
*A Pragmatic Introduction
to the Art of Electrical
Engineering*

Paul H. Dietz

*A Pragmatic Introduction to the Art of
Electrical Engineering* **i**

LICENSE

Rights and Obligations **vii**

How it Works **vii**

A Disclaimer **viii**

CREDITS

How Did We Get Here? **ix**

A Book is Born **ix**

And I Want to Thank All the Little People... **x**

PROLOGUE

*Electrical Engineering for Fun and
Profit* **xi**

Cold Sandwiches, again? **xi**

Electrical Engineering as Programming and
Interfacing **xii**

The Basic Stamp 2 **xiii**

About This Book **xiv**

CHAPTER 1

*Getting Started with the BASIC Stamp
2* **1**

The Problem **1**

What You Need to Know **1**

What is a BASIC Stamp 2? **2**

How Do I Wire it Up? **2**

How Do I Get to the Software? **5**

A First Example Program **5**

A Second Example Program **6**

CHAPTER 2

Lights and Switches **8**

- The Problem **8**
- What You Need to Know **8**
- What is Voltage? **9**
- What is Current? **10**
- What is an LED? **12**
- How Do I Interface a Switch? **16**
- What is a Seven Segment Display? **18**
- Where Do We Go Next? **20**

CHAPTER 3

Maybe **21**

- The Problem **21**
- What You Need to Know **22**
- What is a Voltage Divider? **22**
- How Do I Solve More Complex Resistive Circuits? **24**
- Are There Any Tricks That Can Make This Easier? **27**
- What is an Independent Source and What is Superposition? **30**
- What is a Digital to Analog Convertor? **32**
- What's Next? **33**

CHAPTER 4

Guess the Number **34**

- The Problem **34**
- What You Need to Know **34**
- What are the limitations on our DAC? **35**
- What is an Amplifier? **39**
- How do you build an Analog to Digital Convertor? **44**
- What's Next? **47**

CHAPTER 5	<i>Timing is Everything</i> 48
	The Problem 48
	What You Need to Know 49
	What is a Serial Interface? 49
	What is a Capacitor? 53
	How Do I Use a Capacitor in a Circuit? 56
	What is an Oscilloscope? 64
	What's Next? 67
CHAPTER 6	<i>Déjà Vue</i> 68
	The Problem 68
	What You Need to Know 69
	What is an Inductor? 69
	How Do I Use an Inductor in a Circuit? 71
	How Do I Handle Nonzero Initial Conditions? 78
	What is an LC Circuit? 80
	What is a Loop Detector? 83
	What's Next? 85
CHAPTER 7	<i>Off the Wall</i> 86
	The Problem 86
	What You Need to Know 87
	What is AC Power? 87
	What is a Transformer? 90
	What is a Rectifier? 92
	What is a Voltage Regulator? 96
	What's Next? 97
CHAPTER 8	<i>Taking Control</i> 98
	The Problem 98
	What You Need to Know 99

	How Do I Measure Temperature? 99
	What is an Appropriate Type of A/D Conversion for Measuring Temperature? 101
	What is a Relay, and How Do I Drive It? 106
	How Do I make Noise? 108
	What Algorithm Do I Use to Control the Pumps? 108
	What's Next? 109
CHAPTER 9	<i>Clap On... 110</i>
	The Problem 110
	What You Need to Know 111
	How Do I Detect Sound? 111
	How Do Linear Systems Respond to Sinusoids? 113
	How Do I Generalize Ohm's Law? 114
	How Do I Detect a Clap? 120
	What's Next? 121
APPENDIX A	<i>The BASIC Stamp 2 Serial Cable 123</i>
	Roll Your Own 123
	The Connections 124
APPENDIX B	<i>Equipment 125</i>
	The Kit 125
	Other Supplies 127
	The Smart Shower 127
	Test Equipment 128

Rights and Obligations

How it Works

I have often been frustrated by the terribly high cost of textbooks. As an author, this is my chance to do something about it. Rather than seeking a traditional publisher, I am distributing this book electronically. However, this book is neither free, nor in the public domain. I retain all rights except those specifically granted below. Please be aware that I have considerable legal resources at my disposal, and I will use these to ensure compliance with this agreement.

That said, here are the terms of the agreement:

Schools, businesses and other institutions are required to pay a license fee for the use of this text, except in the case of evaluation as discussed below. If the text is to be used in a class, seminar, training session or similar group educational setting or individual study, a fee of \$5 (US currency) per student is required. Alternatively, if this text is used in such a setting, and students are required to purchase a physical copy as a course requirement, a fee of \$10 (US currency) per a copy should be remitted. Rights to make these copies or otherwise use this text are given only if these fees are paid within 30 days of the first learning session. Failure to submit the fees within the allotted time indicates an agreement to pay a fee of \$1000 (US currency) per student or copy as described previously, as well as all collection

expenses incurred by the author and his agents due to said failure, including legal fees.

Individuals may download and print one copy for personal use only. There is no required fee for this use. However, if you find this text interesting/useful, a voluntary donation of \$4 (US currency) is requested.

Course instructors and reviewers are permitted to download and print one copy for evaluation purposes only. There is no fee for this.

Any use not explicitly indicated here must be approved in writing by the author.

All copies of this book, whether physical or electronic, must be complete, including this license agreement.

Fees should be paid in United States dollars, in cash, or by check drawn on a U.S. bank and mailed to:

Paul H. Dietz

6 Prestwick Drive

Hopkinton, MA 01748

USA

These rules are in effect until January 1, 2002. After that date, no further copies of this text may be downloaded, copied or printed without express permission of the author. (The intent is to have a revised edition available by that date.)

A Disclaimer

Although I have made a good faith effort to ensure the accuracy of the content in this text, I can not absolutely guarantee any of the information contained herein. Persons and institutions are instructed to refrain from basing critical systems upon circuits or ideas in this text, especially systems where a failure could result in human harm or serious financial loss.

How Did We Get Here?

A Book is Born

For the Fall of 1996, I was given the assignment of teaching the required introductory EE course for other engineering majors. Usually, visiting faculty were relegated to this unseemly task, but we were shorthanded so some of us on the tenure track would have to pay our dues. My fellow faculty warned me to expect terrible student evaluations, since most of the students were only taking the course because it was required, and really didn't want to be there. It seemed pretty grim.

Knowing that I would soon be leaving, I decided to throw caution to the wind, and teach a radically different kind of introductory course - one based totally on projects, yet with a sound theoretical underpinning. I couldn't find an appropriate text, and in any case, I knew my students couldn't afford both a text and the serious lab kits I had in mind. So I resolved to write this book "on-the-fly" over the course of the semester. Each weekend, I would build, write and draw like crazy, hand it to my editor in chief, my wife Cathy, you would rather bluntly tell me how bad it was. Then I would start again, often from scratch, and churn out something that she could reasonably fix up. The result is this text.

And I Want to Thank All the Little People...

Obviously, this book only exists due to the wonderful support of my wife, Cathy, who not only tolerated losing many weekends to this effort, but also provided detailed technical suggestions, did significant rewrites, and cleaned up most of the more outrageous runs on sentences, like this one.

A great deal of credit goes to my fabulous teaching staff, Pat Malloy and Bill Glenn, who worked far above and beyond the call of duty. They put in absolutely insane hours in the lab, helping all of our students to successfully complete all of the projects. They ran review sessions, prepared many post-lab handouts (“here’s what you learned”) and generally made the course a smashing success. They also made invaluable suggestions, many of which are incorporated in this version of the text.

Finally, I’d like to thank Ken Gracey of Parallax, who has been pushing me to make this book more widely available. Hopefully, somebody out there will find this useful...

Electrical Engineering for Fun and Profit

Cold Sandwiches, again?

On those days when I was sick enough to stay home from school, my Mom would let me watch mid-day TV. One of the most common commercials of those time slots began with the depressed husband complaining, “Cold sandwiches, again?” The wife suggests technical training in electronics. In no time at all, the happy couple is gorging on roast something or other. This could be you.

Well, maybe an understanding of electronics won’t change your life quite this dramatically, but it certainly couldn’t hurt. Look around you. There are electronic gadgets everywhere. Wouldn’t you like to know how they function? After just one semester of study with this text, you’ll have - I guess I have to be honest here - absolutely no clue how any of it works.

The problem is that electronic stuff has gotten much too complex. There are now toothbrushes with more complex circuitry than was in ENIAC, the first computer! You can’t possibly understand it all in one semester.

This presents an interesting dilemma for those of us trying to teach an introduction to electrical engineering, especially when it is a terminal course. (No, we don’t mean that it will kill you - we mean that it might be the only EE course you ever take.) What should we teach?

In most introductory EE classes, the emphasis is on abstract fundamental principles. “Here’s a circuit with 26 resistors, 4 voltage sources, and 2 current sources - solve for everything.” Questions like these might build your analytical skills, but quickly deplete your stock of No-Doze. Why in the world would you ever want to solve a problem like that?

(There is actually a reason. If you continue in electrical engineering, and enter the particular subdiscipline of analog circuit design, you can then spend hours checking the result your circuit simulator produced in 0.2 seconds. This is very handy.)

This book takes a totally different approach. Instead of dealing in the abstract with an occasional fabricated “real world” example, we will present real problems, and show you what you need in order to solve them. Fundamentally, we know that given the limited time, there is no way we can explain everything. But we can teach you enough to make you dangerous. (Dangerous, that is, to professional electrical engineering consultants that will typically charge you a fortune for things you can whip up in your basement in 20 minutes.) After a semester, you should be able to create electronic things that will amaze your friends and family. However, you will still have no clue how that electronic toothbrush really works.

Electrical Engineering as Programming and Interfacing

Go find your favorite electronic gadget. We’ll wait.

Okay, open it up, and what you will undoubtedly see are a bunch of small black boxes attached to a board. Most of the black plastic things are integrated circuits. Odds are pretty good that the biggest one is some sort of microprocessor or microcontroller - basically, a computer on a chip. The rest is probably stuff the micro needs to operate, or to talk to the outside world.

The curious thing is that the people who “design” these electronic things are mostly buying parts out of a catalog, and hooking them together, often just as diagrammed on some datasheet. So, as Walter Mondale (warning - archaic reference for the Internet generation!) might have said, “Where’s the beef?” - what did these people really design?

Part of the “design” was in choosing the right parts, but lots of companies use very similar, if not identical parts. What often distinguishes an electronic product is not

its hardware, but its software! Remember the micro, the computer inside? It is a great deal easier and cheaper to write software than to design and build hardware. So the intellectual capital largely goes into the software.

How did we get to this state of affairs? Call it the digital revolution, if you like. Micros got irresistibly cheap. At the time of this writing, 8-bit microrcontrollers are just starting to fall below \$0.50/unit. So rather than designing some tricky circuit to perform some control function, you buy some mass produced micro, interface it to your stuff, and simply program it to do whatever you want. This accurately describes a vast array of modern electronic products. Not everything, but a lot of stuff.

Programming the little computers, while sometimes painful, is fairly straight forward. Hopefully, if you are reading this book, you have some significant programming experience. So this part is easy. The problem is, how do you hook up these little computers to do useful stuff? How do you interface the micros? This is the question we will really be addressing in this text.

(Some of you might be wondering about those people who design the chips - they must really be doing some serious EE. Ironically, these chips have gotten so complex that they are physically laid out by electronic design automation software. How do you tell the software what you want the chip to do? You write programs in a hardware description language. So even here, the problem is largely reduced to programming.)

The Basic Stamp 2

This is a class in electrical engineering, not programming. But it is very difficult to talk about building modern circuits without doing some programming. And, as we implied earlier, programming a micro can be tedious.

Enter Parallax, Inc. They make a series of tiny microcontrollers with built in BASIC interpreters. These micros are relatively expensive, slow, and kind of kludgy (a favorite term of your author), but remarkably powerful and simple to use. Called BASIC Stamps, they are literally postage stamp size.

In this text, we will presume that you have access to a BASIC Stamp 2 and the accompanying documentation. We will use the Stamp as our vehicle to explore electrical engineering, and the problems of interfacing a micro to the real world.

About This Book

Each of the following chapters will begin with a problem - How do you build a system to do such and such. This will be followed by a discussion of the background material you will need to interface the BASIC Stamp so as to solve the problem. The hope is that this approach will not only yield a solid understanding of electrical engineering fundamentals, but will also promote actual skill at designing and building functional electronic systems.

Getting Started with the BASIC Stamp 2

The Problem

Okay. You've got a BASIC Stamp 2. Make it do something.

I guess we can be more specific, but that is the general idea. Go to the lab, hook up your Stamp, and run the two example programs from this chapter. In addition, you should write a program that counts seconds in the debug window.

What You Need to Know

In order to solve this problem, here are a few things you need to know:

- What is a BASIC Stamp 2?
- How do I wire it up?
- How do I get to the software?
- A first example program
- A second example program

What is a BASIC Stamp 2?

A BASIC Stamp 2 is actually a PIC microcontroller with a BASIC interpreter in ROM. It also includes EEPROM for program storage, a voltage regulator, and a handful of other components to make it useful for embedded control applications.

Let's try that again in English.

A BASIC Stamp 2 is a small computer which you can easily program with very simple commands. Programs can be stored or erased without special hardware, and the programs remain in memory even after you remove the battery. Although most of the circuitry on the Stamp actually runs off of 5 Volts, there is a device which allows you to power it from a 9 Volt battery, automatically converting this into the required 5 Volt supply. In addition, there's lots of other stuff on the Stamp to make it useful for controlling everything from airplane servos to your bedroom lights, however, you can't run Windows98 on it.

There's a lot one could say about the Stamp. Why, you could even write a book about it. And as luck would have it, you should each have a copy of just such a book - the BASIC Stamp 2 manual.

Rather than trying to summarize the manual here, you really should read it for yourself. On a first reading, try to make it through the first section (about 25 small pages), and just skim through the commands section which follows. The manual ends with some applications information, which you can safely skip over for now.

(Note: The entire manual is available on-line from the Parallax web site at: <http://www.parallaxinc.com>.)

How Do I Wire it Up?

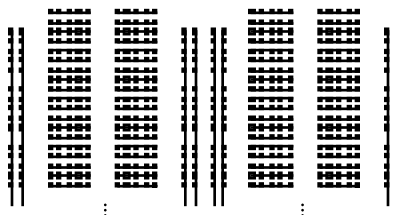
To do anything with the Stamp, you must connect it to the serial port of an appropriate computer, and also connect power. Parallax sells a nice little carrier board to help you do this, but they are expensive, and difficult to use when adding additional circuitry. Instead, we will put the Stamp into a solderless breadboard (which we'll explain momentarily), and use a custom made serial cable and a 9V battery clip.

The manual shows a picture of the Basic Stamp 2 and details all the connections. This figure also appears below. You might want to make a photocopy of this and



paste it someplace handy because you will need to refer to it quite frequently. Looking at the Stamp, pin 1 is right next to where it says “Parallax.”

A solderless breadboard (sometimes referred to as a Proto-Board, the brand name of a particular manufacturer) allows you to make connections by simply pushing components and wires into little holes which connect in a well organized pattern. A diagram of the connections in a typical breadboard is shown below.



The long connected runs are generally used for power and ground connections since they must be routed so many places. A note of caution here: on some boards, these longer runs, often called *buses*, are broken into unconnected segments in a less than obvious manner. Make certain that you really understand the connection pattern for your particular board before you begin wiring.

The Stamp is a 24 pin DIP, or Dual In-Line Package. That means it has two rows of pins. The breadboard is designed so that DIPs can straddle across to two sets of horizontal rails, allowing you to make easy connections to any pin independently. This is done by cutting little pieces of wire, stripping the ends, and inserting them into the proper holes.

Since we’re on the topic of breadboards, this is an appropriate time to say something about wiring style. Every wire should be cut to the appropriate length, no less and no more, and neatly placed on the board. Personally, I like to see wires running only horizontally and vertically - no angles. Careful wiring will make your circuit

infinitely easier to debug, should there be a problem. If your circuit looks like a bird's nest, you have almost no chance of finding mistakes.

Once you become experienced with the basic tools of the trade (wire cutters, strippers and needle nose pliers), there is a great little trick for making up correct length wires very rapidly. What you really want is a piece of insulation that runs from one connection point to the other, with some excess wire sticking out both ends. So instead of cutting a length of insulated wire and then stripping the ends, strip off a long piece to give some working room, and then strip a piece of insulation the correct length and slide it to about 1/2 cm from the end of the wire. Then you just cut the wire so as to leave a 1/2 cm of wire sticking out the other end. and you're done. To make more wires, you just keep stripping the right length, sliding it to almost the end, cutting, and inserting the result in the right place. For very short wires (e.g. adjacent pins), don't even bother with the insulation - just use the bare wire.

To connect up your Stamp, carefully insert it into the breadboard making certain not to bend any of the pins. Although the Stamp will run nicely from a 9V battery, flipping the battery connections, for even a moment, will destroy the Stamp. (I call that a cancelled Stamp.) Unfortunately, it is far too easy to do this with a 9V battery when fumbling to get the clip on the right way. It is highly recommended that you place a diode (a kind of electronic one-way valve) in series with the battery to prevent a reverse connection from destroying your Stamp. We'll learn all about diodes in the coming chapters, but for now, locate a 1N4007 diode, and connect the end with the band to the Stamp pin 24 (PWR) and the other end to the red wire (i.e. the positive side) of a 9V battery clip. The black wire (i.e. the negative side) of the battery clip should go to pin 23 (GND).

Next, you need to connect the serial cable. One end of the cable gets connected to a PC serial port. The other end should have 4 connections that go to pins 1 - 4 (TX, RX, ATN and GND respectively) on the Stamp in the correct order. Because cables vary, you should refer to the appendix, *Making a Stamp Serial Cable*.

This completes the wiring.

How Do I Get to the Software?

The PC needs special software to let you program and communicate with the Stamp. Copies of this software are available from the Parallax web site: <http://www.parallaxinc.com>. Check out the Stamp documentation for a full description of what you need (probably stamp2.exe), and how to use it.

A First Example Program

Now that everything is wired up, we're ready to start on our first program. Generally, we will have all sorts of goodies wired up for the Stamp to control. But since we want to try out the connections we have so far, we will content ourselves with just sending some text back to the computer to see that things are working.

So, without further ado, here's the first program - Hello, Good-bye World!

```
loop:  debug cls
        debug "Hello, World!"
        pause 1000
        debug cls
        debug "Good-bye, World!"
        pause 1000
        goto loop
```

First, connect the battery. Then follow the instructions in the manual for entering the program. When you have completed this, type Alt-R (which means to type the R key while holding down the key labeled Alt). You should briefly see a message saying that the program is downloading, and then the debug screen should appear alternately flashing the two messages. If you get an error message about not being able to locate the hardware, you have probably miswired something, put the diode in backwards, or forgotten to connect the battery.

Examine this program carefully. "loop" is a label, and could have been called anything. "debug", "pause" and "goto" are all commands that you should look up in your manual. Make sure you understand what is going on here. ("debug" is the moral equivalent of the "print" statement you may have seen in other versions of BASIC, and is very useful for - surprise - debugging!)

If you wish to modify your program, hit a key other than space to remove the debug window, modify your code, and then type Alt-R again. This will replace the old program with the new one. It really is that simple.

A Second Example Program

Our first program was pretty self explanatory. Next, we are going to try something with a little bit more interesting syntax. The goal of this next program is to add two numbers together, and then display the result. Here's the code:

```
'Our second example program
'Paul H. Dietz
'8/28/96

'-----
'This program adds two numbers together, and displays
'the result in the debug window.
'-----

'First, define the variables
num1 var byte 'add this number
num2 var byte 'to this number
result var byte 'to get this number

num1 = 5
num2 = 7

'add the numbers
result = num1 + num2

'show the result
debug dec num1, " + ", dec num2, " = ", dec result

stop
```

As in most programming languages, you need to define your variables in advance of using them. Here we have declared our numbers to all be bytes. Thus, they can range from 0 to 255. (On a Stamp, all numbers are integers - no fractions.) Next, we assigned values to these variables. Since we never change them, we could have used constants instead. After adding them, we display the result using some of the nice

formatting features of the debug statement. Finally, we tell the processor to stop, which means to sit there and do nothing.

After you feel you understand what is going on here, it's time to write the timer program. It's just a simple combination of some of the things we demonstrated in these two programs. If you find yourself with extra time, you might want to try out some of the other functions. Play around with the Stamp, and enjoy!

The Problem

So far, we needed a computer to see the results of our Stamp programs. Wouldn't it be nice to see the Stamp do something by itself?

Your task for this chapter is to interface the Stamp to switches and lights. Specifically, you should build a system which indicates on an seven segment LED display the number of times a button has been pressed.

What You Need to Know

In order to solve this problem, here are a few things you need to know:

- What is voltage?
- What is current?
- What is an LED and why do I need a resistor?
- How do I interface a switch?
- What is a seven segment display?

What is Voltage?

As you may have noticed in the Stamp documentation, there are commands to set a pin to an output or an input, and when an output, to set it high or low. Output and input are pretty intuitive, but what actually goes high or low?

The short answer is that setting a pin high means that it is driven to 5 volts, and when it is low, it is driven to 0 volts. But what does that mean? What is a volt? To answer that question, we'll have to delve into a little science.

Voltage is related to potential energy. From physics, you might recall potential energy as that stuff you got when you lifted something off the ground. Let go of that something, and it falls back to the ground, releasing that potential energy you stored up in lifting it in the first place.

Instead of lifting a weight off the ground, imagine you had a positive charge and a negative charge. (What's charge? We have no idea. You can only go so deep...) As opposite charges, they attract each other. If they are initially together, and you start to pull them apart, your effort is being stored as potential energy. If you let go, the charges jump back together, just like the weight crashing to the floor.

To calculate the potential energy of the weight, you would measure the distance it was raised, and plug into the equation $E = mgh$. For the charges, the quantity analogous to height is voltage. In physicist's terms, voltage is a potential. To calculate the potential energy, you must multiply the voltage by the amount of charge raised to this potential. In other words, voltage is a measure of potential energy per unit charge. Think of voltage as the electronic "height" of some charge.

Let's go back to the weight for a moment. If you lift it off the ground, your work has given it some amount of potential energy. While you are holding it up in the air, an annoying friend, we'll call him Paul, shoves a table under your hand. If you were to let go of the weight, it would only fall a little bit down before it hit Paul's table. That doesn't release much energy. So the question Paul asks you is, if your weight will release less energy when you drop it, doesn't it now have less potential energy? And if it does, where did this energy go? Did it just disappear?

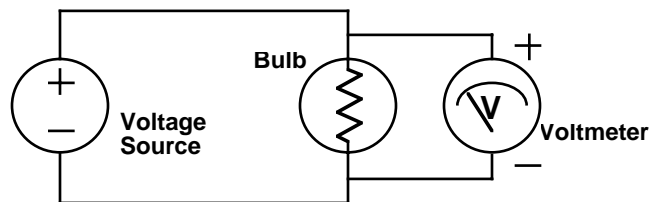
Rather than risk arrest from the Thermodynamic Police, you quickly point out that potential energy is always measured between two points. Your weight still has the same potential energy relative to the floor. It is meaningless to talk about potential energy without respect to some resting position.

The same is true of voltage - it is always measured between two points. But like the original weight example, sometimes the reference point is not specifically mentioned, but is nevertheless presumed to be the ground. In fact, electrical engineers use this exact same word, ground, to refer to a reference level to which all voltages are compared. So when we say that the Stamp has 0 volts or 5 volts on a pin, that is measured with respect to ground. Which, in the case of the BASIC Stamp 2, happens to be on pin 23, and is labeled “GND”. In more generic terms, voltage is an *across* variable - it is always measured across two points.

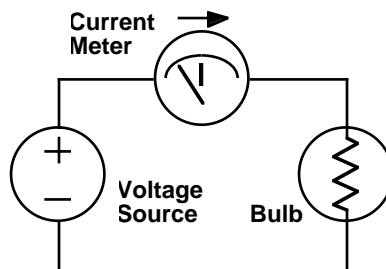
What is Current?

Voltages only tell part of the story. Circuits provide paths for those charges that have been raised to some potential to flow back down to ground. The flow of charge is called *current*. It is measured in *Amperes*, or *amps* for short, and is literally the number of Coulombs of charge that pass a point per second. (A Coulomb is about 6×10^{18} electrons.)

Whereas voltage was an across variable, current is a *through* variable. You measure voltage across a lightbulb by putting the leads of a voltmeter on each side of the bulb. To measure current, you break the circuit, and insert the current meter (sometimes called an Ammeter) into the circuit so that all of the current must flow through it. The two configurations are shown below.



Note that the current meter is labeled “I”. Electrical engineers use the letter “v” when referring to voltages, and “i” when referring to currents. Why “i” for current?



If you're curious, go to the library, and read up on the early history of electrical engineering. (Which is my way of saying that I really don't know why...)

You should also note that both types of meters care about direction. In the earlier figure, the voltmeter has plus and minus signs showing which polarity of voltage gives positive readings. This is a reference direction. It does not necessarily imply that the voltage has the shown polarity. The point is that by having a reference, you will be able to tell by the sign of the readings the true polarity. Of course, flipping the leads of the meter will flip the sign of the results. A similar argument holds for the arrow on the current meter - it is just a reference direction. As we will see, it is essential that you keep careful track of the directions.

There are some basic laws that you can apply to across and through variables. These are called Kirchhoffs Laws, and we will later give lofty definitions for these. But for now, we can easily present where they come from. *Kirchhoff's Current Law*, or *KCL* for short, simply says that current is a conserved quantity - if some amount of current flows into a part of a circuit, the exact same amount must flow out. *Kirchhoff's Voltage Law*, or *KVL* for short, just says that the voltage that you drop on one side of a circuit must be equal to the amount of voltage you raised on the other side. Or in other words, you can only fall the height you were raised. These laws may seem painfully obvious, but when shrouded in mathematics, they appear obscure, yet shockingly powerful.

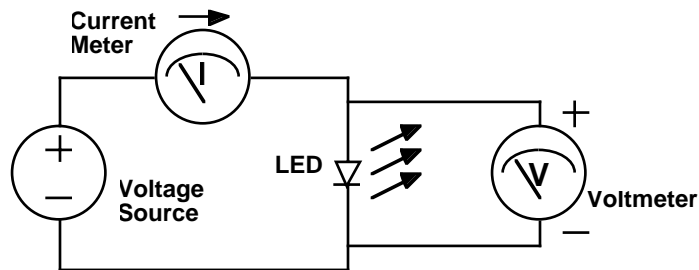
Before we move onto applying our understanding to the problem at hand, we need to mention one more concept - power. As you may recall from basic physics, power is the first derivative of energy with respect to time, i.e. energy per unit time. It is generally measured in *Watts*. If potential energy is just voltage times charge, and current is the first derivative of charge with respect to time, then voltage times current must be power. Or in equation terms, $P = I V$. A circuit element can either pro-

duce power, or dissipate power. To make life easier, we will define P to be positive for things which dissipate power (like lightbulbs), and negative for those things which produce power (like batteries). This may sound a little odd, but it is the standard convention. To make this work out, when measuring the current into an element, we define positive current flow as that going into the positive terminal. Although this choice was initially arbitrary, you must follow it meticulously. Failure to do so is the most common source of errors for people first learning how to solve circuits.

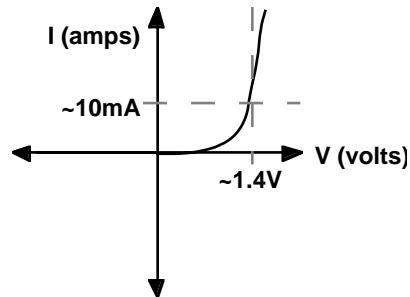
What is an LED?

A Light Emitting Diode, or LED, can be thought of as a semiconductor light bulb. They come in a variety of colors - red, green, orange, yellow, and occasionally, blue - and they are ubiquitous. From power lights to brake lights, LEDs are everywhere.

