ISIM Book

Tuesday, February 9, 2021 7:43 PM

ISIM book

Introduction to Sensors, Instrumentation, and Measurement

Brian D. Storey

Olin College

Contents

1
Preface

These notes are meant to be a supplement the laboratory based course at Olin College; Introduction to Sensors, Instrumentation, and Measurement (ISIM). ISIM is a first-semester, first-year course at Olin College. The course at Olin has a substantial hands-on component of building circuits and conducting experiments. That component of the course will hopefully provide context, applications and allow the students to see some of the basic concepts at work. These notes will lean towards covering the fundamental concepts for circuit analysis and leave the fun stuff for the lab. While the notes will emphasize circuits, in the course we like to emphasize the broader themes which are making measurements, doing real experiments, and analyzing data.

This course will emphasize making electronic measurements. While not all experiments need to involve a circuit (optical measurements, for example), many measurements we make in science and engineering are through sensors, circuits, and computer-based data acquisition. In this course, we will deal only with the basic concepts and skills related to building and analyzing analog circuits. These notes are meant to be the bridge from the applications to the fundamentals that will enable you to not only understand the circuits you build but hopefully *design* your own.

There is no attempt here to write a comprehensive circuits introduction that is general enough for all students of engineering. Many interesting circuits topics are left out and many good textbooks exist that cover the topics I do, as well as those I ignore. As for topics I ignore, one of the great inventions of the 20^{th} century, the transistor, is totally neglected. In our modern world, digital circuits play a critical role and that entire realm is completely ignored in our course. The topics I cover herein are aimed towards the ones we cover in this specific course at Olin. If you are excited by circuits after taking this course, then there are many excellent follow on courses at Olin and throughout the world that cover more material.

The motivation for writing these notes here was mainly an attempt to provide an intuitive introduction to only the concepts we use in our course in a unified manner. The content is directed by the course at Olin and is written as such. If others outside of Olin find them useful, then I will be very pleased. I am happy to provide our laboratory materials to anyone who comes across these notes and wants to see the rest of the course material.

I have taught some aspects of this course consistently to first-year students since 2002, despite my background being in mechanical engineering. My long-standing research interests in the mechanics of fluids will show up here as I will make the hydraulic analogy continuously. I don't mean to bias things in this direction so heavily, but I think the analogy is a good one and one that students at the introductory level might have more intuition for.

The content of these notes and manner of explanations provided here have been greatly influenced by two real electrical engineers who I co-taught with many times over the years. They have taught me pretty much everything I know about circuits. They have patiently explained the ins and outs and all the subtleties that you can only learn through years of experience. I owe a great debt of gratitude to my colleagues, Gill Pratt and Brad Minch.

2
Resistance

2.1 Hydraulic analogy

Let's start with a simple system that you should have some physical intuition about. Imagine a pipe where one end is connected to tank of water and the other end is open to the atmosphere. A schematic is shown in Figure 5.1.

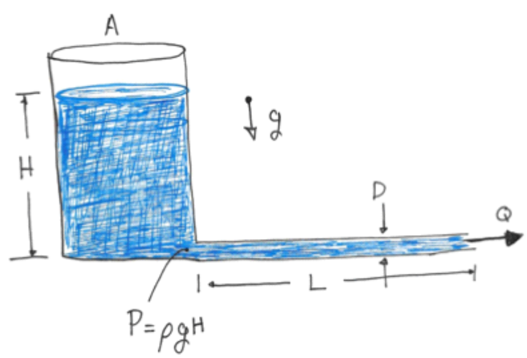

Figure 2.1 Schematic of a simple hydraulic system. A tank of water is drained by a pipe. We want to understand the relationship between the volumetric flow rate and the height of the water.

In this case, water will flow through the pipe, left to right, and out of the tank. The tank has a constant cross sectional area, A, and the water height is H. The pipe has diameter D and length L. The volumetric flow rate (measured in liters per second, for example) through the pipe is Q. We will consider a system sized such that the volumetric flow rate is small compared to the volume of the tank; the height of the water can be considered constant.

In this system it is the high pressure of the water at the bottom of the tank that drives the flow. High pressure results in high flow. Pressure, P, is force per unit area. At the bottom of the tank the pressure is proportional to the height of the water and is given by $P = \rho g H$ where ρ is the density of water (~ 1000 kg/m^3), g is 9.8 m/s^2, and H is the height of the water in meters. Note that the area of tank does not affect the pressure.

Aside: You may have seen the formula for pressure at the bottom of a tank previously. If you forget this formula it is easy to derive - see Figure 2.2. Imagine the bottom of the tank is a piston that must be held in place by a force F. Since the bottom of the tank is held still, the total force must balance the weight of the water above it; $F = mg$. The mass of the water is given as the density times the volume of water, $m = \rho A H$. Therefore the total force is $F = \rho A H g$. This total force divided by the area, $F/A = P = \rho g H$, is the pressure P at the bottom of the tank.

2.1.1 Resistance

A natural question to ask is what is the volumetric flow rate out of the tank? You would probably guess that the flow rate would depend upon the height of water in the tank - higher pressure should give higher flow. To answer the question experimentally, all you need is a beaker to measure volume and a stop watch. For a table-top system, we conducted an experiment in the kitchen with two different length of pipes. This experiment is simple enough that the data in Figure 2.3 were collected for this course by my children when they were 8 and 5 years old. They measured the time required to fill a 10 ml beaker with a stop watch for different heights of water in the tank. After every measurement we wrote the number down on a sheet of paper and were able to create the plot by hand.

We find that for the two experiments in Figure 2.3 the relationship

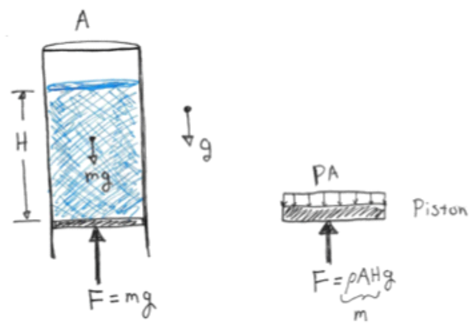

Figure 2.2 Schematic for derivation of hydrostatic pressure equation. If the bottom of the container is a sliding piston, then the force used to hold the piston in place is equivalent to the weight of the water. The pressure is this force distributed over the bottom of the container.

between water height and flow rate seems to be a line. With a little more care than a 5 year old is willing to tolerate the fit would be even better. The slope can be determined from the experimental data. In this experiment we find that the slope depends upon the length of the pipe; a longer pipe has a steeper slope. We can guess that the slope would also depend upon the pipe's diameter. A thinner pipe will have a lower flow rate for the same pressure, thus we would expect that the slope would be steeper for a thinner pipe.

If we repeated the same experiment for a variety of different pipes, we would find the same trends reliably occur across a wide range of experiments. Enough experiments would lead us to conclude that our system has a linear relationship between pressure and flow. Mathematically, this relationship could be described by,

$$P = QR.$$

The constant R is called the *resistance* and depends upon the pipe.

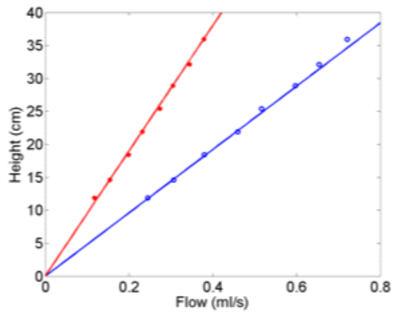

Figure 2.3 Experimental data for pipe of 1/16 inches (1.6 mm) in diameter and two different lengths. The data in red have a pipe with twice the length ($L = 1.4$ m) of the data in blue ($L = 0.7$ m). The points are experimental data and the solid line is an approximate linear fit to the data.

R is called the resistance since it is the constant that tells us about the resistance to flow through the pipe. When resistance is high, even a large pressure can only drive a small flow rate. For any given pipe, we could measure the resistance from the slope of the line in a similar way as Figure 2.3. In Figure 2.3 I found R by estimating values and changing them until the fit looked about right; it really can be that simple.

So why is the relationship between pressure and flow linear? Well, it turns out it need not be. The theory for water flow in a pipe would show us that the linear relationship should hold when the flow rate is relatively low. It is beyond the scope of our analogy to worry about this right now, but realize that a linear relationship between pressure and flow isn't a given or some fundamental law of nature. It is empirical. It is what we observe for our data. For our experimental system, we measured a linear relationship and we will assume that is the only case of interest for now.

Imagine the case when we have a pipe connecting two tanks of water as shown in Figure 2.4. Thinking about this example should lead us to realize that the pressure we use in our formula should be the pressure difference applied *across* the pipe, $\Delta P = P_{inlet} - P_{outlet}$. If the water height and thus the pressure of both tanks is equal, then there will be no flow through the pipe. To be a more precise we should write our relationship as,

$$\Delta P = QR,$$

where ΔP implies the change in pressure from the pipe's inlet to outlet. As far as the water flow is concerned, the overall pressure is unimportant - it is the difference from inlet to outlet.

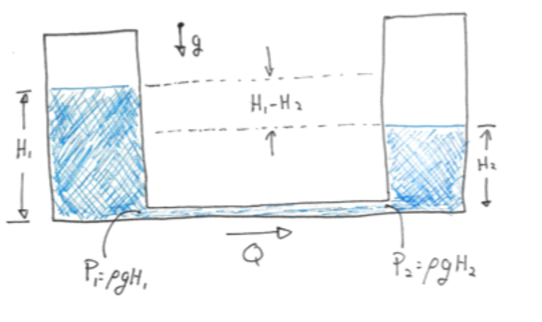

Figure 2.4 Two tanks connected by a pipe. Water flows through the pipe only when there is a pressure difference across the pipe - i.e. when the water heights are not equal in the two tanks.

So note a few very important things;

- Pressure is applied *across* the pipe.
- Water flows *through* the pipe.
- High pressure difference *across* means high flow *through*.
- The ratio of pressure to flow is the resistance, R.

It will be important to remember these points when we move the electrical domain as the analogy will directly carry over.

2.1.2 Resistors in series

In Figure 2.3, one set of data was taken for a pipe twice the length of the other. We can think about the longer pipe as taking two equal lengths of pipe and adding them in *series*. If we carefully look at the data, we find that the resistance of two pipes in series is exactly twice that of the single pipe.

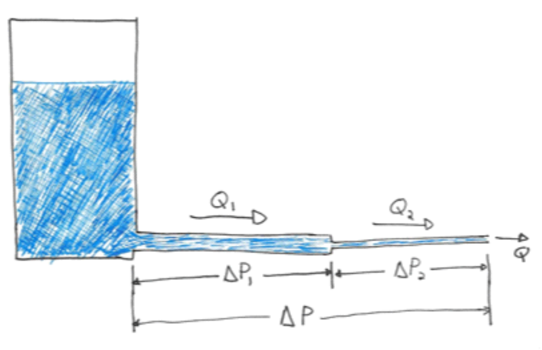

Figure 2.5 Schematic of a simple hydraulic system with two pipes in series.

We can understand this result by looking at a more general case of adding pipes of different resistance in series, shown schematically in Figure 2.5. Note here we have distinguished between the pressure drop across and the flow through each pipe individually. The pressure drop across each section would simply add to equal the total pressure applied across both sections of pipe,

$$\Delta P = \Delta P_1 + \Delta P_2.$$

Substituting the relationship for pressure and flow for each of the two individual pipes we have,

$$\Delta P = Q_1 R_1 + Q_2 R_2.$$

When two pipes are added in series, they must have the same flow rate through them - the water cannot change its volume and there is nowhere else for water to go. Therefore $Q_1 = Q_2 = Q$. There is only one flow rate in this setup. We can rewrite the overall pressure-flow relationship as

$$\Delta P = QR_1 + QR_2 = Q(R_1 + R_2) = QR.$$

Whenever we have two pipes in series, we simply add the resistances to get the total effective resistance to flow out of the tank; $R = R_1 + R_2$. The total pressure and total flow are related to the sum of the individual resistances.

The schematic also gives us some intuition about the extreme limits. Imagine one pipe was pretty fat and the other was really thin. It should make sense that the flow rate out of the system would be about the same as if it there were only the thin pipe. The effective resistance is dominated by the largest of the two values. The formula gives this. Imagine $R_2 = 100$ and $R_1 = 1$ (in whatever unit system we are working in). The effective resistance of the two in series is $R = 101$, a 1% change from just being R_2. If you have two resistances in series and the resistances are very different sizes, the effective resistance is close to that of the *largest* of the two resistors.

While we only considered 2 resistors here, we could generalize this expression and would find that if we added several pipes in series, the total resistance would equal the sum of the individual resistances;

$$R = R_1 + R_2 + R_3 ... \quad \text{Series resistance}$$

2.1.3 Resistors in parallel

Now imagine we take the same tank and put two pipes out as shown in Figure 2.6. In this case, the *total* flow out of the tank would be the sum the flow out of each independent pipe,

$$Q = Q_1 + Q_2.$$

Substituting the relationship for pressure and flow for each if the two pipes we have,

$$Q = \frac{\Delta P_1}{R_1} + \frac{\Delta P_2}{R_2}.$$

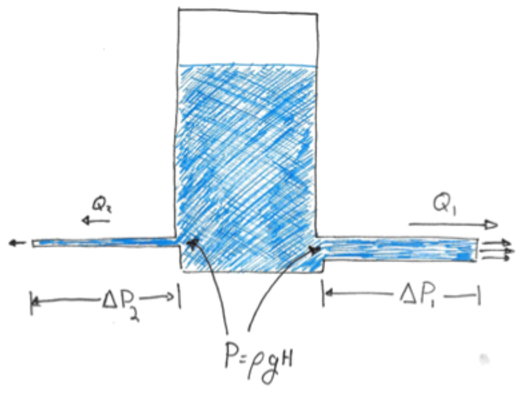

Figure 2.6 Schematic of a simple hydraulic system with two pipes in parallel.

In this case, the pressure applied across each of the two pipes is the same, thus

$$Q = \Delta P \left(\frac{1}{R_1} + \frac{1}{R_2} \right).$$

Rearranging this expression to get pressure on the left side of the equation,

$$\Delta P = Q \frac{1}{\frac{1}{R_1} + \frac{1}{R_2}} = QR,$$

where the total effective resistance of the two pipes is

$$R = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2}} = \frac{R_1 R_2}{R_1 + R_2}$$

We will define

$$R_1 || R_2 = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2}}$$

as shorthand notation for meaning R_1 and R_2 in *parallel.*

There are a few useful cases to think about. First, when $R_1 = R_2$ the effective resistance would become $R = R_1/2$. This result should make sense because if we take two equal pipes draining the tank at the same time, the flow rate out would double (or the overall resistance would be halved) from the case of a single pipe.

We can also rewrite the effective resistance expression as,

$$R = \frac{R_1}{1 + \frac{R_1}{R_2}}.$$

This form lets us see that if R_1 is tiny relative to R_2 (i.e. $R_1 << R_2$) then the effective resistance is just a bit lower than R_1. For example if $R_1 = 1$ and $R_2 = 100$; $R = 0.99$ a 1% error if we just forget about the large resistor. This limit should make physical sense because the thin high resistance pipe will drip while the fat low resistance pipe will really gush - so who cares about the thin pipe's overall contribution to the total flow. With resistors in parallel, if the two resistors are very different sizes then the effective resistance is close to the *smallest* of the two resistors.

The above result for effective resistance would easily generalize to more than two pipes draining the tank,

$$R = \frac{1}{\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} \cdots} \quad \text{Parallel resistance}$$

2.2 Circuits - electrical resistance

These concepts and equations carry over to the analysis of our first electrical circuit element, the resistor. A picture of a resistor of the style we will use in this course and the symbol used in circuit drawings is shown in Figure 2.7. Like a pipe, a resistor works equally as well which ever way it is oriented in a system. Resistors used in modern electronics are much smaller than the ones we use in lab, but they work the same way. The basic equation for the resistor is known as Ohm's law,

$$\Delta V = IR.$$

Here, ΔV is voltage *difference* measured in volts (V), I is current measured in amps (A), and R is the resistance in ohms (Ω).

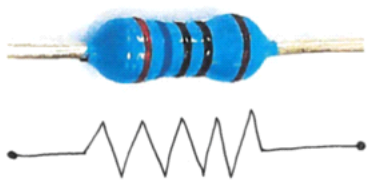

Figure 2.7 Picture of a resistor of the style we will use in this course. The diagram shows the schematic symbol used in circuit drawings.

Making the analogy to the hydraulic example, voltage is like pressure, current is like volumetric flow rate of water, and electrical resistance is like the pipe's resistance. Just like with pressure, it is the voltage *difference* across the resistor that we use in Ohms law. Just like volumetric flow rate, the current flows *through* the resistor.

Electrical current is measured in amperes or amps for short. The ampere is equivalent to a coulomb per second. A coulomb is the unit of electric charge and is equivalent to the charge on 6.24×10^{18} electrons. The current, I, is the amount of charge per unit time passing through. Just like the flow of water, current flows through a circuit in a conserved way. For any part or node in a circuit, the amount of current flowing in must equal the amount of current flowing out. The analogy with the water flow in the pipe is a good one. It is fine to think of current as the flow of electrons (even though the electrons are actually moving in the opposite direction since they are negatively charged!).

Just like resistance of a pipe can change depending on the diameter and length, the resistance of a electrical resistor depends on its size, material, and how it is made. Just like with a pipe, resistance of a wire increases linearly with the length and inversely with the cross section area. A long skinny wire has higher resistance than a short fat one. While resistors come in all shapes and sizes for different reasons, the physical form factor of resistors we use in lab will typically look like those in Figure 2.7. The different values of resistance depend upon the details of how the resistor was manufactured even though they may look the same from the outside. Resistors are designed to target a particular value. Resistance in practical circuits can span many orders of magnitude. In this course we will use resistors ranging from 10 Ω to 10,000,000 Ω, The range of resistors that one can purchase is much wider than this. We use the kilo and mega prefixes to denote the size; a 1,000 Ω resistor would be 1 kilo-ohm or kΩ and a 1,000,000 Ω resistor would be 1 mega-ohm or MΩ. In class when we are speaking, we will usually refer to these values a "1 K" and "1 Meg". In lab, we typically use 1 % resistors, meaning the manufactured value is specified within 1 percent. One can buy higher precision if you need it. The style resistor we use costs around 1 cent each and the small ones found in modern electronics are typically much less than this. Resistors are inexpensive components.

A resistor has a linear relationship between voltage and current and is just like we saw in the hydraulic analogy. While we can explain that the relationship is often linear by appealing to other physics that is beyond this course, it really is just an empirically observed relationship. Resistors that we purchase and will use in lab obey Ohm's law to a very high accuracy. They were designed to do so. Not all materials and resistive devices necessarily obey Ohm's law.

2.2.1 Resistors in series and parallel

The rules we derived for pipes in series and parallel work equally as well here. A circuit with two resistors connected to a constant voltage source such as a battery or power supply is shown in Figure 2.8. In the left figure the two resistors are in series and on the right they are in parallel. Just as with the hydraulic system, we want to come up with a relationship between the applied voltage drop, V, and the resulting total current, I, through the two resistors. There are two circuit symbols used in the schematic. One is the resistor that we discussed already. The symbol with the two lines, one longer than the other represents a voltage source. Think of this as a battery where the long side is the positive terminal of the battery and the short side is the negative terminal. The voltage across the battery is, V. If it were a real battery, V would equal the voltage written on the side of the battery such as 1.5 volts for a AA battery. Since the two terminals of the battery are connected across the resistors, this total voltage is applied across both resistors.

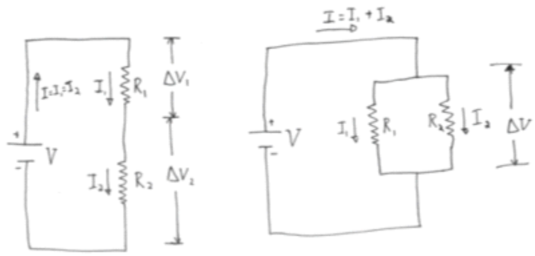

Figure 2.8 Schematic of two resistors connected to a constant voltage source, V. On the left the resistors are in series and on the right the resistors are in parallel.

When we have resistors in series, the total voltage drop across two resistors is set by the battery and is equal to the sum the voltage drops across each resistor individually,

$$V = \Delta V_1 + \Delta V_2.$$

The current flowing through the two resistors, just like the water flow, must be same. Since charge flows through the circuit in a conserved way, $I_1 = I_2 = I$. Using Ohm's law in our above expression for V we obtain,

$$V = I_1 R_1 + I_2 R_2 = I(R_1 + R_2),$$

which was identical to what we derived for the water flow. The battery sees the two individual resistors as an equivalent resistance which is simply the sum.

When we have resistors in parallel, the total current from the battery must equal to the sum of the current flowing through the two resistors,

$$I = I_1 + I_2.$$

Using Ohm's law,

$$I = \frac{\Delta V_1}{R_1} + \frac{\Delta V_2}{R_2}.$$

However, the voltage drop across each resistor is the same and is set by the battery, therefore,

$$I = V\left(\frac{1}{R_1} + \frac{1}{R_2}\right),$$

or

$$V = I\frac{1}{\frac{1}{R_1} + \frac{1}{R_2}}.$$

The battery sees the same current as though there were a resistor with an equivalent resistance of $R = 1/(1/R_1 + 1/R_2)$. The equivalent resistance of two resistors in parallel is exactly as we found previously in the hydraulic case. In both cases, the analysis of the hydraulic and electrical circuits are the same.

