation 5.9. When reaches $V_{s\,sat}$ as given in equation 5.12, $V_{CE} = V_{CEsat}$. It then remains at the saturation voltage as V_s is increased.

5.2.4 The Eber-Moll Model

In discussing transistors, we have been saying that the current into the base controls the current into the collector:

$$I_C = \beta I_B$$

In fact, this is not quite right. It is actually better to think of the voltage difference between the base and the emitter, V_{BE} as the control parameter for the current out of the emitter rather than that into the collector. A better parametrization for the current out of the emitter is then given by the Eber-Moll equation:

$$I_E = I_s\left[e^{V_{BE}/V_T} - 1\right] . \tag{5.13}$$

The saturation current, I_S, is a constant with a value of about $10^{-12}\,A$. The voltage V_T is

$$V_T = k_B T/e \tag{5.14}$$

where k_B is the Boltzmann constant ($k_B = 8.617 \times 10^{-5}\,eVK^{-1}$) and T is the temperature in kelvins. The term $k_B T$ is the thermal energy associated with particles at temperature T. Considering charge carriers of charge e, we can associate a thermal voltage, V_T, with the thermal energy:

$$eV_T = k_B T \tag{5.15}$$

which then yields equation 5.14. At room temperature, we have:

$$\begin{aligned} V_T &= \left(8.617 \times 10^{-5}\, eV/K\right) \cdot (293.2\, K) \,/\, (e) \\ V_T &\approx 25.3\, mV \end{aligned}$$

We can rearrange equation 5.13 and solve for V_{BE},

$$V_{BE} = V_T \cdot \ln\left(\frac{I_E}{I_S} + 1\right) \tag{5.16}$$

which tells us how V_{BE} changes with I_E:

$$\begin{aligned} \frac{dV_{BE}}{dI_E} &= V_T \left(\frac{1/I_S}{I_E/I_S + 1}\right) \\ \frac{dV_{BE}}{dI_E} &\approx V_T/I_E \end{aligned}$$

This looks like an additional resistance, r_E through which current coming out the emitter must flow.

$$r_E = V_T/I_E \tag{5.17}$$

At this point, it is convenient to note that $I_C \approx I_E$, and rewrite equation 5.17 in terms of I_C as

$$r_E = V_T/I_C\,. \tag{5.18}$$

At room temperature, we can write that

$$r_E = (25.3\, mV)\,/I_C\,.$$

This resistance depends on the temperature of the transistor, because of V_T, and on the current into the collector, I_C. For currents on the order of $10\, mA$, we have r_E on the order of Ohms.

5.2.5 Time-dependent Voltages

Previously, we looked at the static, or DC properties of a transistor. In addition to the three DC voltages, let us now allow for time-dependent voltages at the three terminals to the transistor. We will write the total voltage on each pin as $v_b^{tot}(t)$, $v_c^{tot}(t)$ and $v_e^{tot}(t)$. We will also observe that it is possible to break the total voltages into a DC piece and a time-varying piece as follows.

$$\begin{aligned} v_b^{tot}(t) &= V_B + v_b(t) \\ v_c^{tot}(t) &= V_C + v_c(t) \\ v_e^{tot}(t) &= V_E + v_e(t) \end{aligned}$$

If the time-dependent voltages are zero, we will refer to this as the *quiescent operating point*, or simply the *quiescent point*. Our discussion will largely focus on the behavior of the purely time-varying parts of the voltages: $v_b(t)$, $v_c(t)$ and $v_e(t)$, but we will see that setting up a reasonable quiescent point is important in having a well-behaved transistor circuit.

In addition to voltage, we can also look at the current flowing along each leg of the transistor. As above, we will define a DC part represented by a capital I, and a time-varying part represented by a lower case i. As before, we can write the following.

$$\begin{aligned} i_b^{tot}(t) &= I_B + i_b(t) \\ i_c^{tot}(t) &= I_C + i_c(t) \\ i_e^{tot}(t) &= I_E + i_e(t) \end{aligned}$$

In order for the transistor to be in its normal operating range, the total voltage on the inputs must satisfy the previously established relations, namely, the base-emitter junction must be forward biased and the base-collector junction must be reverse biased. In addition, in normal operation, we will have approximately one diode drop across the base-emitter junction. This gives:

$$\begin{aligned} v_b^{tot}(t) &\sim v_e^{tot}(t) + 0.65\,V \\ v_c^{tot}(t) &> v_b^{tot}(t). \end{aligned} \tag{5.19}$$

In fact, we can rewrite equation 5.19 as

$$(V_B - V_E) - (v_b(t) - v_e(t)) \sim 0.65\,V\,, \tag{5.20}$$

which can be decomposed into a time-independent and time-dependent part. Noting that the DC voltages are both constant (that is what DC means), the only way that equation 5.20 can be satisfied for all time is if we have that

$$(V_B - V_E) = 0.65\,V \tag{5.21}$$

and that

$$v_e(t) = v_b(t)\,. \tag{5.22}$$

What is significant here is that $v_e(t)$ is just a copy of $v_b(t)$. At first glance, this may seem fairly useless, as a wire also has $v_o = v_i$, but as we continue we will see why this is useful. It is actually one of the important features of the emitter-follower circuit, which will be discussed in the next section.

Finally, we should note that the internal resistance of the transistor given in equation 5.19 can affect this latter relation. There will be a voltage drop across r_E, but the size of such a drop depends on the current, i_e^{tot}. Thus we may find that $v_e(t)$ is somewhat smaller than $v_b(t)$.

5.3 Simple Transistor Circuits

5.3.1 The Emitter-follower Circuit

Let us proceed with our discussion of transistor circuits by adding a resistor, R_E, that connects the emitter to ground. We will then connect an external DC voltage, V_{CC}, to the collector. This is shown in Figure 5.12. Also shown in the figure is that we have set a DC voltage at the base, V_B, and then added a time-varying voltage, $v_b(t)$, to this. The DC voltage at the emitter will be $V_E = V_B - 0.65\,V$. With the resistor, R_E, connected to ground, there must be a DC current, $I_E = V_E/R_E$. From this, we determine the DC (or average or steady-state) currents flowing in the transistor.

$$\begin{aligned} I_E &= (V_B - 0.65\,V)\,/R_E \\ I_C &= I_E\beta/\,(1+\beta) \\ I_B &= I_E/\,(1+\beta) \end{aligned} \tag{5.23}$$

We want to choose R_E to make sure that we are far below the power dissipation limits of the transistor, $I_C \cdot V_{CC} < P_{max}$. This will keep the transistor from burning out .

Let us now look at the equivalent circuit that $v_b(t)$ sees when it looks into the base of the transistor. The equivalent impedance of the transistor will just be some resistance, R_{in}. In order to determine what this is, let us start with the current and voltage at the emitter. If the emitter voltage changes from its DC value, V_e, by some time-varying amount, $v_e(t)$, then the current I_E will change by some time-varying amount:

$$i_e(t) = v_e(t)/R_E.$$

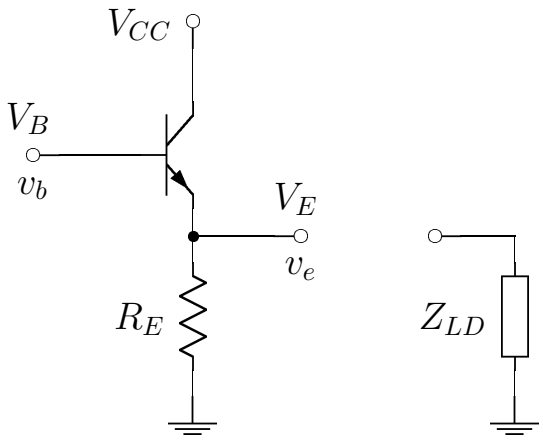

Figure 5.12: A transistor hooked up as an emitter-follower. In such a circuit, $v_e(t) = v_b(t)$. Z_{LD} is a load that could be hooked up to the output of the circuit.

However, we know that v_e exactly tracks the time-varying voltage at the base, v_b, as long as the transistor stays in its normal operating state.

$$v_e(t) \quad = \quad v_b(t)$$

We can write i_e in terms of i_b as in equation 5.23, which gives us

$$(1+\beta)\, i_b \quad = \quad v_b / R_E \,.$$

This can be rewritten as:

$$\frac{v_b}{i_b} \quad = \quad (1+\beta) R_E.$$

Which tells us that the equivalent input impedance, Z_{in}, of the transistor is

$$Z_{in} \quad = \quad (1+\beta) R_E. \tag{5.24}$$

As β is on the order of 100, the input impedance of the transistor appears to be quite large. If we have $R_E = 1\,k\Omega$, then $Z_{in} \sim 100\,k\Omega$, which is hopefully large enough not to "load down" the circuit driving the transistor.

In an application using this circuit, we would be driving some load, Z_{LD}. To describe this, we need to replace R_E with the parallel combination of R_E and Z_{LD}:

$$Z_{in} \quad \sim \quad \beta \left(R_E \,||\, Z_{LD} \right) \,.$$

In our discussion of chaining circuits together, we found that having the input impedance of a circuit very large is good because it does not load down the previous stage.

Now we need to check the output impedance of the emitter-follower circuit, Z_{out}. Looking back into the transistor through the emitter, we see the Thèvenin equivalent of the circuit driving the transistor, but viewed through the transistor. In order to proceed, we have to assume that the driving circuit has a resistance r_s which gives rise to its output impedance. For many applications, r_s might be something like $50\,\Omega$. For a filter circuit, the output impedance can be just about anything.

As with determining the Thèvenin equivalents of DC circuits, we will want to determine the open-circuit voltage and the short-circuit current. The ratio of these gives us Z_{out},

$$Z_{out} \quad = \quad \frac{v_{oc}}{i_{ss}} \,.$$

The open circuit voltage is just $v_{oc} = v_e$. The (AC) short-circuit current is determined from the short-circuit current into the base:

$$i_E = (\beta + 1)\, i_B \,.$$

Nominally, the current into the base is

$$\begin{aligned} i_b &= \frac{v_b}{r_s + Z_{in}} \\ i_b &= \frac{v_b}{r_s + (1+\beta)\, R_E} \,. \end{aligned}$$

But if we short the output, then R_E goes to zero and the short-circuit current into the base becomes

$$i_b = \frac{v_b}{r_s} \,.$$

If we note that $v_e = v_b$, then the short-circuit current from the emitter will be

$$i_{ss} = \frac{v_e\,(1+\beta)}{r_s} \,.$$

Putting all of this together, we find that the Thèvenin equivalent impedance as seen from the emitter, Z_{out}, will be:

$$Z_{out} = \frac{r_s}{1+\beta} \,. \tag{5.25}$$

The output impedance of the circuit is much smaller than the source impedance!

In other words, the emitter-follower is a circuit whose output is equal to the input, whose input impedance is very large, and whose output impedance is very small. This is exactly the behavior that we needed so that we could connect circuits together, but keep all the components of reasonable size. It may be more useful to think of the emitter-follower as an *impedance shifter*. Figure 5.13 shows two different representations of the emitter follower circuit.

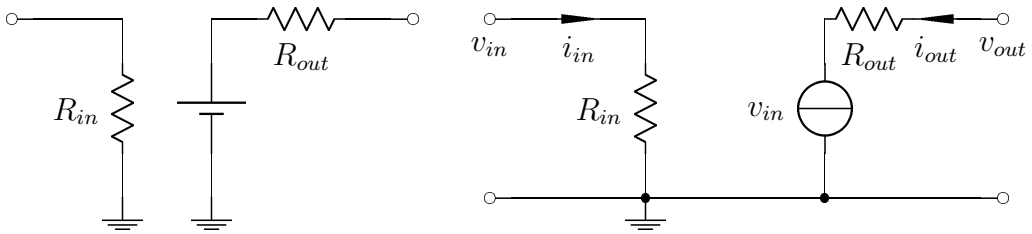

Figure 5.13: Two versions of the equivalent circuit showing the input and output of an emitter-follower circuit. The input impedance, $Z_{in} = (1+\beta)\, R_E$, is typically very large. The output impedance, $Z_{out} = r_s/(\beta+1)$, is typically very small. The diagram on the right emphasizes the fact that the emitter is common in both circuits. It also shows that the output voltage source follows the input voltage.

We now understand the important properties of the emitter-follower. However, there are some operating restrictions on the circuit. First, the lowest voltage to which the emitter can go is zero. This means that if $v_b^{tot}(t)$ falls below $0.65\,V$, the diode turns off and the output, $v_e^{tot}(t)$, goes to zero. Figure 5.14 shows an example of this. There the maximum value of the emitter voltage is given by the external power, V_{CC}. If v_b^{tot} becomes larger than V_{CC}, the emitter voltage will just go to V_{CC}. (When v_b^{tot} becomes larger than V_{CC}, we are forward-biasing the BC junction and the transistor goes into a mode known as *saturation*.) An example of this is shown in Figure 5.15.

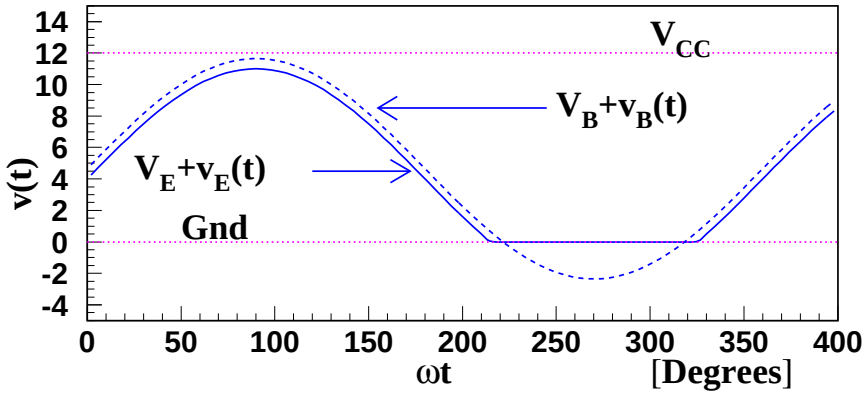

Figure 5.14: The dashed curve shows the input voltage into an emitter-follower circuit, while the solid curve shows the output. Note that when the input falls below $0.65\,V$, the transistor turns off and the output is clipped at $0\,V$.

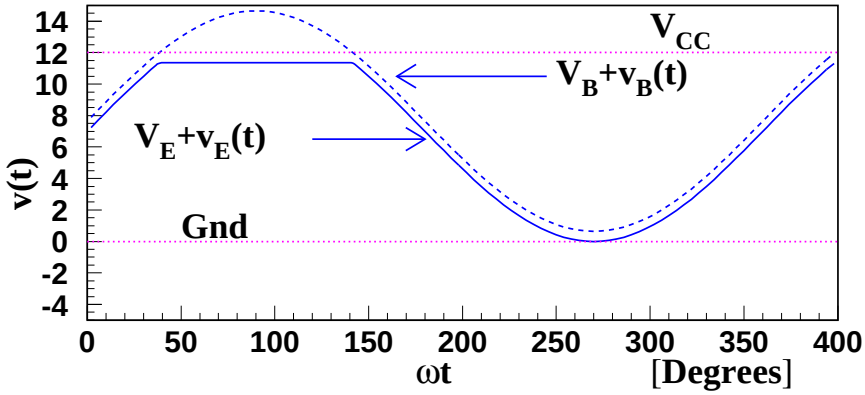

Figure 5.15: The dashed curve shows the input voltage into an emitter-follower circuit, while the solid curve shows the output. Note that when the input goes above V_{CC}, the output of the transistor saturates.

Example: Let us now design an emitter-follower circuit using a number of the tools that we have developed in this course. The circuit that we are going to build is shown in Figure 5.16. The power-supply voltage, V_{CC}, serves two purposes in this circuit. It sets the voltage at the collector, but also, through the R_1-R_2 voltage divider, sets the DC voltage level at the base. We have two *blocking* capacitors, C_1 and C_2, the purpose of which is to block the DC voltages from v_{in} and v_o. In both cases, they are part of a high-pass filter circuit. We will assume that $V_{CC} = 12\,V$ and that the maximum current that the transistor can handle is $10\,mA$.

Let us start with the emitter voltage. Because $v_o = v_{in}$, we would like to allow for the maximum possible voltage swing in v_o. If we choose the nominal DC level at the emitter to be $\frac{1}{2}\,V_{CC}$, then v_o can swing between $-\frac{1}{2}\,V_{CC}$ and $\frac{1}{2}\,V_{CC}$. With $V_E = \frac{1}{2}\,V_{CC}$, we can choose the value of R_E to limit the maximum current through the transistor. If we do not want to risk burning out our transistor, we should set the nominal value of I_E to be about 10% of the maximum current, or $1\,mA$. From this we get that

$$\begin{aligned} R_E &= V_E/I_E \\ R_E &= (6\,V)\,/\,(1\,mA) \end{aligned}$$

or

$$R_E = 6\,k\Omega$$

In order to have the nominal $V_E = \frac{1}{2}V_{CC}$, we must set $V_B = 0.65\,V + \frac{1}{2}V_{CC}$. Using our voltage divider equation, we have

$$0.65\,V + \frac{1}{2}V_{CC} = \frac{R_2}{R_1 + R_2}V_{CC}.$$

This can be solved to yield an expression which relates R_1 and R_2.

$$\begin{aligned} R_1 &= \frac{(V_{CC}/2) - 0.65\,V}{(V_{CC}/2) + 0.65\,V} \cdot R_2 \\ R_1 &= 0.80 \cdot R_2 \end{aligned} \tag{5.26}$$

In order to set the exact values of R_1 and R_2, we need to consider the input impedance of the transistor, $R_{in} = \beta R_E$, and we need to compare it to the Thèvenin equivalent resistance of the voltage divider, or $R_1 \,||\, R_2$. In order for the transistor not to load down the voltage divider, we require that

$$\beta R_E >> R_1 \,||\, R_2$$

Using equation 5.26 and approximating β as 100, we can write this as

$$(100) \cdot (6\,k\Omega) >> \frac{(0.8R_2)\,(R_2)}{(0.8R_2) + (R_2)}$$

which yields

$$R_2 << 1350\,k\Omega.$$

Taking "much less" to mean a factor of 100, we would get the following values for all the resistors in the circuit:

$$\begin{aligned} R_E &= 6\,k\Omega\ , \\ R_1 &= 10.8\,k\Omega \quad \text{and} \\ R_2 &= 13.5\,k\Omega\ . \end{aligned}$$

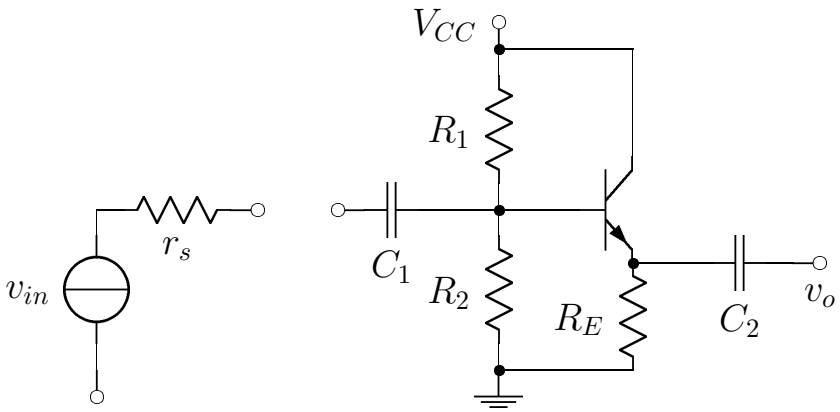

Figure 5.16: An example emitter-follower circuit as described in the text.

Now let us examine the capacitors. Remember that they are intended to form high-pass filters that let through all frequencies larger than some low value. In this example, let us assume that $f = 10\,Hz$. We next need to determine what resistance to include. To do this, we need to realize that for an AC voltage, the DC voltage source, V_{CC}, can also serve as a *sink* or ground. This means that as far as our high-pass filter is concerned, the equivalent circuit is shown in Figure 5.17(a). Figure 5.17(b) shows the explicit high-pass filter with capacitance C_1 and resistance of $R_1 \,||\, R_2 \,||\, \beta R_E$. This gives the characteristic frequency for the circuit:

$$\omega_{RC} \;=\; \frac{1}{(R_1 \,||\, R_2 \,||\, \beta R_E) \cdot C_1}$$

Expanding this, we get

$$\begin{aligned} 2\pi\,(10\,Hz) &> \frac{1}{(13.5\,k\Omega \,||\, 10.8\,k\Omega \,||\, 600\,k\Omega) \cdot C_1} \\ 0.016\,s &< (5.9\,k\Omega) \cdot C_1 \end{aligned}$$

or

$$C_1 \;>\; 2.7\,\mu F\,.$$

Finally, we need to determine the value of C_2. This should also form a high-pass filter with the same

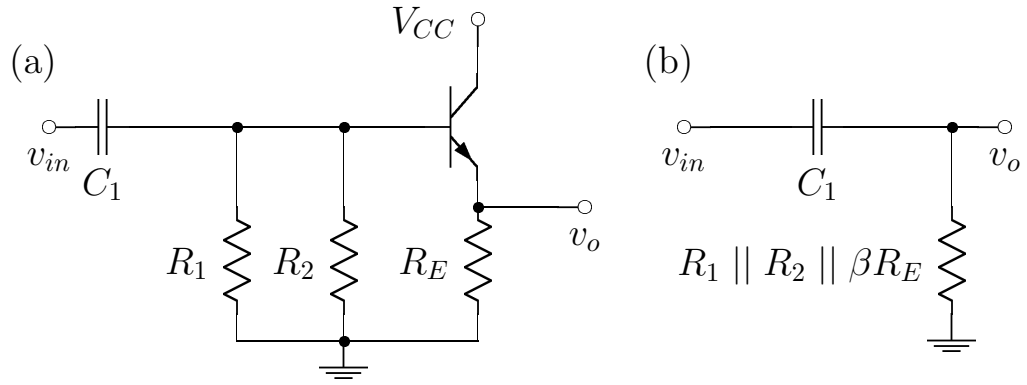

Figure 5.17: The AC equivalent circuit showing the input to the emitter-follower circuit of Figure 5.16. The DC voltage source, V_{CC}, serves as a *sink* or ground for an AC input voltage, v_{in}. Circuit (a) shows the transistor explicitly, while circuit (b) shows the equivalent high-pass filter.

frequency as before, $\omega = 62.8\,s^{-1}$. As before, we need to identify which resistor completes the high-pass filter. This resistance needs to be to the right of the capacitor. The only resistance that this can be is the load connected to the output of the circuit (Figure 5.18). We know that the output impedance of the circuit is roughly $R_{out} = (R_1 \,||\, R_2)\,/\beta$, so choosing a factor of 100 for "much larger", we get that the load must be larger than $R_1 \,||\, R_2$, or $R_{LD} \approx 6.\,k\Omega$. This then gives:

$$\begin{aligned} 62.8\,s^{-1} &> \frac{1}{R_{LD} \cdot C_2} \\ 0.016\,s &< (1.2\,k\Omega) \cdot C_2 \\ C_2 &> 2.7\,\mu F \end{aligned}$$

However, this may not tell us everything. If the source impedance feeding our circuit, r_s, is smaller than $R_1 \,||\, R_2$, then it is possible to safely drive an even larger load, or smaller resistance. For example, if $r_s = 50\,\Omega$, then the load we can drive is also $50\,\Omega$. Replacing the $6.\,k\Omega$ above with r_s, we would find that $C_2 > 320\,\mu F$. The bottom line is that this capacitor's value is set by the minimum load that we expect to drive, which itself may depend on the input to the circuit, and not the circuit itself.

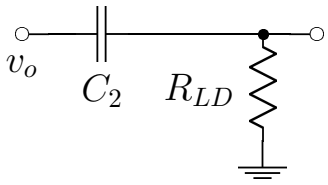

Figure 5.18: The output capacitor and the load resistance attached to the emitter-follower circuit.

5.3.2 The Common Emitter Amplifier

In the last section, we examined a circuit that changed input and output impedance, but whose output signal was just a copy of the input signal. We now want to examine circuits that amplify the input signal, thereby creating an output voltage that is larger than the input voltage. A simple way to do this is the common-emitter, or inverting amplifier. Such a circuit is shown in Figure 5.19, and we will now examine the normal operation of this amplifier circuit.

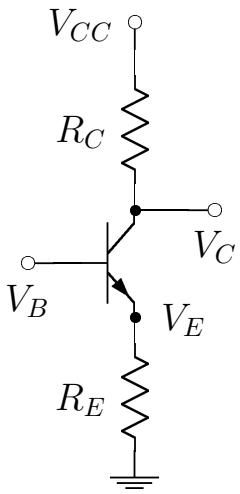

Figure 5.19: The common-emitter amplifier circuit. A supply voltage, V_{CC}, is connected to the collector. The time-varying voltage at the collector is related to that at the base by $v_c(t) = -(R_C/R_E)v_b(t)$.

A supply voltage, V_{CC}, is connected through a resistor, R_C, to the collector of the transistor. In addition, the output of the circuit is taken directly from the collector. As with the emitter-follower, a resistance R_E connects the emitter to ground. In order for the circuit to work, the transistor needs to be in its normal operating state: $V_C > V_B$ and $V_B - V_E = 0.65\,V$, with $V_E > 0$.

The DC voltage, V_B, at the base leads to a DC voltage at the emitter of $V_E = V_B - 0.65\,V$. The supply voltage is connected through the resistor R_C. If a current, I_C, flows through the resistor, then the collector voltage will be

$$V_C \;=\; V_{CC} - I_C R_C\,. \tag{5.27}$$

As with the emitter-follower, the current that flows from the emitter is $I_E = V_E/R_E$. Let us now examine the effect of time-varying voltages on the circuit.

Consider a time-varying input voltage $v_b(t)$ connected to the base. From our discussion of the emitter-follower, we know that $v_e(t) = v_b(t)$. We also know that the time-varying current through R_E is given as

$$i_e(t) \;=\; v_e(t)/R_E\,,$$

which can be rewritten in terms of $v_b(t)$ as:

$$i_e(t) = v_b(t)/R_E\,.$$

The current into the collector is approximately equal to the current out of the emitter, $i_c(t) \approx i_e(t)$. By looking how equation 5.27 behaves when v_c is varied, we obtain:

$$v_c(t) = -R_C \cdot i_c(t),$$

that is, the time-varying collector and the base voltages are related by

$$v_c(t) = -\frac{R_C}{R_E} v_b(t). \tag{5.28}$$

The time-varying voltage at the collector is the inverse of that at the base, but multiplied by the ratio R_C/R_E. For $R_C = 10 \cdot R_E$, the output would have an amplitude 10 times larger than the input[2].

The typical application for such a circuit is to amplify a small signal. By properly choosing R_E and R_C, the amplification can be tuned to the desired value. In such a circuit, we want the largest variation possible for v_c, which means that V_C should be about $V_{CC}/2$. This also means that we need to set V_E as close to zero as possible to allow v_c the largest range possible. This then defines the optimal value of V_B, which tends to be small—not much larger than a single diode drop.

Example: Let us design an amplifier circuit with a gain of -10. We will use a supply voltage of $V_{CC} = 12\,V$, and are told that the maximum input signal will have an amplitude of $0.5\,V$. We want $v_b^{tot}(t)$ to be no lower than $0.65\,V$ so that the transistor does not shut off. This gives us a nominal value of $V_B = 1.15\,V$, and of $V_E = .50\,V$.

Figure 5.20 shows the maximum sinusoidal oscillations for v_b and v_e. Because we want v_c to be -10 times v_b, we find that it has an amplitude of $5\,V$. We also know that $v_c^{tot}(t)$ must remain larger than $v_b^{tot}(t)$ for the transistor to stay on. This means that $v_c^{tot}(t)$ must never be less than $1.65\,V$, which allows us to set the nominal value of the collector voltage:

$$V_C = \left[1.65 + \frac{1}{2}\,(12 - 1.65)\right]\,V,$$

or

$$V_C = 6.825\,V\,.$$

This leads to the output, $v_c(t)$, shown in Figure 5.20.

Input and Output Impedance

We should now examine the input and output impedance of the common-emitter amplifier, so the analysis that we carried out for the input to the emitter-follower is still valid for the inverting amplifier. The input impedance is

$$Z_{in} = \beta R_E. \tag{5.29}$$

We can obtain the output impedance from equation 5.27. If we differentiate this with respect to the current, I_C, we get

$$\frac{dV_C}{dI_C} = -R_C,$$

[2]For fans of the movie *Spinal Tap*, one can change R_C to be eleven times R_E. Thus an amplifier that goes to 11 can be built.

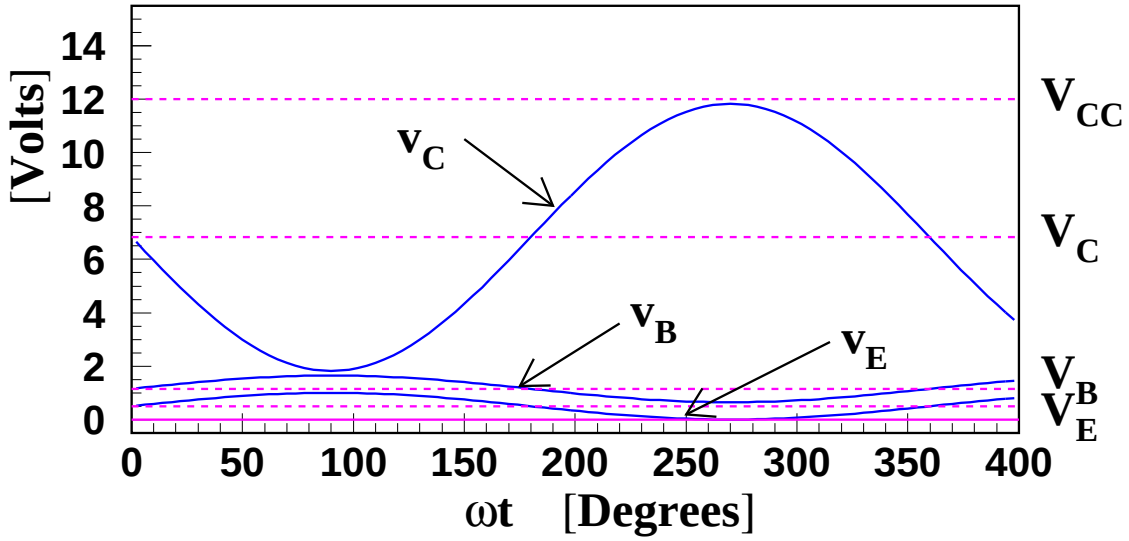

Figure 5.20: Optimized voltage setting for the inverting amplifier described in the example.

so the output impedance is just

$$Z_{out} \quad = \quad R_C \, . \tag{5.30}$$

Figure 5.21 shows the equivalent input and output circuits for the inverting amplifier. While we can arrange that $\beta \cdot R_E$ is reasonably large, we may be limited to a large value of R_C, which might lead to a larger output impedance than we desire.

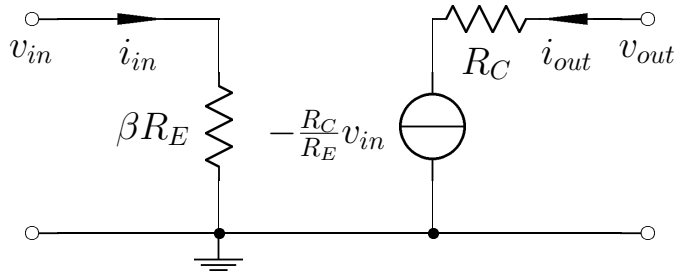

Figure 5.21: The input and output of the common-emitter amplifier. The input impedance is βR_E and the output impedance is R_C.

Example: Let us now carry out the design of the inverting-amplifier circuit shown in Figure 5.22. As we did with the emitter-follower, we use a voltage divider to set the nominal level of V_B, and then use capacitors to form high-pass filters on both the input and output of the circuit. We will use the same $V_{CC} = 12\,V$ as above, and design an amplifier with a gain of -10 for a maximum input signal amplitude of $0.5\,V$. We can take the nominal voltage levels that we found in the previous example as our starting point:

$$\begin{aligned} V_B &= 1.15\,V \\ V_E &= 0.50\,V \\ V_C &= 6.85\,V \end{aligned}$$

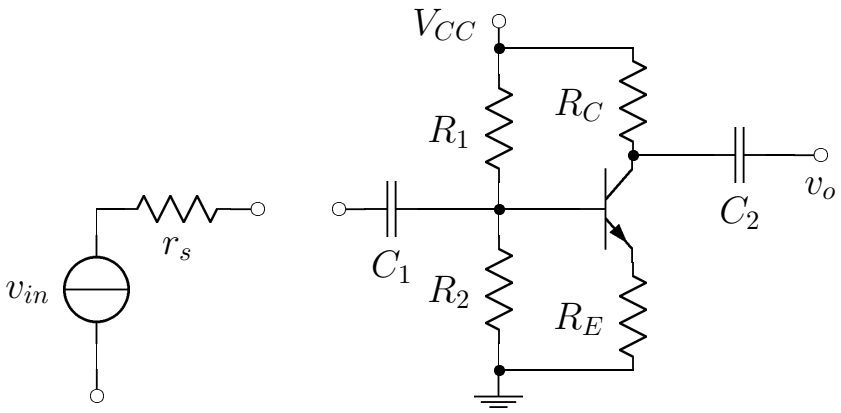

Figure 5.22: An example inverting-amplifier circuit. The output voltage is related to the input voltage by $v_o(t) = -\frac{R_C}{R_E} \cdot v_{in}(t)$.