LEDs are nonlinear devices. What is meant by that is that as you vary the voltage across them, the current does not vary proportionally. In fact, it varies exponentially! Here's a schematic of an experiment where we will vary the voltage across an LED, and measure the current through it. Note the symbol for the LED. The arrows indicate the light coming from the LED. By varying the voltage source and noting the meter readings, we produce something called an I-V curve which is shown below.



Examine the shape of the I-V curve. It has some rather interesting properties. First, you will notice that there is essentially no current flow for negative voltages. For small positive voltages, there is a very small current flow, which rapidly increases at higher voltages. For most LEDs, this rapid increase in current generally occurs



around 1.4V, although this varies widely depending upon the type of device. Near this turn-on voltage, very small changes in voltage can cause dramatic changes in current.

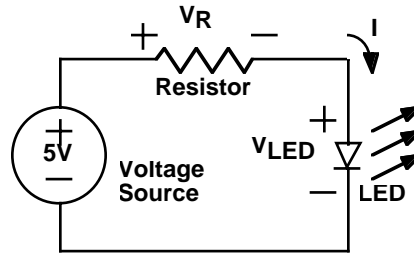
For an LED to operate properly, we would like to provide a current somewhere between 1 to 10 mA. (A milliamp, or mA, is 1/1000 of an amp.) From our BASIC Stamp 2, we have a 5 V supply. If we put this across our LED in the positive direction a massive current would flow, possibly damaging the LED, the Stamp or/and the power supply. (This might well produce what Bob Pease, celebrated circuit guru, calls a DED - Darkness Emitting Diode.)

A very bad idea is to try and build a voltage source which is at just the right voltage. The reason this idea is so bad is that the I-V curve, and thus the required voltage, vary depending upon the particular LED, the temperature, the phase of the moon, etc. and slight shifts cause dramatic changes in current. It would be much better to set up the circuit such that these minor variations do not cause substantial current changes.

It turns out that a circuit which accomplishes this goal is very simple. The idea is to use a *resistor* to set the current. A resistor is a linear device - its I-V curve is a straight line with current proportional to voltage. The constant of proportionality is called the *resistance*, and is measured in *ohms*, which is often denoted by the Greek letter Omega (Ω). You may have heard of *Ohm's Law* which states that $V=IR$. This is not so much a law, as an empirical observation which applies well to most, but not all materials. (Obviously, it does not apply to LEDs!)

The preferred LED circuit is shown below. It consists of three circuit elements: a 5V power supply, a resistor and an LED. Note that we have labeled the voltage across the resistor as V_R , and the voltage across the LED as V_{LED} . We have also

labeled the current flowing through the circuit as I . Like a collection of pipes tied end-to-end in a circle, the way the circuit is wired demands that the exact same current flows through the voltage source, the resistor and the LED. (Remember KCL?) So there is only one current label.



KVL is useful in analyzing this circuit. If we start with 5V from the voltage source, then we must drop 5V in the rest of the circuit. Clearly, this is done in two steps - V_R across the resistor, and then V_{LED} across the LED. The sum of these two must equal 5V. In equation form:

$$V_R + V_{LED} = 5V$$

We can also write equations corresponding to the I-V curves for a resistor, and for an LED. For the resistor:

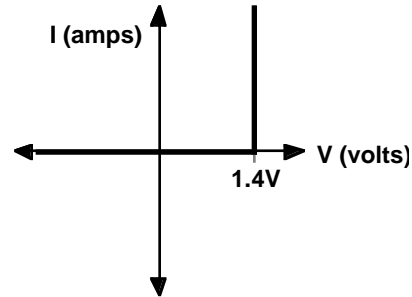
$$V_R = IR$$

The LED I-V curve is an exponential. A reasonable equation for it might be:

$$I = I_0 \left(e^{\frac{V_{LED}}{V_T}} - 1 \right)$$

where I_0 and V_T are constants. This is kind of ugly, so we will make a fairly reasonable simplification. Looking at the LED I-V curve, you can approximate it with two straight lines as shown below:

For $I > 0$, the equation becomes $V_{LED} = 1.4V$, which is the vertical segment of the curve. Since we want some current to flow in our circuit, we will be operating the LED on this portion of the curve.



We now have all the equations we need to complete our analysis of this circuit. Plugging in our value for V_{LED} yields:

$$5V = V_R + V_{LED} = V_R + 1.4V$$

Solving for the resistor voltage, we get:

$$V_R = 5V - 1.4V = 3.6V$$

Finally, we can plug this back into the Ohm's Law equation for the resistor to find what size resistor we need to get $I = 10\text{mA}$.

$$R = \frac{V_R}{I} = \frac{3.6V}{10\text{mA}} = \frac{3.6V}{0.01A} = 360\Omega$$

Thus, by using a resistor of 360Ω , we will get a current of 10mA .

The remaining question is, did this solve our original problem about sensitivity to the variation in the LED turn-on voltage - the point where the I-V curve turns upwards. To get at this, we must write the equation for the current, I , in terms of this voltage which we shall denote, V_D .

$$I = \frac{V_R}{R} = \frac{5V - V_D}{360\Omega}$$

From this equation, we can calculate how the current changes with the LED turn-on voltage. For example, if V_D varied from 1.3 to 1.5V , then the current would vary from 10.3mA to 9.7mA . This is an acceptable variation.

We can use this wonderful little circuit with the Stamp. Any of the 16 I/O pins can serve as a voltage source which can be programmed to deliver either 0V or 5V using the “high” and “low” commands. Of course, these pins are not ideal voltage sources, but at these current levels, they will be quite adequate. (Please keep the current supplied by any one output pin below 20mA to avoid overstressing the Stamp.)

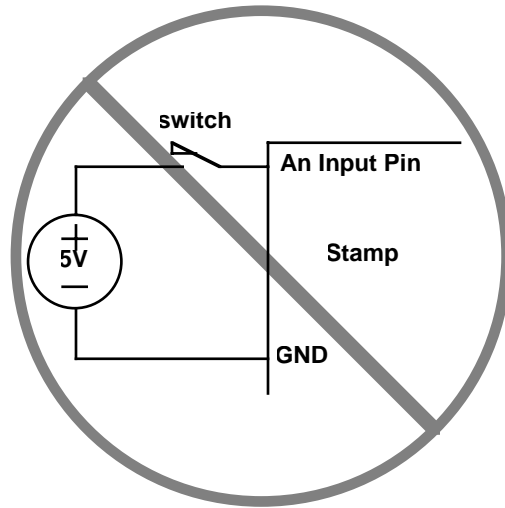
Before you can wire this up in the lab, you need to know a little bit more about the components. Typically, LEDs have two leads, one longer than the other, and one side of the case is flattened slightly. These are used to indicate which are the positive and negative (or, as they are called in the case of diodes, the *anode* and the *cathode*, respectively). Unfortunately, not all manufacturers have uniformly agreed upon which is which. Typically, the lead near the flattened side is the negative one. However, a better way to tell is to look inside the LED. The lead with the “flag” is almost always the negative one. I say “almost always” because there are so many new LED structures coming out now (including organic ones!) that I’d be afraid to call anything an absolute certainty.

Hopefully, you have access to various size resistors, along with a chart explaining the resistor color codes. However, we will be using DIP (Dual-In-line-Package) and SIP (Single-In-line-Package) resistors, which might make your wiring job easier. The way the resistors are connected in these sorts of packages varies widely. You should have DIP resistors which are electrically separated, with each connecting to pins horizontally across the package. In addition, you will need SIP resistors which have one side of each resistor connected to pin 1, with the other sides connecting separately to the remaining pins. Usually, the resistor values are all the same in a given package, and this is probably stamped as a three digit code on the top somewhere. The first two digits are read directly, followed by some number of zeros indicated by the third digit. For example, 102 is a 1000 Ω (1k Ω) resistor.

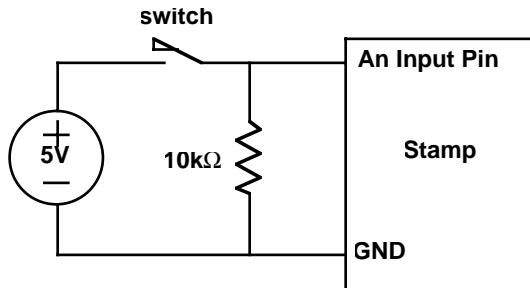
How Do I Interface a Switch?

To get signals into the Stamp, you must provide valid logic levels of either 0V or 5V. Given a simple switch which is either open or makes a connection, how do you do this?

You might think you could use the following circuit:



This circuit will not work well. The problem is beyond the scope of our current discussion, so we will leave it to say that the input must be actively driven to both 5V and 0V. With the addition of a single resistor, we can make this happen as shown below:



The resistor provides a path to ground, removing any residual charge on the input pin that could leave it floating at 5V. This solves the electrical problem.

Unfortunately, we still have a mechanical problem. When the two contacts of the switch hit together, they tend to bounce off of each other a few times before settling down. It's very similar to what happens when you drop a ping-pong ball on a hard surface. Most mechanical switches have this bounce problem. However, you probably never noticed it because the bouncing generally only lasts about 1/100th of a second. The problem is that the BASIC Stamp is plenty fast to see each of these

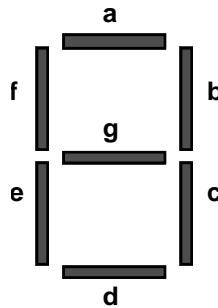
bounces as separate hits. So if you are trying to create a system which counts switch presses, you might count individual presses as multiple ones.

Solutions to this problem are known as *debouncing*. While you can do this in hardware, the Stamp can do it in software much more efficiently. The idea is to ignore the switch for some amount of time after the initial hit to allow the bouncing to stop. The “button” command does this automatically. See the Stamp documentation for more details.

What is a Seven Segment Display?

At this point, we’ve figured out how to turn on LEDs using current-limiting resistors, and how to interface a switch. However, the initial problem called for a seven segment LED display. What is that?

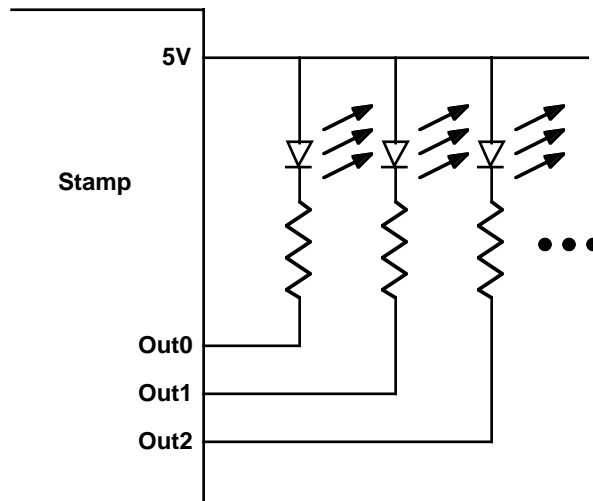
In fact, you’ve probably seen them many times on cash registers, on digital clocks, on the front panel of personal computers indicating the clock speed, etc. They are a collection of seven LEDs (not counting the decimal places) combined in a single package, and arrange to display numbers. The typical arrangement is shown below:



By turning on the appropriate segments, the digits 0-9 can be displayed. For example, to display the number 3, segments a, b, c, d, and g would be lit. Many displays also have LEDs in the appropriate place for decimal points (often labeled “dp”).

With all these LEDs, the number of wires could get very large. Since one side of all the LEDs will generally be connected to ground (or 5V in some cases), these connections are made internally. So there should be a common connection, and one for each segment.

Our displays are *common anode*. The common connection is the positive side (anode) of the LEDs. This changes the circuit slightly. Instead of having the Stamp output supply the 5V, we will connect to the 5V supply and have the Stamp output pull down to 0V. This really is equivalent, because all that matters is the voltage difference. The schematic below shows the connections:



With this scheme, you set an output low in order to turn on the corresponding LED. When an output is at 5V, the LED is off because there is no potential difference across the circuit - no current can flow.

Because you will be driving many LEDs, you may find it useful to learn some of the Stamp instructions which will make it easy for you to store patterns, and output them to the pins. The “lookup” command makes it easy to select a pattern which you can use to drive the pins. This works particularly well when the patterns are specified in binary, and you define a variable which is the pins.

Here’s an example program which will light up one LED, two LEDs, three LEDs, four LEDs, and repeat that pattern forever. With your Stamp manual in hand, study this program and make certain you understand how it works.

```
`Flashing some LEDs  
  
`Paul H. Dietz  
`9/2/96
```

```
`-----`
`This program will turn on pin 0, pins 0 and 1, pins
`0 - 2, and finally pins 0 - 3. This cycle repeats
`forever.
`-----`

`First, define the variables
cnt    var    byte    `for/next counter

`Define a variable which is really the LED outputs
LEDS   var    OUTA    `I could have just used OUTA...

`Set pins high (LEDs off) to start...
`(This also sets them to outputs!)
high 0
high 1
high 2
high 3

`Main program loop...
loop:  for cnt = 0 to 3
        `lookup pattern, and display...
        lookup cnt,[%0111, %0011, %0001, %0000], LEDS
        `wait...
        pause 500
        next

goto loop
```

Where Do We Go Next?

This chapter looked at ways to interface the Stamp to lights and switches. To really understand what was going on, we had to get a little down and dirty, looking into the physics of what voltage and current are. Gratefully, we can leave the physics portion of this class behind, at least until we get to capacitors.

The real world consists of more than on/off lights and switches. Most frequently, we will be dealing with quantities which vary over a continuum. In the next several chapters, we will be exploring ways of having the Stamp's on/off world talk with our continuous world. In the process, we're going to be learning quite a bit about circuit theory. Hold on to your hats!

The Problem

The Stamp is a digital device. It talks in terms like on/off, true/false, 1/0, high/low, or yes/no. But what about that gray area in-between? The real world is filled with quantities that span some continuum. It's not always yes or no - sometimes the answer is maybe.

Your task for this chapter is to build a simple 3-bit digital to analog convertor using your Stamp and a resistor network. With three bits, you should be able to select eight different voltage output levels. These levels should be equally spaced. To select a level, you should have an up switch and a down switch which increment or decrement respectively the current setting. A seven segment display should indicate which level has been chosen. Verify the functionality of your circuit using a digital multimeter.

Note: You will probably be able to reuse a great deal of your wiring from the previous chapter!

What You Need to Know

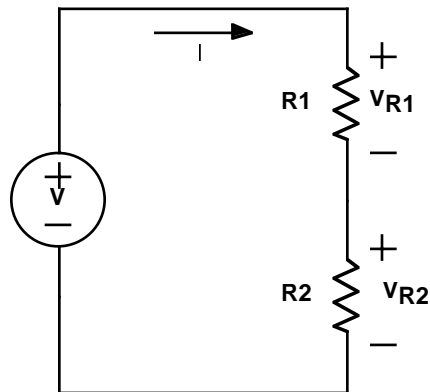
In order to solve this problem, here are a few things you need to know:

- What is a voltage divider?
- How do I solve more complex resistive circuits?
- Are there any tricks that can make this easier?
- What is an independent source and what is superposition?
- What is a digital to analog convertor?

What is a Voltage Divider?

The Stamp lives in a digital world of 5V or 0V. But what if you wanted to use the Stamp to generate voltages in-between? How do you go about this?

In the previous chapter, we saw that a voltage drop occurs as current flows through a resistor, and that the magnitude of this voltage is given by Ohm's Law. By adjusting the current, we change the voltage. We also saw that a resistor could be used in the LED circuit to help set the current. This suggests a circuit utilizing two resistors as shown below.



This circuit is known as a voltage divider. From KVL, we know that $V_{R1} + V_{R2}$ must equal V . By adjusting the two resistor values, we can change the split of how much of V drops across each resistor, but the sum of the voltages remains the same.

Let's solve the voltage divider circuit. Because this network consists of a single loop, the current, I , must be the same everywhere. (That's KCL in action!) Each of the resistors obeys Ohm's Law, giving two equations: $V_{R1} = I R1$ and $V_{R2} = I R2$. Substituting these into our KVL equation, we get:

$$V = V_{R1} + V_{R2} = IR1 + IR2 = I(R1 + R2)$$

Turning this around, we get an equation for I :

$$I = \frac{V}{R1 + R2}$$

This looks an awful lot like Ohm's Law. In fact, two resistors stacked end to end forcing equal currents, acts just like a single resistor with a resistance equal to the sum of the two. This is called a *series* connection. *Resistors in series add.*

We can plug this current back into Ohm's Law for each resistor to find the voltage across them. For $R1$:

$$V_{R1} = IR1 = V \frac{R1}{R1 + R2}$$

Similarly, for $R2$ we get

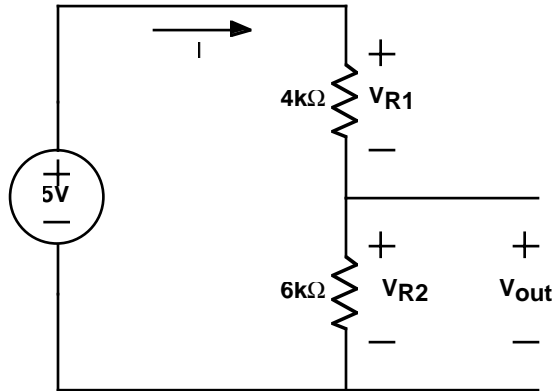
$$V_{R2} = IR2 = V \frac{R2}{R1 + R2}$$

As expected, V is divided between the two resistors. You should note that the bigger resistor will have the larger voltage drop. (With the same current flow, a bigger resistor has a bigger voltage drop.) If the two resistors were the same size, you would get the same voltage drop (half the supply voltage) across each of them.

While we have calculated the voltage drop across each resistor, these may not be the voltages we are seeking. In most circuits, we like to talk about the voltage at various points in the circuit with respect to ground. (We are always very respectful of ground.) The decision of what to call ground in this circuit is fairly arbitrary, but the negative terminal of the voltage supply would be a customary choice. From here, there are really only two other points to measure to - the positive of the supply, and in-between the two resistors. (Although we never stated it explicitly, all terminals directly connected by ideal wires are at the same voltage. We call these *nodes*. The voltage divider circuit has three nodes, counting the ground node.) Of course, the voltage across the supply is just the supply voltage, so that leaves only

one interesting voltage to measure - the voltage in-between the resistors. With respect to our ground, this is just V_{R2} .

Let's consider a specific example. In the circuit shown below, we wish to calculate the voltage, V_{out} . (In the figure, we show the test wires coming out where we would measure V_{out} . Of course, V_{out} is identical to V_{R2} .)



We begin by noting that the resistors are in series. Applying our result from before, we know that resistors in series effectively add together. This gives us 5V across a 10kΩ resistance, yielding a current, I , of 0.5mA. This current flows through each resistor. The voltage across the first resistor is given by $V_{R1} = I R1 = 2V$. For the second resistor, $V_{R2} = I R2 = 3V$. This is V_{out} . We could also have calculated this directly from the voltage divider equation:

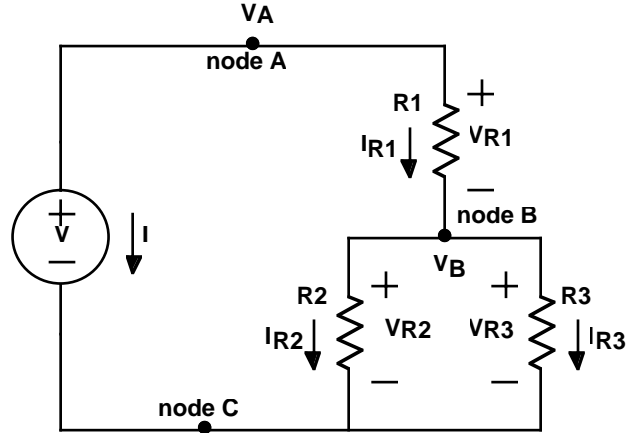
$$V_{out} = V \frac{R2}{R1 + R2} = 5V \frac{6k\Omega}{6k\Omega + 4k\Omega} = 3V$$

It is also good to verify that KVL works. $V_{R1} + V_{R2} = 2V + 3V = 5V$.

How Do I Solve More Complex Resistive Circuits?

By using a voltage divider, we can generate virtually any voltage between 0V and 5V using two carefully chosen resistors. However, the problem for this chapter asks us to build a system that will produce a number of different voltages under program control. Before we can do that, we need to learn more about network theory.

In general, a circuit consists of a network of interconnected elements. Nodes are the points where the elements connect together. *Branches* refer to the paths currents can take through the circuit. Consider the circuit shown below:



This circuit has three nodes, labeled A, B and C. Since we measure voltage between two nodes, let us choose node C as our ground. That leaves only two voltages to measure - the voltage between nodes A and C, which we'll call V_A , and the voltage between nodes B and C, V_B . The voltage between nodes A and B could also be measured, but it will just be the difference $V_A - V_B$. In general, a circuit with n nodes has only $n - 1$ unique voltages associated with it. The voltages across each element can be expressed in terms of these unique voltages. For example, $V_{R1} = V_A - V_B$, and $V_{R2} = V_{R3} = V_B$.

Let's reconsider Kirchhoff's laws, and state them more formally.

Kirchhoff's Current Law: The sum of currents into or out of a node must be zero.

Kirchhoff's Voltage Law: The sum of the voltages around any closed loop must be zero.

To illustrate KCL, consider node B. The sum of the currents into node B are:

$$I_{R1} - I_{R2} - I_{R3} = 0$$

Note the signs of the currents for R2 and R3 in the equation. These are negative because these currents are defined as flowing out of node B.

We can put this equation in terms of voltages by applying Ohm's Law. This gives:

$$I_{R1} - I_{R2} - I_{R3} = \frac{V_A - V_B}{R_1} - \frac{V_B}{R_2} - \frac{V_B}{R_3} = \frac{V_A}{R_1} - V_B \left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} \right) = 0$$

This can be rearranged, yielding:

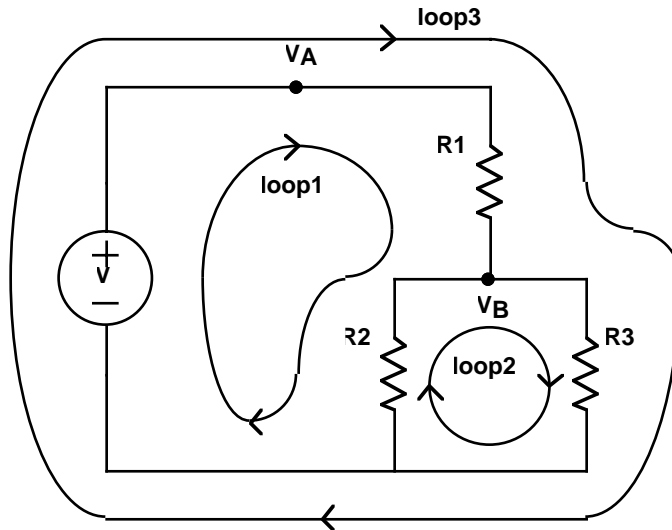
$$\frac{V_A}{R_1} = V_B \left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} \right)$$

Noting that $V_A = V$, we can solve for V_B :

$$V_B = \frac{V}{R_1} \left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} \right)^{-1} = V \frac{R_2 R_3}{R_1 R_2 + R_2 R_3 + R_1 R_3}$$

Knowing V_B , we can now easily solve for all circuit variables. Since such calculations are tediously straight forward, we will leave them as an exercise for the reader.

Just as there are many ways to skin a cat, there are many ways to solve a circuit. KVL provides us with another way of arriving at a complete set of equations. There are three distinct loops in the circuit. These are illustrated in the figure below. For



each of these loops, we can write a KVL equation. In the case of loop1, the equation contains terms for the voltage source, R1 and R2. It is:

$$-V + V_{R1} + V_{R2} = 0$$

Note the sign of the first term. As we go around the loop, we hit the negative end of the supply first, so this term has a negative sign. For loop2, we get the simple equation:

$$V_{R3} - V_{R2} = 0$$

Applying Ohm's Law, these equations become relations in terms of currents. Specifically:

$$V_{R1} + V_{R2} = I_{R1}R1 + I_{R2}R2 = V$$

and:

$$I_{R3}R3 = I_{R2}R2$$

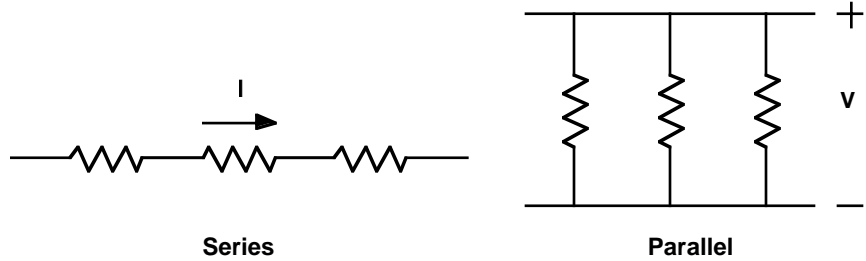
Combining this with the KCL equation $I_{R1} = I_{R2} + I_{R3}$ gives us three equations in three unknowns. Solving these equations will give the same solution as we obtained earlier, but only after considerable manipulation. In fact, extensive use of loop equations is avoided because they almost always result in more cumbersome algebra.

Are There Any Tricks That Can Make This Easier?

By now you must be thinking that solving circuits is a pretty tedious business. In the most general cases, it is. However, if you learn a few tricks, most practical circuits can be solved very rapidly. For those that do not succumb to these techniques, there are computer programs which can quickly solve any resistive circuit.

One of the most basic tricks we have already seen. Resistors in series add. We really only showed this for two resistors, but it does apply to any number of resistors in series. (You can convince yourself of this with an iterative argument: combine the first two resistors into one, then combine that with the third, and so on.)

The *dual* of a series connection is a *parallel* connection. Two resistors are in parallel if they both connect to the same two nodes. The example circuit of the last section had R2 and R3 in parallel. In a series connection, all elements have the same current running through them. In a parallel connection, all elements have the same voltage across them.



It turns out that there is a convenient formula for calculating the effective resistance of a bunch of resistors connected in parallel. Given the same voltage, V , across all the resistors, the total current is the sum of the individual resistor currents:

$$I = \frac{V}{R_1} + \frac{V}{R_2} + \frac{V}{R_3} + \dots$$

Bringing the resistances to the other side of the equation produces:

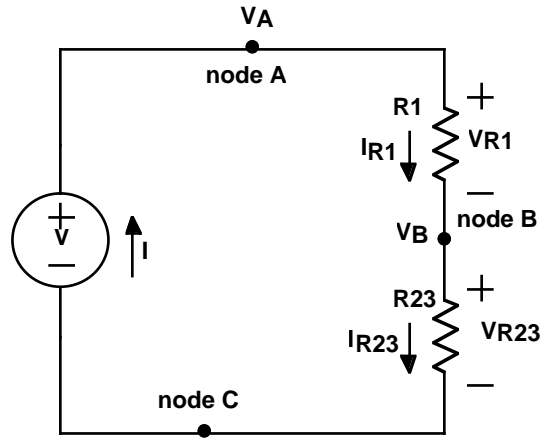
$$V = I \left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \dots \right)^{-1}$$

So the effective resistance is the reciprocal of the sum of the reciprocals of the resistances. (Try saying that quickly ten times...) This relationship is so useful that it is worth wasting the brain cells to commit it to long term memory.

Using these two tricks, we will now demonstrate how quickly they can be used to solve a circuit. Once again, we consider our example circuit from the previous section. In that circuit, R2 and R3 were in parallel. In terms of the rest of the circuit, these two resistors look like a single resistor of value:

$$R_{23} = \left(\frac{1}{R_2} + \frac{1}{R_3} \right)^{-1} = \frac{R_2 R_3}{R_2 + R_3}$$

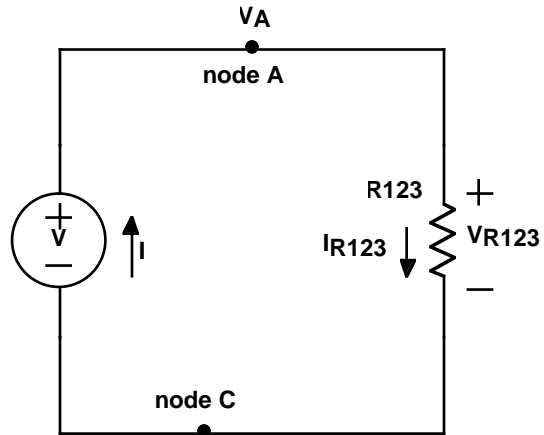
We can redraw the circuit with this effective resistance, R_{23} , replacing R2 and R3 as shown below.



