Electronics: A Guide for Physicists

Version 0.1 (In Active Development)

Electronics: A Guide for Physicists

Version 0.1 (In Active Development)

Derek J. Thuecks
Washington College

July 27, 2026

Website: Textbook

Website: Project Source Files on GitHub

Contents

Part I

Circuit Analysis

Chapter 1

Preliminaries

Electronic devices surround us. Computers, phones, televisions, watches, cars, building wiring, and many other examples all rely on electricity. All of these applications ultimately rely on energy from electricity, and this electricity must be controlled using circuits comprising wires, resistors, capacitors, and many other components.

This text serves as a survey of electronics principles and design. While you will not learn to build a smartphone in this text, you will instead learn about the fundamental building blocks of electronic circuits. Focus will be placed on how electronic components work, both alone and in tandem, as we learn the rules that govern the behavior of circuit elements. While we will occasionally delve into the physical principles underlying the behavior of circuit elements, such topics are more appropriate for a course in undergraduate electrodynamics. So, while we will not focus much on the physics principles behind electronic components, we will instead focus on the rules governing circuit design and behavior.

1.1 Review from Introductory Physics

Electrical devices are powered by electricity, the movement of charges (which we call **current**) through a circuit. These charges move due to **voltage** differences applied by batteries or power supplies. Thus, if we intend to understand the behavior of electrical circuits, we must first remember what we learned in our introductory physics course.

In the presence of charged objects, the space around the objects is altered. This alteration can be represented as a vector field which we call the **electric field** (which has SI units of V/m). The electric field at a position $\vec{r}$ relative to a point

charge source Q is given by

$$\vec{E} = \frac{Q}{4\pi\varepsilon_0 r^2}\hat{r}$$

where ε_0 is the permitivity of free space (a constant of nature). The electric field has SI units of V/m. While the electric field is a useful construct, it can be convenient to represent this space modification using a scalar field called the **electric potential** (with SI units of Volts). For a point charge Q, the electric potential some distance r away from the source charge is

$$V(\vec{r}) = \frac{Q}{4\pi\varepsilon_0 r}.$$

When discussing circuits, this scalar field is often called the **voltage**. The relationship linking these two representations is described by

$$\vec{E} = -\nabla V = -\left[\frac{\partial V}{\partial x}\hat{x} + \frac{\partial V}{\partial y}\hat{y} + \frac{\partial V}{\partial z}\hat{z}\right].$$

While both representations have their uses, voltage will be much more commonly used and referenced when analyzing the behavior of electrical circuits.

In circuits, voltage differences lead to electric fields that cause charges to move. **Current** quantifies the movement of charge through circuit components. More precisely, current is the amount of charge passing by some point in a circuit per unit time, or

$$I = \frac{dQ}{dt}.$$

In many materials (e.g. metals), positive charges are largely immobile since they are associated with the nuclei of atoms in a material. Thus, electrons are typically the charge carrier in circuits, constituting the current. Unfortunately, there was still much unknown about atomic theory when initial experiments on electricity were conducted, with the consequence that the direction of current is defined to be in the direction that positive charge carriers would be moving if they were responsible for the current. This means that current often has a direction that is opposite to the direction in which the actual charge carriers are moving.

1.2 Circuit Schematic: A Visual Model of a Circuit

Before we can jump into analyzing the behavior of a circuit, we must first identify features of the circuit that are likely to contribute most importantly to the overall circuit behavior to within some desired degree of accuracy. Such a process is an example of model building, a common activity in physics and one that is a bit of an art form. Model building requires practioners to balance model simplicity (a

desirable quality that simplifies circuit analysis) with the model's ability to represent the real behavior of a physical circuit if it were to be built. A model that is too complex, it may become difficult, costly (in time or money), or impossible to use in generating predictions of circuit behavior. On the other hand, a model that is too simple may leave out important qualities, leading to predictions that have little semblance to reality. Finding the proper balance between simplicity and completeness is an activity that practitioners learn over time and through experience.

A **circuit schematic**, like the one shown in Figure 1.2.1 is a visual model of a circuit. Rather than presenting an artistic representation or photograph of a real circuit, a circuit schematic represents the circuit construction and includes enough information about the behavior of circuit components and the connections between them while leaving out information that is unimportant for our particular purposes.

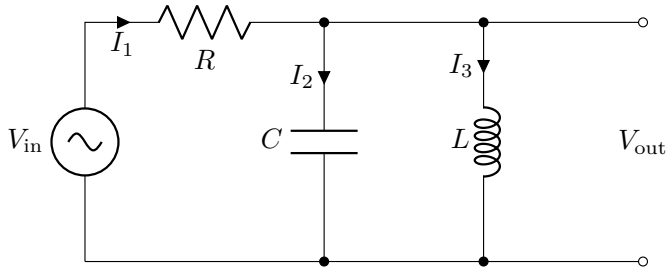
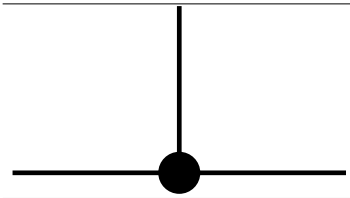
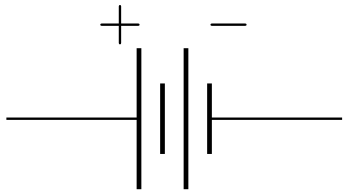
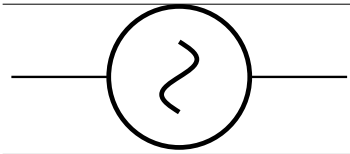
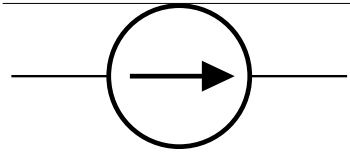
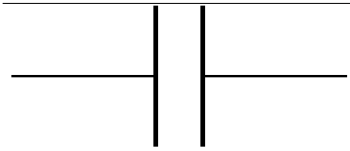
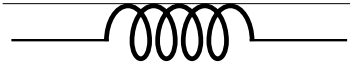
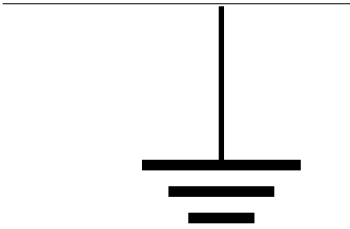
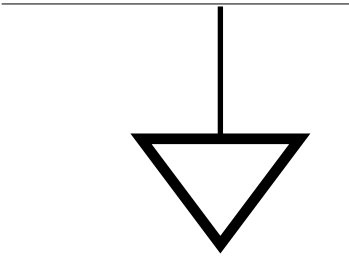


Figure 1.2.1

In Figure 1.2.1, circuit components are represented using standard symbols that are defined in Table 1.2.2. We find an AC voltage source, a resistor, a capacitor, and an inductor, all labeled with variable names. Lines in the circuit schematic represent wire connections between components. If a dot is present where two lines intersect, it indicates a junction where two wires that are connected. Arrows (and associated labels) are used to represent the current present in various parts of the circuit.

Table 1.2.2 Circuit Symbols

Symbol	Name	Notes
	wire	
	junction	
	Battery	SI Unit: Volt (V), Long line is + terminal
	AC voltage source	SI Unit: Volt (V)
	DC current source	SI Unit: Ampere (A)
	Resistor	SI Unit: Ohm (Ω)
	Capacitor	SI Unit: Farad (F)
	Inductor	SI Unit: Henry (H)
	Ground	Physically connected to Earth. Acts as charge source/sink.
	Common	Circuit reference. Assume V=0 at common.

Chapter 2

DC Circuits

In this chapter, we will start our analysis of circuits by examining the behavior of circuits in which voltages and currents are independent of time, and circuits that contain resistors, voltage sources, and/or current sources. We will call these DC (or direct current) circuits.

Define voltage sources (or provide examples): batteries, power supplies Define or provide examples of current sources. Define resistors and ohmic components.

2.1 Kirchhoff's Laws

There are three laws that will make up the foundation of our DC circuit analysis. In truth, these laws will extend beyond the DC circuit analysis presented in this chapter, but we will leave discussion of these extensions to future chapters.

Kirchhoff's Current Law (KCL), or Junction Law, relies on conservation of charge to say that the rate at which charges enter a junction must be equivalent to the rate at which charges leave the junction

$$\sum I_{\text{in}} = \sum I_{\text{out}}.$$

for a scenario as pictured in Figure 2.1.1. The KCL can also be represented mathematically as

$$\sum_i I_i = 0$$

where $I_i < 0$ for currents leaving a node and $I_i > 0$ for currents entering a node.

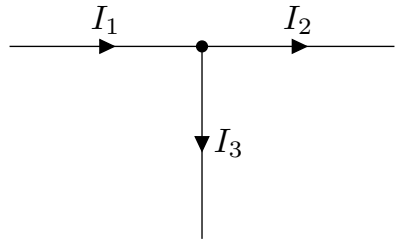


Figure 2.1.1 A circuit junction.

Another conservation law, the conservation of energy, is ultimately responsible for what we call Kirchhoff's Voltage Law (KVL), or the Loop Law. If one walks around any loop in a circuit (such as that shown in Figure 2.1.2), the sum of the voltage changes as one moves in a single direction around a complete loop must add up to zero, or

$$\sum_i \Delta V_i = 0 \text{ around a closed loop.}$$

When using the KVL equation, voltages rise as one crosses a voltage source from the negative terminal to positive terminal and voltages fall in the opposite direction. Additionally, voltages fall across a resistor in the direction of current and rise in the opposite direction.

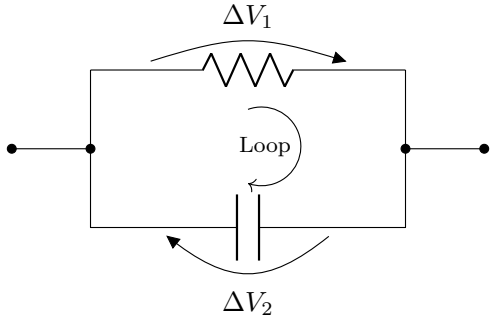


Figure 2.1.2 A circuit loop.

The last rule we'll introduce here is Ohm's Law. For the situation shown in Figure 2.1.3, the magnitude of the voltage change across a resistor is dependent on the resistance and the magnitude of current in the resistor according to

$$\Delta V = IR$$

where $\Delta V = V_A - V_B$. A note on signs: When using Ohm's law, ΔV and I actually represent $|\Delta V|$ and $|I|$. It is conventional to use Ohm's law to relate magnitudes of ΔV to I. The signs on each variable are linked by the fact that current is directed

from high to low voltage. So, if current is in the direction indicated in Figure 2.1.3, then $V_A > V_B$.

Figure 2.1.3 Ohm's Law circuit.

EXAMINE THE VOLTAGE DIVIDER CIRCUIT.

2.2 Voltage Source Symbols

In the following section, we will analyze our first circuit. Before we analyze this circuit, let's first take a moment to discuss a notation convention. It is popular convention to label batteries or voltage supplies with the symbol V with an identifying subscript as we have with the symbol V_b in the next section. Using this convention, one might hypothetically say $V_b = 9\text{V}$ even though what is really meant is that $\Delta V = 9$ Volts across the battery terminals. Though this notation may require a small bit of additional care of the reader, we will use this popular convention of labeling batteries and voltage sources with a V rather than ΔV. When discussing voltage changes across any other type of circuit element, we will return to the ΔV notation.

2.3 Series and Parallel Resistors

Let's begin by applying Kirchhoff's laws and Ohm's law to the circuit shown in Figure 2.3.1.

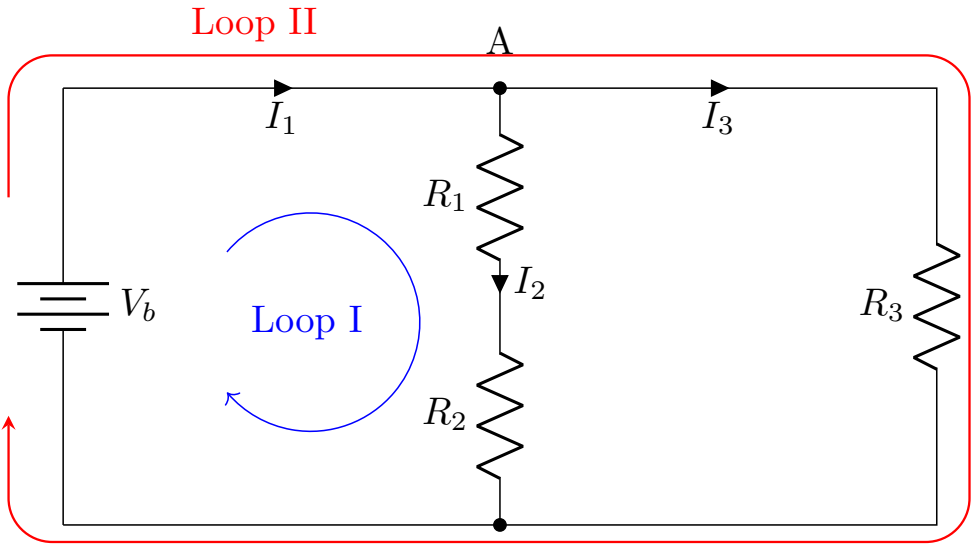


Figure 2.3.1 Circuit example.