We will start by choosing R_E such that I_E is limited to $1\,mA$, a number that we will take as a safe operating point of the transistor. The maximum value of $V_E + v_E$ will be $1.0\,V$, so we want

$$\begin{aligned} 1\,mA &= (1.0\,V)\,/R_E \\ R_E &= 1\,k\Omega\,. \end{aligned}$$

From this, we get that $R_C = 10\,k\Omega$. At the quiescent point, $V_E = 0.5\,V$, which gives that $I_E = 0.5\,mA$. Then

$$\begin{aligned} V_C &= 12.0\,V - 0.5\,mA \cdot 10\,k\Omega \\ V_C &= 7.0\,V\,. \end{aligned}$$

We are not able to quite set V_C at $6.85\,V$, but $7.00\,V$ should be good enough. Next, we need to choose R_1 and R_2 to give us $1.15\,V$ at V_B. We can find the ratio of the two that gives us the right voltage:

$$\begin{aligned} 1.15\,V &= \frac{R_2}{R_1 + R_2} \cdot 12\,V \\ 0.0958 \cdot R_1 &= 0.904 \cdot R_2 \\ R_1 &= 9.44 \cdot R_2\,. \end{aligned}$$

In order to proceed, we need to make sure that $R_1 \,||\, R_2$ is small in comparison to $\beta R_E \approx 100\,k\Omega$. To do this, we would like to choose $R_1 \,||\, R_2 \approx 1\,k\Omega$. Putting this all together, we find that

$$\begin{aligned} R_1 &= 1.11\,k\Omega \\ R_2 &= 117\,\Omega\,. \end{aligned}$$

As we did with the emitter-follower, we will insert capacitors C_1 and C_2 into the circuit to "block" DC voltages. The same logic applies as before, and the equivalent high-pass filters for both the input and the output are shown in Figure 5.23. On the input side, the parallel combination is dominated by the smallest of the three resistors, namely R_2, while on the output side, we need to consider the load, R_{LD}, that we will be driving with this circuit. If we want to set the frequency to be $f_c = 5\,Hz$, then we have $\omega_c = 31.4\,s^{-1}$, which should equal $1/\,(R \cdot C)$. On the input side, we have

$$31.4\,s^{-1} \leq \frac{1}{117\,\Omega \cdot C_1}$$

which yields that

$$C_1 \quad \geq \quad 270\,\mu F.$$

On the output side, we do not necessarily know the load, but it probably should be at least 10 times larger than R_C. Assuming this, we get

$$31.4\,s^{-1} \quad \leq \quad \frac{1}{10\,k\Omega \cdot C_2}$$

so that

$$C_2 \quad \geq \quad 3.2\,\mu F.$$

In the following, we list the component values that we would like for this circuit. However, we often may not be able to get those exact values. We also list possible compromise values that are probably easily obtained. If we do make these changes, we will have to ask what effects they will have on our circuit. We hope that they are negligible, and leave this as an exercise for the reader.

Component	Desired Value	Obtainable Value
R_E	$1\,k\Omega$	$1\,k\Omega$
R_C	$10\,k\Omega$	$10\,k\Omega$
R_1	$1.1\,k\Omega$	$1\,k\Omega$
R_2	$117\,\Omega$	$100\,\Omega$
C_1	$270\,\mu F$	$330\,\mu F$
C_2	$3.2\,\mu F$	$3.3\,\mu F$

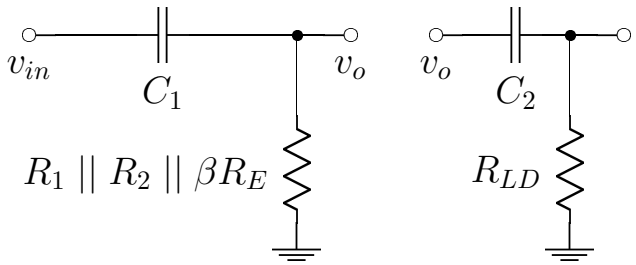

Figure 5.23: The equivalent high-pass filters in the input (left) and the output (right) circuits.

Chaining Together Transistor Circuits

In the previous sections, we have looked at two types of transistor circuits: the emitter-follower circuit and the common-emitter amplifier circuit. Both of these circuits have large input impedance, but only the former is guaranteed to have a small output impedance. By chaining such circuits together, we can build circuits with large input impedance and small output impedance. A feature that we have found to be quite useful.

As an example of this, let us consider the common-emitter amplifier that we discussed in section 5.3.2. We found that while the input impedance of the circuit was large, ($Z_{in} = \beta R_E$ from equation 5.29). The output impedance also tended to be large, ($Z_{out} = R_C$ from equation 5.30). In order to be able to do additional processing on the output of such an amplifier, it would be useful to have a smaller output impedance. This can be accomplished by attaching an emitter-follower circuit to the output of the amplifier. Such a combined circuit is shown in Figure 5.24. The output impedance of this circuit will be given approximately as:

$$Z_{out} \quad \approx \quad \frac{R_3}{\beta} || \frac{R_4}{\beta} || \frac{R_C}{\beta},$$

which is approximately given by the term for the smallest of the three resistors.

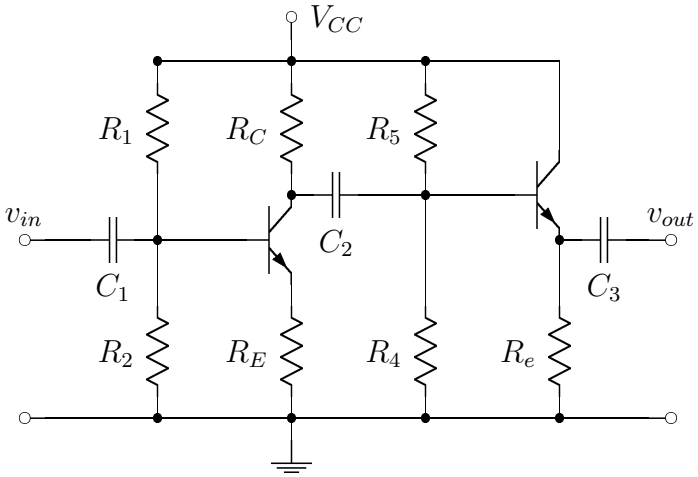

Figure 5.24: A common-emitter amplifier followed by an emitter-follower circuit.

5.3.3 High-Gain Common-Emitter Amplifiers

In the last section, we looked in detail at the common-emitter amplifier, whose gain is found to be $G = -R_C/R_E$. We would now like to explore the consequences of making this gain large—much larger than the nominal factor of ten that we had in the last section. There are two obvious ways to do this: we can increase R_C or we can decrease R_E. As the output impedance of the circuit is given as $Z_{out} = R_C$, increasing R_C immediately leads to a much larger output impedance. Large output impedances are typically not desirable characteristics of circuits. The alternative of decreasing R_E is the typical approach to increasing the gain of the amplifier. However, this can also affect the performance of the amplifier.

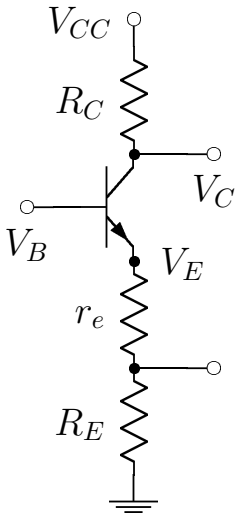

Figure 5.25: The inverting-amplifier circuit. The gain of the amplifier is $G = -R_C/(R_E + r_e)$.

In this section, we will examine the consequences of decreasing R_E. We will find that we can indeed get large gains, but the gain may no longer be linear in input voltages. Before proceeding, we should

recall the results of the Eber-Moll equation (5.19). In particular, the emitter behaves as if there is a small resistance, r_E, in series with the external resistance, R_E (see Figure 5.25). In equation 5.19 we saw that r_E is dependent on both the temperature of the transistor and the total current flowing into the collector, i_c^{tot}. We explicitly summarize that dependence in equation 5.31.

$$r_E = \frac{k_B T/e}{i_c^{tot}} . \tag{5.31}$$

At room temperature ($68^\circ F$), $T = 293.2\,K$, we can simplify this to yield

$$r_E = \frac{25.3\,mV}{i_c^{tot}} . \tag{5.32}$$

Grounded-Emitter Amplifier

In order to increase the gain of the circuit, we will now decrease R_E. In fact, because r_E is finite, we can set $R_E = 0$. Our gain then becomes $G = -R_C/r_E$, which can be quite high.

For a current $i_c^{tot} = 1\,mA$, and $R_C = 10\,k\Omega$, we have a nominal gain of -400. For a $10\,mA$ current, the gain can be 10 times larger. Such a circuit is known as a grounded-emitter amplifier.

Let us now look at this in a bit more detail. As before, there will be DC voltages on the terminals of the transistor, V_B, V_C and V_E. There will also be DC currents flowing and out of the transistor, I_B and I_C in and I_E out. If we consider the case where we only have these DC voltages and currents, then we find the nominal value of the internal resistance, r_E^o, from equation 5.32:

$$r_E^o = \frac{25.3\,mV}{I_C} \tag{5.33}$$

We can now add a time-dependent input voltage, $v_b(t)$ to V_B. This will lead to time-dependent voltages at the emitter and collector:

$$\begin{aligned} v_e(t) &= v_b(t) \\ v_c(t) &= -\frac{R_C}{r_E(t)} v_b(t) . \end{aligned} \tag{5.34}$$

The resistance, $r_E(t)$, can be written as

$$r_E(t) = r_E^o \left[\frac{I_C}{I_C + i_c(t)} \right] .$$

Using the fact that $v_c(t) = -i_c(t) R_C$, we can expand equation 5.34 as

$$-\frac{i_c(t)}{v_b(t)} = \frac{-1}{r_E^o} \left[\frac{I_C}{I_C + i_c(t)} \right] .$$

We can solve this for $i_c(t)$ to yield

$$i_c(t) = \left[\frac{v_b(t) I_C}{r_E^o} \right] \cdot \left[\frac{1}{I_C - v_b(t)/r_E^o} \right]$$

which can be rewritten to yield the gain of the circuit as a function of the input voltage, $v_b(t)$:

$$G = G^o \cdot \left[\frac{1}{1 - v_b(t)/25.3\,mV} \right] \tag{5.35}$$

where $G^o = -\frac{R_C}{r_E^o}$. Here we assume the temperature of the transistor remains constant. Such a function is plotted in Figure 5.26 for values of v_b between $-25\,mV$ and $25\,mV$. This function is both nonlinear and asymmetric about $v_b = 0$.

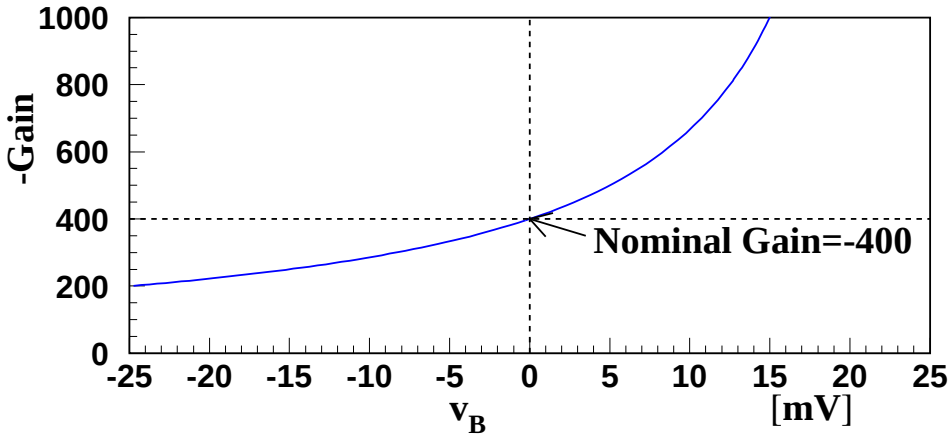

Figure 5.26: The negative of the gain as a function of the input voltage to the emitter-grounded amplifier. The *expected* or *nominal* gain is indicated at $0\,V$ input voltage. As the input goes to the maximum allowed voltage, the magnitude of the gain gets very large.

Let us now consider some numerical examples. We will assume that the nominal gain $G^o = -400$ as expected for $I_C = 1\,mA$ and $R_C = 10\,k\Omega$. The time varying input voltage will be taken as

$$v_b(t) \quad = \quad 25\,mV\cos(\omega t)\ .$$

From this, we nominally expect an output voltage

$$v_c(t) \quad = \quad -10\,V\cos(\omega t)\ .$$

The true gain as a function of v_b is shown in Figure 5.26. Because of the large variation in gain as a function of v_b, this is not a good design for achieving large amplifications. Figure 5.27 shows the output voltage, v_c, as a function of the input voltage, v_b. The dashed line is what we would like for a good amplifier; the solid line is what this circuit delivers. Figure 5.28 shows the result of feeding two triangle waves, of different amplitudes, into the circuit. The dashed line shows the desired triangle wave output, while the solid curve shows what actually comes out of the circuit.

Not only is the gain of this circuit very nonlinear, but the input impedance can also change significantly. In the above example, we would find that

$$Z_{in} \quad = \quad (1+\beta)r_E\,. \tag{5.36}$$

For typical values of $r_E = 25\,\Omega$ and $\beta = 100$, this yields that $Z_{in} \sim 2.5\,k\Omega$. This may already be uncomfortably small for some purposes. However, the large variations in r_E as a function of i_C^t will lead to large variations in the input impedance. This is clearly not a desirable feature.

Lastly, while we have assumed that the temperature of the transistor is constant, in fact this may not be the case. The power dissipated in the transistor depends on the current. In equation 5.31, we see that the value of r_E varies linearly with T. If the transistor temperature increases by about $30°C$, r_E will increase by about 10%.

The reason the gain is so nonlinear is that r_E changes by so much. This, in turn, is because I_C and $i_c(t)$ are of similar size, which leads to large variations in $i_c^{tot}(t)$. In the common-emitter amplifier in section 5.3.2, this was limited by choosing values of R_E that were large compared to r_E. While the changes in r_E were large, the changes in $R_E + r_E$ were very small, and the amplifier had a very linear gain. For very small (relative to $25\,mV$) signals, this circuit could provide an approximately linear response, but even though it is sometimes used, it is generally considered a poor circuit design.

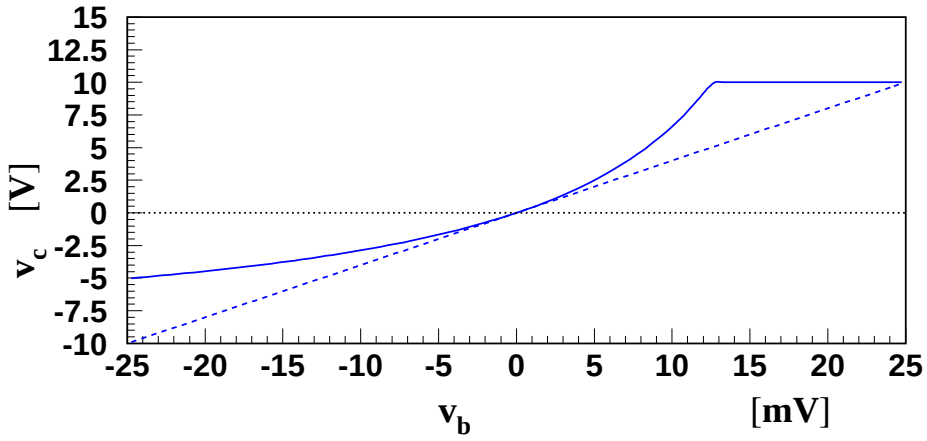

Figure 5.27: The output voltage, v_c as a function of the input voltage, v_b. The dashed line is what is expected for the *nominal gain*. The plateau for large input voltage corresponds to the output saturating and reaching the supply voltage.

Bypassed Common-Emitter Amplifier

The problems that we encountered in the grounded design traced back to the fact that changes in $i_c^{tot}(t)$ led to large changes in r_E. If we could set some finite level for I_E which did not depend on the exact value of r_E, we could improve the situation. This is the idea behind the bypassed common-emitter amplifier. Rather than shorting the emitter directly to ground, we can modify the circuit slightly, to that shown in Figure 5.29.

In this circuit, we still have an emitter resistor, R_E, which is chosen to be about $\frac{1}{10}R_C$. However, we bypass the emitter resistor, R_E, with a capacitor, C_E. The value of C_E is chosen such that the magnitude of the impedance, Z_C, for all interesting frequencies is much smaller than the nominal value of r_E. The capacitor provides an AC ground, while the resistor provides the DC connection to ground. If a total current i_e^{tot} flows out of the emitter, then I_E flows through the resistor to ground and $i_e(t)$ flows through the capacitor to ground. For the AC signal, the gain is again given by $G = -R_C/r_E$. However, unlike the previous example, the variations in the possible values of r_E are much less extreme because there is a large DC current, I_C, to which we add the AC current, $i_c(t)$. Unfortunately, as we will see in the following, this circuit still suffers from a non-linear gain, but because we have moved V_E away from ground, we do not have the transistor in a situation where it is very close to the on/off point.

To understand the behavior of this circuit, we again choose some specific values. We set $V_{CC} = 20\,V$, $R_C = 10\,k\Omega$ and $R_E = 1\,k\Omega$. We also want to set the nominal DC level at the emitter to be $1\,V$. From this, we get that $I_C \approx I_E = 1\,mA$ and the nominal collector voltage is $V_C = 10\,V$. Next, we want to choose a capacitor, C_E whose impedance is very small compared to r_E for all frequencies of interest. If we take a cutoff frequency of $\omega = 10\,s^{-1}$, then $C_E > 1/(\omega r_E)$. For a nominal $r_e = 25\,\Omega$, we choose $C_E > 4\,mF$. The nominal AC gain is then $G = -R_C/r_E \approx -400$, while for very low frequencies, it will be $-R_C/R_E = -10$.

However, now let us look at a little more detail. If we would like to amplify a $25\,mV$ signal, $v_b(t)$, then we have that $v_e(t)$ follows v_b, and the current, $i_e(t)$, will just be $v_e(t)/r_E$. Putting in the numbers, we find the amplitude of $i_e(t)$ is also $1\,mA$. This will lead to exactly the same problems that we saw before. However, for signals much smaller than $25\,mV$, this amplifier will have much better behavior than the simple grounded version.

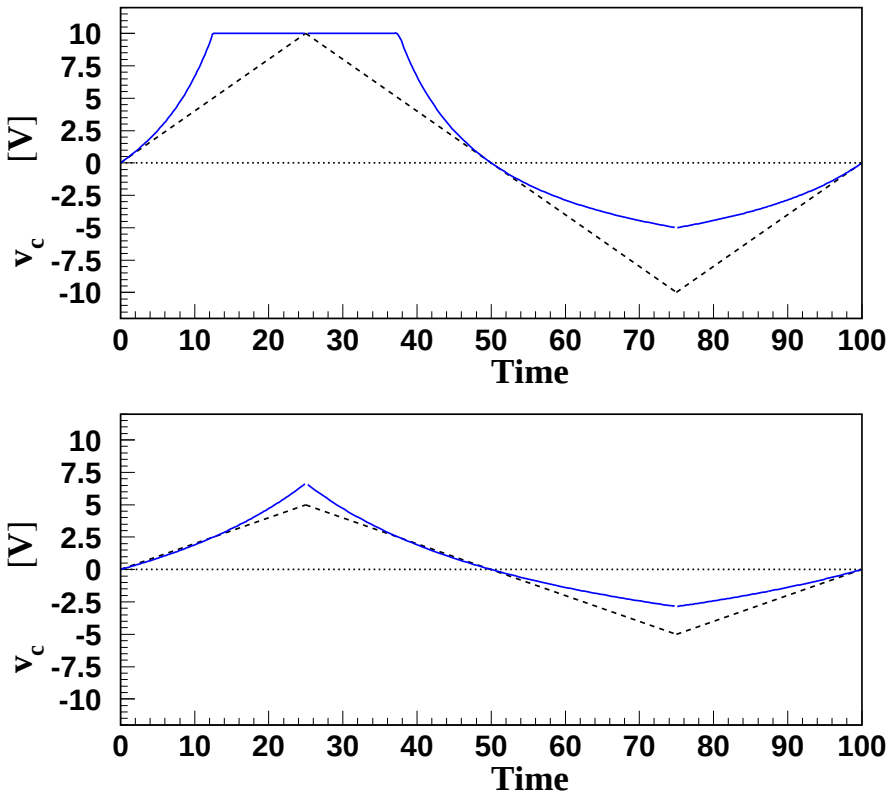

Figure 5.28: The response of the grounded-emitter amplifier to triangle wave inputs. The expected triangle waves are shown as dashed lines while actual responses are shown as solid lines.

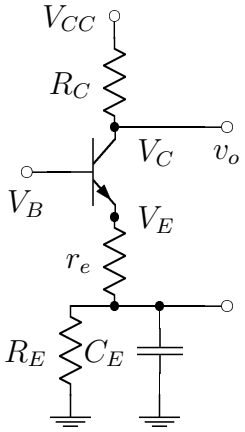

Figure 5.29: A bypassed common-emitter amplifier circuit.

A Partially-Bypassed Amplifier

Figure 5.30 shows a partially-bypassed common-emitter amplifier circuit. We have put the resistor R_B in series with the capacitor C_E, which, according to our earlier analysis, will give us a gain of

$$G = -\frac{R_C}{R_B}$$

in the high-frequency region, while the gain will be $-R_C/R_E$ in the low-frequency region. If we choose R_B large relative to r_E, but smaller than R_E, we can produce a much more linear response with a higher gain.

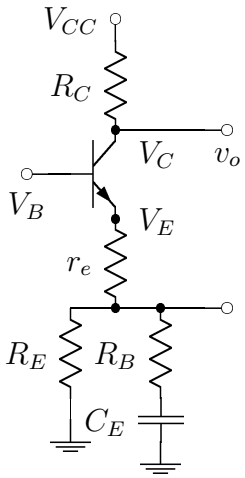

Figure 5.30: A partially-bypassed common-emitter amplifier.

5.4 Field Effect Transistors

In addition to the bipolar transistors that we discussed in the earlier part of this chapter, other types of transistors are also in common use. In this section, we will discuss a class of transistors known as *field effect transistors*, or simply FETs. As the name implies, their output is controlled by the strength of an electric field inside the transistor.

In fact, FETs are much more common in the semiconductor industry than bipolar transistors[3]. One advantage is that they have significantly higher input impedance than the bipolar transistor, which in turn means that the current drawn at the input to the transistor is smaller.

All FETs rely on the behavior of moving charges in semiconductors. The conductance of a semiconductor material can be found by combining equations 4.4 and 4.5 into a single expression. If we limit ourselves to an n-type semiconductor with only majority carriers, then the conductance is

$$G = q\mu_e n\left(\frac{A}{L}\right) . \tag{5.37}$$

In order to to control the current flow in a semiconductor using the relation

$$I = G \cdot V , \tag{5.38}$$

we need to be able to control the value of conductance. In equation 5.37, we see that this can be accomplished by varying the mobility, μ_e, the concentration of charge carriers, n, or the geometry of

[3]It is interesting to note that the first patent for a field effect transistor was awarded to Julius Edgar Lilienfeld in 1926, nearly twenty years earlier than the bipolar transistor was invented. See U.S. Patent numbers 1,745,175 and 1,900,018.

the conductor, A/L. Field effect transistors use the electric field in the semiconductor to change the geometry of the conducting path, and hence the conductance.

Like bipolar transistors, field effect transistors have three terminals, though they are named differently. Figure 5.31 shows the symbol for an n-channel (a) and p-channel (b) FET. The three terminals are known as the *gate* (G), *source* (S), and *drain* (D). The potential difference between the gate and the source, V_{GS}, is used to control the current flowing between the drain and the source, I_D. For certain regions of operation, we will find that I_D does not depend on the voltage between the drain and source, V_{DS}—it only depends on V_{GS}. Hence, we will arrive at an I-V curve similar to that of the bipolar transistor.

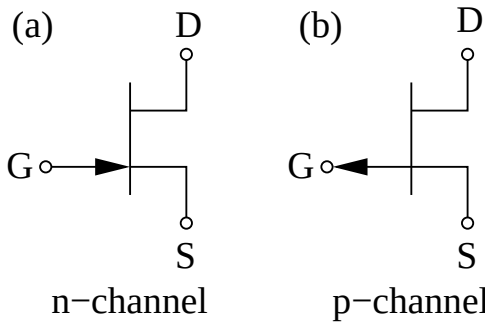

Figure 5.31: The symbol for an n-channel (a) and p-channel (b) field effect transistor. The three terminals are the gate (G), the source (S) and the drain (D). Control over the conductance of the FET is achieved by changing the voltage difference between the gate and the source.

5.4.1 Junction Field Effect Transistors

The junction field effect transistor, or JFET, is based on a single semiconductor junction. Figure 5.32 shows a cross-sectional slice of an n-channel JFET. The source and drain are connected to the ends of the same n-type semiconductor. A p-type semiconductor, to which the gate is connected, is deposited on top the n-channel. If the potential difference between the gate and the sink is zero, then the conductance, G, is a maximum value, and the relation between the current I_D and the drain-source voltage, V_{DS}, is given by equation 5.38.

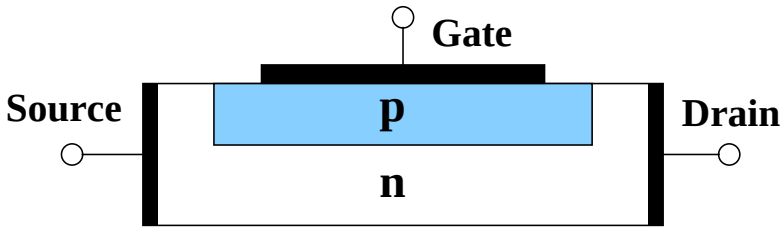

Figure 5.32: A junction field effect transistor (JFET). Current normally flows through the n-type semiconductor between the drain and the source in the conduction channel. The gate is a p-type semiconductor into which electrons from the conduction channel can diffuse.

However, the junction field effect transistor uses the voltage difference, V_{GS}, to control the conductance, G. If we reverses-bias the junction such that the gate is at a lower potential than the source, then the electrons in the n-channel will be pulled into the p-type semiconductor, thus depleting the conductors in the channel. This *depletion region* will decrease the cross-sectional area of the conduction path, and, based on equation 5.37, will decrease G. As V_{GS} becomes more negative, we will eventually

reach a point where the n-channel is entirely depleted and the conduction will go to zero. The voltage at which this happens is known as the *pinch voltage*, V_P (note that in an n-channel JFET, V_P is always negative). When $V_{GS} \leq V_P$, then $G = 0$. The pinching off of the conduction channel is shown in Figure 5.33(a). Similarly, Figure 5.33(b) shows the I-V curves for three different values of V_{GS}: maximum conduction at $V_{GS} = 0$, some intermediate value when V_{GS} is some fraction f of the pinch-voltage, and finally $G = 0$ when $V_{GS} \leq V_P$.

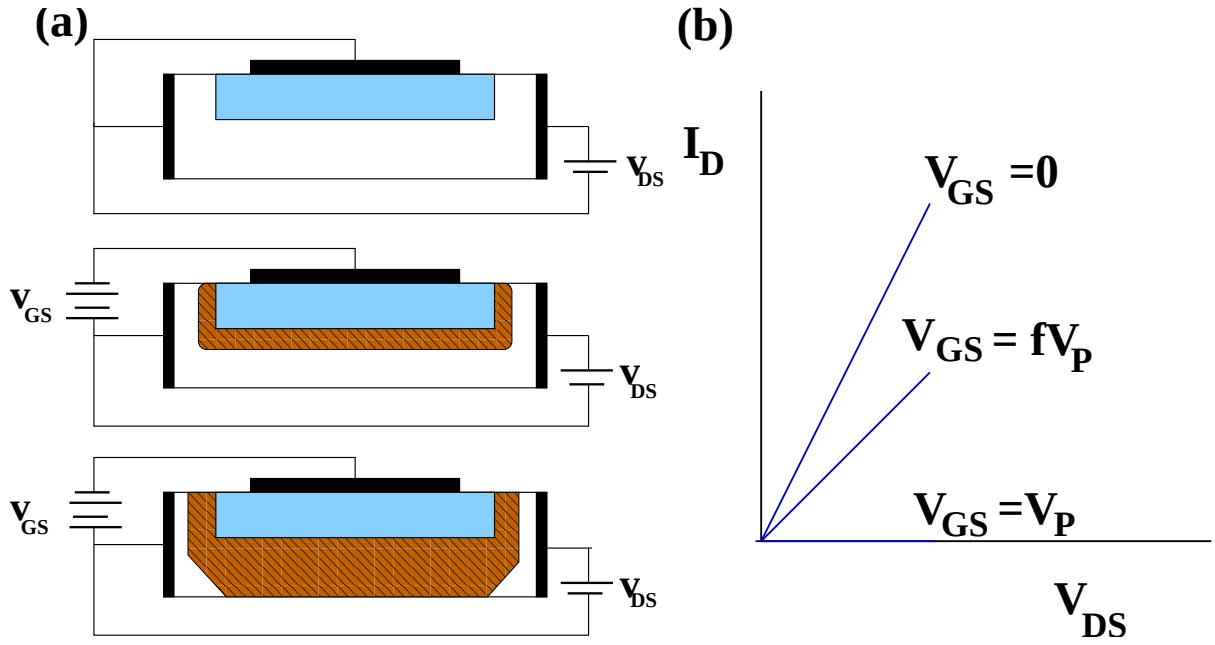

Figure 5.33: (a) A biased JFET. In the upper figure, there is no potential difference between the gate and the source. The conduction channel has its maximum width. The center figure has the gate-source bias at some fraction of the pinch voltage. The conduction channel is narrower. In the bottom picture, the gate-source voltage is equal to the pinch voltage and the conduction channel is closed. (b) I-V curve for the JFET whose gate is at the same potential as the source.

As a note, if we forward-bias the junction with V_{GS}, the conduction does not increase beyond what it is at $V_{GS} = 0$ as there is no easy way to increase the concentration of conductors in the n-channel beyond its unbiased value. Hence V_{GS} for a JFET ranges between V_P and 0 (recall that V_P is negative).

In order to have current flow from the drain to the source, we need to apply a voltage, V_{DS}. For V_{DS} smaller than the magnitude of the pinch voltage, the previous description is accurate. However, as V_{DS} approaches $\mid V_P \mid$, the biasing causes the n-channel to start to pinch off at near the drain. When V_{DS} reaches $\mid V_P \mid$, the n-channel will completely pinch off as shown in Figure 5.34. However, rather than the current going to zero, we find that I_D becomes fixed at a constant value. In fact, this value is controlled by the gate-source voltage, V_{GS}, as shown in Figure 5.35. We have effectively achieved the same I-V curve that we had for a junction bipolar transistor. The maximum current in the JFET occurs when $V_{GS} = 0$, and is known as I_{DSS}. This value is typically quoted on the specification sheets for the JFET.

Though we will not quantitatively explain the flattening out of the I-V curve of the JFET, we can at least understand it qualitatively. While the reverse bias of the junction has depleted all the mobile charge carriers in the pinched-off region, it does not prevent charge carriers from flowing across the "pinch" as long as there is a sufficiently large potential difference. As such, current can continue to flow through the channel. It becomes constant because the increased potential between the drain and the source nearly cancels the increased resistance due to the growing pinched region. If $R_{DS} = \alpha R_0$ and

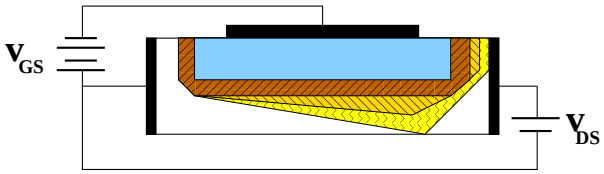

Figure 5.34: A biased JFET with a fixed gate-source voltage that is some fraction of the pinch voltage. The progressively larger pinched-off regions correspond to increasing values of the drain-source voltage. Eventually this gets large enough to pinch off the conduction channel.

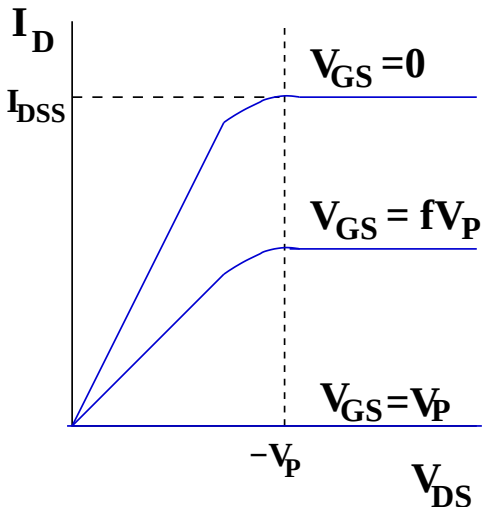

Figure 5.35: The I-V curve for a JFET. Once V_{DS} becomes larger than V_P, the current I_D becomes independent of V_{DS}. However, it can still be controlled by V_{GS}. The constant current for $V_{GS} = 0$ is known as I_{DSS}.

$V_{DS} = \alpha V_0$, then the current, given as V_{DS}/R_{DS}, is just a constant. We simply state that this is what happens, without further justification. In Figure 5.36 are shown the drain current, I_D as a function of the gate-source voltage, V_{GS}. For sufficiently large V_{GS}, I_D is proportional to V_{GS}. It is in this proportional region that I_D is independent of the drain-source voltage.

Finally, we note that the bipolar junction transistor worked with a forward-biased pn-junction, while the JFET works with a reverse-biased pn-junction. According to the characteristic diode curve shown in Figure 4.17, this means that the input current drawn by the JFET will be significantly smaller than than for the bipolar transistor. Hence, the input impedance of the JFET will be much larger. However, because we have a reverse-biased junction, we know that we eventually reach the breakdown voltage of the junction, at which point significant current will begin to flow into the gate. This breakdown limits the maximum value of V_{DS} in the JFET.

Example: As with the bipolar transistor, a follower circuit can be built using a JFET. Such a circuit is shown in Figure 5.37. An external supply voltage, V_{DD}, is connected to the drain, and the source is connected to ground via a resistor, R_S. While we will not show it, the output impedance of this circuit tends be larger for the JFET circuit than for the bipolar transistor. This arises because the resistance corresponding to r_E in the bipolar transistor is larger in the JFET, and this value sets the output impedance. Values of a few hundred ohms are typical.

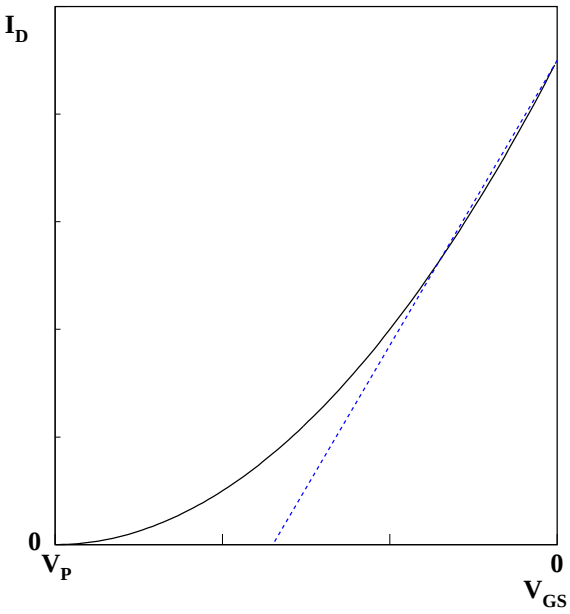

Figure 5.36: The transfer curve for a JFET. The drain current, I_D as a function of the gate-source voltage, V_{GS}. For large enough V_{GS}, I_D is proportional to V_{GS} as shown by the dashed line in the figure.

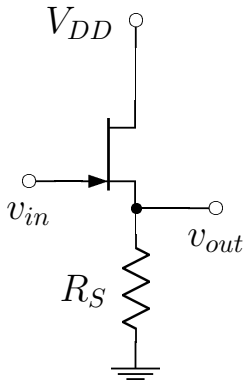

Figure 5.37: A follower circuit built using an n-channel JFET. The time-varying output voltage will follow the time-varying input voltage.