2.2.2 Ground and circuit schematics

In lab, we will often draw our schematics a little differently than the last section; more typical is Figure 2.9. The triangle symbol pointing downward means *ground*. In our lab ground will refer to the common point where we define the voltage to be zero. Note that in the right half of Figure 2.9 that the two ground symbols are *electrically connected*. Even though the grounds are not drawn connected by a wire, it is implied. We are allowed to let current flow into and out of our ground connection.

Since voltage differences matter, where we consider voltage to be zero has some arbitrariness to it. Think about potential energy. When we talk about the potential energy to what height do we ascribe "zero"? The height of the object above the floor I stand on? If I am on the second floor of the building do I use the ground level outside? Or maybe sea level? Voltage is the same way, we need to have an agreement on what we want zero to be, but we have some freedom in what the zero is. We will consider ground to just be where we have decided to set zero for voltage.

The reason that we call the zero voltage ground, is that a building's power system ground is set by a metal spike driven into the earth. The third prong on the electrical power system in the building is tied to the voltage of the actual earth's ground. If we are powering our circuit using a power supply that is plugged into building power, then typically the ground for the power supply would be connected to the actual ground of the earth. If we are running a circuit off a battery, then we are not connected to the earth. If you are an electrician, for safety reasons, you better get your circuit ground right. For us, whether the zero voltage is tied into the actual earth is generally not going to be relevant.

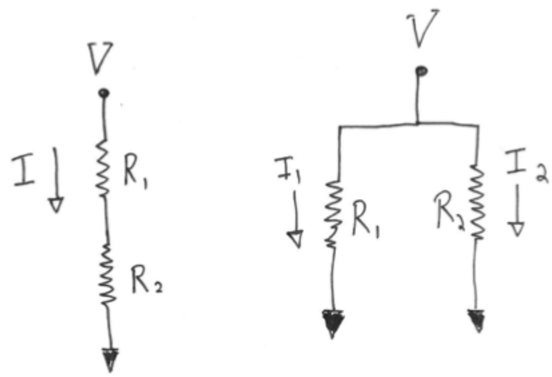

Figure 2.9 Schematic of circuit with battery and power supply. It is the same circuit but one case is drawn with a common ground.

2.2.3 Voltage divider

Let's look at a simple example of two resistors in series and ask what the voltage is between the two resistors (left of Figure 2.9). This is a really simple circuit, but one that we will find has a lot of application and shows up repeatedly in lab. The applied voltage (from a battery or power supply) is V_{in}. The voltage between the resistors is V_{out}. The voltage at ground is taken to be to zero. The circuit schematic is shown in Figure 2.9 left. The total current flowing through the two resistors is found by Ohm's law with the effective series resistance,

$$I = \frac{V_{\text{in}} - 0}{R_1 + R_2}.$$

Using Ohm's law for the second resistor,

$$V_{\text{out}} - 0 = IR_2.$$

Putting the previous two equations together we have,

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

If $R_2 >> R_1$ then $V_{\text{out}} \approx V_{\text{in}}$. If $R_2 << R_1$ then $V_{\text{out}} \to 0$. If $R_1 = R_2$ then $V_{\text{out}} = V_{\text{in}}/2$ The above circuit is known as a voltage divider because it, well, divides the input voltage V_{in}.

2.3 Kirchhoff's circuit laws

In the previous sections we derived laws for resistors in series and in parallel. Generalizing some of the ideas we have already used leads us to Kirchhoff's laws, which are attributed to Gustav Kirchhoff around 1845.

Kirchhoff's current law (KCL) states that the sum of the currents flowing into any circuit node must equal the sum of the currents flowing out. The law is based on conservation of charge. We already used this law when we analyzed resistors in series and parallel.

Kirchhoff's voltage law (KVL) states that the *directed* sum of the voltage differences around any closed loop in a network is zero. The only tricky part about KVL is keeping the signs straight (which is what we mean by the directed sum). For example, as we sum the voltages around a loop, we count a resistor voltage drop as positive if we are summing in the direction of the current. The resistor voltage drop is counted as negative if we are summing in a direction against the current.

When doing circuit analysis, we will we always invoke KCL. We did this when we first derived resistors in series and parallel. In simple circuits you will often find that you can get at the result you want without explicitly calling out KVL as we did in our early examples. KVL is very useful as it provides a systematic way of solving complex circuits. However, we will not need to call out its use explicitly when doing circuit analysis. When we draw circuits with the ends going into ground, we don't even see any loops. We implicitly invoke KVL when we analyze two resistors in series and say that the voltage drop across each resistor sums to the total applied voltage. We will not explicitly use KVL, I only introduce it here as it is a term you will often hear. We will save KVL for another class.

2.4 Measurement input impedance

In everything we have discussed thus far we have assumed that the measurement of the voltage is something we can do without disturbing the circuit itself. Generally, this is can be done (you will see how a little later) but we need to be careful. When making a measurement of voltage, the device you use will very often have a number somewhere on the specifications called the "input impedance". Impedance is a more general term than resistance that we will discuss in detail later. For now, consider impedance to be resistance. Note that units for impedance will be given in Ohms. For the device we currently use in our labs, the input impedance is 1MΩ; that would be a pretty typical value. What input impedance means is that inside the measurement device, there is essentially a resistor connecting the two measurement inputs. The resistor is part of the device and is there nothing you can do but account for it's presence.

Let's consider, as an example measuring the voltage in a real voltage divider with a device that has an input impedance in Figure 2.10. In the ideal scenario, when the measurement device does not draw any current then the voltage between the two resistors is $V_{\text{in}}/2$. However, when we make the measurement with a real device, current flows into the measurement device itself due to the internal resistance. To analyze the circuit, we can just take the resistor R and the input resistor R_I to be an equivalent resistance with the value of R and R_I in parallel. This equivalent resistor is now in series with the upper resistor. Using the voltage divider rule, the measured voltage is

$$V_{out} = V_{in}\frac{R||R_I}{R||R_I + R}.$$

When reading things like this, work out the result yourself on a piece of paper and make sure you agree with me. Using our formulas for resistors in parallel we have

$$\frac{V_{out}}{V_{in}} = \frac{\frac{1}{\frac{1}{R}+\frac{1}{R_I}}}{\frac{1}{\frac{1}{R}+\frac{1}{R_I}}+R} = \frac{\frac{1}{1+\frac{R}{R_I}}}{\frac{1}{1+\frac{R}{R_I}}+1} = \frac{1}{1+1+\frac{R}{R_I}} = \frac{1}{2+\frac{R}{R_I}}$$

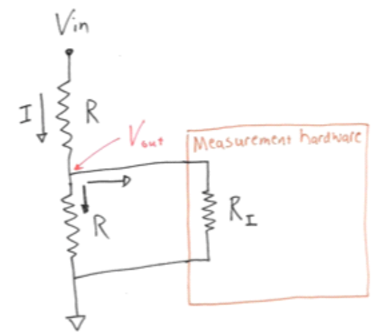

Figure 2.10 When measuring voltage, very often the measurement device will have a finite *input impedance*. This input impedance allows current to flow into the measurement device and thus perturbs the voltage measured in the circuit from what would be measured in an ideal scenario where the measurement device draws no current; i.e. infinite input impedance.

From this relation we see when the device input impedance is really large compared to the resistors in the circuit, the measured voltage is 1/2 of the input voltage as we would expect. When $R = R_I$ then

$$\frac{V_m}{V_{in}} = \frac{1}{2+1} = \frac{1}{3}$$

not what we would might have expected.

Example: In the lab, we can vary the values of the resistor in the voltage divider by keeping both resistors to have the same value as each other, but varying the value of R in Figure 2.10. If we set $V_{in} = 1$ V, then for an ideal measurement we would measure $V_{out} = 0.5$ V regardless the value of the resistors. For our device we obtain the results in Table 2.4. What is the value of the devices input impedance?

R	V measured
1 kΩ	0.4998 V
10 kΩ	0.4975 V
100 kΩ	0.4762 V
1 MΩ	0.3333 V
10 MΩ	0.0833 V

2.5 Application: scale using strain gauges

We will see the simple voltage divider pop up throughout various labs in our course. Just to get a preview of where we can use a voltage divider, let's consider a useful circuit for making a physical measurement. A strain gauge is a device used for measuring the mechanical strain, or deformation, of a material. All solids deform when they are subject to a force. If I take a bar of metal and pull on both ends the metal will elongate even if it seems imperceptible to us. This elongation is proportional to the force I use to pull on the bar, thus if I can measure the deformation I can infer the force. This principle can be used to make a scale. In fact if you open up a bathroom scale you will find strain gauges. As shown in Figure 2.11, the bar of metal has an initial length L when there is no weight applied and a length $L + \delta L$ when the weight is applied and the bar of material is loaded.

The strain gauge is nothing more than a thin, flat wire embedded into a plastic film. The strain gauge is super-glued to the material we want to measure the elongation of. When the bar is stretched, the wires in the strain gauge are stretched and become longer and skinnier, thus their electrical resistance increases. When the bar is squished, the wires in the strain gauge become shorter and fatter and the electrical resistance decreases. The nominal resistance of the unloaded strain gauge is R where δR is the change when deformed, thus the resistance is $R_s = R + \delta R$. It turns out that the change in length of the bar is directly related to the change in resistance through

$$\frac{\delta R}{R} = G\frac{\delta L}{L}$$

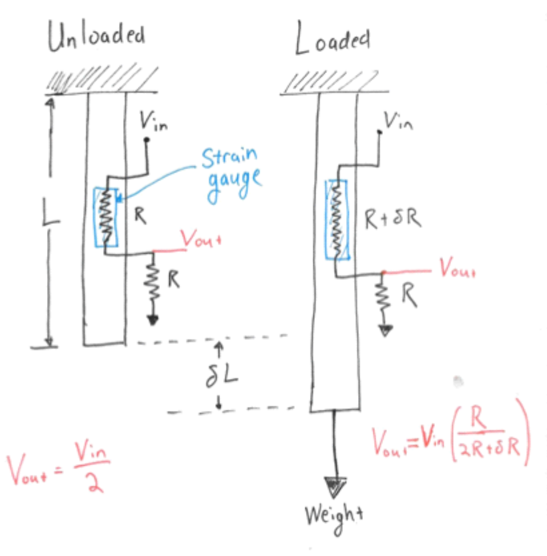

Figure 2.11 Concept of using a strain gauge to create a scale. A bar of material will elongate when weight is applied. This mechanical elongation will stretch the thin wires of the strain gauge, causing a small increase in electrical resistance.

where G is a constant for the strain gauge which depends in the width and thickness of the wires and is called the "gauge factor". Typically, this gauge factor is around 2 and is a measured empirical number that depends on the strain gauge design.

A simple circuit for measuring the electrical resistance change is the voltage divider shown in Figure 2.11. We set the lower resistor to be a fixed value equal to that of the strain gauge when it is unloaded, R. When the scale is not loaded we have two equal resistors in series the voltage that we measure is half the applied voltage V_{in}. When the weight is applied, the resistance of the strain gauge increases and thus the measured voltage decreases. The relationship between measured voltage comes directly from the voltage divider equation,

$$V_{\text{out}} = V_{\text{in}} \frac{R}{R_s + R} = V_{\text{in}} \frac{R}{R + \delta R + R} = V_{\text{in}} \frac{1}{2 + \delta R / R}$$

Rearranging such that we can compute the relative change in resistance from the measured voltage we have a simple formula,

$$\frac{\delta R}{R} = \frac{V_{\text{in}}}{V_{\text{out}}} - 2.$$

Now in a real application, the resistance and thus the voltage change between the loaded and unloaded configuration is tiny - maybe a few mV at best. There are many sources of error that complicate measuring small changes. The resistor in the circuit is not exactly the values they state (we usually use % 1 resistors). The applied voltage may not be exactly what you think it is - i.e. the power supply may say 5V but it is not exactly 5.000 volts. Thus, in practice the real circuit we will use in lab has a few features that need to go beyond what we describe here to account for these difficulties. However, the principle of of building a scale with strain gauges is simple and the same regardless of any additional real world complexities.

Example: For the circuit shown in Figure 2.11 we set V_{in} to 2.5 volts. The nominal resistance of the strain gauge and the resistor in the measurement circuit are both 100 Ohms. When the bar is loaded it stretches a little and the resistance of the strain gauge changes by 0.1 % to 100.1 Ohms. When the bar is unloaded we measure $V_{out} = 1.250$ V. The measured voltage is reduced by 0.625 mV (check this yourself).

In a real circuit the input voltage will not be precisely 2.5 V and the resistors will not precisely 100 Ohms. Imagine that the actual value of the input voltage was 2.513 volts, the fixed resistor was 101.5 Ohms and the nominal resistance of the strain gauge was 99.5 Ohms. Upon loading the strain gauge changes by 0.1 % to 99.6 Ohms. For this case compute the *change* in the measured voltage from the loaded to the unloaded configuration. You should find it to be very close to the same as with the ideal values. Therefore the *change* in voltage is not that sensitive to uncertainty in the component values.

3

Capacitance

3.1 Hydraulic analogy

As with the resistor, let's continue with the hydraulic analogy for our next electrical circuit component the capacitor. We made the analogy in the last chapter that the electrical resistor was like the pipe through which water flows. The capacitor is analogous to the tank that can store and discharge water. Just like the tank can store water, we will see that the capacitor can store electrical charge.

The volume of water *stored* in the tank, V, at any instant is related to the height, H, and the tank's cross sectional area, A, through $V = HA$. We previously saw that height is relate to pressure as $P = \rho g H$. Putting this together, the pressure at the bottom of the tank (which will be available to drive flow) is related to the volume of water stored in the tank through,

$$P = \rho g H = \rho g \frac{V}{A}.$$

Recall that ρ is the water's density and g is the acceleration due to gravity; constants that we cannot change in this example. We could rearrange this expression to read

$$V = \left(\frac{A}{\rho g}\right) P = CP$$

where we define $C = A/(\rho g)$ to be the *capacitance*. The word makes sense because the area A represents the capacity, or size, of the tank. If the tank has a large area it has a large capacity - the volume of water stored is large even when the pressure at the bottom of the tank is low.

The capacitance linearly relates the amount of water stored in the tank to the pressure the tank has available to drive flow.

Consider two tanks of different cross sectional area being filled with a faucet at constant flow as in Figure 3.1. The flow rate, Q, through the two faucets are set to be the same. It seems clear from the schematic that the volume of water stored in the tanks will be the same (since the flow rate in is the same), however the pressure at the bottom of the tanks will be quite different.

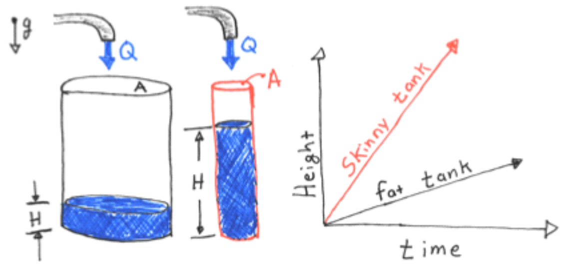

Figure 3.1 Two tanks of different cross sectional area being filled with faucets with the same flow rate. The volume of water in the two tanks is equal, while the height of the water depends on the cross sectional area of the tank.

The flow rate in or out of the tank need not be constant but can vary in time. Let's be explicit and use the notation $V(t)$, to emphasize that the volume is a function of time. If we *integrate* the volumetric flow rate, $Q(t)$, with respect to time we obtain the volume of water that has gone into (or out of) the tank. Therefore the volume of water in the tank at any instant is the initial value plus what has been added,

$$V(t) = V(0) + \int Q(t)dt.$$

We can express the same relationship between flow rate and volume by taking the derivative with respect to time,

$$\frac{dV(t)}{dt} = Q(t).$$

In these expressions the volumetric flow rate is positive for flow into the tank and negative for flow out.

3.2 Draining the tank

Imagine a tank of water after it is filled up connected to a pipe as shown in Figure 5.1. As we saw previously, the high pressure at the bottom of the tank drives water flow through the pipe from the left to the right. When we studied resistance, we empirically found a linear relationship between the height of the water in the tank (or the pressure at the bottom) and the flow rate out of the tank. Our empirically validated hydraulic version of Ohm's law for the pipe was,

$$\Delta P = RQ,$$

where ΔP was the pressure applied across the pipe and R is the hydraulic resistance.

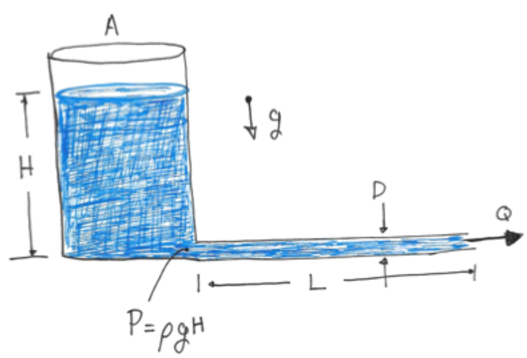

Figure 3.2 Schematic of a tank draining through a single tube.

In the chapter on resistors, we took the tank to be very large relative to the flow rate out such that the water height remained constant. Now we consider a case where the tank is not so large and as the water flows out, we can see the height in the tank drop. Imagine what will happen if we watch the tank drain. Initially, when the water level is high, the flow rate out is rapid and the level changes quickly. As the water level drops, so does the flow rate. As the flow rate through the tube drops the *rate* that the level drops slows down. As the level drops further, the *rate* that the water level drops slows even more.

This simple idea is very important. As the water flows out of the tank, the pressure applied *across* the pipe decreases. As the pressure decreases, so does the flow rate *through* the pipe. Experimental data were recorded by simply timing the water height, H, with a stopwatch and the result is shown in Figure 3.3.

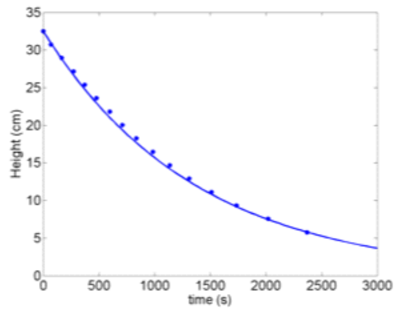

Figure 3.3 Behavior of tank draining problem. The points are experimental measurements using nothing more than a stopwatch and the solid curve is from the theoretical prediction.

Now let's explain our data quantitatively. For the draining problem in order to find the volume of water in the tank at any instant, let's start with the relationship between flow rate and the rate of change of volume stored,

$$\frac{dV(t)}{dt} = -Q(t).$$

The negative sign denotes that the volume decreases when the flow is out of the tank through the pipe.

Substituting the capacitance relation for the tank into our rate equation yields,

$$\frac{dV(t)}{dt} = -Q(t) \quad \text{or} \quad C\frac{dP(t)}{dt} = -Q(t).$$

Using Ohm's law for the relation between pressure across and flow through the pipe, $\Delta P(t) = Q(t)R$, we arrive at,

$$C\frac{dP(t)}{dt} = -\frac{\Delta P(t)}{R}.$$

If we take atmospheric pressure as our zero reference, then the pressure drop across the tube is equivalent to the pressure in the tank.

We can rewrite this differential equation as,

$$\frac{dP(t)}{dt} = -\frac{P(t)}{RC} = -\frac{P(t)}{\tau}.$$

From the units of the time derivative we note that the parameter, RC, must have units of time. We call this the time constant for the system, $\tau = RC$, which is related to the *product* of the tube resistance and the tank's capacity. Large resistance and capacitance result in a large time constant (i.e. a long time to drain). A higher resistance should decrease the *rate* that the tank height falls. If the tank has a large *capacity*, then the water level also falls slowly.

Since pressure and water height are linearly related, $P(t) = \rho g H(t)$, we can divide our equation by the constant ρg and get the same expression for the rate of change of the height of the water,

$$\frac{1}{\rho g}\frac{dP(t)}{dt} = -\frac{P(t)}{\tau \rho g} \quad \longrightarrow \frac{dH(t)}{dt} = -\frac{H(t)}{\tau}$$

This expression is useful so that we can compare to our experimental data in Figure 3.3 more easily.

You might be familiar with solving an expression like this; if not we will walk through the solution here. Note that since the right side of the equation is a currently unknown function of t, you just can't integrate directly. However, a simple trick allows for easy integration. In order to solve this differential equation, we separate the variables,

$$\frac{dH(t)}{H(t)} = -\frac{dt}{\tau},$$

and integrate both sides,

$$\ln(H(t)) = -\frac{t}{\tau} + B,$$

where B is a constant of integration. We find the constant by using the initial state of the system, namely we know the initial height, $H(t=0)$.

$$\ln(H_0) = -\frac{0}{\tau} + B.$$

Therefore using the now known value of B,

$$\ln(H(t)) = -\frac{t}{\tau} + \ln(H(0)).$$

Taking the exponential of both sides,

$$H(t) = \mathrm{e}^{-t/\tau + \ln(H_0)} = \mathrm{e}^{\ln(H_0)}\mathrm{e}^{-t/\tau} = H_0\mathrm{e}^{-t/\tau},$$

the final result is thus,

$$H(t) = H_0\mathrm{e}^{-t/\tau}.$$

The height decays exponentially. The value of τ controls how quickly the exponential falls. When τ is a large number the tank takes along time to drain and when it is a small number, it takes less time to drain.

In the previous experiment on resistance we measured R. Using this measured value of R and the cross sectional area of the tank our value of τ for the experimental measurements was ~ 1373 seconds. The predicted curve is shown as the solid curve in Figure 3.3 and we see quite excellent agreement.

The data and the model are a near perfect fit and the only empirical parameter is the tube resistance which was determined in a separate experiment. It seems that our equations are able to make accurate *quantitative* predictions.

3.3 Exponentials

The exponential function is a wonderful function. It shows up in a a number of systems from hydraulic, electrical, mechanical, chemical, thermal, biological, financial, and many more. Even though it is likely you have encountered this function before, it is important to make sure you understand a few properties.