Examining this schematic, we can now see that $R1$ is in series with $R23$. We can now replace these two resistors with their effective resistance:

$$R_{123} = R1 + R23 = R1 + \frac{R2R3}{R2 + R3}$$

This has now become a trivial circuit that can be solved by inspection.



The current, I , is:

$$I = \frac{V}{R_{123}} = \frac{V}{R_1 + \frac{R_2 R_3}{R_2 + R_3}} = V \frac{R_2 + R_3}{R_1 R_2 + R_2 R_3 + R_1 R_3}$$

Presuming we wanted to calculate V_B , we have a number of methods available. For example, we could calculate the voltage drop across R_1 , and subtract this from V . Another possibility would be to directly use the voltage divider equation. This would be:

$$V_B = V \frac{R_{23}}{R_1 + R_{23}} = V \frac{\frac{R_2 R_3}{R_2 + R_3}}{R_1 + \frac{R_2 R_3}{R_2 + R_3}} = V \frac{R_2 R_3}{R_1 R_2 + R_2 R_3 + R_1 R_3}$$

which is the same answer we obtained previously.

What is an Independent Source and What is Superposition?

In journalism, an independent source is someone that can verify information in a fashion that does not depend upon your other sources. In circuits, an independent source is a voltage source or a current source which does not depend upon other circuit currents or voltages. Yes, this is a horribly strained metaphor. Let's put this in simpler terms.

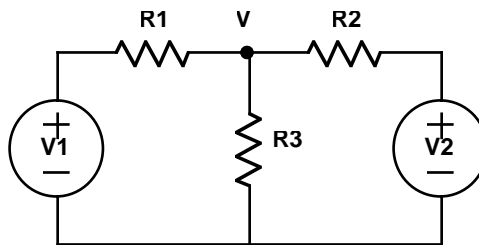
If you had a bag full of batteries, you would have a bag full of independent voltage sources. It's that straight forward. Dependent sources, which we will be discussed in the next chapter, put out a voltage or current depending upon some other voltage or current in the circuit. An example would be an amplifier where the output voltage is so many times the input voltage. But we're getting ahead of ourselves here. Independent sources are just simple power supplies.

We've used independent voltage sources throughout the last two chapters. There are also independent current sources which force a given current through a circuit. Batteries and other inexpensive power sources are mostly voltage sources. However, more expensive lab power supplies can often run as either fixed voltage or fixed current sources. Whereas the I-V curve for a voltage source is a vertical line, a cur-

rent source is a horizontal line. If you think about this, that means that voltage sources are kind of similar to 0Ω resistors (a vertical I-V curve through the origin), and current sources are like infinite value resistors (a horizontal I-V curve through the origin). We'll come back to this point later.

In order to use your Stamp to create different voltages (your problem for the week), you will be combining several output pins into a single circuit. You could solve these circuits using the techniques we have already discussed. However, because you will be using independent sources with a *linear* network, there is another useful trick to learn that will make the analysis much easier.

Consider the circuit shown below:



Using KCL, we can write the equation at V as:

$$\frac{V_1 - V}{R_1} + \frac{V_2 - V}{R_2} = \frac{V}{R_3}$$

Solving for V is straightforward, producing:

$$V = \frac{\frac{V_1}{R_1} + \frac{V_2}{R_2}}{\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}} = \frac{R_3(V_1R_2 + V_2R_1)}{R_3R_2 + R_3R_1 + R_1R_2}$$

If the problem had more nodes and more sources, the algebra gets unwieldy rather quickly. There must be a better way.

If you examine this solution, you will notice that it really has two parts: one due to V_1 , and the other due to V_2 . Imagine for a moment that V_2 was set to $0V$. A voltage

source of 0V is essentially a wire, so we may now consider R2 and R3 to be in parallel. In fact, this is exactly the same problem we solved earlier. The voltage is:

$$V = V1 \frac{R3R2}{R3R2 + R3R1 + R1R2}$$

This suggests a general technique. *A circuit with multiple independent voltage sources can be solved by considering the effect of each voltage source independently, and then adding the solutions together.* When a voltage source is not being considered, it is replaced by a short. This technique is known as *superposition*, and it really does work!

Basically, superposition succeeds because the elements are all linear. Every increment in current in a resistor creates proportional increases in the voltage. So what ever amount each source adds in, adds in linearly, in addition to what all the other sources are contributing.

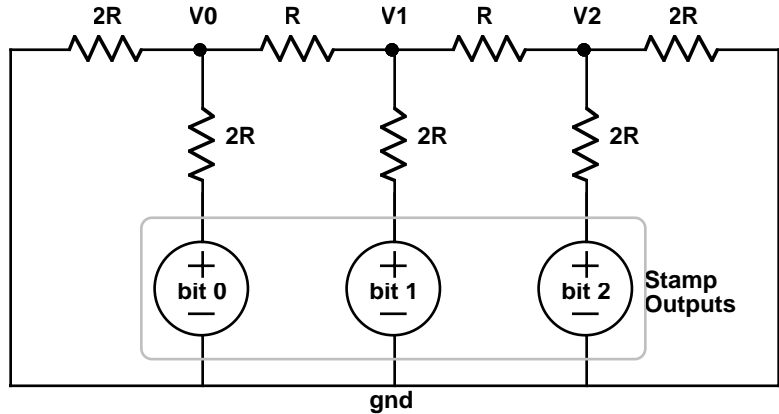
Unfortunately, superposition doesn't always work. It depends upon the circuit being linear. If you have a nonlinear element in your circuit, like an LED, this just plain doesn't work. (Equal increments in voltage across an LED can cause wildly varying changes in current.) Another limitation is dependent sources. We will discuss how to deal with those in the next chapter. For the moment though, you may revel in the elegance of this technique.

What is a Digital to Analog Convertor?

A digital to analog convertor is a device which takes digital inputs, and outputs a number of different voltage levels. For example, in your compact disk player, the 16-bit digital to analog convertor, or DAC, takes the binary data from the disk, and puts out a rapid sequence of voltages chosen from a list of 65536 different levels. These levels are so closely spaced, that they are termed analog - meaning that they are a continuum of voltages rather than simply high or low.

Your goal for this chapter is to build a 3-bit DAC. So you will be using your Stamp to produce various 3-bit numbers (that means eight different possibilities) which will drive a resistor network so as to produce different voltage levels.

The circuit configuration you will be using is called an R-2R ladder. The basic circuit is shown below. Resistors of two different values are used: R and $2R$. The volt-



age sources shown are the Stamp outputs.

As part of your preparation for this lab, you will need to analyze this circuit. You should calculate V_2 as a function of the Stamp outputs switching from 0V to 5V in the eight possible patterns.

Hint: This circuit has some wonderful properties that make it very easy to analyze. Start at one end, and think parallel and series combinations.

What's Next?

In this chapter, we learned how to generate a range of output voltages. That's called digital to analog conversion. But what if you wanted to go the other way and do analog to digital conversion? That's the subject for the next chapter.

The Problem

In the last chapter, we saw how you could use the Stamp to generate an analog voltage. In this chapter we consider the inverse problem: how do you measure an analog voltage using the Stamp?

Your task for this chapter is to build a simple 3-bit analog-to-digital convertor. Specifically, you will use a potentiometer to generate voltages ranging from 0V to 5V. Your circuit will measure these, and drive a seven segment display with the numbers 0 - 7 to indicate the degree of rotation of the potentiometer.

Note: You will probably be able to reuse a great deal of your wiring from the previous chapter. The heart of your analog-to-digital convertor will be your digital-to-analog convertor!

What You Need to Know

In order to solve this problem, here are a few things you need to know:

- What are the limitations on our DAC?

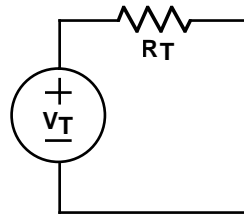
- What is an amplifier?
- How do you build an Analog to Digital Converter?

What are the limitations on our DAC?

In the previous chapter, we built a simple digital-to-analog converter, or DAC, using a network of resistors known as an R-2R ladder. While it is true that this produces a digitally controllable voltage, it is not easy to make use of it. The problem is that if we were to hang a load (such as a resistor) from the output to ground, this changes the network, and hence the output voltage. In this section, we will examine a way of analyzing this problem.

Voltage sources, current sources, and resistors all have I-V curves which are straight lines - they are all linear elements. In fact, it is not too difficult to show that if you were to measure across any two points in a circuit consisting only of linear elements, the resulting I-V curve is linear. In the case of our simple DAC, we have a linear network, and we can safely presume that if we were to measure the I-V curve between the output and ground, we would get some straight line.

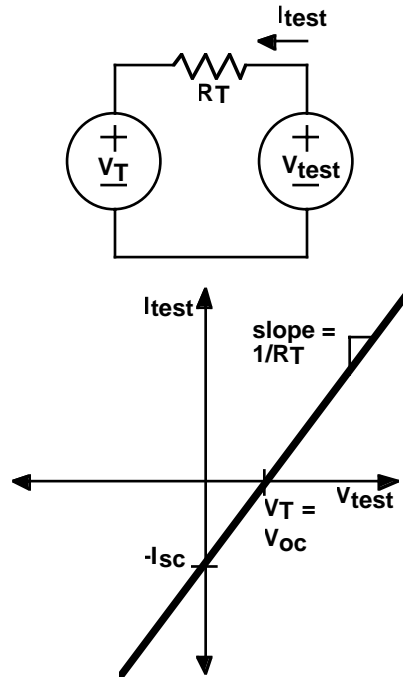
Now here's the kicker - a straight line has only two degrees of freedom. For example, the slope and the x-intercept completely define a line. With the circuit shown below, by choosing the right resistor value and the right voltage, we can produce any linear I-V curve. This is called a *Thevenin equivalent* circuit.



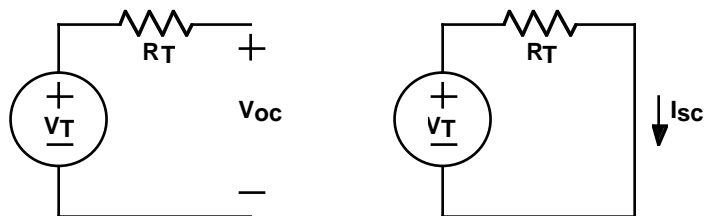
To prove this, let's find the I-V curve produced using this circuit. The test setup is shown below. A test voltage is forced across the two terminals in question, and the current is measured. The current through the resistor will be:

$$I_{\text{test}} = \frac{V_{\text{test}} - V_T}{R_T}$$

Plotting V_{test} versus I_{test} gives the I-V curve shown here.



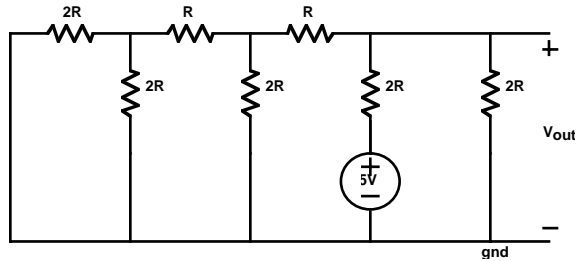
In fact, we could have found this line even more easily. Since it is a straight line, any two points will suffice. Two convenient ones are the short-circuit current, I_{sc} , and the open-circuit voltage, V_{oc} . The short-circuit current is the current which flows when the test voltage is zero - i.e. the current when the terminals are shorted together. By convention, this current is defined as going the opposite direction of I_{test} . (This makes the signs work out more easily.) The open-circuit voltage is the voltage at which no current flows - i.e. the voltage when the terminals are left unconnected. These connections are shown in the next figure. From these two mea-



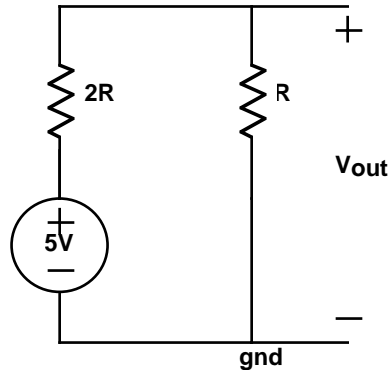
surements, we can quickly conclude that $V_T = V_{oc}$ and $R_T = V_{oc}/I_{sc}$.

The upshot of all this is that the output of any linear circuit, no matter how complex, can be modeled by a trivial Thevenin equivalent circuit consisting of a single voltage source and resistor. Linearity is a wonderful thing!

Let's consider modeling our DAC from the last chapter. We'll consider the case when the MSB (Most Significant Bit, i.e. the biggest one) is on, but the other bits are off. The circuit, redrawn for this particular case, is shown below.

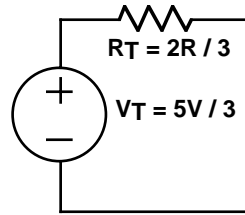


Applying our simple resistor combination tricks, this circuit quickly reduces to one with only two resistors.



We can now easily find the short-circuit current, and the open-circuit voltage. In the short-circuit case, R is shorted out, so $I_{sc} = 5V/2R$. For the open-circuit case, we can use the voltage divider relationship to quickly calculate $V_{oc} = 5V R/(2R+R) = 5V/3$. The Thevenin resistance, $R_T = V_{oc}/I_{sc} = (5/3 V) / (5V/2R) = 2R/3$. The Thevenin equivalent circuit is shown below. Looking into the terminals, this circuit behaves identically with the far more complex circuit we started with.

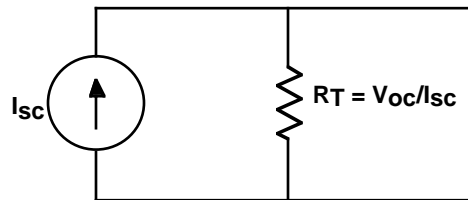
This simple Thevenin equivalent circuit lets us quickly evaluate what would happen with different loads on the output of our DAC. For example, placing a load resistor



of $2R/3$ on the output would cut the output voltage in half! Lower resistances would depress the output voltage even further. Clearly, the only load resistances that would not grossly affect the output would have to be much bigger than R_T . This is a pretty serious limitation.

The *output impedance* of a linear circuit is just its Thevenin resistance. Basically, this tells us how much current we can draw before we start to change the output voltage significantly. For our DAC, this was not very much current.

Before we leave this topic, we should note that there is another circuit that can create all possible linear I-V curves. It is the *dual* of the Thevenin equivalent circuit, and is called a *Norton equivalent* circuit. This circuit consists of a current source in parallel with a resistor, as shown in the figure. You should convince yourself that



this circuit does exactly the same thing as the Thevenin equivalent.

In case you should ever find yourself wanting a Ph.D. in Electrical Engineering, you should be warned that a favorite qualifying exam question involves these equivalent circuits. You are given two identical looking boxes with two terminals sticking out. One box contains the Thevenin circuit with a one volt source and a one ohm resistor. The other box contains the Norton circuit, made with a one amp current source, and a one ohm resistor. How can you tell which is which? The answer: you can't tell electrically. However, the Norton box has power dissipating in the resistor. So you feel the boxes, and the one that is getting hot is the Norton circuit.

What is an Amplifier?

At this point we have a DAC, and we have shown that it has substantial output impedance - i.e. its voltage will change dramatically depending upon the load we put on the output. This is not good. If the DAC is going to be at the heart of our ADC, then we want it to be well behaved, acting more like a real voltage source (which has zero output resistance).

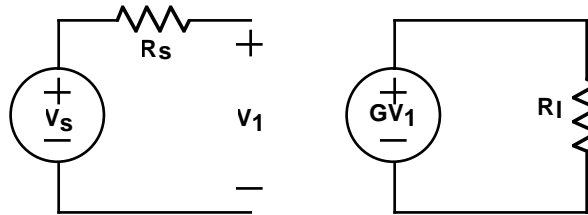
The nifty little gadget which solves this problem is called an *amplifier*. Now, you may have thought you understood what an amplifier was, but you probably didn't think about the units of what got amplified. Ideal amplifiers can take four basic forms. They can produce an output voltage which is a bigger version of the input voltage. Or they might take in a current, and produce an enlarged output current. More confusing still, they might take in a voltage, and produce a proportional current. And finally, they might take in a current, and produce a proportional voltage. These are called voltage amps, current amps, transconductance amps and transimpedance amps respectively.

In terms of network theory, we call all of these *dependent sources*. Their outputs depend upon some other measured quantity. A linear dependent source is one where the output is proportionally related to the input. For now, we will only consider linear dependent sources. This allows us to use the full power of superposition and equivalent circuits.

There is one small catch. When solving circuits using superposition, we considered each independent source separately, suppressing the others. However, this doesn't work with dependent sources. The problem lies in the fact that dependent sources can cause a much richer variety of changes to a circuit. For example, a resistor can be modeled as a dependent source (the current depends proportionally on the voltage). Clearly, removing a resistor from a circuit changes the circuit dramatically, and is not a reasonable option when applying superposition. The moral of the story: never suppress a dependent source when doing superposition calculations.

Let us begin by considering the simplest possible amplifier; one where the output voltage is equal to some input voltage. Consider the circuit below. Here, $G = 1$. This is often referred to as a unity gain buffer.

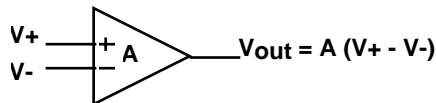
You may have thought that there wasn't much point to an amplifier with a gain of one, but if you examine the figure closely, you'll notice that it has a dramatic effect. If the load resistor, R_L , had been placed across V_1 , a voltage divider would have



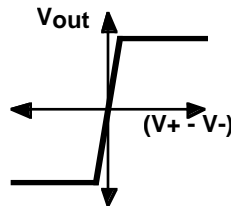
been formed, and the load would have seen only a fraction of the voltage V_s . With the buffer, the load is directly driven by the dependent source, bypassing the source impedance, R_s . No voltage divider is formed, and the load, R_l , sees the full voltage, V_s . In this case, the unity gain buffer allows higher current draw without changing the output voltage. (It allows infinite current in this ideal case. Of course, in the real world, the buffer would also have a small, but nonzero output impedance which would limit the current available.)

By placing a unity gain buffer on the end of our DAC, we solve the problem of loading. However, the bad news is that while unity gain buffer chips are available, we are not recommending that you use one. In fact, unity gain buffers are usually constructed from *operational amplifiers*, and hopefully you have some of these.

Operational amplifiers, or *op amps* for short, are the work horses of the analog circuit world. They measure the voltage difference between two nodes, and output a



voltage which is A times greater. Or at least that is the ideal. The gain, A , is so big, generally 10^6 or more, that even very small input differences quickly overwhelm the output voltage range. In this case, the op amp clips, generally within a few volts of the op amp supply voltage. (Although we don't often show it on these ideal sche-

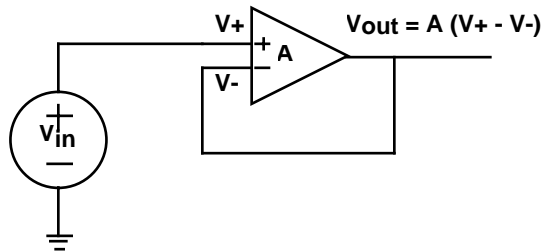


matics, op amps must be connected to a power supply to work!)

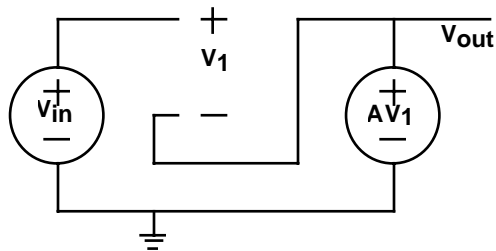
Modern op amps (meaning those built after the 1960's or so) are made from transistors. We're not going to get into a detailed discussion of transistors here, but we had to say something about building precision circuits with transistors. Unfortunately, transistors are awfully inconsistent little beasts. The β of a transistor, which is basically its current gain, can vary wildly from transistor to transistor. This is even true for supposedly identical transistors. (For example, ranges of 50 to 5000 for a given device type are not unheard of!) Circuits built from these devices can be similarly imprecise. So when we say that an op amp has a gain of about 10^6 , it might be anywhere from 5×10^5 to 10^7 , or some similarly broad range. The trick is to use the op amp in such a way so as the precise gain doesn't matter, so long as it is large enough.

The magic that creates precise circuits out of horribly imprecise components is known as feedback. Let's look at an example to see how it works.

Consider the circuit shown below. For an ideal op amp, no current flows into the +



and - input terminals. So we can redraw this circuit using the symbology of dependent sources leaving these connections open. The op amp is a voltage controlled



voltage source. Implicitly, the output is referenced to ground.

The next step is to find the output voltage, V_{out} , as a function of the input voltage, V_{in} . We can write the equation by inspection of the figure.

$$V_{out} = A(V_+ - V_-) = A(V_{in} - V_{out})$$

Collecting the V_{out} terms on the left yields:

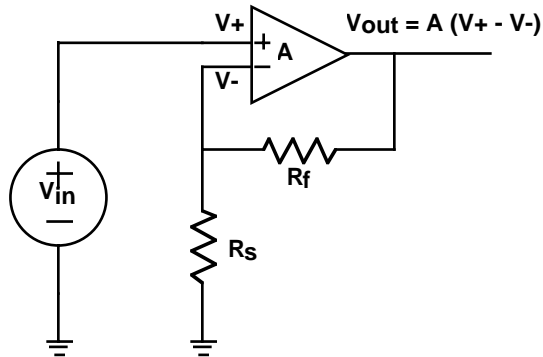
$$V_{out}(1 + A) = AV_{in}$$

This reduces to:

$$V_{out} = \frac{AV_{in}}{(1 + A)} \cong V_{in}$$

for sufficiently large gain, A . Thus, the circuit is a unity gain, noninverting, buffer that is independent of A , so long as A is large.

Let's consider another op amp circuit. Shown below is a more general noninverting amplifier. In this case, instead of feeding back the output to the negative input, we



first reduce the output with a voltage divider, and then feed back the reduced voltage. This has a rather curious effect.

From the voltage divider equation, we can write:

$$V_- = V_{out} \frac{R_s}{R_s + R_f}$$

Plugging this into the equation for V_{out} , we get:

$$V_{\text{out}} = A(V_+ - V_-) = A\left(V_{\text{in}} - V_{\text{out}}\frac{R_s}{R_s + R_f}\right)$$

Collecting the terms yields:

$$V_{\text{out}}\left(1 + \frac{AR_s}{R_s + R_f}\right) = AV_{\text{in}}$$

Solving for V_{out} and approximating for large A gives:

$$V_{\text{out}} = \frac{AV_{\text{in}}}{1 + \frac{AR_s}{R_s + R_f}} \cong V_{\text{in}}\frac{R_s + R_f}{R_s} = V_{\text{in}}\left(1 + \frac{R_f}{R_s}\right)$$

Oddly, the result is an amplifier with gain equal to the inverse of the voltage divider's gain. In fact, there is a general principle here - that you get the inverse of whatever you put in the feedback path. We'll be discussing this more later.

Let's take a closer look at what's going on here. The voltage on the negative terminal of the op amp is:

$$V_- = V_{\text{out}}\frac{R_s}{R_s + R_f} \cong V_{\text{in}}\frac{R_s + R_f}{R_s}\frac{R_s}{R_s + R_f} = V_{\text{in}} = V_+$$

In fact, the positive and negative inputs on the op amp had better be about equal. If this were not the case, the output of the real op amp would be pegged near one of the supply rails. Basically, the feedback is doing all that it can to try and make the voltages on the plus and minus inputs as close to equal as possible. All negative feedback op amp circuits should have this characteristic.

This suggests a fast way of solving op amp circuits. If the op amp is setting the two inputs equal, why don't we just assume that has happened, and solve from there! Applying this technique to the previous circuit, we see that $V_- = V_{\text{in}}$. If no current goes into the op amp inputs, then the current through the two resistors must be equal. This gives:

$$I_f = I_s = \frac{V_{\text{out}} - V_{\text{in}}}{R_f} = \frac{V_{\text{in}}}{R_s}$$

Collecting terms gives:

$$\frac{V_{\text{out}}}{R_f} = V_{\text{in}} \left(\frac{1}{R_s} + \frac{1}{R_f} \right)$$

Solving for V_{out} yields:

$$V_{\text{out}} = V_{\text{in}} \frac{R_f + R_s}{R_s} = V_{\text{in}} \left(1 + \frac{R_f}{R_s} \right)$$

Assuming the inputs are equal turns out to be equivalent to saying that A is very large.

Before we leave this section, there is one more thing to say about high gain differential amplifiers. If you look at the input versus output characteristic for an op amp, you will note that if the positive input is bigger than the negative input, the output quickly saturates near the positive rail. Similarly, if the situation is reversed, the output dives down near the negative rail. An op amp without feedback to keep it in the tiny linear region is a *comparator*. A comparator is a device which tells you if one input is bigger than another. While real comparators are differential amplifiers that have been optimized for the comparison task, a generic op amp will work almost as well for our low speed applications. In fact, the symbol for a comparator is exactly the same as that for an op amp.

How do you build an Analog to Digital Convertor?

At this point, we return to our original objective - building an analog-to-digital convertor, or ADC. There are many different ways of creating an ADC. For now, we will use a classic technique called successive approximation.

Have you ever played a game called “Guess the Number”? One person thinks of a number within some range. The other person tries to guess it. After each wrong answer, the guesser is told whether the guess was too high or too low. Eventually, the guesser zeros in on the correct number.

This is exactly how a successive approximation analog-to-digital convertor works. You have some unknown input voltage. Using a DAC, you generate a guess. Then a comparator tells you whether you were high or low. You modify the guess, and continue the game until you home in on the value of the input.

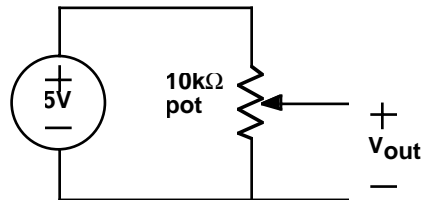
If you are very smart, you play the game cleverly. For example, if you know the number is between 0 and 100, your first guess should be 50. This evenly splits the range, giving you the maximum amount of information from the guess. If you split the remaining range every time, it will take you $\log_2 n$ guesses to find the correct neighborhood of n numbers.

For your assignment, you are to build an ADC which reads a 0V - 5V input, and tells you which of eight ranges it is in. These ranges should be equally spaced and are shown below.

0V .625V 1.25V 1.875 2.5V 3.125 3.75 4.375 5V



The input will be created with a *potentiometer*. That's a big name for a variable resistor. Sometimes you'll hear the older generation refer to them as *rheostats*. Personally, we prefer to just call them *pots*. In any case, these are resistors with connections at the two ends, plus a third connection, called a *wiper*, which can be adjusted from one end to the other. You should have two 10k Ω trim pots. (These small ones are usually used to trim, or make small corrections to a circuit, hence the name.) The wiper is the middle terminal. You will connect this as a variable voltage divider as shown below.