Example: We can also build an amplifier circuit using a JFET. This is shown in Figure 5.38. As with the follower circuit, the internal resistance of the n-channel will cause the gain of this circuit to be smaller than the corresponding gain of the bipolar transistor circuit.

5.4.2 Metal-Oxide-Semiconductor Field Effect Transistors

Similar to the JFET is the *metal-oxide-semiconductor* field effect transistor, the MOSFET. In this device, the p-type semiconductor is replaced with a thin insulating layer. On top of this layer is a thin metal layer that is connected to the gate. It may also be true that different n-type semiconductors

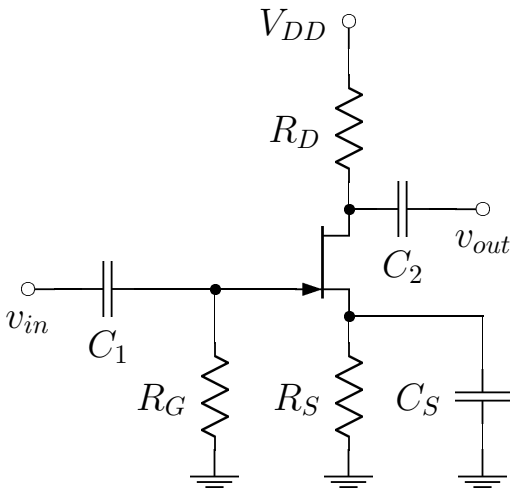

Figure 5.38: An inverting amplifier built using an n-channel JFET.

are used near the drain and source than in the conduction channel. Figure 5.39 shows a side view of a typical MOSFET device. The symbol for the MOSFET (Figure 5.40) is slightly different from that for the JFET (Figure 5.31). In particular, many MOSFETs have a fourth connection to the bulk, or substrate, material in the transistor. This is indicated as B in Figure 5.40. However, many MOSFETs have an internal connection between B and S which might also be indicated on the symbol.

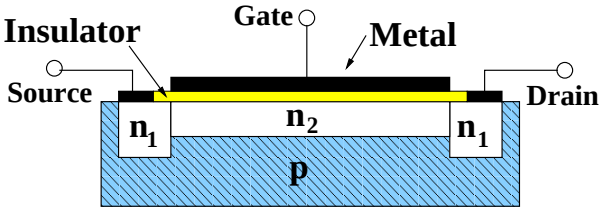

Figure 5.39: A metal-oxide-semiconductor field effect transistor (MOSFET). The two two n-regions labeled n_1 have a larger concentration of charge carriers than the region labeled n_2. The p-region only serves as a mechanical substrate. A metal electrode is separated by a thin insulator from the conduction channel, n_2, which functions as a capacitor. When the gate is at lower voltage than the source, the capacitor creates a net positive charge along the insulator-n_2 junction. This removes negative charge carriers from the conduction channel.

The difference between a JFET and a MOSFET is that when a voltage is applied between the gate and the source, V_{GS}, the transistor behaves as if the metal plate and the n-channel are the two sides of a capacitor. When the source is at a higher potential than the gate ($V_{GS} < 0$) the same description that we had for the JFET applies. The is known as the *depletion mode*. However, in the case where the gate is at higher potential, then we can actually inject additional charge carriers into the conducting channel. The is known as *enhancement mode*.

We will now introduce without proof two equations which describe the I-V curve of the MOSFET. Equation 5.39 describes the region where the conduction channel has not been pinched off, while equation 5.40 describes the transistor behavior when the channel has been pinched off.

$$i_D = A_0 \cdot \left[(v_{GS} - V_P) \cdot v_{DS} - \frac{1}{2} v_{DS}^2 \right] \tag{5.39}$$

$$i_D = \frac{1}{2} A_0 \left(v_{GS} - V_P \right)^2 \tag{5.40}$$

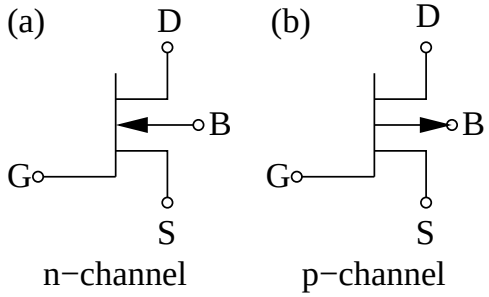

Figure 5.40: The symbols for n- and p-channel MOSFETs. The fourth connection indicated in the figures (B) connects to the bulk substrate of the transistor. This is often internally connected to the source terminal of the transistor.

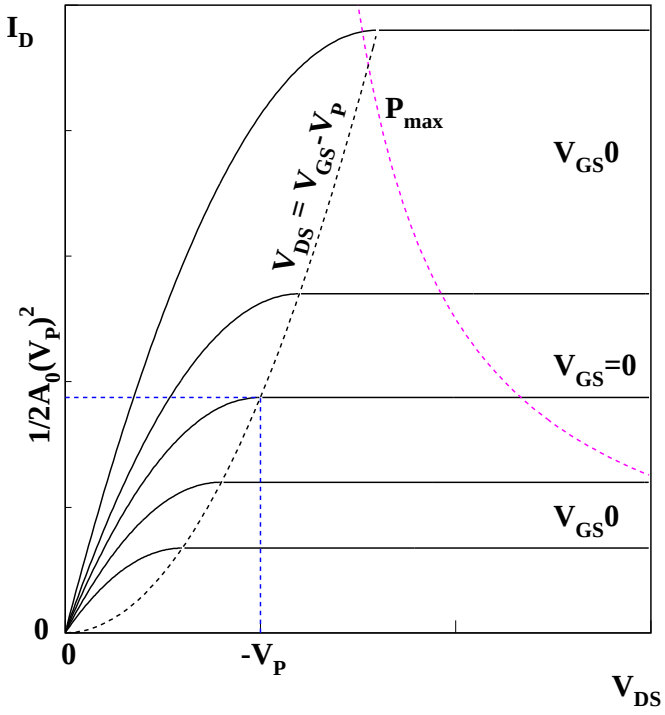

Figure 5.41: The I-V curve for a MOSFET. The horizontal dashed line indicates where $V_{GS} = 0$. Above this line, the transistor is in the enhancement regime. Below the line, it is in the depletion regime. The curved line labeled P_{max} indicates the maximum power dissipation of the transistor.

In the depletion regime, equation 5.39 is valid when $v_{DS} \leq (v_{GS} - V_P)$ and $(v_{GS} - V_P) \geq 0$. Equation 5.40 is valid for $v_{DS} > (v_{GS} - V_P)$ and $(v_{GS} - V_P) \geq 0$. When the MOSFET is in the enhancement regime, the pinch voltage is replaced with a threshold voltage V_T. The same equations and limits apply after this substitution.

Finally, we consider the maximum power of the MOSFET. This is indicated by the P_{max} curve in Figure 5.41. The maximum power curves indicate when breakdown begins. Either for the pn-junction

in a JFET, or the insulating layer in the MOSFET. Unfortunately, breakdown of the insulating layer in a MOSFET usually results in permanent damage, it not destruction of the insulating layer. Related to this is the extremely large input impedance at the gate of a MOSFET. This can be damaged by the accumulation of static charge. When MOSFETs are not in use, the leads should be shorted together and the the device only handled by its case. Similarly, inserting a MOSFET into a powered circuit is considered a "bad idea".

Problems

1. An npn transistor will be used in a circuit where its nominal base voltage is $V_B = 2.0\,V$. If a resistor R_E connects the emitter to ground, what should be chosen for the value of R_E if the nominal value of the current I_E should be $10\,mA$?

2. The transistor in problem 1 is connected to an external supply of voltage $V_C = 10\,V$. How much power is dissipated in the transistor at its normal operating point?

3. An npn transistor is operating with a nominal base current of $I_B = 10\,\mu A$ and a nominal emitter current of $I_E = 1\,mA$. What are β and α for the transistor?

4. An npn transistor has a nominal operating point of $I_C = 100\,mA$. What is the value of internal emitter resistance r_E at room temperature? What is r_E if the transistor heats up to $50°C$?

5. An npn transistor has a nominal operating point of $I_C = 1\,mA$. What is the value of internal emitter resistance r_E at room temperature? What is r_E if the transistor heats up to $50°C$?

6. An npn transistor has a nominal operating point of $I_C = 10\,mA$. If there is also a time dependent current $i_C(t)$ of amplitude $8\,mA$, what is the ratio of minimum to the maximum value of r_E as i_C is changing?

7. An emitter-follower circuit is built using a transistor with $\beta = 100$ and with $R_E = 1000\,\Omega$. What is the input impedance of the open circuit? If a load of $100\,\Omega$ is connected to the output of the circuit, what is the input impedance?

8. An emitter-follower circuit is built using a transistor with $\beta = 150$ and with $R_E = 1500\,\Omega$. It is connected to a source with output impedance $1000\,\Omega$. What is the output impedance of the emitter-follower?

9. The emitter-follower in problem 8 is connected to a source with output impedance $100\,k\Omega$. For what load resistance will the output sag below one-half its nominal value?

10. A common-emitter amplifier is built with $R_E = 100\,\Omega$ and $R_C = 1.5\,k\Omega$. What is the gain of the amplifier?

11. A common-emitter amplifier is to be driven with a power supply of voltage $V_{CC} = 10\,V$. If the maximum amplitude of the input signal is $1\,V$, what should be the nominal value of V_C to have the maximum output range on $v_C(t)$?

12. A common-emitter amplifier is to be built with a gain of -25. This implies that the ratio of R_C to R_E should be 25. What criteria other would you consider in choosing actual values for R_C and R_E?

13. In the discussion of the grounded emitter amplifier, we found the gain expression as given in equation 5.35. What is the maximum value of v_b such that the gain does not vary by more than 5% from its nominal value?

14. In the discussion of the grounded emitter amplifier, we found the gain expression as given in equation 5.35. If the temperature of the transistor is $50°C$, what is the maximum value of v_b such that the gain does not vary by more than 5% from its nominal value?

15. You have built the circuit shown in Figure 5.42 and are using it to make measurements. The supply (DC) voltage, V_{CC}, is applied at point **A**. Unless otherwise noted, express your answers in terms of V_{CC}, R_1, R_2, R_E, R_C, C_1 and the β of the transistor. (**a**) Correctly label the *base*, *emitter* and *collector* on the npn transistor in the above circuit. (**b**) If the nominal (DC) voltage level at the base, V_B, is supposed to be $\frac{1}{10}$ of the controlling voltage, V_{CC}, what is the ratio, $\frac{R_2}{R_1}$?

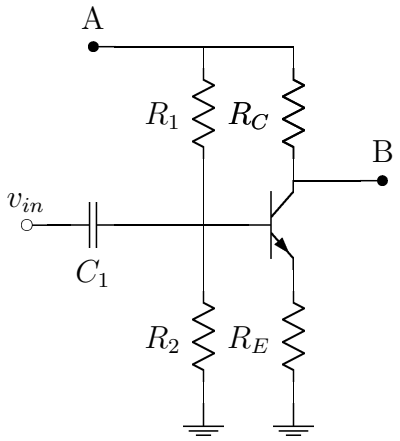

Figure 5.42: The circuit for problem 15.

(**HINT**: Only consider R_1 and R_2 in the voltage divider.) (**c**) What are the (DC) voltage level, V_E, and the current , I_E, flowing through the resistor R_E? Express your answers in terms of V_B and R_E. (**d**) What are the current, I_C, flowing through the resistor R_C, and the voltage at the collector of the transistor, V_C? Express your answer in terms of V_B, R_E, R_C and V_{CC}. (**HINT**: is it safe to ignore I_B?) (**e**) A small changing voltage, v_{in}, is now added to the base voltage, V_B. What is the *change* in the collector voltage away from V_C, v_o, due to this small input change? Now assume that $V_{CC} = 10.0V$ and that the static value of $V_B = 1.0V$. In addition, let $R_C = 10 \times R_E$. (**f**) A sinusoidal input voltage, $v_{in} = 0.5\cos(\omega t)$, is applied to the circuit as shown in the figure. On the same plot of voltage versus time, sketch v_{in} and v_o. Sketch them both for exactly one period starting at time $t = 0$. Be sure to accurately label your plot. Also, make sure to identify any possible clipping of the output voltage. (**g**) The capacitor, C_1, forms a high–pass filter for the input voltage, v_{in}. Describe the method by which you would determine the resistance, R_{eff}, that goes into this filter (do not evaluate R_{eff}). What resistor **in the circuit** is probably closest in value to R_{eff}? (**h**) If you wanted the $3dB$ point on the filter to be at $f = 10\,Hz$, what values would you choose for C_1? Express your answer in terms of R_{eff} from the previous part. (**HINT**: A numerical answer is NOT possible).

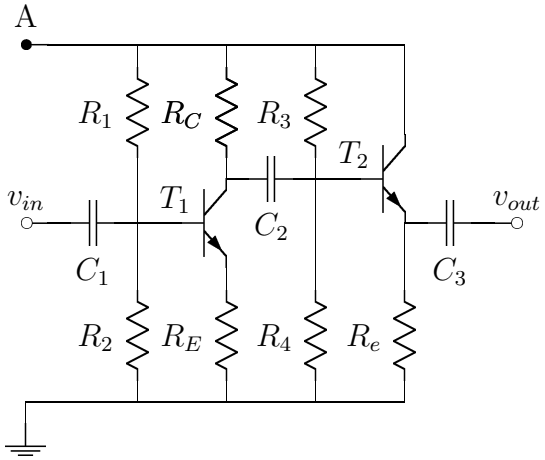

Figure 5.43: The circuit for problem 16.

16. The transistor circuit shown in Figure 5.43 is constructed using two identical transistors, T_1 and

T_2 (current gain $\beta \approx 100$). It contains three capacitors (C_1, C_2, C_3) and seven resistors (R_1, R_2, R_3, R_4, R_C, R_E, R_e). Initially, you know that $R_C = 10R_E$. A *DC* voltage , V_{CC}, is connected between **A** and ground such that point **A** is positive. A small, time-varying input voltage, $v_{in}(t)$, is connected as shown. This produces an output voltage, v_{out}, as shown in the figure. (**a**) If $v_{in} = 0$, what are the approximate voltages at the bases of the two transistors, T_1 and T_2? (Express your answer in terms of V_{CC}, R_E, R_e, R_1, R_2, R_3, R_4, C_1, C_2 and C_3; *assume a well designed circuit!*) (**b**) Give reasonable estimates for the voltage at the base of the transistor T_2 which would allow this circuit to function optimally. Briefly explain your choices. (**c**) Explain the purpose of the three capacitors shown in the circuit including any relevant frequencies. (**d**)Assume that both transistors are biased such that they are in their linear operating range, and an input voltage of $v_{in}(t) = 0.001V_{cc}\cos\omega t$ is applied (where ω is chosen to be in the normal operating range for the circuit). What is the open circuit output, $v_{out}(t)$? Explain your reasoning. (**e**) As seen from the output terminal, v_{out}, what is the Thèvenin equivalent circuit? Estimate the values of the Thèvenin impedance and voltage. For what frequency range is your answer applicable?

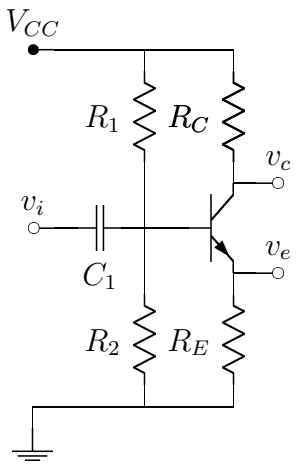

Figure 5.44: The circuit for problem 17.

17. Consider the transistor circuit shown in Figure 5.44. Assume the transistor is in its normal operating range, unless otherwise specified. The transistor has a current gain β. V_{CC} is a DC voltage supplying power to the circuit and v_i is a time-varying input voltage. v_c and v_e are the time-varying parts of the output voltage (what you would see if you AC-coupled your scope). You may assume that the circuit has well-chosen resistor values. Express your answers in terms of β, V_{CC}, v_i, R_1, R_2, R_C and R_E and state approximations that you make due to the circuit being well designed. (**a**) Assume that $v_i = 0$. What are the voltages at the base of the transistor, V_B, at the emitter of the transistor, V_E, and at the collector of the transistor, V_C? (**b**) Capacitor C_1 forms an RC filter with some total resistance R. What is R and what kind of RC filter have we formed? (**c**) Assume that $R_C = R_E \equiv R$. What are v_e and v_c in terms of v_b? (**d**) Using this circuit, we want to drive a load R_L which is similar in size to $R_E = R_C = R$. Is it better to drive the load using v_c or v_e? Explain why.

18. A circuit is built as shown in Figure 5.46. Assume the transistor is in its normal operating range, unless otherwise specified. The transistor has a current gain β and V_{CC} is a DC voltage to control the circuit. You are told that $R_1 = R_2 = R_E = R$. You may assume that the circuit has well-chosen resistor values. Be sure to list any approximations that you make due to the circuit being well designed. (**a**) What are V_B, V_C and V_E in terms of β, V_{CC} and R? (**b**) What are I_B, I_C and I_E in terms of β, V_{CC} and R? Be sure to indicate if these currents flow "into" or "out of" the transistor. (**c**) Define I_1 as the current through R_1 and I_2 as the current through R_2. Assuming

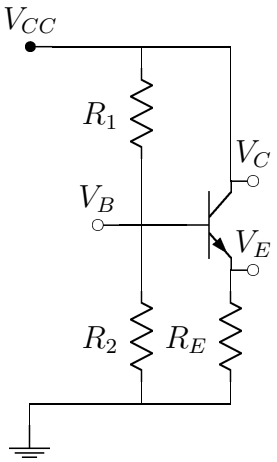

Figure 5.45: The circuit for problem 18.

that there is no current flowing into the base of the transistor, compute I_1 and I_2. Compare your values to I_B from part **b**. In particular, is I_B smaller than, similar to, or larger than I_1 and I_2? (**d**) What is the Thèvenin equivalent output resistance if we use V_E and ground as our output terminals?

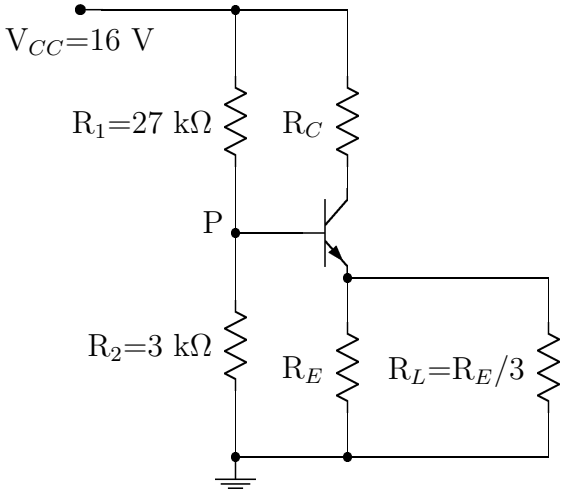

Figure 5.46: The circuit for problem 19.

19. An npn transistor, having $\beta = 100$, is connected as shown to $V_{CC} = 16\,V$ and to ground using resistors $R_1 = 27\,k\Omega$, $R_2 = 3\,k\Omega$, R_C, R_E, and $R_L = R_E/3$. You may assume the transistor circuits in this problem are properly biased to operate normally. You may make **reasonable** approximations, but explain what approximations you are making. (**a**) What would the be the value of the potential, V_P, expected at point P if the base of the transistor were *not* connected to point P? (**b**) What are the requirements (if any) on the values of R_E and R_C if V_P is to not to be loaded down significantly when the base of the transistor is connected to point P? (Remember that $R_L = R_E/3$.)? Now R_E and R_L are chosen as $R_E = 600\,\Omega$ and $R_L = 200\,\Omega$. Capacitors C_1, C_2, and C_3 are added to allow a small A.C. signal, $v_i(t)$ to enter the circuit and to allow A.C. signals v_C and v_E to be extracted from the circuit. (**c**) If v_i is known to be made up of a range of frequencies between $f_{min} = 100\,Hz$ and $f_{max} = 40\,kHz$, what are the requirements on the value

of C_1 to allow the signals to pass through the circuit without being distorted? (Clearly indicate the value of any relevant resistance used in the calculation and don't forget factors of 2π.) Assume C_1, C_2, and C_3 have been correctly chosen so the signals won't be distorted. (You may neglect the impedance of the capacitors at the frequencies of interest.) (**d**) If $R_C = 0$ (i.e. the resistor is replaced with a piece of wire) , find v_E and v_C. (**e**) If $R_C = 2400\,\Omega$, find v_E and v_C. (**f**) The output impedance of the device which **generates** v_i is $R_i = 50\,\Omega$. Approximately what is the output impedance at the point where v_E is measured? (Neglect the impedance of all capacitors at the frequencies of interest.)

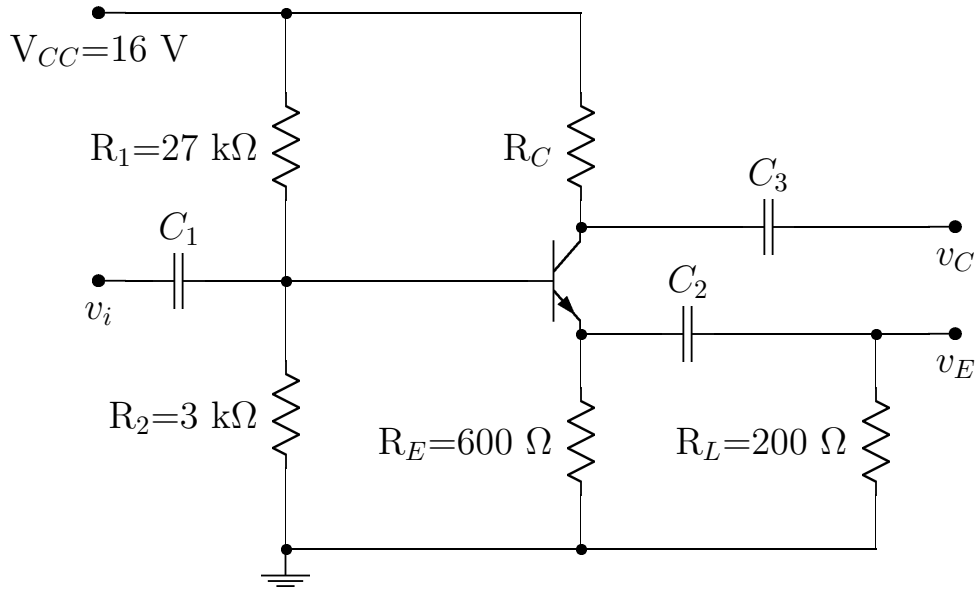

Figure 5.47: The circuit for problem 19 c.

Chapter 6

Feedback and Operational Amplifiers

6.1 Introduction

In many situations, the mechanism of *feedback* is used to stabilize the behavior of a system. The basic idea is that part of the output of the system is fed back into the input, either in-phase or out-of-phase with the input signal. The in-phase case is referred to as positive feedback, while the out-of-phase is known as negative feedback. In general, positive feedback tends to be unstable, with the output being driven away from its non-feedback values. Negative feedback has the opposite effect. It tends to make systems more stable, holding an output near a particular desired value.

We consider systems in which the output is some function of the input, but is limited between two values. One use of positive feedback is to quickly drive the system to one of the two limiting values.

6.2 Negative Feedback

Negative feedback is the process of subtracting some fraction, β, of the output from the input signal. This is shown schematically in Figure 6.1. Some input signal, v_{in}, goes into a device with gain A_o, which produces an output (with no load) of $v_{out} = A_o v_{in}$. We now subtract a portion of the output, βv_o, from the input. In the figure, this is shown as

$$v_{sum} = v_{in} - \beta v_{out} .$$

This allows us to write

$$v_{out} = A_o \cdot v_{sum} ,$$

or

$$\frac{v_{out}}{v_{in}} = \frac{A_o}{1 + \beta A_o} .$$

We refer to A_o as the open-loop gain of the device, and can define the overall gain (with feedback) to be A_f, where

$$A_f = \frac{A_o}{1 + \beta A_o} . \tag{6.1}$$

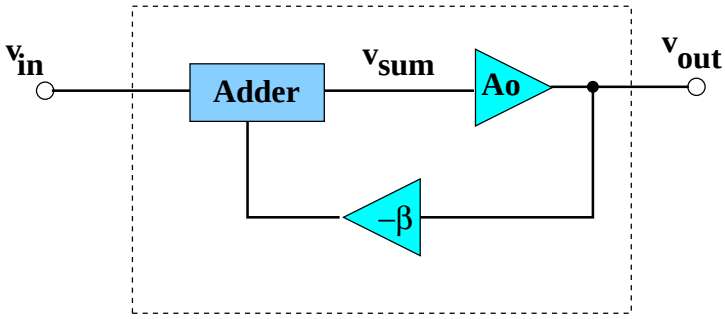

Figure 6.1: A negative feedback loop. A fraction, $-\beta$, times the output signal is added back to the input signal: $v_{sum} = v_{in} - \beta v_{out}$.

Typical circuits in which we employ feedback have very large open-loop gains. If βA_o is much larger than one, then the overall gain can be written as

$$A_f \approx \frac{1}{\beta} . \tag{6.2}$$

At first, this might not seem very useful. However, β is usually something that we can control very precisely, while the exact value of A_o may not be stable. In the case of high-gain transistor circuits, this was clearly true. A_o might also have some frequency dependence. What equation 6.2 says is that we can give up some of the gain, A_o, to get a very stable gain, A_f. This is generally a very good thing, and one benefit of negative feedback.

As an example of this, consider a situation where we have set $\beta = \frac{1}{100}$, while A_o varies from 5000 up to 500000 in some fashion over which we have little or no control. If we put the two limiting values of A_o into equation 6.1, we get:

$$\begin{aligned} A_f(A_o = 5000) &= \tfrac{5000}{1+5000/100} &= 98.0 \\ A_f(A_o = 5000000) &= \tfrac{5000000}{1+5000000/100} &= 99.98 \end{aligned}$$

As A_o changes by a factor of 1000, A_f changes by less than 2% . This is an extreme situation; more typically A_o might change by 10 to 20%, which would then lead to an extremely small change in A_f as long as βA_o is large compared to one.

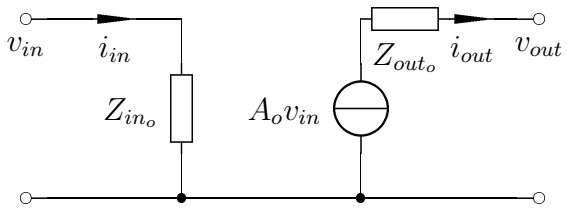

Figure 6.2: The equivalent circuit for a common-emitter amplifier circuit with gain A_o, input impedance Z_{in_o} and output impedance Z_{out_o}.

We now consider an amplifier whose equivalent circuit is shown in Figure 6.2. The circuit has an open-loop gain of A_o, input impedance Z_{in_o} and output impedance Z_{out_o}. We have already examined what happens to the gain of the circuit when we add negative feedback. Let us now look what happens to the input and output impedance of the circuit. Figure 6.3 shows the same circuit with negative feedback.

The voltage across the input impedance drops from v_{in} to $v_{sum} = v_{in} - \beta v_{out}$. We can also write

$$v_{sum} = i_{in} \cdot Z_{in_o} \ .$$

Combining this equation with the ones from above, it is possible to write that:

$$v_{in} \quad = \quad i_{in} Z_{in_o} + \beta \frac{A_o}{1 + \beta A_o} v_{in} \ .$$

The input impedance is just the ratio of v_{in}/i_{in}, which allows us to solve for Z_{in_f}:

$$Z_{in_f} \quad = \quad (1 + \beta A_o) \, Z_{in_o} \ . \tag{6.3}$$

The input impedance with feedback is substantially larger than without feedback. This is certainly a very good property. It means that the input of such a circuit will draw very little current, which in turn will not load down the output of the previous stage. It is also relatively easy to understand this. Because the negative feedback is working *against* the input, the system draws less current than it would without the feedback. This then yields a larger input impedance.

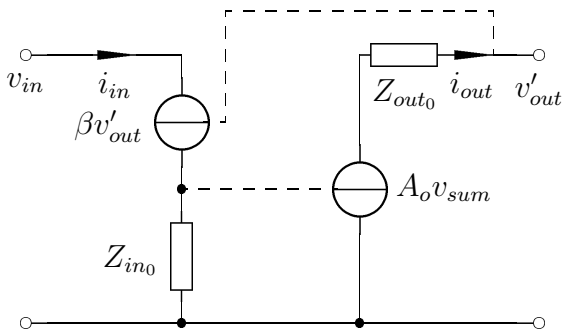

Figure 6.3: The equivalent circuit for an amplifier with gain A_o and a fraction β of the output voltage connected to the input as negative feedback. The input impedance, Z_{in_0}, and output impedance, Z_{out_0}, are those for the circuit without feedback.The output voltage is $A_f v_{in}$ rather than the open-loop $A_o v_{in}$.

Similarly, we can examine the output impedance of the circuit. If we look at the voltage drops on the output side of the circuit, we can write

$$v'_{out} \quad = \quad A_o v_{sum} - i_{out} Z_{out_o} \ .$$

Expanding v_{sum}, we find

$$v'_{out} \quad = \quad A_o \left(v_{in} - \beta v'_{out} \right) - i_{out} Z_{out_o} \ ,$$

which can be solved for the current

$$i_{out} = \frac{A_o v_{in}}{Z_{out_o}} - \frac{1 + \beta A_o}{Z_{out_o}} v'_{out} \ .$$

From this we see that the output impedance with feedback is

$$Z_{out_f} \quad = \quad \frac{Z_{out_o}}{1 + \beta A_o} \ . \tag{6.4}$$

The output impedance with feedback is smaller than that without feedback by $1/\left(1 + \beta A_o\right)$. In the approximation that we have high open-loop gain, this is quite a bit smaller than one.

Thus, negative feedback on a high-gain amplifier gives us very large input impedance, very small output impedance, and precise control of the gain.

6.2.1 A High-Gain Amplifier

Figure 6.4 shows a high-gain amplifier built using three transistors. The first two stages are common-emitter amplifiers, while the third stage is an emitter-follower circuit. The small-signal gain in the first the stages has been increased by partly bypassing the emitter resistors with capacitors.

This design incorporates negative feedback, which is accomplished with a voltage divider consisting of R_{f_1} and R_{f_2}. The output signal that is fed back into the input is in phase with the input signal. This feedback is negative because it causes v_{BE} of the first stage transistor to be *smaller* than it would be without the feedback. The feedback is controlled by the two resistors in the feedback divider:

$$\beta = \frac{R_{f_2}}{R_{f_1} + R_{f_2}}. \tag{6.5}$$

We could well imagine building up ever more complicated amplifiers with larger numbers of transistor and feed-back stages. In fact, this has been done for us, and it has been packaged in a convenient form. So, rather than worrying what we would need to do to build such a device, we will simply add it to our list of convenient components. This device is discussed in the following sections.

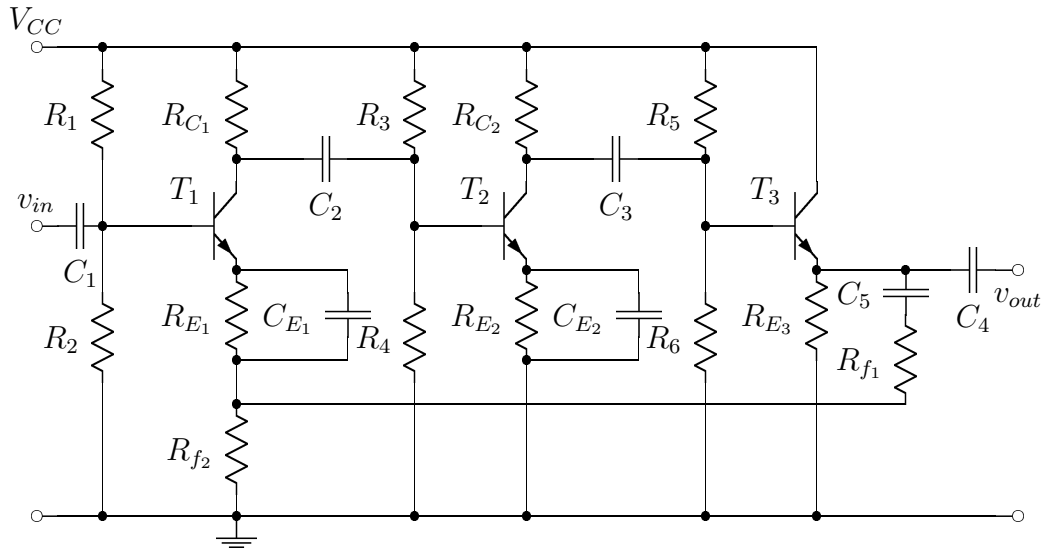

Figure 6.4: Two stages of high-gain transistor amplifiers followed by an emitter-follower circuit. The output signal goes through C_5 and R_f and feeds back into the emitter of the first stage as negative feedback.

6.3 Op-Amps

In the preceding sections, we found that by adding negative feedback to a high-gain amplifier, we can build a very stable amplifier whose gain is precisely controlled by external circuit elements. While we have examined transistor circuits that can function as basic amplifiers, a commonly-used type of high-gain amplifier is the *Operational Amplifier* or *Op-Amp*. These are integrated-circuit amplifiers which internally have a large number of transistors. Op-Amps are designed to have very large open-loop gain, very large input impedance and very small output impedance. Figure 6.5 shows an op-amp with two inputs *non-inverting* and *inverting* and and output. The figure also explicitly shows the external power connections: V_{CC} and V_{EE}. Typically, these are not shown in a circuit, but they do need to be hooked up in order for the op-amp to work. The output of the op-amp can be driven very close to either of these two supply voltages. While in the symbol shown in Figure 6.5, the inverting input is below the non-inverting input, these are also drawn reversed. The + and − in the figure indicates what the two inputs

are. In this book, we will draw the op-amp symbol in both ways—to make the overall circuit clearer. In building such circuits, it is important to identify which input is inverting and which is non-inverting.

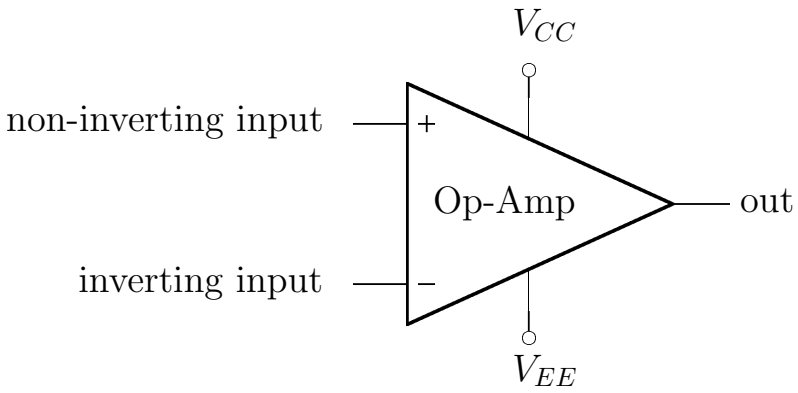

Figure 6.5: An op-amp explicitly showing the two power connections, V_{CC} and V_{EE}. There are two inputs, labeled *inverting* and *non-inverting*, and the output.