Let's just plot the function,

$$y(t) = \mathrm{e}^{-t/\tau}.$$

The time constant, τ, determines the characteristic time for decay. In Figure 3.4 we show a time constant of $\tau = 1$, $\tau = 2$ and $\tau = 6$. If one draws a straight line from the initial value at $t = 0$ to the value of τ on the time axis, this straight line will have the same slope as the exponential at $t = 0$. This linear extrapolation is shown as the dashed line in Figure 3.4. Using this type of extrapolation of the slope is an excellent way to estimate the time constant from experimental data. Note that when $t = \tau$ the function has the value of $1/\mathrm{e} \approx 0.37$. The value of $y = 1/\mathrm{e}$ is shown as the dotted line in Figure 3.4.

Another trick is that instead of plotting the x axis of our figures as time in units of seconds, we could plot the time axis in units of τ. This change might sound odd, but is no different than changing our units from seconds to minutes to hours. When we plotted the exponential function in these time units all the curves would look the same. Therefore, any system with exponential behavior has the same basic function, it can just be stretched out in time or in its vertical scale.

3.4 Capacitors in circuits

After the resistor, the capacitor is the next electrical component we will introduce. Physically, a capacitor can be created from two parallel plates separated by a thin nonconducting layer such as a gap of air or a polymer film. The symbol on a circuit schematic is two parallel lines, representative of the capacitor's physical nature. A capacitor can act as an energy storage device where the capacitor stores the energy in the electrical field between the two plates. A detailed description of the mechanism behind the capacitor's behavior is probably best left for another day, we will just describe the result here. When you buy capacitors, they come in different shapes and sizes. A few common ones that we will use in lab are shown in Figure 3.5. There are many different styles, though they all have two wires for connection in the circuit. We will typically use non-polarized capacitors which mean that they work the same no matter what direction they are placed in the circuit. Some capacitors are polarized, meaning that the direction in the

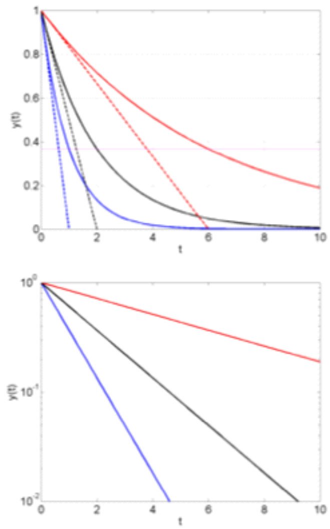

Figure 3.4 Effect of an increasing time constant on the resulting dynamics of exponential decay. Here we show results for $y = \mathrm{e}^{-t/\tau}$ for $\tau = 1,\ 2,\ 6$ as the solid lines. The dashed lines are for the initial linear extrapolation. On the upper figure the y-axis is linear and on the lower it is logarithmic.

circuit matters (i.e. the positive side marked on the capacitor needs to be connected to the positive terminal of the battery).

The capacitor has a simple relationship between the *charge* stored in the capacitor and the voltage across the capacitor as

$$Q(t) = C\Delta V(t).$$

Now I have to admit that the symbols here with the hydraulic analogy are starting to get confusing. I am using standard notation for the two different fields, but there is some overlap in the symbols. In the electrical world Q is charge measured in coulombs, C is the capacitance

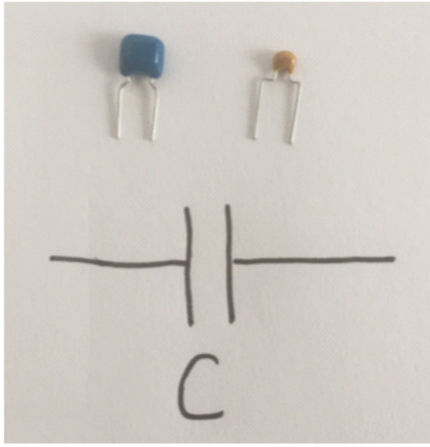

Figure 3.5 Picture of two styles of capacitor we will use along with the circuit schematic symbol. Capacitors come in many other shapes and sizes.

measured in farads, and ΔV is the voltage across the capacitor. I know this is confusing, but we will soon drop the hydraulic analogy and just stick with circuits. The more charge stored in the capacitor the higher the voltage. Note the analogy with the hydraulic world where the volume stored in the tank is proportional to the pressure. With the electrical capacitor the charge stored in the capacitor is proportional to the voltage. Note that I have not proven this capacitance expression in any way, I am just stating it.

The charge in the capacitor goes up or down as current flows and at any time the charge is given by

$$Q(t) = Q(0) + \int I(t)dt$$

Recall the current has units of amps, or coulombs per second. Thus when I integrate current with respect to time I get the total charge. This expression is just like with the tank where the volume of water stored in the tank goes up or down according to how much water flows. Here the charge stored in the capacitor goes up or down according to how much current flows.

The only breakdown of our analogy is that the current flows *through* the capacitor. Whatever current goes into one lead comes out of the other. This is not the way our hydraulic analogy worked; the tank has just one outlet. It is important to remember that analogies are just that. I find the hydraulic analogy useful at the beginning of circuits because we can literally see with our eyes the concepts related to flow, storage, and pressure.

Taking the time derivative of the charge expression we obtain,

$$\frac{dQ(t)}{dt} = I(t).$$

Using the relationship between charge and voltage we have

$$C\frac{dV(t)}{dt} = I(t).$$

The rate of change of the voltage across the capacitor is proportional to the current flowing through. In the hydraulic case we had the analogous expression where the rate of change of pressure in the tank was proportional the flow rate out of the tank.

The numerical value of C has units of farads. One farad is equal to one coulomb of charge accumulated when the capacitor is charged to one volt. Capacitors come in all sorts of physical packages and sizes. Like resistors, the values of the capacitance can vary by many orders of magnitude. The smallest capacitors we will use in this course are about 10 picofarads (pF). A picofarad is 10^{-12} farads. The largest we will use are about 100 microfarads (μF) or 0.0001 farads. A one farad capacitor is really large for our considerations.

For your reference, the capacitor values are stamped with a number for their value. For example, you might pick one up and it is stamped 104 or some other 3 digit code. The 104 means 10×10^4 pF. This is equivalent to $10 \times 10^4 \times 10^{-12}$ farads or 10^{-7} farads or 0.1 μF. A capacitor stamped 473 means 47×10^3 pF. Below in table 3.4, are some values we will often use,

Label	C value
102	1 nF
103	10 nF=0.01 μF
104	100 nF=0.1 μF
105	1 μF
106	10 μF

3.5 RC circuits

To complete our hydraulic analogy, we need to add the pipe to the tank in order to drain it - a resistor in our circuit world. Consider a resistor and capacitor arranged in series as shown in Figure 3.6. The voltage between the resistor and capacitor will be considered the output of the circuit, V_{out}. Initially, the voltage at the left side of the resistor will be set by some external source to V_{in}. Once the capacitor is charged up in this initial setup, the voltage at V_{out} is no longer changing and thus no current is flowing. When there is no current flow through the resistor $V_{\text{out}}(0) = V_{\text{in}}$. At $t > 0$ the switch is instantaneously closed (the valve is opened) and current can begin to flow. At the instant we close the switch V_{in} is set to zero.

At any instant, the current through the resistor is given by Ohm's law. Since Ohm's law requires that we use the voltage drop across the resistor, we have,

$$I_R(t) = \frac{V_{\text{in}} - V_{\text{out}}(t)}{R}.$$

At $t > 0$, $V_{\text{in}} = 0$ and therefore the current is given as

$$I_R(t) = \frac{-V_{\text{out}}(t)}{R},$$

where we explicitly remind ourselves for now that the current and output voltage are functions of time.

Using our law for the capacitor, the current through is given as,

$$C\frac{dV_{\text{out}}(t)}{dt} = I_C(t).$$

Since the components are in series, the current through the resistor, I_R,

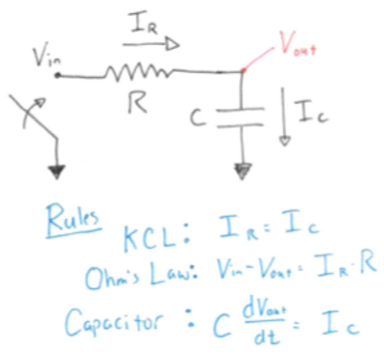

Figure 3.6 Circuit that we will analyze. Initially all voltages are set to 1 and then suddenly the input voltage, V_{in}, is pulled down to zero. A measurement is made at the voltage between the resistor and capacitor, V_{out}.

must equal that through the capacitor, I_C. Recall that this statement is invoking Kirchhoff's current law. We therefore have,

$$C\frac{dV_{\text{out}}(t)}{dt} = -\frac{V_{\text{out}}(t)}{R}.$$

The equation can be rewritten as

$$\frac{dV_{\text{out}}(t)}{dt} = -\frac{V_{\text{out}}(t)}{RC}.$$

We can notice from the units of the time derivative, that the product RC has units of time. If we define the time constant as $\tau = RC$, then you should notice the analogy with the draining tank.

$$\frac{dV_{\text{out}}(t)}{dt} = -\frac{V_{\text{out}}(t)}{\tau}.$$

The equation is exactly the same, only the symbols and physical meaning is completely changed. Just as with the hydraulic example, the time constant $\tau = RC$ increases with an increase in the both resistance and capacitance.

Even though we could use the hydraulic analogy to get the solution, let's work through the details again. Separating the variables of this equation,

$$\frac{dV_{\text{out}}}{V_{\text{out}}} = \frac{-dt}{\tau},$$

and integrating yields,

$$\ln(V_{\text{out}}) = \frac{-t}{\tau} + B,$$

where B is a constant of integration. Applying the initial condition $V_{\text{out}}(0) = 1$ yields,

$$\ln(1) = \frac{-0}{\tau} + B \quad \text{or} \quad B = 0.$$

Taking the exponential of both sides of the equation we have

$$V_{\text{out}} = \mathrm{e}^{-t/\tau}.$$

The solution to the problem is the same as in the hydraulic analogy. Experimental data are shown in Figure 3.7.

What is also interesting is that if we start the system where $V_{\text{out}} = V_{\text{in}} = 0$ initially and then raise the input voltage to 1 volt, the mathematical solution to this problem is the same, only the capacitor charges up, exponentially approaching 1 volt. Let's work this out in detail.

The equation in this case is the same as above, except that $V_{\text{in}} = 1$, thus

$$\frac{dV_{\text{out}}(t)}{dt} = \frac{1 - V_{\text{out}}(t)}{RC}.$$

Separating the variables of this equation (and using $\tau = RC$),

$$\frac{dV_{\text{out}}}{1 - V_{\text{out}}} = \frac{dt}{\tau},$$

and integrating yields,

$$-\ln(1 - V_{\text{out}}) = \frac{t}{\tau} + B,$$

where B is a constant of integration. Applying the initial condition

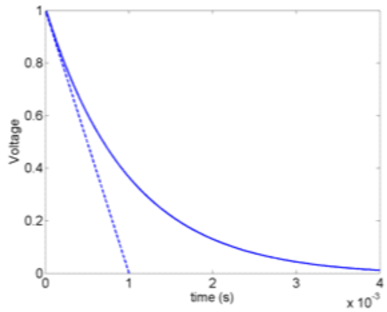

Figure 3.7 Measurement of voltage between resistor and capacitor in circuit of Figure 3.6. The value of the time constant was selected to be $RC = 0.001$ s. The dashed line shows the linear extrapolation using the 1 ms time constant. The measured behavior is extremely close to the predicted behavior.

$V_{\mathrm{out}}(t = 0) = 0$ yields,

$$\ln(1) = \frac{-0}{\tau} + B \quad \text{or} \quad B = 0.$$

Taking the exponential of both sides of the equation we have

$$1 - V_{\mathrm{out}} = \mathrm{e}^{-t/\tau}.$$

or

$$V_{\mathrm{out}} = 1 - \mathrm{e}^{-t/\tau}.$$

The solution for filling the capacitor is identical to the draining, only the solution is "flipped upside down".

3.6 Square wave driving: filling and draining the tank

Imagine the following experiment, shown schematically in Figure 3.8. Initially the tank is empty. The pressure at the inlet is suddenly set to a high value and held constant; the tank starts to fill. As the tank fills, the pressure drop across the pipe decreases and thus the rate of filling slows. The height of the water exponentially approaches an equilibrium state where the pressure across the pipe is zero.

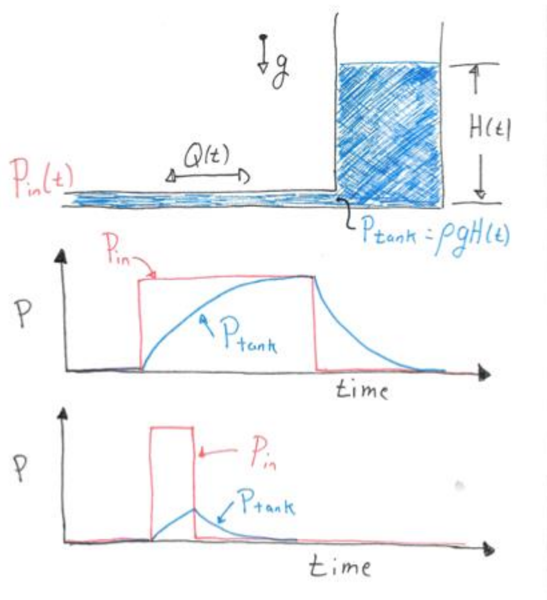

Figure 3.8 Hydraulic experiment where the pressure at the inlet of the pipe varies with time. The pressure starts at zero and then is held constant and high for some time, and then reduced back to zero.

Once the tank fills and reaches equilibrium, imagine we suddenly reduce the inlet pressure to zero. The tank will then start to drain following the exponential decay we saw previously. The behavior of the height of the water as a function of is shown schematically in Figure 3.8.

Now imagine that when we start filling the tank in the first step, we hold the pressure high for only a short amount of time before reducing it back to zero. The water height has insufficient time to reach equilibrium and only fills up a little before we pull the plug. It should seem intuitive that the shorter the amount of time we hold the pressure high, the less water goes into the tank.

How we define a "short" amount of time depends on the system. If the tank is very large and has a massive capacity, then we will need to hold the pressure high for a long time to fill the tank. Likewise, if the tube is very narrow and of high resistance, then the flow rate is so slow that regardless the size of the tank it may take a very long time to fill. The mathematical solution showed that RC (the product of resistance and capacitance) set the time constant for how long it takes the system to equilibrate. So to be precise, if the time we hold the pressure high is much greater than RC we can expect the tank to fill all the way and come to equilibrium before we drain the tank. If the pressure is held high for a time much shorter than RC, then we would expect that there is insufficient time for the water level to change much.

We can test this idea out quite simply in the lab. We can construct the circuit in Figure 3.6. With the lab hardware, we can easily set the voltage on the input V_{in}, to follow a periodic square wave. For the experiment we set $RC = 1$ ms (millisecond). The behavior at different periods of the driving is shown in Figure 3.9. Note that in this figure we show one full cycle of the square wave input, so for each plot the time axis gets shorter and shorter. The behavior we see is what we would expect from our physical argument of the hydraulic system above. When the voltage is held high for 5 ms we see the capacitor "fills up" to the top before the voltage is reversed and the capacitor drains. When the voltage is held high for 2.5 ms, we see that the capacitor almost, but doesn't quite fill up. As we continue changing the time, in the final experiment when we hold the voltage high for 0.05 ms before switching again, you see that the output voltage on the capacitor has insufficient time to change much.

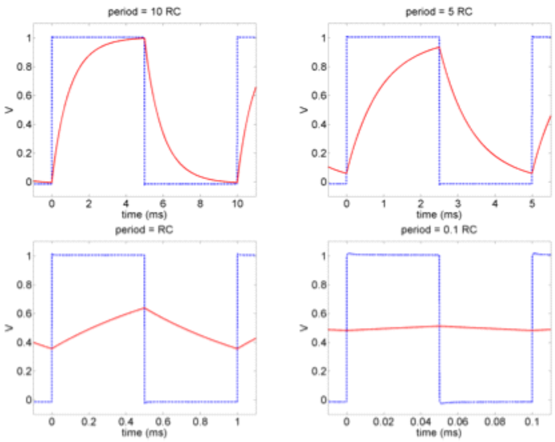

Figure 3.9 In this experiment $RC = 1$ ms. In each experiment, the switching time is varied. Right to left and then down, the period of the square wave is 10 ms, 5 ms, 1 ms and 0.1ms. The dashed blue line is $V_{\rm in}$ and the solid red line is $V_{\rm out}$. Note that the time axis change in each figure.

3.6.1 Transient effects

It is important to note here that the results shown in Figure 3.9 are for the *periodic steady state* of the system. Meaning I have been blinking the voltage up and down for some time, such that the memory of the initial charge on the capacitor has been erased. When you are looking at the periodic steady state, then the output voltage in our circuit becomes equal to the *average* value of $V_{\rm in}$.

We can understand the transient effect due to the initial charge on the capacitor (or level of water in the tank) by conducting the experiment. The result shown in Figure 3.10 starts with the capacitor charged to 1 volt and then we start the square wave at a high frequency, 5 times RC. In Figure 3.10 we see two time scales; one time scale is the natural RC time for the circuit and the other time scale is set by the driving. If we ignore the wiggles set by the driving, we observe that $V_{\rm out}$ decreases exponentially toward average value of V_{in}. It takes just a few RC times to settle into the periodic steady state. This idea of a periodic steady state will be an important one throughout the course.

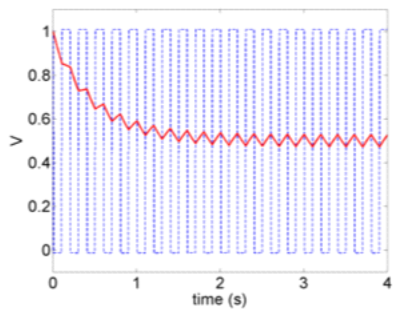

Figure 3.10 Transient effect when we start the capacitor in a fully charged state. Here in order to capture the transient, the RC time was set to 1 second. The period of the square wave driving is set to 5 times RC.

3.7 Pulse width modulation

The averaging effect seen here with the RC circuit is related to *pulse width modulation*, PWM. PWM is sometimes used in dimmer switches on lighting. Your eyes are unable to respond instantaneously to changes in light intensity. Like a capacitor or the tank of water, your eyes take time to respond. So if a light blinks very rapidly on and off, your eyes cannot perceive the blinking light. When the light is rapidly blinking, you perceive the intensity to be constant but dimmer than when the light is steadily on. The perceived intensity of light is controlled by varying the amount of time the light is on relative to the amount it is off in one cycle; known as the duty cycle. You see the average light intensity, not the individual flicker.

The same averaging effect can be used in control the heating element on an electric stove. If the heater is rapidly switched on and off, the temperature of the water in the pot has insufficient time to respond to such changes. It takes time to heat and cool. As with the other examples, the water in the pot only "perceives" the duty cycle of the heating element. The knob that controls high and low setting simply adjusts the duty cycle. These two examples are showing the same averaging effect we see in Figure 3.10.

4
RC circuits: Sinusoidal driving

In the last chapter with RC circuits, we found that the output of a simple circuit with a resistor and capacitor depended on the frequency of the square wave driving. We saw different behavior at high and low frequency. We also found the notion of what counts as "high" frequency in an application depends upon the period relative to the RC charging time of the circuit.

In this chapter we will start to analyze the case of sinusoidal variations rather than the square wave. The response of circuits to sine waves is extremely important to us and the ideas presented in this chapter emerge in many other physical, engineering, and natural systems. We will focus on the circuit applications here, but be aware that the ideas here are much broader than electronic circuits. If you never touch another circuit after this class, the concepts we start in this chapter will reappear in many other contexts. At this point in our discussion, I will drop the hydraulic analogy. It has served us well and can be a useful thinking tool for understanding RC circuits qualitatively. However, analogies can only be taken so far before they lose their effectiveness and we have reached that point.

The reason that sine wave forcing is so important to us is that, as you will learn later in your education, we like to represent arbitrary *signals* as a sum of sine waves of different frequencies. You are likely familiar with this idea at least intuitively. In an orchestra, each instrument gives out tones at different frequencies and then your ear hears all of them combined to make the music.

In measurements, there are often different frequencies which can be embedded in our measurement. Many times, we want to isolate particular frequencies in order to separate signal from noise. In lab, we will see this idea many times and thus you will see the seemingly abstract ideas and trigonometric relations presented in this chapter put to practice.

4.1 Individual resistor and capacitor: sine wave

Let's start with looking at an individual resistor and capacitor. We vary the voltage across the part sinusoidally and monitor the current through the part. See Figure 4.1. For both cases, let's set the amplitude of the voltage to 1 V and the vary the frequency ω, which here will be in units of radians per second. We can represent the input voltage as

$$V_{\text{in}}(t) = \sin(\omega t).$$

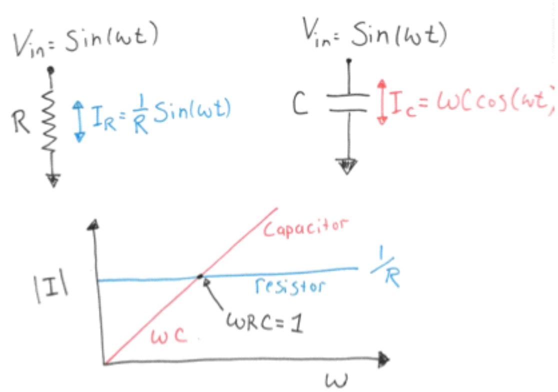

Figure 4.1 Sinusoidal driving of a resistor and capacitor individually. In the plot we show schematically the magnitude of the current as a function of frequency. Note that the current magnitude through the resistor and capacitor are equivalent when $\omega RC = 1$.