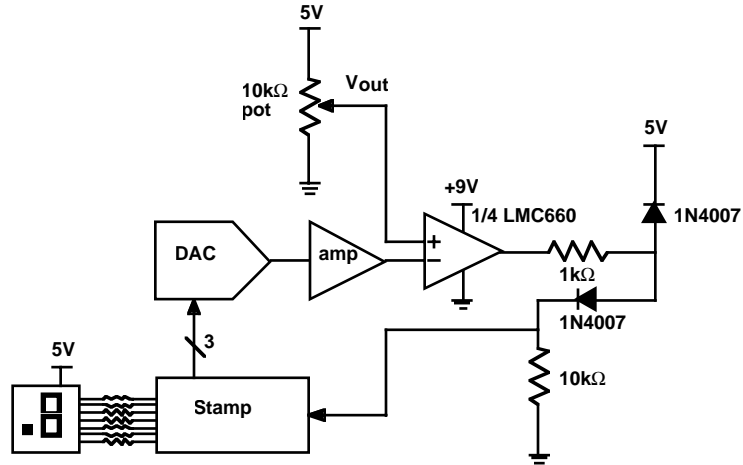


By the way, there are many types of potentiometers. Most are rotary, but some are moved linearly, and are sometimes called *sliders*. On some pots, the resistances change linearly with angle. On others, the resistances change logarithmically. These later ones are used to control volume levels, and are *audio tapered*, as opposed *linear tapered*. When buying potentiometers, make sure you get the right type - both are very common!

(Hearing, like most human senses, is inherently logarithmic. For example, the just noticeable increases in volume from a given level increase proportionally with the

volume level. Audio taper pots allow equal increments on a volume control to sound like equal increments in loudness, even though the power has to increase dramatically more with each further setting.)

So, without further ado, here's the big picture:



The idea here is that you will use the Stamp to choose DAC outputs. These will be amplified up to give the appropriate comparison voltages. (You will have to design this amplifier. Use a noninverting amplifier as discussed in the previous section with resistor values on the order of $10\text{k}\Omega$.) The output of the amp is compared with the potentiometer output. After some protection circuitry, the output of the comparator is sent to a Stamp input.

Unfortunately, another nonideality of op amps is that they will only function properly when the inputs are within the common-mode input range. For the LMC660, this extends below the negative rail up to about 2.5V below the positive rail. So to make this work with a $0\text{V} - 5\text{V}$ input, you will be running the op amp directly off of the 9V battery - not the 5V supply! This gives you far more headroom to play around in. However, this also means the output can swing above 5V , especially since this op amp features rail-to-rail outputs. If you were to connect the output directly to the Stamp, you would run the risk of damaging the Stamp. The two diodes and resistors form a clamp circuit, which holds this output within the 5V supply range. You should double check your wiring before powering up your board!

The op amp chip you will be using is an LMC660, designed by Dennis Monticelli of National Semiconductor. It is a quad op amp, which means that there are four op

amps for you to use in this one package. Please read the datasheet carefully. Not all of it will make sense, but you will most likely find it helpful.

What's Next?

In the next chapter, we will be examining another technique for performing analog-to-digital conversion. However, this new approach will use time varying circuits, including capacitors. This would be a good time to review first order differential equations!

The Problem

As you might have noticed, we are starting to use up most of the I/O pins available on the Stamp. This is unfortunate because we still have more functionality we'd like to build into our systems. The problem for this week is to dramatically decrease the number of I/O pins that are being used, while simultaneously increasing the capabilities of the system.

Like the previous chapter, your task is to digitally display the setting of a potentiometer. However, instead of displaying a single digit, we would like you to build a two digit display which ranges from 0 to 99. If we approached this similarly to the ADC from last week, we would need 14 I/O lines to drive the LED displays, and at least 7 I/O lines to handle the required 7-bit DAC. In addition, we need one more line for the comparator output. That totals to 22 I/O lines - 6 more than we have!

To get around this problem, you will be using a serial-to-parallel driver which reduces the required number of I/O lines for the LED display to just three. That leaves the DAC as the major I/O line hog. While we could use another serial-to-parallel conversion here, we will take an entirely different approach to the problem. A simple circuit will be used to convert the resistance measurement problem into a time measurement problem - something the Stamp does easily with precision.

What You Need to Know

In order to solve this problem, here are a few things you need to know:

- What is a serial interface?
- What is a capacitor?
- How do I use a capacitor in a circuit?
- What is an oscilloscope?

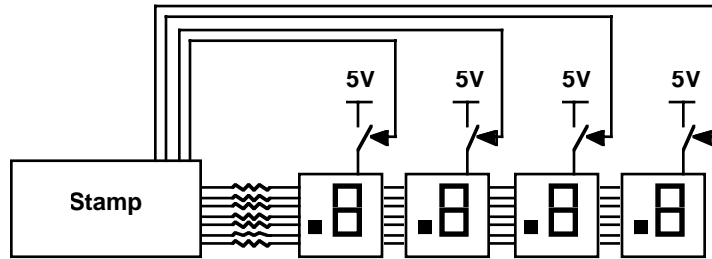
What is a Serial Interface?

When driving the LED displays in our previous problems, we used one Stamp I/O line for each LED segment. To display a particular number, we needed to set seven output lines to the appropriate levels. As long as we wanted the number displayed, those seven outputs had to remain at the correct levels.

At a more abstract level, what we were doing was sending seven bits of information from the Stamp to the display, using seven individual wires. While this was fairly straightforward, this technique does not scale up well to larger displays. Imagine trying to build a calculator with a ten digit display. Counting the decimal points, that would take 80 outputs - far more than are available on our microcontrollers.

The usual solution to this problem is a *multiplexed-display*. To build one, all of the like segments of different digits are connected together. No matter how many digits, this will result in seven (or eight if you count the decimal points) connections. The common terminal of each digit goes to its own electronic switch, which connects it to the appropriate supply (the positive supply for common anode displays, ground for common cathode displays). There is thus one additional connection for each digit in order to control the switches, in addition to the seven segment connections. This is shown in the figure below.

The operation of a multiplexed display depends upon a feature of human visual perception known as *flicker fusion*. If a light is flashed quickly enough, individual flashes become imperceptible, and the illusion of a steady light is created. This is the basis of movies, TV and florescent lighting. All of these flash quickly enough to give the illusion of steady light. (If you ever have the chance to look closely at a movie projector you will see that there is a mechanical shutter which opens and closes several times for each movie frame. This is because flicker fusion requires a



faster repetition than our perception of motion requires.) A good rule of thumb is that any light flashing greater than 50 times a second will appear fairly steady.

Multiplexed displays only show one digit at a time. But by cycling through all the digits repetitively, very quickly, a multi-digit display is perceived. To drive a particular digit, its common terminal is connected to the appropriate supply, and the segments are driven as needed for the number desired in that place. None of the other common terminals are connected, so the others are all off.

Unfortunately, the Stamp is a poor choice for creating a multiplexed display. The way it is set up makes it very difficult to cycle through the digits while the processor is trying to figure out what should be displayed. Many microcontrollers support something called *interrupts*, which allow a program to be periodically interrupted to perform other tasks, like driving a multiplexed display. (If you've ever tried to install hardware in a PC, you may have run into *IRQ* problems. *IRQ* stands for interrupt request, and it is how a board signals the processor to stop what it is doing and handle data from the board.) The upshot is that we will not be using a multiplexed display with the Stamp.

Even though we are not using multiplexing, there is still much to be learned from the example. Most importantly, we note that the data, i.e. which segments to light, is sent for one digit, then the next, and the next, and so on. The same data lines are being used to transmit different data for the different digits at different times. Of course, we also have the switches for the commons which indicate which digit the data is for. But the key idea here is *time division multiplexing* - to reuse the same wire at different times to emulate multiple wires.

Reducing this to its most minimal form, we could use a single signal wire (and a ground, of course) and somehow transmit all the required bits one after another. This general scheme is called a *serial interface*. Serial interfaces are very common: even your Stamp talks to the PC via a serial interface! Other examples include

infrared remote controls, modems, Apple's LocalTalk, many PC mice, Ethernet, USB, FireWire (an upcoming standard for connecting peripherals to PCs), etc.

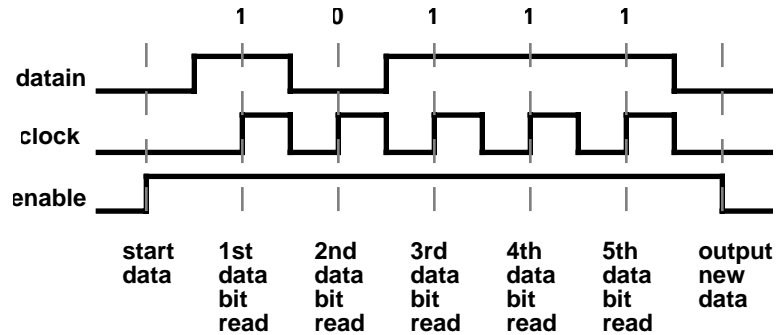
The trouble with serial interfaces is knowing where the data is. For example, before you start sending data, let's presume the line is set low. That's called the *idle channel* condition. If you now wanted to transmit a series of zeros, how would you be able to distinguish this from the idle channel?

The way of approaching this problem is to create a *protocol* which uniquely identifies where the data is. There are many types of protocols. For example, *RS-232*, the standard PC serial protocol, uses a *start bit* which is opposite the idle channel condition to indicate the beginning of data. The receiver sees this start bit, and starts to look for data at specified times after it.

In this chapter, we will be using a National Semiconductor MM5484 16-segment LED display driver. (Unfortunately, National recently obsoleted this part, so you may or may not be able to find it. More on this in a few paragraphs...) This chip receives serial data using a special protocol, and appropriately drives 16 output pins. These outputs each look like switches to ground which can be turned on or off. Once data is received, the switches stay in that state until new data has been received. So unlike multiplexed displays, the LEDs are on continuously, instead of flashing very quickly. The operation being performed by the chip is sometimes called a *serial-to-parallel* conversion because it takes in serial data, and presents multiple outputs at the same time. Find the data sheet for the MM5484, and look it over.

The MM5484 is a fairly simple interface chip. It uses a particularly easy serial protocol which uses two additional wires to indicate when the data is available. One of these wires is the *clock* signal. The other is an *enable* signal. With enable high, each bit is read off of the *datain* signal each time the clock signal transitions from low to high. When all of the data you wish to display at the next moment has been transmitted, enable is brought low, and the output changes at that moment to reflect the new data. This type of protocol, where there are extra signal lines to indicate when to look for the data, is said to be *synchronous*. On the other hand, *asynchronous* protocols must somehow structure the data stream so the data can be found without the extra signal lines, typically a much harder task.

Let's consider an example. In the figure shown below, we are transmitting the data 10111. We begin by setting enable high. Next we present the first data bit, a 1, on *datain*. Switching the clock from low to high causes this bit to be read. However, the display will not be updated until we bring enable low, which we won't be doing



until we've sent all the data bits. We do this by successively placing the bits on datain, and transitioning the clock from low to high. Finally, when all the bits have been received, we set enable low, updating the display. The data will appear at the outputs, setting D1, D2, D3, and D5 to open, and pulling D4 low. (Note that the data goes in most significant bit first.)

This may sound like a complex series of operations to program. Fortunately, your Stamp has special instructions specifically for talking with serial devices that greatly simplify the task. In particular, *SHIFTOUT* is a command which outputs a variable bit by bit on a pin, while creating the appropriate clock signals on another pin. The Stamp manual includes an application note on the use of *SHIFTOUT*, and its counterpart, *SHIFTIN*. Please read this documentation. You will have to program the enable signal yourself, but that is fairly minor. In order to update your display you will have to prepare the data (most likely, using the *LOOKUP* command twice - once for each digit), set enable high, *SHIFTOUT* the data, and finally output the data by setting enable low again.

We strongly suggest that you wire up your two-digit display right away. While it is fairly straightforward, it is as much wiring as you have done in any of the labs to date. It should now be clear why we used common anode LED displays rather than common cathode. Because the MM5484 connects outputs to ground, it only works with a common anode display. Of course, you will still need resistors on each of the outputs to limit the current through the LED segments. Each output in the MM5484 is rated for a maximum current of 15mA, so be careful not to exceed this specification.

As noted previously, you may not be able to find an MM5484. Unfortunately, semiconductor companies don't necessarily make the same parts forever, and this particular National part has been "obsoleted". In other words, National Semiconductor no longer makes this product. This is a major problem for companies that have

designed this part into their systems. Either an alternative must be found, or the system must be redesigned. Typically, semiconductor companies will send out “Lifetime Buy” notices shortly before obsoleting a part. This gives a user one last chance to stock up before the part goes out of production. Rarely, enough new demand surfaces for an obsoleted part that the company goes back into production. Somewhat more frequently, the obsoleted part is transferred to another company, such as Rochester Electronics, which specializes in once discontinued parts. (You’ve got to love their company slogan, “Leaders on the Trailing Edge of Technology”...)

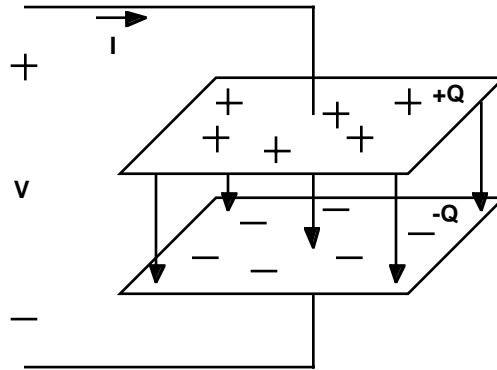
As luck would have it, we have not been able to find a “pin compatible” replacement for the MM5484. However, there are many parts which offer similar functionality. These include: Micrel MM5450, MIC5821, Allegro A6276EA, UCN5832A, and Texas Instruments TPIC6C595. These are all serial input, parallel latched sinking driver output parts, but they vary in number of outputs, interfacing, cascadability, current limiting, and drive capability. As always, examine the data sheets carefully before substituting these parts!

What is a Capacitor?

In comedy, the stock market, and life in general, timing is everything. But up until now, we have carefully avoided talking about timing when analyzing circuits. In fact, all of the circuits we solved fall under the rubric of “DC circuits”, that is, no voltages or currents have been specified as functions of time. However, as we mentioned in the problem statement, we will be building a circuit that will give us a pulse, the width of which will vary according to the pot setting. You can’t do that with just resistors and voltage sources.

In this chapter, we introduce the capacitor, often called a *cap* for short. Unlike the devices we have study so far, it can not be characterized by a simple I-V curve. In fact, if you attached a capacitor to almost any curve tracer, you would get a loop instead of a single line. Clearly, this is not a single valued function, so something weird is going on.

In order to understand this, we’ll have to once again wallow in some basic physics. The prototypical capacitor consists of two parallel conductive plates separated by some insulator as shown in the figure below. Although charges do not actually flow between the plates, placing some positive charge $+Q$ on the top plate will push away an equivalent amount of charge from the bottom plate, leaving it charged to $-Q$. So in fact, from the outside, it looks as if current can flow through the capacitor.



The fundamental relationship for a capacitor is that the voltage across it will be proportional to the charge stored on it. Recall that when we defined voltage, we said it was the potential of separated charges. You might think that adding more charge doesn't change the separation, and thus it wouldn't change the voltage. However, adding additional charges becomes increasingly more difficult because the plate is already charged, and repels the new charges. Thus, adding each new charge takes progressively more work, and this energy ultimately increases the potential.

An analogy with a fluid system might be helpful here. A capacitor is like a tank of water. If you are pumping water in at the bottom of the tank you will need increasingly more pressure as the tank starts to fill. The pressure at the bottom, a measure of the potential, rises with the rising fluid level. In fact, for a tank of constant cross section, the pressure will rise linearly with the amount of fluid in the tank just as the voltage across a capacitor rises linearly with the amount of charge stored in it.

In equation form, the defining relationship for a capacitor is:

$$Q = CV$$

The constant of proportionality, C , is called the *capacitance*, and is measured in *farads*, often abbreviated F. One Coulomb of charge stored on a one farad capacitor produces one volt. One Farad is a huge amount of capacitance. In the lab, we will generally be talking about capacitances ranging from 10^{-12} farads (or simply, one picofarad, written 1 pF, and often pronounced “one puff”) to 10^{-3} farads (generally written as 1000 μ F).

As nice as this equation may be, it doesn't tell us what we want to know: namely what is the current as a function of the voltage. Current is the rate of change of charge with respect to time, so we can take the derivative of this equation to yield:

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

In plain english, the current through a capacitor is proportional to the rate of change of voltage across it. We can rewrite this in integral form as:

$$V = \frac{1}{C} \int i dt$$

This just says that the voltage across the cap will be proportional to the integral of the current into the cap - just like the height of water in the constant cross section tank is proportional to the integral of the fluid flow into it. So the voltage across the capacitor depends upon the whole history of current flow into it. Now you can see why the I-V curve was not just a simple, single valued curve.

An ideal cap does not dissipate energy. No charges actually ever fall back together through the insulator, so there is no way to give up the potential energy within the device. However, energy is clearly stored in the capacitor, in particular, in the electric field between the plates. Imagine a cap which has no charge across it that we then charge up to some voltage, V . We know that the power delivered to the cap will be:

$$P = VI = VC \frac{dV}{dt}$$

Integrating this with respect to time gives the total energy stored in the cap. This is:

$$E = \int P dt = \int VC \frac{dV}{dt} dt = C \int V dV = \frac{1}{2} CV^2 \Big|_0^V = \frac{1}{2} CV^2$$

Before we move on to considering the use of capacitors in circuits, let's figure out how capacitors in parallel and series combine. By definition of parallel, caps in parallel have the same voltage across them. The total charge across two parallel caps is just:

$$Q_p = Q_1 + Q_2 = C_1 V + C_2 V = (C_1 + C_2) V$$

So the capacitance of caps in parallel is just the sum of their capacitances.

For caps in series, they must have the same current going through them. Assuming we started with no initial charge on the caps, this means that the caps must have the same charge since this is just the integral of the same current for each cap. Applying KVL, the total voltage across the series connected caps is:

$$V_s = V_1 + V_2 = \frac{Q}{C_1} + \frac{Q}{C_2} = Q \left(\frac{1}{C_1} + \frac{1}{C_2} \right)$$

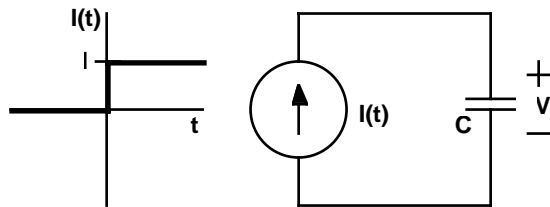
With a slight rearrangement, we get:

$$Q = \left(\frac{1}{C_1} + \frac{1}{C_2} \right)^{-1} V_s$$

So the capacitance of caps in series is the reciprocal of the sum of their reciprocals. You may have noticed that this is just the opposite of what we got for resistors. The easiest way to remember these equations is to note that capacitors in series add like resistors in parallel, and vice versa.

How Do I Use a Capacitor in a Circuit?

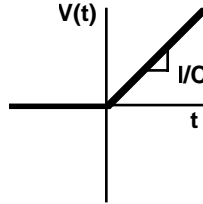
Obviously, we want to use the capacitor in a circuit. The simplest possible interesting circuit consists of a cap and a current source as shown below.



The current source puts out a current which varies over time. In particular, this current source is zero until $t=0$, at which point it instantaneously switches to a current I . By KCL, we know that this must be the current through the capacitor. (We'll be pretending that the current actually passes through the cap, since that's what it looks like from the outside.) Assuming that the cap has no charge across it initially, the voltage across is described by the equation:

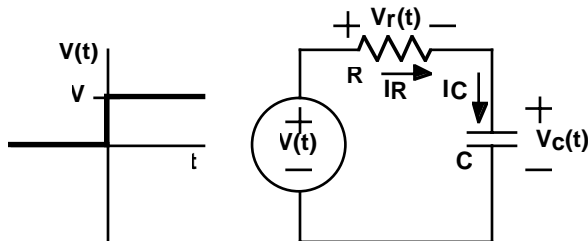
$$V = \frac{1}{C} \int_0^t I dt = \frac{It}{C}$$

So the voltage increases linearly with time, starting at $t = 0$ as shown below.



Of course, as we noted before, the Stamp outputs are well modeled as voltage sources. However, a circuit with just a cap and a voltage source quickly leads to trouble. The problem is that a step change in voltage makes dV/dt go to infinity, requiring infinite current. Caps cannot change their voltage instantaneously! Of course, if we were to hook up a cap to a Stamp output, we wouldn't get infinite current. This is where our model of the Stamp output as a simple voltage source breaks down. In fact, the current would be limited by the small output impedance of the Stamp output that we generally ignore.

To get around this problem, we next consider a circuit with a time varying voltage source, a cap and a resistor as shown below.



By KCL, we know that we have the same current flowing through all of these devices. In addition, we will need to know the initial condition of the cap. We will presume that $V_C(0) = 0$.

Writing the KCL equation at the node shared by the resistor and the capacitor yields:

$$I_R = I_C \Rightarrow \frac{V(t) - V_C(t)}{R} = C \frac{dV_C(t)}{dt}$$

For $t > 0$, $V(t) = V$, allowing us to rewrite this as:

$$V_C(t) + RC \frac{dV_C(t)}{dt} = V$$

From previous study of differential equations, you will hopefully recognize this as a separable equation. Rearranging the terms gives:

$$\frac{RC}{V - V_C(t)} dV_C(t) = dt$$

Integrating both sides and moving some terms yields:

$$\int \frac{RC}{V - V_C(t)} dV_C(t) = \int dt$$

$$-RC \log|V - V_C(t)| = t + K$$

$$\log|V - V_C(t)| = \frac{-t}{RC} + K'$$

where K and K' are constants. Next, we exponentiate both sides giving:

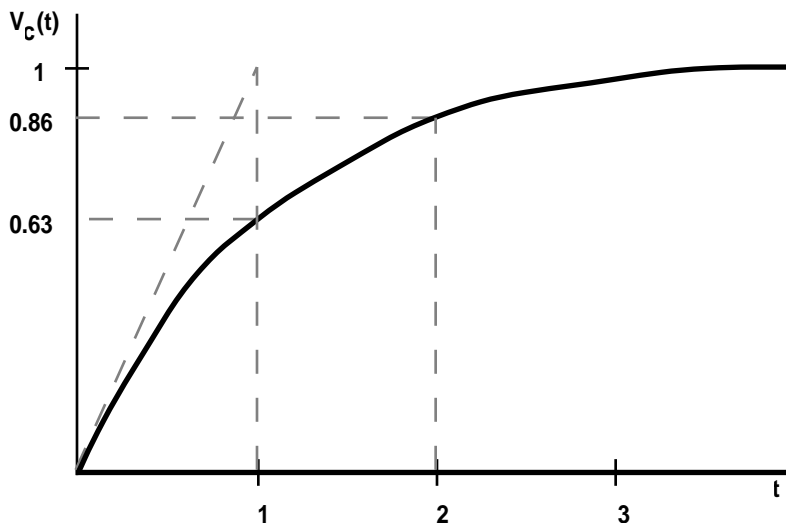
$$V - V_C(t) = K'' e^{\frac{-t}{RC}}$$

$$V_C(t) = V - K'' e^{\frac{-t}{RC}}$$

where K'' is yet another constant. Since the initial condition tells us that $V_C(0) = 0$, then $K'' = V$. The final result is:

$$V_C(t) = V \left(1 - e^{\frac{-t}{RC}} \right)$$

for $t > 0$, and $V_C(t) = 0$ for $t < 0$. A graph of this equation for $V = 1$ and $RC = 1$ appears below.

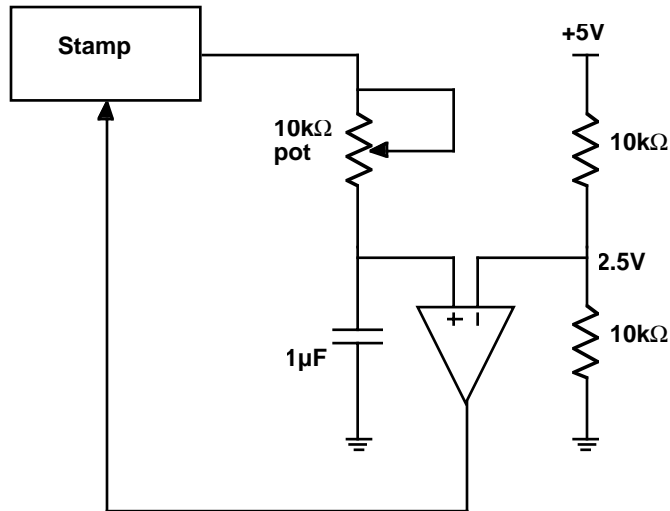


First of all, note the general behavior. At first, the voltage on the capacitor increases fairly rapidly. But this rate of increase slows down, and the curve approaches $V_C(t) = V$ asymptotically. If you think about this, it makes tremendous sense. At first, the resistor sees the full voltage V across it, resulting in a fairly high current to charge the cap. As the cap voltage starts to rise, this reduces the voltage across the resistor. Thus, the current decreases, and the cap voltage increases more slowly. Eventually, the cap voltage is almost identical to V , and the current flow is reduced to almost nothing. This is classic behavior for a first order system.

Second, we should note the role of the term RC . This is called the *time constant* of the system. If you plug this into the equation, you will see that after one time constant, the system reaches about 63% of its final value. After two time constants, it's up to about 86% of the final value. After ten time constants, it's up to 99.995% of its final value.

Another way of looking at the time constant is with respect to the initial slope. If you were to continue a straight line from the beginning part of the waveform, it would intersect the final value at the time constant. This is a handy way of estimating the time constant or the final value from a graph. We recommend that you take a moment to prove this to yourself - it's quite straightforward.

How do you apply this to the problem posed at the beginning of this chapter? It

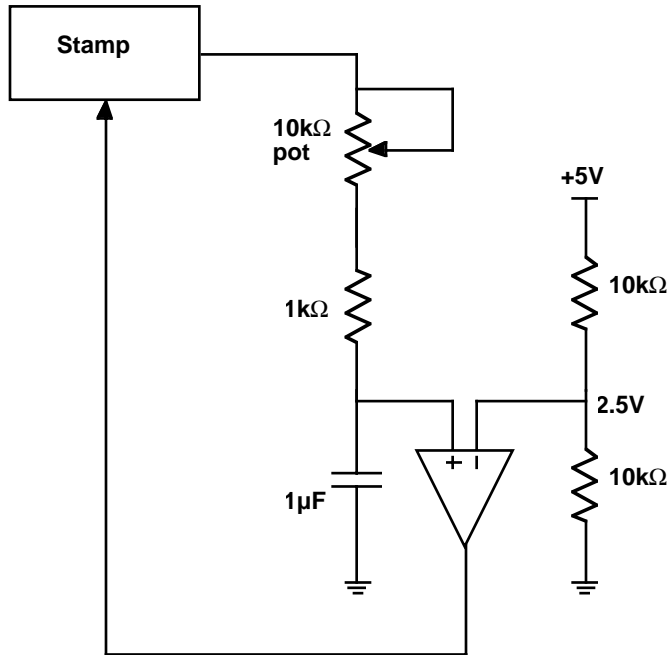


actually is quite easy. We can use one of the Stamp outputs to generate a voltage which turns on at a given time. By connecting a cap and a pot, we can build a circuit which gives a first order response. Varying the pot varies the RC time constant. Finally, we can use a comparator to determine when the waveform passes a certain level, for example, 2.5V. This time will vary according to the time constant. As you can quickly calculate, to reach halfway takes approximately $0.69 \cdot RC$ seconds.

A circuit for doing exactly this is shown above. This circuit allows you to charge and discharge a cap with a pot connected to a Stamp output. The voltage on the capacitor is compared with a voltage divider which provides 2.5V. Looking at the comparator output, the Stamp can measure the time it takes to reach 2.5V.

This is not a good circuit. It has a significant flaw. If the pot is turned all the way down, it provides zero resistance, and the Stamp output is then directly connected to the capacitor. Recall that you cannot instantaneously change the voltage on a capacitor because that would require infinite current. This problem can be fixed by placing a small resistor in series with the pot to make sure that the resistance never gets too small. The figure below illustrates the fix.