As long as the desired output is within the range defined by the supply voltages, the output is given as

$$v_{out} = A_o (v_+ - v_-) \tag{6.6}$$

where v_+ is the voltage at the non-inverting input and v_- is the voltage at the inverting input. The open-loop gain, A_0, of an op-amp can be very high, with $A_o \sim 10^5$ not being unusual. If the desired output is outside of this range, the output will be driven close to one of the two limits and the equation 6.6 is not valid.

Op-amps are normally only used with feedback, where we take advantage of the high open-loop gain to build a very stable, lower-gain amplifier. To understand how op-amp circuits work, we start with some approximations that allow us to get a very good idea what a particular circuit will do. These assume that we have *ideal op-amp behavior*, which is defined as follows:

1. The open-loop voltage gain is infinite, $A_0 = \infty$.
2. The input impedance to the op-amp is infinite, $Z_{in} = \infty$, which means that the two op-amp inputs draw no current.
3. The output impedance of the op-amp is zero, $Z_{out} = 0$.
4. The op-amp can change the output voltage instantaneously (the slew rate is infinite).
5. When both inputs are at the same voltage, the output voltage is zero.
6. The op-amp performance does not depend on variations in temperature.
7. The op-amp performance does not depend on variations in supply voltage.

From these approximations , we arrive at two rules that allow us to understand op-amp circuits.

The first rule is based on the fact that the op-amp has a very large input impedance. Typical values are in the range of 10^6 to $10^{12}\,\Omega$. This means that very little current actually flows into the input of an op-amp. While it is not exactly zero, we can approximate it as zero for our calculations. This leads to our first rule.

1. *No current flows into the inputs of an op-amp.*

The second rule comes from the fact that the open-loop gain of an op-amp is very large. The only way to get an output that is not at one of the two supply voltages is to have the difference in the two inputs in equation 6.6 be very small. Otherwise, when the difference is multiplied by the very large A_o, the output will be huge. In an op-amp with feedback, we can approximate the difference as exactly zero, which leads to our second rule.

2. *The feedback in the op-amp circuit drives the voltage differences between the inverting and non-inverting inputs to zero.*

These two rules are both easy to apply, and give us a very good first-order understanding of circuit behavior. They are of course both approximations, so it is possible to describe the system more precisely, but often we do not need to do that. In the following, we will apply these rules to several op-amp circuits. In a section 6.4 we will revisit the basic assumptions that went into these rules and look in detail at what consequence the actual ope-amp behavior has on an op-amps operation.

6.3.1 The Op-Amp Follower Circuit

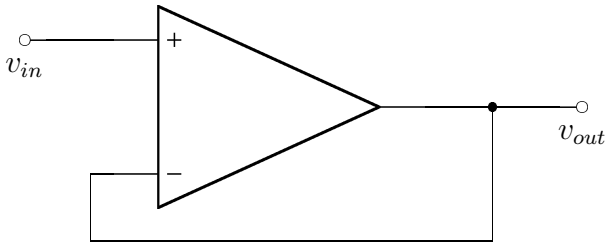

Figure 6.6: A follower circuit built with an op-amp.

Figure 6.6 shows an op-amp follower circuit. Our first rule states that the neither of the two inputs draws any current. While this does not help in analyzing the behavior of this circuit, it does tell us that the op-amp will not load down the circuit driving it. The second rule states that the difference between the two inputs is zero. Since the output is fed directly into the inverting input, this means that

$$0 \quad = \quad v_{in} - v_{out}\,,$$

which then yields

$$v_{out} \quad = \quad v_{in}\,.$$

The follower circuit is useful in the same situations where we used the transistor emitter-follower circuit. Its important property is that it acts to transform impedance. Figure 6.7 shows the equivalent circuit. The crucial point is that Z_{in} is very large (approximated as ∞) while Z_{out} is very small (approximated as 0).

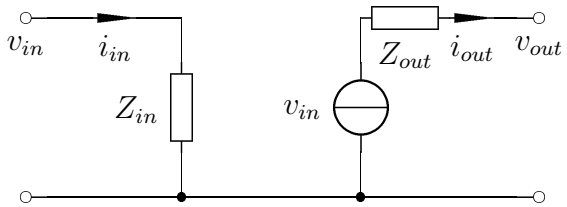

Figure 6.7: The equivalent for an op-amp follower circuit.

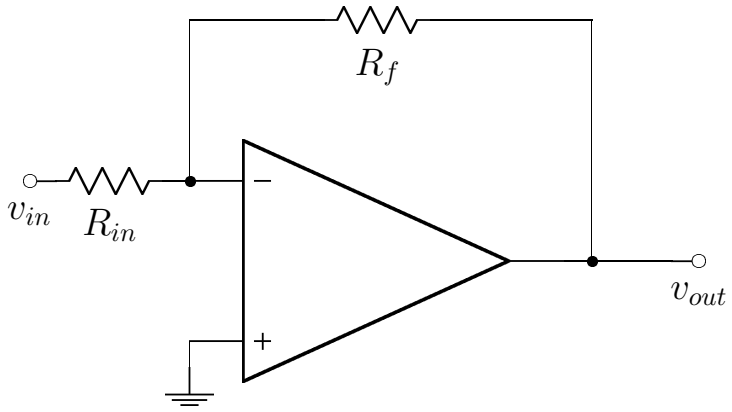

Figure 6.8: An inverting amplifier circuit built with an op-amp. The gain is $G = -R_f/R_{in}$. Note that the op-amp here has its inputs reversed relative to the symbol. In particular, the non-inverting input is grounded.

6.3.2 An Inverting Amplifier

The circuit shown in Figure 6.8 is an inverting amplifier (notice that the input signal is now applied to the inverting input). The feedback is accomplished with resistors R_f and R_{in}. To understand the behavior of this circuit, we again apply the two op-amp rules. The second rule tells us that the voltage at the inverting input should be the same as that at the non-inverting input. The latter is grounded, so we say that the inverting input is a *virtual* ground. We now know that there is a potential drop, v_{in}, across the input resistor, R_{in}, which means that a current

$$i_{in} \quad = \quad v_{in}/R_{in}$$

flows through the resistor, R_{in}. From the first op-amp rule, none of this current flows into the op-amp, so it must all flow through R_f to the output. The voltage drop across R_f gives us the output voltage

$$v_{out} \quad = \quad 0 - i_{in}R_f \,,$$

which means that

$$v_{out} \quad = \quad -\frac{R_f}{R_{in}} v_{in} \,.$$

With appropriate choice of resistor values, the output can be much larger than the input with the gain given as

$$G \quad = \quad -\frac{R_f}{R_{in}} \,.$$

6.3.3 A Non-inverting Amplifier

The circuit shown in Figure 6.9 is a non-inverting amplifier. To understand the behavior of this circuit, we take advantage of the second op-amp rule to note that the feedback must make the voltage at the inverting input be v_{in}. The two resistors, R_1 and R_2, then form a voltage divider. The first rule says that no current flows into either input, so with input v_{out} across both resistors, the output across R_2 is just v_{in}. This gives

$$v_{in} \quad = \quad \frac{R_2}{R_1 + R_2} v_{out} \,.$$

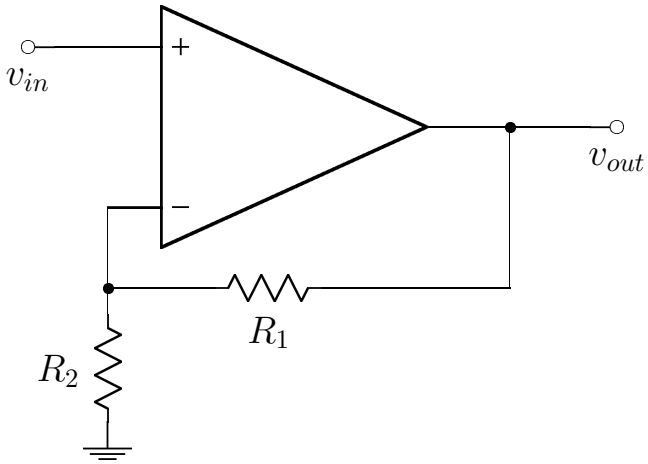

Figure 6.9: A non-inverting amplifier circuit built with an op-amp. The gain is $G = 1 + R_1/R_2$.

This can be rearranged to yield the gain of the circuit

$$G = \frac{(R_1 + R_2)}{R_2} \text{ or}$$
$$G = 1 + \frac{R_1}{R_2}.$$

6.3.4 A Differential Amplifier

In many applications, small signals need to be sent through relatively long cables before reaching electronics which will process them. Unfortunately, such signals are susceptible to noise during transmission. One way to reduce this problem is to transmit a differential signal. Two lines are run from the source to the receiver. The signal, v_s, is transmitted in one line, while a negative copy of the signal, $-v_s$, is transmitted along the other. It is normal for the two lines to be wound around each other in what is referred to as a *twisted pair*. If noise, v_n, is picked up during transmission, then both lines will have picked up the same noise. Rather than v_s and $-v_s$, the two signals will now be $v_1 = v_n + v_s$ and $v_2 = v_n - v_s$.

If, at the receiver end, we now combine the two signals back together such that $v_r = v_1 - v_2$, then with a little math, we see that $v_r = 2 \cdot v_s$. The noise has been subtracted out. This is the power of a differential signal.

In order to do the subtraction, we need to have a differential amplifier. An example of such a circuit is shown in Figure 6.10. The signal, v_1, goes into the non-inverting input, while the inverted signal, v_2 goes into the inverting input. We will find that the output voltage is

$$v_{out} = \frac{R_2}{R_1}(v_1 - v_2) \ .$$

It is very important to match the two pairs of resistors. In building this circuit, the use of precision resistors is probably warranted. It might also be useful to hand-select the resistors as well.

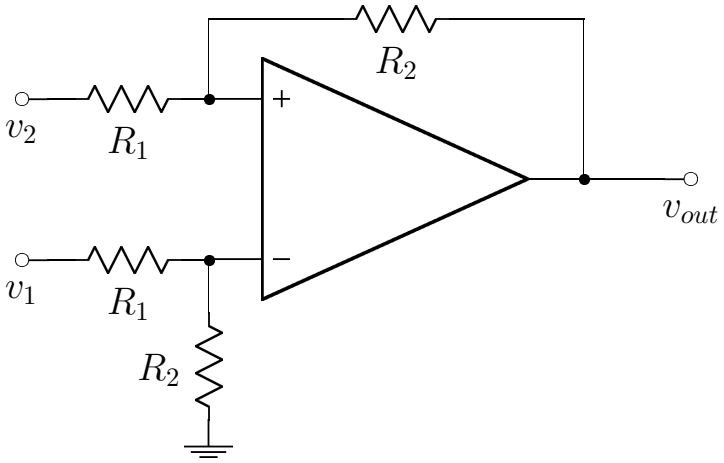

Figure 6.10: A differential amplifier circuit built with an op-amp. The gain is $G = R_2/R_1$.

We can analyze the behavior of our circuit using our two op-amp rules as before. At the inverting input, we have a voltage given as

$$v_{inv} = v_1 \cdot \frac{R_2}{R_1 + R_2} \ .$$

This means that the voltage at the non-inverting input will be equal to this. The voltage drop across R_1 going into the non-inverting input is

$$v_{noninv} = v_2 - v_1 \cdot \frac{R_2}{R_1 + R_2} \ .$$

This gives that the current through R_1 is:

$$i_{noninv} = \frac{v_2}{R_1} - \frac{R_2}{R_1}\frac{v_1}{R_1 + R_2}$$

This then yields that the output voltage is

$$v_{out} = v_{noinv} - i_{noinv} R_2 \,,$$

which can be simplified to yield

$$v_{out} = \frac{R_2}{R_1} (v_1 - v_2) \,. \tag{6.7}$$

6.3.5 The Adder Circuit

The circuit shown in Figure 6.11 produces an output equal to a weighted sum of the input voltages and is known as an *adder* circuit. Applying the second op-amp rule, we have that the inverting input must be a virtual ground because the non-inverting input is grounded. This means that the voltage drop across each of the input resistors must be equal to their corresponding input voltage. Thus the current through each resistor is

$$i_{1,2,3} = v_{1,2,3}/R_{1,2,3} \,.$$

As no current flows into the inputs of the op-amps, the sum of all the currents must flow through R_f to the output.

$$v_{out} = 0 - (i_1 + i_2 + i_3) R_f$$

Putting these together, we get

$$v_{out} = 0 - \frac{R_f}{R_1} v_1 - \frac{R_f}{R_2} v_2 + - \frac{R_f}{R_3} v_3 \,.$$

In the case where $R_1 = R_2 = R_3 = R_f$, the above equation reduces to

$$v_{out} = -(v_1 + v_2 + v_3) \,. \tag{6.8}$$

The output is the negative sum of all the inputs to the op-amp.

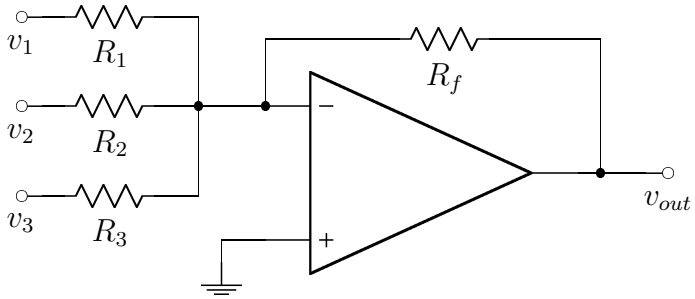

Figure 6.11: An op-amp adder circuit. Note that the op-amp symbol has the non-inverting input grounded.

6.4 Limits on Op-Amp Performance

In the previous section, we treated op-amps as if they had ideal characteristics. As one may expect, physical op-amps deviate from this ideal in several ways. In the following, we discuss the nature of these deviations and the consequences of some of them. We also note that there are a large number of different op-amps on the market, many of which are carefully designed to minimize the effects of one or more of the following problems. In designing a real circuit, it is quite likely that one can obtain an op-amp that is well suited to the task at hand.

6.4.1 The Voltage Gain

First, we assumed that the op-amp had infinite open-loop gain. In fact, while the gain is large, it is finite and also falls off with increasing frequency. Typical op-amp open-loop gains are very large in the range of $\approx 10\,Hz$, and fall off about 20 dB per decade. The frequency at which the open-loop gain is equal to 1 is referred to as f_T. Figure 6.12 shows the gain for a couple of typical op-amps. The figure also shows a line for a typical feedback gain, $A_f = 100$. The feedback gain cannot exceed the open-loop gain. When the curves get close together, the A_f curve will follow the open-loop gain curve, but will always be smaller than A_0.

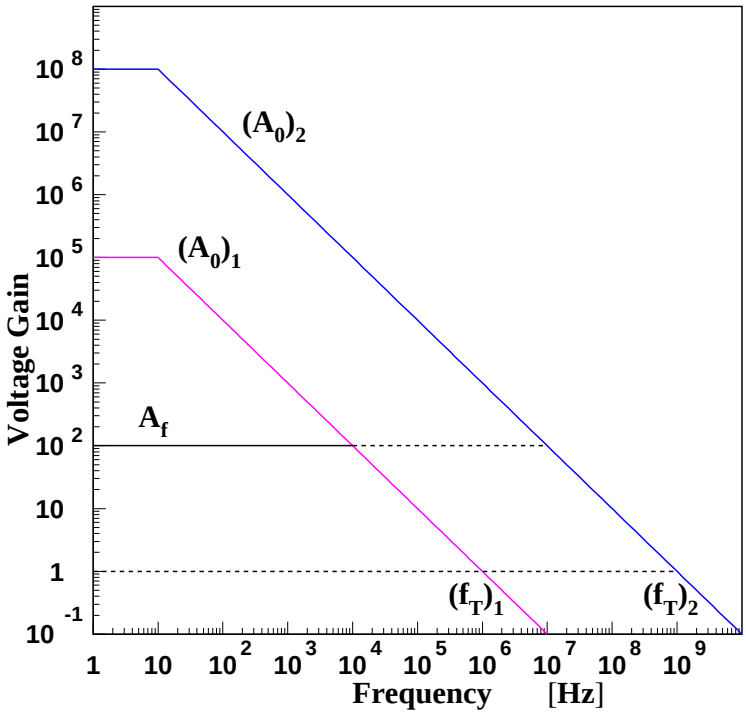

Figure 6.12: The open-loop gain as a function of frequency for two different op-amps. The frequency at which the open-loop gain is equal to 1 is known as f_T. Also shown is a curve for $A_f = 100$. When this curve gets close to the open-loop gain, it will fall off, staying just slightly below the A_0 curve.

6.4.2 Phase Shifts

In an open-loop configuration, the output is about 90° out of phase with the input at the point where the voltage gain begins to fall. The phase shift rises towards 180° as input frequency increases. At f_T, where the open-loop gain is one, the difference between the phase shift and 180° is known as the phase margin. While we will not use the phase margin in this text, it is a number that one would find on the specification sheet for an op-amp.

6.4.3 The input impedance and input current

Because the op-amps are built out of transistors, they need to draw some small but non-zero input current. This current is known as the input bias current, I_B. This is one half the sum of the two input currents with the inverting and non-inverting inputs held at the same voltage. These currents are just the base or gate currents of the input transistors of the op-amp. For bipolar junction transistors, typical values are measured in tens of nA, while field effect transistors have values typically measured in tens of pA.

Related to this is the input impedance of the op-amp, which has been approximated as infinite. With the use of feedback, the input impedance tends to be very large. The fact that it is not infinite does not tend to be a very important parameter in op-amp performance.

Example: Consider the inverting amplifier shown in Figure 6.8. Let us assume that we are using a 741 op-amp whose base current is about $30\,nA$. If we have a $1\,mV$ input signal that we want to amplify to $100\,mV$, we choose $R_{in} = 1\,k\Omega$ and $R_f = 100\,k\Omega$. Assuming that the inputs need to be at nearly the same potential, we find that the current through the input resistor is

$$i_{in} \quad = \quad 1.0\,\mu A\,.$$

This current is split, with $30\,nA$ flowing into the op-amp input and the remaining $970\,nA$ flowing through the feedback resistor. This means that the output voltage is

$$v_{out} \quad = \quad 97\,mV\,,$$

which is about 3% lower than the expected ouput. This could be reduced by using smaller resistors so that there is a larger input current, or by using an op-amp with a smaller base current such as the 411 ($I_{base} = 0.2\,\mathrm{nA}$).

6.4.4 Input Range

The inputs to an op-amp are limited to a finite voltage range. This is typically something less than the supply voltages of the op-amp. It also may not be symmetric. Many op-amps can run closer to the positive supply than the negative supply. If these ranges are exceeded, the behavior of the op-amp can be quite unexpected. Things like phase reversal, or output being pushed to one of the limits can occur.

6.4.5 The Slew Rate

The rate at which the output voltage of an op-amp can change is known as the slew rate. This is typically measured in $V/\mu s$. If we have a square wave input to an op-amp circuit, then the output will be limited by the slew rate as shown in Figure 6.13(a). While this is an extreme case, the same problem can apply to other types of inputs. If an input voltage has time dependence of $\sin \omega t$, and produces an output voltage of $v_{out} = V_o \sin(\omega t)$, then the rate of change of this output voltage is

$$\frac{dv_{out}}{dt} \quad = \quad \omega V_0 \cos(\omega t)\ .$$

The maximum rate of change is ωV_0. If this is larger than the slew rate of the op-amp, then the signal will be distorted near the zero-crossings of the output voltage (Figure 6.13(b)).

6.4.6 Input Offset Voltage

The two op-amp inputs are generally not identical, with the differences due to manufacturing. This imbalance can be easily seen by setting an op-amp up with the same voltage applied to both inputs and no feedback. If the inputs were identical, the output would be zero. However, it is normally either

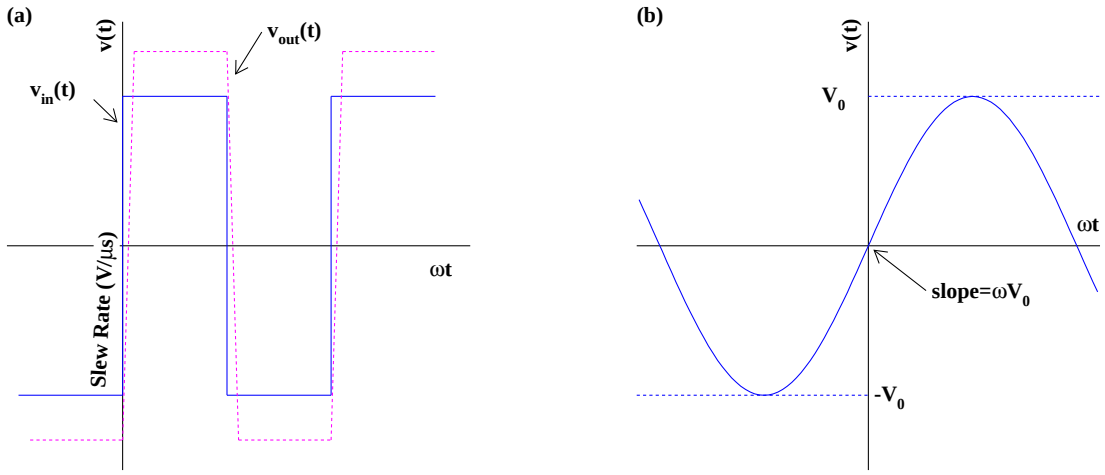

Figure 6.13: The slew rate limits how fast an op-amp can change the output voltage. For the square-wave input (a), the output has some maximum rate of change ($V/\mu s$). This is known as the slew rate. For a sinusoidal function (b), the maximum rate of change is at the zero crossing, ωV_o. If this is larger than the slew rate of the op-amp, the op-amp will distort the sine wave.

V_{CC} or V_{EE}. In order to compensate for this, most op-amps have two offset inputs. Using a 10 kΩ potentiometer as shown in Figure 6.14, it is possible to *trim* the op-amp so that the voltage at the two inputs are the same. The output will still not be zero, but the two inputs will be much closer to identical.

For example, the LF411 op-amp has an average offset voltage of 0.8 mV and a maximum of about 2 mV. Unfortunately, these values also have a temperature drift of about 0.007 mV/K. The offset voltage might also drift with time. If this drift is detrimental to the circuit, one should select an op-amp in which it is much smaller. Precision op-amps exist whose voltage offset is measured in microvolts and whose drifts are quite small.

Figure 6.15 shows the pin outs for a couple of common op-amps. Pins 1 and 5 are used to adjust the voltage offset.

6.4.7 Input Offset Current

Not only are the input voltages not quite identical in an op-amp, the currents drawn by the two inputs can also differ. Like the offset voltages this offset current the result of manufacturing. If both inputs are driven by identical sources, the small difference in input current will lead to slightly different voltages at the two inputs.

6.4.8 Op-amp Specifications

Table 6.1 gives the specifications of a couple of commonly used op-amps, the LM741 and the LF411. Both of these are known as *jelly beans*. Wikipedia gives the following definition of "jelly bean" in the semiconductor industry.

> In the semiconductor industry, a jelly bean component is one which is widely available,

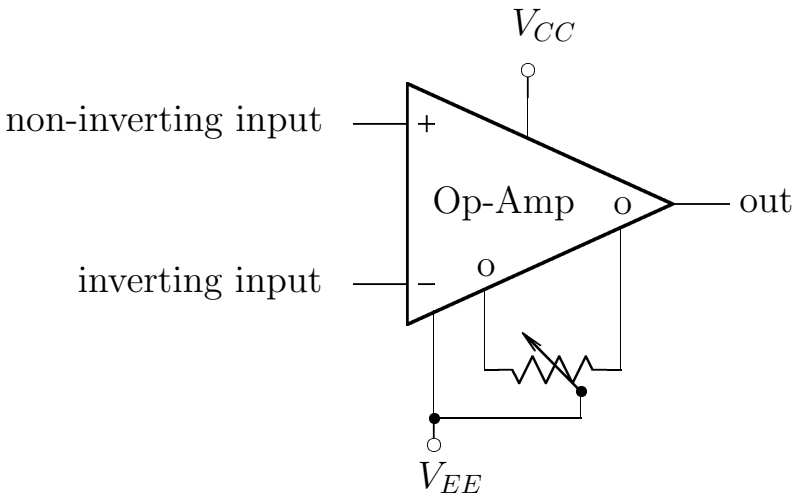

Figure 6.14: A circuit to compensate for the offset voltage in an op-amp. The adjustment is made by grounding the two inputs and adjusting the potentiometer such that the output is very close to the transition from V_{CC} to V_{EE}.

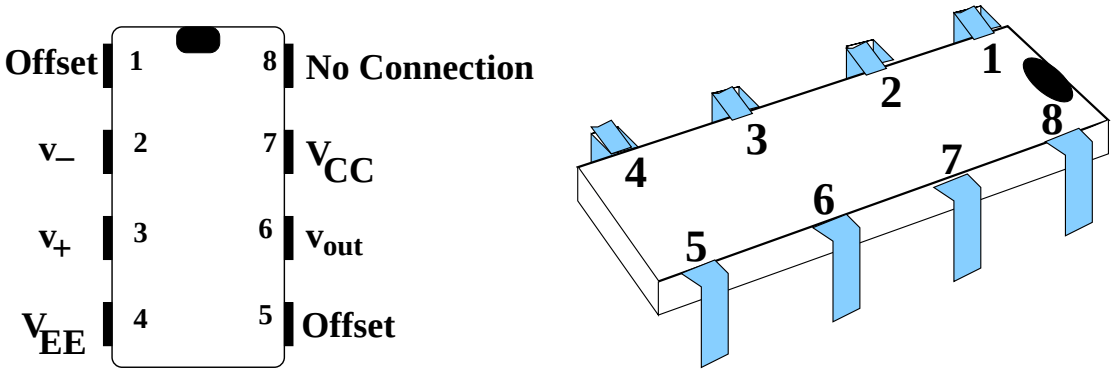

Figure 6.15: The eight-pin flat package for both the LM741A and the LF411 op-amp. The two pins labeled *Offset* are used to correct for non-ideal input properties of the op-amp.

used generically in many applications, and has no very unusual characteristics - as though it might be grabbed out of a jar in handfuls when needed, like jelly beans. For example, the 741 might be considered a jelly bean opamp.

Device	Input Trans.	V_{offset}	I_{Bias} Typ.	Slew Rate Typ.	f_T Typ.	Gain Min.	Gain Typ.
LM741A	Bipolar	$0.8\,mV$	$30\,nA$	$0.7\,V/\mu s$	$6\,MHz$	50000	200000
LF411	JFET	$0.8\,mV$	$0.2\,nA$	$15\,V/\mu s$	$4\,MHz$	25000	200000

Table 6.1: Properties of two jelly bean op-amps, the LM741A and the LF411.

6.5 Active Filters

6.5.1 Integrating and Differentiating Circuits

So far, we have looked at circuits in which the feedback was controlled by purely resistive elements. However, it is also possible to build circuits in which the feedback is controlled by complex impedances. We can start with the the inverting amplifier in section 6.3.2. Here we will replace the resistors R_{in} and R_f with impedances Z_{in} and Z_f as shown in Figure 6.16. We can use the same analysis that we did

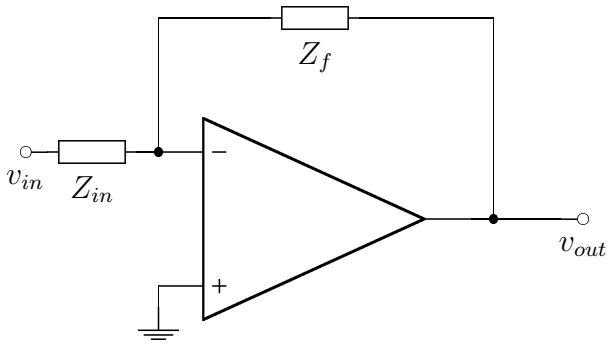

Figure 6.16: An inverting amplifier circuit using complex impedances Z_i and Z_f to control the feedback. The gain of the circuit is given as $G(\omega) = -Z_f(\omega)/Z_{in}(\omega)$. Note that the non-inverting input is grounded.

before. The inverting input is at a virtual ground so that the two inputs are at the same voltage. This means that the current flowing through Z_{in} is

$$i_{in}(\omega) \quad = \quad v_{in}/Z_{in}(\omega)\,.$$

Because no current flows into the input of the op-amp, all this current flows through Z_f, which then gives that the output voltage is

$$v_{out}(\omega) \quad = \quad 0 - \frac{Z_f(\omega)}{Z_{in}(\omega)} v_{in}(\omega)\,.$$

From which the gain is

$$G(\omega) \quad = \quad -\frac{Z_f(\omega)}{Z_{in}(\omega)}\,. \tag{6.9}$$

Let us now consider some particular components: Z_f is a capacitor, C, and Z_{in} is a resistor, R. Putting this into equation 6.9, we find

$$G(\omega) \quad = \quad -\frac{1}{j\omega RC}$$

or

$$G(\omega) \quad = \quad -\frac{\omega_{RC}}{j\omega} \tag{6.10}$$

where as before $\omega_{RC} = \frac{1}{RC}$. The factor of $\frac{1}{j\omega}$ in the above equation is the same factor that we find when we integrate a voltage. If $v_{in} = V_0 e^{j\omega t}$, then we can integrate this as follows:

$$\int_0^t v_{in}(\tau)d\tau \quad = \quad \int_0^t V_0 e^{j\omega\tau} d\tau$$
$$\int_0^t v_{in}(\tau)d\tau \quad = \quad \frac{V_0}{j\omega} e^{j\omega\tau} \,|_0^t$$

which gives

$$\int_0^t v_{in}(\tau)d\tau = \frac{1}{j\omega}v_{in}(t) + \text{Const}\,. \tag{6.11}$$

The constant of integration is just some offset that can be neglected here. For a sinusoidal input voltage, the output is just the integral of the input. This can be generalized to any function by use of Fourier expansions.

We can also analyze this specific RC case without choosing a particular form for the input voltage. The current through the resistor, R_{in}, is $i = v_{in}/R_{in}$. This same current must flow through the capacitor. However, the voltage drop across a capacitor is $v_C = q/C$, or $\frac{dv_C}{dt} = i/C$. This means that we have an expression for the derivative of the output voltage:

$$\begin{aligned} \frac{dv_{out}}{dt} &= -\frac{i}{C} \\ \frac{dv_{out}}{dt} &= -\frac{v_{in}}{RC} \end{aligned}$$

which gives

$$\frac{dv_{out}}{dt} = -\omega_{RC}v_{in}\,. \tag{6.12}$$

Integrating equation 6.12 and using equation 6.11 yields

$$v_{out} = -\omega_{RC}\cdot\int_0^t v_{in}(\tau)d\tau\,. \tag{6.13}$$

Thus, this circuit acts as an integrator. This is similar to the analysis of the low-pass filter that we did in chapter 3. There, the circuit integrated when the frequency was much greater than ω_{RC}. The op-amp circuit has no such limitation. It integrates over all frequencies as long as we can safely approximate the op-amp performance as ideal.

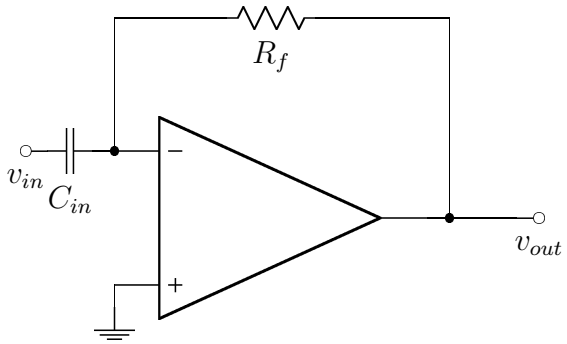

Figure 6.17: An op-amp differentiating circuit. Note that the non-inverting input is grounded.

If we reverse the resistor and capacitor, we get the circuit in Figure 6.17. This changes equation 6.10 to

$$G(\omega) = -\frac{j\omega}{\omega_{RC}}\,. \tag{6.14}$$

Carrying out a similar analysis to what we did above, we note that the factor of $j\omega$ in equation 6.14 is the factor that we get if the circuit is differentiating:

$$v_{out}(t) = \frac{d}{dt}v_{in}(t)\,.$$

Because the non-inverting input is grounded, the inverting input must be a virtual ground. Therefore, the voltage drop across the input capacitor is $v_C = v_{in}$. From this, the current through the input capacitor must be

$$i_{in} = C_{in}\frac{dv_{in}}{dt}\,.$$

Because no current flows into the input of the op-amp, all of it must go through the feedback resistor, R_f, leading to a voltage drop of $i_{in}R_f$ across the feedback resistor. The output voltage is then

$$v_{out} = -\omega_{RC}\cdot\frac{dv_{in}}{dt}\,. \tag{6.15}$$

As with the op-amp integrator, the op-amp differentiator is not limited to a particular frequency regime.

6.5.2 High-pass and Low-pass Filters

The previous two circuits integrate and differentiate signals, apparently without a cut-off frequency. It is also useful to reproduce the behavior of our high-pass and low-pass filters from chapter 2. The circuits shown in Figure 6.18 do this. In both cases, equation 6.9 allows us to determine the gains of these circuits.

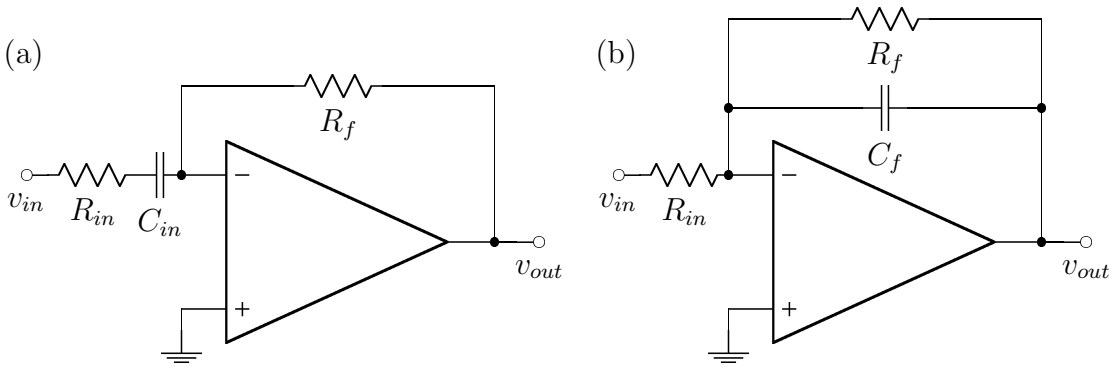

Figure 6.18: A high-pass (a) and low-pass filter (b) built with op-amps. Unlike the passive filters which had a gain of 1 in the pass-band, these circuits have a gain of $-R_f/R_{in}$ in their respective pass-bands. Note that the non-inverting inputs are grounded.