Aside: For analysis it is convenient to use the natural unit for frequency, radians per second. In lab we will use the more common unit for frequency - the Hertz (Hz). A frequency of 1 Hz is one complete cycle per second. The two units differ by a factor of 2π. If you forget about the units, you will start noticing that a lot of your results tend to be off by about a factor of 6.

For the case of the resistor, we have Ohm's law, $\Delta V = IR$, and thus the current is

$$I_R(t) = \frac{1}{R}\sin(\omega t).$$

The current is instantaneously related to the voltage and there is no effect of the frequency. The magnitude of the current (i.e the number in front of the sine term) is the inverse of the resistance, $|I_R| = 1/R$.

For the capacitor where $CdV/dt = I$,

$$I_C(t) = C\omega\cos(\omega t).$$

This result is more interesting. The amplitude of the current depends on the frequency. We therefore say that the magnitude of the current is $|I_C| = C\omega$. At very low frequency, the current is small and thus the capacitor effectively has very high "resistance". At very high frequency the current is high and the capacitor has very low "resistance".

I say resistance in quotes, because it is not a true resistor. Later we will generalize this idea, but let's hold off on that discussion for now. It is important that the *phase* of the current relative to the voltage is different in the resistor and the capacitor. In the resistor, voltage and current are perfectly in phase. Sine for voltage, sine for current. For the capacitor, voltage and current are 90 degrees out of phase. Sine for voltage, *cosine* for current. Phase refers to how much the two otherwise identical signals are shifted in their cycles relative to each other.

4.2 RC driven by a sine wave

Now consider the resistor and capacitor in series as in the last chapter and shown in Figure 4.2. Based on what we saw with the square wave voltage we would expect the following behavior. When the frequency is low then $V_{\text{out}} \approx V_{\text{in}}$; there is sufficient time for the capacitor to charge. We can understand this behavior using our voltage divider. In the last section we argued that when the frequency was low the capacitor acted like a very *large* resistor. In a voltage divider of the arrangement in Figure 4.2 the output voltage is close to the input voltage as shown schematically.

At high frequency, we know from the last chapter that the output voltage doesn't fluctuate much. We have argued that at high frequency the capacitor had very low "resistance". Using the voltage divider concept would imply that at high frequency the measured output voltage would be closer to ground than V_{in}

Figure 4.2 Sinusoidal driving of a resistor and capacitor in series. As seen in Figure 4.1, at low frequency the capacitor behaves as very large "resistor" and at high frequency the capacitor behaves like a very small resistor.

The aim so far is to provide a little intuition about how the basic RC circuit behaves. Let's explore the system experimentally. In Figure 4.3, we show the input and output voltages at different frequencies. This figure is analogous to the one we looked at in the previous chapter where the input was a square wave. We do in fact see the behavior described above. At low frequency the input and output voltages are the same. At high frequency the amplitude of the output voltage goes to zero.

As described in the previous chapter, everything here is assuming the sinusoidal steady state. We have waited until the system has come to equilibrium and is only responding to the sinusoidal driving. Looking at the experimental data we notice a few things. If I drive an RC circuit with a sine wave, then the output voltage

- is a sinusoidal wave of the same frequency.

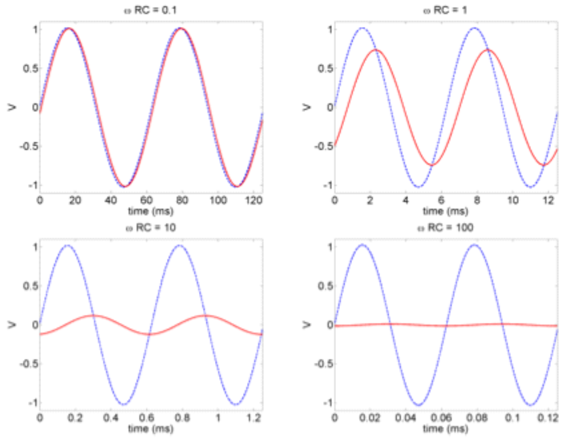

Figure 4.3 In this experiment $RC = 1$ ms. In each experiment, the frequency varied. Right to left and then down, the frequency increases by a factor of 10 each figure. The dashed blue line is V_{in} and the solid red line is V_{out}. Note that the time axis change in each figure. Note that even though the curves look computer generated, this is real experimental data.

- can have a different amplitude than the input voltage.
- can have a phase shift relative to the input.
- has an amplitude and phase shift that depend on the frequency.

Since the circuit is composed of *linear* components, then we could actually show more rigorously that the output voltage must be the same frequency as the input signal. By linear components we mean that the voltage and current are linearly related - not that the current is proportional to the square of the voltage or some other such law. The four bullet points above are very important and common to all linear systems in other physical domains. They provide a methodology for analyzing the circuits and compactly displaying the results as we will now discuss.

4.3 Analysis of the low-pass filter

Now let's consider this same problem of the resistor and capacitor in series mathematically. The resistor is driven with an external source as $V_{\text{in}} = \sin(\omega t)$ where the frequency ω would be in radians/second and is something we control. For simplicity I am assuming the amplitude of the input voltage is 1 volt, though this assumption does not cause a loss of generality.

Now let's use our circuit laws to analyze this case in detail. Since the two parts are in series, Kirchoff's current law would state that the current through the resistor must equal that through the capacitor. There is only one current, I. The current through the resistor is,

$$I = \frac{V_{\text{in}} - V_{\text{out}}}{R}.$$

and the current through the capacitor is,

$$I = C\frac{dV_{\text{out}}}{dt}.$$

Equating the terms we obtain,

$$RC\frac{dV_{\text{out}}}{dt} = V_{\text{in}} - V_{\text{out}}.$$

The resulting equation is exactly what we saw in the previous chapter.

Experimentally, when we checked the output of this circuit we found a sine wave of a certain amplitude and a phase relative to the input sine wave. This output sine wave repeats and repeats with no change. We are interested in figuring out the amplitude and phase once the output has reached (here is this term again so you don't forget) the *periodic steady state.*

We assume the generic sinusoidal forms of the input and output voltage and substitute those into our equations, $V_{\text{in}} = \sin(\omega t)$ and $V_{\text{out}} = A\sin(\omega t + \phi)$, we have

$$RC\omega A\cos(\omega t + \phi) = \sin(\omega t) - A\sin(\omega t + \phi).$$

We can now solve this expression for the amplitude and phase, A and ϕ. It might not be immediately obvious how to solve this equation. For now, let's just present and discuss the result and then follow up with a pretty simple (but somewhat detailed) derivation. The result of the

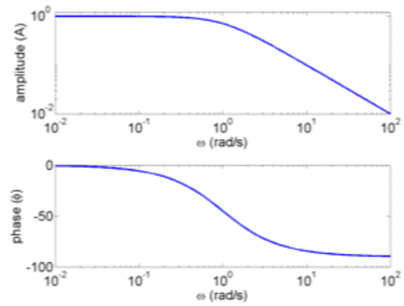

Figure 4.4 Low pass filter. Relative amplitude as a function of frequency and the phase angle in degrees as a function of frequency when $RC = 1$.

analysis is

$$A = \frac{1}{\sqrt{1 + R^2C^2\omega^2}},$$

$$\phi = \tan^{-1}(-RC\omega).$$

We can get a better handle on what is happening by generating a plot of A and ϕ as a function of frequency, shown in Figure 4.4.

This plot tells us everything we ever would hope to know about this circuit. It tells us how the output amplitude and phase vary as a function of frequency. Showing this type of plot is much more succinct than showing hundreds of snapshots as in Figure 4.3 - since each output sine wave is fully characterized by these two numbers, the amplitude and phase. This type of plot is referred to as a Bode plot. This circuit is called a low-pass filter as it allows low frequency signals to pass through unchanged. At low frequency the output amplitude is the same as the input ($A \to 1$) and the phase between output and input goes to zero. At high frequency the output amplitude decreases as the frequency decreases. Notice that by convention, I have used log-log and linear-log coordinates for the amplitude and phase. The use of log coordinates on the x-axis allows us to more clearly see what is happening over many orders of magnitude in frequency (4 in this case). The use of log coordinates on the y-axis in the amplitude plot allows us to see how the signal decreases over many orders of magnitude. In this case we can clearly see from the plot, that at high frequency, a factor of ten increase in frequency gives a factor of 10 decrease in amplitude. Power law functions show up as straight lines on log-log plots. While the exact functional behavior might not have been predictable, it is exactly what we have described up to now and consistent with our trusty hydraulic analogy.

Notice that the product RC must have units of time (seconds if we express R in ohms and C in farads). The term $RC\omega$ must have no units. We can also easily see the low and high frequency limits from our expressions for A and ϕ. When

- $\omega \to 0$, then $A \to 1$ and $\phi \to 0$
- $\omega \to \infty$, then $A \to 1/(RC\omega)$ and $\phi \to -90$ degrees.
- when $RC\omega = 1$ then $A = 1/\sqrt{2}$ and $\phi = 45$ degrees. This point will be the "knee" in the Bode plot.

The important frequency of the filter is the value of $\omega = 1/(RC)$ rad/s or $\omega = 1/(2\pi RC)$ Hz. This frequency is the knee of the amplitude plot and will be often referred to as the characteristic or cutoff frequency. It is called the cutoff frequency because above this value is where attenuation of the signal really sets in.

4.3.1 Low-pass derivation

Let's quickly derive the result for the low-pass filter. If you skip the details here for the first time reading you will be fine. I just want you to understand that the expressions for A and ϕ are something you can calculate and not something I pulled from thin air. Let's start with the expression from above,

$$RC\omega A\cos(\omega t + \phi) = \sin(\omega t) - A\sin(\omega t + \phi).$$

Using trigonometric identities often called the "product and sum" formulas (I always forget the exact from but I can easily look them up), we can obtain

$$RC\omega A\left(\cos(\omega t)\cos(\phi) - \sin(\omega t)\sin(\phi)\right) \quad =$$

$$\sin(\omega t) - A\left(\sin(\omega t)\cos(\phi) + \cos(\omega t)\sin(\phi)\right).$$

The only way this equation can be true at all times is that if all the terms with $\cos(\omega t)$ (blue terms) and $\sin(\omega t)$ (red terms) balance separately. Think about this and make sure you understand.

Grouping all the terms with cosine together we have

$$RC\omega A\cos(\omega t)\cos(\phi) = -A\cos(\omega t)\sin(\phi)$$

Canceling $A\cos(\omega t)$ from both sides gives,

$$RC\omega\cos(\phi) = -\sin(\phi)$$

or

$$\tan(\phi) = -RC\omega.$$

We already have derived the expression for the phase.

Now we can group the $\sin(\omega t)$ terms to obtain,

$$-RC\omega A\sin(\omega t)\sin(\phi) = -A\sin(\omega t)\cos(\phi) + \sin(\omega t).$$

Canceling the sin term and rearranging we obtain,

$$A = \frac{1}{\cos(\phi) - RC\omega\sin(\phi)}.$$

This expression gives the amplitude of the output divided by the amplitude of the input. It is a complicated looking function and hard to interpret since the amplitude depends upon the phase.

A better form for our answer uses the trigonometric identities to reduce the amplitude equation to a more explicit form. We are doing nothing more than re-arranging the equations into a simpler form. Namely we can use the trigonometric relations that state,

$$\cos(-\tan^{-1}(RC\omega)) = \frac{1}{\sqrt{1 + R^2C^2\omega^2}},$$

$$\sin(-\tan^{-1}(RC\omega)) = \frac{-RC\omega}{\sqrt{1 + R^2C^2\omega^2}}.$$

Again, these are trigonometric identities that I would not expect you remember but are easy to look up.

Substituting these trig identities into the amplitude equation provides,

$$A = \frac{1}{\frac{1}{\sqrt{1+R^2C^2\omega^2}} + \frac{R^2C^2\omega^2}{\sqrt{1+R^2C^2\omega^2}}} = \frac{1}{\sqrt{1+R^2C^2\omega^2}}$$

4.4 High-pass filter

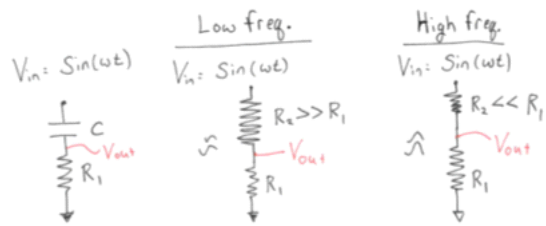

Figure 4.5 High-pass filter circuit arrangement for R and C. As seen in Figure 4.1, at low frequency the capacitor behaves as very large "resistor" and at high frequency the capacitor behaves like a very small resistor.

Now let's change the order of the parts as shown in Figure 4.5 and find out what happens. Let's see what using the idea that we can think of a capacitor acting as a high resistance at low frequency and low resistance and high frequency gives us. Using our voltage divider logic, at *low frequency*, the high capacitor "resistance" means that the output voltage should be close to ground. At *high* frequency then the low capacitor "resistance" means the measured output voltage should be close to the input. These two limits of the behavior are exactly the opposite as in the low-pass filter.

If we conduct the analysis as before and assume that $V_{\text{in}} = V\sin(\omega t)$ and $V_{\text{out}} = AV\sin(\omega t + \phi)$, when we do the analysis we obtain for the amplitude and phase,

$$A = \frac{RC\omega}{\sqrt{R^2C^2\omega^2 + 1}}$$

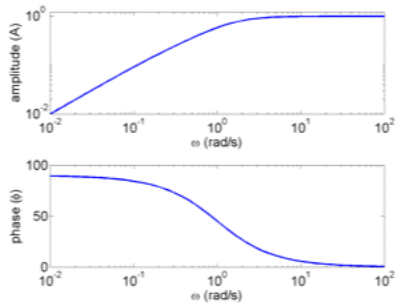

Figure 4.6 High pass filter. Relative amplitude as a function of frequency and the phase angle in degrees as a function of frequency when $RC = 1$. Note the logarithmic axis.

$$\phi = \tan^{-1}\left(\frac{1}{RC\omega}\right).$$

These functions are plotted as the Bode plot in Figure 4.6.

The behavior is opposite from before. Namely, when

- $\omega \to 0$, then $A \to RC\omega$ and $\phi \to 90$ degrees.
- $\omega \to \infty$, then $A \to 1$ and $\phi \to 0$.
- $RC\omega = 1$ then $A = 1/\sqrt{2}$ and $\phi = 45$ degrees. This point represents the knee in the Bode plot.

The filter is called a high-pass filter since it allows high frequency signals to pass through unmodified. Again, the frequency given by $\omega = 1/(2\pi RC)$ Hz is the important one and is called the cutoff or characteristic frequency. While of course the behavior is a continuous transition, but generally below the cutoff frequency and the signal is attenuated and above it, the signal is mostly unchanged.

4.4.1 High-pass derivation

The derivation for the high-pass proceeds exactly as the low-pass and I include the details here for your reference. Again, if you skip through the derivation the first time through you will be OK. I am including these details so you can go back and understand where all the results we use come from.

The current through the resistor is

$$I = \frac{V_{\text{out}}}{R}.$$

and the current through the capacitor is,

$$I = C\frac{d\,(V_{\text{in}} - V_{\text{out}})}{dt}.$$

Thus

$$RC\frac{d\,(V_{\text{in}} - V_{\text{out}})}{dt} = V_{\text{out}}.$$

Substituting in the generic sinusoidal form as before gives us,

$$RC\omega\,(V\cos(\omega t) - AV\cos(\omega t)\cos(\phi) + AV\sin(\omega t)\sin(\phi)) \quad = \quad AV\,(\sin(\omega t)\cos(\phi) + \cos(\omega t)\sin(\phi))$$

Grouping the $\sin(\omega t)$ terms (red) together gives,

$$\tan(\phi) = \frac{1}{RC\omega}.$$

and grouping the $\cos(\omega t)$ terms (blue) give,

$$A = \frac{RC\omega}{\sin(\phi) + RC\omega\cos(\phi)}.$$

At this point we could use these two expressions to create the plot of amplitude and phase as a function of frequency.

We could also rearrange the equation and get a different form by using our trig identities again,

$$\cos\left(\tan^{-1}\left(\frac{1}{RC\omega}\right)\right) = \frac{1}{\sqrt{1+\frac{1}{R^2C^2\omega^2}}} = \frac{RC\omega}{\sqrt{R^2C^2\omega^2+1}},$$

$$\sin\left(\tan^{-1}\left(\frac{1}{RC\omega}\right)\right) = \frac{\frac{1}{RC\omega}}{\sqrt{1+\frac{1}{R^2C^2\omega^2}}} = \frac{1}{\sqrt{R^2C^2\omega^2+1}},$$

Combining the expressions gives the amplitude as

$$A = \frac{RC\omega}{\sqrt{R^2C^2\omega^2+1}}.$$

4.5 Experimental Bode plots

As we will see in the lab, it is quite easy to create a Bode plot of a circuit experimentally. Using our hardware in the lab, it is easy to create a sine wave of voltage of a known amplitude and frequency to drive our circuits; i.e. we can set V_{in} to whatever we like. To make the Bode plot we just adjust the frequency of the driving and record the amplitude and phase of the output. We can imagine doing this if we refer back to Figure 4.3. In each case the amplitude of the input is one volt. We could work through data such as this and extract the amplitude of the output sine wave and the time deference between the peaks on the input and output. The time difference between the peaks is related to the phase by taking the ratio of the delay between the peaks to the total period. For the 4 sub-figures shown in Figure 4.3 where $RC = 1$ ms, we could extract the following 4 data points from the experimental data.

ω (Hz)	ωRC	Period (ms)	A	Δt (ms)	ϕ (degrees)
15.9	0.1	62.8	0.99	1	6
159	1	6.28	0.72	0.73	42
1,590	10	0.628	0.11	0.15	85
15,900	100	0.0628	0.099	0.0155	89

Completing such a table for a wide range of frequencies would allow us to capture the experimental Bode plot. It would take just a little work to record enough data to plot the experimental Bode plot. We would just plot the A and ϕ from the table as a function of frequency. Fortunately, the hardware we use in lab will allow us to create an experimental Bode plot with the click of a button. The software interface will conduct the experiment automatically. The software will set the frequency of the input, measure the amplitude and phase of the output, and then adjust the frequency to another value and repeat. The software controls the hardware to do what we would do by hand but in an automated way. While the automated generation of the experimental Bode plot makes our life easier, never lose sight of what is really happening inside.

One difference we find from what I have shown you thus far is that the hardware will default to plot the amplitude in decibels. The decibel, dB, in this case is defined as

$$\mathrm{dB} = 20\log_{10}\left(\frac{V_{\mathrm{out}}}{V_{\mathrm{in}}}\right)$$

For every factor of ten decrease of the output divided by the input we would see -20 dB drop. For every factor of ten *increase* in the output relative to the input, we would see 20 dB increase. Since the dB is already is taking the logarithm, we plot the dB on a *linear* scale. Plotting dB on a linear scale or plotting $V_{\mathrm{out}}/V_{\mathrm{in}}$ on a logarithmic scale are equivalent.

Truth be told, I am personally not a fan of dB. To me it has always seemed extraneous and we could accomplish the same thing more clearly by just generating plots on a log scale, as I have done in the previous figures. However, the dB has been around a long time and is in common use, so I have come to accept it. You just can't make the rest of the world do what you would want (perhaps this is the most useful advice in this entire book).

One very important feature of reading the Bode plot is how the amplitude decreases with frequency. Look back at the Bode plot for the low-pass filter. At high frequency, well above where the amplitude curve transitions from flat to dropping off, notice that the amplitude of the output decreases a factor of 10 for every factor of 10 in amplitude. This is called a *first order* filter. In dB the amplitude decreases by 20 dB for every factor of ten change in frequency for a first order filter. A *second order* low-pass filter would be where for every factor of 10 in frequency, the amplitude would drop by a factor of 10^2 or 100. A *third order* filter, the amplitude would drop by a factor of 10^3 or 1000 for every factor of 10 in frequency. In this course, we will very often pay the most attention to the amplitude part of the Bode plot.

4.6 Use of filters for noise reduction

Filters have a lot of uses and we will see them in many of our labs. I will let you see the behavior and uses for yourself rather than spending too much time explaining them in words. We will use filters often to separate signals from noise. All real measurements are subject to electrical noise which can come from many sources. One of the most

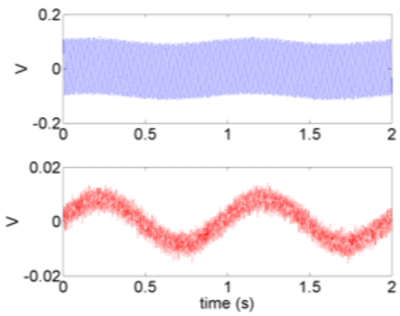

Figure 4.7 Example signal input (upper) and output (lower) to an RC low-pass filter with a cutoff frequency of 1.6 Hz. In this example the input is the sum of a 100 mV, 60 Hz sine wave and a 10 mV, 1 Hz sine wave.

common sources of noise we will encounter is noise at 60 Hz. The AC power in our buildings operate at 60 Hz. All the wires on your circuit act like antennae that pick up this strong 60 Hz signal. You will find a lot of other sources of noise at different frequencies coming from the lights (which in our room will provide a signal at 120 Hz), nearby radio stations (which are quite close to us), and other electronic equipment in the room. This noise will contaminate our final results.