At node A, we apply the KCL to find

$$\sum_i I_i = 0 \Rightarrow I_1 - I_2 - I_3 = 0. \tag{2.3.1}$$

The KVL can then be applied to Loop I and Loop II. Examining Loop I first, we will start in the lower left corner and traverse the loop in the direction indicated by the arrow. Traversing this loop has us pass through battery V_b from the negative to positive terminal, resulting in $\Delta V = +V_b$. The next component that we encounter is R_1, which we travel across in the direction of current, so Ohm's law tells us that the voltage falls by $\Delta V = I_2 R_1$ in the direction of current. Likewise, the voltage falls by $\Delta V = I_2 R_2$ as we travel across R_2 in the direction of current on the way back to our starting point. The KVL for Loop 1 will thus give

$$\sum_{i \text{ in Loop I}} (\Delta V)_i = 0 \Rightarrow V_b - I_2 R_1 - I_2 R_2 = 0 \tag{2.3.2}$$

where there are negative signs in front of $I_2 R_1$ and $I_2 R_2$ because voltage falls by these amount in the direction that we've traversed the loop. A similar application of the KVL to Loop II gives

$$\sum_{i \text{ in Loop II}} (\Delta V)_i = 0 \Rightarrow V_b - I_3 R_3 = 0. \tag{2.3.3}$$

Often, voltages provided by batteries (or a DC voltage source) and the resistances are known as these were the physical components assembled into the circuit.

Thus, (2.3.1) - (2.3.3) represent a series of equations that must be solved to determine currents I_1, I_2, and I_3 in Figure 2.3.1.

Rather than solving this system of equations for the three currents, we will instead perform some intermediate analysis to investigate the behavior of resistor combinations. Starting with (2.3.2), we can see that

$$V_b - I_2 (R_1 + R_2) = V_b - I_2 R_{1+2}.$$

Resistors R_1 and R_2 are considered to be in **series**. Resistors are in series if the current in one resistor must be the same current in the other series resistors. In this case, R_1 and R_2 are in series because I_2 must be the current through each of the resistors since there is no junction between them. We see that the voltage drop across the pair of series resistors is just $I_2 R_{1+2}$ where $R_{1+2} = R_1 + R_2$, meaning that resistors in series act together as a single resistor that has a resistance equal to the sum of the individual series resistances. This relationship holds for any number of resistors that are in series such that

$$R_{\text{eq}} = \sum_{\text{series}} R_i \quad \text{(Resistors in series).} \tag{2.3.4}$$

This means that currents I_1, I_2, and I_3 will have identical values in all three circuits displayed in Figure 2.3.2

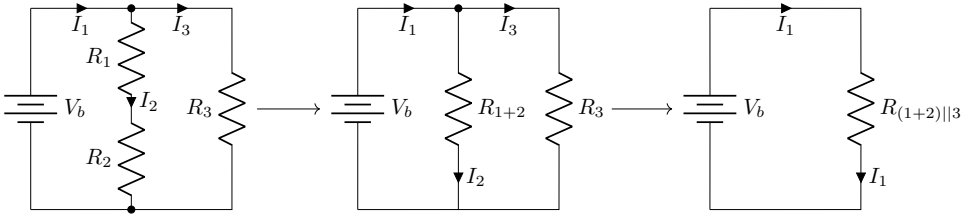


Figure 2.3.2 Original circuit (left) is simplified into an intermediate circuit (center) by combining series resistors R_1 and R_2 into effective resistance R_{1+2}. The circuit is further simplified (right) by combining parallel resistors R_{1+2} and R_3 into effective resistance $R_{(1+2)||3}$.

If we now examine the simplified circuit in Figure 2.3.2b, we can see that both R_{1+2} and R_3 both have the same voltage drop across them

$$I_2 R_{1+2} = V_b$$
$$I_3 R_3 = V_b$$

so that

$$I_2 = \frac{V_b}{R_{1+2}}$$

$$I_3 = \frac{V_b}{R_3}.$$

These expressions can be inserted into (2.3.1) to find

$$\begin{aligned} I_1 &= I_2 + I_3 \Rightarrow \\ &= \frac{V_b}{R_{1+2}} + \frac{V_1}{R_3} \\ &= V_b \left(\frac{1}{R_{1+2}} + \frac{1}{R_3} \right) \end{aligned}$$

so that

$$V_b = I_1 \left(\frac{1}{R_{1+2}} + \frac{1}{R_3} \right)^{-1}.$$

This has the form of Ohm's Law where R_{1+2} and R_3 combine into an equivalent resistance

$$R^{-1}_{(1+2)||3} = R^{-1}_{1+2} + R^{-1}_3.$$

We say that resistors R_{1+2} and R_3 are in parallel because the potential difference across each resistor must be identical as each side of the resistors are connected together by wires. An equivalent resistance for a combination of any number of parallel resistors can be found using

$$R^{-1}_{eq} = \sum_{\text{parallel}} R_i^{-1} \quad \text{(Resistors in parallel).} \tag{2.3.5}$$

2.4 Circuit Analysis Methods

In the following sections, we will analyze the DC circuit found in Figure 2.4.1 using a variety of different methods. Each of these methods can be useful in different circumstances, so these exercises are meant to introduce you to a variety of tools that will remain in your circuit-analysis toolbox for future use.

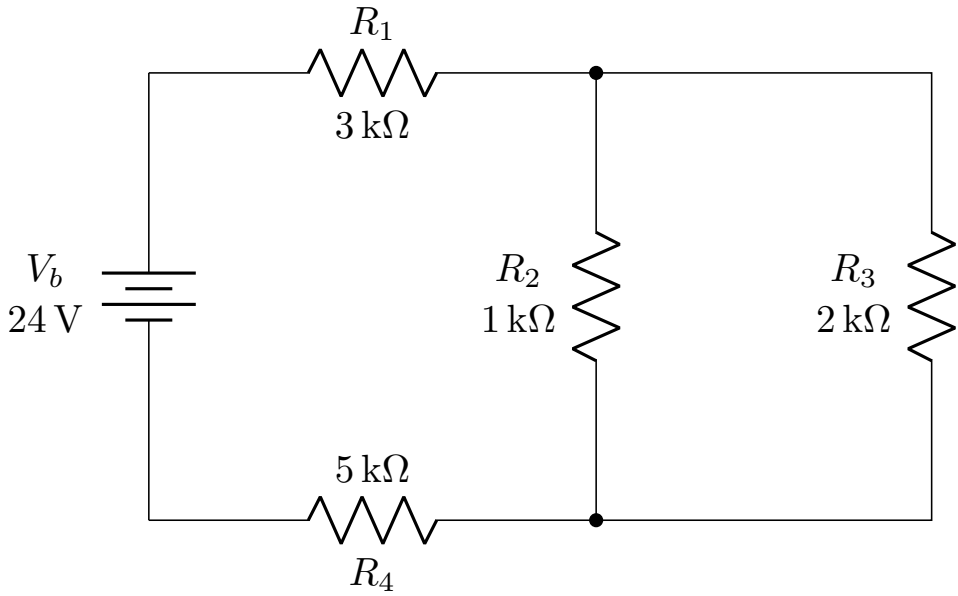


Figure 2.4.1 The circuit that will be analyzed using multiple different methods over the course of the following sections.

2.4.1 Analysis: Circuit Reduction

In the method of circuit reduction, we use the rules of series and parallel resistors to find all currents present in the circuit. We will start by drawing the circuit and labeling all currents as in Figure 2.4.2a.

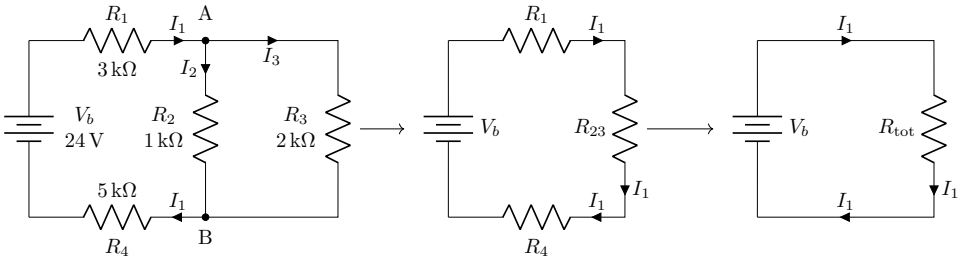


Figure 2.4.2 Circuit reduction for Figure 2.4.1.

In our first simplification, we note that R_2 and R_3 are in parallel, so

$$R_{23} = \left(\frac{1}{R_2} + \frac{1}{R_3}\right)^{-1} = 666.67\Omega.$$

With that simplification done, we now find that R_1, R_{23}, and R_4 are in series, so

$$R_{\text{tot}} = R_1 + R_{23} + R_4 = 8666.67\Omega.$$

Ohm's law can be used to find I_1

$$V_b = I_1 R_{\text{tot}} \rightarrow I_1 = \frac{(\Delta V)_{\text{tot}}}{R_{\text{tot}}} = 2.77 \text{ mA}$$

where $(\Delta V)_{\text{tot}}$ is the voltage change across R_{tot}. We must also check the direction of I_1. When examining a circuit, you must sometimes guess at the directions of currents. If you guess correctly and the guessed current direction flows from high to low potential, then your value for current will be positive. If you guessed incorrectly at the outset, you will instead find that the value for current will be negative. In this situation, we note that the I_1 direction pictured does indeed pass through R_{tot} from high to low potential given how R_{tot} terminals are connected directly to battery terminals. Thus, the value that we calculate for I_1 should be positive. So, $I_1 = 2.77$ mA.

Now, in order to find currents I_2 and I_3, we'll have to find the voltage changes across R_2 and R_3. We can do this by taking our calculated value for I_1 and using Ohm's Law in Figure 2.4.2(left). Thus, we find that

$$\begin{aligned}\Delta V_1 &= V_{b+} - V_A \\ \Delta V_4 &= V_B - V_{b-}.\end{aligned}$$

In other words,

$$\begin{aligned}V_A &= V_{b+} - \Delta V_1 \\ V_B &= V_{b-} + \Delta V_4\end{aligned}$$

so that

$$\begin{aligned}\Delta V_2 = \Delta V_3 &= V_A - V_B \\ &= (V_{b+} - \Delta V_1) - (V_{b-} + \Delta V_4) \\ &= (V_{b+} - V_{b-}) - \Delta V_1 - \Delta V_4 \\ &= \Delta V_b - \Delta V_1 - \Delta V_4 \\ &= V_b - I_1 R_1 - I_1 R_4 = 1.85\text{V}.\end{aligned}$$

Once we have $\Delta V_2 = \Delta V_3$, we can use Ohm's law to find I_2 and I_3

$$\begin{aligned}I_2 &= \frac{\Delta V_2}{R_2} = 1.85\text{mA} \\ I_3 &= \frac{\Delta V_3}{R_3} = 0.925\text{mA}.\end{aligned}$$

If we define $V_{b-} = 0$, then

$$\begin{aligned} V_{b+} &= 24\text{V} \\ V_A &= V_{b+} - I_1 R_1 = 15.69\text{V} \\ V_B &= V_{b-} + I_1 R_4 = 13.85\text{V}. \end{aligned}$$

2.4.2 Analysis: Branch Method

In the **branch method**, we perform the following steps:

1. Draw the circuit.
2. Define a current for each circuit segment that may have a different current. Use your intuition to try to guess the correct directions for the currents. Don't worry about guessing wrong. It is just important to clearly identify your assumed direction for each current.
3. Write out KCL equations at nodes. Remove duplicate or redundant equations. Note that the number of independent KCL equations is one fewer than the number of junctions present in the circuit.
4. Write down KVL equations for enough loops to include all circuit components at least once.
5. Solve the system of KCL and KVL equations for all unknown currents. If you were wrong in your initial guesses for any current directions, those currents will just come out with negative values.
6. Use Ohm's law to find voltage changes through each resistor and use those changes to find the voltage at any point of interest in the circuit.

To analyze the circuit in Figure 2.4.1 using the mesh method, we draw the circuit and define currents as in Figure 2.4.3.