On the software side, you will alternately need to set the Stamp output high and low. You can then use the PULSIN command which will track the amount of time required to reach the appropriate input. It has a resolution of $2\mu\text{S}$, and can time up to about 130mS. With the component values shown, the maximum time constant is

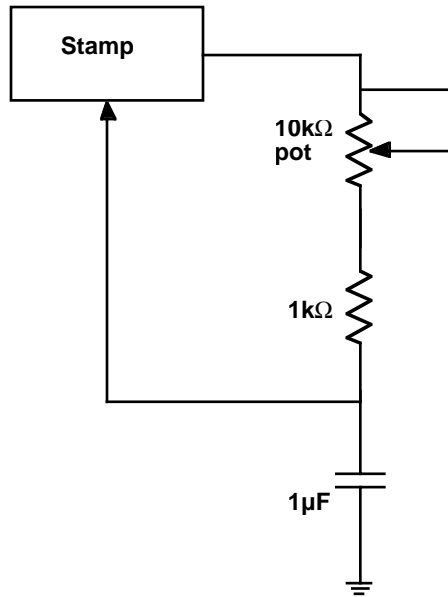


11mS. The minimum is 1mS. You should probably pause 5 to 10 time constants in-between measurements to make sure that the cap is very close to the final value before starting a new measurement.

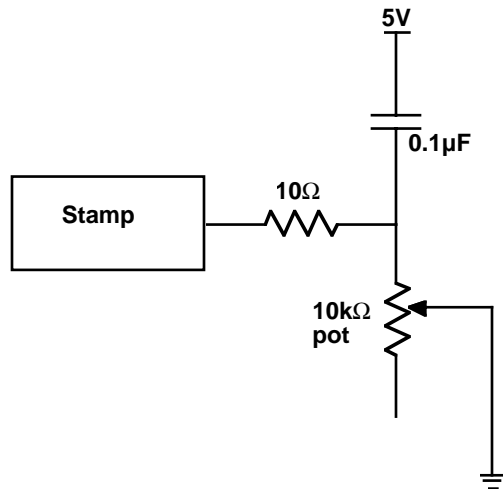
This all said, we can now simplify the circuit considerably. The inputs of the Stamp switch rather abruptly between recognizing a low or high around 1.4V. So you can think of the Stamp I/O pin as looking a lot like a comparator which is comparing to 1.4V when it is set to an input.

This suggests removing the comparator, and having the Stamp input do the comparison to 1.4V. The figure below shows the configuration. To use this circuit, you should probably only measure one direction, resetting the cap the correct way after each measurement.

Incredibly enough, we can simplify this circuit even further. Since a pin can switch between input and output, we can use it to serve double duty: at one moment charging up a cap, at another, timing when it falls back down past a given level, namely 1.4V.



Unfortunately, the Stamp documentation shows such a circuit, but we caution against its use. The schematic is reproduced below. This circuit can damage your

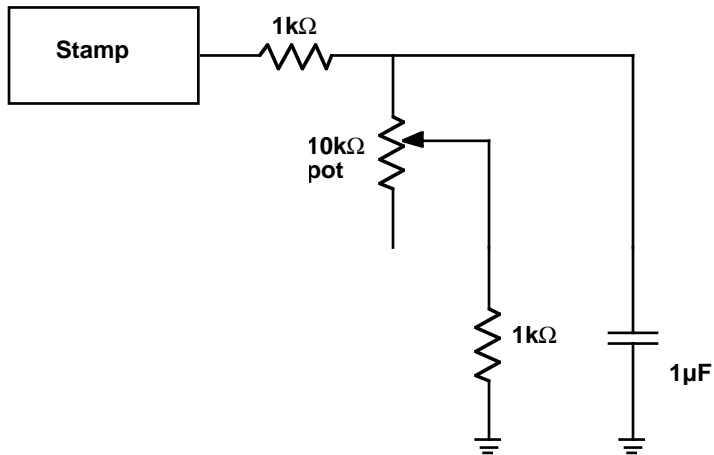


Stamp! If the pot turned all the way the one direction, the bottom of the cap is effectively tied to ground. That leaves a 10Ω resistor between the Stamp output and

ground. When the output is set high, that implies a current of $1/2$ A, a good order of magnitude above the maximum specified current for a pin.

Increasing the size of the 10Ω resistor to $1k\Omega$ eliminates the excessive current problem. However, this introduces other problems. Ignoring the cap for the moment, you will note that this resistor forms a voltage divider with the pot. The voltage of the divider is the final voltage the cap will settle to. For low resistance pot settings, this voltage will be below $1.4V$. Thus, the cap voltage might never get above the threshold! In fact, if you sit down and figure it out, you'll find that this is true for the first 28% of the pots full swing. So you will be unable to differentiate any of the pot settings for over a quarter of its range.

It is possible to work around this problem too. Adding yet another $1k\Omega$ resistor to the wiper lead guarantees that the divider will be at a minimum of $2.5V$. This is shown in the schematic below. Also, you will note that the other end of the cap is



now tied to ground. Theoretically, it doesn't make much difference where that end is tied. However, it is easier to visualize what is going on this way. Also, there are some subtle reasons to prefer this configuration, but these are beyond the scope of our present discussion.

This is a significantly better circuit than that shown in the Stamp documentation. However, it does not yield a time which is strictly linear with respect to the setting on the pot, although it is pretty close. This nonlinearity could be corrected in software if it was a major concern.

All of these circuits have accuracy problems. First of all, capacitors are notoriously inaccurate. While it is possible to buy 2% capacitors, $\pm 20\%$ is far more common. Worse yet, these values may change dramatically with temperature. For example, Z5U caps typically drop their capacitance in half with a 50°C shift in either direction away from room temperature! Adding to the problem is the fact that the 1.4V threshold of the Stamp input may also vary with temperature. The moral of the story is that if you wish to build an accurate circuit, you will have to pay close attention to both your design and choice of components!

What is an Oscilloscope?

We conclude this chapter with a quick look at one of the electrical engineers most valuable instruments: the *oscilloscope*. You will be using oscilloscopes in the lab to examine the various voltages in your circuit. But unlike a simple meter, an oscilloscope, or just *scope* for short, shows you a graph of the voltage versus time.

Early analog scopes consisted of a cathode ray tube (or CRT), a sweep generator, and a vertical amplifier. The sweep generator caused the electron beam in the CRT to sweep horizontally across the screen. The vertical amplifier moved the beam vertically in response to an input signal. As a result, a picture of the voltage as a function of time would be presented on the screen.

Unfortunately, CRTs are not like a piece of paper. When you write on the screen, the phosphor will emit light for only a short time. The picture must be constantly refreshed. You have absolutely no hope of this happening unless the waveform you are trying to look at constantly repeats. As luck would have it, many interesting waveforms are periodic. With a repeating waveform, you have some hope that the same image will be continuously refreshed so that your fairly slow eyes can perceive it.

The earliest scopes had a knob which would let you adjust the sweep frequency in an attempt to match the frequency of the input signal. As you might guess, this made using scopes a rather tricky business. Fortunately, a much better system soon emerged, called *triggered sweep*, and this revolutionized the oscilloscope.

The idea behind triggered sweep was to try to get the multiple sweeps to start at the same place in the waveform each time. This way, the pictures would all line up. When the scope sees the input go past a certain level, it would trigger the horizontal

sweep. Soon after a sweep was complete, the circuit would rearm, and await another trigger event.

Of course, this system has some limitations. For example, a complex waveform, which has multiple waves in a single period can easily confound the trigger circuit. To the trigger circuit, all the crossings of the trigger point look the same so it will choose them fairly randomly. Often there is another signal somewhere in the circuit with the same period, but without the extra bumps. Most scopes will allow you to trigger the sweep for one channel from a different channel so you can synchronize to the cleaner signal.

Another problem is that the sweep starts after the trigger event, so you can't see what happened right at the trigger, or before it. A clever solution to this problem was the invention of *delayed sweep*. With delayed sweep you could tell the scope not to start the sweep until some amount of time after the trigger. If you adjusted the knobs carefully, you could usually set the delay so that the scope was triggering just before what would have been the next trigger event.

One of the annoying side effects of triggered sweep is that if you don't get a trigger, there is no sweep, and hence no picture. Thus, you have no way to tell how you should be adjusting your trigger. *Auto triggering* was a convenience feature that was added to make your life a bit easier. If the scope doesn't receive a trigger for a long time, it will automatically trigger, just so you have something on the screen to work with. If you don't want this to happen, you can turn this feature off by setting the trigger to *normal* mode.

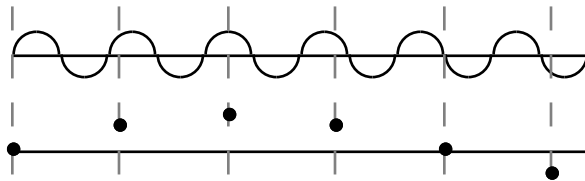
Until about ten years ago, that pretty much summed up what scopes were all about. But, as you might have guessed, something happened that completely changed how oscilloscopes work. They became digital.

The *digital storage oscilloscope*, or *DSO*, was yet another inevitable result of the digital revolution. A modern DSO is basically a computer optimized for data acquisition. At the heart of the DSO is one or more high speed analog-to-digital converters. These sample the analog voltages and display them on a computer screen.

When DSOs first appeared on the scene, they looked more like computers than scopes. This didn't sit well with the hundreds of thousands of EEs and technicians who were used to analog scopes. Eventually, the companies got wise to this, and did everything they could to make their DSOs look and feel like analog scopes. So in case you were wondering why DSOs have such an odd interface, that's why: marketing.

That said, digital scopes can do many things analog scopes can't. But they also have some surprising limitations. First of all, digital scopes can digitize a waveform and hold it on the screen essentially forever. With old analog scopes, exotic technologies were used to attempt this feature, but they were crude and expensive.

Another problem with digital scopes is that they truly sample the waveform at various instants, and you may not be seeing what's really going on in-between. Digital scopes can *alias*. Imagine a repeating sine wave with some period, which is sampled with a slightly longer period. The samples look like a slowly moving sine wave, but the actual waveform is moving much faster. An example is shown below.



The most serious problem with some digital scopes is a very clever trick generically known as non-real-time sampling. Fast ADCs are very expensive. So if you are going to be looking at repetitive waveforms anyway, why go nuts trying to sample the waveform incredibly quickly. Instead, use aliasing to your advantage! For example, take one sample per a period, each time taking it further after the trigger, and then recreate the original waveform. This is often called *repetitive interleaved sampling*. It works great for periodic waveforms assuming a good trigger. Otherwise, all bets are off. If you really want to see a *transient waveform* (one that may only happen once), you'll need a scope with real-time capability. Unfortunately, there is much confusion here because some scopes will automatically switch from real-time to repetitive sampling if the ADC can't keep up. Watch out for this!

Probably the single greatest boon to undergraduate lab measurements was the inclusion of an Auto Setup button on most digital scopes. If you press this, the scope will try and guess what the appropriate settings should be by looking at the input signals. Generally this works pretty well. But it is a far over used crutch, and you are advised not to depend upon it exclusively. There is no substitute for a true understanding of what's going on.

Finally, we have to say a word about scope probes. These are the fancy looking test leads that are used with a scope. Unfortunately, some people think these are just wires. In reality, they generally contain circuitry that helps reduce the loading

effects of the scope on the circuit you are trying to measure. Scope probes can cost anywhere from \$15 to over \$1000 each, so treat them with respect.

Now that you have some idea of the basic principles, you will need to get a demonstration on the use of the particular scopes in your lab. You should make sure that you view all of the interesting waveforms in your circuit for this chapter.

What's Next?

In the next chapter, we will examine the inductor. Whereas a capacitor stores energy in an electric field, the inductor stores it in a magnetic field. Circuits with both a capacitor and an inductor become particularly interesting because they don't just monotonically approach some steady state in response to a step input. These circuits can ring like a bell. Be forewarned that the math will be getting a little harder, but the results will be more interesting!

The Problem

Imagine you are an employee of Small Town Engineering. Your company has been asked by the city of Small Town to bid on a proposed upgrading of the main intersection in town, Rt. 660 and Out Street. Rt. 660 is a 50 mph, two-lane divided highway, and is a major thoroughfare through this part of the state. Out St. is a lightly travelled, 25 mph, one-way road which crosses Rt. 660. (In St. parallels Out St., but is one-way the other direction, and crosses under Rt. 660 via a narrow tunnel.) Currently, there is a Stop sign on Out Street at the intersection. However, the constant flow of traffic on Rt. 660 means long waits for those traveling through the intersection via Out Street. The city engineer wants to install a traffic light to solve this problem.

A number of other companies have also been asked to bid on the project. Knowing that different companies will propose somewhat different systems, the city engineer has requested that each company present a working model of their intersection. You have been selected to build the model for your company.

Your task for this chapter is to build the model intersection. In particular, you will build a small *loop detector* which will detect the presence of a “car” (actually, a small block of steel) and actuate a “traffic light” (actually six LEDs) accordingly.

What You Need to Know

In order to solve this problem, here are a some things you need to know:

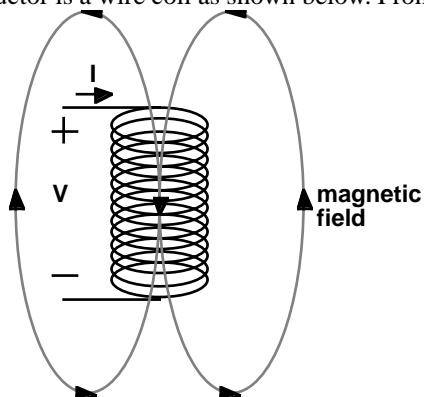
- What is an inductor?
- How do I use an inductor in a circuit?
- How do I handle Nonzero Initial Conditions?
- What is an LC circuit?
- What is a loop detector?

What is an Inductor?

In the previous chapter, we introduced the capacitor. We said that a cap stores energy in the form of an electric field, and that the voltage across it was related to the integral of the current through it.

An inductor is the dual of a capacitor. An inductor stores energy in the form of a magnetic field, and the current through it is related to the integral of the voltage across it.

The prototypical inductor is a wire coil as shown below. From physics, recall that a



coil with a current running through it produces a magnetic field. The direction of the magnetic field can be determined by the *right hand rule*. With the fingers of your right hand wrapped in the direction of positive current flow, your thumb points in the direction of the magnetic field.

Generating a magnetic field takes energy. However, maintaining one need not. Magnets don't require a power source to keep being magnetic. The same is true for an inductor. Energy is used by the inductor to create the magnetic field. Ideally, this energy can later be returned to the system. Ideal inductors do not dissipate power - they are energy storage devices.

The defining relationship for an inductor is:

$$V = L \frac{dI}{dt}$$

in differential form, or:

$$I = \frac{1}{L} \int V dt$$

in integral form. The constant L is the *inductance*, and is measured in units of *Henries*. A one Henry inductor is a very large inductor. Usually, we'll be talking in terms of mH or μH .

Assuming that there is initially zero current through an inductor (which is the dual of saying that there is no voltage across a capacitor), we can calculate the amount of energy it takes to get to up to a given current level, which implies a certain level of magnetic field. This will mirror our derivation for the energy stored in a capacitor. We will first calculate the power.

$$P = IV = IL \frac{dI}{dt}$$

Next, we integrate this to find the energy.

$$E = \int P dt = \int IL \frac{dI}{dt} dt = L \int I dI = \frac{1}{2} LI^2 \Big|_0^I = \frac{1}{2} LI^2$$

This is the energy stored in the magnetic field created by the inductor at a given current level. Note how similar this is to the expression for the energy stored in a capacitor at a given voltage level.

As circuit elements, we can consider the effect of series and parallel combinations of inductors. Let's consider a series connection of two inductors. Since the inductors are in series, they must have the same current, and the voltage across the com-

combination is just the sum of the voltages across each inductor. Solving for this voltage yields:

$$V = V_1 + V_2 = L_1 \frac{dI}{dt} + L_2 \frac{dI}{dt} = (L_1 + L_2) \frac{dI}{dt}$$

Clearly, inductors in series act like a single inductor of a value equal to the sum of the inductances.

Inductors in parallel must have the same voltage across them. The total current through the combination will just be the sum of the individual currents as given below.

$$I = I_1 + I_2$$

If this is true, it must also be true that the first derivatives match up.

$$\frac{dI}{dt} = \frac{dI_1}{dt} + \frac{dI_2}{dt} = \frac{V}{L_1} + \frac{V}{L_2} = V \left(\frac{1}{L_1} + \frac{1}{L_2} \right)$$

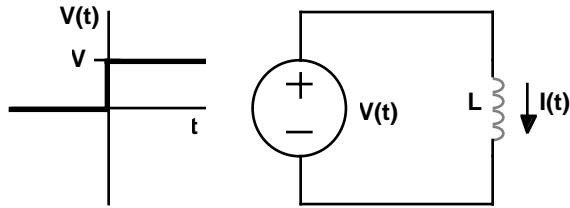
This can be written as:

$$V = \left(\frac{1}{L_1} + \frac{1}{L_2} \right)^{-1} \frac{dI}{dt}$$

which shows that inductances in parallel act like the reciprocal of the sum of their reciprocals. Or, in other words, inductors in series and parallel combine just like resistors. It is important to recall that caps did just the opposite.

How Do I Use an Inductor in a Circuit?

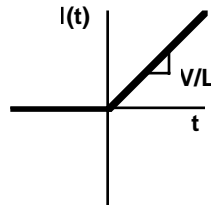
Once again, our discussion here will be the dual of our discussion concerning capacitors. The simplest circuit we can consider is an inductor connected to a voltage source as shown below. Note the symbol for an inductor - a stylized coil.



To solve this circuit, we need an initial condition. In this case, we will presume that the initial current through the inductor for $t < 0$ is zero. We can now directly apply the integral form of the inductor relationship.

$$I = \frac{1}{L} \int_0^t V dt = \frac{Vt}{L}$$

This result applies for $t \geq 0$. A rough sketch is shown below.

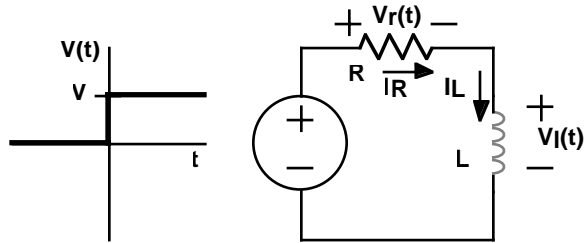


This should all look very familiar. It is just like the result for the capacitor, except that I and V are flipped, and we now have L instead of C . Just as caps can't change their voltage instantaneously without infinite current, inductors can't change their current instantaneously without an infinite voltage. Of course, you can't really have infinite voltages, but you can get very high ones. In fact, rapidly changing the current in an inductor is frequently used as a way of generating high voltages. (The ignition coil in an automobile is a wonderful example.)

Let's consider a circuit consisting of an inductor in series with a resistor and a voltage source.

Writing the KVL equation for this single loop gives:

$$V(t) = V_R(t) + V_L(t) = I(t)R + L \frac{dI(t)}{dt}$$



Once again, we will assume that the current through the inductor is zero for $t \leq 0$. Solving the equation for $t > 0$ allows us to say that $V(t > 0) = V$, making this a separable equation. Rearranging the terms gives:

$$L \frac{dI(t)}{V - I(t)R} = dt$$

Integrating both sides yields:

$$\int L \frac{dI(t)}{V - I(t)R} = \int dt$$

$$\frac{-L}{R} \log|V - I(t)R| = t + K$$

$$\log|V - I(t)R| = \frac{-tR}{L} + K'$$

where K and K' are constants to be determined from the initial condition. Next we exponentiate both sides, creating yet another constant K'' .

$$V - I(t)R = K'' e^{\frac{-tR}{L}}$$

Solving for $I(t)$ produces:

$$I(t) = \frac{V - K'' e^{\frac{-tR}{L}}}{R}$$

Applying the initial condition that $I(0) = 0$ sets $K'' = V$, yielding:

$$I(t) = \frac{V}{R} \left(1 - e^{-\frac{tR}{L}} \right)$$

This is another exponential, and says that the current starts at zero, and asymptotes to V/R with a time constant of L/R . This is essentially the same relationship as we got for the voltage across the capacitor in the previous chapter, with the exception that the time constant there was RC .

We can complete our analysis by calculating the voltage across the inductor as a function of time. From KVL, we have:

$$V_L(t) = V - V_R(t) = V - I(t)R = V - V \left(1 - e^{-\frac{tR}{L}} \right) = V e^{-\frac{tR}{L}}$$

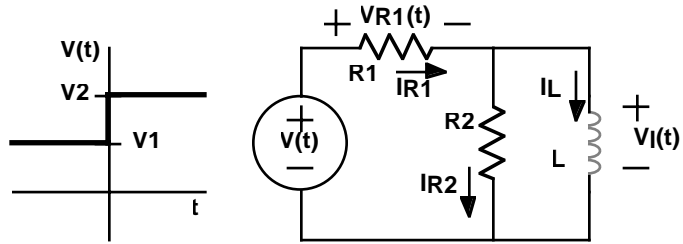
which shows that the voltage starts at V and falls exponentially to zero.

At this point, it would be prudent to examine the differences and similarities between inductors and capacitors. If you examine the circuit solution, you will see that when we first turn on the voltage, the inductor looks like an open circuit - no current flows through it. As time progresses, the current through the inductor increases, and asymptotes at a level that you would expect for a short circuit. This is a general principle. To sudden changes, the inductor looks like an open, and eventually (with the scale determined by the time constant L/R), it looks like a short.

If you examine the results of the last chapter, you will see that caps do exactly the opposite. To sudden changes, the cap looks like a short circuit, but after a number of time constants (RC in this case), it looks like an open circuit.

If you keep these results in mind, you can rapidly solve for the time response of any network consisting of resistors, a stepped source, and a single capacitor or inductor. Solving an example problem will help demonstrate how you can pull all of this together. Consider the schematic below.

We will assume that $V(t)$ has been at V_1 for a long time (long, that is, compared to any time constants in the system) before $t = 0$. If that is true, then the inductor will look like a short circuit just before $t = 0$, giving $V_L(t < 0) = 0$, $I_L(t < 0) = V_1/R_1$ and $I_{R2}(t < 0) = 0$. At $t = 0$, the voltage source suddenly rises to V_2 . Eventually, the inductor will again look like a short circuit. At that point, we will have $V_L(t \gg 0) =$



0, $I_L(t \gg 0) = V_2/R_1$ and $I_{R2}(t \gg 0) = 0$. All that remains is to determine the behavior shortly after $t = 0$. We know that this will be an exponential, but we need to know the time constant.

We have already shown how to find the time constant for simple R-C and R-L circuits. If you look at those circuits again, you will notice that they are really solutions for caps and inductors tied to Thevenin equivalent circuits. Since we can model any I-V-R network by its Thevenin equivalent, and we know how to solve for the time constants in these cases, we can solve for the time constant of any I-V-R network with a single cap or inductor.

Looking back into the circuit from the inductors terminals, we can quickly find the Thevenin equivalent circuit for $t > 0$. $I_{SC} = V_2/R_1$ and $V_{OC} = V_2 R_2 / (R_1 + R_2)$ giving $R_T = R_1 R_2 / (R_1 + R_2) = R_1 \parallel R_2$. So the time constant will be:

$$\tau = \frac{L}{R_T} = L \frac{R_1 + R_2}{R_1 R_2}$$

Putting this all together, we can now write down all the relevant equations. For example, the voltage across the inductor will be:

$$V_L(t > 0) = V_2 e^{-\frac{t}{\tau}} = V_2 e^{-\frac{t R_1 R_2}{L(R_1 + R_2)}}$$

The current through the inductor is:

$$I_L(t > 0) = \frac{V_1}{R_1} + \frac{V_2 - V_1}{R_1} \left(1 - e^{-\frac{t R_1 R_2}{L(R_1 + R_2)}} \right) = \frac{V_2}{R_1} + \frac{V_2 - V_1}{R_1} e^{-\frac{t R_1 R_2}{L(R_1 + R_2)}}$$

These are simply the equations which match our initial conditions and our final values with the right time constant.

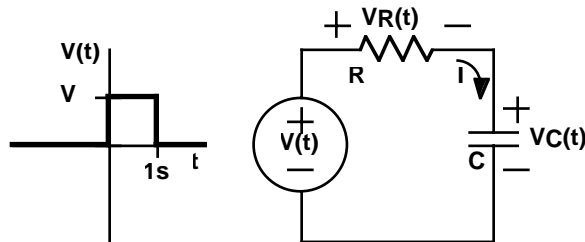
Finally, we can add in one more of our tricks to extend this even further. Ideal inductors and capacitors are linear components, allowing the use of superposition. You can easily prove this for the capacitor by finding the current due to the sum of two voltages, i.e. $V(t) = V_1(t) + V_2(t)$.

$$I(t) = C \frac{dV(t)}{dt} = C \frac{d(V_1(t) + V_2(t))}{dt} = C \frac{dV_1(t)}{dt} + C \frac{dV_2(t)}{dt}$$

A proof for the inductor is similar, and is left as an exercise for the reader.

The upshot is that you can now find the time response of any network of resistors, stepped sources, and a single inductor or capacitor by finding the response to each source individually, and summing these together.

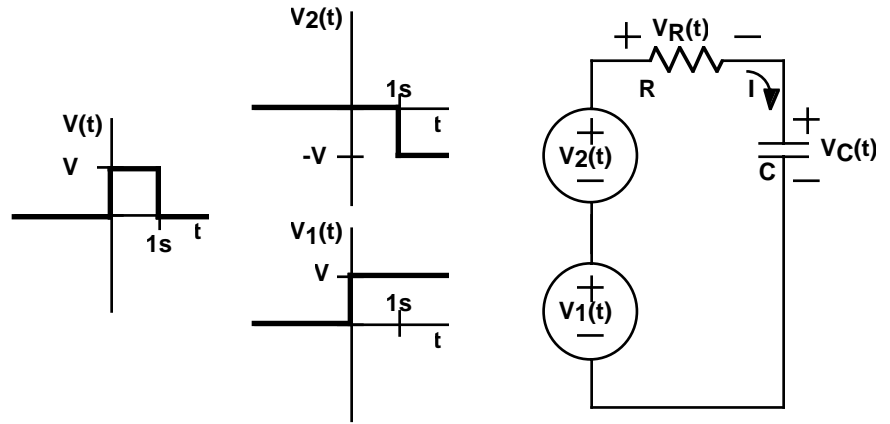
In this last example, we will be using superposition in a particularly clever way. In the circuit below, the voltage source will produce a 1 second pulse. We would like to calculate the voltage across the capacitor as a function of time.



What has this got to do with superposition? Surprisingly, a great deal. The problem with solving this circuit is that we have only solved for the case of a single step in the voltage; not a pulse! But, you can think of a pulse as the sum of two steps as shown in the figure.

This circuit is completely equivalent, and easily solved using superposition. Suppressing $V_2(t)$ and considering the effect of $V_1(t)$, we get:

$$V_C(t > 0) = V \left(1 - e^{-\frac{t}{RC}} \right)$$



and $V_C(t \leq 0) = 0$. In the other case, suppressing $V_1(t)$ and considering $V_2(t)$, we have:

$$V_C(t > 1s) = -V \left(1 - e^{-\frac{t-1s}{RC}} \right)$$

and $V_C(t \leq 1s) = 0$.

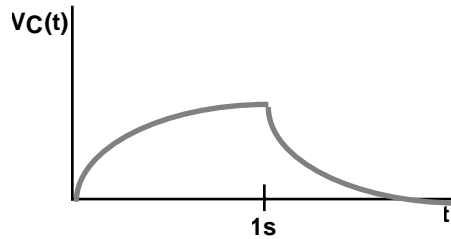
Summing these two solutions together gives the complete answer:

$$V_C(t \leq 0) = 0$$

$$V_C(0 < t \leq 1s) = V \left(1 - e^{-\frac{t}{RC}} \right)$$

$$V_C(t > 1s) = V \left(1 - e^{-\frac{t}{RC}} \right) + -V \left(1 - e^{-\frac{t-1s}{RC}} \right) = V e^{-\frac{t}{RC}} \left(e^{\frac{1s}{RC}} - 1 \right)$$

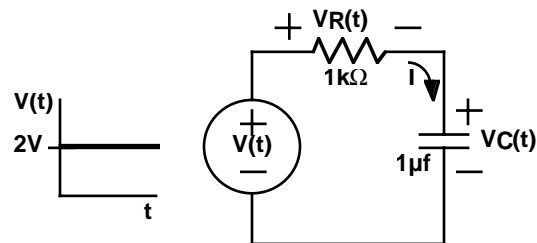
A sketch of this looks roughly like this:



How Do I Handle Nonzero Initial Conditions?