Consider the controlled experiment shown in Figure 4.7. Here I took a 1 Hz sine wave at 10 mV and added it to a 100 mV sine wave at 60 Hz. In this example the 1Hz is the "true" signal and the 60 Hz is "noise". Notice that the noise is larger than the signal. In this example I have controlled the signal precisely for demonstration purposes, while in reality the noise would be at a wide range of frequencies. I created a single RC low-pass filter where the product of RC gives a frequency of 1.6 Hz. The upper figure is the input signal to the filter. While we can see a little evidence of the 1 Hz sine wave, we mostly see the 60 Hz noise. In the lower figure, we show the signal after the filter and can now clearly see the 1 Hz sine wave and the 60 Hz noise has been greatly attenuated. If these source of noise and the signal are more widely separated in frequency, then our single RC filter would really be able to isolate the signal with little evidence of the noise.

While the example above provides a clean demonstration, it is not what we will experience in practice. In practice the noise is not always so regular. The frequencies and magnitude may depend on where you are sitting and what else is going on in the room. Over the years I have learned that if I build a demo at home it will not always work when I bring into the lab to show the class. The overall electrical noise is much greater in the lab than in my house.

4.7 Filters in series

We will often find that the frequency of the signal and noise are close to each other and thus a single RC filter is not sufficient to remove the noise. We can combine filters in series to get different effects and provide stronger attenuation of particular frequencies. To be concrete, let's look at four filters shown schematically in Figure 4.8. If we put four RC filters in series, the voltage output is just the *product* of the output voltage of each filter, see Figure 4.8. You might be tempted to think that we could just analyze the filters individually and then multiply them to get the overall effect. This would be wonderful, but note that V_1 in the chain of four filters is not the same V_1 as you would get if you only had the first filter in isolation.

When building filters with passive analog components like resistors and capacitors, current always flows from one filter to the next and adding a second filter to the output of the first will change the behavior of the whole system. In principle, we need to analyze the whole circuit together - all four resistors and four capacitors simultaneosly. In practice, if we are careful we can try to minimize the current that flows from one filter to the next and thus the overall behavior is close to treating each filter independently. We can limit the interaction by increasing the resistance as we move down the filter chain. If we are careful we can approach the ideal case where filters in series are predicted by analyzing each filter independently and then multiplying the results all together.

To demonstrate this idea let's consider a specific case in detail. Figure 4.9 shows an example of two low pass filters in series. Each low pass

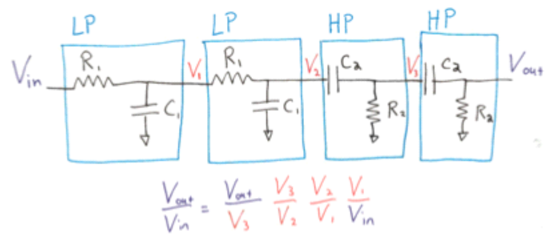

Figure 4.8 Schematic of two low-pass and two high pass in series. The final output is the product of each individual filter. The filters can be treated as isolated components only when the current flow from one to the other is very small.

filter with have the characteristic frequency (product of RC). When we build the circuit, we can vary R and C simultaneously to keep their *product* the same. When the second filter has a resistance 10 times the first, then approximately 1/10 of the current flowing through the first filter leaks to the next. The second filter only influences the first by about 10 percent and the two filters are nearly independent. However, if the resistor is smaller on the second filter than the first, then there is significant interaction between the two filters.

To compare the measured behavior to the expected (with the assumption of no coupling) we just take the product of individual low-pass filters with the same RC value,

$$\frac{V_{\text{out}}}{V_{\text{in}}} = \left(\frac{V_{\text{out}}}{V_1}\right)\left(\frac{V_1}{V_{\text{in}}}\right) = \left(\frac{1}{\sqrt{1+R^2C^2\omega^2}}\right)\left(\frac{1}{\sqrt{1+R^2C^2\omega^2}}\right),$$

$$\frac{V_{\text{out}}}{V_{\text{in}}} = \frac{1}{1+R^2C^2\omega^2}.$$

This predicted result for amplitude versus frequency is shown compared to the experimental data in Figure 4.9. When we compare this model to the experiment, we see excellent agreement when the resistance of the second filter in series is 10 times greater than the first (blue curve). When the second filter in series has a lower resistance (red curve), then the simple model does not work because we would have to analyze

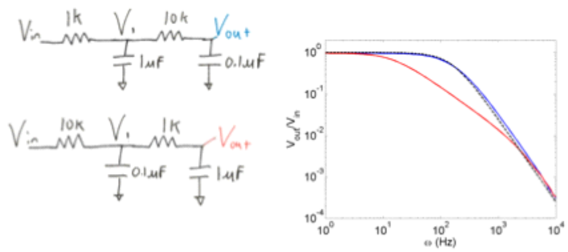

Figure 4.9 Example of 2 RC low-pass filters in series. The product of RC in both filters is the same. On the left are the circuits built and on the right are the experimental data. The dashed black curve is the theoretical result assuming the two filters are independent of each other. When the resistor in each filter in series increases as we move from left to right, we can assume the filters are independent of each other and the result is just the product of each filter individually.

all four components simultaneously. We can't just chain our analysis together.

We can abstract this idea to any number of filters in series that we like, just keeping in mind that the true behavior will depart from the simple product when there is significant current flow from one filter block to the next. If we can build the chain such that the resistor increases value significantly for each additional filter, then we can use our simple model that the final result is just the product of the blocks. In the next chapter, we will introduce the operational amplifier, which is a component that will allow us to solve this filter coupling issue. Once we have the op-amp, we can truly design independent functional circuit blocks.

4.8 Application example: EKG

An electrocardiogram (EKG or ECG) is a diagnostic test that looks for problems with the electrical activity of the heart. The test involves attaching several electrodes to the patient and monitoring small voltage "blips" when the heart fires. The voltages that we measure through skin contact are usually pretty small and other sources of electrical noise can be quite prominent. The EKG is an example where we want to make sure we are getting the true signal to the doctor.

When we build an EKG, we want to analyze the electrical activity during a heart beat cycle which repeats at about 1 Hz. While the heart beats around 1 Hz, there is some higher frequency ($\sim$ 10 Hz) characteristics in the signal that are important to the clinician. The point of an EKG is not to measure your heat rate (that can be done with your finger tips) but for the clinician to look closely at the electrical activity of the heart and diagnose any problems in the timing of the firing sequence. Retaining the true physiological signal and removing the artificial ones is really important. Since the true signal from the heart is relatively low voltage when measured by skin contact, we will find that what we measure a lot of 60 Hz noise. Your body acts like a big antennae and noise dominates the signal. A well designed filter is needed to squish out the noise which is at a known frequency, and try to leave the signal we want which is at a lower frequency.

In addition to high frequency noise, there are low frequency variations in the signal too. You will measure a voltage difference across your body which can be constant or slowing varying. This difference can depend on your body position and the contact of the electrodes with you so any motion will change the overall level of the voltage you measure, regardless of the fluctuations with each heartbeat. The signal will drift around as you move ever so slightly. This part of the signal also needs to be removed for the doctor. To extract the true EKG signal, we will need some combination both a low and high pass filter.

An example is shown in Figure 4.10. In the figure on the left, the blue signal is the voltage measured by two electrodes across my two wrists. The red signal is after passing through a set of filters. The amplitude Bode plot for the filter I designed is shown on the right. Here I used 3 low-pass and 3 high-pass filters in series to try to isolate frequencies in the range of 1 to 10 Hz. Now please realize that I am not qualified to read an EKG so I am not certain if the filter that I have designed is really that clinically useful. However, I am quite certain that the filtered signal is much closer to what I would want than the original blue one.

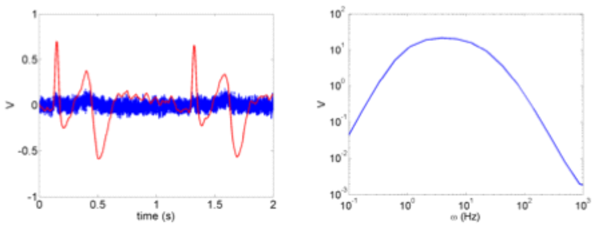

Figure 4.10 On the left, an example of EKG data before and after a filter to remove the noise and highlight the true signal. On the right we show the amplitude part of the Bode plot used to generate the EKG data. The filter was designed to try and keep frequencies between 1 and 10 Hz.

5
Operational amplifiers

Operational amplifiers (op-amps) are an important *active* circuit element. We say the op-amp is an active element since it is externally powered and ultimately it behaves the way it does because that is the way it was designed and built. Resistors and capacitors are passive elements. They both have two wires to connect to and *physics* sets the relationship between voltage across and current through. True, a resistor is engineered to have a specific value of resistance but ultimately Ohm's law (or a material's departure from it) can be explained through the theory of solid state physics. The same is true of the capacitor, while we can engineer for specific values of capacitance, the general relations for voltage and current are determined by the laws of electromagnetism.

It is important to keep in mind that the rules for op-amps are not derivable from physics. The rules are what they are because people manufactured and sold you a device to act this way. The device follows the rules it does because many years ago (dating back to around 1927) people were clever enough to begin to realize that a device that had the op-amp's behavior could be used to make a lot of interesting circuits. Over the decades, the op-amp has become an important element and is an essential *building block* in many of the electronic devices that surround you today. As we will see, the op-amp can be used to construct circuits which do various mathematical operations - hence where the name arrived from.

It is also important to realize that not all op-amps are created equal. They can have a range of specifications such as the voltage range that must power them, the maximum amount of current they can supply, and the maximum speed with which they can operate. The rules we will discuss in this chapter are generally followed any op-amp, but how closely a given op-amp follows the ideal behavior depends on the specific one used. We will largely ignore these practical issues in this reading and discuss only the idealized behavior. You will experience some of the practical realities and frustrations first hand in lab. It is important to keep in mind that we are just scratching the surface of op-amps in this course.

5.1 Schematic, inputs and outputs

The op-amp is a powered device that has two inputs and one output. Three example circuit schematics are shown in Figure 5.1. The two inputs on the left of the schematic are marked "+" and "-". The reason that the inputs are labeled with the plus and minus is that the device essentially takes the difference of the two inputs (more on this in the next section). There is a single output labeled on the right pointy part of the triangle. Thus the op-amp has two input voltages, V_+ and V_-, and one output voltage, V_{out}.

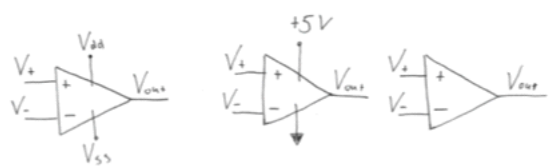

Figure 5.1 Op-amp schematics shown three ways, with a generalized power supply, a specific 5 volt supply, and with no power supply shown.

The device is powered, which is denoted by the vertical wires going to V_{dd} (the positive power supply voltage) and V_{ss} (the negative power supply voltage). The output voltage of the device is limited to be between these two power supply values. You will see different conventions with respect to drawing the power of the op-amp on schematics. Sometimes we will be explicit and draw the circuit with the generic power supply voltage marked (left). Usually when actually building a circuit for testing, we will label the power explicitly as in the center figure where we show a supply of 0 to 5 volts, as an example. Many times we will draw without the power supply as on the right of Figure 5.1. When analyzing op-amp circuits for general behavior the power supply values are not so important and it is often a good first step to ignore them. *Paying attention to the values on the power supply rails becomes critical when building real circuits in the lab.*

One of the very important aspects to understand is that op-amps are designed such that the inputs draw no current. The inputs can *measure* the input voltages but do so with essentially no current draw. You can think of the op-amp inputs as eyes that only observe the voltage and they do not disturb the circuit. In reality, there is a very small current flow into the inputs but it is really small. The exact amount depends on the op-amp design, but for the devices we will use, the current flowing into the op-amp inputs is less than the current leakage between adjacent *insulated* rows on your circuit prototyping board. We will always make use of this very good assumption, **the inputs draw no current**.

Since the op-amp is a powered device, the output can source or sink current. We can hook the output up to other things such as a resistor, a light bulb, a motor, or a speaker and the op-amp can push or pull current through the output. The maximum amount of current is set by the particular op-amp design and can be found on the specification sheet for the device.

Actual op-amp chips are packaged in many ways. In lab, we will use a particular package that can easily fit into your circuit prototyping board. There are pins that allow for easy connection of the input, output and supply voltages. In modern electronics, op-amps would typically be in much smaller packages than the ones we use such that they may be soldered directly to the surface of a printed circuit board. We will typically use a "quad" package meaning that in the single 14 pin chip, there are four independent op-amps. Each op-amp has 2 inputs and 1 output - which used up 12 of the pins. The other 2 pins on the chip are used for the power supply connection which is shared among the four op-amps. The pin diagram and a photo of a quad op-amp package is shown in Figure 5.2. You can see how the four individual op-amps are accessed through the 14 pins.

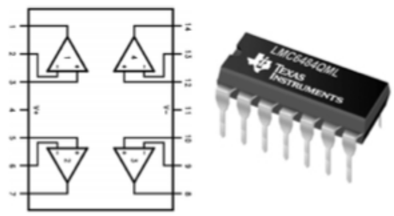

Figure 5.2 Op-amp pin diagram for a standard quad chip and a picture of the chip itself in the package we will use with the bread-board. Both images taken from the Texas Instruments datasheet. This op-amp is the particular model we use in lab at the time of this writing - different models would look the same and have the same connection based on convention.

5.2 Basic op-amp behavior

The simplest thing we can do in the lab is to power up the op-amp and adjust the input voltages while monitoring the output voltage to see what happens. An example of the time dependent, experimental behavior of the op-amp is shown in Figure 5.3. In this experiment, we hold V_- fixed and vary V_+ as a function of time using a triangle wave, as shown in the schematic. One particular time series is shown. From this data we note that V_{out} has two states, either 5 or 0 volts. We find that when $V_+ > V_-$ then $V_{\text{out}} = 5$ V and when $V_+ < V_-$ then $V_{\text{out}} = 0$. At this point the op-amp looks like some kind of switch to find out if V_+ is greater than V_-.

We can generalize the behavior of Figure 5.3 when we plot the output voltage as a function of the *difference* of the input voltages, where time is now parametric, Figure 5.4. When reading this graph note that since every cycle is the same, in time we move back and forth along this curve. In this figure, I am showing you 10 different experiments where I have changed the fixed value of V_- or some aspect of the time varying V_+. When plotted in this way, the behavior for each experiment collapses to this single curve. You can't even tell I conducted 10 experiments - the

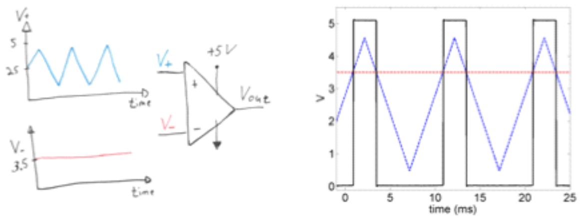

Figure 5.3 Voltage as a function of time. The negative op-amp input is held fixed (red dashed), the positive input is varied as triangle wave (blue dashed) and the output is monitored (black solid). This data was taken for the LMC6484 op-amp powered between 0 and 5 volts.

behavior is always the same. If we conducted even more experiments, we would see this behavior is very consistent.

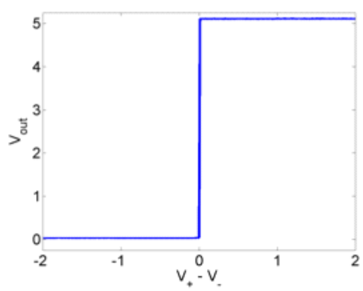

Figure 5.4 Plot of output voltage versus the difference between the positive and negative inputs. This plot shows ten experimental runs where the V_- is changed to different values. All experiments collapse to a single behavior.

This behavior that we see experimentally is the behavior the op-amp was designed to have. If we investigate the region right around where the $\Delta V \approx 0$ we see a very sharp transition for the output voltage. The smallest difference between the input voltages and the output changes state. It is important that the change of state of the output voltage is *continuous*, but happens over a very very narrow range of $\Delta V = V_+ - V_-$.

We can draw some conclusions from Figure 5.4, which are

- If we observe $V_{\text{out}} = 5$ V, then $V_+ > V_-$.
- If we observe $V_{\text{out}} = 0$ V, then $V_+ < V_-$.
- If we observe $0 < V_{\text{out}} < 5$ V, then $V_+ \approx V_-$.

These conclusions here are written with our op-amp powered from 0 to 5 volts. For different power supply ranges, we would find the same behavior. Note that since the op-amp output can't exceed the power supply, how closely the op-amp reaches the power supply limits is a function of the op-amp's design.

The third bullet item on our list is the most important one for analyzing circuits in this course. If the op-amp circuit is doing anything interesting, i.e. not just outputting a constant voltage which corresponds to either of the power supply limits, then the **input voltages are equal**.

Note that we are ignoring time in our description here. If we conduct this experiment paying attention to time, we would find that the op-amp does its thing very rapidly and that the behavior we described here holds up until the rate of change of V_+ becomes quite high. We will discuss dynamics in more detail later.

5.3 Feedback

Op-amps are usually used with *feedback*. By feedback we mean that the output is in some form or fashion connected back to the input. Typically we will use *negative feedback*, meaning there will be some path comprising wires, resistors, and capacitors which ties the *output* of the op-amp back to the *negative input* of the op-amp. Feedback is used to control the output voltage and hold it somewhere in between the power supply rails.

Feedback is a concept that is ubiquitous in science and engineering. Feedback is used in all control systems. A simple example of feedback in a control system is the thermostat in your house. In the winter in Boston, when my house is too cold, the heat kicks in to warm it up. When the room is too hot, the heat turns off. In this case the measured temperature of the room feeds back through the thermostat to change the state to the system (whether the heater is on) in order to regulate the temperature. Feedback is used in man-made and natural systems to control a system and hold it as steady as possible. Your body uses feedback to regulate your body temperature.

My first home in Boston did not have automatic feedback in the heating system. The building was old and we had a steam boiler in the basement which pushed hot steam through the pipes and radiators inside the home. There was no thermostat, just manual valves you could adjust to make the radiators flow more or less steam through them. To get the air temperature inside comfortable, you would need to get the valves in the exact right position so the steam was flowing at just the right rate. Everything was good as long as the weather didn't change. However, once the weather would suddenly become much colder outside then the valves would be in the wrong position and the house would be too cold. With no automated feedback it was very hard to keep the temperature inside comfortable in light of disturbances in the weather.

In op-amps, feedback *from* the output *to* the negative input can *control* the op-amp's output voltage. Without negative feedback the op-amp can't do anything but output a high or low voltage. In principle I could sit there and try to adjust the input voltages by hand (just like my old radiator valves) and try to hold the input voltage exactly right to give the output I want. But it would be a fruitless exercise. For the op-amp to do anything other than slam to the maximum or minimum voltage, we need feedback. In popular culture, the words "negative feedback" sound bad and we all like to receive positive feedback. In control systems, it is reversed. Negative feedback is good. Things can be stable with negative feedback - while positive feedback causes things to spin out of control.

The simplest example of feedback with op-amps is the circuit shown in Figure 5.5. Here we just use a wire to connect the output voltage to the negative input. The experimental data for this circuit is shown on the right in Figure 5.5. Here V_+ is controlled as a sine wave and the output voltage is monitored. The figure shows both V_+ and V_{out} as a function of time. You see only one curve because they are right on top of each other. Regardless what value or function we set for V_+, we see that V_{out} is a direct copy. You can confirm this behavior for yourself in the lab.

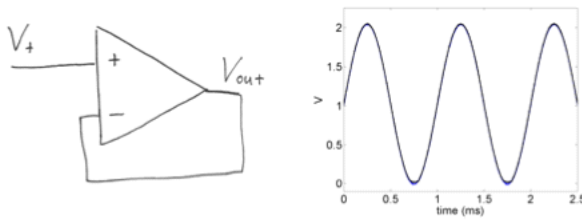

Figure 5.5 Follower circuit schematic and example experimental output. In the experimental data on the right the input at V_+ is set to a 1 volt sine wave at 1 kHz and centered at 1 volt. The output is an exact copy and we cannot see that two curves are plotted in this figure.

If we go back and look at the last section on open loop (i.e. no feedback) behavior, we found that for the output to be anything between the power supply rails, then $V_+ = V_-$. Since we observe that $0 < V_{\text{out}} < 5$, then we expect these two input voltages to be equal. Since there is a wire connecting the negative input to the output, then $V_{\text{out}} = V_-$. There is just one voltage on that wire. With this simple use of negative feedback, we now see that output following the input is consistent with the conclusions we drew from the basic op-amp behavior with no feedback. This circuit is often called the "follower", because the output follows the input.

5.4 Why the follower?

You might ask, who cares about a circuit that follows a voltage. How is this different than a wire? What a stupid circuit! In order to understand the importance of the follower, you must remember that the current going into the op-amp input is zero. Note that we now have a device that can measure the voltage with essentially no current draw, and then replicate that voltage out of a device which is allowed to *source* or *sink* current. This is actually very important feature which I hope you will come to appreciate.

As an example of why the following feature is important, recall in Chapter 2 on resistors when we discussed that most practical measurement devices have an input resistance (impedance). Recall what happens in the simple example of measuring the voltage between two resistors when the device's internal resistance is the same as the two resistors. The same amount of current flows into the measuring device as through the lower resistor in the divider. We measured 1/3 the applied voltage instead of the expected 1/2 if the device could make the measurement with no current draw. The act of measurement changes the expected voltage.

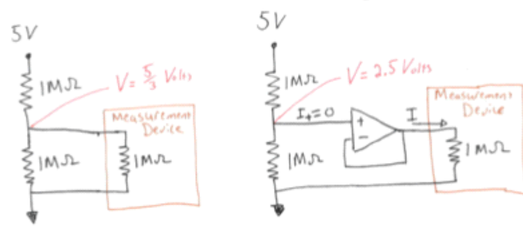

Figure 5.6 Example to show the utility of the simple follower circuit. The follower completely decouples the measurement from the circuit such that the output passes through the follower with no disturbance since no current goes into the op-amp input.