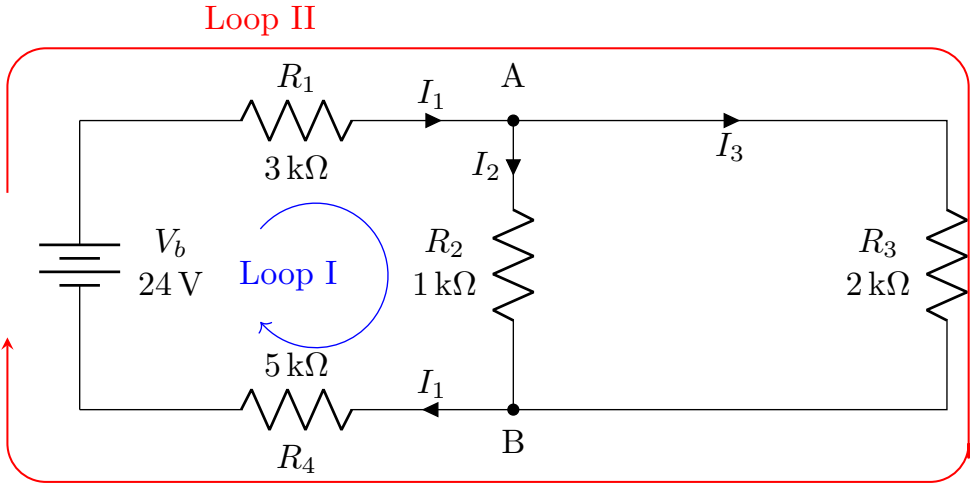


Figure 2.4.3 Branch method circuit analysis.

Examining the KCL at junction A and the KVL for loops I and II, we find the following equations

$$0 = I_1 - I_2 - I_3 \tag{2.4.1}$$
$$0 = V_b - I_1R_1 - I_2R_2 - I_1R_4 \tag{2.4.2}$$
$$0 = V_b - I_1R_1 - I_3R_3 - I_1R_4 \tag{2.4.3}$$

You can now find values for I_1, I_2, and I_3 using your favorite method for solving systems of equations. We will illustrate several such methods.

Example 2.4.4 Solving Branch Method Equations - Analytic. Solve (2.4.1)-(2.4.3) for the three unknown currents using analytic methods.

Solution. Solving Equation (2.4.1) for I_1,

$$I_1 = I_2 + I_3,$$

and inserting into Equations (2.4.2) and (2.4.3), we get

$$0 = V_b - (I_2 + I_3)\, R_1 - I_2R_2 - (I_2 + I_3)\, R_4$$
$$0 = V_b - (I_2 + I_3)\, R_1 - I_3R_3 - (I_2 + I_3)\, R_4$$

which can be rearranged into

$$0 = V_b - I_2(R_1 + R_2 + R_4) - I_3(R_1 + R_4)$$
$$0 = V_b - I_2(R_1 + R_4) - I_3(R_1 + R_3 + R_4).$$

Solving the top equation for I_2, we find

$$I_2 = \frac{V_b}{R_1 + R_2 + R_4} - \frac{R_1 + R_4}{R_1 + R_2 + R_4} I_3$$

which can be substituted into the bottom equation to give

$$0 = V_b - \frac{V_b\,(R_1 + R_4)}{R_1 + R_2 + R_4} - \frac{(R_1 + R_4)^2}{R_1 + R_2 + R_4} I_3 - I_3(R_1 + R_3 + R_4).$$

Plugging in values for known components, we can solve for I_3. Once we have a value for I_3, we use the equations above to calculate I_2 and then I_1, finding

$$\begin{aligned} I_1 &= 2.77\text{mA} \\ I_2 &= 1.85\text{mA} \\ I_3 &= 0.92\text{mA}. \end{aligned}$$

Then, just as in the circuit reduction method, Ohm's law can be used to find

$$\begin{aligned} V_A &= 15.69\text{V} \\ V_B &= 13.85\text{V}. \end{aligned}$$

▲

Example 2.4.5 Solving Branch Method Equations - Python. (2.4.1)(2.4.3)
Solution. The Python programming language can be used to solve a system of equations through the use of the scipy.linalg package. If presented with an equation $\mathbf{Ax} = \mathbf{v}$ where $\mathbf{A}$ is a matrix and $\mathbf{x}$ and $\mathbf{v}$ are vectors, we can solve the equation using `x = scipy.linalg.solve(A,v)`.

As an example, let's start with the system of equations (2.4.1)-(2.4.3)

$$\begin{aligned} 0 &= I_1 - I_2 - I_3 \\ 0 &= V_b - I_1 R_1 - I_2 R_2 - I_1 R_4 \\ 0 &= V_b - I_1 R_1 - I_3 R_3 - I_1 R_4. \end{aligned}$$

These equations can be re-expressed as a matrix equation if we remember the rules of matrix multiplication (see Appendix B for details). The resulting matrix equation is expressed as

$$\begin{pmatrix} 1 & -1 & -1 \\ R_1 + R_4 & R_2 & 0 \\ R_1 + R_4 & 0 & R_3 \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \\ I_3 \end{pmatrix} = \begin{pmatrix} 0 \\ V_b \\ V_b \end{pmatrix}.$$

Since we have numerical values for V_b and all resistors, we can plug these into the matrices. This matrix equation is in the necessary form for `x = scipy.linalg.solve(A,v)`.

```
import numpy as np
import scipy.linalg as linalg

Vbb=24 # Volts
R1, R2, R3, R4 = [3000, 1000, 2000, 5000] # Ohms
A = np.array([[1, -1, -1],[R1+R4, R2, 0],[R1+R4, 0, R3]])
v = np.array([0, Vbb, Vbb])

x = linalg.solve(A,v)
VA = Vbb - x[0]*R1;
VB = VA - x[1]*R2;
print('I1 = {0:0.2f} mA, I2 = {1:0.2f} mA, I3 = {2:0.2f} mA'.format(x[0]*1000,x[1]*1000, x[2]*1000)) # 1000 mA in one Amp
print('VA = {0:0.2f} V, VB = {1:0.2f} V'.format(VA, VB))
```

```
I1 = 2.77 mA, I2 = 1.85 mA, I3 = 0.92 mA
VA = 15.69 V, VB = 13.85 V
```

The results of this method are identical to the analytic result. ▲

Example 2.4.6 Solving Branch Method Equations - Sympy. (2.4.1)(2.4.3)

Solution. One can also use functionality from SymPy, the symbolic math package for Python, to solve systems of linear equations. As before,

$$\begin{aligned} 0 &= I_1 - I_2 - I_3 \\ 0 &= V_b - I_1 R_1 - I_2 R_2 - I_1 R_4 \\ 0 &= V_b - I_1 R_1 - I_3 R_3 - I_1 R_4 \end{aligned}$$

becomes

$$\begin{pmatrix} 1 & -1 & -1 \\ R_1+R_4 & R_2 & 0 \\ R_1+R_4 & 0 & R_3 \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \\ I_3 \end{pmatrix} = \begin{pmatrix} 0 \\ V_b \\ V_b \end{pmatrix} \quad \Rightarrow \quad \begin{pmatrix} I_1 \\ I_2 \\ I_3 \end{pmatrix} = \begin{pmatrix} 1 & -1 & -1 \\ R_1+R_4 & R_2 & 0 \\ R_1+R_4 & 0 & R_3 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ V_b \\ V_b \end{pmatrix}$$

which can be solved symbolically with the following code:

```
# Using Sage since Python/SymPy output as LaTeX is not playing
    nice
import sympy as sym
from IPython.display import display
sym.init_printing(use_latex=True)

R1, R2, R3, R4 = sym.symbols('R_1 R_2 R_3 R_4')
Vb = sym.symbols('V_b')
matrix = sym.Matrix([[1, -1, -1],
                     [R1+R4, R2, 0],
                    [R1+R4, 0, R3]])
invmatrix = matrix.inv()
I = invmatrix * sym.Matrix([[0],[Vb],[Vb]])
sym.simplify(I)
```

This analysis gives the result

$$\begin{bmatrix} \frac{V_b(R_2+R_3)}{R_1R_2+R_1R_3+R_2R_3+R_2R_4+R_3R_4} \\ \frac{V_bR_3}{R_1R_2+R_1R_3+R_2R_3+R_2R_4+R_3R_4} \\ \frac{V_bR_2}{R_1R_2+R_1R_3+R_2R_3+R_2R_4+R_3R_4} \end{bmatrix}$$

which simplifies to the results of the previous methods when numerical values are inserted for V_b and resistances. ▲

Example 2.4.7 Solving Branch Method Equations - Gaussian elimination. Solve (2.4.1)-(2.4.3) for the three unknown currents using Gaussian elimination.

Solution. We can also use a linear algebra method called **Gaussian elimination** to solve the matrix equation

$$\begin{pmatrix} 1 & -1 & -1 \\ R_1+R_4 & R_2 & 0 \\ R_1+R_4 & 0 & R_3 \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \\ I_3 \end{pmatrix} = \begin{pmatrix} 0 \\ V_b \\ V_b \end{pmatrix}.$$

Gaussian elimination relies on two facts:

1. If we multiply one row of $\mathbf{A}$ and the same row of $\mathbf{v}$ by some constant q, the solutions for $\mathbf{x}$ remain unchanged.

2. Any linear combination of two rows will give a new correct row.

Using the two rules above, our goal is to find an equivalent $\mathbf{A}'\mathbf{x} = \mathbf{v}'$ such that

$$\mathbf{A}' = \begin{pmatrix} 1 & A'_{12} & A'_{13} \\ 0 & 1 & A'_{13} \\ 0 & 0 & 1 \end{pmatrix}.$$

Often, the first step in using Gaussian elimination is to divide the first row by A_{11}. In this case, both $\mathbf{A}$ and $\mathbf{v}$ remain unchanged since $A_{11} = 1$. Next, we'll subtract $[A_{21} * (\text{first row of } \mathbf{A} \text{ and } \mathbf{v})]$ from row 2 to get

$$\begin{pmatrix} 1 & -1 & -1 \\ 0 & R_2 + R_1 + R_4 & R_1 + R_4 \\ R_1 + R_4 & 0 & R_3 \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \\ I_3 \end{pmatrix} = \begin{pmatrix} 0 \\ V_b \\ V_b \end{pmatrix}.$$

Then, we'll subtract $[A_{31} * (\text{first row of } \mathbf{A} \text{ and } \mathbf{v})]$ from row 3 to get

$$\begin{pmatrix} 1 & -1 & -1 \\ 0 & R_2 + R_1 + R_4 & R_1 + R_4 \\ 0 & R_1 + R_4 & R_3 + R_1 + R_4 \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \\ I_3 \end{pmatrix} = \begin{pmatrix} 0 \\ V_b \\ V_b \end{pmatrix}.$$

If we now divide row 2 by A_{22}, we get

$$\begin{pmatrix} 1 & -1 & -1 \\ 0 & 1 & \frac{R_1+R_4}{R_2+R_1+R_4} \\ 0 & R_1 + R_4 & R_3 + R_1 + R_4 \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \\ I_3 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{V_b}{R_2+R_1+R_4} \\ V_b \end{pmatrix}$$

Subtracting $[A_{32} * (\text{second row of } \mathbf{A} \text{ and } \mathbf{v})]$ from row 3 will eliminate A_{32} while leaving $A_{31} = 0$

$$\begin{pmatrix} 1 & -1 & -1 \\ 0 & 1 & \frac{R_1+R_4}{R_2+R_1+R_4} \\ 0 & 0 & R_3 + R_1 + R_4 - \frac{(R_1+R_4)^2}{R_2+R_1+R_4} \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \\ I_3 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{V_b}{R_2+R_1+R_4} \\ V_b - \frac{V_b(R_1+R_4)}{R_2+R_1+R_4} \end{pmatrix}.$$

Then, divide row 3 by A_{33} to get

$$\begin{pmatrix} 1 & -1 & -1 \\ 0 & 1 & \frac{R_1+R_4}{R_2+R_1+R_4} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \\ I_3 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{V_b}{R_2+R_1+R_4} \\ \left[V_b - \frac{V_b(R_1+R_4)}{R_2+R_1+R_4}\right] / \left[R_3 + R_1 + R_4 - \frac{(R_1+R_4)^2}{R_2+R_1+R_4}\right] \end{pmatrix}$$

which simplifies to

$$\begin{pmatrix} 1 & -1 & -1 \\ 0 & 1 & \frac{R_1+R_4}{R_2+R_1+R_4} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \\ I_3 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{V_b}{R_2+R_1+R_4} \\ \frac{V_b R_2}{R_2+R_1+R_4} \frac{R_3+R_1+R_4}{R_3^2+2R_3(R_1+R_4)} \end{pmatrix}.$$

Now, we can read off an expression for I_3 and subsequently calculate I_1 and I_2

$$\begin{aligned} I_3 &= \frac{V_b R_2}{R_2 + R_1 + R_4} \frac{R_3 + R_1 + R_4}{R_3^2 + 2R_3 (R_1 + R_4)} = 0.92\text{mA} \\ I_2 &= \frac{V_b}{R_2 + R_1 + R_4} - \frac{R_1 + R_4}{R_2 + R_1 + R_4} I_3 = 1.85\text{mA} \\ I_1 &= I_2 + I_3 = 2.77\text{mA}. \end{aligned}$$

▲

2.4.3 Analysis: Mesh Method

In the **mesh method**, we perform the following steps:

1. Draw the circuit.
2. Define **mesh currents**. Each current passes around a complete loop. Every component must be in at least one loop. Typically, it is best to use the simplest loops possible. The total current passing through any point in the circuit is the sum of the mesh currents through that point. Choose and clearly indicate each current loop's direction; the direction chosen doesn't really matter as long as it is clearly defined.
3. Write down a KVL equation for each loop. If there are multiple mesh currents passing through a resistor, the voltage change is the sum of voltage changes for each mesh current through that component. The signs for each voltage drop contribution result in voltage drops for the mesh currents aligned with the direction we go around the loop for the loop law, while they result in voltage rises for the mesh currents aligned against the direction we go around the loop when creating the loop law equation.
4. Solve the system of KVL equations for all unknown mesh currents.
5. The current through any point in the circuit is the sum of mesh currents through that point. Note that two mesh currents through a point on the circuit will add together with opposite signs.
6. Use Ohm's law to find voltage changes through each resistor and use those changes to find the voltage at any point of interest in the circuit.