In the last section we developed a pretty good understanding of how circuits respond to step waveforms presuming everything started at zero. Now we consider the question of what happens when we have a nonzero initial condition for a cap or inductor. We have actually answered this question, indirectly, when we talked about superposition. Here we address it specifically.

Imagine at $t = 0$, a $1\mu\text{f}$ cap has 1V across it. This cap appears in the circuit shown below.



If we are trying to find what happens for $t > 0$, we don't care what $V(t < 0)$ was. All of that past history has simply resulted in $V_C(t = 0) = 1\text{V}$. This is what we call a *state variable*. The state variables of a circuit completely describe the state of a circuit and no further information about the past is required to calculate what happens in the future.

So this leaves us the question of how to calculate how this circuit will respond to the flat 2V input, knowing that $V_C(0) = 1\text{V}$. Assuming at $V_C(t = -\infty) = 0\text{V}$, the volt-

age across the cap can be written as the integral of the charge over all time up until t .

$$V_C(t) = \frac{1}{C} \int_{-\infty}^t I_C(t) dt$$

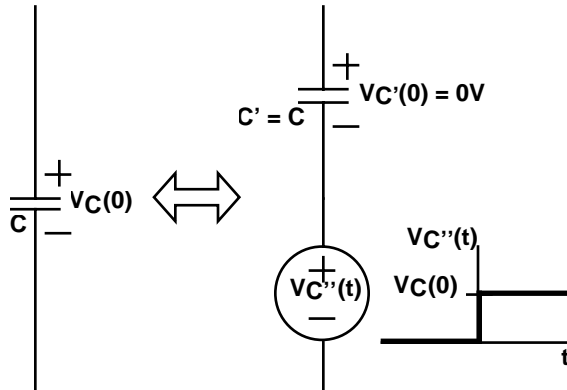
However, we know $V_C(t)$ at $t = 0$. This is

$$V_C(0) = \frac{1}{C} \int_{-\infty}^0 I_C(t) dt$$

which allows us to rewrite $V_C(t)$ as

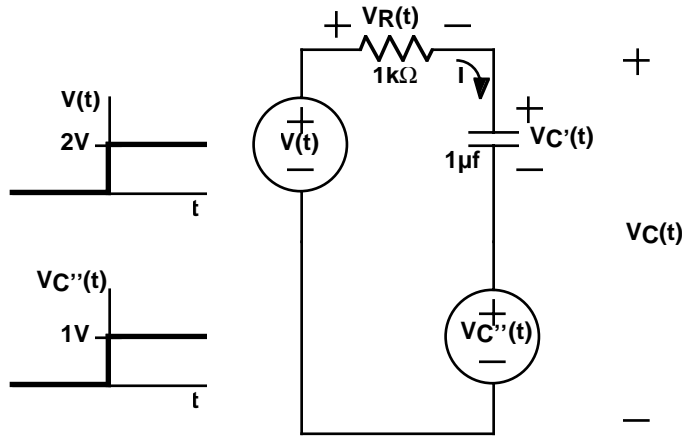
$$V_C(t) = V_C(0) + \frac{1}{C} \int_0^t I_C(t) dt$$

This just says that a capacitor with an initial charge behaves just like a cap with no initial charge in series with a voltage source equal to the initial voltage on the cap. We can even show this schematically.



We can now reform our original problem, adding in a non-time-varying voltage source and a cap with zero initial charge at $t = 0$ as shown below.

This circuit is now easily solved by superposition. (Well, maybe not easily, but at least fairly straightforwardly.) Considering either source, the cap initially looks like



a short (and thus has zero volts across it), and eventually looks like an open. For $V(t)$ that will set $V_C(t \gg 0) = 2V$. For $V_C''(t)$, $V_C(t \gg 0) = -1V$. Summing the two solutions gives $V_C(t \gg 0) = 1V$. But $V_C''(t > 0) = 1V$, so $V_C(t \gg 0) = 2V$. Knowing this started at 1V, and that the time constant is RC , we can write:

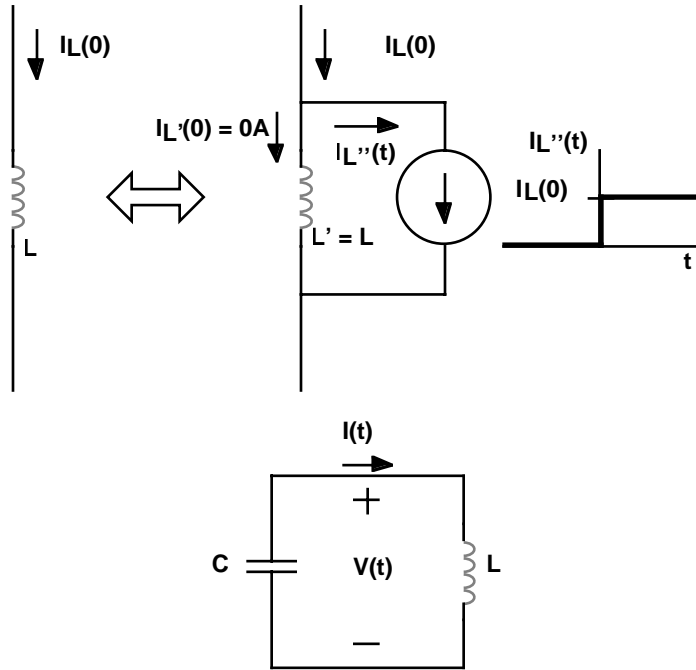
$$V_C(t > 0) = 1V + 1V \left(1 - e^{-\frac{t}{1 \times 10^{-3}}} \right)$$

A very similar technique applies for inductors. An inductor with an initial current behaves just like an inductor with no initial current in parallel with a current source with current equal to that in the original inductor. Again, we end up with an elegant way of handling nonzero initial energy storage.

What is an LC Circuit?

So far, we have restricted our discussion to circuits with single inductors or capacitors. In this section we consider what happens when you connect a cap to an inductor. Mathematically, this will result in a second order differential equation. But rather than focus on the mathematics, we will first try to build some intuition.

First, let's make some assumptions. To make life easy, let $C = 1F$, $L = 1H$, and at time $t = 0$, we will say the $V(0) = 1V$, and $I(0) = 0A$. Since this circuit consists only



of nondissipative elements, the total energy stored in the circuit must be constant over time. At $t = 0$, all the energy is stored in the cap. With $E = 1/2 CV^2$, that works out to $1/2$ joule.

With 1V across the capacitor, current will eventually start to flow through the inductor. This current will continue to flow (in the positive direction on the diagram) at least until $V = 0$. At this point, the capacitor is completely discharged, and all of the energy must be stored in the inductor. If $E = 1/2 LI^2$, then the current through the inductor at that moment must be 1A. But that's not the end of the story. The inductor is still pushing 1A of current around the loop, so the cap starts to charge the other direction, with a negative V . This will continue until the inductor current goes to zero. At that point, all the energy is once again stored in the cap, but now the voltage is reversed!. Of course, this -1V across the cap causes a negative current to flow into the inductor. And as you might have guessed, this moves all the energy from the cap back to the inductor, but with the magnetic field the opposite direction from before. Finally, the current through the inductor recharges the cap back to the original state, and the process continues forever.

It might be easiest to understand this via an analogy. A mass bouncing on an ideal spring transfers energy back and forth between kinetic and potential forms. If there are no losses (e.g. friction), the mass will theoretically bounce forever. The cap-inductor system behaves analogously. Energy is completely transferred back and forth between the electric field in the cap and the magnetic field in the inductor.

If you were to work out the differential equation, you would find that the voltage and the current each oscillate sinusoidally, with a period:

$$\tau = 2\pi\sqrt{LC}$$

The equation for the voltage across the capacitor would be:

$$V(t) = \cos(\omega t)$$

where:

$$\omega = \sqrt{\frac{1}{LC}}$$

ω is usually referred to as the *frequency* of the oscillation in radians per second. (2π rads/sec = 1 cycle/sec = 1 Hertz)

It is easy to show the current through the circuit must be:

$$I(t) = \sin(\omega t)$$

and proof of this is left as an exercise for the reader.

Of course a real LC circuit does not oscillate forever because of various losses that this simplified analysis hasn't consider. These losses will cause the oscillations to gradually die away, and will even shift the frequency somewhat. If you go on for further study in Electrical Engineering, you will probably study this extensively. But for our purposes, we've said just about as much as we'd like to about the theory behind LC oscillators.

What is a Loop Detector?

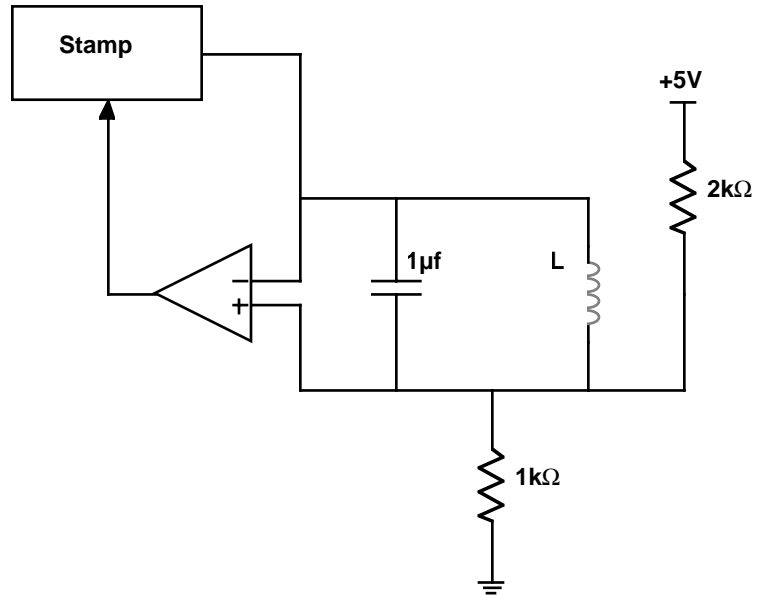
Returning to the problem for this chapter, you are asked to build a *loop detector*. It would probably help if you knew what one was. These are the most common sensors that are buried in the road to detect the presence of an automobile in a particular location. They are frequently used at traffic lights to detect if a car is waiting for the light to change. But they are also used on highways to gather data on congestion.

The principle is fairly simple. A coil of wire is buried in the road. A car passing over the coil interacts with the magnetic field, which slightly modifies the inductance. An electronic circuit detects this change. You may have noticed cuts in a road which were made to insert a loop detector.

For your inductor, you should wind 100 turns of 28 gauge magnet wire in an approximately 1" diameter loop. (A paper towel or toilet paper tube makes a good form for winding around.) Magnet wire uses a lacquer coating for insulation, which you will have to carefully scrape off the ends to make contact. This should give you an inductor with a value of approximately 1/2mH.

While you could use a simple R-L circuit, similar to the last lab, to measure the inductance, we would like you to use the L-C circuit shown below. A Stamp I/O pin is used to provide current to the L-C circuit. Set this pin to 5V for some time period, waiting for everything to settle out. This will result in some current flow through the inductor, but no voltage across the capacitor. Next, set this pin to an input. That has the same effect as disconnecting it. This will allow the L-C circuit to oscillate. The comparator will amplify these oscillations into valid logic levels. (Be sure to run the comparator off of the 5V supply this time to avoid protection circuits!) These can then be timed to get a relative measure of the change in inductance.

There is some simple analysis you can do on this circuit to understand its behavior. With the Stamp output delivering 5V for a long time, the current through the inductor should be 5mA. (Recall that in the long term, the inductor looks like a short!) Disconnecting this pin (by setting it to an input) leaves the L-C circuit hanging off of a voltage divider. The voltage at the divider will be 5/3V, and the other end of the L-C circuit will oscillate up and down relative to this. With all the energy initially



stored in the inductor, $E = 1/2 LI^2 = 6.25 \times 10^{-9}$ joules. When this moves to the capacitor, it will generate a voltage:

$$V = \sqrt{\frac{2E}{C}} = \sqrt{\frac{2 \times 6.25 \times 10^{-9}}{1 \times 10^{-6}}} \approx 0.1 \text{ V}$$

Of course, with nonideal components, there will be losses, so this will quickly dissipate.

You should also calculate the period of the oscillation. This will be about $140 \mu\text{s}$. With $2 \mu\text{s}$ timing resolution on the Stamp, you should be able to detect inductance changes on the order of 2%. This should be sufficient to detect a steel block placed on top of the inductor.

What's Next?

Up until now, all of our circuits have been battery powered. And as you may have noticed, batteries do eventually go dead. Wouldn't it be nice to be able to power your circuits from a standard 110V AC power line?

In the next chapter we will be exploring exactly that issue - how do you generate a 5V power supply from 110V AC? We will do this in several steps, and along the way, introduce some new components. This will include specially made inductor pairs which share a common magnetic field. These are called *transformers*. We will also be revisiting diodes and capacitors. So you may want to review your understanding of those elements.

The Problem

Batteries are an incredibly expensive source of power. In 1998, electric utilities in the United States charge between 5 and 20 cents per kilowatt-hour. On the other hand, you would be hard pressed to find a 9V battery that could put out a watt for much more than an hour. And they cost a great deal more than 20 cents. Power out of the wall socket is at least several orders of magnitude cheaper than buying batteries.

Given this economic reality, we would like to be able to power our 5V circuitry from a standard 110VAC wall outlet. (Standard, at least, in the U.S.) But connecting the Stamp to such a high voltage produces only a momentarily spectacular display, ending with a depressing puff of smoke. How does one efficiently convert 110VAC to a suitable 5V supply?

Your goal for this chapter is to build an isolated, regulated, linear 5V power supply, which runs off of 110 VAC. Not only will this eliminate the need for batteries, but it has the side benefit of delivering significantly more current than the Stamp's on-board 5V regulator which was limited to 50mA. This will allow you to use higher current devices such as relays, light bulbs, etc.

What You Need to Know

In order to solve this problem, here are some things you need to know:

- What is AC power?
- What is a transformer?
- What is a rectifier?
- What is a voltage regulator?

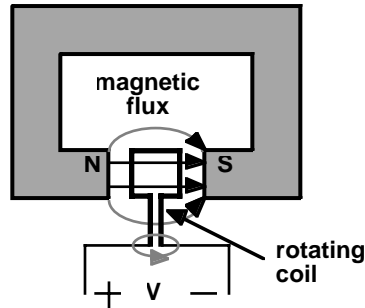
What is AC Power?

In the previous chapters, our power source has generally been a constant value voltage source, such as a battery. However, batteries are a remarkably inefficient and often inconvenient way of generating electricity. Much of the world's accessible energy is most easily released in thermal or mechanical form.

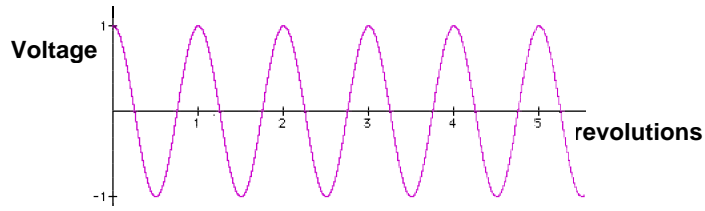
Arguably one of the greatest discoveries of all time was made essentially simultaneously by Michael Faraday (of Farad fame) and Joseph Henry (of Henry fame) around 1831. They noticed that when a magnet was passed by a piece of wire, it generated a small voltage. In particular, they found that this voltage was proportional to the rate of change of magnetic flux - the faster they moved the magnet, the higher the voltage. In physics, this is known as *Faraday's Law of Induction*.

The flip side of this may be more familiar. As we reviewed in the last chapter, a coil of wire with a current passing through it creates a magnetic field. This will allow us to create a device which converts a voltage signal into a magnetic field and then converts it back into another voltage signal. But we're getting ahead of ourselves.

Faraday's Law of Induction provides a basis for converting mechanical energy into electrical energy. The basic idea is to move a coil of wire relative to a magnetic field, so as to cross lines of magnetic flux. This will generate a current in the wire. Such a device is called a *generator* and is depicted below. To make things simpler mechanically, the coil is usually made to rotate within the field. As the coil rotates, it cuts through the flux lines, generating a voltage across the coil terminals. When the face of the coil is parallel to the field as shown in the figure, it cuts rapidly through the flux lines. But when the coil has turned 90° and is perpendicular to the field lines, at that moment the motion of the coil is tangential to the field, and produces no voltage. As the coil continues to turn past this point, it cuts through the



field in the opposite direction, generating a negative voltage. The upshot of all this is that as the coil turns, the voltage produced will vary as the cosine of the angle as shown below.



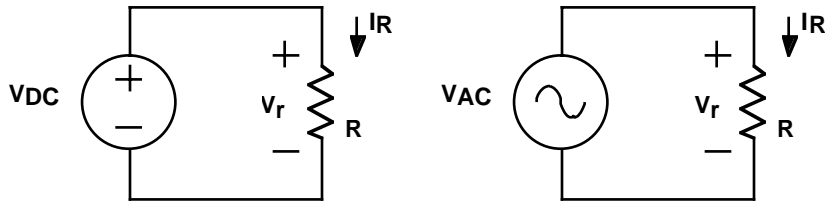
This sinusoidal waveform is referred to as *alternating current* or AC for short. The equation for a waveform of this type is:

$$V = A \cos(\omega t + \theta)$$

where A is the maximum *amplitude*, ω is the *frequency* in radians per second, and θ is the *phase* in radians. These three parameters completely specify a sine wave.

Since generators naturally produce sine waves, it is not surprising that these waveforms play an important role in circuit design. (In fact, sine waves are the basis for much of linear circuit analysis for reasons that we will explore to a limited extent in a later chapter.) So we need to have a good handle on describing them.

An interesting question to ask might be what size sine wave do you need to deliver the same amount of power as a steady source (often referred to as *DC* for *direct current* - to distinguish it from AC). Consider the two circuits shown here. For the DC circuit, the power dissipated in the resistor will just be $IV = V^2/R$.



For the AC circuit, the power being dissipated in the resistor will fluctuate with the voltage and current. This will give us something of the form:

$$P = IV = \frac{V^2}{R} = \frac{(A \cos(\omega t + \theta))^2}{R}$$

To compare this with the DC expression, we will need to find the average power delivered by this AC source. We can find this by integrating the power over one period, and dividing by the length of that period. This gives:

$$P_{ave} = \frac{\omega}{2\pi} \int_0^{\frac{2\pi}{\omega}} \frac{(A \cos(\omega t + \theta))^2}{R} dt = \frac{\omega A^2}{2\pi R} \int_0^{\frac{2\pi}{\omega}} (\cos(\omega t + \theta))^2 dt$$

Applying the trigonometric identity $\cos^2 A = 1/2 + \cos(2A)/2$ yields:

$$P_{ave} = \frac{\omega A^2}{2\pi R} \int_0^{\frac{2\pi}{\omega}} \left(\frac{1}{2} + \frac{\cos(2(\omega t + \theta))}{2} \right) dt$$

The second term in the integral is a sine wave that we are integrating for exactly 2 cycles, and thus sums to zero. Integrating the first term leaves us with:

$$P_{ave} = \frac{\omega A^2 t}{2\pi R 2} \Big|_0^{\frac{2\pi}{\omega}} = \frac{A^2}{2R}$$

Comparing this with the DC equation reveals that to get the same average power:

$$\frac{V^2}{R} = \frac{A^2}{2R} \Rightarrow V = \frac{A}{\sqrt{2}} \approx 0.707A$$

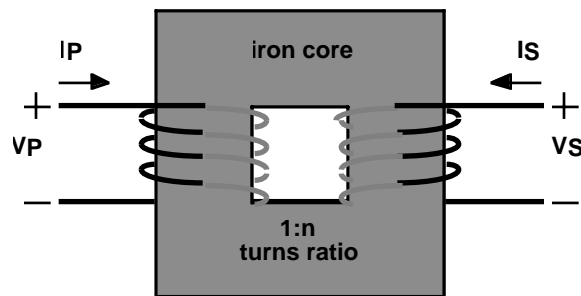
This average power concept is so useful that AC voltages are general expressed in terms of it. When we say that the utility provides 110 VAC, we really mean that they provide a sine wave with average power equivalent to a DC 110V. Plugging into the equation above, we see that means the actual voltage varies between $\pm 155.56V$.

Rather than calling this something like the DC equivalent voltage, this is usually call the *root-mean-square*, or *RMS* voltage. If you examine closely the steps we went through, you'll see that we in fact squared the voltage, took the average, and then finally applied the square root function.

AC voltages are not always measured in terms of their RMS value. Sometimes, we simply talk about their amplitude - a much easier value to read off of a scope. You'll also see the term peak-to-peak which is measuring from the lowest to the highest voltage, i.e. two times the amplitude. When making AC voltage measurements, be sure you carefully note which form you are using. Most AC meters will directly read out in RMS values. (Note: many meters presume that the waveform is sinusoidal, measure the amplitude, and scale by a factor of square root of 2. If you want to find the RMS value of a nonsinusoidal signal, you will need to use a meter with *true RMS* capability.)

What is a Transformer?

A *transformer* is a device where two or more coils share a common magnetic flux. The idea is that the primary coil takes in a changing voltage and creates a changing magnetic field, and the secondary coil converts this changing magnetic field back into a changing voltage. Typically, the coils are wrapped around a common iron core, which serves to channel the magnetic field through each coil. An example is shown here.



An ideal transformer obeys a simple relationship: the voltage on the secondary side will be proportional to the voltage on the primary side, with the constant of proportionality determined by the ratio of the number of turns on the secondary side to the number of turns on the primary side (known as the *turns ratio*). In equation form, this can be stated as:

$$V_S = nV_P$$

where n is the turns ratio.

Another key feature of ideal transformers is that they do not dissipate power. So whatever power goes in one side must come out the other. This too can be written in equation form as:

$$I_S V_S = I_P V_P$$

Combining this with the first equation yields:

$$\frac{I_P}{I_S} = \frac{-V_S}{V_P} = -n \Rightarrow I_S = \frac{-1}{n} I_P$$

So the current is also scaled by n , but in the opposite direction, and with a sign flip.

Ideally, a transformer would follow these equations, but for a variety of reasons, real transformers have some serious limitations. Most important among these is that they really only work for AC waveforms, and more specifically, AC waveforms within a fairly narrow frequency range. But with that caveat, transformers are remarkably ideal.

Transformers are a major reason why electric power is generally transmitted in AC form. With an inexpensive transformer, you can efficiently convert a high current, low voltage signal to a low current, high voltage one, and vice versa. Not only is this handy, but it has dramatic implications for power transmission. Electric power lines are run at very high voltages, and relatively low currents. That way, resistive losses in the long wires do not cause significant voltage drops, and hence wasted power. Transformers allow utility companies to use far thinner and cheaper wires.

For our purposes, we will be using a transformer to efficiently convert the 110VAC (that's RMS!) available from a wall socket to a more manageable level. You should have a wall transformer (often called an AC adaptor) designed to convert 120VAC to 12VAC, which means it should have a turns ratio of about 10:1 (i.e. $n=1/10$).

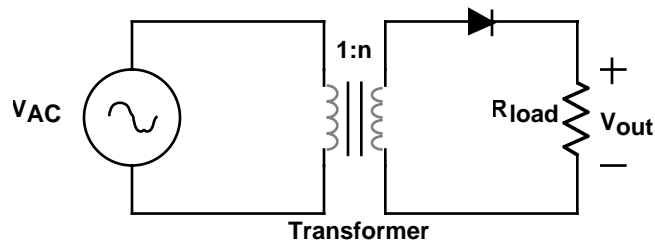
Examine the transformer output voltage on a scope in lab. Be very careful not to connect your Stamp to these voltages.

What is a Rectifier?

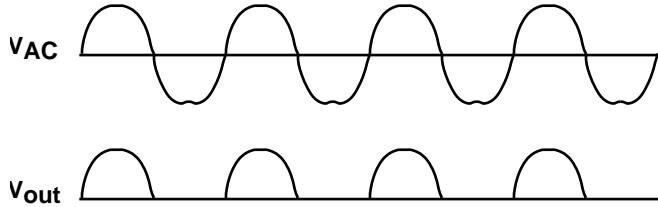
Now that we know how to step down AC voltages to a manageable level, we are left with the problem of converting 12VAC into our desired 5V DC supply. We'll approach this in two steps: first we'll convert this AC voltage into a DC one via a process called *rectification*. Then we'll step this down to 5V using a *regulator*.

Fortunately, you have already been exposed to the key component needed for rectification: the diode. Recall that diodes are two-terminal devices which allow current to pass only in one direction. Unlike LEDs, silicon diodes typically drop about 0.6 or 0.7V when conducting. But for our purposes in this chapter, we'll tend to think of them as ideal, one-way current valves.

The simplest possible circuit for converting AC into DC is a *half-wave rectifier*. It basically consists of a single diode which only allows current flow in one direction. A possible circuit is shown below. Note the symbol for the transformer.

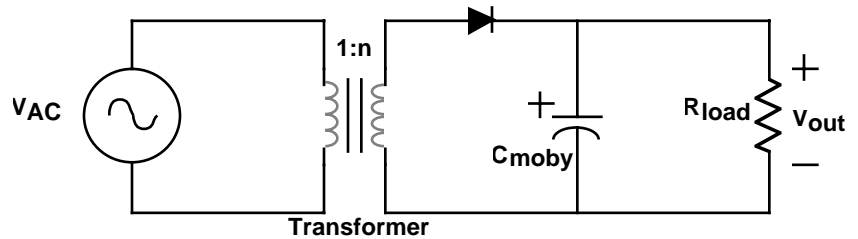


The operation of this circuit is straightforward. When V_{AC} is in the positive part of the cycle, a positive voltage is produced on the secondary side of the transformer. This turns on the diode, allowing most of that voltage to be seen across the load. When V_{AC} goes negative, the secondary side is also negative, and the diode is off. No current flows in the load during the negative half of the cycle. Hence the name half-wave rectifier. The figure below shows typical waveforms. Diodes used in these sorts of circuits are specially constructed to handle the large currents, and are often called rectifier diodes, or just rectifiers for short.

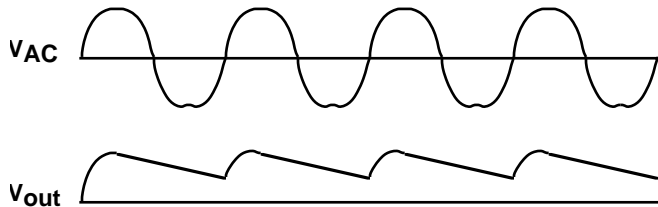


Looking at the output voltage, you should be struck by how little it resembles the output of a battery. While this output voltage is now positive only, it certainly is very bumpy, and not even on half the time. What we need is a way to smooth things out - something which will store some of the energy during the on times, and release it during the off times. The ideal solution is a capacitor.

In the circuit below, we have added a large capacitor (labelled C_{moby} , after the whale of enormous dimensions) which can store energy during the on times, and release it during the off times. What happens now is that the diode only turns on



when the voltage on the cap is about 0.7V below that coming out of the transformer. The transformer only charges the cap upwards - the diode insures this. Meanwhile, the load discharges the cap with our standard RC time constant. The circuit must be carefully designed so that the time constant is much longer than the AC cycle time. In any case, there will be some amount of *ripple* on the output as the cap discharges between pulses. The sketch below illustrates what happens.