With an op-amp follower (on the right side of Figure 5.6) the voltage measured between the two resistors is done by the op-amp. The inputs to the op-amp draw no current. The copied voltage on the follower circuit output is supplied to the measurement device. The op-amp can supply whatever current the measuring device wants for its internal resistance and therefore the measured voltage is exactly what we would expect under ideal conditions - 1/2 the applied voltage.

This simple example illustrates the utility of the follower. A voltage from one part of the circuit can be fed to another such that there is no backward coupling. The measurement device to the right cannot impact what happens in the voltage divider on the left. The op-amp acts as a *buffer* that allows information to pass from the input to the output - but not the other way. This behavior allows us to build functional circuit blocks and then connect them together where adding more blocks to the end does not effect what happens upstream. With the behavior of the op-amp buffer, we can build (and sell) functional circuit blocks that can be easily chained together since each block can be analyzed, characterized, and built independent of the other ones. This ability to build and isolate functional blocks is very important in the rapid developments seen in modern electronics.

Conceptual Isolation

If you recall in our last chapter on filters, we discussed connecting filters in series. We saw that the ideal behavior of chaining RC filters in series was not possible in reality. Now we have a mechanism to reach the ideal scenario. We can use the op-amp follower as a buffer between different functional filter blocks.

5.5 Why negative feedback works

So now let's think about how and why negative feedback is good and positive feedback is bad (in circuits). Let's look carefully at an experiment where we build the follower circuit and drive the input with a square wave; the experimental result is shown in Figure 5.7. When we drive the positive input with a 1 kHz square wave and monitor the output, we see the "following" behavior if we look at the "long" time scales. On the left in Figure 5.7, both the input and output are plotted though we can't really see any difference between the two at this time scale. On the right we zoom in right around the transition. We see that the output takes a few microseconds to catch up to the input. Two important observations are

- It takes time for the op-amp's output voltage to change
- The output voltage changes continuously.

The rate that an op-amp can change its output depends on the chip, but rates on the order of volts per microsecond like we see here are pretty common.

Think carefully about what happens right at the transition in the

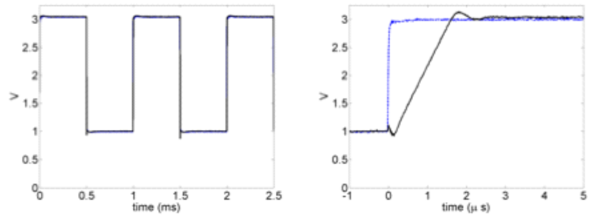

Figure 5.7 Response of the follower circuit to a 1 kHz square wave input. On the left when we look at long time scales the output seems to instantly follow the input. If we zoom into the microsecond time scale (on the right) we see there is a very short lag before the output settles to the input. On the right the input voltage is shown as the dashed blue line and the output is shown as the solid black.

zoomed in part of the Figure on the right. At $t < 0$, all three voltages, V_+, V_-, and V_{out} are equal and at 1 volt. The system is stable and happy. At t=0, we suddenly pull $V_+ = 3$ V. The instant that $V+ > V_-$, the op-amp wants to change its state toward 5V. However, the output voltage has to change in a continuous way - the output cannot just jump to 5 V. V_{out} starts increasing very rapidly from 1 V. Since the output is connected to V_-, for all times $V_{\text{out}} = V_-$. The positive input remains greater than the negative input and the op-amp keeps working to increase its voltage.

Somewhere around 1.5 microseconds, the op-amp output *exceeds* 3 volts. Remember that for all times $V_{\text{out}} = V_-$. At this time the negative input voltage is now greater than the positive input voltage. Thus remembering our open loop behavior, the op-amp's output wants to be driven back towards zero volts. The output has to change continuously, so the output voltage begins rapidly decreasing.

You should start to see the stabilizing effect of negative feedback. When the output voltage (equivalently V_-) is less than V_+, the output wants to increase toward V_+. When the output voltage is more than V_+, the output wants to decrease toward V_+. Feedback stabilizes things such that the two input voltages become equal very quickly. That the op-amp does this stabilization in a few microseconds is a function of the op-amp's design. There is little bit more to real story of the op-amp dynamics, and that is a topic we will pick up in the last chapter.

If you follow the same argument for the follower circuit with *positive feedback*, i.e. the output tied to the positive input, you would see the feedback is unstable. A small difference in V_+ and V_- would drive the output in a manner that would reinforce the difference until the op-amp get's stuck at one of the source voltage limits; 0 or 5 volts in this case.

5.6 Op-amp circuits with negative feedback

If we build and test several circuits with feedback we start to notice the behavior of the follower circuit is pretty general. When we build circuits with negative feedback, the output adjusts itself on a very rapid time scale to force the input voltages to be equal. This is completely consistent with the "rules" we found for open-loop behavior. There we found that for the op-amp to be outputting any voltage other than the power supply limits, $V_+ = V_-$. This "rule" of equal input voltages holds whenever the voltages are changing at rates much less than a volt per microsecond (or thereabouts, depending on the op-amp). Since many of the circuits we will build in lab, such as the EKG, where the changes are on the time scale of a second assuming that the inputs voltages are equal and that the rate of the op-amp is unimportant is a pretty good assumption. For frequencies less than about 10 kHz (and often much higher), we can typically ignore any of the op-amp dynamics.

For analyzing circuits with negative feedback, we always begin by making the following two assumptions which allows us to quickly and easily understand an op-amp circuit behavior.

1. **The inputs draw no current.** $I_+ = I_- = 0$.
2. **The input voltages are equal.** $V_+ = V_-$.

It is important to always try and remember that assumption 1 holds regardless of the frequency and that assumption 2 holds at "low" frequency where low typically means something less than 10 or 100 kHz.

5.7 Some example application circuits

Two classic application circuits are shown in Figure 5.8. Here I am going to ignore the power supply for now, and then we will deal with it in the next section. In general, I like this procedure of analyzing the op-amp circuit by ignoring the power supply first. Then, I'll add the power supply back into my analysis before actually moving toward constructing the circuit in the lab. As we will see, when we take the power supply into account we may need to adjust things a bit, but fundamentally the circuit is the same.

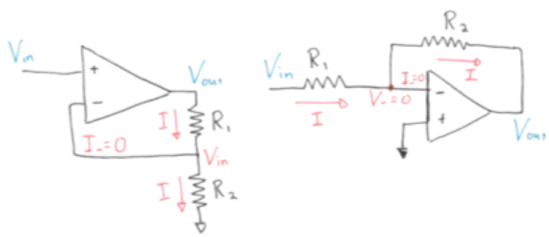

Figure 5.8 Simple (non-inverting) amplifier on the left and inverting amplifier on the right.

The amplifier circuit is shown on the left in Figure 5.8. This circuit is often called the non-inverting amplifier, but the double negative in the name always seems insane to me, so I am going just going to call it an amplifier. I have caved in accepting the decibel as a unit so at this point I have to stand up for something.

The labels in blue are usually all you might see in a circuit schematic, where the annotations in red show what we would learn from applying our two rules. The consequence of our two rules are;

- The inputs draw no current, $I_+ = I_- = 0$. For the amplifier, the current through R_1 is equal to the current through R_2.
- The input voltages are equal, $V_+ = V_-$. For the amplifier, the voltage at the negative input is equal to that set at the positive input; $V_- = V_{\text{in}}$.

Recognizing the voltage divider off the output of the op-amp (since no current goes into V_-), we can use our voltage divider analysis to get the answer quickly. The voltage divider equation would say that

$$\frac{V_{\text{in}}}{V_{\text{out}}} = \frac{R_2}{R_1 + R_2} \quad \text{or} \quad V_{\text{out}} = V_{\text{in}} \frac{R_1 + R_2}{R_2}$$

The circuit multiplies V_{in} by a constant, $(R_1 + R_2)/R_1$. In this case the output is always larger than the input and the output is always the same sign as the input. Hence the circuit is an amplifier.

The circuit on the right in Figure 5.8 is called the *inverting amplifier.* The annotations in blue are usually what you would see in a circuit schematic, where the annotations in red show what we would learn from applying our two rules. The consequence of our two rules are;

- The inputs draw no current. $I_+ = I_- = 0$. For the inverting amplifier, the current through R_1 is equal to the current through R_2.
- The input voltages are equal. $V_+ = V_-$. For the inverting amplifier, the voltage at the negative input is equal to ground, $V_- = 0$.

Once we apply our two rules, the rest of the analysis falls into place. We just need to apply Ohm's law to the two resistors, knowing that the currents are equal,

$$I_{R_1} = I_{R_2} \quad \text{or} \quad \frac{\Delta V_1}{R_1} = \frac{\Delta V_2}{R_2} \quad \text{or} \quad \frac{V_{\text{in}} - 0}{R_1} = \frac{0 - V_{\text{out}}}{R_2}$$

which can be written as

$$V_{\text{out}} = -V_{\text{in}} \frac{R_2}{R_1}$$

The circuit multiplies V_{in} by a constant, R_2/R_1, and changes the sign. The change of sign and amplitude change is why this circuit is called an inverting amplifier. Note that we will usually use the circuit where $R_2 > R_1$, which would amplify the output. If $R_1 > R_2$ then the circuit would attenuate the input.

5.8 Accounting for the power supply

Let's examine the *inverting amplifier* in a little more detail and now pay closer attention to the power supply limits. We will often use a power supply of 0 to 5 volts in the lab. Let's take the input voltage to be a 1 volt sine wave centered around ground; $V_{\text{in}} = \sin(\omega t)$. In Figure 5.9, we show the ideal behavior if we could ignore the power supply limits as well as the true behavior. The ideal behavior would be to invert the input signal and multiply it by 3. The op-amp is unable to output voltage outside $0 < V_{\text{out}} < 5$, thus the signal gets clipped when it tries to go negative.

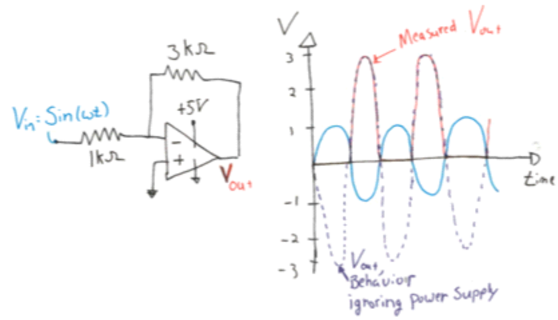

Figure 5.9 Inverting amplifier with a 0 to 5 volt power supply. The actual output voltage is shown in red and the dashed line shows the expected output if you ignore the power supply limits.

So we spot two issues here. When working with time varying signals and op-amps, the "zero" value for the signal needs to be somewhere in the middle of the op-amp's power supply. In our labs we will use 0 to 5 volts as the power supply, thus for time varying signals we will center them at 2.5 V. An input signal that goes outside the op-amp's power supply range is no good. If we are going to work with a 0 to 5 volt range we will want to shift the input voltage to $V_{\text{in}} = 2.5 + \sin(\omega t)$.

In Figure 5.10 we have shifted the op-amp circuit to be centered at 2.5 V. Both the reference value at V_+ and assumed that the input voltage has been shifted up by 2.5 Volts together. Everything is now right in the middle of the 0 to 5 volt supply. If we apply the rule that

$V_+ = V_-$, we find immediately that $V_- = 2.5$. This means the voltage *drop* across R_1 is $V_{\text{in}} - V_- = \sin(\omega t)$ - exactly as it was previously. Since no current flow through the op-amp inputs, the current through R_1 must equal the current through R_2. Therefore

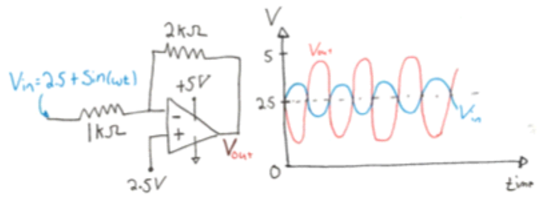

Figure 5.10 Inverting amplifier with a 0 to 5 volt power supply. Here, the input and reference in the circuit are shifted by 2.5 V.

$$I_{R_1} = I_{R_2} \quad \text{or} \quad \frac{\Delta V_{R_1}}{R_1} = \frac{\Delta V_{R_2}}{R_2} \quad \text{or} \quad \Delta V_{R_2} = \Delta V_{R_1}\frac{R_2}{R_1}$$

Finally, we can substitute for the voltage drops and obtain

$$V_{\text{out}} - 2.5 = -(V_{\text{in}} - 2.5)\frac{R_2}{R_1}.$$

The resulting behavior is sketched on the left in Figure 5.10. Notice that everything in the analysis is exactly as before, except we are now offset to 2.5 V relative to ground. So for the analysis and basic understanding of the circuit, the power supply doesn't matter. In practice if you build a circuit and don't properly account for the power supply limits, you might get what initially seems to be unexpected behavior.

5.8.1 Shift your frame of reference

Notice that the analysis with the circuit referenced to ground and with the symmetric voltage supply to the op-amp (i.e. something like +5 and -5 volts) is usually easier. The constant 0 volts shows up a lot and therefore can make the analysis a little simpler and more intuitive.

However, if in the lab we use the 0 to 5 volt power supply then we need to pay attention. In many of our circuits we can use 2.5 volts relative to ground to really act like "zero" for the signal. This is a fine way to think, because remember that voltage level is arbitrary and the difference is the only thing that matters.

What I tend to do when I work with the 0 to 5 volt reference for op-amps, is to shift my frame of reference such that I shift all the voltages to center the signal and the op-amp's power. An example is shown in Figure 5.11 for the amplifier circuit. On the left is the circuit with a symmetric power supply and the signal centered at 0. On the right, I can just shift everything by 2.5 volts. The one on the left is convenient for analysis and design. The one on the right is one I need to build. All we really have to do to shift our frame of reference is to move all the signal voltages that are referenced to ground and reference them to 2.5 V.

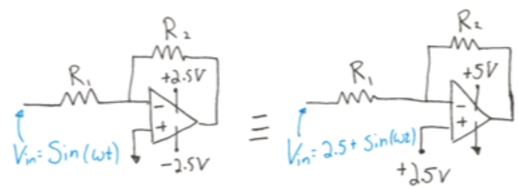

Figure 5.11 Amplifier with a 0 to 5 volt power supply. The two circuits are equivalent, we have just shifted the reference. The circuit on the left is generally a little easier to analyze, while the one on the right is what you are stuck with when working on a single sided power supply.

This change of reference can be confusing for students just learning to build circuits. It still messes me up sometimes. Checking and thinking about what is your "zero" point for the signal you want is always a critical step when debugging a non-functional circuit.

6
Complex impedance

We will now go over a more sophisticated but ultimately simpler method for analyzing circuits with sinusoidally varying voltages and currents. This method will involve us using complex and imaginary numbers to represent voltages and currents. Likely, the idea that we can use imaginary numbers in this way seems pretty odd at first. In this chapter I will attempt to be complete in my attempt to demonstrate that it really is fine (and powerful) to use the seemingly abstract idea of complex numbers to represent physical things like voltage and current.

Some of the sections provided in this chapter contain short proofs. These are not essential to the core idea, but they do try to address the inevitable questions around whether it is in fact acceptable to use complex numbers. I will try to convince you that it is perfectly fine and that in fact it makes our lives easier. Many of the sections here provide some detail that may not be important on your first reading, and I will try to denote those as we go along.

6.1 Imaginary numbers

The imaginary unit is defined as

$$i^2 = j^2 = -1.$$

In many disciplines i is used as the imaginary unit. In our class (and electrical engineering in general) j is preferred to avoid confusion with our accepted symbol for current.

Some students balk at the imaginary unit (probably because of the name) and think that it is a totally made up thing. Just because we call it imaginary makes it no less "real" than many concepts in math you take for granted. The first math that you are exposed to as a child is counting with the natural numbers, 1, 2, 3, 4, and so on. Very soon you are led to addition. You have 4 apples and your friend has 3. How many do you have together? As you get a little older you are led to other shorthand operations such as multiplication and powers. For example, if in a class of 25 students and each student has four apples, rather than writing $4 + 4 + 4... + 4 = 100$, you simply write that the class has 25×4 apples. Likewise when you start multiplying numbers, rather than writing for some problem, $4 \times 4 \times 4 \times 4 \times 4$ you write it more compactly as 4^5. So far there is no need for us to expand our concept of numbers beyond the natural numbers.

When you started to learn these concepts above, you found that the *inverse* of these operations did require you to extend your concept of numbers beyond the natural (or counting) numbers. Subtraction is the inverse of addition. When you subtract a large number from a small number, you learned that you could get *negative* integers. When you take the inverse of the power, you get operations such as the square root. When you solve $x^2 = 2$, you find that the square root of 2 is an *irrational* number. While these two examples are simple, they are times in your life when you had to extend your thinking about what a number was. You probably encountered this long ago and have forgotten that it may have seemed confusing at the time. Maybe you just took this for granted.

The imaginary numbers comes up quite naturally in the same way when we want to solve $x^2 = -1$. To account for the square root of negative numbers, we have to extend our thinking just a little bit. Despite the name, the imaginary number is just as real as any of the extensions to our numbers you are used to.

6.2 Complex numbers

A complex number, z, is one that has a real and imaginary part,

$$z = x + jy.$$

The number x is called the *real part* and y is called the *imaginary part*. Both x and y could be positive or negative, integers or irrational numbers. We often like to plot our complex number on a plot where the x-axis is real and the y-axis is imaginary. When we plot the point, as in Figure 6.1, we see that we could also fix the point in the complex plane by using a radius and angle where,

$$r = \sqrt{x^2 + y^2}$$

$$\theta = \tan^{-1}\left(\frac{y}{x}\right)$$

Using these definitions, we can rewrite our complex number as

$$z = x + jy = r\left(\cos\theta + j\sin\theta\right)$$

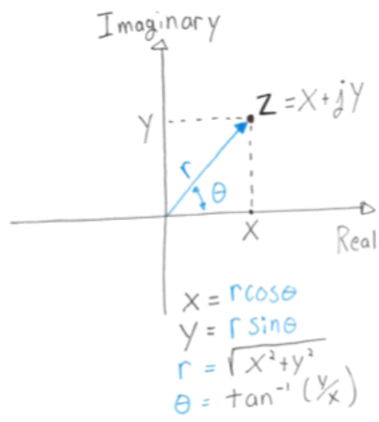

Figure 6.1 Schematic of the complex plane and some different forms for writing out complex numbers.

Adding complex numbers is pretty easy, we just add the real and imaginary parts separately. If we have $z_1 = x_1 + jy_1$ and $z_2 = x_2 + jy_2$

$$z_1 + z_2 = (x_1 + jy_1) + (x_2 + jy_2) = (x_1 + x_2) + j(y_1 + y_2).$$

Graphically, the vectors pointing to the complex numbers follow vector addition as shown in Figure 6.2.

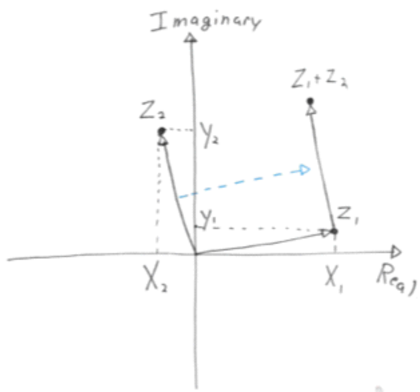

Figure 6.2 Schematic of the complex addition.

6.3 Euler identity

The most useful definition when working with complex numbers is the Euler identity,

$$e^{j\theta} = \cos\theta + j\sin\theta.$$

You may have encountered this identity before - if not it might seem strange that complex exponentials are equivalent to your usual trigonometric functions. To make sure you understand there is no mystery here, let me offer a little demonstration of how you can derive this relationship. The remainder of this section is to prove this relation to you.

This section can skipped on the first reading if you want to accept that Euler was correct and you first want to see the consequences of this relationship.

To derive the identity you need to remember the Taylor series, which says that if we want to approximate a one-dimensional function of x around zero as a polynomial, you just need to match the value of the function and all the derivatives. The Taylor series gives us the polynomial expansion,

$$f(x) \approx f(0) + \left.\frac{df}{dx}\right|_{x=0} x + \left.\frac{d^2f}{dx^2}\right|_{x=0} \frac{x^2}{2} + \left.\frac{d^3f}{dx^3}\right|_{x=0} \frac{x^3}{3!} + \dots \left.\frac{d^nf}{dx^n}\right|_{x=0} \frac{x^n}{n!}.$$

The notation $\left.\frac{df}{dx}\right|_{x=0}$ is written to denote explicitly that you take the derivative of the function and then evaluate the derivative at $x = 0$. The exclamation point is the *factorial* - i.e. $4! = 4 \times 3 \times 2 \times 1 = 24$.

Now let's apply the Taylor series to $f(x) = e^x$ (with no imaginary number). This function is nice because all derivatives of e^x are also e^x. Thus around $x = 0$,

$$e^x \approx e^0 + e^0 x + e^0 \frac{x^2}{2} + e^0 \frac{x^3}{3!} + \dots e^0 \frac{x^n}{n!}$$

since $e^0 = 1$ we have

$$e^x \approx 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \dots \frac{x^n}{n!}$$

Now, let's do the same thing for $f(x) = \cos x$. Recall that

$$\frac{df}{dx} = -\sin x, \quad \frac{d^2f}{dx^2} = -\cos x, \quad \frac{d^3f}{dx^3} = \sin x, \quad \text{and} \quad \frac{d^4f}{dx^4} = \cos x.$$

Substituting these relations into the Taylor series gives the approximation around $x = 0$,

$$\cos x \approx \cos 0 - \sin 0\, x - \cos 0\, \frac{x^2}{2} + \sin 0\, \frac{x^3}{3!} + \cos 0\, \frac{x^4}{4!} - \sin 0\, \frac{x^5}{5!} + \dots$$

Using the fact that sine and cosine at $x = 0$ are 0 and 1 respectively,

$$\cos x \approx 1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6} + \frac{x^8}{8!} + \dots.$$

Cosine is approximated by an even polynomial.