One of the benefits of the mesh method over the branch method is that there are fewer equations to solve in the resulting system of equations (since there is no KCL equation), though this comes at the cost of having to add multiple mesh currents together to find the total current through any point in the circuit.

The analysis begins with drawing the circuit and defining our mesh currents as demonstrated in Figure 2.4.8.

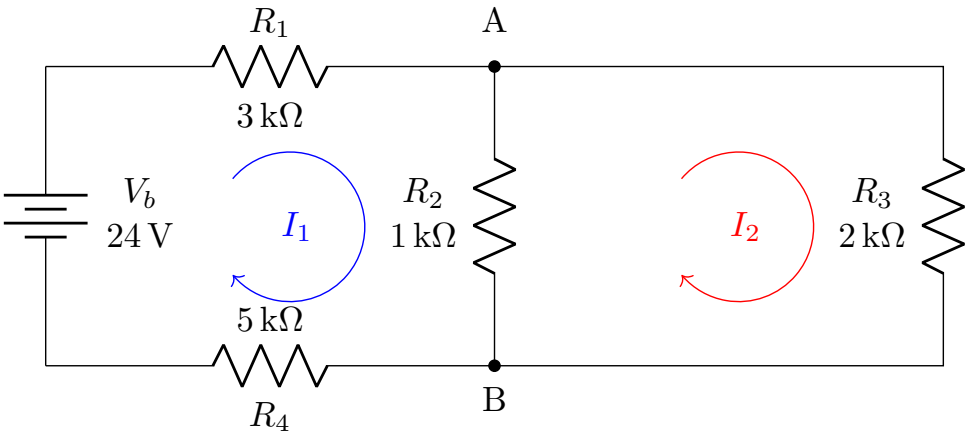


Figure 2.4.8 Mesh method circuit analysis.

Then, the KVL equations associated with mesh currents I_1 and I_2 are

$$0 = V_b - I_1R_1 - (I_1 - I_2)R_2 - I_1R_4$$
$$0 = (I_1 - I_2)R_2 - I_2R_3$$

which can be arranged into the form

$$0 = V_b - I_1(R_1 + R_2 + R_4) + I_2R_2$$
$$0 = I_1R_2 - I_2(R_2 + R_3).$$

Once again, your favorite method for solving systems of equations can be used to find the mesh currents $I_1 = 2.77$mA and $I_2 = 0.923$mA. Therefore, the current passing through R_4, V_b, and R_1 is $I_1 = 2.77$mA clockwise, the current through R_3 is $I_2 = 0.923$mA downward, and the current through R_2 is $I_1 - I_2 = 1.85$mA downward. With these current values, we can once again use the same Ohm's law method from the previous sections to find the voltages at any location in the circuit.

Note that the mesh method leaves us with a system of two equations to solve and compare that to the system of three equations that the branch method provided. This simplification comes at a cost, though. In the branch method, the variables representing currents are the actual currents passing through each part of the circuit. In the mesh method, we instead solve for mesh currents. The real currents can then be calculated as sums and differences of these mesh currents.

2.5 Superposition Theorem

The superposition theorem isn't often used on its own in circuit analysis, but is the foundation for some other very important rules and theorems. Before stating the

superposition theorem, we must first establish a few definitions.

First, **linear circuit elements** are circuit elements for which the relationship between voltage and current is a linear function. Examples of linear circuit elements include resistors, capacitors, and inductors. A **linear circuit** contains only linear elements.

With these definitions in place, the superposition theorem states that for any linear circuit containing more than one independent source, the circuit can be solved for one source while turning the others `off'. Then, this can be repeated for each source and add the results together. When we talk about turning a voltage source off, we mean that we replace it in our analysis with a wire. Turning a current source off means replacing the current source with a break in the circuit. The superposition theorem in electronics arises directly from use of the superposition principle in electricity and magnetism.

2.6 Thévenin's and Norton's Theorems

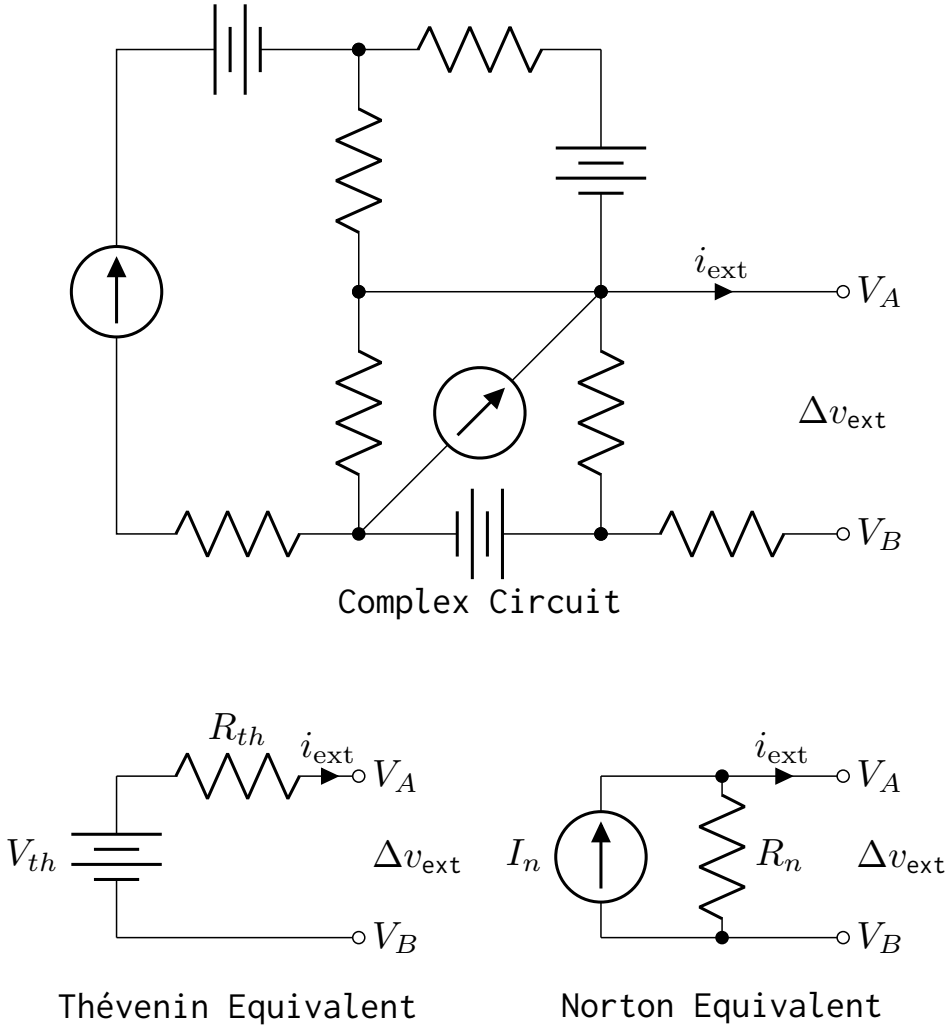


Figure 2.6.1

Circuits controlling devices that you use every day are often very complex with complicated scematic diagrams. This complexity can make it difficult to analytically determine the behavior of the output voltage and current when this complex circuit is connected to an external load. DEFINE LOAD. NOTE SOMEWHERE THAT $\Delta v_{\text{ext}} = V_A - V_B$.

In this section, we introduce two theorems that allow us to replace a complicated circuit with a simplified circuit. For the moment, we will limit ourselves to resistor-based DC circuits and will generalize these theorems next chapter.

Thévenin's theorem states that any complicated circuit containing comprised of only ideal voltage sources, ideal current sources, and resistors will demonstrate a relationship between i_{ext} and Δv_{ext} that is identical to the relationship demonstrated by a Thévenin equivalent circuit containing a single voltage source (the Thévenin voltage V_{th}) in series with a single resistance (the Thévenin resistance R_{th}) for any load (DEFINE!) connected across the circuit output. Similarly, **Norton's theorem** states that any complicated circuit containing comprised of only ideal voltage sources, ideal current sources, and resistors will demonstrate a relationship between i_{ext} and Δv_{ext} that is identical to the relationship demonstrated by a Norton equivalent circuit containing a single current source (the Norton current I_n) in parallel with a single resistance (the Norton resistance R_n) for any load (DEFINE!) connected across the circuit output.

Below, we will start by analyzing the Thévenin and Norton equivalent circuits and determine the relationship between i_{ext} and v_{ext}. We will then show that any complicated circuit containing only ideal sources and resistors must demonstrate this same behavior, thus proving these two theorems.

2.6.1 Behavior of Thévenin/Norton circuits

Applying Kirchhoff's voltage law to the Thévenin equivalent circuit in Figure 2.6.1, we find that

$$V_{\text{th}} - i_{\text{ext}} R_{\text{th}} - \Delta v_{\text{ext}}.$$

Rearrangement of this equation shows

$$i_{\text{ext}} = -\frac{1}{R_{\text{th}}} \Delta v_{\text{ext}} + \frac{1}{R_{\text{th}}} V_{\text{th}}, \tag{2.6.1}$$

a linear relationship between i_{ext} and v_{ext}.

Application of Kirchhoff's current law to the Norton equivalent circuit in Figure 2.6.1 shows

$$I_n - i_{\text{ext}} - \frac{1}{R_n} \Delta v_{\text{ext}}$$

where Ohm's law was used to express the current through R_n in terms of Δv_{ext}. Rearrangement of this equation shows that

$$i_{\text{ext}} = -\frac{1}{R_n} \Delta v_{\text{ext}} + I_n \tag{2.6.2}$$

which again demonstrates a linear relationship between i_{ext} and Δv_{ext}.

If both Thévenin's theorem and Norton's theorem are to be true, then Equations (2.6.1) and (2.6.2) must be equivalent. This can only be true if

$$I_n = \frac{V_{\text{th}}}{R_{\text{th}}}$$

and

$$R_n = R_{\text{th}}.$$

The linear relationship between i_{ext} and Δv_{ext} for these circuits is shown in Figure 2.6.2. Here we see that the slope is the negative inverse of R_{th} and R_n. The Thévenin voltage V_{th} is equal to Δv_{ext} when $i_{\text{ext}} = 0$ (an open-circuit where nothing is connected across the output terminals), and the Norton current I_n equals i_{ext} when the outputs are shorted by a wire (making $\Delta v_{\text{ext}} = 0$).

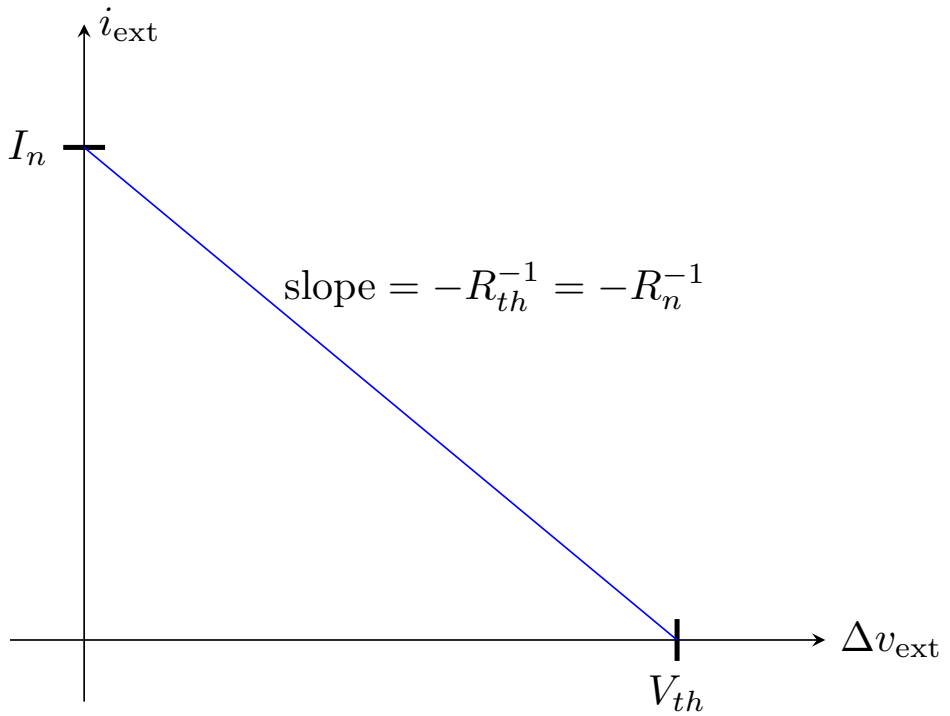


Figure 2.6.2 I-V plot for linear circuits. The norton current I_n is equal to i_{ext} when the output is shorted ($v_{\text{ext}} = 0$). We call this the short-circuit current I_{sc}. The Thevenin voltage V_{th} is given by v_{ext} when $i_{\text{ext}} = 0$. This occurs when there is no load placed across the output terminals, so we call it the open-circuit voltage V_{oc}.