We start by considering the high-pass filter shown Figure 6.18(a). This circuit has input and feedback impedances of

$$\begin{aligned} Z_{in} &= R_{in} + \frac{1}{j\omega C_{in}} \\ Z_f &= R_f\,. \end{aligned}$$

If we define the characteristic frequency to be

$$\omega_{RC} = \frac{1}{R_{in}C_{in}}\,,$$

then it is straightforward to show that the gain of the circuit is

$$G(\omega) = -\frac{R_f}{R_{in}}\cdot\frac{1}{1-j\omega_{RC}/\omega}\,. \tag{6.16}$$

This has the same frequency dependence as the gain of the high-pass filter built entirely out of passive components. However, the gain in the pass-band is $-R_f/R_{in}$ rather than 1. This allows us to build

a filter with gain. As with the passive filter, we can look at the low- and high-frequency limits of this circuit.

$$\begin{aligned} \omega << \omega_{RC} \qquad & G(\omega) \to j \cdot \frac{\omega}{\omega_{RC}} \cdot \frac{R_f}{R_{in}} \\ \omega >> \omega_{RC} \qquad & G(\omega) \to -\frac{R_f}{R_{in}} \end{aligned}$$

Similarly, we can look at the gain of the low-pass filter shown in Figure 6.18(b). This circuit has the input and feedback impedances of

$$\begin{aligned} Z_{in} &= R_{in} \\ Z_f &= \frac{R_f}{1 + j\omega R_f C_f} \, . \end{aligned}$$

If we define the characteristic frequency to be

$$\omega_{RC} = \frac{1}{R_f C_f} \, ,$$

then the gain of the circuit is

$$G(\omega) = -\frac{R_f}{R_{in}} \cdot \frac{1}{1 + j\omega/\omega_{RC}} \, . \tag{6.17}$$

This is the same as the low-pass filter with passive components, except that the gain in the pass-band is $-R_f/R_{in}$.

As a final note, one should always keep in mind that the gain of any filter cannot exceed the open-loop gain of the op-amp. In fact, the gain will always be somewhat smaller than the open-loop gain. This means that at some large-enough frequency, many filters will reach the falling open-loop gain of the op-amp. At such a limit, the filter stops doing what it is supposed to do.

6.6 Functional Feedback

Let us now consider the circuit shown in Figure 6.19. The feedback occurs through some device such that the voltage drop across the device is given by the function $f(i)$, where i is the current flowing through the device. In such a circuit, the current is

$$i = v_{in}/R_{in}$$

so the output voltage will be

$$v_{out}(t) = -f\left(\frac{v_{in}}{R_{in}}\right) \, .$$

In other words, the output is some defined function of the input.

Example: Consider the circuit shown in Figure 6.20 where the feedback element is a diode. We know from Chapter 4 that the current through a diode is

$$I(V) = I_S \left(e^{v/V_T} - 1\right) \, ,$$

so the voltage drop across the diode is

$$v_{diode} = V_T \cdot \ln\left(1 + i_{diode}/I_S\right) \, .$$

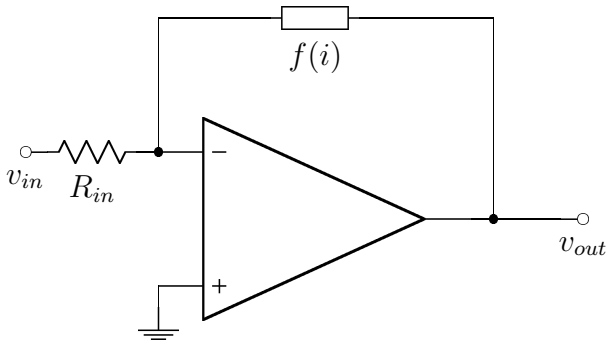

Figure 6.19: An op-amp circuit with functional feedback.

Recalling that I_S is typically pA, for currents through the diode that are mA, we can drop the 1 inside the logarithm, (which is added to the ratio of currents). We also know that the inverting input of the op-amp is a virtual ground, thus the output voltage is zero minus the voltage across the diode. We also know that the current flowing through the diode must be the current flowing through R_{in}, or v_{in}/R_{in}. Putting things together, we find that the output voltage of the circuit is

$$v_{out} \quad = \quad -V_T \cdot \ln\left[v_{in}/(R_{in} \cdot I_S)\right] \; .$$

Thus, the output is proportional to the logarithm of the input.

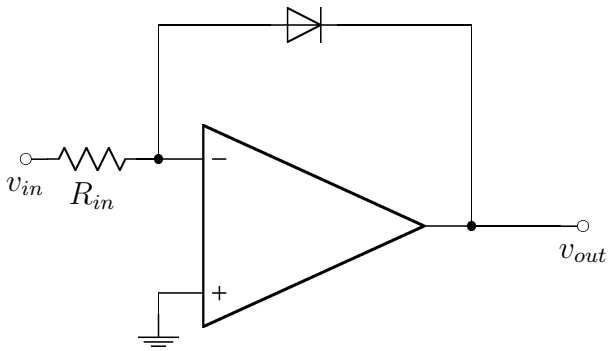

Figure 6.20: An op-amp circuit with diode feedback.

6.7 NICS and Gyrators

6.7.1 Negative-Impedance Converters

Consider the circuit shown in Figure 6.21, where matched resistors provide feedback to both inputs. The non-inverting input is also connected to ground through some impedance Z. We can use our two rules of op-amp performance to understand the behavior of this circuit.

The non-inverting input is at a voltage $v_+ = v_{in}$, which means that the inverting input must also be at the same voltage, $v_- = v_{in}$. To determine the output voltage, we note that from v_{out} to ground, there is a voltage divider consisting of R and Z, where the divided voltage is v_{in}. This gives us

$$v_{in} \quad = \quad \left(\frac{Z}{R+Z}\right) v_{out} \; .$$

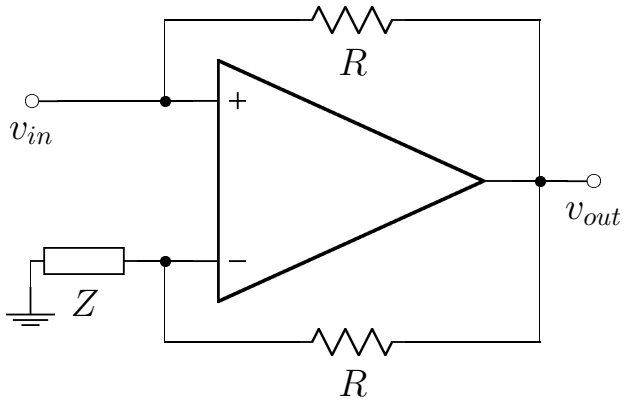

Figure 6.21: A negative-impedance converter. The input impedance, $Z_{in} = -Z$.

This can be rearranged to yield an expression for v_{out},

$$v_{out} = \left(1 + \frac{R}{Z}\right) v_{in} \,,$$

so the gain of the circuit is

$$G = 1 + \frac{R}{Z} \,. \tag{6.18}$$

Let us now consider the input impedance, Z_{in}, of this circuit. This can be found from

$$Z_{in} = \frac{v_{in}}{i_{in}} \,.$$

The input current, i_{in}, depends on the voltage drop across the feedback resistor to the non-inverting input.

$$\begin{aligned} i_{in} &= \frac{1}{R} \cdot \left[v_{in} - \left(1 + \frac{R}{Z}\right) v_{in} \right] \\ i_{in} &= -\frac{v_{in}}{Z} \end{aligned}$$

Putting all of this together, we find that the input impedance is

$$Z_{in} = -Z \,. \tag{6.19}$$

The input impedance is the **negative** of the impedance connecting the inverting input to ground. These circuits are known as negative-impedance converters, or NICS.

Example: Let us assume that the impedance Z is a resistor, R_0. Then the input resistance of the circuit is $-R_0$. If we apply some voltage V_0 to the input, the current is $I_{in} = -V_0/R_0$—that is the current flows from the output to the input. The bigger the voltage we apply, the more current flows out of the input.

Example: Assume that the impedance Z is a capacitor, C. In this case, $Z = \frac{1}{j\omega C}$, and the input impedance is $Z_{in} = -\frac{1}{j\omega C}$. Noting that $\frac{-1}{j} = j$, we can rewrite this as

$$Z_{in} = \frac{j}{\omega C} \,.$$

The key feature of this is that the current lags the applied voltage, rather than leading it as we would normally see for a capacitor. In fact, the current leading the voltage is an effect that we get for inductors. Thus, this circuit, using a capacitor, behaves in some sense like an inductor.

6.7.2 Gyrators

Expanding on the last example of the NIC using a capacitor, if we could build a circuit that not only changed the leading current to a lagging one, but could also invert the frequency dependence of the capacitor, it would be possible to mimic an inductor. A circuit that does this is known as a *gyrator*; an example is shown in Figure 6.22. The input impedance of the gyrator can be determined from the ratio of v_{in}/i_{in}, as shown in the circuit.

To analyze this circuit, we will start with our simple op-amp rules. The voltage at the non-inverting input is v_{in}, which means that the voltage at the inverting input must also be v_{in}. This tells us that the output voltage is also v_{in}. If we have v_{in} at both inputs, we can determine the currents flowing through the circuit by looking at the right-hand plot in Figure 6.22. The current i_{in} flows through resistor R_1, the current i_c flows through the capacitor and the sum of these two flow through R_2.

If we define the voltage at the middle of the divider to be v_a, then we can write the following three equations:

$$\begin{aligned} i_{in} &= (v_{in} - v_a)/R_1 \\ i_c &= (v_{in} - v_a)/Z_C \end{aligned} \tag{6.20}$$

$$i_{in} + i_c = v_a/R_2$$

Adding the first two and substituting into the third, we obtain an expression that can be solved for v_a.

$$v_a = v_{in}\left(\frac{R_2/R_1 + R_2/Z_C}{1 + R_2/R_1 + R_2/Z_C}\right)$$

We can substitute this into equation 6.20 to obtain that the input impedance of the gyrator circuit is given as:

$$Z_{in} = j\omega C R_1 R_2 + R_1 + R_2\,. \tag{6.21}$$

This looks exactly like an inductor with $L = R_1R_2C$ and internal resistance $R_L = R_1 + R_2$. The gyrator circuit behaves like an inductor connected to ground.

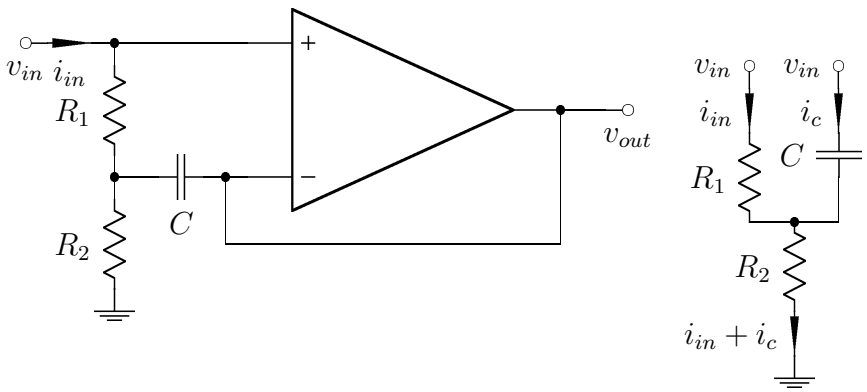

Figure 6.22: A simple gyrator circuit with input $Z_{in} = j\omega R_1R_2C + R_1 + R_2$. The left-hand figure is the actual circuit, while the right-hand figure shows how currents flow through the circuit.

A gyrator is useful because it is possible to have both resistors and capacitors on a semiconductor chip. However, it is not possible to make a true inductor on a chip. By using a gyrator, it is possible to create a circuit element that behaves like an inductor. Thus, it is possible to build oscillators directly on a chip.

6.8 Comparators

6.8.1 Simple Comparators

In many situations, it is useful to be able to compare an input voltage to some reference voltage. When the input is larger than the reference we get one output, while when the input is smaller than the reference, we get a different output. Such a *comparator* is a basic element of many electronic switches, or devices that convert analog signals to digital. A digital voltmeter is based on such a circuit.

Figure 6.23 shows a simple circuit that does such a comparison. The reference voltage is connected to the inverting input of the op-amp, while the input voltage goes into the non-inverting input. Looking at the circuit, the output voltage is nominally

$$v_{out} = A_0 \left(V_{ref} - v_{in} \right) ,$$

however, the output can be no smaller than V_{EE} and no larger than V_{CC}. This means that in reality v_{out} will be either V_{CC} or V_{EE}. If v_{in} is larger than V_{ref}, then $v_{out} = V_{EE}$. When v_{in} is smaller than V_{ref}, then $v_{out} = V_{CC}$.

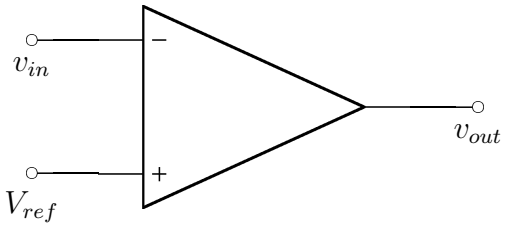

Figure 6.23: A simple comparator circuit. If the input voltage is larger than V_{ref}, the output will be V_{EE}. If the input is smaller than V_{ref}, the output will be V_{CC}.

A slight modification to this circuit is shown in Figure 6.24. Here a potentiometer provides an adjustable reference voltage between 0 and V_{ref}. Such a circuit might form the basis for a thermostat, where we want to be able to adjust the temperature at which the heat turns on an off in a house.

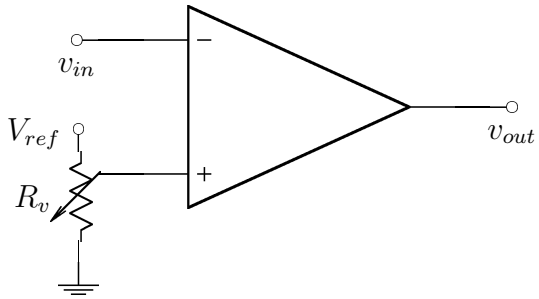

Figure 6.24: A comparator circuit where reference can be adjusted to be between V_{ref} and ground using the potentiometer, R_v.

6.8.2 The Schmitt Trigger

Unfortunately, the simple comparators described in the last section have a potentially undesirable feature. If the input is both very close to the reference and noisy, then the comparator may be continually changing states. This situation is shown in Figure 6.25. Such a problem can be solved by the circuit shown in Figure 6.26, the Schmitt trigger. This circuit is very similar to the comparator, except that it employs feedback into the input with the reference voltage.

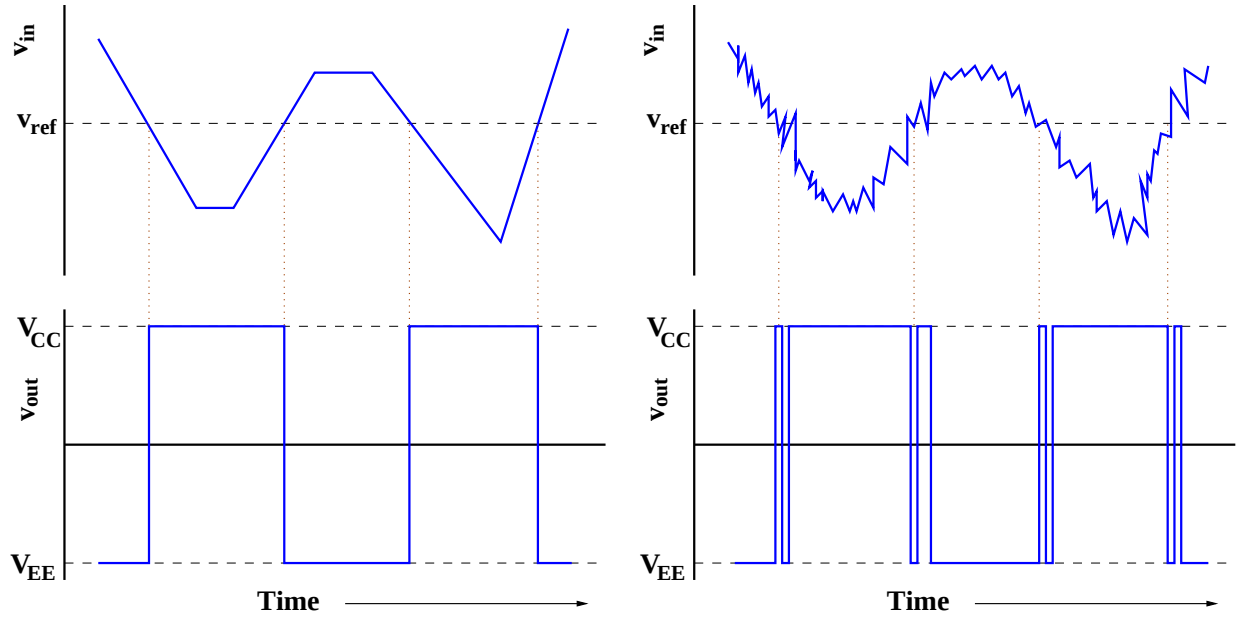

Figure 6.25: The response of a comparator circuit for an idealized input voltage (left) and a noisy input voltage (right). In the case with noise, the output oscillates rapidly between the two states as the input crosses the reference voltage.

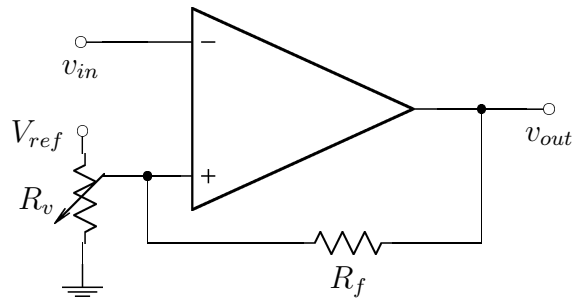

Figure 6.26: A Schmitt-trigger circuit whose reference can be adjusted to a value between V_{ref} and ground using the potentiometer, R_v.

To understand the behavior of this circuit, let us look at the inverting input which is connected to the reference voltage. We anticipate that the output voltage will still be either V_{CC} or V_{EE}. We will also explicitly replace the potentiometer with two resistors, R_1 and R_2, which will then function as a voltage divider for the reference voltage. and R_2. Figure 6.27 illustrates the two possible states of the system. The voltage to which the input is compared is v_a when the output is V_{CC}, and v_b when the output is V_{EE}.

The reference voltage, V_r, without feedback would be

$$V_r = \frac{R_2}{R_1 + R_2} \cdot V_{ref}$$

We can solve for v_a and v_b by considering the currents flowing through the circuit and then see what happens to the reference voltage, v_r in the two cases. If we define the quantity α to be:

$$\alpha = \frac{R_1 R_2}{R_f(R_1 + R_2)},$$

then we can solve for our two voltages:

$$v_a = V_r \cdot \left(\frac{1}{1+\alpha}\right) + V_{CC} \cdot \left(\frac{\alpha}{1+\alpha}\right) \tag{6.22}$$

$$v_b = V_r \cdot \left(\frac{1}{1+\alpha}\right) + V_{EE} \cdot \left(\frac{\alpha}{1+\alpha}\right) . \tag{6.23}$$

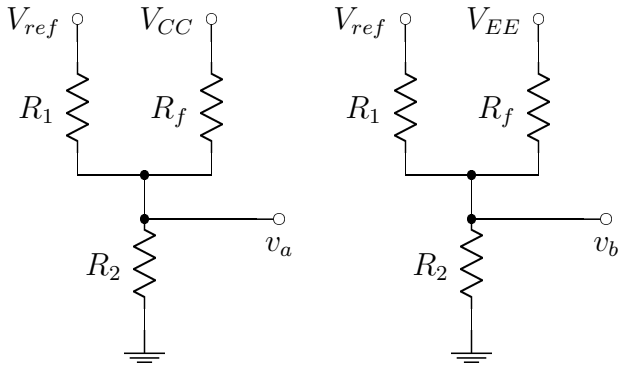

Figure 6.27: The two stable states of the Schmitt trigger. The left-hand case has the output voltage at V_{CC} and the voltage at the non-inverting input is at v_a. The right-hand case has the output voltage at V_{EE} and the non-inverting input is at v_b.

Let us now consider the limit when α is less than 1. We will also take that $V_{EE} = -V_{CC}$. In this case, we find the two voltages to be:

$$v_a = V_r + \alpha V_{CC}$$

and

$$v_b = V_r - \alpha V_{CC} .$$

Thus v_a is slightly larger than V_r and v_b is slightly smaller than V_r. To understand what the Schmitt trigger is doing, we consider the voltage response as shown in the left-hand plots in Figure 6.28. The right-hand plots show a noisy version of the same input. The output of the Schmitt trigger, unlike that of the simple comparator does not rapidly between the two output states when the input is near the reference voltage.

The system starts out with v_{in} larger than v_a and the output voltage at V_{EE}. The input voltage then drops. When the input falls below v_b, the output switches to V_{CC}. The voltage continues changing and eventually starts to rise. When it rises above v_a, the output switches to V_{EE}. It remains there until the input falls below v_b again. The key feature is that system is much less sensitive to noise when the input is near the reference voltage.

Example: Let us consider a Schmitt trigger in which we have a potentiometer of total resistance R_f, the same as the feedback resistor. This means that $R_1 + R_2 = R_f$, which allows us to rewrite α in terms of R_f and R_1.

$$\alpha = \frac{R_1}{R_f}\left(1 - \frac{R_1}{R_f}\right)$$

The ratio R_1/R_f can vary from 0 to 1. The value of α is 0 at both endpoints and a maximum of $\frac{1}{4}$ when $R_1 = \frac{1}{2}R_f$. Finally, we will assume that $V_{ref} = \frac{1}{2}V_{CC}$, so the voltage to which we are comparing

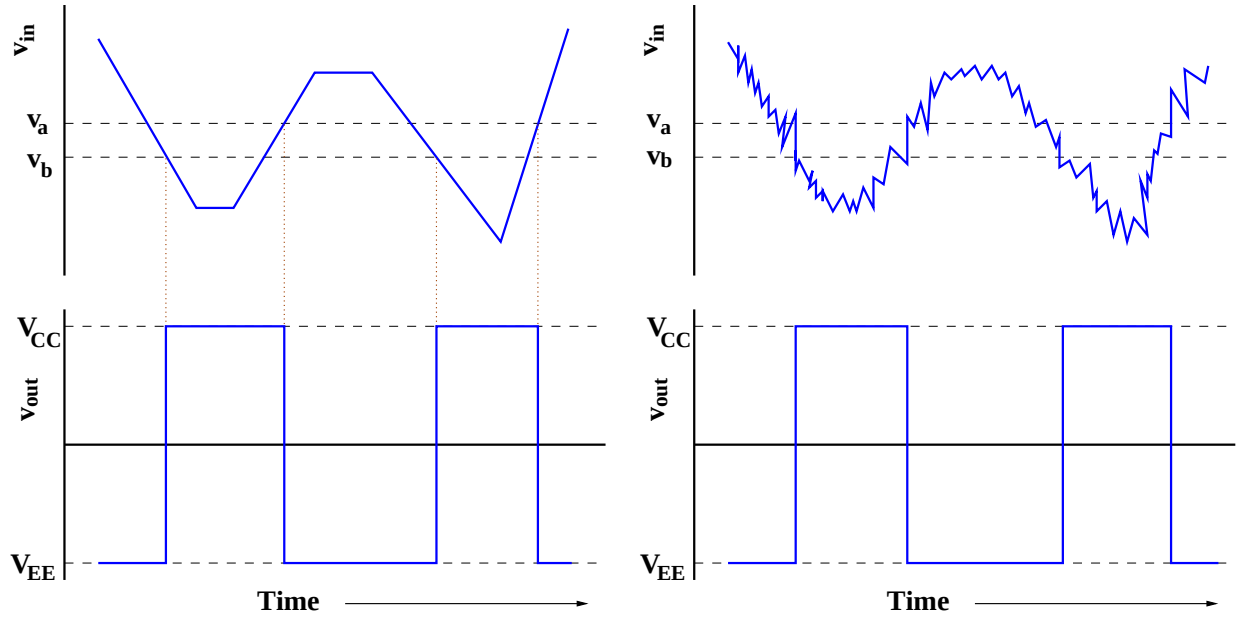

Figure 6.28: The response of a Schmitt trigger to an input voltage v_{in}. The upper plots show the input voltages while the lower plots show the output voltages. The system changes state at different voltages depending on whether the output is V_{CC} or V_{EE}. This eliminates the rapid output oscillations seen in the simple comparator when there is noise on the input signal.

our signal varies from 0 to $\frac{1}{2}V_{CC}$. Figure 6.29 shows the reference voltages, V_a and V_b, as a function of the ratio R_1/R_f. The dashed line in the figure is what the reference would be without feedback in the circuit. The main feature to note is that the separation between the two lines is a function of the exact values of the resistors in the circuit. Also, as we go to the two extremes, 0 and 1, the separation goes to zero. This removes the ability of the circuit to filter out noise.

The Schmitt trigger is an example of a circuit that exhibits *hysteresis*. If the input voltage is between v_b and v_a, then the only way that we can predict v_{out} is to know how the input voltage got to its current value. If it got there by falling from above v_a, the the output will be V_{EE}. However, it it got there by increasing from below v_b, then the output will be V_{CC}. Without some knowledge of the history system, we are unable to predict its current behavior. This "memory" of where the system has been is known as hysteresis.

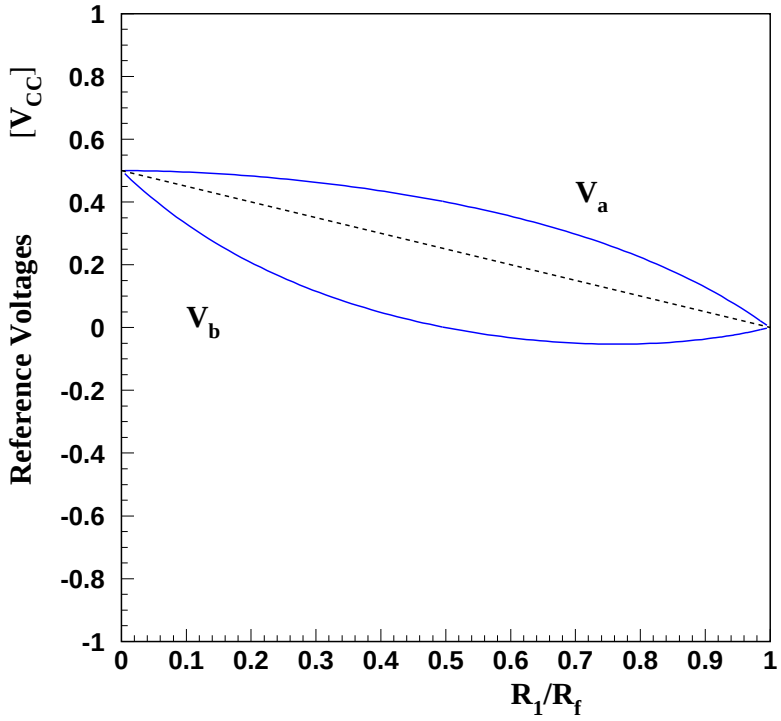

Figure 6.29: The reference voltages for a Schmitt trigger as a function of R_1/R_f. The dashed line is what the reference voltage would be without feedback.

Problems

1. A negative feedback circuit has an open-loop gain of $1000 \pm 15\%$ and a feedback fraction of $-\beta$. How small can β be such that the gain of the circuit is stable to $\pm 1.5\%$?

2. A negative feedback circuit is designed assuming a constant open-loop gain of 10^6 with a desired gain of 100. Unfortunately, the open-loop gain is not constant, but rather falls off as $1/\omega$, and reaches 100 at $\omega = 100000\, s^{-1}$. What is the gain of the circuit for frequencies above $100000\, s^{-1}$? By how much does it differ from the open-loop gain?

3. An op-amp circuit is built with $V_{CC} = +10\, V$ and $V_{EE} = 0\, V$. The inverting input is grounded and a signal $v(t) = 1.0\, V \cos(\omega t)$ is connected to the non-inverting input. Assuming that $\omega = 100\, s^{-1}$, plot the output of the op-amp as a function of time.

4. An op-amp circuit has $V_{CC} = +10\, V$ and $V_{EE} = 0\, V$. The non-inverting input is grounded and a signal $v(t) = 1.0\, V \cos(\omega t)$ is connected to the inverting input of the op-amp. Assuming that $\omega = 100\, s^{-1}$, plot the output of the op-amp as a function of time.

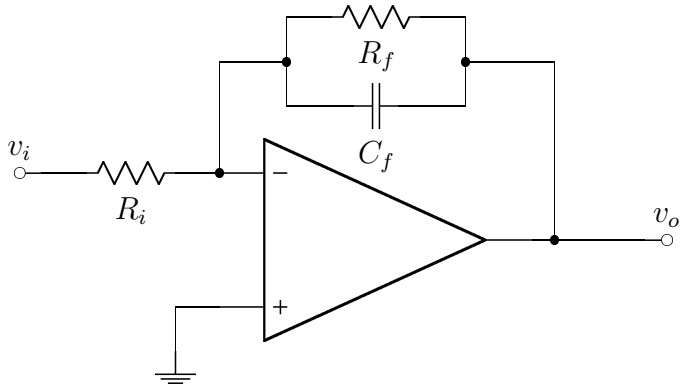

Figure 6.30: The circuit for problem 5. Note that the non-inverting input is grounded.

5. This question deals with the op-amp circuit shown in Figure 6.30. This has an input resistor, R_i, and a feedback impedance, Z_F, built from a resistor, R_f, in parallel with a capacitor, C_f. The circuit has an input voltage of $v_i(t)$, and an output voltage of $v_o(t)$. You may assume that the op-amp is correctly biased with the appropriate DC voltage such that it is *on*. (**a**) The resistor, R_f, and capacitor, C_f, in the feedback loop are in parallel with each other. Which one will dominate in the limit of low frequencies and which will dominate in the limit of high frequencies? (**b**) Write the impedance of the parallel pair, Z_f, as R_f times a dimensionless quantity. (**c**) What is the gain, G, of the circuit in the low-frequency regime? (**d**) What is the gain, G, of the circuit in the high-frequency regime? (**e**) Using your results for **c** and **d**, plot the low-frequency and high-frequency **limits** of $\mid G(f) \mid$ as a function of f on a log-log plot. Sketch a transition curve that could connect these two limits, and label a frequency (as expressed in terms of known Rs and Cs) that falls in this transition region. (**f**) For an arbitrary time-dependent input voltage, $v_i(t)$, what will the input current, $i_i(t)$, be? (**g**) What will be the current in the feedback loop (through the parallel R-C pair)?

6. The resistor R_f from problem 5 is now removed so that we have the circuit shown in Figure 6.31. (**a**) Use the results of parts **f** and **g** in problem 5 to show that if we remove R_f, the output voltage, $v_o(t)$, will be proportional to the integral of the input voltage, $v_i(t)$. (**b**) Restore the feed back resistor, R_f, so that you have the original circuit as shown in Figure 6.30 again. In what range of frequencies, f, will the output voltage, v_o be the integral of the input voltage, v_i? (**HINT:** There is no exact answer, but you should be able to quantitatively indicate a transition point).

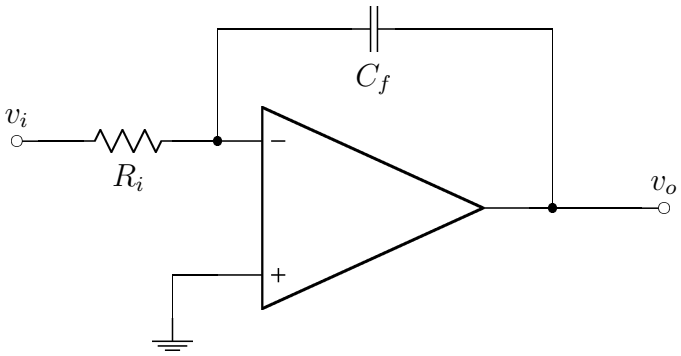

Figure 6.31: The circuit for problem 6. Note that the non-inverting input is grounded.

7. You have constructed the idealized op-amp circuit as shown in Figure 6.32. Answer the following in terms of R_i, L_f, C_f and v_i.

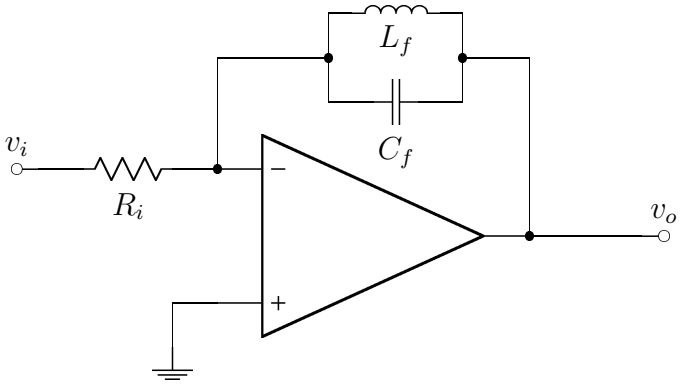

Figure 6.32: The circuit for problem 7. Note that the non-inverting input is grounded.

(**a**) Which circuit elements are important in determining the gain of this circuit in the limit of low frequencies ($\omega \to 0$, but not quite equal to zero)? (**b**) Which circuit elements are important in determining the gain of this circuit in the limit of high frequencies, $\omega \to \infty$? (**c**) What is the complex gain function in the limit of low frequencies, $\omega \to 0$? (Give the limiting dependence on ω, not just the behavior at $\omega = 0$.) (**d**) What do you expect to happen when $\omega = \frac{1}{\sqrt{L_f C_f}}$? What physical limitations of real components would limit this behavior? (**e**) Sketch a Bode plot for the amplitude of the gain function. Be sure to label your axes and indicate the relevant frequencies, ω.

8. In the op-amp circuit in Figure 6.33, resistors $R_1 = R_2 = R_3 \equiv R$, while R_4 has a value that is smaller than R. The input and output voltages are v_i and v_o respectively. It is possible to measure the voltage at the points **a** and **b**. Answer the following questions in terms of R, R_4, v_i and v_o. Be sure to state any assumptions that you make, including op-amp rules and approximations. You may assume that the op-amp is in its normal operating conditions. (**a**) What are the currents through resistors R_1 and R_2 and the voltages at points **a** and **b**? Be sure to indicate the directions of these currents on the circuit. (**b**) What is the current through R_4? (**c**) Based on what you found in parts **a** and **b**, express the gain of the amplifier in terms of R and R_4. (**d**) Assume that $R : R_4 = 100 : 1$ and consider the currents through the feedback resistors R_2, R_3 and R_4 (i_2, i_3 and i_4). These currents are similar to the three currents for a transistor, I_B, I_C, I_E. What is the correspondence of currents, i_2, i_3, i_4 and I_B, I_C, I_E? (**f**) What determines the analogue of β?