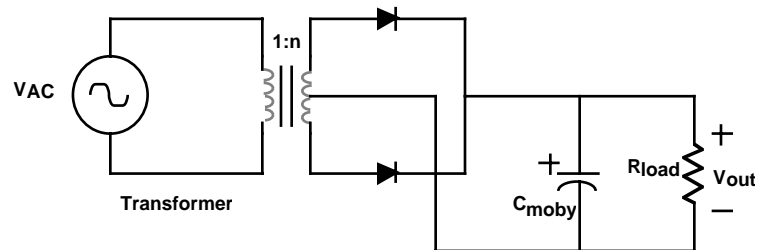


There is something else new in this circuit. Notice how the bottom plate of the capacitor is shown with a curve, and the top is marked with a plus sign. That is

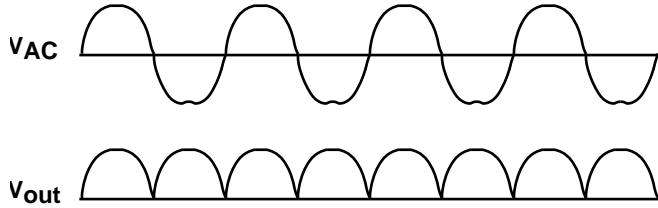
because special capacitors are required to get a very high capacitance in a reasonable space. In particular, you will be using *electrolytic capacitors* which are constructed using a paper soaked with an electrolyte. This gives enormous capacitances in a small volume, but also results in the cap being *polarized* - i.e. it only works with one polarity of voltage. If you reverse the polarity, hydrogen can dissociate from the internal anode, resulting in a nasty tendency to explode. Therefore, it is imperative that you guarantee via circuit design that electrolytic caps never see the wrong polarity. Electrolytic caps always have their polarity clearly marked, often with a bunch of minus signs pointing to the negative terminal. You should have a 1000 μ f electrolytic cap for use in power supply circuits. Caps have a maximum voltage rating which should never be exceeded.

(A side note on electrolytic caps: there are now specially constructed, nonpolarized electrolytic caps available. These are somewhat more expensive, but at least eliminate the problem of accidentally inserting caps backwards.)

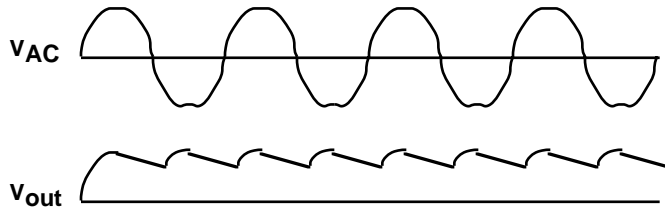
While the half-wave rectifier has the virtue of simplicity, it lacks efficiency and elegance. A better solution would use the power available from the transformer on both sides of the cycle. Systems of this type are called *full-wave rectifiers*. In the days when transformers were cheaper than diodes (shortly after the Jurassic period), center-tapped transformer circuits using two diodes were particularly common. A center-tapped transformer adds a connection in the middle of the secondary coil. This allows you to treat it as two separate secondary coils, with a common middle connection. In the circuit shown below, each of these halves are used on opposite parts of the cycle, basically stringing two half-wave rectifiers together.



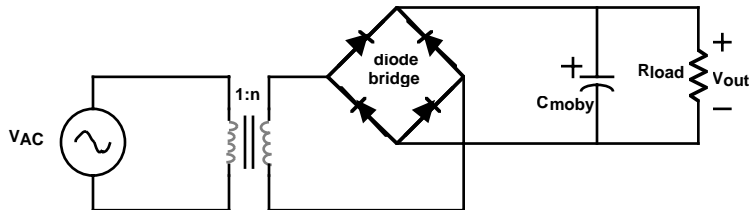
The output of this power supply has significantly less ripple because the capacitor is refreshed twice as often. If we were to remove the cap and look at the waveform, it would look something like the figure shown below. The AC waveform is said to have been rectified. However, this is hardly a constant voltage, and while the addition of a cap helps clean things up, there may still be a fair amount of ripple. The



next figure shows how this is cleaned up by the capacitor. You should compare this with the ripple of the half-wave circuit.



In an era where diodes sell for pennies, a better solution is a *full-wave bridge rectifier*. This particularly clever circuit uses four diodes to always turn the output voltage the right direction and is shown in the figure below. If you examine this closely,

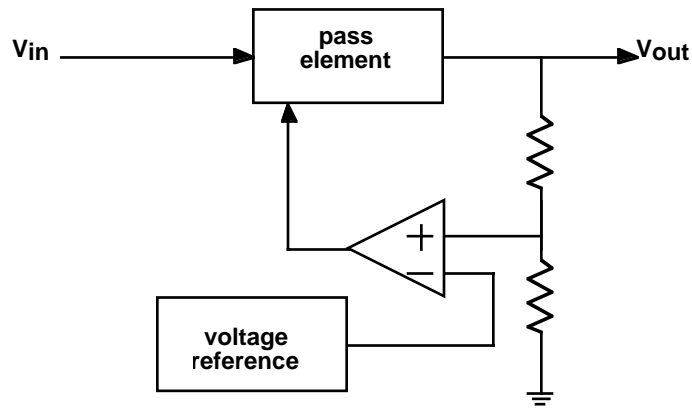


you will notice that which ever side of the transformer output is positive has a path through a diode to the positive side of the cap. Similarly, which ever side is negative has a path to the negative side of the cap. At any given moment, only two diodes on opposite sides are conducting, and this switches every half-cycle.

What is a Voltage Regulator?

The full-wave bridge rectifier is an elegant little circuit for converting an AC signal into a DC one. However, it still has some significant shortcomings. While better than a half-wave circuit, it still suffers from significant ripple due to the cap discharging in the short space between pulses. How do we get a really clean 5V output?

The solution is a voltage regulator. This is a circuit which actively monitors the output voltage, and only allows enough current to pass to reach precisely the correct voltage. A rough block diagram is shown below. The main components are a con-



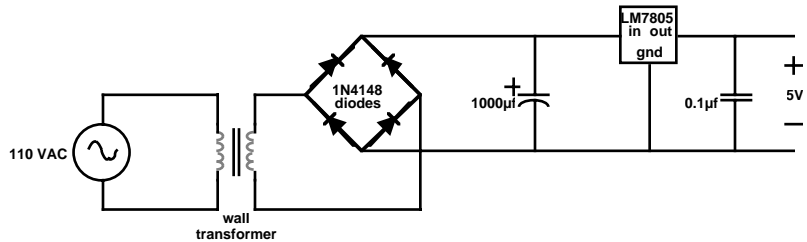
trollable pass element (a transistor), a voltage reference circuit, and an op amp to drive the pass element such as to force the output to be the proper voltage. (Usually, the voltage reference puts out about 1.2V, so the output is first divided down by a voltage divider to this level before the comparison in the op amp.) Of course, the input voltage must be higher than the output voltage in order for this scheme to work.

This type of voltage regulator is called a *linear regulator*. All it does is selectively restrict the flow of current from the input to the output. Thus, the current on the input must be at least as high as the current on the output. (It will be a little higher because some current is needed to drive the regulator electronics.) If the input voltage is much higher than the output voltage, there is a large voltage drop across the pass element, and possibly most of the power from the input gets wasted right here. Linear regulators are notorious for their poor efficiency.

Switching regulators have recently become quite popular. These are far more complex circuits which use energy storage components to regulate the output voltage in a very efficient manner. They commonly reach efficiencies over 90%. Most personal computers use switching power supplies for this reason. While it is beyond the scope of this class to discuss how these work, you should be warned that switching supplies can have some rather nasty surprises for you, such as minimum loading requirements, and lots of high frequency noise.

Returning to the problem at hand, your design of a linear regulator is made trivial by the existence of IC linear voltage regulators. These are typically three pin devices - an input, an output, and a ground. They could hardly be easier to use. You should have an LM7805, which is a 5V voltage regulator. Take a moment to find this part, and examine the data sheet. Applying a voltage between 7.5V and 35V to the input will give an output voltage of 5V.

The complete power supply circuit looks like this:



After you have thoroughly tested this circuit, you can use it to power your Stamp. However, unlike the 9V battery connection, the output of this supply will be connected directly to the +5V pin on the Stamp.

What's Next?

At this point, we've pretty much covered most of what we wanted to say about basic components. In the next section we'll be looking at using these elements to solve instrumentation and signal conditioning problems.

The Problem

You reach into the shower and turn on the water. It feels too cold, so you turn up the hot, and turn down the cold. A few moments later, the really hot water has made it up to the shower, and you readjust again. Finally, you step under the water and put shampoo in your hair. Somewhere, somebody flushes a toilet, and you scream slightly as you attempt to jump out of the path of the now scalding water.

Does this sound all too familiar? While indoor plumbing may be what separates the United States from third world countries, it has hardly been perfected. Your goal is to create a better shower - one that automatically adjusts the flow of hot and cold water to achieve the desired temperature. Sometimes, this won't be possible (for example, if there is no hot water left), and you should provide a warning to the unfortunate bather as soon as possible. But most of the time, you should be able to maintain the temperature with some precision.

We will presume that you have access to a model shower with pump driven hot and cold water supplies which you can turn on and off electrically. Because the pumps draw significant current, they will have separate power supplies and you will use the relays to control them. By turning the pumps on and off, you can control the relative amounts of hot and cold water.

The model showers also have a small mixing chamber. This allows the water to reach a uniform temperature before being expelled. You will construct a temperature measurement device to place in this mixing chamber. The temperature should be indicated on your 7-segment LED displays most of the time. You will also need to provide a mechanism for the user to be able to select the desired temperature. An audible signal should indicate when the temperature cannot be properly maintained. Finally, a float switch will indicate when the mixing chamber is full, so you can maintain an appropriate flow rate without overflowing.

What You Need to Know

This project brings together many concepts we have explored in previous chapters. In addition, here are a some other things you need to know:

- How do I measure temperature?
- What is an appropriate type of A/D conversion for measuring temperature?
- What is a relay, and how do I drive it?
- How do I make noise?
- What algorithm do I use to control the pumps?

How Do I Measure Temperature?

There must be literally hundreds of different ways to generate an electrical signal which depends upon temperature. For example, the resistance of some materials varies dramatically with temperature and these are used in the construction of *thermistors* - temperature variable resistors. *Thermocouples* are created by joining two dissimilar metals in a circuit. These create a voltage difference via the Seebeck effect by having the two junctions at different temperatures. *Thermostats* use two dissimilar pieces of metal with different thermal coefficients of expansion which have been welded together, and thus tend to bend in response to temperature changes. Simple contacts, potentiometers, etc. can then convert the motion into an electrical signal. There are temperature sensitive capacitors, and diodes change their I-V characteristics with temperature. Infrared detectors can measure temperature optically. The list goes on and on.

With all of these possibilities, it might be hard to decide what type of temperature sensor to use in a given application. However, to simplify your life, we have already selected a device. It is an LM35D temperature sensor integrated circuit, which is remarkably easy to use. It has three leads, power, ground and output, and produces a voltage which is proportional to temperature in degrees Centigrade. In fact, it is calibrated to output $10\text{mV}/^\circ\text{C}$, with high repeatability and accuracy. This would be an excellent time to examine the specifications in the LM35D datasheet.

Integrated sensors like the LM35D are revolutionizing the way physical quantities are measured. Recent advances in micromachining technology allow companies to build the mechanical structures of sensors along with sophisticated electronics on the same integrated circuit. For example, automobile air bags require a method of quickly detecting when a collision has occurred. A typical integrated accelerometer has a tiny mass, held in position by electrostatic forces. The circuitry automatically varies the strength of the electric field to hold the mass in a constant position via a feedback control system. A voltage is output which is directly proportional to acceleration, with almost no delay, and no mechanical ringing. This incredibly sophisticated electromechanical system fits on a single chip, is extremely reliable and costs very little in large quantities.

In the future, you can expect integrated sensors to get smarter and smarter. A recent single chip temperature sensor made by Dallas Semiconductor is a complete temperature measurement system. In addition to measuring temperature, it does the analog to digital conversion and provides a network interface so that a single wire (plus ground) can be used to transmit temperature data from many different sensors, each with a unique ID. And it does all this for under \$10. There are no circuits to design - only software to write for the micro trying to read these sensors.

Unfortunately, you are learning about Electrical Engineering in the midst of this tumultuous transition. So while sensors with digital interfaces are becoming more common, they are by no means the rule. It is still important that you know something about reading an analog sensor - properly conditioning the signal, and converting it to digital form. Given current economics, the LM35D is a cheaper solution than most of the direct digital reading temperature sensors. But be forewarned that this is unlikely to last. However, for many sensor types, it will be quite a bit longer before the conversion to digital is complete.

To use the LM35D to measure temperature could hardly be easier. Connect up power and ground, and measure the output voltage with a meter. With $10\text{mV}/^\circ\text{C}$, a room temperature of 20°C will register as 0.2V.

What is an Appropriate Type of A/D Conversion for Measuring Temperature?

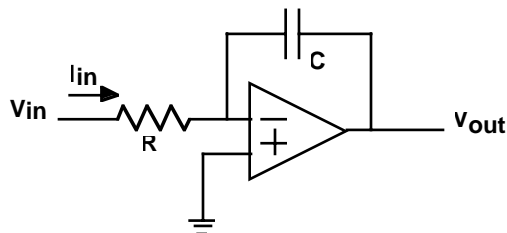
While it was amazingly easy to generate a voltage proportional to °C, reading this with the Stamp is a bit more challenging. In previous chapters, we discussed two types of A/D converters - a successive approximation converter which played high-low comparing the input to a DAC, and a simpler technique where we varied an RC time constant to measure a changing resistance. Unfortunately, neither circuit can be directly applied to this problem. The former depended upon an accurate R-2R ladder, which is painful to construct for a large number of bits. The latter measured a resistance rather than a voltage. We need yet another technique.

Temperature changes quite slowly. So you don't need a high speed converter. However, you might want a fair degree of precision. And, as always, you want it as inexpensively as possible.

The classic solution to this problem is the *Dual-slope Integrating ADC*. This technique is about as elegant as they come - simple and precise. Most digital meters use it (or one of the many variants), achieving accuracies up to one part in 100,000,000, which is 8 digits of precision! (If that doesn't impress you, try measuring something easy, like distance, to 8 digits of precision...)

Before we can explain this wonderful circuit, you need to know what an integrator is. So we return for a moment to the subject of op amp circuits.

The circuit shown below is an inverting integrator. Recall that the high gain of the



op amp will try to force $V_+ = V_-$, and that there is no current into the op amp inputs. In this case, we refer to the V_- input as the *virtual ground* node because the op amp will try to keep it at ground. Thus, the current into the circuit, I_{in} , is just V_{in}/R .

Since this current does not flow into V_- , it all goes to charging the capacitor. For a positive current I_{in} , the V_- side of the cap becomes positively charged with respect

to the V_{out} side. To maintain the V- node at ground, V_{out} must be going increasingly negative. In equation terms:

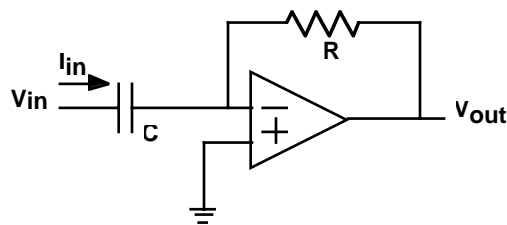
$$I_{in} = \frac{V_{in}}{R}$$

$$V_{out} = -\frac{1}{C} \int I_{in} dt = -\frac{1}{C} \int \frac{V_{in}}{R} dt = -\frac{1}{RC} \int V_{in} dt$$

The output voltage is the integral of the input voltage, times $-1/RC$. That is why this is called an inverting integrator.

It is important to get the feeling for what is going on here. By tying one end of the resistor to virtual ground, the current through it will only be a function of the input voltage. (In contrast, the current depends upon both the input and the cap voltage for simple RC circuits.) This current then charges the cap, since it has nowhere else to go. Of course, the virtual ground concept depends upon high gain negative feedback. If you really wanted to truthfully explain what was going on, the argument is more complex. As the cap starts to charge, the V- terminal is driven positive. The op amp correspondingly swings the output in the negative direction, lowering V- until it again approximately equals $V+$.

While we're on this topic, we should point out that it is equally easy to build a circuit which differentiates. Basically, this just involves swapping the cap and the resistor, as shown below. The analysis of this circuit is simplified by assuming a vir-



tual ground at the V- terminal. In this case, the current I_{in} is given by:

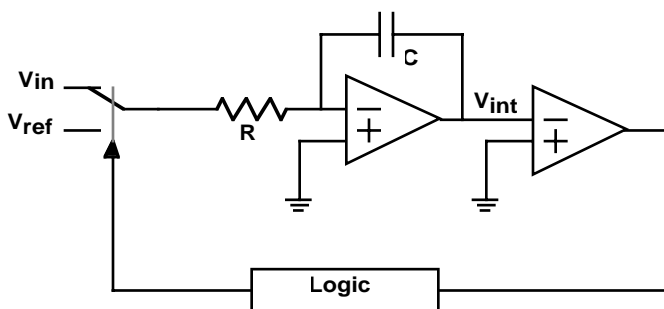
$$I_{in} = C \frac{dV_{in}}{dt}$$

This current must flow through the resistor, giving:

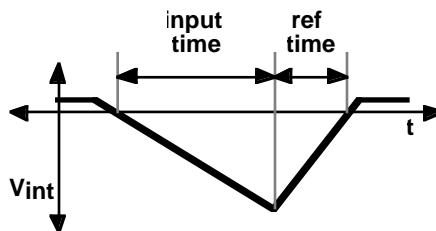
$$V_{out} = -I_{in}R = -RC \frac{dV_{in}}{dt}$$

Thus the output is proportional to the derivative of the input.

Now that we understand how easy it is to build an integrator, we can describe how dual-slope integration works. The basic circuit is shown below. Initially, assume



that the integrator output is slightly positive. The top switch is closed, and the input of the integrator is connected to V_{in} . (V_{ref} is left unconnected at this point.) Assuming that V_{in} is a positive voltage, the inverting integrator output will fall at a rate proportional to the input voltage. After the comparator has detected the zero crossing, the integration of the input is allowed to continue for a fixed amount of time. After that, V_{in} is disconnected, and a negative reference voltage, V_{ref} , is applied, causing the inverting integrator output to head upwards. The logic then times how long it takes until the comparator indicates that zero has been crossed in the opposite direction. The length of time it takes for this second integration will be proportional to the input signal. These signals are diagrammed below.



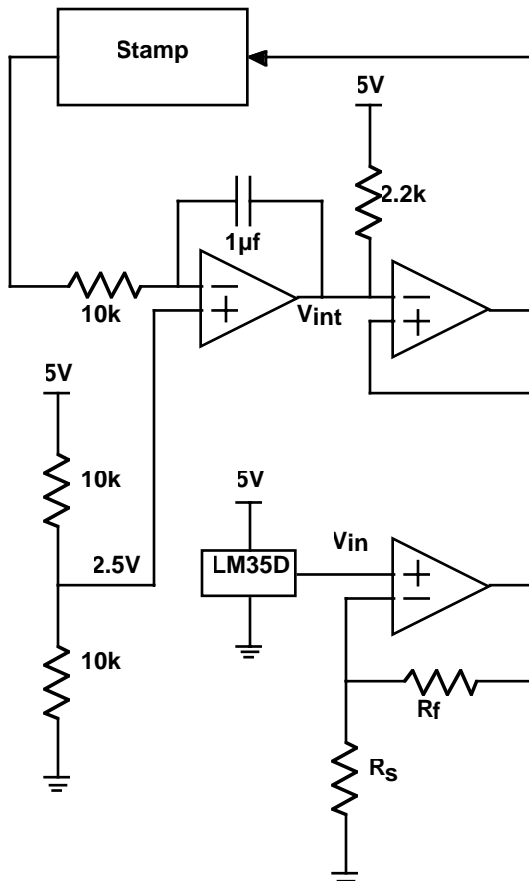
The beauty of this system is its insensitivity to all sorts of errors. It does not depend upon the specific values of the cap or the resistor. And even if these change slowly with temperature, so long as they are fairly constant over a single conversion cycle, the system is still accurate. Offset errors in the comparator (which cause it to flip at something other than equal inputs) also don't matter because this equally affects where the integration starts and stops. This circuit is a wonderful example of how analog circuit designers manage to build precision systems with imprecise components.

Unfortunately, we have not provided you with a key component to make a dual-slope convertor - good, electronically controllable switches. JFET and CMOS switches are commercially available which perform as almost ideal switches for these sorts of applications. But since we didn't make these available, you have probably surmised that we had something else in mind. We do.

Rather than building a dual-slope convertor, we are asking you to build a somewhat simpler, but less accurate system. The idea is straightforward: use an integrator to generate a linearly increasing voltage of known slope. Compare this to the input voltage. When they cross will be proportional to the input signal. This circuit suffers from most of the errors that a dual-slope avoided, but is far easier to build on a single 5V supply with our selection of components.

The full temperature conversion circuit appears below. It uses three op amps (3/4 of an LMC660 running off a single ended 5V supply). The first op amp is used to create the inverting integrator. In order to work well off of the 5V supply, the V+ terminal is set to 2.5V. (You may wish to adjust this slightly downwards since this is at the edge of the common-mode input spec.) So instead of creating a virtual ground, we are virtually at 2.5V on the V- input. Switching the Stamp output between 5V and 0V will generate positive and negative currents respectively into the resistor, which will in turn cause the output to integrate down and up. Of course, on a 5V supply, you can't integrate beyond the 0-5V range.

Note that there is a 2.2k resistor to 5V on the output of this amplifier. This is just one of those nasty little reasons why analog circuit designers have job security. Although an ideal op amp, and many real ones, would not require this resistor, the particular quirks of the LMC660 require it in this case to achieve stable operation. If you try running your circuit without it, you may notice some fuzz when you examine the output of the integrator on an oscilloscope. If you zoom in on the fuzz, you will see that it is in fact a small amplitude, high frequency oscillation. A full understanding of what is going on here is a topic worthy of graduate study in analog integrated circuit design, and hence far beyond the scope of this book. We will leave



it to say that adding the resistor increases the current in the output stage of the op amp, and this increases the speed of that stage, making it less likely that the feedback signal will return late, and possibly in phase, creating an undesired oscillation. If you read the datasheet carefully, you will see that it specifically mentions that adding the pull-up resistor can help cure oscillations when driving heavy capacitive loads, so you didn't necessarily need a graduate degree to think of this.

Our old friend, the non-inverting amplifier, is used to increase the signal level coming from the LM35D. This helps on several fronts. First, larger signals are easier to work with. Second, the offsets of the amplifiers are of the same order as the low end of the signal, so gaining the signal up makes these appear relatively smaller. However, you should choose the resistors so that the maximum signal does not exceed 2.5V - the common mode range limit. Finally, you will see that the integrator does

not go all the way down to zero, and this limits the effect of this missing part of the scale.

The last amplifier is just used as a simple comparator. This signals when the amplified sensor output no longer exceeds the integrator output.


To use this circuit, you will have to set the Stamp output high for a time long enough to integrate all the way down. (Extra time won't do any harm.) Then you set the output low, causing the integrator output to ramp up. Timing how long it takes for the comparator output to trip will be a measure of the temperature. Of course, you will have to calibrate this appropriately to get temperature in standard units.


What is a Relay, and How Do I Drive It?

Now that we have figured out how to measure temperature, we need a way of controlling the pumps. The pumps are designed to run off of 12V supplies, and draw a couple of amps. Obviously, this is not directly compatible with the output drive capabilities of the Stamp, which is limited to 25mA at 5V. Worse yet, these loads look inductive, with frequent voltage spikes. We will want to be very careful about how we approach this problem.

Rather than drive the pumps directly, we will use *relays* to isolate the pump circuits from the Stamp circuits. Relays are mechanical switches which are actuated via electromagnets. Relays and switches are described by the number of *poles* and the number of *throws*. A single pole, single throw (SPST) switch is a simple switch. A single pole, double throw (SPDT) switch connects a common contact alternately to a *normally closed* contact and a *normally open* contact. The common configurations are diagrammed below.

 SPST

 SPDT

 DPST

 DPDT

You should have two 5V, double pole, double throw (DPDT) relays. Frequently, the tops of the relays are imprinted with the connection diagram. Otherwise, you will need to look at the datasheet. You should confirm that the relay contacts can handle the currents and voltages required by the pumps. Assuming you can figure out how to wire these as simple on/off switches, this reduces the problem to driving the relays from the Stamp.

Although your relays are meant to be driven from a 5V supply, they are likely to require about 100mA to function. (If your relays require substantially more current than this, you should obtain ones with a higher coil resistance.) This is a lot closer to the Stamp's drive capability, but is still too high. You will need some type of amplifier. You should have a DS2003 seven channel, high current/voltage Darlington driver chip. It falls into the category of *peripheral driver* chips. This IC contains seven individual power amplifiers which can each drive 350mA loads - more than sufficient to drive the relays. Once again, we ask that you carefully examine the datasheet for this part.

Using the DS2003 is not difficult, but requires some explanation. There are seven inputs, and these are compatible with your Stamp outputs. That part could not be easier. Pin 8 is a ground connection, and that's pretty simple too. The outputs are a little trickier. Effectively, they are switches to ground. Setting the input to a channel high allows current to flow into the output to ground. (That may sound odd, but that's what happens.) So if you are going to drive a relay, one end gets connected to 5V, and the other gets connected to a channel output.

Finally, there is the matter of pin 9, labelled common. There is a diode from each output to this common pin, which should be connected to the positive supply. Why? At first glance, this may not make much sense. The direction the diode is pointed means that current can only flow when the output is higher than the supply. Does this ever happen?

The short answer is generally no. However, if the load is an inductor, suddenly turning off the current flow will cause a large voltage spike ($V=Ldi/dt!$) which could drive the output much higher than the positive rail, and potentially fry the driving transistors inside the DS2003. In this case, these *free-wheeling diodes* turn on, and give the current someplace to go, minimizing the voltage spike. Circuits meant to suppress the voltage spike which occurs when turning off the current flow in an inductor are called *snubber circuits*.

How Do I make Noise?

Now that you know how to drive a relay, driving a speaker is not much more difficult. A speaker is usually an electromagnet which pushes a magnet in and out against a cone, generating sound pressure waves. Not unlike the relay, it also looks like an inductive load, and can be effectively driven by the DS2003.

Hopefully, you have a small 8Ω speaker. You may connect this to the Stamp via the DS2003 just like the relays. However, you may wish to add a small series resistance to make the tones a little less obnoxiously loud. On the Stamp side, there are special instructions for generating tones. Examine the *freqout* command in your Stamp documentation.

Many small speakers today are actually piezoelectric. If you happen to have one of these, you can directly drive it from a Stamp output. There are also piezoelectric buzzers which contain internal oscillator circuits. These are easier to use, however, you don't get to change the frequency of the sound.

What Algorithm Do I Use to Control the Pumps?

You now have all of the pieces, but putting them all together to make an effective system can be challenging. There is a great deal of room for creativity here, but we will provide some general hints as to how to proceed.

The key loop in your program should turn on the hot and cold water pumps for some amount of time. There are two factors to consider: keeping the temperature correct, and keeping the level in the mixing chamber correct. The temperature will be most affected by the ratio of hot water to cold water. The level will be most affected by the percentage of time the pumps are operating. We suggest you keep these two things as separate as possible in your program. For example, to change the temperature you should change the amount of time running the hot water pump versus running the cold water pump - not the total amount of pump running time.

What's Next?

In this chapter, we touched on circuits that used energy storage elements in a more complex fashion than an RC circuit. In the next chapter, we will expand upon this greatly, showing that we can use our old friend, Ohm's Law, to quickly solve circuits containing multiple inductors and capacitors using the concept of complex impedance.

The Problem

In science fiction movies, you will often see people conversing with their appliances. Commands such as “Lights on!” or “Screen off!” cause the appropriate actions to occur. Wouldn’t it be cool to build a system using your Stamp which recognized spoken commands, and performed actions accordingly?