In the next section, we show that Figure 2.6.2 must represent the behavior of any circuit comprised of ideal voltage sources, ideal current sources, and resistors.

2.6.2 Proof of Thevenin's and Norton's Theorems

The branch method of circuit analysis is the basis for our proof. Let's assume that inside of a box we have a circuit with two leads exiting the box (Figure 2.6.3). The circuit in the box is comprised of a complicated network of resistors R, ℓ DC voltage sources V, and k ideal current sources I. We will also assume j unknown currents i that are dependent on the size of the voltage difference Δv_{ext} across the output terminals of the circuit.

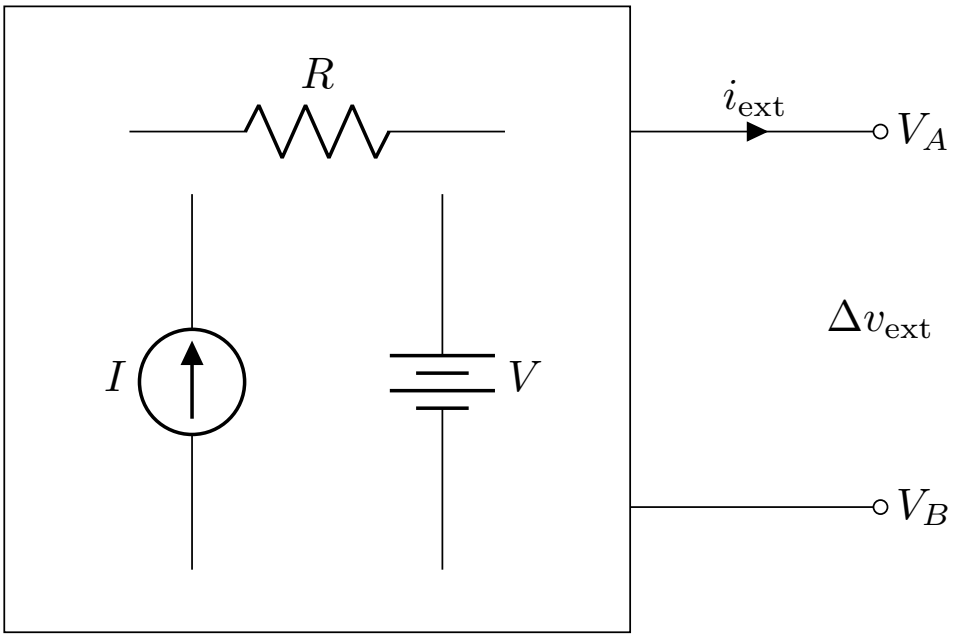


Figure 2.6.3

Assume that our analysis results in N_1 independent KCL equations. Each KCL equation can be expressed as

$$a_{r,\text{ext}} i_{\text{ext}} + \sum_{n=1}^{j} a_{rn} i_n + \sum_{n=1}^{k} b_{rn} I_n = 0 \tag{2.6.3}$$

where a_{rn} and b_{rn} have values of +1, -1, or 0 depending on whether the currents i_n and I_n are entering, exiting, or not present at a junction and r is an index counting over KCL equations.

Analysis will also provide N_2 independent KVL equations, analyzing loops that are absent known current sources I_n. This restriction will not cost us anything because this proof relies on knowledge of the currents through each circuit segment. Each KVL equation has the form

$$m_{r,\text{ext}}\Delta v_{\text{ext}} + \sum_{n=1}^{\ell} m_{rn}V_n + c_{r,\text{ext}}i_{\text{ext}} + \sum_{n=1}^{j} c_{rn}i_n = 0. \tag{2.6.4}$$

The variables m_{rn} have values of either +1, or -1 depending on the battery orientation relative to the direction traveled around the loop, or a value of 0 for batteries not present in a given loop. Variable c_{r_n} has units of resistance because each nonzero term in these sums represents a voltage change described by Ohm's Law. Here, the index r counts over KVL equations.

These N_1 instances of (2.6.3) and N_2 instances of (2.6.4) can be re-expressed as a matrix equation $\mathbf{Ax} = \mathbf{v}$. First, let's rearrange these equations so that all unknowns are on the left side of the equation and all of the source terms are on the right

$$\begin{aligned} a_{r,\text{ext}}i_{\text{ext}} + \sum_{n=1}^{j} a_{rn}i_n &= -\sum_{n=1}^{k} b_{rn}I_n \\ c_{r,\text{ext}}i_{\text{ext}} + \sum_{n=1}^{j} c_{rn}i_n &= -m_{r,\text{ext}}\Delta v_{\text{ext}} - \sum_{n=1}^{\ell} m_{rn}V_n. \end{aligned}$$

We can express these equations in the form $\mathbf{Ax} = \mathbf{v}$ if we let

$$\mathbf{A} = \begin{pmatrix} a_{1,\text{ext}} & a_{11} & \cdots & a_{1j} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N_1,\text{ext}} & a_{N_1 1} & \cdots & a_{N_1 j} \\ c_{1,\text{ext}} & c_{11} & \cdots & c_{1j} \\ \vdots & \vdots & \ddots & \vdots \\ c_{N_2,\text{ext}} & c_{N_2 1} & \cdots & c_{N_2 j} \end{pmatrix},$$

$$\mathbf{x} = \begin{pmatrix} i_{\text{ext}} \\ i_1 \\ \vdots \\ i_j \end{pmatrix} \quad , \quad \mathbf{v} = \begin{pmatrix} -\sum_{n=1}^{k} b_{1n} I_n \\ \vdots \\ -\sum_{n=1}^{k} b_{N_1 n} I_n \\ -m_{1,\text{ext}} \Delta v_{\text{ext}} - \sum_{n=1}^{\ell} m_{1n} V_n \\ \vdots \\ -m_{N_2,\text{ext}} \Delta v_{\text{ext}} - \sum_{n=1}^{\ell} m_{N_2 n} V_n \end{pmatrix}.$$

Now, every element in matrix $\mathbf{A}$ is either a unitless number or a function of resistances only. More to the point, there is no voltage or current dependence in $\mathbf{A}$. If we multiply both sides of our matrix equation by $\mathbf{A}^{-1}$, we get

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{v}.$$

The solution that is most important to us is i_{ext} as we wish to examine the relationship between output current and voltage. If we actually perform the calculations necessary to invert $\mathbf{A}$, use matrix multiplication to find an expression for i_{ext}, and then group source terms and rename constants multiplying each source ($v_{\text{ext}}, V,\ I$), the resulting solution is of the form

$$i_{\text{ext}} = \alpha_0 v_{\text{ext}} + \sum_{n=1}^{\ell} \alpha_n V_n + \sum_{n=1}^{k} \beta_n I_n$$

where all α_n and β_n terms are functions of resistances only and are thus constants. This result demonstrates that i_{ext} is a linear function of v_{ext} with slope α_0 and a y-intercept that equals the two summations on the right side of the equation. So, we have now shown that any circuit comprised of solely linear circuit elements is itself a linear circuit.

At this point, it is important to discuss that the internal behavior of the complicated circuit and simplified circuit are different. EXPAND ON THIS THOUGHT.

2.6.3 Finding Thevenin Equivalent Circuits

Given a complicated circuit, we can use the procedure provided in the proof above to find Thevenin and Norton equivalent circuit parameters.

Example 2.6.4 Thevenin Example - Matrix Approach. Coming soon.

Here, we develop (2.6.3) and (2.6.4), convert them into a matrix equation, and use Python to solve for the Thevenin and Norton equivalent circuit parameters.

Solution. Coming soon ▲

If one wishes to find Thevenin and Norton equivalent parameters analytically, it is often easier to follow this procedure:

1. Use circuit analysis to find the open-circuit voltage V_{oc} which is equivalent to $V_{oc} = V_A - V_B$ when nothing is connected externally between output terminals A and B. Then, $V_{th} = V_{oc}$.
2. Find $R_{th} = R_n$ which is equivalent to R_{AB} when all internal source voltages and currents are `turned off'.
   - Turning a source `off' means replacing voltage sources with wires and replacing current sources with breaks.
   - This method works as a consequence of application of the superposition theorem to the complicated circuit.
3. Use your values for V_{th} and R_{th} to find $I_n = V_{th}/R_{th}$

Let's look at an example.

Example 2.6.5 Thevenin Example - No Linear Algebra. Find V_{th}, R_{th}, and I_n for the circuit in Figure 2.6.6.

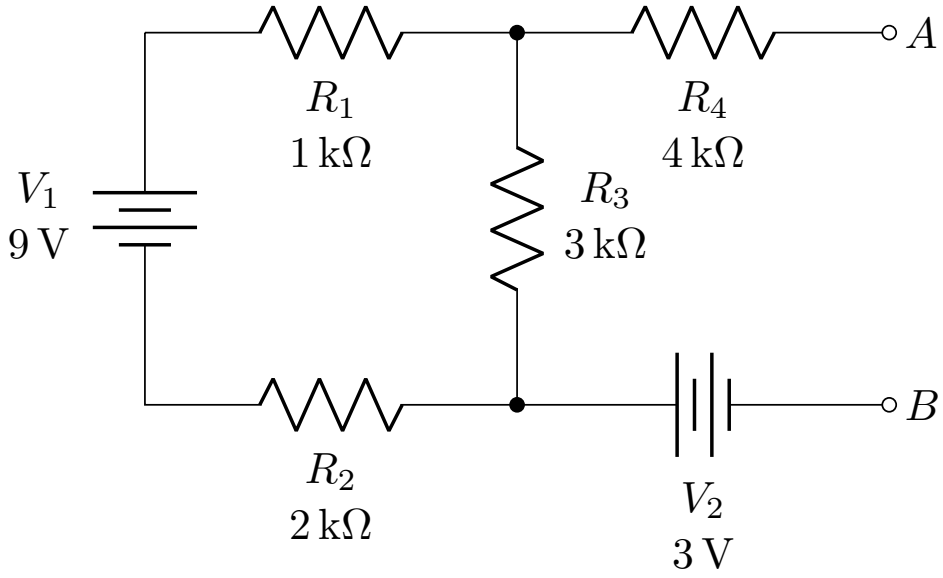


Figure 2.6.6

Answer.

$$V_{th} = 7.5\text{V}$$

$$I_n = 1.36\text{mA}$$
$$R_{th} = 5.5\text{k}\Omega$$

Solution.