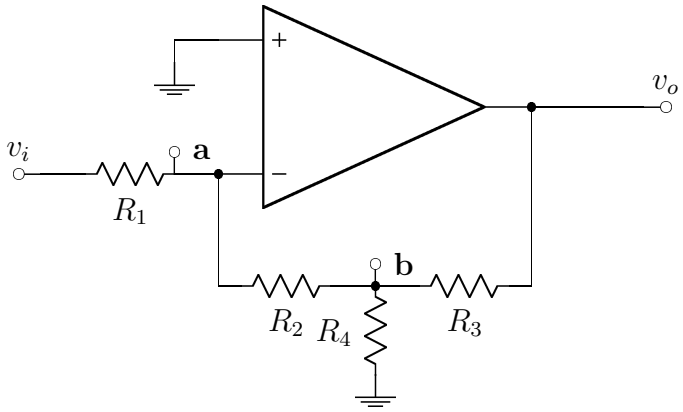

Figure 6.33: The circuit for problem 8.

9. Consider the op-amp circuit as shown in Figure 6.34. It is sometimes referred to as an *all-pass filter* and often used as a *phase shifter*. You will see why when we discuss the gain, $G(\omega)$. (**a**)

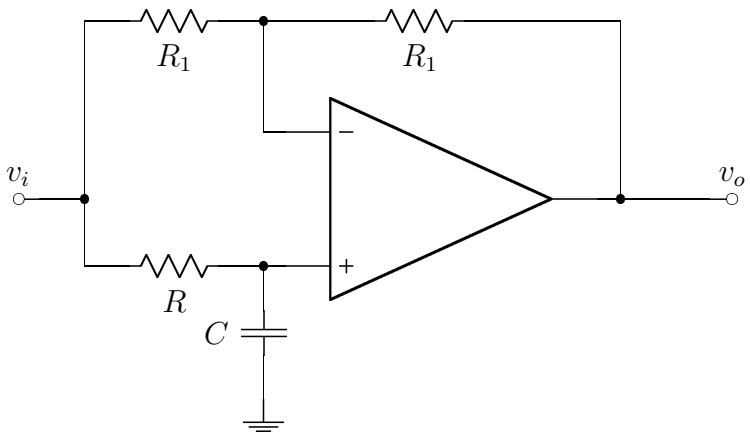

Figure 6.34: The circuit for problem 9.

Approximately how much current flows into the + and − terminals of the op-amp? (**b**) What is the approximate relation between V_+ and V_-? (**c**) Does the sum of the currents *into* the three terminals (+, −, and output) equal zero? Explain why or why not. (**d**) You are told that the gain is

$$G(\omega) \quad = \quad \frac{1 - j\omega RC}{1 + j\omega RC}\,.$$

Graph both the magnitude of the gain (in dB) vs. log frequency, and the phase of the gain vs. log frequency. Compute the numerical values for the magnitude and phase of the gain in table 6.2 below. In doing so, choose a third frequency value in between the two limits.

Frequency	$\lvert G(\omega)\rvert$	ϕ_G
$\omega = 0$		
$\omega \to \infty$		

Table 6.2: Complete this table as part of problem 9.

10. Instead of deriving the full expression for the gain from part **d** in problem 9, let's consider the simpler case of just the high-frequency limit of our circuit. In this case, the capacitor may be treated as a short, giving us the circuit shown in Figure 6.35. (**a**)What is the current through the

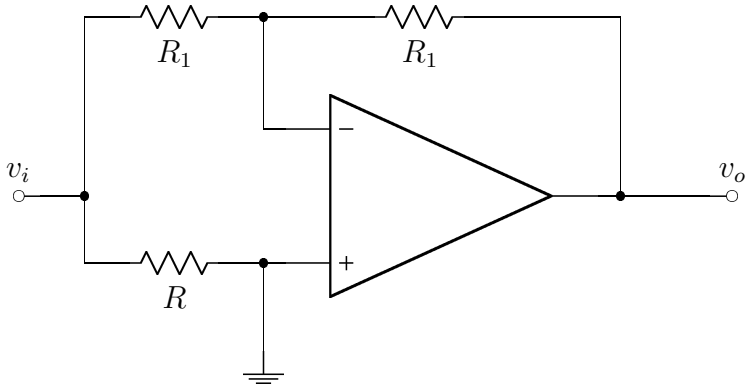

Figure 6.35: The circuit for problem 10.

lower resistor, R? What is the current through the resistor R_1 closest to the input? Both answers should be expressed in terms of v_i, R, R_1. (**b**) Calculate the gain of the new high-frequency equivalent circuit (without the capacitor). HINT: Your answer should agree with the high-frequency limit of the full-gain expression from part **d**, in both magnitude and phase!

11. Determine the input impedance of the circuit shown in Figure 6.36. Use this to show that the circuit is a gyrator.

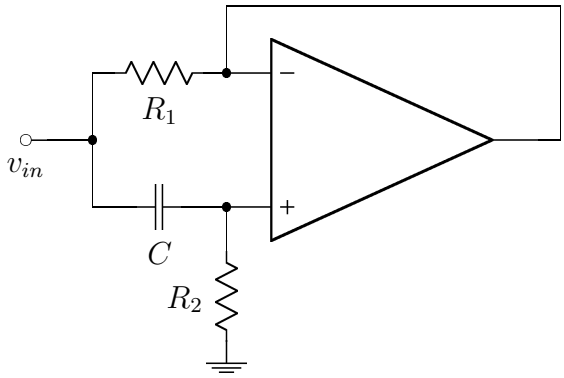

Figure 6.36: The circuit for problem 11.

12. In this problem, we will be looking at the op-amp circuit shown in the Figure 6.37. The input voltage is connected to the non-inverting input while the inverting input is connected to ground through an impedance, Z. Feedback to both the inverting and non-inverting input is through a pair of resistors, R_1 and R_2. Answer the following questions in terms of v_{in}, Z and R_1 and R_2. In all cases, you can assume that the op-amp input does not cause the output to be driven to either V_{CC} or V_{EE}. You may also assume that the two rules of op-amp operation apply. (**a**) What are the voltages at the inverting and at the non-inverting input of the op-amp? (**b**) How much current flows into the inverting and into the non-inverting input of the op-amp? (**c**) What is the complex gain, G, of the circuit? (**d**) If $R_1 = R_2 = R$, then it can be shown that the gain of the circuit is $G = 1 + R/Z$. What is the input impedance of the circuit when both resistors are equal to R? (**Hint:** $Z_{in} = v_{in}/i_{in}$).

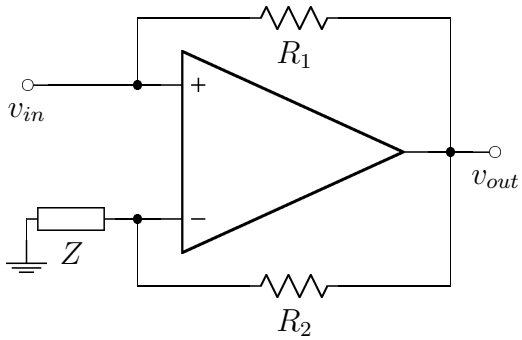

Figure 6.37: The circuit for problem 12.

13. You have constructed the idealized op-amp circuit as shown in the circuit in Figure 6.38. Answer the following in terms of R, L, C and v_i and assuming idealized op-amp behavior. (**a**) What is the magnitude of the gain of the circuit in the limit of $\omega \to \infty$? (**b**) What is the magnitude of the gain of the circuit in the limit of $\omega \to 0$? (**c**)What is the complex gain of the circuit as a function of the angular frequency ω? Express your answer in the form of $G = A + jB$ where A and B are real numbers. (**d**) As ω goes from close to zero to a very large value, by how much does the relative phase between v_{in} and v_{out} change?

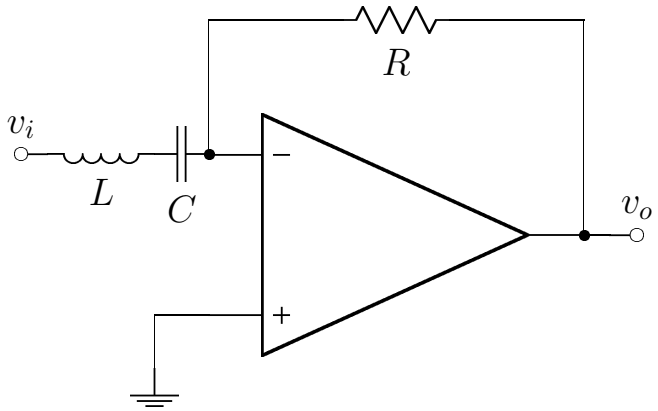

Figure 6.38: The circuit for problem number 13.

Chapter 7

Digital Electronics

7.1 Introduction to Digital Electronics

Our previous discussions have considered *analog* electronics, in which a system is allowed to have any voltage, usually within some given limits. Such a system can therefore be in any of an infinite number of possible states. Another type of electronics deals with devices having only two states, for example, the comparator of the previous chapter, whose output can take on the value of either supply voltage, but nothing in between. This two-state model is the basis of *digital* electronics.

In digital electronics, we can give names to the two states. Possible pairs of names include *off* and *on*, *open* and *closed*, or *zero* and *one*. This latter scheme is directly related to counting in base 2, or binary. Each electronic two-state system is associated with a *bit*, or *binary digit*, and putting together the bits allows one to count as shown in Table 7.1.

base-10	base-2	base-10	base-2
0	0000	8	1000
1	0001	9	1001
2	0010	10	1010
3	0011	11	1011
4	0100	12	1100
5	0101	13	1101
6	0110	14	1110
7	0111	15	1111

Table 7.1: Counting to 16 in base 2.

Because the two states are reasonably well separated in voltage, digital systems are much less sensitive to noise than analog. That is, noise is unlikely to cause one state to be confused with the other.

7.2 A Simple Two-state System

While the comparator, as mentioned, is a fairly simple two-state system, we can examine an even simpler one in the form of the transistor circuit shown in Figure 7.2. If the input voltage here is less than about $0.65\,V$ (V_L), then the transistor is turned off, and no current flows. In this case, the output voltage will be $v_{out} = V_{CC}$. If the input voltage is larger than $0.65\,V$ (V_H), the transistor will turn on and allow current to flow. Because the emitter is connected directly to ground, the current depends on the internal resistance, r_E, of the transistor, but in general it will be large. This drives the transistor into saturation, with the output voltage falling to the saturation voltage, V_{CEsat} above zero. Figure 7.2

shows the transfer curve for the transistor. This system can switch between two output voltage states, given approximately as V_{CC} and 0.

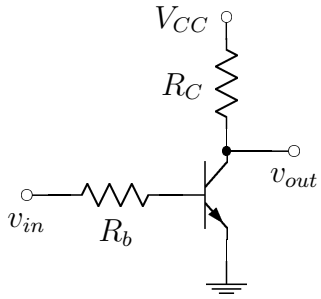

Figure 7.1: A two-state transistor circuit.

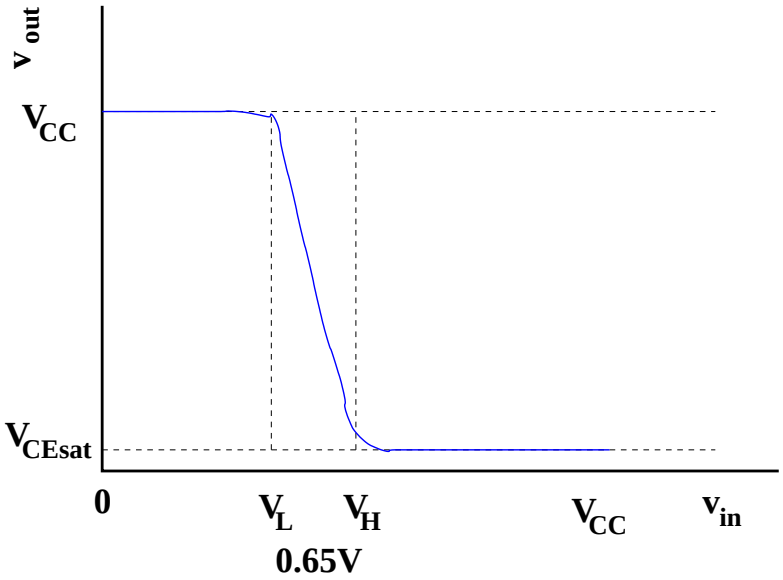

Figure 7.2: The transfer curve for the two-state transistor circuit shown in Figure 7.1.

We might note that, for such a device, there is a range of input voltages for which the output will not be well-defined. Specifically, if v_{in} is between V_L and V_H, we cannot predict the output of the system. This means that when we describe a digital device, we must define a range of valid input voltages which will give rise to particular outputs. When an input is outside these limits, the system is in an ill-defined state. Normally, the output ranges will be more restrictive than the inputs. This allows for the possibility of the output suffering electronic noise or distortion before going to the input of the next stage. In our example, we could define the valid input voltages as shown in Table 7.2, and we would define valid outputs as being very close to either V_{CC} or V_{CEsat}.

state	v_{in}
0	$0 < v_{in} < V_L$
1	$V_H < v_{in} < V_{CC}$

Table 7.2: State definitions for the transistor logic in Figure 7.1.

Another consideration is the speed at which the system can change from one state to the other. Like

op-amps, which are limited by the slew rate, digital circuits have a finite upper limit on their switching rate. This limits how fast such circuits can run.

Finally, we might want to examine the currents in our transistor circuit. In one case, the current is large, while in the other case, it is zero. This means that the power dissipation in the two states is different, and so noise may be generated when the circuit changes state, because a large current is changing rapidly.

7.3 Common Digital Families

Numerous different physical compositions, or *families*, of digital devices exist. As we saw in the last section, each of these will have well-defined ranges of valid input and output voltage. Three fairly common families are *transistor-transistor logic*, TTL, *complementary metal-oxide semiconductor*, CMOS, and *emitter-coupled logic*, ECL. Their input and output levels are given in Table 7.3, and shown graphically in Figure 7.3.

Logic	State	Input Range		Output Range	
TTL	0	0.00	0.80	0.00	0.50
	1	2.00	5.00	2.70	5.00
5V CMOS	0	0.00	1.50	0.00	0.05
	1	3.50	5.00	4.95	5.00
ECL	0	−0.81	−1.13	−0.81	−0.98
	1	−1.48	−1.95	−1.63	−1.95

Table 7.3: Input and output voltage limits for some common digital logic families.

TTL is based on voltage levels of 0 and $5\,V$, while ECL uses $-0.8\,V$ and $-2\,V$. CMOS has scalable levels, with the external voltage being referred to as V_{DD}. Table 7.3 lists CMOS levels for $V_{DD} = 5\,V$, but note that V_{DD} can be either larger or smaller than this. It is this capability which makes CMOS very important in modern digital electronics, where V_{DD} can be pushed down to around $1\,V$. In addition, since CMOS is based on MOSFETS, rather than bipolar transistors, it draws significantly less current.

While one might think that the high and low input voltage limits could abut, this would in fact significantly reduce the noise rejection inherent in digital circuits. If for some reason a signal was very close to the boundary level, the output state could bounce back and forth. However, a large gap between the two input voltages, only a very large amount of noise would cause this behavior. Instead, the device's behavior would be undefined. While we might not be able to receive the signal then, at least we would not be misinterpreting a noisy signal.

Given the different behaviors of the digital families mentioned above, it is usually not possible for the output from one family to serve as the input for another. In general, we need a level converter to accomplish this. One exception is that the output of a $5\,V$ CMOS system can be used as input to TTL (though the converse is not true).

7.4 Digital Logic

With digital signals, the processing is not done with filters and amplifiers, but rather with logical operations. These are performed by circuit elements referred to as gates. The simplest gate is a NOT, which when applied to digital signals produces an output that is in the opposite state from the input. In fact, the transistor circuit that we saw in Figure 7.1 is an example of such a gate. We represent the NOT operation by drawing a bar over the state, so that $\overline{1} = 0$ and $\overline{0} = 1$.

While the NOT operates on a single input, most operations handle two or more inputs. The logical OR has two inputs. If one or more inputs are high, then the output is high, otherwise it is low. The symbol for OR is a plus sign. The next common operation is a logical AND. This also has two or more

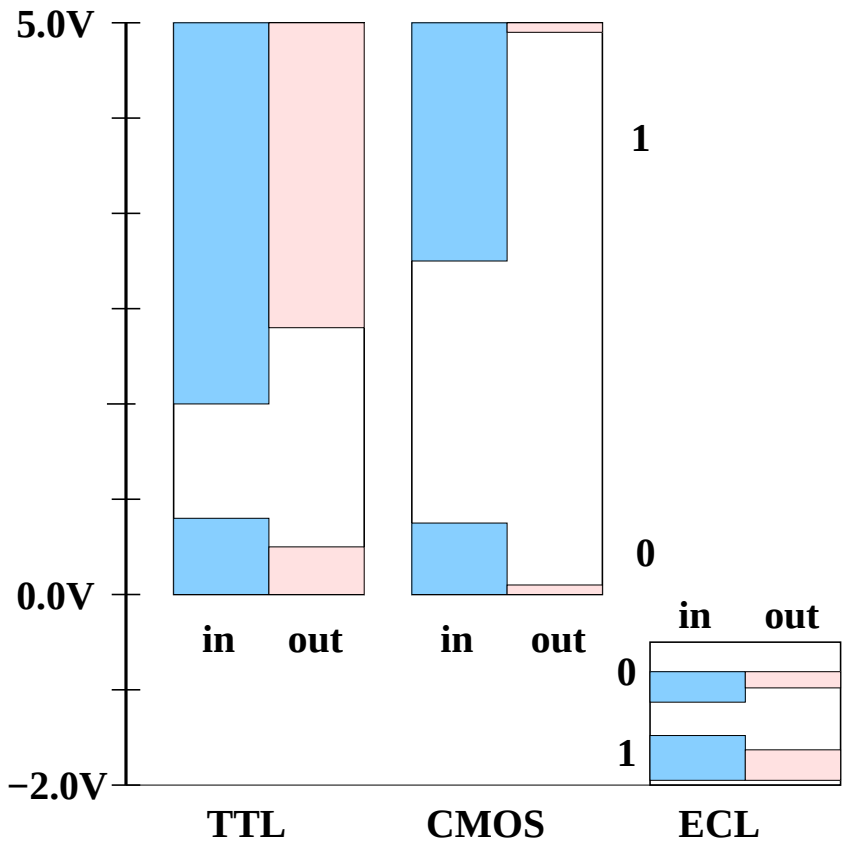

Figure 7.3: A graphical representation of the digital logic levels listed in Table 7.3.

inputs and the output is high if all of the inputs are high, otherwise it is low. The symbol for a AND is a multiplication dot.

There is a second kind of OR known as an exclusive-or (XOR). With two inputs, an XOR outputs a high if exactly one input is high, otherwise it outputs a low. Finally, all of the common operations can have their output negated, or operated on by a NOT. These operations have an N prepended to their names: NOR, NAND, and XNOR.

If we represent the *high* state as 1 and the *low* state as 0, then then we can create a truth table which shows the result of each logical operation on a pair of inputs, A and B. Examples of this for individual logic gates are given in the following sections.

7.4.1 De Morgan's Theorem

In working through logic circuits, *DeMorgan's Theorem* can be used to determine equivalent logical circuits. This theorem states that the complement of an AND is the OR of the complements and vice versa. This theorem yields the following two logic equations.

$$\begin{aligned} \text{NOT (A AND B)} &= \text{(NOT A) OR (NOT B)} \\ \text{NOT (A OR B)} &= \text{(NOT A) AND (NOT B)} \end{aligned}$$

These can be expressed in logic notation as

$$\begin{aligned} \overline{A \cdot B} &= \overline{A} + \overline{B} \\ \overline{A + B} &= \overline{A} \cdot \overline{B}. \end{aligned}$$

7.4.2 Digital Gates

We now want to consider the electronic gates that perform the logical functions which we have discussed. Some of the earliest digital logic was done with transistors and resistors, and some of these circuits are still illustrative. With the advent of integrated circuits, things changed considerably and the basic logic gates are now considered fundamental elements of circuit design. However, because it is illustrative of how a gate could be built, Figure 7.4 shows three complementary gates built using transistors: a NOR gate, a NAND gate and a NOT. In the case of the NOR gate, if either of the inputs are high, then the corresponding transistor goes into saturation and the voltage at the output O goes low. If both inputs are low, then the output O is high. One interesting fact is that NAND gates and NOR gates are generally simpler than AND and OR gates. As will be seen later, it is possible to build all the logic gates out of either a NAND or a NOR. This is not true of the AND and OR.

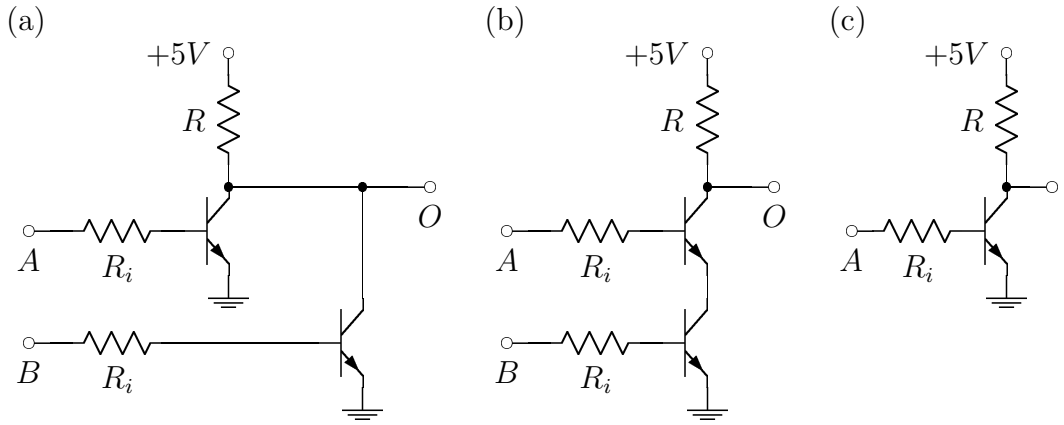

Figure 7.4: Circuit (a) is a NOR gate built using two transistors. Circuit (b) is a NAND gate built using two transistors. Circuit (c) is a NOT gate.

The NOT Gate

The NOT gate takes a single input, A and returns the complement of A as its output, $Q = \bar{A}$. The logic symbol for the NOT is shown in Figure 7.5along with the IEEE rectangular symbol. (Note that an open circle at the input or output of a logic gate is used to represent the complement of the signal.) The truth table for the NOT gate is given in Table 7.4.

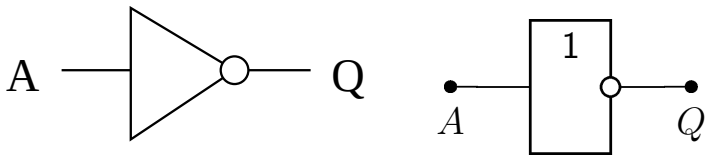

Figure 7.5: The logic symbol for a NOT gate (left) and its IEEE rectangular symbol (right).

A	Q
0	1
1	0

Table 7.4: The truth table for the NOT gate.

The AND Gate

The AND gate takes two inputs, A and B, and returns an output which is the logical AND of the two inputs. The logic symbol and the IEEE rectangular symbol for the gate are shown in Figure 7.6. The truth table for the AND is given in Table 7.5.

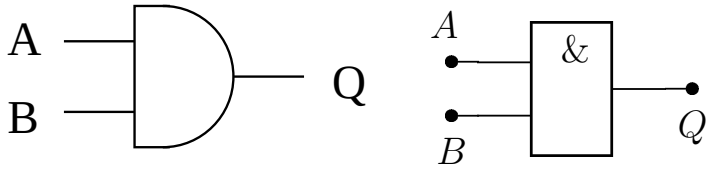

Figure 7.6: The logic symbol for a AND gate (left) and the IEEE rectangular symbol (right).

A	B	Q
0	0	0
1	0	0
0	1	0
1	1	1

Table 7.5: The truth table for the AND gate.

The NAND Gate

The NAND gate takes two inputs, A and B, and returns an output which is the logical NAND of the two inputs. The logic symbol and the IEEE rectangular symbol for the gate are shown in Figure 7.7. The truth table for the NAND is given in Table 7.6.

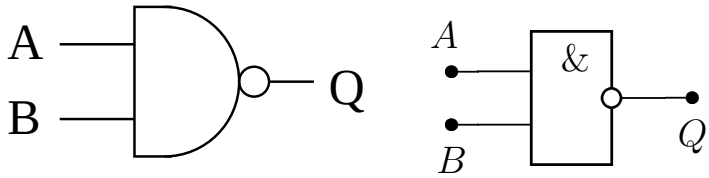

Figure 7.7: The logic symbol for a NAND gate (left) and the IEEE rectangular symbol (right).

A	B	Q
0	0	1
1	0	1
0	1	1
1	1	0

Table 7.6: The truth table for the NAND gate.

The OR Gate

The OR gate takes two inputs, A and B, and returns an output which is the logical OR of the two inputs. The logic symbol and the IEEE rectangular symbol for the gate are shown in Figure 7.8. The truth table for the OR is given in Table 7.7.

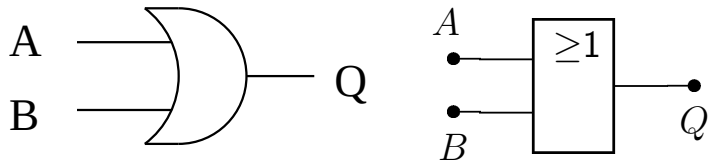

Figure 7.8: The logic symbol for a OR gate (left) and the IEEE rectangular symbol (right).

A	B	Q
0	0	0
1	0	1
0	1	1
1	1	1

Table 7.7: The truth table for the OR gate.

The NOR Gate

The NOR gate takes two inputs, A and B, and returns an output which is the logical NOR of the two inputs. The logic symbol and the IEEE rectangular symbol for the gate are shown in Figure 7.9. The truth table for the NOR is given in Table 7.8.

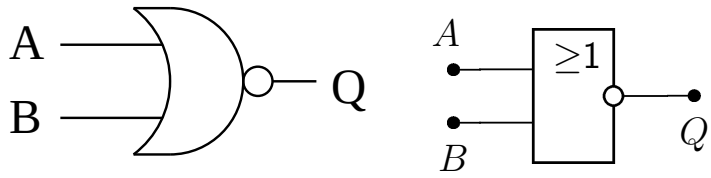

Figure 7.9: The logic symbol for a NOR gate (left) and the IEEE rectangular symbol (right).

A	B	Q
0	0	1
1	0	0
0	1	0
1	1	0

Table 7.8: The truth table for the NOR gate.

The XOR Gate

The XOR gate takes two inputs, A and B, and returns an output which is the exclusive OR (XOR) of the two inputs. The logic symbol and the IEEE rectangular symbol for the gate are shown in Figure 7.10. The truth table for the XOR is given in Table 7.9.

The XNOR Gate

The XNOR gate takes two inputs, A and B, and returns an output which is the logical XNOR of the two inputs. The logic symbol and the IEEE rectangular symbol for the gate are shown in Figure 7.11. The truth table for the XNOR is given in Table 7.10.

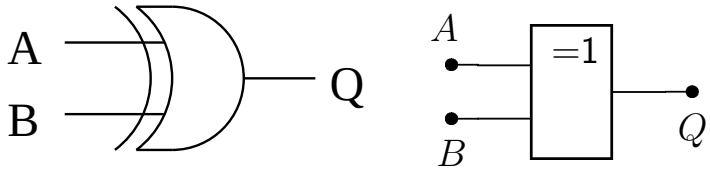

Figure 7.10: The logic symbol for a XOR gate (left) and the IEEE rectangular symbol (right).

A	B	Q
0	0	0
1	0	1
0	1	1
1	1	0

Table 7.9: The truth table for the XOR gate.

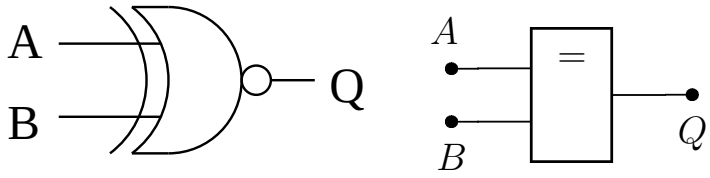

Figure 7.11: The logic symbol for a XNOR gate (left) and the IEEE rectangular symbol (right).

A	B	Q
0	0	1
1	0	0
0	1	0
1	1	1

Table 7.10: The truth table for the XNOR gate.

Gates With More Than Two Inputs

It is possible to construct logic gates that have more than two inputs. In principle, AND and OR gates can have as many inputs as desired. The truth tables just extend those in Tables 7.5 and 7.7. The AND will have a high output if **all** of its inputs are high, otherwise, the output will be low. The OR will give a high output if **at least** one input is high; if all inputs are low, it will return a low output. The standard symbols for these are just the logic symbols in Figure 7.6 and 7.8 with as many inputs as desired. Related to these gates are so-called *Majority Logic Gates* [1]. These return a high output if a *majority* of the inputs are true.

7.4.3 Putting Gates Together

Logic gates typically come in integrated circuit (IC) packages with more than one gate per chip. In addition to the gate inputs and output, the ICs have external power and a ground connections. As with op-amps, these external connections are typically not shown in logic diagrams.

As we mentioned earlier, it is possible to build gates out of other gates. De Morgan's theorem gives us ways to relate one gate to another. Figure 7.12 shows how it is possible to build an AND gate using two NAND gates in series, and how to build an OR gate using two NOR gates in series. In both cases, we

[1] As far as the author knows, the complement of a majority logic gate is not a *Minority Empowerment Gate*.

rely on the fact that if both the A and B inputs to one of these complementary gates are the same, the output is the complement of the input, or a NOT operation.

$$\begin{aligned} \overline{A \cdot A} &= \bar{A} \\ \overline{A + A} &= \bar{A} \end{aligned}$$

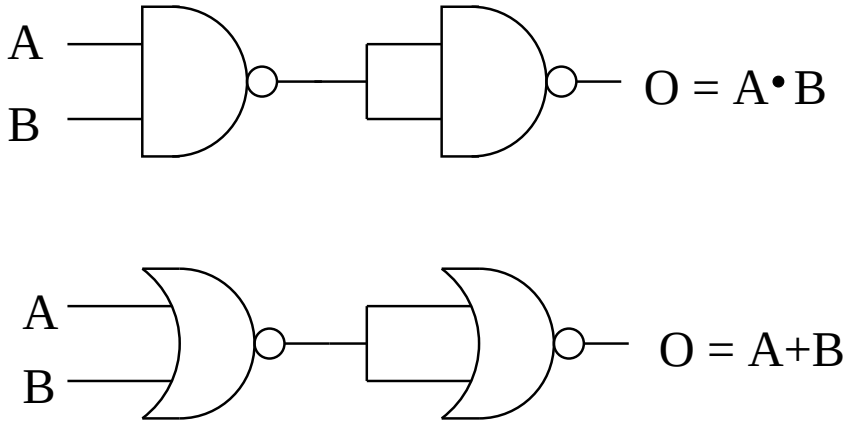

Figure 7.12: The upper figure shows an AND gate built using two NAND gates. The lower figure is an OR gate built using two NOR gates.

We can continue to push this a bit. If we have only NAND gates, it is also possible to build a NOR gate, and hence an OR gate as well. The idea is to use de Morgan's theorem to convert an AND into an OR. To show this, we start with

$$\bar{A} \cdot \bar{B} = \overline{A + B}\,.$$

If we now complement both sides, we get

$$\overline{\bar{A} \cdot \bar{B}} = A + B\,.$$

In Figure 7.13 we show how both an OR gate and a NOR gate can be built entirely out of NAND gates.

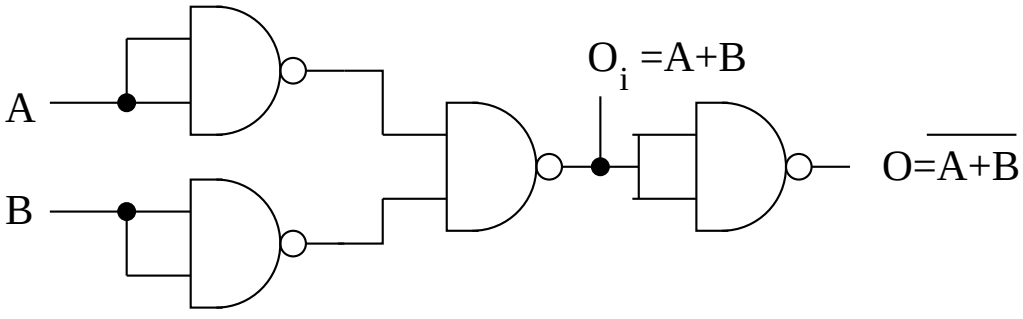

Figure 7.13: An NOR gate built out of four NAND gates. By tapping the output at O_i, we have an OR gate instead.

Finally, we could look at XOR gates. To create these out of other gates requires both AND and OR gates. An example of this is shown in Figure 7.14 where we have created an XOR using the following logic:

$$A \wedge B = (A + B) \cdot \overline{(A \cdot B)}\,,$$

A	B	$\bar{A}$	$\bar{B}$	O_i	O
0	0	1	1	0	1
1	0	0	1	1	0
0	1	1	0	1	0
1	1	0	0	1	0

Table 7.11: A truth table that walks us through the logic of Figure 7.13. Inputs A and B are complemented in the first stage, and then put into an AND gate to yield the intermediate output O_i. This in turn is complemented to yield O.

where the $\wedge$ symbol is used to represent the XOR operation.

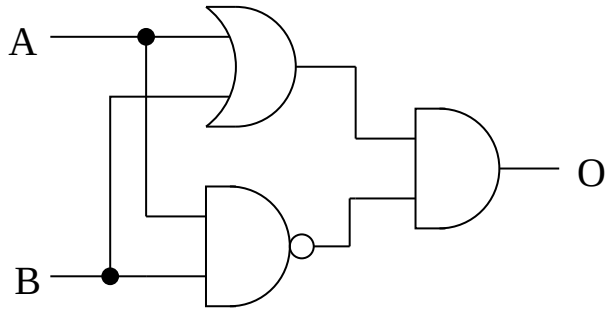

Figure 7.14: An XOR gate built using other logic gates.

7.5 Digital Memory Circuits: Flip-flops

So far we have examined gates which allow us to perform simple logic operations on one or more inputs. Going beyond this, it would be useful to have a logic circuit that could remember the state it is in, even its input changes. This is the basic idea behind a *memory* circuit. Essentially, we have a device that can be set into one of two states by appropriate inputs. The input can then go off, but the output remains in the set state. While there are many different circuits that can carry out this function, we will examine the so-called *flip-flop*, which as the name implies can go between two states.