Now, let's do the same thing for $f(x) = \sin x$. Recall that

$$\frac{df}{dx} = \cos x, \quad \frac{d^2f}{dx^2} = -\sin x, \quad \frac{d^3f}{dx^3} = -\cos x, \quad \text{and} \quad \frac{d^4f}{dx^4} = \sin x.$$

Substituting these relations into the Taylor series gives,

$$\sin x \approx \sin 0 + \cos 0 \; x - \sin 0 \; \frac{x^2}{2} - \cos 0 \; \frac{x^3}{3!} + \sin 0 \; \frac{x^4}{4!} + \cos 0 \; \frac{x^5}{5!} + \ldots$$

Using the fact that sine and cosine at $x = 0$ are 0 and 1 respectively,

$$\sin x \approx x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \frac{x^9}{9!} + \ldots$$

Sine is an odd function.

Finally, let's do the Taylor series for $f(x) = e^{jx}$, the complex exponential. Here the derivatives are

$$\frac{df}{dx} = je^{jx}, \quad \frac{d^2f}{dx^2} = -e^{jx}, \quad \frac{d^3f}{dx^3} = -je^{jx}, \quad \text{and} \quad \frac{d^4f}{dx^4} = e^{jx}.$$

Evaluating these derivatives at $x = 0$ gives,

$$\left.\frac{df}{dx}\right|_{x=0} = j, \quad \left.\frac{d^2f}{dx^2}\right|_{x=0} = -1, \quad \left.\frac{d^3f}{dx^3}\right|_{x=0} = -j, \quad \text{and} \quad \left.\frac{d^4f}{dx^4}\right|_{x=0} = 1.$$

Now, using the above relations into our general expression for the Taylor series gives,

$$e^{jx} = 1 + jx - \frac{x^2}{2} - j\frac{x^3}{3!} + \frac{x^4}{4!} + j\frac{x^5}{5!} - \frac{x^6}{6!} - j\frac{x^7}{7!} + + \frac{x^8}{8!} + j\frac{x^9}{9!} + \ldots$$

or upon grouping the real (blue) and imaginary (red) parts,

$$e^{jx} = 1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6} + \frac{x^8}{8!} + \ldots \; j\left(x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \frac{x^9}{9!} + \ldots.\right)$$

Now you should see why I was coloring the equations with blue and red all along. Substituting in the blue and red equations for sine and cosine, we are left with our proof of the Euler identity,

$$e^{jx} = \cos x + j\sin x.$$

6.4 Polar form

Previously, we showed that a complex number could be written in two equivalent forms,

$$z = x + jy = r(\cos\theta + j\sin\theta)$$

Using the Euler identity, we see that we can also represent the complex number in *polar form*

$$z = r(\cos\theta + j\sin\theta) = re^{j\theta}.$$

While we aren't quite there yet, in the coming sections we will find that this form is extremely convenient. In polar form we will call r the *magnitude* of the complex number and θ the *phase*. Refer back to Figure 6.1 to visualize the magnitude and phase. The magnitude is the distance from the origin and the phase is the angle we make with the real axis.

It is worth noting that the polar form makes multiplication and division quite easy. For example, take two complex numbers $z_1 = x_1 + jy_1$ and $z_2 = x_2 + jy_2$. Multiplying the two numbers in rectangular form gives,

$$z_1 z_2 = (x_1 + jy_1)(x_2 + jy_2) = x_1 x_2 - y_1 y_2 + j(x_1 y_2 + y_1 x_2).$$

While this is fine to work with, you can see that as you increase the number of multiplies, the algebra gets messier. In polar form we have $z_1 = r_1 e^{j\theta_1}$ and $z_1 = r_1 e^{j\theta_1}$. So the complex multiplication becomes,

$$z_1 z_2 = r_1 e^{j\theta_1} r_2 e^{j\theta_2} = r_1 r_2 e^{j(\theta_1 + \theta 2)}$$

In polar form, the magnitudes multiply and the phases add. You can see this form is no more difficult if you are multiplying several complex numbers together.

Division is similarly easy in polar form, namely,

$$z_1 z_2 = \frac{r_1 e^{j\theta_1}}{r_2 e^{j\theta_2}} = \frac{r_1}{r_2} e^{j(\theta_1 - \theta 2)}.$$

For division, the magnitudes divide and the phases subtract.

6.5 Sinusoidal signals

In Chapter 4, we wanted solve for a circuit's sinusoidal steady state; we assumed voltage of a form such as

$$V(t) = 2\sin(\omega t)$$

where 2 is the amplitude of the sine wave in volts and ω would be the frequency in radians per second. By now, in the laboratory, you should be familiar with the idea that we can control the amplitude and the frequency of the sine wave of voltage.

We are now going to say that since complex exponentials are just sines and cosines in disguise, that we can represent voltages as

$$V(t) = \mathbf{V}e^{j\omega t}$$

where the bold $\mathbf{V}$ means that this number itself is a complex number of the form $\mathbf{V} = x + jy$. This may seem odd that you can represent a physical thing like a voltage with a complex number. It really is OK to do. Let's proceed so you can see how easy circuit analysis gets and then in the next section, somewhat tediously, I will try to demonstrate that using the complex representation is totally fine.

Let's go back to our trusty low-pass filter shown in Figure 6.3. If you look back you will see that our analysis using the fact that we know the relation between voltage and current for the two parts and that the current through the resistor and capacitor are the same that our equation is,

$$I(t) = C\frac{dV_{\text{out}}}{dt} = \frac{V_{\text{in}} - V_{\text{out}}}{R}.$$

We will now represent both V_{in} and V_{out} as complex exponentials,

$$V_{\text{in}}(t) = \mathbf{V}_{\text{in}}e^{j\omega t} \quad \textit{and} \quad V_{\text{out}}(t) = \mathbf{V}_{\text{out}}e^{j\omega t}$$

where both $\mathbf{V}_{\text{in}}$ and $\mathbf{V}_{\text{out}}$ are themselves complex numbers. Taking our equation,

$$RC\frac{dV_{\text{out}}}{dt} = V_{\text{in}} - V_{\text{out}},$$

and substituting in the complex exponentials we get,

$$RCj\omega\mathbf{V}_{\text{out}}e^{j\omega t} = \mathbf{V}_{\text{in}}e^{j\omega t} - \mathbf{V}_{\text{out}}e^{j\omega t}.$$

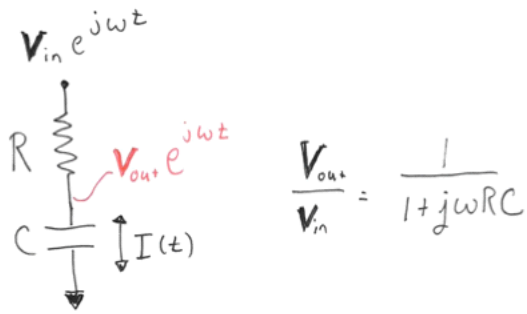

Figure 6.3 Schematic of the low pass filter.

We can immediately cancel e^{jwt}

$$j\omega RC\mathbf{V}_{\text{out}} = \mathbf{V}_{\text{in}} - \mathbf{V}_{\text{out}}.$$

and rearrange to get

$$\frac{\mathbf{V}_{\text{out}}}{\mathbf{V}_{\text{in}}} = \frac{1}{1 + j\omega RC}.$$

Note that in polar form that

$$1 + j\omega RC = \sqrt{1 + (\omega RC)^2}e^{j\tan^{-1}(RC\omega)}$$

Therefore, our result for the output voltage over the input voltage can be written as

$$\frac{\mathbf{V}_{\text{out}}}{\mathbf{V}_{\text{in}}} = \frac{1}{\sqrt{1 + (\omega RC)^2}}e^{-j\tan^{-1}(RC\omega)}.$$

Therefore, the *magnitude* of the ratio of the output to the input is

$$\text{magnitude} = \frac{1}{\sqrt{1 + (\omega RC)^2}}$$

and the phase between the input and output is

$$\text{phase} = -\tan^{-1}(RC\omega)$$

If you look back at the analysis of the low-pass filter with sines and cosines and all the trigonometric identities, you will see that we have obtained the same result. This is really cool. The magnitude of the complex number,

$$\frac{\mathbf{V}_{\text{out}}}{\mathbf{V}_{\text{in}}} = \frac{1}{1 + j\omega RC} = \frac{1}{\sqrt{1 + (\omega RC)^2}} e^{-j\tan^{-1}(RC\omega)}.$$

is the magnitude that we measure in the Bode plot. The phase of the complex number is the phase the we measured. The Bode plot *is* the complex function of frequenct we find here. The complex number naturally allows us to capture the magnitude and the phase of the signal.

It is important to note that we haven't actually derived anything new or obtained a new result. We obtained the same result, just much more easily. By the end of the chapter, I will show you an even easier way. Just hold on, it will all be worth it.

6.6 Is it really ok to use complex numbers?

The first time I came across using complex exponentials it seemed really weird and abstract. How can you represent a real physical thing with "imaginary" numbers? In this section I will work through the low-pass filter in extreme detail, trying to demonstrate the the result you get for the complex exponentials is equivalent to using sines and cosines. You can skip this section on first reading if you like and get to the applications. However, if representing physical things with complex numbers bugs you, I will try to address your concerns in this section. Ultimately, it is up to you to work through such things and convince yourself that they are true.

For our circuit analysis, let's start by assuming that we can represent a sinusoidal voltage as a complex number multiplied by the complex exponential,

$$V(t) = (x + jy)e^{j\omega t}.$$

Recall the

$$r = \sqrt{x^2 + y^2} \quad \text{and} \quad \phi = \tan^{-1}\left(\frac{y}{x}\right)$$

How did it go from scalar to vector?

and using polar form the expression becomes,

$$V(t) = (x + jy)e^{j\omega t} = re^{j\phi}e^{j\omega t} = re^{j(\omega t+\phi)}.$$

Going back to sine and cosine form using the Euler identity, we have,

$$V(t) = r\cos(\omega t + \phi) + jr\sin(\omega t + \phi).$$

We start to get the indication from here, that using a complex number in front of the $e^{j\omega t}$ term gives us the ability to build in the amplitude and phase of our signal.

To be concrete, let's work through the low-pass filter in Figure 6.3 yet again. Let's assume without loss of generality that the input voltage magnitude is set to one and thus,

$$V_{\text{in}}(t) = e^{j\omega t} \quad \text{and} \quad V_{\text{out}}(t) = \mathbf{V}_{\text{out}}e^{j\omega t}$$

In terms of sines and cosines this means that

$$V_{\text{in}}(t) = \cos(\omega t) + j\sin(\omega t)$$

and

$$V_{\text{out}}(t) = |\mathbf{V}_{\text{out}}|\cos(\omega t + \phi) + j|\mathbf{V}_{\text{out}}|\sin(\omega t + \phi)$$

I have denoted the real terms to be blue and the imaginary terms as red. The notation $|\mathbf{V}_{\text{out}}|$ means the *magnitude* (or radius) of the complex number $\mathbf{V}_{\text{out}}$.

Let's put these long expressions into our equation

$$RC\frac{dV_{\text{out}}}{dt} = V_{\text{in}} - V_{\text{out}}.$$

and we obtain

$$\omega RC|\mathbf{V}_{\text{out}}|\sin(\omega t + \phi) + j\omega RC|\mathbf{V}_{\text{out}}|\cos(\omega t + \phi) =$$

$$\cos(\omega t) + j\sin(\omega t) - |\mathbf{V}_{\text{out}}|\cos(\omega t + \phi) - j|\mathbf{V}_{\text{out}}|\sin(\omega t + \phi).$$

In order for the equation to be true, both the real and imaginary parts must be satisfied. Independently. So we must solve the real "blue" problem,

$$\omega RC|\mathbf{V}_{\text{out}}|\sin(\omega t + \phi) = \cos(\omega t) - |\mathbf{V}_{\text{out}}|\cos(\omega t + \phi)$$

and the imaginary "red" problem.

$$\omega RC|\mathbf{V}_{\text{out}}|\cos(\omega t + \phi) = \sin(\omega t) - |\mathbf{V}_{\text{out}}|\sin(\omega t + \phi).$$

Expanding the blue problem using the product-to-sum trigonometric identities just like we did when we first encountered the sinusoidal forcing, we have

$$\omega RC|\mathbf{V}_{\text{out}}|\left(\sin(\omega t)\cos(\phi)+\cos(\omega t)\sin(\phi)\right)=$$

$$\cos(\omega t)-|\mathbf{V}_{\text{out}}|\left(\cos(\omega t)\cos(\phi)-\sin(\omega t)\sin(\phi)\right)$$

For this to be true at all times we must cancel the $\sin(\omega t)$ and $\cos(\omega t)$ terms independently. Grouping the $\sin(\omega t)$ terms we have

$$\omega RC|\mathbf{V}_{\text{out}}|\cos(\phi)=-|\mathbf{V}_{\text{out}}|\sin(\omega t)\sin(\phi)$$

or

$$\tan^{-1}\phi=-\omega RC.$$

Grouping all the $\cos(\omega t)$ terms we have,

$$\omega RC|\mathbf{V}_{\text{out}}|\cos(\omega t)\sin(\phi)=\cos(\omega t)-|\mathbf{V}_{\text{out}}|\cos(\omega t)\cos(\phi)$$

or

$$|\mathbf{V}_{\text{out}}|=\frac{1}{\cos\phi-\omega RC\sin\phi}$$

Now notice that we solved the "red" problem already back in Chapter 4. Going back you will find that the solution proceeded just like the "blue" problem and the answer was

$$\tan^{-1}\phi=-\omega RC.$$

$$|\mathbf{V}_{\text{out}}|=\frac{1}{\cos\phi-\omega RC\sin\phi}$$

We obtain the same result for the real and imaginary problems. While I have gone through the pain to show this, it should not surprise you. At sinusoidal steady state I assumed the voltage was a sine wave input. Should I really expect a different result if I assume a cosine wave?

Through this somewhat tedious exercise I hope I have started to convince you that using the complex exponentials is equivalent to simultaneously getting your circuits response to a sine and cosine forcing. The real part is the cosine wave, the imaginary part is the sine wave. Since no new information comes from considering sine and cosine forcing, we have the same result from the real and imaginary parts. However, as you saw in the last section, once we are ready to accept the complex exponential's role in our lives, we can make progress much quicker as we can eliminate the need for a lot of trigonometric identities.

6.7 Complex impedance

Now that we went through all the analysis in detail with sines and cosines, let's introduce an even faster way to analyze circuits. Let's look at one component, the capacitor. Let's assume I can control the voltage *across* the capacitor and I want to vary the voltage sinusoidally, shown schematically in Figure 6.4. Hopefully I have convinced you that for analysis it is convenient to represent the voltage as

$$V(t) = \mathbf{V}e^{j\omega t},$$

where again $\mathbf{V}$ is a complex number. Let's now monitor the current

$dt = \frac{1}{j\omega}$?

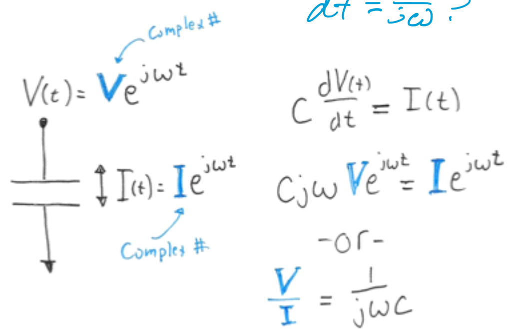

Figure 6.4 Schematic of the complex impedance for a capacitor. Impedance, like resistance, is voltage divided by current

through the part. It will also be sinusoidal and be at the same frequency as the voltage so we can represent as

$$I(t) = \mathbf{I}e^{j\omega t}.$$

b/c R is const

A capacitor has a relationship between voltage and current as

$$C\frac{dV}{dt} = I$$

or for our sinusoidal forcing,

$$j\omega C\mathbf{V}e^{j\omega t} = \mathbf{I}e^{j\omega t}.$$

Canceling the exponential terms we have,

$$j\omega C\mathbf{V} = \mathbf{I}.$$

Using the concept of resistance, or voltage divided by current, as our inspiration let's define the ratio of the voltage of the current to be called the impedance $\mathbf{Z}$,

$$\frac{\mathbf{V}}{\mathbf{I}} = \frac{1}{j\omega C} = \mathbf{Z}.$$

This formula tells us that the impedance *increases* at low frequency and that the current is always 90 degrees out of phase with the voltage (If the voltage were real, the current would be imaginary, thus a 90 degree difference in the complex plane). We discussed this general behavior of capacitors previously - high "resistance" at low frequency. Now we have the proper language to say that the capacitor has high *impedance* at low frequency. It is interesting that the impedance is a complex number. It captures the amplitude and phase relationship between voltage and current. Wat?

Since a resistor has an instantaneous relation between voltage and current, the impedance of a resistors is just the resistance, $\mathbf{Z} = R$. The nice thing about working with impedances is that all the familiar rules of resistors in series and parallel port over to the impedance world. Impedances in series add, and impedances in parallel add the reciprocals, see Figure 6.5 The analysis is now really simple. Resistors have $\mathbf{Z} = R$. Capacitors have $\mathbf{Z} = 1/j\omega C$. Once we use impedances, we can use our generalized version of Ohm's law, $V = I\mathbf{Z}$, and our usual rules of resistors in series and parallel. Our analysis with time dependent voltages is now really easy.

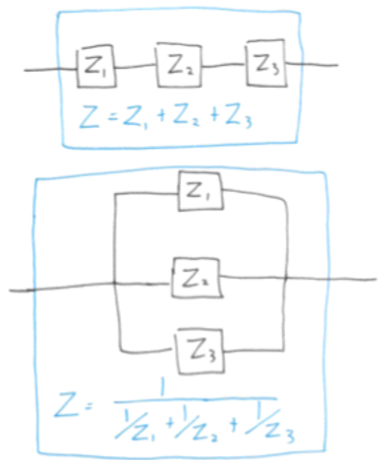

Figure 6.5 Schematic of impedances in series and parallel.

6.8 Examples: low-pass, high pass

Using the impedance idea, we can now analyze the low-pass and high-pass filter quite readily and see both cases as a general extension of the voltage divider. See Figure 6.6. For the general case of impedances, the output voltage between the two parts is given simply by the voltage divider equation using impedances instead of resistances,

$$\frac{\mathbf{V}_{\text{out}}}{\mathbf{V}_{\text{in}}} = \frac{\mathbf{Z_2}}{\mathbf{Z_1} + \mathbf{Z_2}}$$

For the case of the low-pass filter, let's insert the impedance values for the resistor and capacitor,

$$\frac{\mathbf{V}_{\text{out}}}{\mathbf{V}_{\text{in}}} = \frac{\frac{1}{j\omega C}}{R + \frac{1}{j\omega C}} = \frac{1}{1 + j\omega RC}.$$

This result is identical to the one derived early in the chapter. Plotting this complex number as a function of frequency *is* the Bode plot. Let that sink in again. The magnitude of the complex number will be the amplitude part of the Bode plot and the phase of the complex number is the phase between the input and output signals.

For the case of the high-pass filter, we just insert the proper impedance values for the switched locations,

$$\frac{\mathbf{V}_{\text{out}}}{\mathbf{V}_{\text{in}}} = \frac{R}{R + \frac{1}{j\omega C}} = \frac{j\omega RC}{1 + j\omega RC}.$$

Again, this complex function of frequency that represents the output relative to the input *is* the Bode plot.

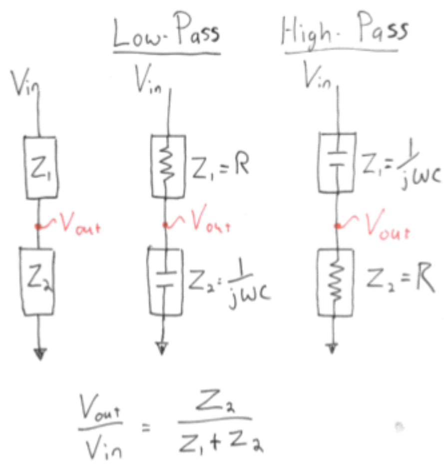

Figure 6.6 Schematic of impedances in high and low pass filter.

This is really something amazing. We have removed all the complexity of Chapter 4 where we assumed sinusoidal driving, wrote down the differential equations, and used all our trigonometric prowess to arrive at the result. We are now down to the capacitor acting as a simple extension of the resistor. We have to accept that we can represent the capacitor's impedance as complex number and we have to be willing to accept that the Bode plot is nothing more than the plot of this complex function of frequency. Once you can believe it, we have a powerful and simple tool at our disposal. In the end, we have done nothing new, but we have a simple and compact analysis tool.

This concept of complex impedance is a really powerful and general idea. Even if you never touch a circuit again in your life after this course, you will come across the ideas in this chapter again and again in many different areas of science and engineering.

6.9 Summary

The specific complex number forms for the low and high pass filter will show up in complex notation so often that it is worth putting them here again for you to remember and refer back to.