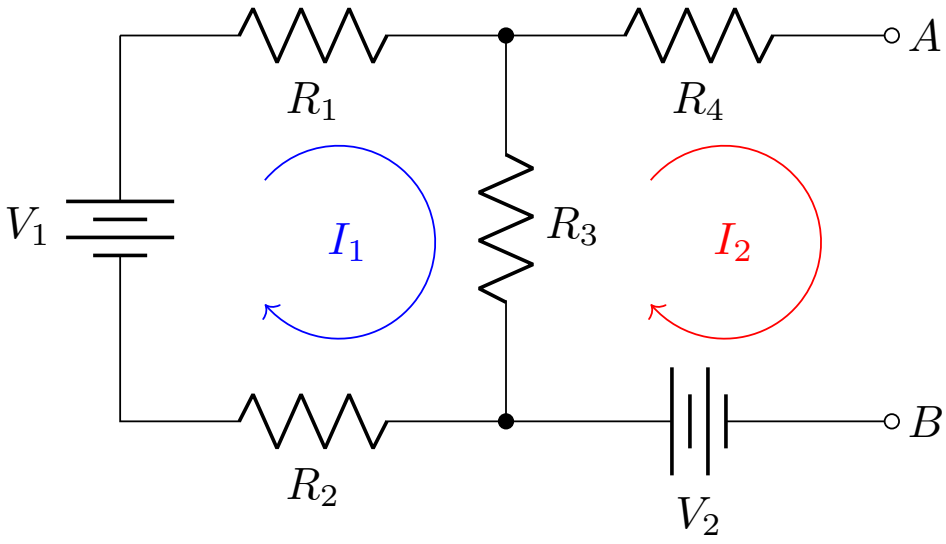


Figure 2.6.7

Find V_{th} : Assume an open circuit between terminals A and B such that $I_2 = 0$. Using mesh analysis, we find

$$0 = V_1 - I_1R_1 - I_1R_3 - I_1R_2$$

or

$$V_1 - I_1(R_1 + R_2 + R_3) = 0.$$

Then

$$I_1 = \frac{V_1}{R_1 + R_2 + R_3}$$
$$= 1.5\text{mA}.$$

Thus, the voltage difference across R_2 is $I_1R_3 = 4.5\text{V}$. Since $I_2 = 0$, there is no voltage change across R_4, but there is an additional voltage change across V_2 so that the open-circuit voltage between the output terminals is

$$V_{th} = V_{oc} = 7.5\text{V}.$$

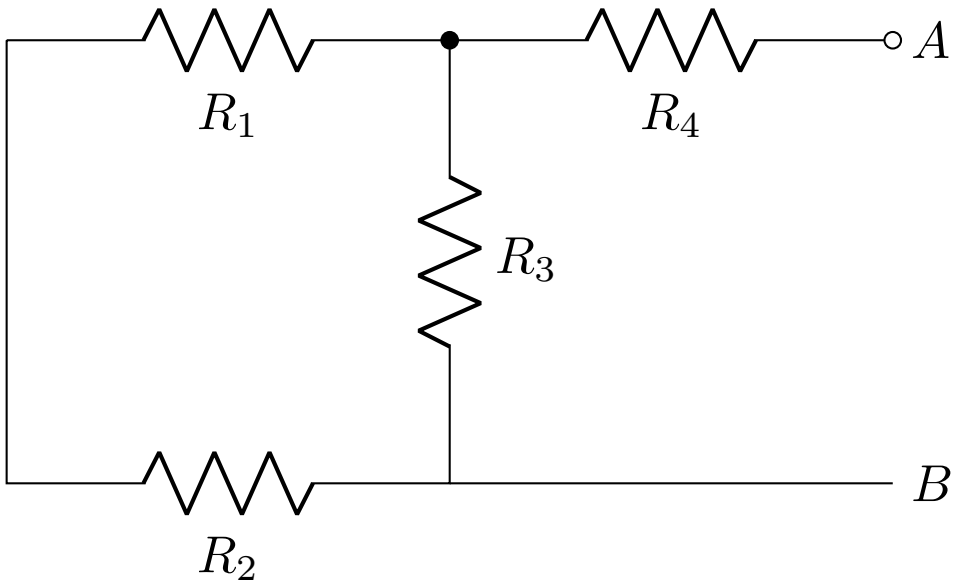


Figure 2.6.8

Find R_{th}: If we turn off all sources in our circuit, we are left to find the resistance between terminals A and B in Figure 2.6.8. Looking for series and parallel resistances, we find that

$$R_{th} = ((R_1 + R_2) \,||R_3) + R_4.$$

Plugging in resistance values, we find that

$$\begin{aligned} R_{th} &= R_4 + \left((R_1 + R_2)^{-1} + R_3^{-1}\right)^{-1} \\ &= 5.5\text{k}\Omega. \end{aligned}$$

Find I_n: The Norton current $I_n = V_{th}/R_{th} = 1.36\text{mA}$. ▲

Sometimes, one does not have a circuit schematic but only has the physical circuit. In this case, a digital multimeter can be used to physically perform the following procedure:

1. Find the open-circuit voltage V_{oc} which is equivalent to $V_{oc} = V_A - V_B$ when nothing is connected externally between output terminals A and B. Then, $V_{th} = V_{oc}$.

2. Find the short-circuit current I_{sc} which is the current from terminal A to terminal B when an external wire connects the two outputs. Then $I_n = I_{sc}$.

3. The Thevenin resistance (which is equivalent to the Norton resistance) is then $R_{th} = R_n = V_{th}/I_n$.

Look at section 41.6 to add color to example blocks. also section 28.2 and similar. xsl stylesheets

2.7 Application of Thevenin Equivalent Cicruits: Batteries

Batteries have a complicated voltage and resistance internally, as do power supplies, function generators, and other voltage sources that may be present in a circuit. All can be represented by a Thevenin equivalent circuit.

The Thevenin equivalent representation of a battery can help us understand the behavior of batteries when connected to resistive loads as pictured in Figure 2.7.1. Here, the chemical reactions inside the battery maintain an internal voltage difference V_{source}. There is an internal resistance R_{internal} that means that the actual battery terminals that we have access to provide a voltage difference of $V_{\text{battery}} \neq V_{\text{source}}$.

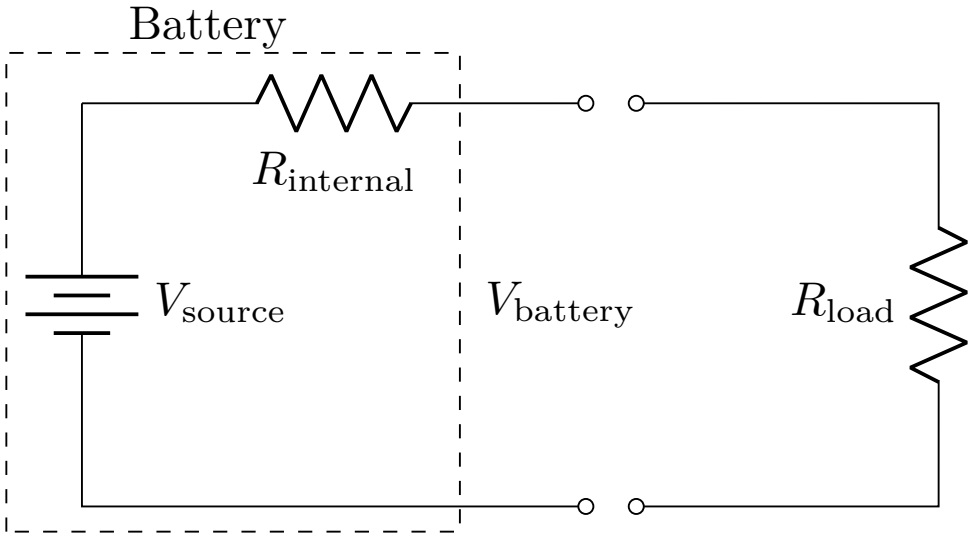


Figure 2.7.1

If we connect a resistive load with resistance R_{load}, we find that we have a voltage divider circuit where

$$V_{\text{battery}} = \frac{R_{\text{load}}}{R_{\text{internal}} + R_{\text{load}}} V_{\text{source}}.$$

So, if R_{load} is very large ($R_{\text{load}} \gg R_{\text{internal}}$), we call this a "light" resistive load because $V_{\text{battery}} \approx V_{\text{source}}$. Conversely, if R_{load} is very small ($R_{\text{load}} \ll R_{\text{internal}}$), we call this a "heavy" resistive load because $V_{\text{battery}} \ll V_{\text{source}}$.

The analysis above shows why the voltage between battery terminals can 'droop' lower than their rated voltage when a heavy resistive load is connected to it. We will discuss this effect in greater detail in Section 3.9.

2.8 DC Circuit Uses

Chapter 3

AC Circuits

When studying DC circuits, resistors and sources were the main components considered. For AC circuits, analysis becomes more interesting and there are additional components to understand. In nature, innumerable measureable quantities have time dependence, and electronics will be used to measure/collect such data. Thus, an understanding of AC signals and AC circuit behavior is critical for a practicing scientist.

3.1 Sinusoidal Signals

In this chapter, we will focus on circuit behavior that results from sinusoidal signals

$$\begin{aligned} V(t) =& V_0 \cos(\omega t + \delta_V) \\ I(t) =& I_0 \cos(\omega t + \delta_I) \end{aligned}$$

where $\omega = 2\pi f = 2\pi/T$ is the angular frequency, f is the frequency, T is the period, and δ_V, δ_I are the phase of the signals at time $t = 0$. (We could have used the sine function instead of cossine, but these are the same function but shifted from each other by $\pm\pi/2$ radians in their arguments.)

While it may seem like focusing on a sinusoidal signal with a single frequency will limit the applicability of the analysis we develop in this chapter, this is not actually the case. Fourier's Theorem states that any well-behaved function can be represented by an integral (sum) of sinusoidal signals.

$$I(t) = \int_{-\infty}^{\infty} a(\omega) \cos(\omega t + \delta_I(\omega)) \, \mathrm{d}\omega$$

where $a(\omega)$ and $\delta_I(\omega)$ are weights and phase shifts for each ω value. We can use the trigonometric identity

$$\cos(\omega t + \delta_I) = \cos(\omega t)\cos(\delta_I(\omega)) + \sin(\omega t)\sin(\delta_I(\omega))$$

to express this in a more common form

$$I(t) = \int_{-\infty}^{\infty} [\alpha(\omega)\cos(\omega t) + \beta(\omega)\sin(\omega t)]\,\mathrm{d}\omega$$

where $\alpha(\omega) = a(\omega)\cos(\delta_I(\omega))$ and $\beta(\omega) = a(\omega)\sin(\delta_I(\omega))$.

Since any general (well-behaved) function of time can be represented as a sum of sinusoidal signals, we can use the superposition theorem to find the response of a linear circuit to a general time signal. If we examine the behavior of a linear circuit in response to a single sinusoidal signal with frequency ω, we can then use superposition to add up (integrate) responses for many sinusoidal signals (each with its own ω), weighting the results in the sum using the same weighting factors that appear above.

Thus, it is sufficient to examine the response of linear circuits to a single sinusoidal signal. INCLUDE FURTHER DISCUSSION OF SINE/COSINE SERIES

3.2 Characterizing Signal Size

With AC signals, there are multiple ways to represent amplitude. Here, we will examine methods of characterizing voltage amplitudes, but all methods discussed apply equally well to any other sinusoidally-varying quantity.

1. Peak voltage V_p corresponds to the traditional definition of amplitude for a sinusoidal waveform where $V_p = V_{\max}$.

2. Peak-to-Peak voltage $V_{pp} = V_{\max} - V_{\min} = 2V_p$.

3. Root-mean-square (RMS) amplitude is calculated by examining a single cycle of data ($V(t_i)$) containing n measurements where t_i ($i = 1..(n-1)$) are the discrete values of time for which measurements were recorded. Then,

$$V_{\mathrm{rms}} = \sqrt{\frac{\sum_{i=1}^{n} V_i^2}{n}}$$

which is $V_p/\sqrt{2}$ for sinusoidal waves.

4. The average power P_{avg} dissipated in a resistor R due to a sinusoidal signal is

$$P_{\text{avg}} = I_{\text{rms}}^2 R = \frac{V_{\text{rms}}^2}{R}.$$

5. Power in a signal P_2 (with amplitude A_2) relative to the power in some other signal P_1 (with amplitude A_1) can be represented using a logarithmic scale called decibels, where

$$\begin{aligned} P_{\text{dB}} &= 10 \log_{10} \left(\frac{P_2}{P_1} \right) \\ &= 20 \log_{10} \left(\frac{A_2}{A_1} \right). \end{aligned}$$

3.3 Complex Signals

Signal analysis with sinusoidal functions can be challenging due to the need for trigonometric identities. The mathematical manipulations in circuit analysis can more easily be done if signals are instead represented as complex exponentials.

Using Euler's formula

$$e^{\pm i\theta} = \cos(\theta) \pm i \sin(\theta)$$

we can see that

$$V_0 \cos(\omega t + \delta')$$

is the real part of

$$V_0 e^{i(\omega t + \delta')}.$$

Using this notation, it is easy to see that the phase information can be absorbed into the amplitude, resulting in a complex amplitude:

$$\tilde{V}(t) = V_0 e^{i\omega t + i\delta'} = V_0 e^{i\delta'} e^{i\omega t} = \tilde{V}_0 e^{i\omega t}.$$

With this notation, the measured signal $V(t)$ is the real part of the complex signal $\tilde{V}(t)$. So, we can use $\tilde{V}(t)$ in equations describing circuit behavior, and the real part of the result represents what we would measure for the parameter of interest.

3.4 Complex Impedance

When analyzing AC circuits with complex signals $\tilde{V}(t)$, we will need to expand beyond the concept of resistance. We define **impedance** Z as

$$\begin{aligned} Z &\equiv \frac{\tilde{V}}{\tilde{I}} \\ &\equiv R + iX \end{aligned}$$

where R is our normal resistance and X is **reactance**. The impedance is the factor that relates $\tilde{I}$ to $\tilde{V}$ as seen in Figure 3.4.1. Generally, Z is frequency-dependent.

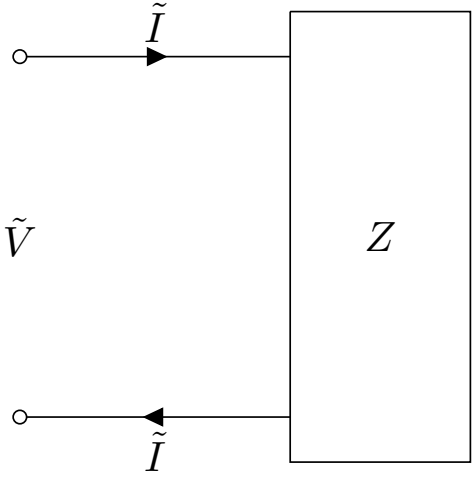


Figure 3.4.1

Let's look at a few examples. Ohm's law continues to hold for AC signals, so it is trivial to demonstrate that $Z_R = R$.