7.5.1 The Reset-Set Flip-flop

This simplest flip-flop circuit is the so-called *reset-set* or RS flip flop. This basic circuit is shown in Figure 7.15, where two versions (related by De Morgan's theorem) are given. The left-hand circuit is built from NAND gates and its function is described by the following recursive logic equations:

$$X = \overline{(A \cdot Y)} \tag{7.1}$$
$$Y = \overline{(B \cdot X)} \tag{7.2}$$

For a given input, A and B, we need to find an output state, X and Y, that satisfies equations 7.1 and 7.2. If we apply De Morgan's theorem to these two equations, we arrive at the following, which describe the circuit on the right-hand side of Figure 7.15.

$$X = \overline{A} + \overline{Y}$$
$$Y = \overline{B} + \overline{X}$$

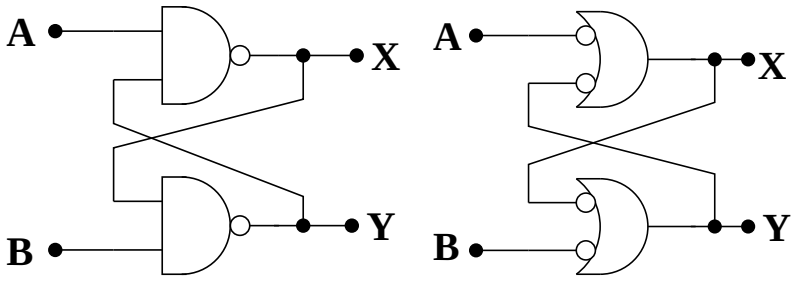

Figure 7.15: Two identical flip-flop circuits. The left-hand figure shows two NAND gates, with the logic of the two gates being $\overline{A \cdot Y}$ and $\overline{X \cdot B}$. By De Morgan's Theorem, this is equivalent to the right-hand circuit where complements of the inputs are put into an OR gate, $\overline{A} + \overline{Y}$ and $\overline{X} + \overline{B}$.

Solving equations 7.1 and 7.2, we obtain the results given in Table 7.12. There are a couple of interesting things to note about this circuit. If B is high and A is low, then we *set* the X output high and Y is the complement of X. If B is low and A is high, then we set the X output low, or *reset* the X output. Again, Y is the complement of X. When we have *set* or *reset* the X output, then if both the A and B inputs are held high, then the flip-flop will retain the X and Y settings. It is these functions that are the most relevant in the functioning of the so-called reset-set or RS flip-flop.

The remaining case where both A and B are low leads to both X and Y being high. If we were to transition from low-low to high-high input, then the output of the flip-flop would be uncertain. It would fall into one of the two states depending on mostly random factors.

A	B	X	Y
0	0	1	1
1	0	0	1
0	1	1	0
1	1	1	0
1	1	0	1

Table 7.12: The truth-table for the flip-flop circuit shown in Figure 7.15. The crucial feature related to memory is that there are two valid outputs for the case when both inputs are high, and these outputs depend on previous states.

Looking at the truth table, one can discern why the complementary logic shown on the right in Figure 7.15 is used. In this case, when both A and B are low, the flip-flop holds its current state. However, independent of how one looks at the circuit, it is the ability to set a *bit*, Q, either high or low, and then to hold that particular value that makes the RS flip-flop a memory circuit. One flip-flop can store one bit of information.

In Figure 7.16 we show the circuit element from Figure 7.15 as an RS flip-flop. The A input is the Reset, the B input is the Set, the X output is Q and the Y is given as the complement, $\overline{Q}$. The RS flip-flop is both the basis of more complicated flip-flops as well as other useful circuits which follow.

A Switch De-bouncer

Aside from its use as a memory circuit, the RS flip-flop finds other useful applications. One common use is in *de-bouncing* switches. When a typical switch is closed, the contacts may actually open and close many times before settling down in the closed state. Generally, this is not a desirable situation. We can take advantage of the memory ability of the RS flip-flop to eliminate this bouncing.

Figure 7.17 shows such a de-bouncing circuit. When the switch is connected to the R input, R goes low and S become high. This causes the output to go high. When the switch is connected to S, then R

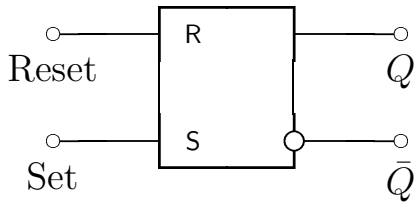

Figure 7.16: An RS flip-flop symbol. The two input are labeled Set and Reset, while it puts out Q and $\bar{Q}$ outputs. If Set is high, then Q is high. If Reset is high, then Q is low.

goes high and S goes low, leading to a low output. When the switch is in between the two terminals, then both R and S are high, and the current state is held. Closing the switch sets the output state and the memory features of the flip-flop prevent the bouncing from changing that state.

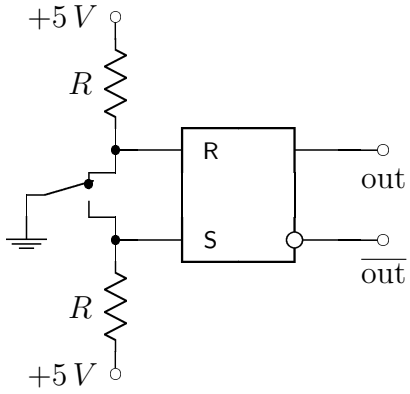

Figure 7.17: An RS flip-flop hooked up as a switch de-bouncer. As the switch is toggled back and forth between the Set and Reset inputs, the Q is toggled between high and low, with a single change of state each time the switch is flipped.

Clocked Flip-flops

The clocked RS flip-flop is shown in Figure 7.18. In this circuit, the S and R input can be considered data lines that get clocked into the flip-flop when the clock pulse is high, and then held while the clock pulse is low. Looking in detail at the circuit, we see that when the clock is low, the output of both NAND gates will be high and the RS flip-flop will be in its hold mode. When the clock is high, it is combined with the two data inputs to produce the output from the RS flip-flop. The truth table for this is given in Table 7.13.

When both of the data lines are low, the output is indeterminate. This forces the flip-flop into the mode where both Q and $\overline{Q}$ are high. When the clock pulse then goes low, the flip-flop drops randomly into one of its two states. We will continue the idea of a clocked flip-flop as we discuss the D flip-flop and the JK flip-flop in the following sections.

7.5.2 The JK Flip-flop

The JK flip-flop is a clocked flip-flop with two inputs, J and K. A simple schematic diagram for a JK flip-flop is given in Figure 7.19. The output side of the circuit is effectively an RS flip-flop, however the crossover feedback in the RS flip-flop is also fed back into the input of the JK flip-flop. Let us now look in detail what happens in the JK flip-flop.

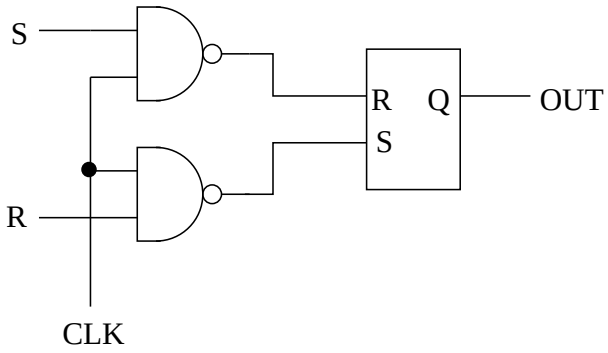

Figure 7.18: A clocked RS flip-flop.

S	R	Q_{n+1}
0	0	Q_n
0	1	0
1	0	1
1	1	Not Defined

Table 7.13: The truth table for the clocked flip-flop (Figure 7.18) when the clock pulse is high. When the clock pulse is low, the current state of the flip-flop is held.

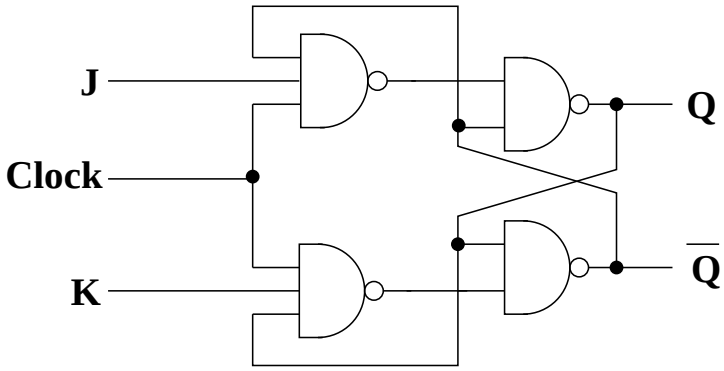

Figure 7.19: A JK flip-flop circuit.

First we comment on the two three-input NAND gates at the input of the JK flip-flop. The output of these will be high *unless* all three input lines are high, in which case it will be low. This means that if the clock input is low, the output of both NAND gates will be high and the RS flip-flop will hold its current state.

If the clock input is high, then the states of the other inputs matter. If both J and K are low, we are also in the hold state. If either is high we set the state of the flip-flop. If J is high, then Q goes high, and if K is high, then Q goes low. If all three inputs to the JK flip-flop are high, then both Q and $\bar{Q}$ flip states. This behavior is summarized in Table 7.14. The toggling behavior has applications in more sophisticated circuits such as counters and shift registers which will be discussed later.

7.5.3 The D Flip-flop

The D flip-flop is an edge-triggered flip-flop that takes the signal input on its *data* line, D, and puts it on the output line, Q, when the clock makes a transition (known as an *edge*). Some flip-flops transfer

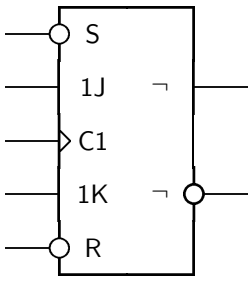

Figure 7.20: A typical circuit symbol for a JK flip-flop.

$\overline{CLR}$	CLK	J	K	Q	$\bar{Q}$	Comment
L	x	x	x	L	H	Default
H	on	L	L	Q_{n-1}	$\bar{Q}_{n-1}$	Hold
H	on	H	L	H	L	Set
H	on	L	H	L	H	Set
H	on	H	H	$\bar{Q}_{n-1}$	Q_{n-1}	Toggle

Table 7.14: The truth table for the JK flip-flop.

the data on the rising edge and others on the falling edge. The logic in a simple D flip-flop is shown in Figure 7.21. Let us look in detail at its behavior.

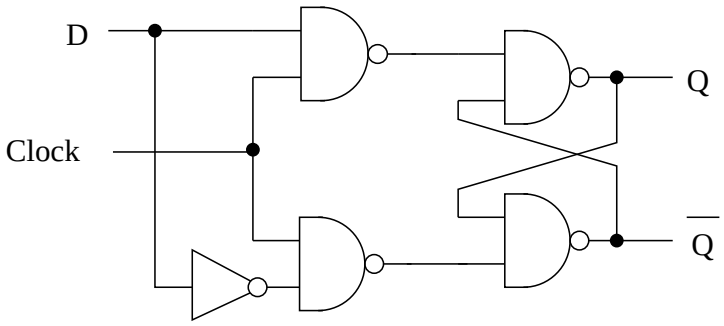

Figure 7.21: A D flip-flop circuit.

We can first note that the two NAND gates connected to the outputs Q and $\bar{Q}$ are actually an RS flip-flop with the truth table given in Table 7.12. To simplify things, we will consider the upper input to this flip-flop as A and the lower input as B, as in Figure 7.15. We now look at the output of the two NAND gates on the input side of the flip-flop.

If the clock signal is low, then both NAND gates will output a high signal (a NAND produces a low output only if both inputs are high). This means that both A and B are high and the RS flip-flop holds its current state.

If the clock goes from low to high, then one of the two NAND gates will output a low and the other a high. If D is high, then A will be low and B will be high. If D is low, then A will be high and B will be low. In other words, A will be $\bar{D}$ and B will be D. This then leads to the truth table in Table 7.15. When the clock is high, Q will take on whatever value D has. When the clock is low, it will remember the value.

A typical symbols for a D flip-flops is shown in Figure 7.22. However, there are likely to be additional inputs which allow one to set or clear the outputs, independent of the input levels.

Clock	D	Q_{n+1}
L	L	Q_n
L	H	Q_n
↑	L	L
↑	H	H

Table 7.15: The truth table for the D flip-flop.

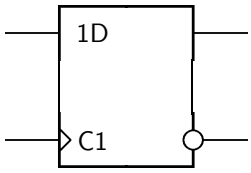

Figure 7.22: A typical circuit symbol for a D flip-flop.

7.6 Clock Circuits

In the previous section, we saw several instances of a clock pulse being used as input to a gate. In fact, if we look at a digital computer, one of the fundamental elements is a clock. Essentially, the computer performs an operation at every clock pulse. Being able to generate and use clock pulses is a crucial element of modern digital electronics. Precision clocks are often driven by crystal oscillators, and these can be easily purchased. In this course, we will look at a clock circuit that can be built out of the circuit elements we have studied. Such a circuit is the basis of a common off-the-shelf clock chip known as the 555.

7.6.1 The 555 Chip

A common timer chip is the 555 family of chips. The pin out of the in-line package is shown in Figure 7.23. Table 7.16 lists the eight pins with their names and the abbreviation used in the circuit diagrams for these pins. The 555 is designed to output a clock pulse from the output terminal. The period and shape of the pulse is controlled by two external resistors and an external capacitor that are connected to the 555. This allows for a choice of time constants, with the period of the clock pulse given

Pin	Name	Abbreviation
1	Ground	
2	Trigger	IT
3	Output	O
4	Reset	R
5	Control Voltage	OK
6	Threshold	IS
7	Discharge	OD
8	V_{CC}	

Table 7.16: A list of the names of the eight pins on the 555 and the symbol used in this text for each of them.

by $(R_1 + 2 \cdot R_2)\, C$.

The main elements in the 555 are two comparator circuits, as shown in Figure 7.24. The upper comparator has a threshold voltage of $V_{ref} = \frac{2}{3}V_{CC}$. When the *Threshold* input is below $\frac{2}{3}V_{CC}$, the comparator has a low (ground) output. When Threshold is above $\frac{2}{3}V_{CC}$, the comparator has a high (V_{CC}) output.

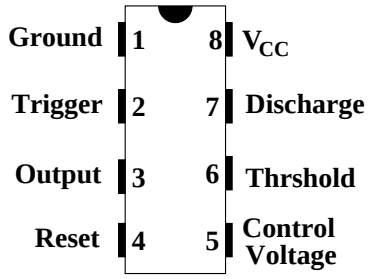

Figure 7.23: The pin out of an in-line 555 timer chip.

The lower comparator has a threshold voltage of $\frac{1}{3}V_{CC}$. When the *Trigger* input is below $\frac{1}{3}V_{CC}$, then the output of the comparator goes high (V_{CC}). If the Trigger is above $\frac{1}{3}V_{CC}$, the the comparator output goes low (ground). The outputs of the two comparators are used as *set* and *reset* inputs to a flip-flop. The output of the flip-flop drives the output pulse, and when it is high, puts the transistor into saturation, which effectively connects the *Discharge* line to ground.

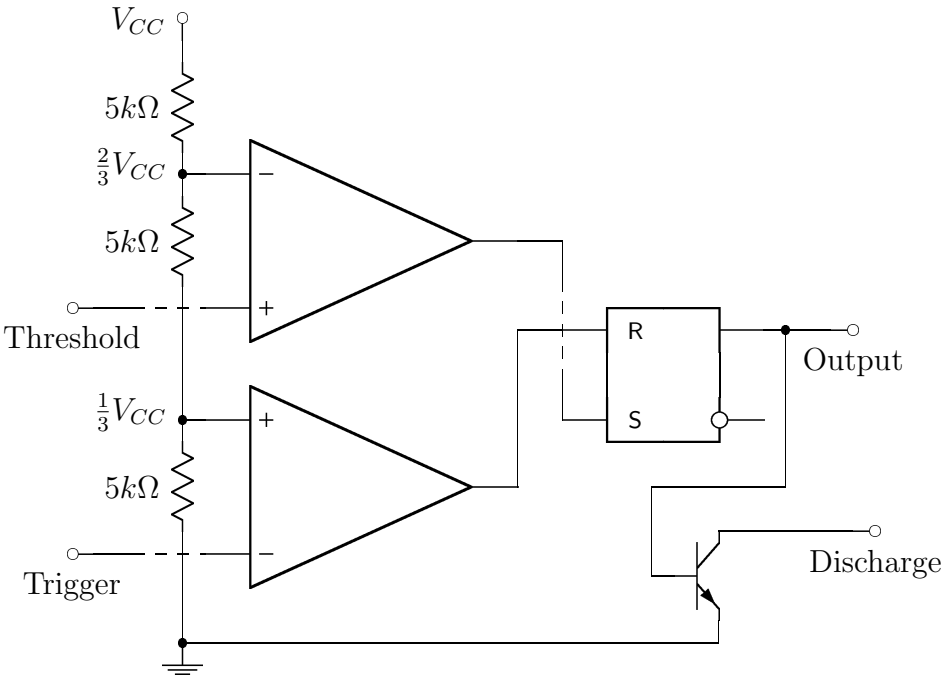

Figure 7.24: The circuit inside the 555, showing the two comparators and RS flip-flop.

7.6.2 A Clock Pulse Generator

We might want to use the 555 used in a circuit to produce a regular output pulse. In order to do this, we need to connect two resistors and a capacitor to the circuit. An example circuit is shown in Figure 7.25.

The external resistors form an RC circuit between V_{CC} and ground. This charges up the capacitor with a time constant of $\tau_1 = (R_1 + R_2) \cdot C$. Let us now look at what happens for various voltage ranges at the point P. Table 7.17 shows the output of the RS flip-flop inside the 555 as a function of the voltage at the point P. If the voltage falls below $\frac{1}{3}V_{CC}$, the clock circuit will output a low signal. If V_P rises above $\frac{2}{3}V_{CC}$, the clock circuit will output a high signal. For voltages in between, the circuit will hold

the current output state.

V_P	R	S	O
0 to $\frac{1}{3}V_{CC}$	L	H	L
$\frac{1}{3}V_{CC}$ to $\frac{2}{3}V_{CC}$	L	L	Hold
$\frac{2}{3}V_{CC}$ to V_{CC}	H	L	H

Table 7.17: The output of the RS flip-flop in the 555 timer (shown in Figure 7.24) as a function of the voltage at point P in the circuit of Figure 7.25. When V_P is between $\frac{1}{3}$ and $\frac{2}{3}$ of V_{CC}, the flip-flop simply holds its current output.

Assume that the capacitor is initially uncharged. It will begin to charge up from zero with time constant $\tau_1 = (R_1 + R_2) \cdot C$. While V_P is smaller than $\frac{1}{3}V_{CC}$, the output will be low and the transistor that connects Discharge to ground will be off. When $V_P = \frac{1}{3}V_{CC}$ the flip-flop will go into its hold mode. The circuit will continue to charge up from $\frac{1}{3}V_{CC}$ to $\frac{2}{3}V_{CC}$ and the output will remain low. When V_P reaches $\frac{2}{3}V_{CC}$, the output will change to high and the transistor will go into saturation, effectively connecting the Discharge terminal to ground. This will cause the potential at point G to go to ground. The capacitor will start discharging through R_2 to ground. When V_P reaches $\frac{1}{3}V_{CC}$, the output will go low, and the capacitor will start to charge up again.

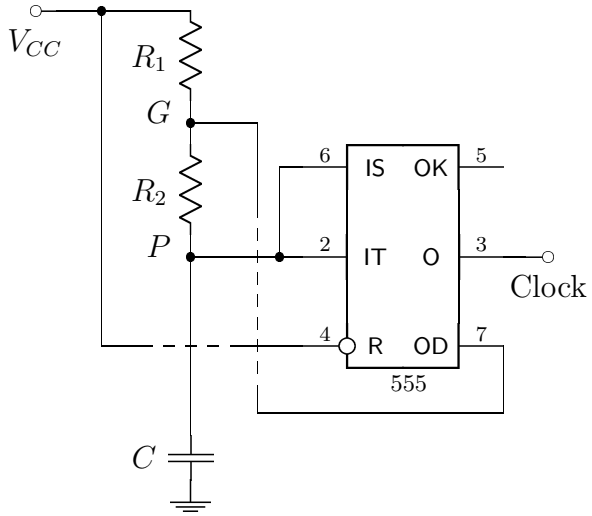

Figure 7.25: This circuit shows a 555 timer hooked up to produce a clock output signal with period $T = 0.693 \cdot (R_1 + 2R_2)\, C$. The output is on the *Clock* line.

The upper plot in Figure 7.26 gives the voltage at P as a function of time. This shows the charge-up with time constant $(R_2 + R_1) \cdot C$ and the discharge with time constant $R_2 \cdot C$. Since charging and discharging of a capacitor go as

$$V(t) \quad \sim \quad e^{-t/\tau}\,,$$

the time to go from $\frac{1}{3}V_o$ to $\frac{2}{3}V_o$ can be found from

$$\begin{aligned} \frac{1/3}{2/3} &= e^{-t/\tau} \\ t &= \ln(2) \cdot \tau\,, \end{aligned}$$

so the period is $0.693 \cdot \tau$. For our circuit, we then have

$$T_{555} \quad = \quad 0.693 \cdot (R_1 + 2R_2) \cdot C\,. \tag{7.3}$$

We would also like to know the ratio of high clock pulse time to low. In the case where R_1 equals R_2, the output clock pulse is shown in the lower plot of Figure 7.26. The fraction of time in which the pulse is low is $f_{low} = \frac{R_1+R_2}{R_1+2R_2}$, while the fraction in which it is high is$f_{high} = \frac{R_2}{R_1+2R_2}$. In the case that R_1 is much smaller than R_2, both of these approach $\frac{1}{2}$.

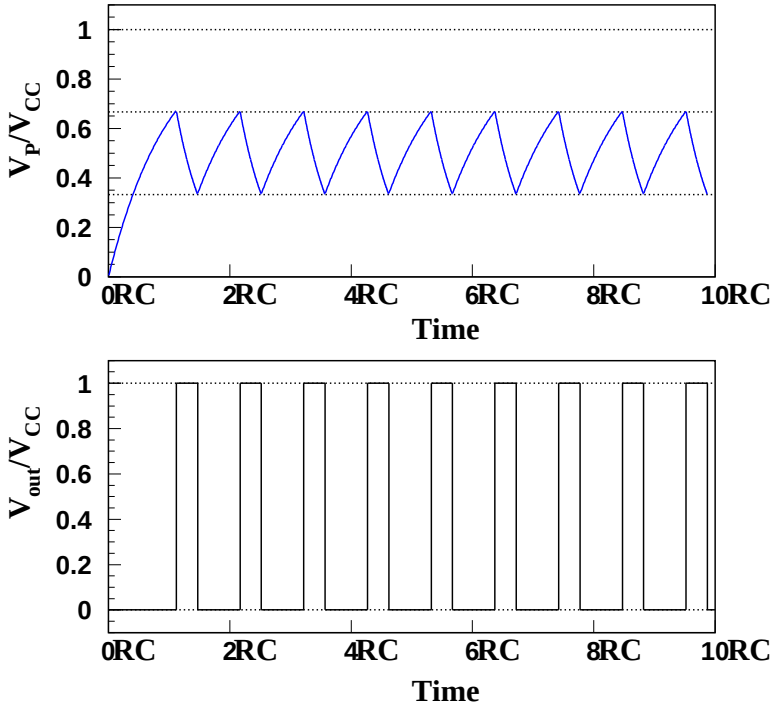

Figure 7.26: The upper plot shows the voltage at point P in the left-hand circuit of Figure 7.25 as a function of time. The lower plot shows the output of the circuit as a function of time. It is assumed that $R_1 = R_2 = R$.

7.6.3 A Gate Generator

Another use for the 555 timer is to produce a single output pulse with a controllable pulse width. Such a circuit is useful in triggering circuits when some specified event has occurred. A circuit to do this is shown in Figure 7.27.

7.7 Counter Circuits

The circuit shown in Figure 7.28 does several things. The easiest to see is based on the fact that with both the J and K inputs of the JK flip-flop high, the output Q toggles between high and low each time the clock pulse goes high, and holds its current state when the clock is low. The output of each successive flip-flop is connected to both the clock input of the next flip-flop in the stage, and to an external bit, Q_i as shown in the circuit.

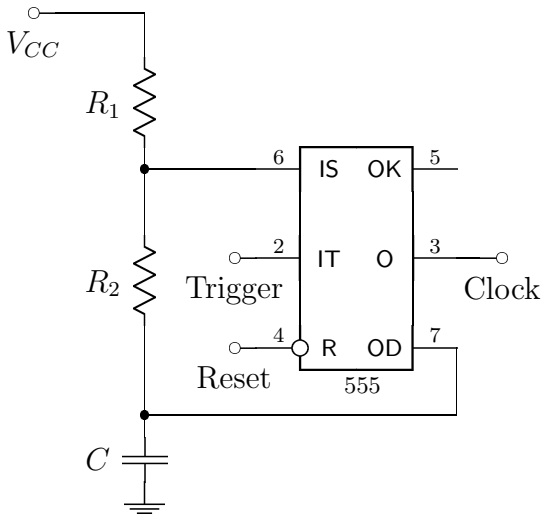

Figure 7.27: A 555 timer hooked up to produce a single output pulse with controllable width.

If we look at the values of Q_i, then we will see the output bits counting up from zero (000) to seven (111), then going back to zero and starting over. We have a simple counting circuit which can be extended as far to the right as we want.

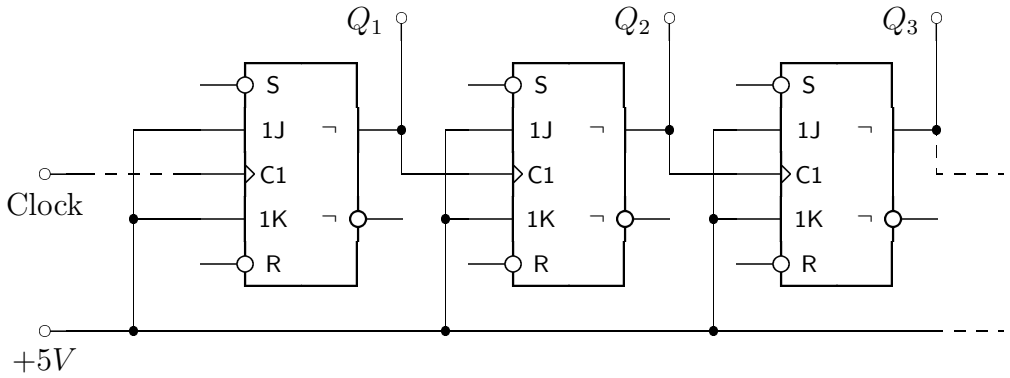

Figure 7.28: A 3-bit binary counter built from three JK flip-flops. The circuit is extensible to the right by adding more flip-flops. The outputs, Q_i, can also be used as clocks themselves. The clock rate at output $i = 1, 2, 3, \ldots$ is $1/2^i$ times the input clock rate.

However, if we examine the outputs at the various Q_i, we see the pulse structure shown in Figure 7.29. The output on Q_1 has $\frac{1}{2}$ the frequency of the clock input, while each successive stage divide the output frequency by another factor of two. Such a circuit is known as a frequency divider. It allows us to take some very large frequency and divide it down by any power of 2.

7.8 Shift Registers

Related to the counting circuit from section 7.7 are the so-called *shift registers*. These circuits shift the bits of a data word one bit to the left or one bit to the right on each clock pulse. An example of such a circuit is shown in Figure 7.30, where we have a 4-bit shift register built from D flip-flops. All of the D flip-flops are clocked from the same clock pulse. On the relevant clock edge, the flip-flops copy their

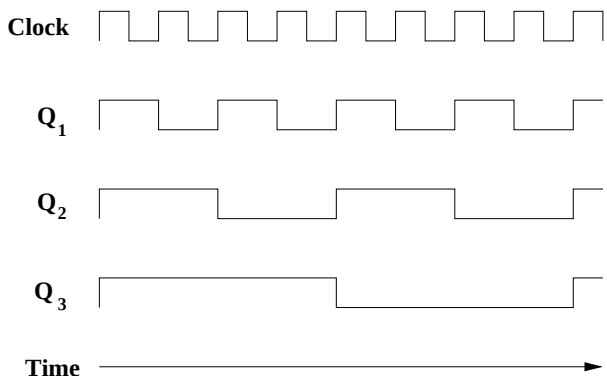

Figure 7.29: The input clock pulse and the signals on the three output bits, Q_1, Q_2 and Q_3, as a function of time.

inputs to their output. If the Data line is high on the first transition, and then subsequently low, the high bit will move to the right through the shift register on each clock pulse.

Typically, the Reset lines of all the D flip-flops are connected to an external Reset line. This allows one to reset all the flip-flops to zero. It is also normal to have a set of data input lines which connect to the Sets of the flip-flops. This allows one to set a particular data word that is shifted through the shift register.

If the bit that falls off the right-hand side of the circuit is fed back into the data input on the left, the shift register is known as *circular*.

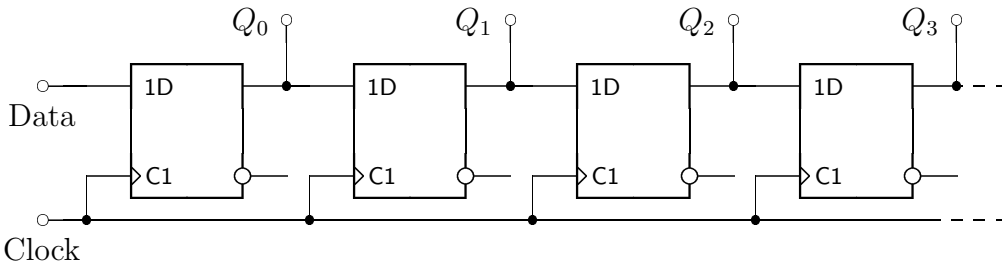

Figure 7.30: A four-bit shift register built with D flip-flops. The data line input is shifted one to the right on each clock pulse input. The system can be extended to the right with additional flip-flops.

7.9 Adder Circuits

7.9.1 The Half Adder Circuit

A half adder takes two bits as input and returns two output bits that represent the sum of the two bits. We refer to these as the *least significant bit*, L, and the *carry bit*, C. The truth table for such a device is given in Table 7.18. If we look at what this is doing, we quickly realize that the carry bit is just the

A	B	C	L
0	0	0	0
0	1	0	1
1	0	0	1
1	1	1	0

Table 7.18: The truth table for the half adder shown in Figure 7.31.

AND of the two inputs and the least significant bit is just the exclusive OR (XOR) of the two inputs. This leads us to the circuit shown in Figure 7.31 for a simple implementation of the half adder.

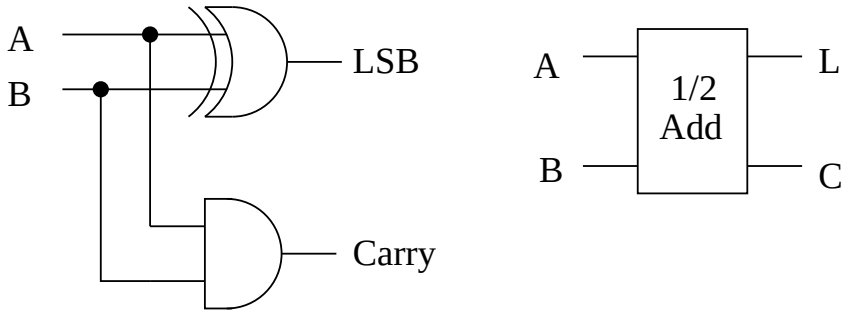

Figure 7.31: A half adder circuit with two inputs, A_1 and B_1 and two outputs, L and C.

While adding two bits is interesting, in fact we would like to be able to do additions on a few more. One implementation of a two-bit adder can be made by using three half adder circuits and an OR gate as shown in Figure 7.32. This circuit has the truth table given in Table 7.19. We could logically continue chaining half adders together increasing the number of bits in the circuit. An unwieldy example is shown in Figure 7.33, where we have put together four layers of half adders to create a four-bit adder circuit.

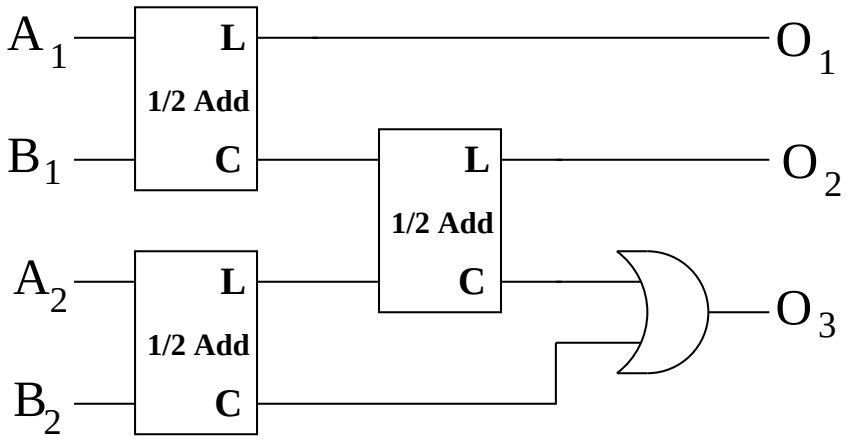

Figure 7.32: A two-bit adder circuit built from three half adders and an OR gate.

A_2	A_1	B_2	B_1	O_3	O_2	O_1
0	0	0	0	0	0	0
0	0	0	1	0	0	1
0	0	1	0	0	1	0
0	0	1	1	0	1	1
0	1	0	0	0	0	1
0	1	0	1	0	1	0
0	1	1	0	0	1	1
0	1	1	1	1	0	0
1	0	0	0	0	1	0
1	0	0	1	0	1	1
1	0	1	0	1	0	0
1	0	1	1	1	0	1
1	1	0	0	0	1	1
1	1	0	1	1	0	0
1	1	1	0	1	0	1
1	1	1	1	1	1	0

Table 7.19: The truth table for the two-bit adder in Figure 7.32.