Unfortunately, it’s not going to happen. The sad truth is that speaker independent, continuous speech recognition is a really hard problem, and requires far more computational horsepower than is available with the Stamp. However, we could try to do something far easier - recognize a simple pattern of sound.

Your goal for this chapter is to create a system which toggles the state of a relay in response to clapping your hands twice in rapid succession. (You may have noticed similar devices advertised on television.) You should endeavor to make this sound sensitive switch as selective as possible: i.e. it should only respond to the appropriate pattern of claps, and not to other noises.

What You Need to Know

Here are some things you need to know:

- How do I detect sound?
- How do linear systems respond to sinusoids?
- How do I generalize Ohm's Law?
- How do I detect a clap?

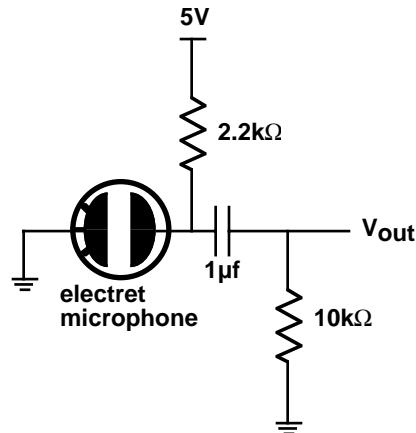
How Do I Detect Sound?

What we perceive as sound is simply rapid variations in air pressure. *Microphones* or *mics* for short, are devices specifically designed to detect these pressure fluctuations, and produce a proportional electrical signal. Typically, they respond only to changes in pressure - not absolute pressure.

There are many types of microphones. Most have some type of flexible diaphragm which moves in response to the pressure changes. In *carbon mics*, the diaphragm presses up against loosely packed carbon granules, changing the resistance according to the pressure variations. *Moving-coil mics* are essentially speakers run backwards - a diaphragm is connected to a coil which moves back and forth in a magnetic field producing a proportional voltage. *Piezoelectric mics* couple the diaphragm to a crystal, which generates a voltage in response to changes in stress. *Condenser mics* use the moving diaphragm as one side of a sound responsive capacitor. *Electret mics* are a type of condenser mic where a permanently polarized material is used for the fixed plate, keeping a charge on the varying capacitor. Typically, piezoelectric and electret mics contain a small built in amplifier.

We presume that you will be using an electret microphone. As always, read the data sheet carefully. There are two common configurations for these sort of mics - three lead and two lead. For most three lead devices, there is a positive supply lead, a ground and a signal output. For two lead devices there is only a ground and a signal output. However, the device still needs power, and this is usually provided by a pull-up resistor between the positive supply and the signal output. For either type of device, the output signal is meant to be *capacitively coupled* to an external amplifier. A typical schematic for a two lead device is shown below.

What does it mean to be capacitively coupled? If you were to examine the signals in



this circuit with a scope, you would see that the output of the electret mic (taken at the bottom of the $2.2\text{k}\Omega$ pull-up resistor) moves around in response to sound, as does V_{out} . However, the electret output bounces around some positive voltage (typically a volt or two) while V_{out} bounces around zero volts!

In a previous chapter, we noted that in response to a transient, a capacitor initially looked like a short, but after a long time, looked like an open. If we consider a sine wave applied to a capacitor, you might imagine that if the voltage was moving very quickly (i.e. a high frequency), then the cap would look like a short. For very slow signals (i.e. low frequency), things are hardly moving, so you would expect the cap to look like an open. This is pretty much what happens.

The output of the electret consists of the audio signal, sitting on top of a constant DC voltage. To this DC voltage, the capacitor looks like an open circuit, and it has no effect on the output. On the other hand, to the fast moving audio signal, the cap looks like a short, and this signal appears as V_{out} . The coupling capacitor has given us a way to pass only the varying part of the signal, and to block the DC, or stationary part of the signal.

This all begs an interesting question: if the cap looks like a short to high frequencies and an open to low frequencies, when does it switch over? To answer this question, we will have to explore AC circuit theory in a bit more depth, and figure out what frequency response and filtering are all about.

How Do Linear Systems Respond to Sinusoids?

Circuits consisting of ideal resistors, capacitors and inductors are linear circuits. Recall that linearity implies that if you were to add up the responses to a number of different inputs that would give the same result as the response if all the inputs were summed together and applied at once.

As we have already seen, sinusoidal signals naturally arise in many practical situations. So we ask, what happens if you apply a sinusoidal signal to a linear system? (We'll assume that you've been applying this signal forever, so we are looking for some sort of *steady-state* response.) The remarkable answer is that all the currents and voltages in the circuit will also be sinusoidal at the same frequency as the input, but possibly with different phases and amplitudes!

Why might this be so? Consider a sinusoid of the form:

$$v(t) = A \cos(\omega t + \theta)$$

Using the all-but-forgotten trig identities, we realize that this can be rewritten as:

$$v(t) = (A \cos \theta \cos \omega t - A \sin \theta \sin \omega t)$$

Examining this equation, we realize that a sinusoid with some phase and amplitude can be thought of as the sum of a sin and cosine at the same frequency, but usually different amplitudes determined by the phase. Letting θ range from 0 to 90° smoothly changes the equation from being all cosine to all sine. The upshot is that any sum of a sine and cosine at the same frequency results in a sinusoid with some generally different amplitude and phase, but at the same frequency.

Taking this a step further, you can show that summing any two sinusoids at the same frequency will sum to another sinusoid at that same frequency. That's because each sinusoid can be broken down into its sine and cosine parts, and summing the sine or cosine parts results in a sine or cosine of some amplitude with everything at the same frequency. This is just another sinusoid, at the same frequency. Mathematically, this is:

$$v(t) = A \cos(\omega t + \theta) + B \cos(\omega t + \phi)$$

$$v(t) = (A \cos \theta \cos \omega t - A \sin \theta \sin \omega t) + (B \cos \phi \cos \omega t - B \sin \phi \sin \omega t)$$

$$v(t) = (A \cos \theta + B \cos \phi) \cos \omega t - (A \sin \theta + B \sin \phi) \sin \omega t$$

which is easily recognized as another sine wave at the same frequency, but with different amplitude and phase.

Ultimately, this explains why linear systems respond to a sinusoidal input with sinusoids at the same frequency, but with varying amplitudes and phases. Since all that linear systems do is sum linearly, every operation ends up with another sinusoid at the same frequency. Therefore, the response of any linear system to a sinusoid at a given frequency can be characterized by just two parameters: the *gain* (which is the ratio of the new amplitude to the old amplitude) and the *phase shift* (which is how much the phase has changed).

We can take this even further. If we supply to a linear circuit an input consisting of two sinusoids at different frequencies, the output can be described as the sum of two new sinusoids at the same frequencies as the input sinusoids. This reasoning can be extended to any number of sinusoids.

We won't prove it here, but it can be shown that any waveform can be modeled as the sum of some number of sine waves. (Unfortunately, the number required for a nearly perfect model can be infinite, but that's mathematics for you.) The result is that if you can predict the behavior of a linear system to sine waves at all frequencies, then you can predict the behavior to any input - not just sinusoids!

How Do I Generalize Ohm's Law?

Sinusoids are so important to understanding linear systems that electrical engineers have devised a very clever scheme for representing them. The basic idea comes from Euler's theorem:

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

where j is the square root of negative one. (Unfortunately, "i" was already the symbol for current, so EEs use "j".)

Consider the equation:

$$Ae^{j(\omega t + \theta)} = A \cos(\omega t + \theta) + Aj \sin(\omega t + \theta)$$

the real part of this expression is a sinusoid with some amplitude, A, and some phase, θ . This can also be written in the form:

$$Ae^{j(\omega t + \theta)} = Ae^{j\theta} e^{j\omega t}$$

Here we have separated out the phase term, $e^{j\theta}$.

The idea is to use complex numbers to represent sinusoids. In particular, we will represent amplitude as the magnitude of a complex number, and phase as the angle of that complex number. To obtain the actual real signal, we just need to look at the real part of our result.

Let's apply this idea to a capacitor. The defining equation for a capacitor is:

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

Applying a sinusoidal voltage gives:

$$I = C \frac{dA \cos(\omega t + \theta)}{dt} = -\omega CA \sin(\omega t + \theta)$$

The current has the same frequency, but the phase leads the input voltage by 90° , and the amplitude is dependent on the radian frequency, ω .

Alternatively, using complex numbers, we have:

$$I = C \frac{dAe^{j(\omega t + \theta)}}{dt} = Aj\omega Ce^{j(\omega t + \theta)}$$

Taking the real part of this expression yields:

$$I = \text{Re}[Aj\omega Ce^{j(\omega t + \theta)}]$$

$$I = \text{Re}[Aj\omega C\{\cos(\omega t + \theta) + j\sin(\omega t + \theta)\}] = -\omega CA \sin(\omega t + \theta)$$

which is identical to the previous result.

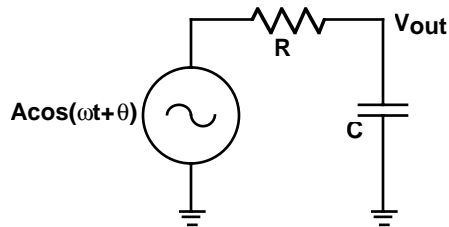
The neat thing about the complex number representation is that it lets us make a simple generalization to Ohm's Law to allow for capacitors. Rather than talking about the resistance of a capacitor, we talk about its *impedance*, which we will denote with the letter Z , rather than R . This will be a complex number which in one fell swoop, relates the voltage and the current in terms of amplitude and phase at a given frequency.

Specifically:

$$Z = \frac{V}{I} = \frac{Ae^{j(\omega t + \theta)}}{Aj\omega C e^{j(\omega t + \theta)}} = \frac{1}{j\omega C}$$

The term, $j\omega$ comes up so often, that we often abbreviate it with the symbol s . (Why s ? If you study linear systems further, you will see how this is all intimately related to Laplace transforms.) So the impedance of a capacitor is $1/Cs$.

Is this really useful? Absolutely! Consider the circuit shown below:



We can solve this circuit just as if the capacitor were a resistor of value Z . This is then quickly recognized as a voltage divider, giving:

$$V_{out} = V_{in} \frac{Z}{R + Z} = V_{in} \frac{\frac{1}{Cs}}{R + \frac{1}{Cs}} = V_{in} \frac{1}{1 + RCs}$$

Next, we wish to find the real part:

$$V_{out} = \text{Re} \left[V_{in} \frac{1}{1 + RCs} \right] = \text{Re} \left[A e^{j(\omega t + \theta)} \frac{1}{1 + RCj\omega} \right]$$

To remove the imaginary part from denominator, we multiply top and bottom by the conjugate.

$$V_{out} = Re \left[A e^{j(\omega t + \theta)} \frac{1 - RCj\omega}{(1 - RCj\omega)(1 + RCj\omega)} \right] = Re \left[A e^{j(\omega t + \theta)} \frac{1 - RCj\omega}{1 + R^2 C^2 \omega^2} \right]$$

Finishing this up yields:

$$V_{out} = Re \left[A (\cos(\omega t + \theta) + j \sin(\omega t + \theta)) \frac{1 - RCj\omega}{1 + R^2 C^2 \omega^2} \right]$$

$$V_{out} = A \left[\frac{\cos(\omega t + \theta) + RC\omega \sin(\omega t + \theta)}{1 + R^2 C^2 \omega^2} \right]$$

This is a complex looking expression, but it is possible to get a reasonable understanding out of it by looking at the limits. For very low frequencies, this reduces to:

$$V_{out} = A \cos(\omega t + \theta)$$

This agrees with our earlier statement that the cap looks like an open to low frequencies, making the output follow the input.

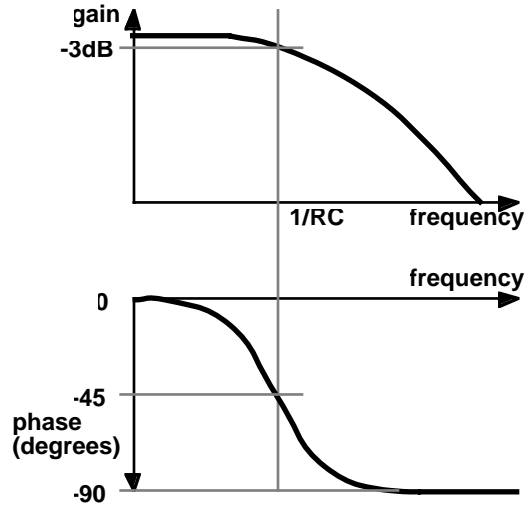
At very high frequencies, the expression becomes:

$$V_{out} = A \left[\frac{\sin(\omega t + \theta)}{RC\omega} \right]$$

This says that at high frequencies, V_{out} is now 90° out of phase with the input, and dropping rapidly in magnitude with increasing frequency. If the input frequency is high enough, the output will be essentially zero, agreeing with our observation that caps at very high frequencies look like shorts. Circuits like this, which pass low frequencies but attenuate high frequencies are called *low-pass filters*.

At what frequency does this filter switch from passing signals to not passing them? That's not really a fair question. In fact, it does not suddenly switch - it is a gradual transition. However, when the magnitudes of the real and imaginary parts of the denominator are equal, we are least able to make our approximations. This occurs at $\omega = 1/RC$. At this frequency, the sine and cosine parts are equal, which gives a phase shift of 45°. If you work through the math, you'll find that the amplitude is down by a factor of $1/\sqrt{2}$.

We can now plot a rough graph of the frequency response of this circuit - i.e. its gain and phase as a function of frequency. Typically, this is plotted on log-log paper, and is called a *Bode plot*. The sketch appears below.

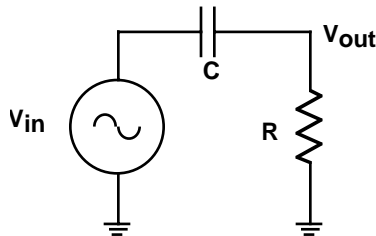


Note that gain is specified in *decibels*, which is defined by the equation:

$$dB = 20 \log \frac{V_{out}}{V_{in}}$$

which is really a unitless quantity.

Let's consider another example. The circuit shown below has the resistor and capacitor flipped from before.



Once again, this is really a simple voltage divider. We can immediately write the equation for V_{out} as:

$$V_{out} = V_{in} \frac{R}{R + \frac{1}{Cs}} = V_{in} \frac{RCs}{1 + RCs}$$

This equation gives the transfer function from V_{in} to V_{out} . First, let's get the imaginary part out of the denominator by multiplying by the conjugate.

$$V_{out} = V_{in} \frac{RCj\omega(1 - RCj\omega)}{(1 + RCj\omega)(1 - RCj\omega)} = V_{in} \frac{RCj\omega + R^2C^2\omega^2}{1 + R^2C^2\omega^2}$$

We can directly get at the magnitude and phase response of this circuit from this equation. The magnitude of a complex number is the square root of the number times its conjugate. So the gain is:

$$gain = \sqrt{\frac{(R^2C^2\omega^2 + RCj\omega)(R^2C^2\omega^2 - RCj\omega)}{(1 + R^2C^2\omega^2)^2}} = \sqrt{\frac{R^4C^4\omega^4 + R^2C^2\omega^2}{(1 + R^2C^2\omega^2)^2}}$$

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

This is a *high-pass filter*. At low frequencies, the gain is very small. At high frequencies, it is roughly unity (or 0dB). Again, the most ambiguous point is where $\omega = 1/RC$. This is the -3dB point.

The phase will just be the angle of the transfer function. This is given by:

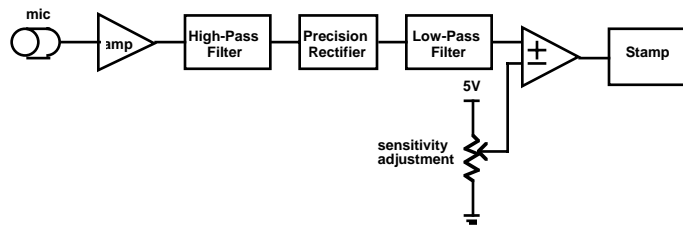
$$\angle T = \operatorname{atan} \frac{\operatorname{Im}[T]}{\operatorname{Re}[T]} = \operatorname{atan} \frac{\left(\frac{RC\omega}{1 + R^2C^2\omega^2} \right)}{\left(\frac{R^2C^2\omega^2}{1 + R^2C^2\omega^2} \right)} = \operatorname{atan} \frac{1}{RC\omega}$$

Before leaving this section, we should mention that an inductor has an impedance of sL . So you now have all the tools you need to solve for the steady state response of any linear circuit to sinusoidal input.

How Do I Detect a Clap?

Quick, sharp sounds tend to have lots of power at high frequencies. So one of the keys to being able to detect a clap will be the ability to detect high frequency sounds. Since most conversation is at a lower frequency, you would like to remove these sounds from consideration. The key to doing this will be the filters we discussed in the last section.

Let's look at a block diagram for a reasonable (though far from ideal) clap-activated switch.

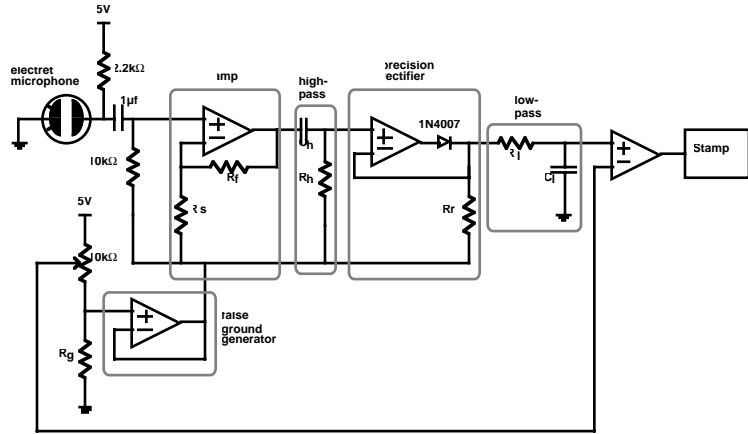


The microphone converts acoustic energy into electrical signals. In addition to the amp in the microphone, an external amp will be needed to get reasonable signal levels. The next step is a high-pass filter, to remove unwanted noises, such as conversation. A precision rectifier then removes the negative parts of the signal. This is so the following low-pass filter can average only the positive peaks to get a measure of the amplitude of the high frequency components of the signal. Finally, this is compared to a reference level. Any loud, high frequency sound will cause the comparator to trip, which can be detected by the Stamp. The pot provides a sensitivity adjustment to change the reference level for the comparator.

The detailed schematic is fairly complicated, and is shown below. A major issue you will have to confront right away is the problem of running off of a single 5V supply. To deal with this, an op amp is dedicated to creating a false ground approximately 1 volt above ground. This will give the op amps some room to swing both negative and positive with respect to our false ground.

The amplifier is a basic noninverting gain stage. We would suggest you start with a gain of approximately 100. This should get you up to a reasonable signal level. Be sure to examine the signals on a scope.

A simple high-pass filter attenuates low-frequencies. You will have to determine reasonable values for the resistor and the capacitor.



The precision rectifier is next. When the V₊ input is higher than the V₋ input, the diode is turned on, and the amplifier pours its silicon heart out trying to make V₊=V₋. When V₊ goes below V₋, the diode turns off, the op amp output shoots to ground, and the rectifier output is held at the false ground by the resistor.

The last piece of circuitry left to design is the final low-pass filter. This smooths out the rectified waveform. If you make the time constant too small, the comparator output will flutter up and down. If you make it too big, the response will be very slow. Getting the time constant right here may well be the trickiest part of the design.

Finally, you will need to write some software for your Stamp to recognize a double clap. We suggest you carefully consider, and actively look for the appropriate on and off times. You will be amazed how much clever software can clean up mediocre hardware!

What's Next?

That's up to you. As for this book, this is the end. We hope you enjoyed it!

Clap On...

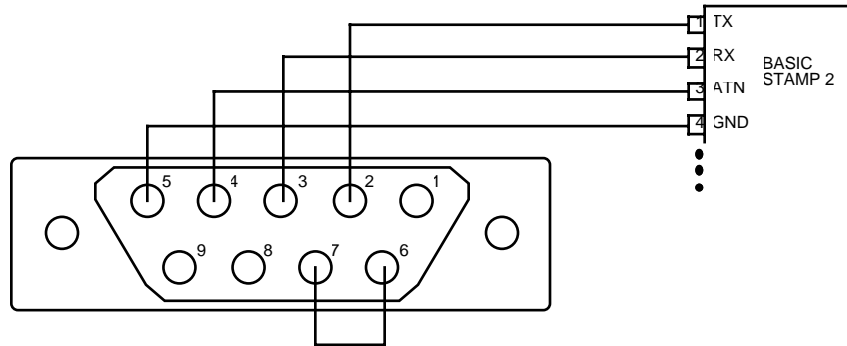
The BASIC Stamp 2 Serial Cable

Roll Your Own

Before you can make a Stamp do anything, you will need a way of connecting it to your development computer. While there are a number of commercial solutions, you can make up your own cable very easily and inexpensively. This discussion presumes that you have some experience with ohm meters and soldering. If you do not, seek out someone that does. It takes less than ten minutes to make up the cable once you have all the parts. I would consider this is a one-beverage level favor to ask a more experienced friend.

The Connections

Most modern PCs have a 9-pin serial port connector on the rear panel. Your goal is to connect pins 2, 3, 4 and 5 on that connector to pins 1-4 respectively on the Stamp 2. In addition, you need to connect pin 6 to pin 7 on the PC serial port. This is explained in the Stamp documentation and is diagrammed below.



The easiest way to do this is to buy an inexpensive 9-pin serial cable, and cut off the far end. This should leave you with a female DB-9 connector running out to a bunch of wires. Here's the hard part: You need to identify which wire goes to which pin. The pin numbers should be labelled on the connector, and you can use an ohm meter to trace out what goes where. Strip and tin the wires. Make the pin 6-7 connection by twisting the wires together and soldering. Then cover with electrical tape. The remaining wires can be directly plugged into the solderless breadboard in the appropriate order. Personally, I like to write this order down directly on the cable, so it can't get lost. If you want to get fancy, you could attach a 4-pin connector with 0.1" pin spacing to plug in next to the Stamp 2. Just remember to mark which pin goes to pin 1 on the Stamp 2.

Equipment

The Kit

There are quite a few parts that are needed to build all the projects in this book. To ease the acquisition problem, I have been working with Parallax to provide a parts kit to accompany this text. See the Parallax Stamps in Class web site, <http://www.stampsinclass.com>, for details.

When I taught this class, the kits which were sold through the bookstore. One kit was required for each lab group of two people. In addition to the kits, selections of resistors and capacitors were made available in the lab. Since this was a project oriented class, I wanted to give my students the tools to do much of the work outside of lab.

In case you wish to make your own kit, a reasonable parts list is shown below.

Count	Description	Comments
1	Plastic Box (Shoebox size)	
1	1660 Connection Solderless Breadboard	RSR 03MB104
1	BS2-IC BASIC Stamp 2	Parallax BS2-IC
1	Wire Stripper	RSR 060220
1	Cutters/Pliers/Screwdrivers Set	Jameco 99565

Equipment

Count	Description	Comments
1	12VAC Wall Transformer ($\geq 300\text{ma}$)	RSR 16AC12
1	9V Battery	
1	9V Battery Clip	RSR 28011290
1	DIP Extractor	RSR 060400
1	25', 22AWG Black Wire	RSR 27022BK
1	25', 22AWG Red Wire	RSR 27022RD
1	25', 22AWG White Wire	RSR 27022WH
1	Speaker (8 ohm or piezo)	RSR 26SPPC or 26TD02
2	Red LEDs	RSR 08L53RD
2	Yellow LEDs	RSR 08L53YD
2	Green LEDs	RSR 08L53GD
1	7805, 5V Regulator (TO-220 style)	RSR 107805-T
1	1000uF, 16V Electrolytic Capacitor	RSR 14ER0251000U
2	1uF Capacitors	
5	0.1uF Capacitors	RSR 14MN050.1U
2	0.01uF Capacitors	RSR 14MN050.01U
2	1 kohm DIP Resistor Network	RSR 13D1K
2	10 kohm DIP Resistor Network	RSR 13D10K
2	4.7 kohm SIP Resistor Network	RSR 13S10,94.7K
2	10 kohm Trimmer Pots	RSR 18STC10K
2	2.2 kohm Resistors	RSR 130052.2K
9	1N4007 Diode	RSR 111N4007
1	LMC660CN Quad CMOS OP Amp	Digi-Key
1	DS2003CN Darlington Driver	RSR 102003
2	Common Anode 7-Segment LED Display	RSR 08MAN72
1	MM5484 16-Segment LED Driver	Second Source Electronics
1	LM35DZ Temp Sensor	Digi-Key
2	5V DPDT Relay (Omron G4D-287P-BT2 or equivalent)	RSR 22RZ5
8	Momentary Switches	Digi-Key
1	Electret Microphone	RSR 26MPFET

Count	Description	Comments
1	30' 28 Gauge Magnet Wire	
1	Serial Download Cable	
2	Steel blocks (1" x 3")	
1	Set of Datasheets for all of above!	

Note: Parallax - <http://www.parallaxinc.com>, RSR - <http://www.elexp.com>, Digi-Key - <http://www.digikey.com>, Jameco - <http://www.jameco.com>, Second Source Electronics - <http://www.2ndsrc.com>.

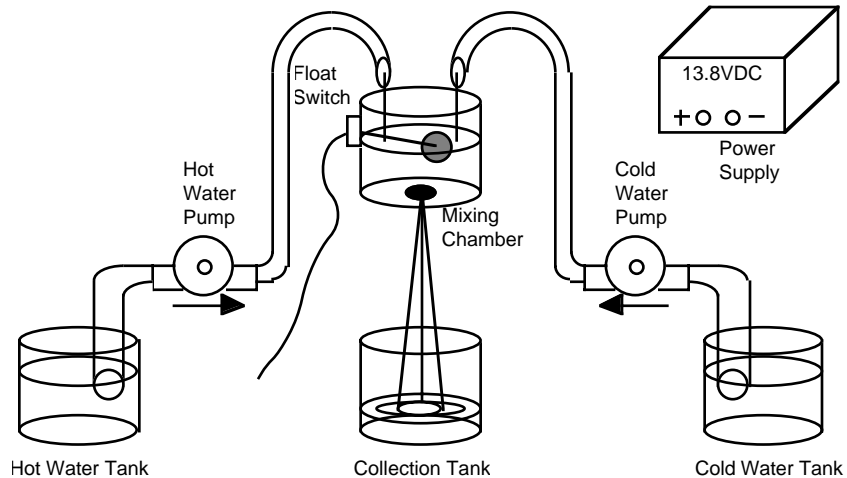
Other Supplies

In addition to the kit, there are some other supplies that will be need to accomplish the projects. However, these can easily be shared among multiple groups. These include some construction materials for building the intersection (a hot glue gun, cardboard, construction paper, etc.) and the Smart Shower setups. I also kept a few extra kits in reserve to replace broken parts.

The Smart Shower

The Smart Shower apparatus is shown below. Hot and Cold water tanks are emptied via two automobile windshield wiper pumps (available at most auto supply stores). These flow into a mixing chamber with a small hole in the bottom. There is also a float switch in the mixing chamber to allow detection of the desired fluid level. Underneath the Mixing Chamber is a Collection Tank which receives the “shower”. A high current automotive type power supply is available which is sufficient to drive both pumps simultaneously. A nice touch is to bring all electrical connections

to a terminal strip for easy wiring. You may also wish to add the LM35D tempera-



ture sensor into the mixing chamber. We found hot glue provided an adequate water tight seal of the electrical connections.

Test Equipment

Some basic test equipment is required to complete the projects. This includes a multimeter and an oscilloscope, which are generally available in undergraduate electronics labs. If it can be afforded, it would be an excellent idea to include an inexpensive DMM in the kit.