Now, let's work to determine the impedance for a capacitor. In introductory physics, you likely learned that capacitance C is a quantity that links the voltage difference V between two objects and the charge Q on each object through the expression

$$Q = CV.$$

Pairing this expression with the definition of current

$$I = \frac{\mathrm{d}Q}{\mathrm{d}t}$$

we find that

$$I = C\frac{\mathrm{d}V}{\mathrm{d}t}.$$

For a voltage $\tilde{V}(t) = \tilde{V}_0 e^{i\omega t}$, we thus see that the resulting current through the capacitor is $\tilde{I}(t) = i\omega C\tilde{V}_0 e^{i\omega t}$. Using these expressions, we find that

$$Z_C = \frac{\tilde{V}}{\tilde{I}} = \frac{\tilde{V}_0 e^{i\omega t}}{i\omega C\tilde{V}_0 e^{i\omega t}}$$

so that

$$Z_C = \frac{1}{i\omega C} = -\frac{i}{\omega C}. \tag{3.4.1}$$

Note that the impedance Z_C is frequency-dependent and complex. Examining this expression in limiting cases, $Z_C \to \infty$ (as $\omega \to 0$) and $Z_C \to 0$ (as $\omega \to \infty$).

Now, let's work to determine the impedance for an inductor. Inductors, fundamentally, are coils and act in a way as an anti-capacitor. The fundamental equation governing inductor behavior is

$$V = L\frac{\mathrm{d}I}{\mathrm{d}t}$$

which means that the voltage drops across an inductor when $\mathrm{d}I/\mathrm{d}t \neq 0$. If we assume a current

$$\tilde{I}(t) = \tilde{I}_0 e^{i\omega t}$$

then

$$\tilde{V}(t) = i\omega L\tilde{I}_0 e^{i\omega t}$$

so that

$$Z_L = \frac{\tilde{V}}{\tilde{I}} = \frac{i\omega L\tilde{I}_0 e^{i\omega t}}{\tilde{I}_0 e^{i\omega t}}.$$

Simplifying, this leaves us with

$$Z_L = i\omega L.$$

Note that the impedance Z_L is frequency-dependent and complex. Examining this expression in limiting cases, $Z_L \to 0$ (as $\omega \to 0$) and $Z_L \to \infty$ (as $\omega \to \infty$).

The results above can be used with a generalized Ohm's law

$$I = \frac{V}{Z}$$

and also with equations for series and parallel impedances

$$Z_{\text{series}} = Z_1 + Z_2 + \cdots$$
$$\frac{1}{Z_{\text{parallel}}} = \frac{1}{Z_1} + \frac{1}{Z_2} + \cdots$$

3.5 Capacitor Charging and Discharging Behaviors

Let's start examining circuits containing resistors, capacitors, and inductors. We'll start with the circuit in Figure 3.5.1. We want to evaluate the voltage across the capacitor (and resistor) as a function of time if the switch starts closed and opens suddenly at time $t = 0$. Assume that the internal battery resistance is much smaller than the circuit resistor so that $r_e \ll R$ and the voltage difference $V_C \approx V_0$ across the capacitor at $t = 0$.

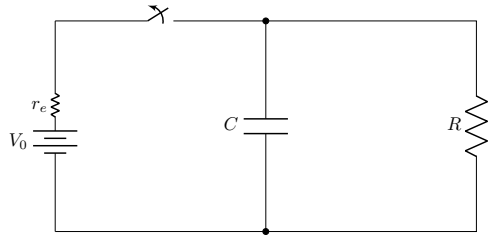


Figure 3.5.1 A circuit for examining a discharging capacitor.

We can use KVL to examine the circuit. Since the switch is opened at $t = 0$, we only have a single KVL equation

$$V_C - IR = 0$$

where V_C is the voltage difference between the capacitor plates. For a capacitor, $Q = CV_c$, so

$$\frac{Q}{C} - IR = 0.$$

Taking the time derivative of the above equation will allow us to eliminate Q dependence in favor of I, so let's revisit the relationship between Q and I. Since the current I causes positive charges to leave the positively-charged capacitor plate and return to the negatively-charged plate, the current will cause a reduction in Q as time advances, meaning $I = -dQ/dt$. With this information, we find that

$$-\frac{I}{C} - R\frac{dI}{dt} = 0$$

or, after rearranging:

$$\frac{dI}{dt} = -\frac{1}{RC}I(t).$$

By inspection, the solution to this differential equation is

$$I(t) = I_0 e^{-t/RC}$$

where $I_0 = I(t = 0)$. Now, this current can be plugged back into our original equation $V_C - IR = 0$ to find

$$V_C(t) = I_0 R e^{-t/RC} = V_0 e^{-t/\tau}$$

where $V_C(t = 0) = V_0 = I_0 R$ at the instant that the switch opens and we've defined something called the **RC time constant** $\tau = RC$ which is a characteristic response time associated with this RC circuit. We plot this discharging behavior in Figure 3.5.2.

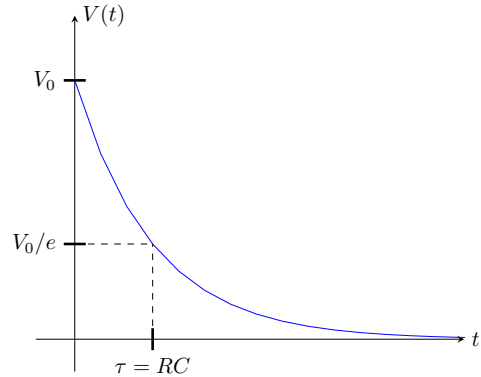


Figure 3.5.2 Discharging capacitor.

Similarly, we can examine the behavior of a charging capacitor by examing the behavior of the circuit in Figure 3.5.3. Here, we do not worry about the battery's internal resistance r_e as $r_e \ll R$ typically, meaning that the charging behavior will be determined by $R + r_e \approx R$.

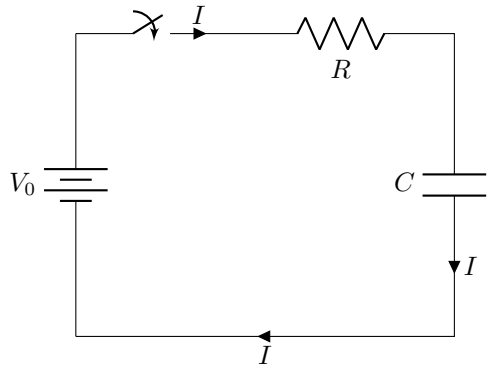


Figure 3.5.3 A circuit for examining a charging capacitor.

Applying the loop law, we find that

$$V_0 - IR - V_C = 0$$

where V_C is the voltage difference across the capacitor. Remembering that

$$I = C\frac{\mathrm{d}V}{\mathrm{d}t}$$

we find that

$$V_0 - RC\frac{\mathrm{d}V}{\mathrm{d}t} - V_C = 0$$

or

$$V_0 - V_C = RC\frac{\mathrm{d}V}{\mathrm{d}t}.$$

Since V_0 is a constant, we can write this as

$$V_0 - V_C = RC\frac{\mathrm{d}}{\mathrm{d}t}\left(V_C - V_0\right).$$

If we define $V' = V_C - V_0$, then this becomes

$$\frac{\mathrm{d}V'}{\mathrm{d}t} = -\frac{1}{RC}V'.$$

This has an exponential solution of the form

$$V' = V_C - V_0 = Ae^{-t/RC}.$$

The initial condition $V_c(t = 0) = 0$ requires that $A = -V_0$. Thus, our result can be rearranged into

$$V_C = V_0\left(1 - e^{-t/RC}\right).$$

This shows that the capacitor charges from 0% to 63% in $\Delta t = RC$.

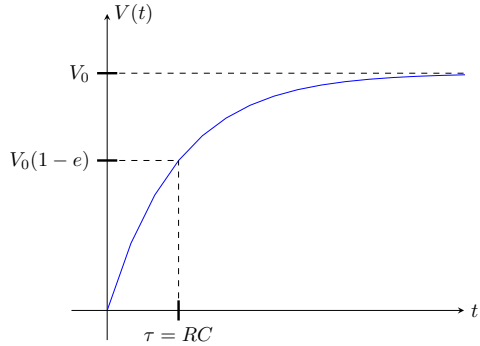


Figure 3.5.4 Charging capacitor.

When charging a capacitor, energy is being stored in the electric field

$$U_C = \frac{1}{2}CV^2 = \frac{Q^2}{2C} = \frac{1}{2}QV.$$

Likewise, inductors can store energy. As the current through an inductor increases, energy is stored i nthe magnetic field

$$U_L = \frac{1}{2}LI^2.$$

3.6 RC High-pass filter

Let's look at some circuit behavior for sinusoidal signals. Figure 3.6.1 shows the circuit that we will examine.

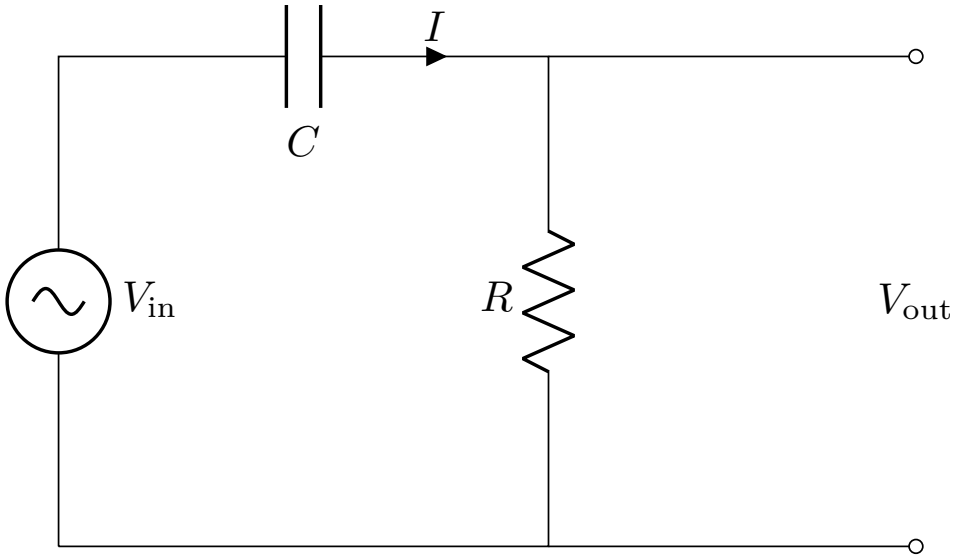


Figure 3.6.1 High-pass RC Filter Circuit.

There is nothing in the formulation of Kirchhoff's laws that would limit them to only DC circuits. In fact, they apply equally well to all circuits (DC, AC, non-linear). Assume that all quantities with a tilde over them are complex quantities expressed as exponentials. So,

$$\tilde{V}_{\text{in}}(t) = \tilde{V}_{\text{in}_0} e^{i\omega t} = V_0 e^{i\delta} e^{i\omega t}$$

where the real signal $V_{\text{in}}(t) = \Re\left[\tilde{V}_{\text{in}}(t)\right]$. Using Kirchhoff's Voltage Law, we find that

$$\tilde{V}_{\text{in}} - \tilde{I}Z_C - \tilde{I}Z_R = \tilde{V}_{\text{in}} - \tilde{I}\left(Z_C - Z_R\right) = 0$$

so that

$$\tilde{I} = \frac{\tilde{V}_{\text{in}}}{Z_R + Z_C} = \frac{\tilde{V}_{\text{in}}}{R - \frac{i}{\omega C}}.$$

We can then find $\tilde{V}_{\text{out}}$ using Ohm's Law

$$\tilde{V}_{\text{out}} = \tilde{I}R = \frac{R}{R - \frac{i}{\omega C}}\tilde{V}_{\text{in}}.$$

Expanding our voltage signals,

$$\tilde{V}_{\text{out}_0}e^{i\omega t} = \frac{R}{R - \frac{i}{\omega C}}\tilde{V}_{\text{in}_0}e^{i\omega t}$$

so that

$$\tilde{V}_{\text{out}_0} = \frac{R}{R - \frac{i}{\omega C}}\tilde{V}_{\text{in}_0}.$$

Then,

$$\frac{\tilde{V}_{\text{out}_0}}{\tilde{V}_{\text{in}_0}} = \frac{R}{R - \frac{i}{\omega C}} = \frac{R}{R - \frac{i}{\omega C}}\frac{\left(R + \frac{i}{\omega C}\right)/R^2}{\left(R + \frac{i}{\omega C}\right)/R^2} = \frac{1 + \frac{i}{\omega RC}}{1 + \left(\frac{1}{\omega RC}\right)^2}.$$