- **Low-pass filter**

$$\frac{\mathbf{V}_{\text{out}}}{\mathbf{V}_{\text{in}}} = \frac{1}{1 + j\omega RC}.$$

- **High-pass filter**

$$\frac{\mathbf{V}_{\text{out}}}{\mathbf{V}_{\text{in}}} = \frac{j\omega RC}{1 + j\omega RC}.$$

7

Active filters and op-amp dynamics

In this final chapter, we will review a few more applications involving op-amp circuits and use our complex analysis to make sense of the observed behavior. While we are just dipping our toe in the water, we will start to see how op-amps can be used to design different low-pass, high-pass and band-pass filters. Op-amps are useful in filter design because we can buffer between different functional blocks such that complicated functionality can be chained together. With just passive components like resistors and capacitors, we have to analyze an entire system at once as all the components interact. You will start to see that once you have the basic analysis tool of complex exponentials, you will have the ability to design and understand interesting new circuits. A new world of analog circuits will be opened for you.

We also find that the frequency response of op-amp circuits starts to deviate from our ideal behavior at high frequency. This effect was alluded to in previous chapters, but was more or less swept aside. We will conclude our study by looking at the internal dynamics of the op-amp to hopefully provide a little more insight into the magic that is contained inside that little black box.

7.1 Active filters

In Figure 7.1 are 4 different op-amp filter circuits. If we group the parts together as impedances, we see that all the circuits can be abstracted to having the same basic topology as the final circuit in Figure 7.1 with two generalized equivalent impedances, shown as circuit E.

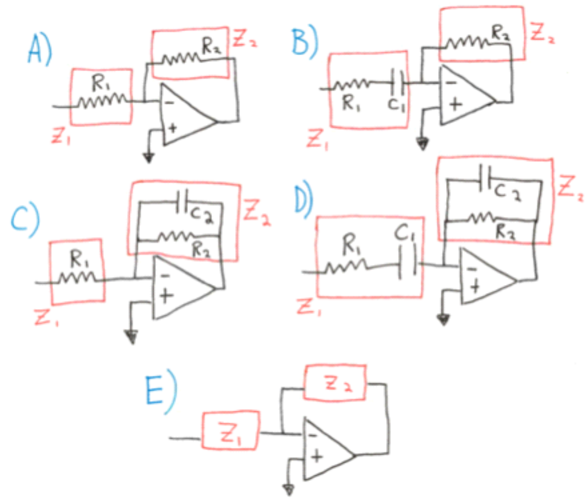

Figure 7.1 Some difference active filter designs.

Here, I will assume a symmetric op-amp power supply and will take all the signals referenced to ground, for simplicity. If you review the op-amp chapter you will recall that you can always shift everything up if you are working with a single sided power supply. The basic results here will remain unchanged, but recall that it is important to get the details right when working in the lab. Let's begin with the analysis of the final circuit in Figure 7.1 E. We see negative feedback so let's instantly apply our two op-amp rules. The rules and their consequence are;

- **The inputs draw no current**. Therefore, the current through Z_1 equals the current through Z_2.
- **The input voltages are equal**. Therefore, the voltage between the two impedances at V_- is set to ground.

Our op-amp rules plus the generalized version of Ohm's law give,

$$I = \frac{V_{\text{in}} - 0}{Z_1} = \frac{0 - V_{\text{out}}}{Z_2} \quad \text{or} \quad \frac{V_{\text{out}}}{V_{\text{in}}} = -\frac{Z_2}{Z_1}$$

For case A, $Z_1 = R_1$ and $Z_2 = R_2$, so the circuit is just the usual inverting amplifier; a result you have already seen. The other cases are a little more interesting.

7.1.1 Case B - high-pass filter

For case B we can use the two impedances in series on the input to obtain

$$Z_1 = R_1 + \frac{1}{j\omega C_1} \quad \text{and} \quad Z_2 = R_2.$$

Using these relations, the output voltage is

$$\frac{V_{\text{out}}}{V_{\text{in}}} = -\frac{Z_2}{Z_1} = -\frac{R_2}{R_1 + \frac{1}{j\omega C_1}} = -\left(\frac{j\omega R_1 C_1}{1 + j\omega R_1 C_1}\right)\left(\frac{R_2}{R_1}\right)$$

Now that you are becoming a little familiar with the form of different filters in complex notation, you might recognize that V_{out} is the product of a high-pass filter with the cutoff frequency set by $1/(R_1C_1)$ rad/s and an amplifier gain whose value is R_2/R_1. So with this topology you can do two operations at one time - filter and amplify.

7.1.2 Case C - low-pass filter

For case C we can combine the two impedances on the feedback loop using our rule for impedances in parallel to obtain,

$$Z_1 = R_1 \quad \text{and} \quad Z_2 = \frac{1}{\frac{1}{R_2} + j\omega C_2} = \frac{R_2}{1 + j\omega R_2 C_2}.$$

Using these relations, the output voltage is

$$\frac{V_{\text{out}}}{V_{\text{in}}} = -\frac{Z_2}{Z_1} = -\left(\frac{1}{1 + j\omega R_2 C_2}\right)\left(\frac{R_2}{R_1}\right).$$

You might recognize that V_{out} is the product of a low-pass filter with the cutoff frequency set by $1/(R_2C_2)$ rad/s and an amplifier gain whose value is R_2/R_1. So with this topology you can do two operations at one time - filter and amplify.

7.1.3 Case D - band-pass filter

For case D we can just combine everything we have done so far to obtain,

$$Z_1 = R_1 + \frac{1}{j\omega C_1} \quad \text{and} \quad Z_2 = \frac{R_2}{1 + j\omega R_2 C_2}.$$

$$\frac{V_{\text{out}}}{V_{\text{in}}} = -\frac{Z_2}{Z_1} = -\frac{R_2}{1 + j\omega R_2 C_2} \frac{1}{R_1 + \frac{1}{j\omega C_1}},$$

or with a little re-arrangement we can obtain

$$\frac{V_{\text{out}}}{V_{\text{in}}} = -\left(\frac{1}{1 + j\omega R_2 C_2}\right)\left(\frac{j\omega R_1 C_1}{1 + j\omega R_1 C_1}\right)\left(\frac{R_2}{R_1}\right),$$

By now, hopefully the pattern is clear. With this circuit, we have three operations,

- Low-pass filter with cutoff frequency of $1/(R_2C_2)$ rad/s.
- High-pass filter with cutoff frequency of $1/(R_1C_1)$ rad/s.
- Amplifier with gain of R_2/R_1.

This circuit is called a band-pass filter as we can set the filter to allow a band (or range) of frequencies in and attenuate anything lower or higher than the band. Typically we would set the frequency of the high-pass filter to be less than or equal to that of the low pass filter.

When designing a circuit for an application, you are free to select the component values to give the frequency cutoffs and amplitude gains that you desire. When designing this circuit to have particular performance, you have two frequencies and one gain you want to set, but you have four degrees of freedom to pick (2 resistors and 2 capacitors). You might need to arbitrarily set one of the component values and then select the other 3 based on your design criteria. In lab we typically have a wider selection of resistor values, so usually it works best to pick a reasonable capacitor value to start with and then see what the other three component values are. You may find you need to iterate this design process until all four components have reasonable values that you can obtain.

An example experimental amplitude Bode plot for cases B, C, and D is shown on the left in Figure 7.2. We can clearly see from this picture that the band-pass filter is just the product of the low and high pass filters. In this example we set the cutoff frequency of the high and low pass filters to all be the same and set at 159 Hz; shown as the vertical line. Notice that we see the classic low and high-pass behavior below 10 kHz. On the right in Figure 7.2 we show the experimental data compared to our model prediction above for the band-pass filter. Remember, the prediction is just the magnitude of the complex number of our expression $V_{\text{out}}/V_{\text{in}}$. We find near perfect agreement below 10 kHz and serious departures at higher frequency. The behavior at high frequency is the topic of the remainder of this chapter.

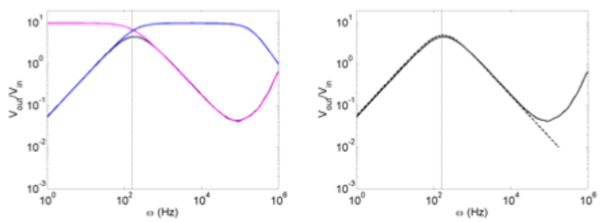

Figure 7.2 Active filter example and resulting experimental Bode plot. On the left we show Case B:high pass (blue), Case C: low pass (magenta), and Case D (black): band-pass. For this circuit $R_1 = 1\ k\Omega$, $R_2 = 10\ k\Omega$, $C_1 = 1\ \mu F$, and $C_2 = 0.1\ \mu F$. The cutoff frequency for both RC filters is 159 Hz, shown as the vertical line. On the left we compare the experimental band-pass circuit (solid blue) to the theoretical result (dashed black). There is excellent agreement below 10 kHz and serious departures as the frequency goes toward 1 MHz. Since the cutoff for the low and high pass are the same frequency, the resulting Bode plot for the band pass shows a peak at the characteristic frequency.

7.2 Op-amp dynamics

Notice in the last section that the measured and expected behavior of the op-amp circuit was not accurate at high frequency. In our op-amp explanations before you may recall that we discussed that we had to be careful to apply our usual rules only when changes were "slow". Let's know look under the hood of the op-amp and get a peak at what is inside (functionally) and hopefully enlighten your understanding just a bit. A *functional* schematic of the inner workings of the op-amp are shown in Figure 7.3. Note these functions are created through further internal circuitry for the chip using resistors, capacitors and transistors (a component we have not discussed). The circuit components that comprise a modern op-amp are quite complex and a topic in a later course for those interested in electrical engineering. In this functional op-amp model we see three basic operations. The op-amp

- takes the difference between the two input voltages.
- multiplies this voltage difference by a large number G.
- integrates this result with respect to time.

It is very important to note that it remains true that **op-amp inputs draw no current**. This rule still holds.

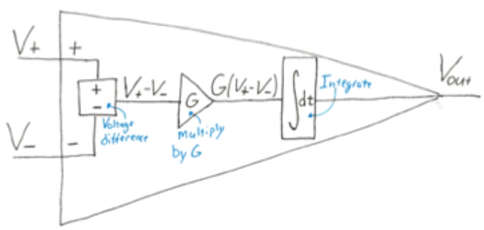

Figure 7.3 Simple model of the functional blocks inside a basic op-amp. The op-amp takes the difference of the input voltages, multiplies by G (which is large) and integrates with respect to time.

It is important to realize that the integrator inside the op-amp cannot integrate forever. It has saturation limits, just like a bathtub. A bathtub is an integrator for the net volumetric flow rate into/out of the tub. The bathtub integrates the volumetric flow rate to get the total volume of water stored in the tub. The net flow rate into/out of the tub is given by the difference between the flow of the faucet and the flow of the drain. If the faucet flow is on full blast, the bathtub will integrate (fill) up and overflow the top. The integrator (the bathtub) cannot accumulate any more. If the drain is open and the faucet off, the bathtub fully drains but cannot integrate to be less than than empty. Similar to the bathtub, the integrator inside the op-amp cannot exceed the limits of the power supply. Once the integrator reaches the power supply limits, the integrator (and voltage increase) saturates.

We can confirm our op-amp model, by conducting experiments in open-loop (with no feedback) as we did previously. We can provide a controlled step input and monitor the output as shown in Figure 7.4. We see that the output of the op-amp changes linearly in time when given a step input (and no feedback). Previously we described the output behavior as slamming between 0 and 5 volts when operating in open loop. Here we zoom our time scale, we see there is a rapid rate of change of the output that depends on the difference of the input voltages. This result fits the integrator model since the integral of a constant (the step change) is a line and the slope of the line depends upon the difference.

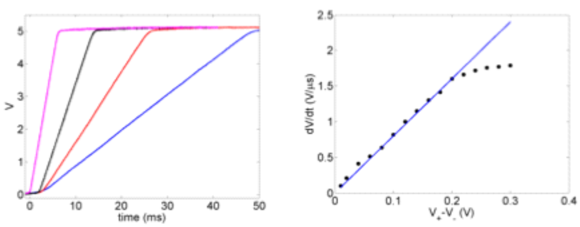

Figure 7.4 Data for an step input to the op-amp in open-loop mode for 10, 20, 40, and 80 mV difference at the input. On the right we show the rate of change of the output as a function of the voltage difference at the input.

If we conduct a number of experiments where we control the input difference and observe the rate, we see that the *rate* of increase of the output is proportional to the input difference. The model presented above, would say that

$$V_{\text{out}} = \int G(V_+ - V_-)dt$$

or equivalently,

$$\frac{dV_{\text{out}}}{dt} = G(V_+ - V_-).$$

Therefore, the rate of increase of the output should depend linearly on the input voltage difference, as the data in Figure 7.4 confirms. We can figure out the constant, G, by plotting our measured dV_{out}/dt against the input voltage difference as shown in Figure 7.4. Below an input voltage difference of about 0.15 V, a value of $G = 8\ \mu s^{-1}$ fits the data pretty well. Note the units - since I am expressing G in inverse microseconds, G is a large number in terms of inverse seconds. The model works really well until a maximum rate is reached.

It turns out the op-amp has a speed limit. This maximum speed is called the slew-rate limit and is the fastest that the output voltage can change. For our data we measure a slew rate limit of about 1.8 V/μs. In the analysis that follows we will ignore this slew rate limit as it makes simple linear analysis impossible but it is important to keep in mind.

7.3 Feedback revisited

Armed with our new model of the op-amp, let's revisit the idea of feedback by looking at the example of the (non-inverting) amplifier. The non-inverting amplifier in light of the functional op-amp model is shown in Figure 7.5.

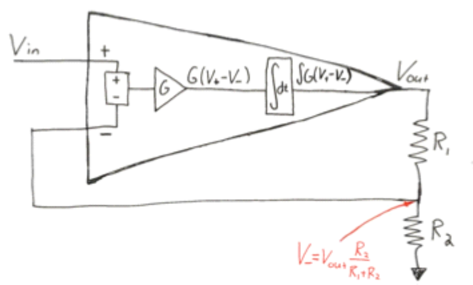

Figure 7.5 The functional diagram for the op-amp used in a basic amplifier circuit.

Our new model of the internal dynamics does not affect our assumption that the inputs draw no current, but it will shed some light on circuits with feedback and the idea that the input voltages are equal with negative feedback. Following our op-amp model we start with,

$$\frac{dV_{\text{out}}}{dt} = G(V_+ - V_-).$$

However for the circuit of interest $V_+ = V_{\text{in}}$ and the voltage divider equation gives $V_- = V_{\text{out}} \frac{R_2}{R_1+R_2}$. Substituting in these expressions we have

$$\frac{dV_{\text{out}}}{dt} = G\left(V_{\text{in}} - V_{\text{out}}\frac{R_2}{R_1+R_2}\right).$$

We can rewrite the equation as

$$\left(\frac{R_1+R_2}{R_2 G}\right)\frac{dV_{\text{out}}}{dt} = \left(V_{\text{in}}\frac{R_1+R_2}{R_2} - V_{\text{out}}\right).$$

or

$$\tau\frac{dV_{\text{out}}}{dt} = \left(V_{\text{in}}\frac{R_1+R_2}{R_2} - V_{\text{out}}\right).$$

where $\tau = \frac{R_1+R_2}{R_2 G}$. This expression might seem familiar. It is exactly the same equation we derived for the RC, low-pass circuit. If you don't believe me, go back and look. We have two results from the RC circuit that we can carry over; the step response and the sinusoidal steady state. We learned previously that the RC circuit will show exponential convergence to steady state when given a step input and will act as a low-pass filter when given a sinusoidal input. However, let's solve the problem again and see that the behavior is as before.

7.3.1 Step response

For the step response, let's take the example value of $(R_1+R_2)/R_2 = 10$. For our op-amp, we measured $G = 8\times10^6$ V/s - thus $\tau = 1.25\times10^{-6}$. Of course these numbers are specific to this exact problem and our op-amp but are indicative of typical numbers. The time constant of the exponential will be about a microsecond - very fast for our normal human time scales.

In the experiment, let's set $V_{\text{in}} = 0$ V and hold it there for some time. When the system reaches equilibrium, there is no change and $dV_{\text{out}}/dt = 0$. Therefore, our model states that *at equilibrium*,

$$\left(V_{\text{in}}\frac{R_1+R_2}{R_2} - V_{\text{out}}\right) = 0.$$

At equilibrium, if $V_{\text{in}} = 0$ V then $V_{\text{out}} = 0$. Now lets switch $V_{\text{in}} = 0.1$ V instantaneously. When the system reaches the new equilibrium state, then the output should become $V_{\text{out}} = 1.0$ V; since $(R_1+R_2)/R_1 = 10$.

You might remember the solution when we first studied the RC circuit is that we should find exponential approach to the new equilibrium state and the time scale for that exponential should be τ, which is a very small number. The data for an experiment are shown in Figure 7.6. Note that for this case, $\tau = 1.25\ \mu s$, and looks like we are onto a very reasonable model. We see from the data that V_{out} approaches 1V with an exponential decay.

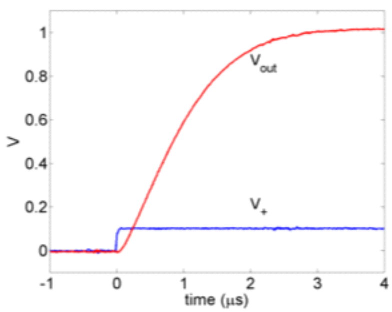

Figure 7.6 Data for an step input to the op-amp in an amplifier circuit shown in Figure 7.5.

To solve the model analytically for these specific values, we solve

$$\tau\frac{dV_{\text{out}}}{dt} = \left(V_{\text{in}}\frac{R_1+R_2}{R_2} - V_{\text{out}}\right) = 0.1(10) - V_{\text{out}} = 1 - V_{\text{out}},$$

with the initial condition $V_{\text{out}}(0) = 0$. Let's derive the solution to the problem as before by separating the variables,

$$\frac{dV_{\text{out}}}{1 - V_{\text{out}}} = \frac{dt}{\tau},$$

and integrating

$$\ln(1 - V_{\text{out}}) = \frac{-t}{\tau} + B,$$

where B is a constant of integration. Applying the initial condition $V_{\text{out}}(0) = 0$ gives us the value of $B = \ln(1)$. Rearranging the expression yields,

$$V_{\text{out}} = 1 - e^{-t/\tau}.$$

We have now really quantified the ability of the op-amp to maintain the two input voltages to be equal. The dynamics of the step response is such that the system reaches equilibrium on the time scale of τ. This time scale is very fast due to the large value of the gain inside the op-amp chip. An integrator comes to equilibrium - i.e. stops integrating - when the *net input* to the integrator is zero. Think of the bathtub example - if the flow rate into the tub via the faucet equals the flow out via the drain then the integrator (the tub) stops changing its state (the amount of water in the tub). It is the same here, when the input to the integrator $V_+ - V_-$ is equal to zero, the system has come to equilibrium and the output voltage stops changing.

You can also start to see why positive feedback is unstable. If we rebuilt the circuit and conducted the analysis for positive feedback, everything would be the same as above. The only change is that the sign would change and we would be solving,

$$\tau \frac{dV_{\text{out}}}{dt} = V_{\text{out}} - 1.$$

The solution procedure would be the same only we would have a positive exponential, i.e. $V_{\text{out}} = e^{t/\tau}$. The output of the integrator blows up exponentially until it saturates at the power supply limits.

7.3.2 Sinusoidal steady state

Let's consider the same amplifier circuit but consider the sinusoidal steady state response. Let's use our complex analysis and take $V_{\text{in}} =$

$\mathbf{V}_{\text{in}}e^{j\omega t}$ and $V_{\text{out}} = \mathbf{V}_{\text{out}}e^{j\omega t}$ where $\mathbf{V}_{\text{in}}$ and $\mathbf{V}_{\text{in}}$ are complex numbers that are constant in time. Taking our dynamic equation of the op-amp

$$\tau \frac{dV_{\text{out}}}{dt} dt = \left(V_{\text{in}} \frac{R_1 + R_2}{R_2} - V_{\text{out}} \right).$$

and substituting in the complex exponentials yields,

$$\tau j\omega \mathbf{V}_{\text{out}} e^{j\omega t} = \left(\frac{R_1 + R_2}{R_2} \mathbf{V}_{\text{in}} e^{j\omega t} - \mathbf{V}_{\text{out}} e^{j\omega t} \right).$$

Canceling $e^{j\omega t}$ leaves,

$$\tau j\omega \mathbf{V}_{\text{out}} = \left(\frac{R_1 + R_2}{R_2} \mathbf{V}_{\text{in}} - \mathbf{V}_{\text{out}} \right),$$

or after some rearrangement,

$$\frac{\mathbf{V}_{\text{out}}}{\mathbf{V}_{\text{out}}} = \left(\frac{R_1 + R_2}{R_2} \right) \left(\frac{1}{1 + j\omega\tau} \right),$$

This expression is the product of the amplifier gain $\frac{R_1+R_2}{R_2}$ and a low pass filter with a cutoff at $\omega = 1/\tau$. The experimental data are shown in Figure 7.7. Now we see that we can quantify the frequency dependent behavior that we only alluded to previously. Below the cutoff frequency of $1/\tau$ rad/sec, we can safely use the simple model that the op-amps input voltages are equal. Above this frequency, we would need to account for the internal dynamics of the op-amp. In this example $\tau = 1.25$ μs which corresponds to 127 kHz, which appears to be about where the "knee" in the amplitude Bode plot lies.

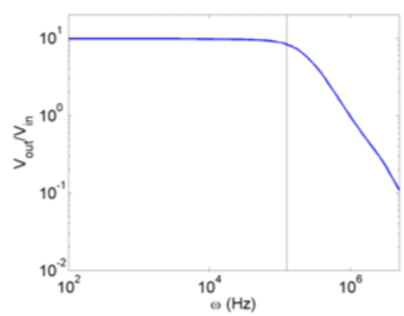

Figure 7.7 Experimental amplitude Bode plot for the op-amp in an amplifier circuit .