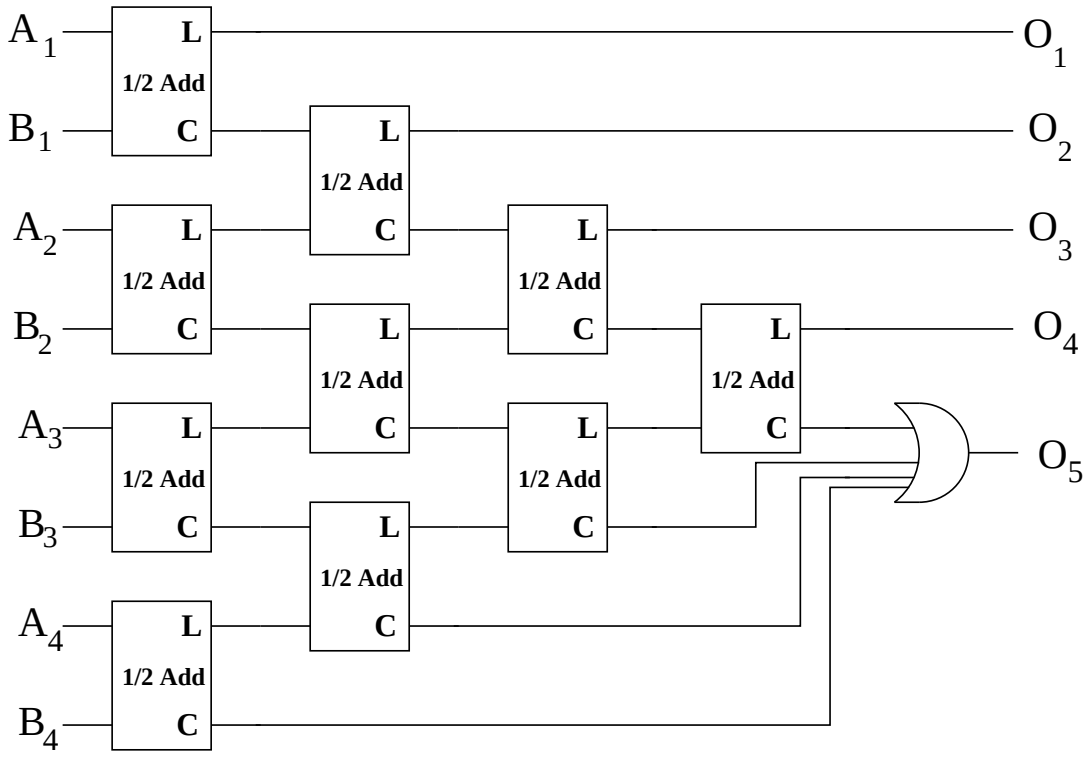

Figure 7.33: A four-bit adder circuit built from half adders and a 4-input OR gate.

7.9.2 Full Adder Circuits

The adder circuits that we have built out of half adders very quickly became quite unwieldy, both in the number of half adders needed, and the increasing "depth" of the logic circuits. It is useful at this point to step back and consider what we really want to do. In fact, we need an adder circuit that accepts three inputs: A, B and an input carry, C_i. The circuit should then produce two output bits, L and C_o, as before. The truth table for such a circuit is given in Table 7.20.

A	B	C_i	C_o	L
0	0	0	0	0
1	0	0	0	1
0	1	0	0	1
0	0	1	0	1
1	1	0	1	0
1	0	1	1	0
0	1	1	1	0
1	1	1	1	1

Table 7.20: The truth table for a full adder. The adder has three inputs, A, B and carry in, C_i, and two outputs, L and carry out, C_o.

The implementation of this circuit is slightly more complicated than the half adder circuit. The full adder is shown in Figure 7.34, where two exclusive ORs are used to determine the least significant bit. The logical OR of a pair of ANDs leads to the carry bit. What makes this adder particularly nice is how it generalizes to further bits. Figure 7.35 shows it expanded to four bits. The logic is quite straightforward, and the depth does not increase as the number of bits are increased. Basically, the depth is limited to three logic gates for any number of bits, whereas the half adder depth goes as the number of bits plus 1.

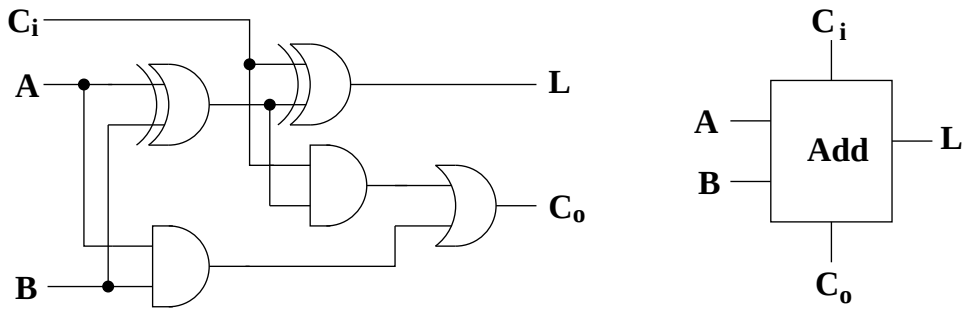

Figure 7.34: A full adder circuit.

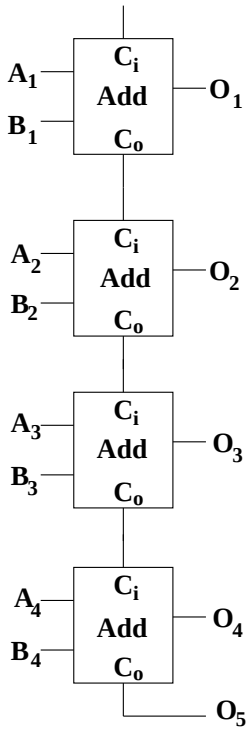

Figure 7.35: A 4-bit adder built using four full adder circuits.

7.10 Converters

In many situations it is desirable to digitize some signal, or to turn some digital signal into an analog signal. In this section, we discuss several converters which carry out this task.

7.10.1 Analog to Digital Converters

An analog to digital converter, ADC, accepts an analog input voltage and outputs a digital signal that indicates the magnitude of the voltage. In section 6.8, we looked at simple comparator circuits that output two different voltages depending on whether the input voltage was larger or smaller than a reference voltage. It is this behavior that is used as the basis of an ADC. The basic idea is to set up a series of comparators that whose output is either a low or a high signal depending on whether the input voltage is larger or smaller than the comparator's reference voltage. The outputs of the comparators are then multiplexed together to form a digital output that represents the input voltage.

To design such a circuit, we must specify an allowed range of allowed input voltages, v_{min} to v_{max}. In addition, we need to specify how many bits of precision we want the output to have. As we saw above, for a single bit of precision, we needed one comparator. For two bits of precision, we would need three comparators. Continuing, for n bits of precision, we would need $2^n - 1$ comparators. Generally speaking, the comparators would be set up to have uniformly spaced reference voltages:

$$v_{ref} = \frac{1}{n}\left(v_{max} - v_{min}\right), \frac{2}{n}\left(v_{max} - v_{min}\right), \ldots, \frac{n-1}{n}\left(v_{max} - v_{min}\right) .$$

Assuming that the input voltage is between v_{min} and v_{max}, then all comparators whose reference is below v_{in} would output a high signal and all which are above would output a low signal. Figure 7.36 shows a three-bit ADC. On the left side we see 7 comparators whose reference voltages are driven by the voltage divider network from V_{ref} to ground.

We now would like to turn the comparator output into a digital signal. This is done with the series of XOR gates which are driven by the comparators. The XOR associated with the comparator with the largest reference voltage has one of its inputs grounded. If the comparator is high, then this XOR will have exactly one high input, and therefore its output will be high. In addition, all the comparators below the highest one will also have a high output. This means that all the XOR gates below the top one will have two high inputs, so their outputs will be low. Only the XOR that sits *above* the highest comparator that has a high output will have a high output.

Since at most one XOR will have a high output, we can have each XOR drive the necessary bits to produce the correct 3-bit output. We add one extra step, which is to put diodes between the XORs and the output bus and then put some pull-down resistors at the bottom of the buses so the voltages at the outputs will effectively be the high signal (minus a diode drop) from the output of the XOR.

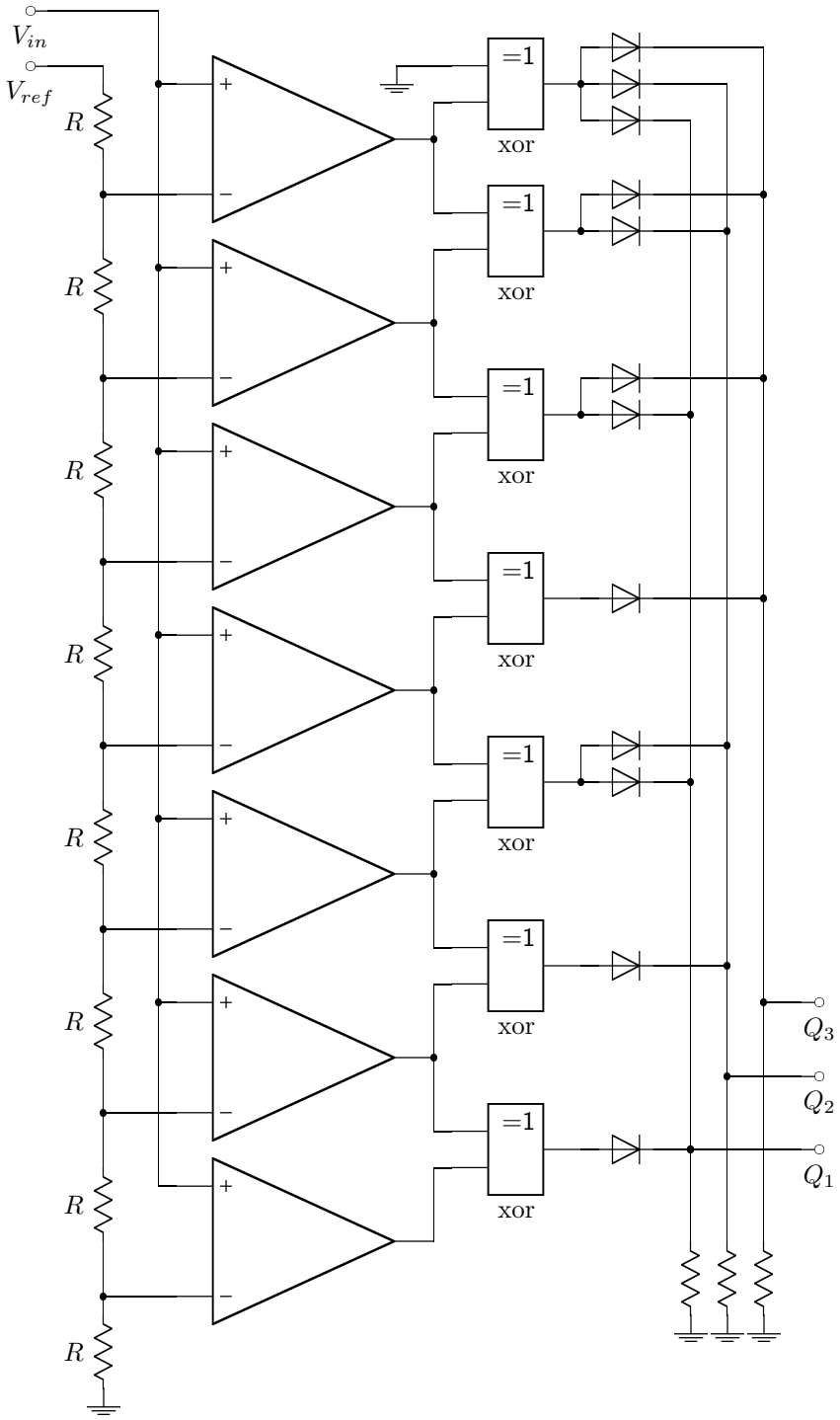

Figure 7.36: An analog to digital converter which is able to output 0 to 7 inclusive in binary on the three output bits.

7.10.2 Digital to Analog Converters

The reverse of the ADC is the digital to analog converter, or DAC. This circuit would take a digital input and produce an analog voltage that is proportional to the digital input. An example of an R2R DAC circuit is shown in Figure 7.37. The resistors need to be accurately matched, but the exact value of R is not crucial. The digital input is set by toggling the switches. A zero has a switch connected to ground, while a one has the switch connected to the op-amp's inverting input.

The left-most switch corresponds to the most significant bit, while the right-most switch is the least. The op-amp circuit is just an adder circuit as discussed in section 6.3.5. Moving across the top row of resistors from left to right, the voltage at the various junctions are $\frac{1}{2}V_{cc}$, $\frac{1}{4}V_{cc}$, $\frac{1}{8}V_{cc}$ and $\frac{1}{16}V_{cc}$. Each of these voltages can be connected to the inverting input of the op-amp through a resistor R, which is matched to the feedback resistor in the circuit. The adder just produces an output which is the negative of the sum of the connected input voltages, that is an analog output which can take on values between 0 and $-V_{CC}$ in $\frac{1}{16}V_{CC}$ volt steps.

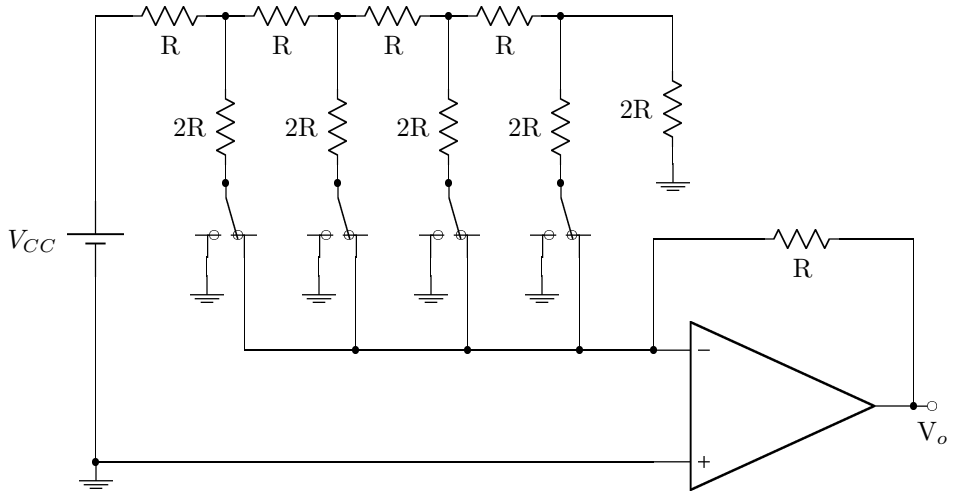

Figure 7.37: A four-bit digital to analog converter. The output voltage will range from 0 to $-\frac{15}{16}V_{CC}$.

Appendix A

Component Labels

Components are labeled in a number of ways. Color codes are typical for resistors, while various combinations of numbers, letters and colors may be used for capacitors. In this section, we show some of the common methods of labeling components. This list is by no means comprehensive, and when in doubt about a particular component value, measuring it is always the right thing to do.

As mentioned above, colors are used in labeling several different components. The color-to-number correlation is given in Table A.1. There are numerous poems and phrases in which the first letter of the color name is the first letter of a word, and the poem or phrase follows the 0-to-9 order in the following table. None of these phrases will be repeated here.

Color	Value
Black	0
Brown	1
Red	2
Orange	3
Yellow	4
Green	5
Blue	6
Violet	7
Grey	8
White	9

Table A.1: The color-numeric correspondence in electrical components.

A.1 Resistor Codes

The most common usage of color codes is in labeling resistors. Typical resistors have four color bands as shown in the left-hand picture of Figure A.1. Precision resistors have a 5th colored band as shown in the right side plot of the figure. To determine the value of a resistor with n bands, the first $n-2$ bands give the numerical part of the resistance, the $(n-1)$th band gives the power-of-ten multiplier, and the nth band gives the tolerance. Table A.2 shows how this works for four-band resistors, while table A.3 explains five-band resistors.

A couple of resistor examples are shown in Figure A.2. The first has four bands: red-black-yellow-gold. From Table A.2, we see that the first two bands give s 20. The third band says that we multiply this by 10^4 and the fourth band indicates that we have a 5% tolerance. The resistor is $R = 200\,k\Omega$ ($\pm 5\%$). The second example is a 5-band resistor with purple-black-brown-red-brown. Table A.3 gives us the

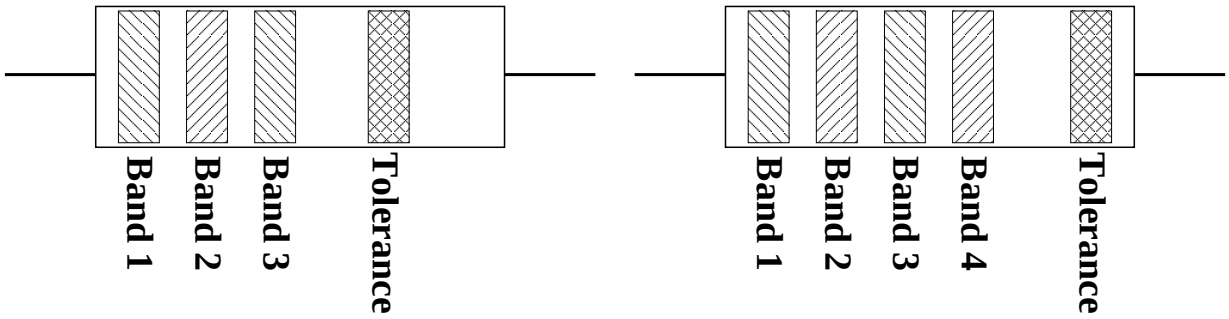

Figure A.1: The left-hand diagram shows a normal resistor with 4 bands. The method for reading these is shown in Figure A.2. The right-hand diagram shows the coding on precision resistors. Table A.3 gives the rules for reading these.

Color	1^{st}	2^{nd}	3^{rd}	4^{th}
Black	0	0	$\times 10^0$	
Brown	1	1	$\times 10^1$	
Red	2	2	$\times 10^2$	
Orange	3	3	$\times 10^3$	
Yellow	4	4	$\times 10^4$	
Green	5	5	$\times 10^5$	
Blue	6	6	$\times 10^6$	
Violet	7	7		
Grey	8	8		
White	9	9		
Gold			$\times 10^{-1}$	$\pm 5\%$
Silver			$\times 10^{-2}$	$\pm 10\%$
(None)				$\pm 20\%$

Table A.2: The meaning of each band in normal (4-band) resistor.

Color	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}
Black	0	0	0	$\times 10^0$	$\pm 1\%$
Brown	1	1	1	$\times 10^1$	$\pm .1\%$
Red	2	2	2	$\times 10^2$	$\pm .01\%$
Orange	3	3	3	$\times 10^3$	$\pm .001\%$
Yellow	4	4	4	$\times 10^4$	
Green	5	5	5	$\times 10^5$	
Blue	6	6	6	$\times 10^6$	
Violet	7	7	7		
Grey	8	8	8		
White	9	9	9		
Gold				$\times 10^{-1}$	
Silver				$\times 10^{-2}$	

Table A.3: The meaning of each band in a precision (5-band) resistor.

numeric part from the first three bands as 701. The fourth band indicates we multiply this by 10^2 and the fifth band says that we have 0.1% tolerance. Thus $R = 70.1\,k\Omega\,(\pm 0.1\%)$.

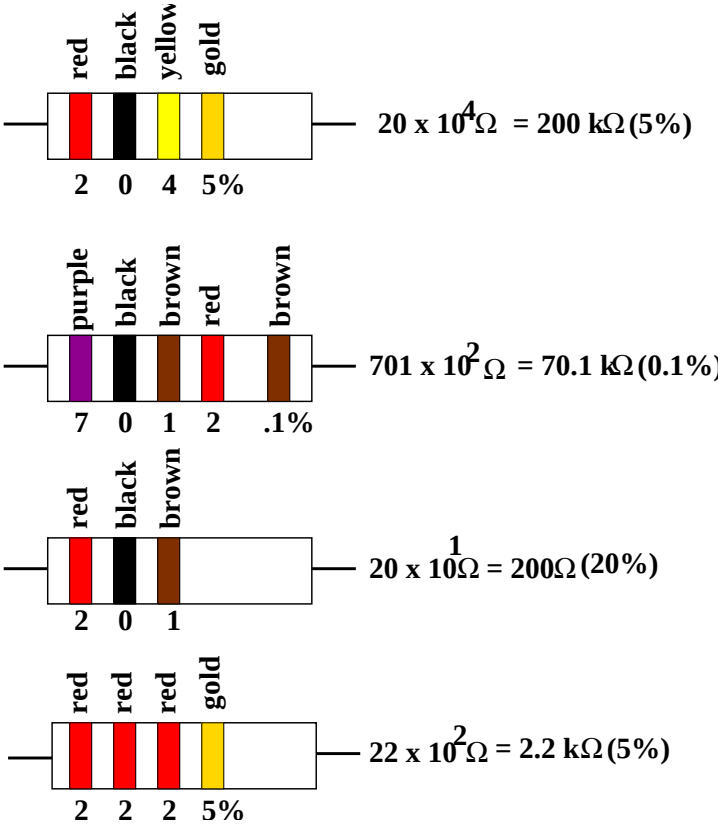

Figure A.2: Examples of reading 4-band and 5-band resistors. Note that in the third example, the color of the fourth band is "none".

A.2 Capacitors

There are probably more ways of labeling capacitors than one can count. In this section, we go over a number of the labeling schemes that one may encounter in building circuits. As with any component, measuring it is probably the most accurate way to determine its actual value. Electrolytic capacitors tend to come in cylindrical cans. These capacitors have a definite parity with the positive terminal (anode) being labeled in some fashion. A mark may be printed on the capacitor, or there may be a band or ring around one end of the capacitor as shown in Figure A.3. The capacitor will also normally have its capacitance printed on the side in μF; however, a $22\,\mu F$ capacitor (for example) may be labeled as **22 M**, where M is used to represent μF. In addition, the capacitor will have a voltage rating which indicates the maximum voltage at which the capacitor can operate. If one of these is hooked up backwards and subjected to a large voltage, one of the ends tends to remove itself from the the capacitor with a loud **bang**.

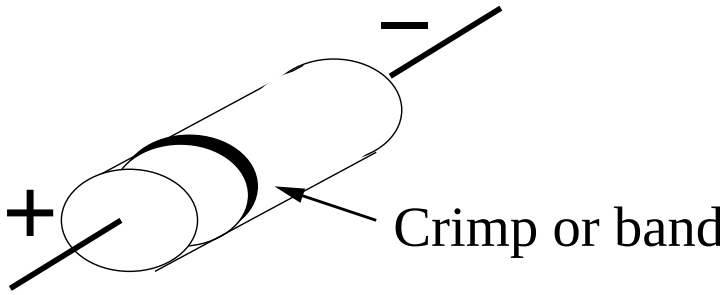

Figure A.3: An electrolytic capacitor. Capacitance is indicated in μF, while the end with the crimp or band is the positive end (anode) of the capacitor.

Another type of capacitor is the ceramic disk capacitor. The are typically flat discs. A couple of typical labeling schemes for these are shown in Figure A.4.

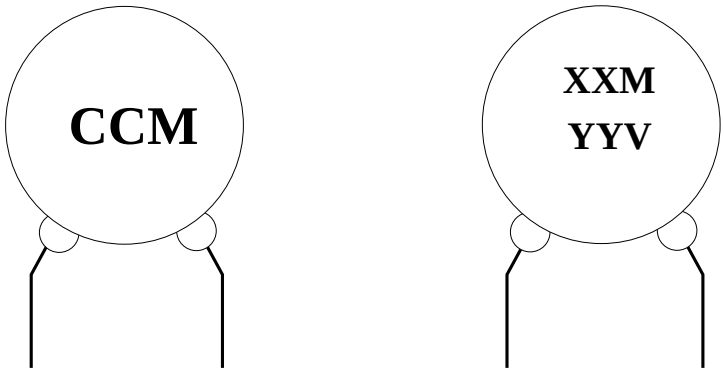

Figure A.4: The ceramic capacitor on the left is labeled with three numbers as shown, **CCM**. The value of the capacitor is given as $CC \times 10^M\,pF$. The label 103 translates to $10 \times 10^3\,pF$, while 501 translates to $50 \times 10^1\,pF$. The capacitor on the right has both a capacitance and a voltage on it. The **XXM** is $XX\,\mu F$, while the voltage is given as $YY\,V$.

You may also encounter tantalum electrolytic capacitors, seen in Figure A.5. Some of these are color-coded as shown in the figure, where Table A.4 shows how to interpret the colors on the capacitor.

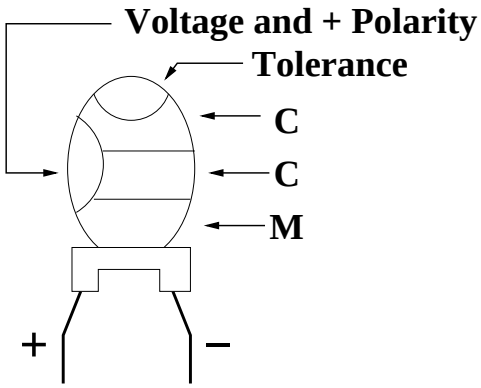

Figure A.5: A tantalum electrolytic capacitor. The capacitance can be written numerically on the capacitor in μF, or the color code in Table A.4 can be used. If the color code is used, the capacitance is in pF.

Color	Voltage	Value	Multiplier
Black	4	0	
Brown	6	1	
Red	10	2	
Orange	15	3	
Yellow	20	4	10^4
Green	25	5	10^5
Blue	35	6	10^6
Violet	50	7	10^7
Grey		8	
White		9	

Table A.4: The color code for the tantalum electrolytic capacitors shown in Figure A.5. The capacitance is given in pF.

A.3 Semiconductor Labels

Semiconductors are labeled with a combination of letters and numbers:

XNYYYY

with the letters and numbers having the following meaning.

X The number of semiconductor junctions. For a diode, this is one. For a bipolar transistor, this is 2.

N The device is a semiconductor.

YYYY The identification number (order of registration) of the device. This also may includes a suffix letter that can indicate matching devices (**M**), reverse polarity (**R**) and modifications (**A**,**B**,**C**,$\cdots$).

An example is the **2N2222** transistor, for which there is also a modified version, the **2N2222A**. A typical diode is the **1N4004**.

A.4 Diodes

In a diode, the pin which is connected to the p-type semiconductor layer is the anode, while that connected to the n-type layer is the cathode. With diodes, one of the most crucial pieces of information is which terminal is connected to the anode and which to the cathode. The diode is said to be forward biased when the anode is at higher potential than the cathode. It is reverse biased when the cathode is at higher potential. Figure A.6 shows a couple of typical diode packages on the left, while on the right are a couple of typical light-emitting diode (LED) packages. For the diode, the pointed end of the can, or the end with the stripe or band around it, indicates the cathode. In the LED, the shorter leg, or the side that has a flat spot on the base of the can, is the cathode. Zener diodes will probably

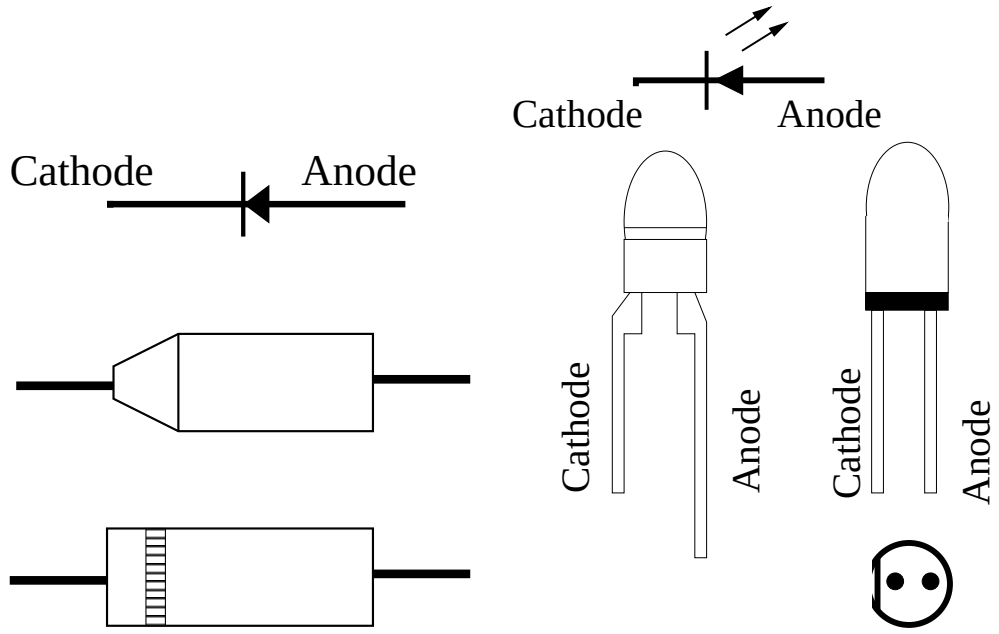

Figure A.6: Typical diode and light-emitting diode containers. Under forward biasing, the anode is at higher potential than the cathode.

have the break-down voltage printed on the can as well. "5.6" would mean that the diode breaks down when it is reverse biased with $5.6\,V$. There may also be colored bands on a diode. This information codes the diode type. Two examples of this are shown in Figure A.7 while Table A.5 shows what each color means. The last band is always the suffix letter, with black indicating that there is no suffix. The remaining bands, reading away from the cathode, give the diode identification number. To get the full number, add **1N** to the start of the code. In the examples, yellow-black-black-yellow-brown corresponds to 4004A. This is then a 1N4004A diode.

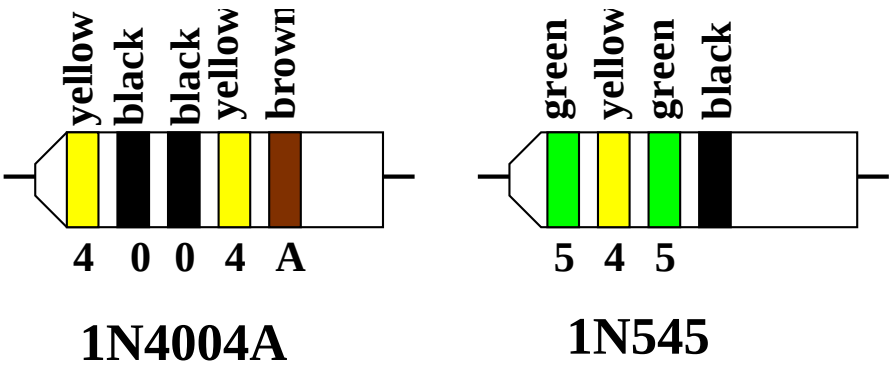

Figure A.7: Colored band labels for diodes.

Color	Digit	Suffix
Black	0	(none)
Brown	1	A
Red	2	B
Orange	3	C
Yellow	4	D
Green	5	E
Blue	6	F
Violet	7	G
Grey	8	H
White	9	J

Table A.5: The color codes used to identify diodes.

A.5 Transistors

The pin labels of typical bipolar transistors are shown in Figure A.8. While it is usually safest to double-check the pins on the transistor spec sheet, Figure A.8 does accurately describe a majority of these transistors. The *base* connection is in the middle. Most bipolar transistors will function, but not as well, if one reverses **E** and **C** in a circuit.

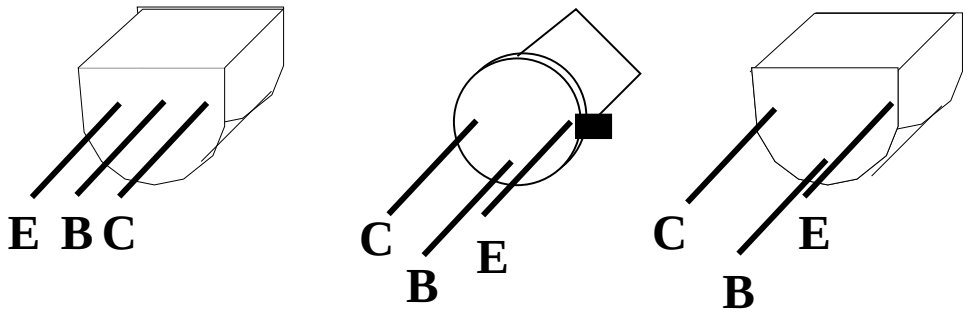

Figure A.8: Typical pinouts for bipolar transistors. The three pins are the emitter (**E**), the base (**B**), and the collector (**C**).

A.6 Integrated Circuits

Typical ICs come in 8- and 14-pin packages. Figure A.9 shows the pin numbering scheme on a 14-pin package. The key identifying mark is the *tab* shown at the center of the right-hand side of the chip. Looking at the top of the package with the tab on the right, pin 1 is above the tab and the highest-numbered pin (14) is below the tab.

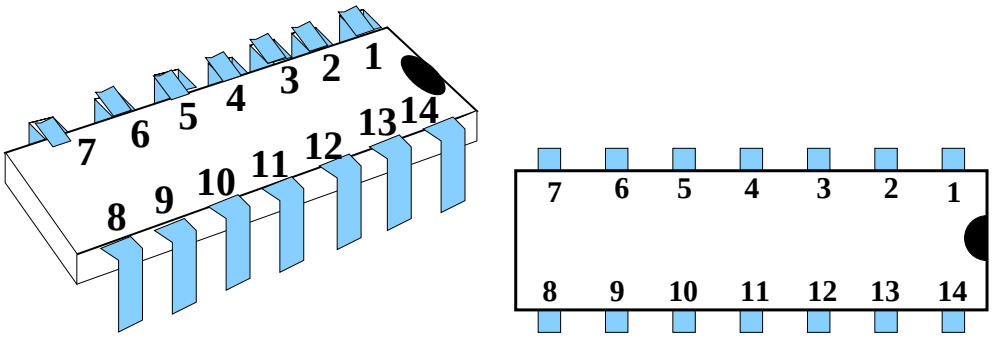

Figure A.9: The pin number scheme on a 14-pin IC package. Pin 1 is to the right of the tab, and pin 14 is to the left of the tab.

Appendix B

Answers to Selected Problems

B.1 Selected Problems in Chapter 1

(1.2) $v \simeq 8 \times 10^{-3} cm/s$.

(1.6) $1:1$, $1:2$, $1:9$.

(1.8) (a) $V_0/11$, (b) 0, (c) $V_0/(8R)$.

(1.10) $I_N = V_0/(8R)$, $R_N = (8/11)R$.

(1.16) $V_{th} = V_0 \frac{R_2}{R_1+R_2}$, $R_{th} = \frac{R_1 R_2}{R_1+R_2}$.

(1.20) $R_1 = 200\Omega$, $R_2 = 100\Omega$.

(1.24) (a) $V_{AB} = \frac{2}{5}V_o$, (b) $I_{AB} = \frac{2}{3}\frac{V_o}{R}$, (e) $R_L = \frac{3}{5}R$.

B.2 Selected Problems in Chapter 2

(2.2) $V_{RMS} = \frac{v_0}{\sqrt{3}}$.

(2.6) $(5+3j) = \sqrt{34}e^{.54j}$.

(2.12) (a) $i(t) = 0.224\, Ae^{j(\omega t - \pi/4)}$, (b) $< P >= 0.792\, W$.

(2.14) $Z_{eq} = R(-j)\frac{1-\omega^2 LC}{\omega C}$, $\frac{1}{\sqrt{LC}}$.

(2.16) $Z_{eq} = \frac{j\omega L}{1-\omega^2 LC}$.

(2.18) $\omega = \sqrt{\omega_{LC}^2 - \omega_{RC}^2}$.

(2.24) $s^{-\frac{1}{2}}$, $10\, dB/decade$.

(2.26) $f(t) = \frac{3}{4}\sin(2\pi t/T) - \frac{1}{4}\sin(6\pi t/T)$.