Expanding the complex amplitude,

$$\frac{\tilde{V}_{\text{out}_0}}{\tilde{V}_{\text{in}_0}} = \frac{V_{\text{out}_0}e^{i\delta'}}{V_{\text{in}_0}e^{i\delta}} = \frac{V_{\text{out}_0}}{V_{\text{in}_0}}e^{i\phi}$$

where $\phi = \delta' - \delta$. We can calculate the voltage gain $G_V = V_{\text{out}_0}/V_{\text{in}_0}$:

$$\begin{aligned} G_V &= \frac{V_{\text{out}_0}}{V_{\text{in}_0}} \\ &= \left|\frac{V_{\text{out}_0}}{V_{\text{in}_0}}e^{i\phi}\right| = \left|\frac{1 + \frac{i}{\omega RC}}{1 + \left(\frac{1}{\omega RC}\right)^2}\right| \end{aligned}$$

so

$$G_V = \frac{1}{\sqrt{1 + \left(\frac{1}{\omega RC}\right)^2}} = \frac{1}{\sqrt{1 + \left(f_0/f\right)^2}} \tag{3.6.1}$$

where $f_0 = 1/2\pi RC$ and $f = \omega/2\pi$. We can also determine the phase difference ϕ between the output and input voltages. Since

$$\frac{V_{\text{out}_0}}{V_{\text{in}_0}}e^{i\phi} = \frac{1 + \frac{i}{\omega RC}}{1 + \left(\frac{1}{\omega RC}\right)^2},$$

we can rearrange to find

$$e^{i\phi} = \frac{V_{\text{in}_0}}{V_{\text{out}_0}} \frac{1 + \frac{i}{\omega RC}}{1 + \left(\frac{1}{\omega RC}\right)^2}.$$

Now,

$$\tan(\phi) = \frac{\sin(\phi)}{\cos(\phi)} = \frac{\Im\left(e^{i\phi}\right)}{\Re\left(e^{i\phi}\right)} = \frac{1}{\omega RC}$$

so that

$$\phi = \tan^{-1}\left(\frac{1}{\omega RC}\right) = \tan^{-1}\left(f_0/f\right). \tag{3.6.2}$$

Using Eqs. (3.6.1) and (3.6.2), we can show the following behaviors:

1. $0° < \phi < 90°$, meaning $\tilde{V}_{\text{out}}$ leads $\tilde{V}_{\text{in}}$.
2. $V_{\text{out}_0} \to V_{\text{in}_0}$ as $f \to \infty$.
3. $V_{\text{out}_0} \to 0$ as $f \to 0$.

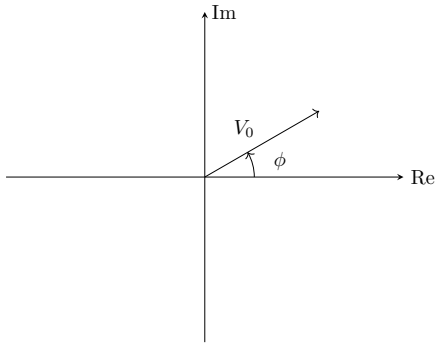
Based on these behaviors, the circuit in Figure 3.6.1 is called a **RC high-pass filter**. Thus, if a voltage input signal contains many sinusoidal signal components with a variety of frequencies, then this filter circuit can eliminate the low-frequency components while leaving the high-frequency components unaffected.

3.7 Phasors

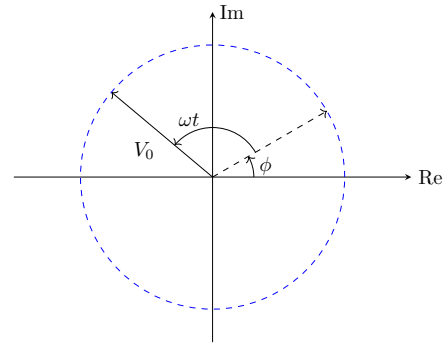
In this section, we will introduce **phasors**, a graphical method of representing complex electronics signals. Take a complex signal, e.g.

$$\tilde{V}(t) = \tilde{V}_0 e^{i\omega t} = V_0 e^{i\phi} e^{i\omega t} = V_0 e^{i(\omega t + \phi)}.$$

For time $t = 0$, we can represent this graphically as a vector at an angle ϕ as shown in Figure 3.7.1(a). This phasor will rotate about a circle with radius V_0 as time advances, pointing at an angle $\omega t + \phi$ at some later time t as shown in Figure 3.7.1(b).



(a) $V(t=0)$



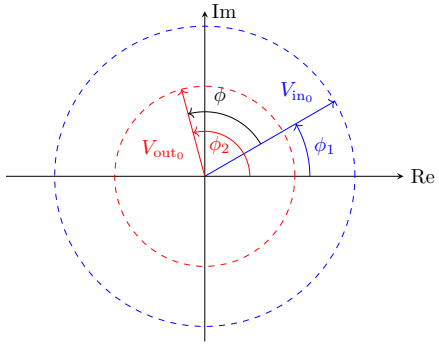
(b) $V(t)$ relative to $V(t=0)$.

Figure 3.7.1 Phasor example.

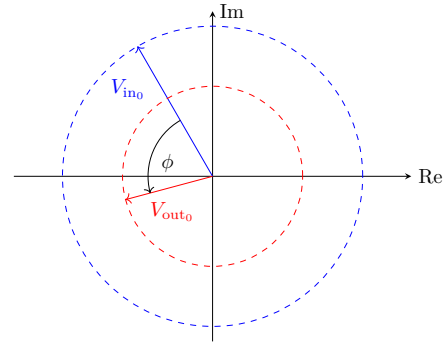
If we have two signals that we would like to compare

$$\tilde{V}_{\text{out}} = V_{\text{out}_0} e^{i(\omega t+\phi_2)}$$
$$\tilde{V}_{\text{in}} = V_{\text{in}_0} e^{i(\omega t+\phi_1)}$$

we can see in Figure 3.7.2 that both vectors rotate at the same rate, leaving the phase $\phi = \phi_2 - \phi_1$ between them unchanged as time advances.



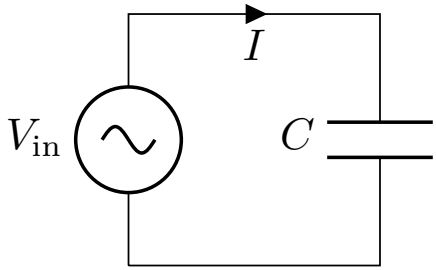
(a) Time $t = 0$



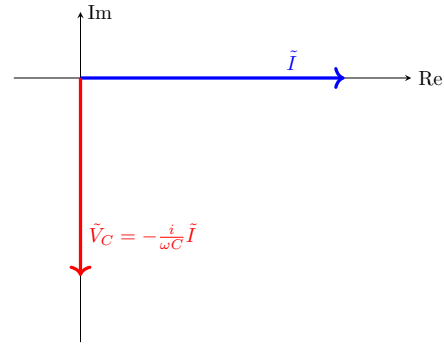
(b) Some later time t

Figure 3.7.2 Two phasors with the same ω advancing in time.

Phasors can help us quickly use graphs to determine the amplitudes and phases of signals relative to some other reference signal. As an example, let's examine the circuit in Figure 3.7.3(a).



(a) Circuit



(b) Phasor diagram.

Figure 3.7.3 Phasor analysis for a capacitor.

The phasor diagram in Figure 3.7.3(b) was generated by completing the following steps:

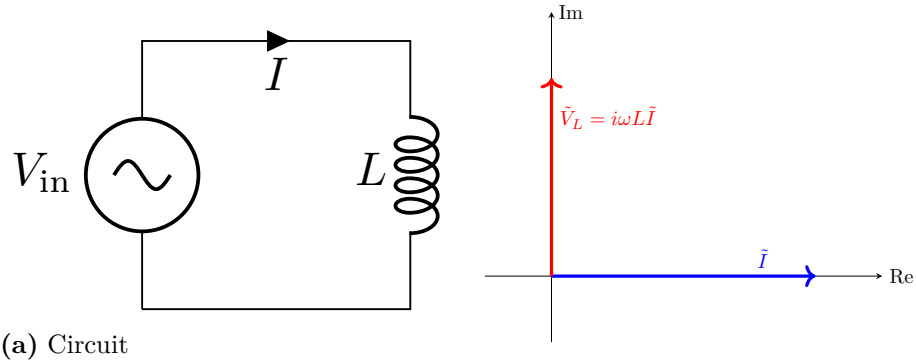
1. Draw your phasor diagram for the instant when one of your signals (voltage or current) is purely real, meaning its phasor is aligned along the real axis. Choose this signal based on how easily it can be related to other phasors. In this case, we'll look at an instant when $\tilde{I}$ is purely real.

2. Use Ohm's law to find expressions for voltage changes across components. In this case, we used $\tilde{V}_C = \tilde{I}Z_C = -(i/\omega C)\tilde{I}$.

3. Plot $\tilde{V}$ changes across all components. In this example, $\tilde{I}$ is purely real which means that $\tilde{V}_C$ is aligned with the negative imaginary axis.

Thus, we find that the phase of the current $\tilde{I}$ through a capacitor leads the phase of the capacitor voltage $\tilde{V}_C$ by 90° and the amplitude of the voltage is $V_{C_0} = I_0/\omega C$.

We can proceed with a similar analysis for an inductor. Examining the circuit in Figure 3.7.4(a).



(a) Circuit

(b) Phasor diagram.

Figure 3.7.4 Phasor analysis for an inductor.

and generating the phasor diagram in Figure 3.7.4(b), we find that the phase of the current $\tilde{I}$ through an inductor trails the phase of the inductor voltage $\tilde{V}_L$ by $90°$ and the amplitude of the voltage is $V_{L_0} = \omega L I_0$.

THINGS TO CHECK: WHICH EQUATIONS SHOULD HAVE TILDE VARIABLES AND WHICH SHOULD NOT.

Example 3.7.5 Phasor analysis of the RC High-pass filter.

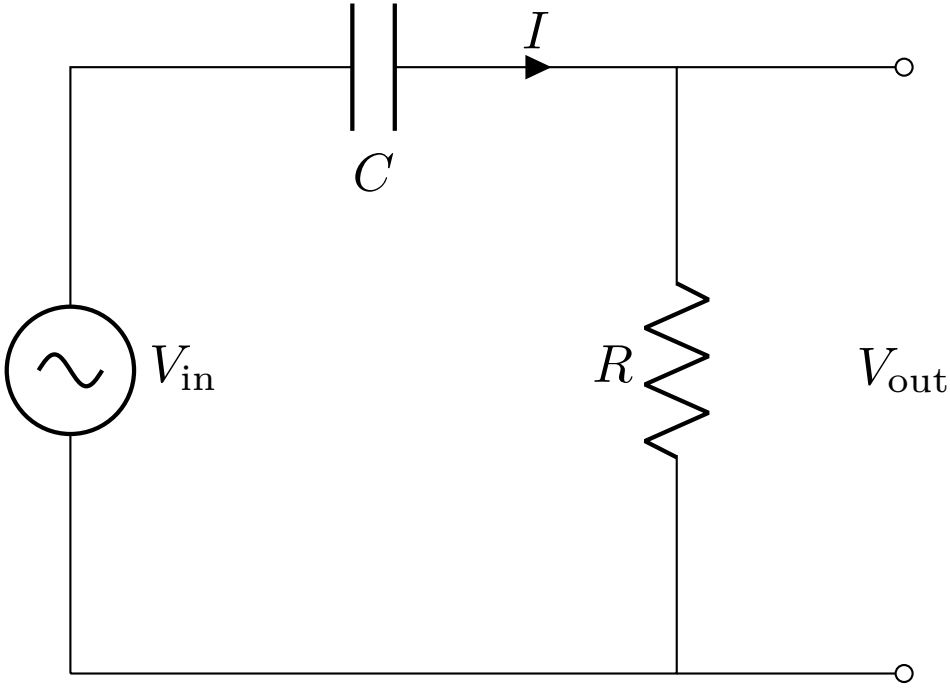


Figure 3.7.6 High-pass RC Filter.

Let's revisit the RC high-pass filter circuit in Figure 3.6.1 and analyze the circuit behavior using phasors. Find the voltage gain G_V and the phase $\phi = \phi_{\text{out}} - \phi_{\text{in}}$.

Answer.

$$G_V = \frac{V_{\text{out}_0}}{V_{\text{in}_0}} = \frac{1}{\sqrt{\left(\frac{1}{\omega RC}\right)^2 + R^2}} = \frac{1}{\sqrt{1 + (f_0/f)^2}}$$

$$|\phi| = \tan^{-1}\left(\frac{1}{\omega RC}\right) = \tan^{-1}(f_0/f)$$

$$f_0 = 1/2\pi RC$$

Solution.

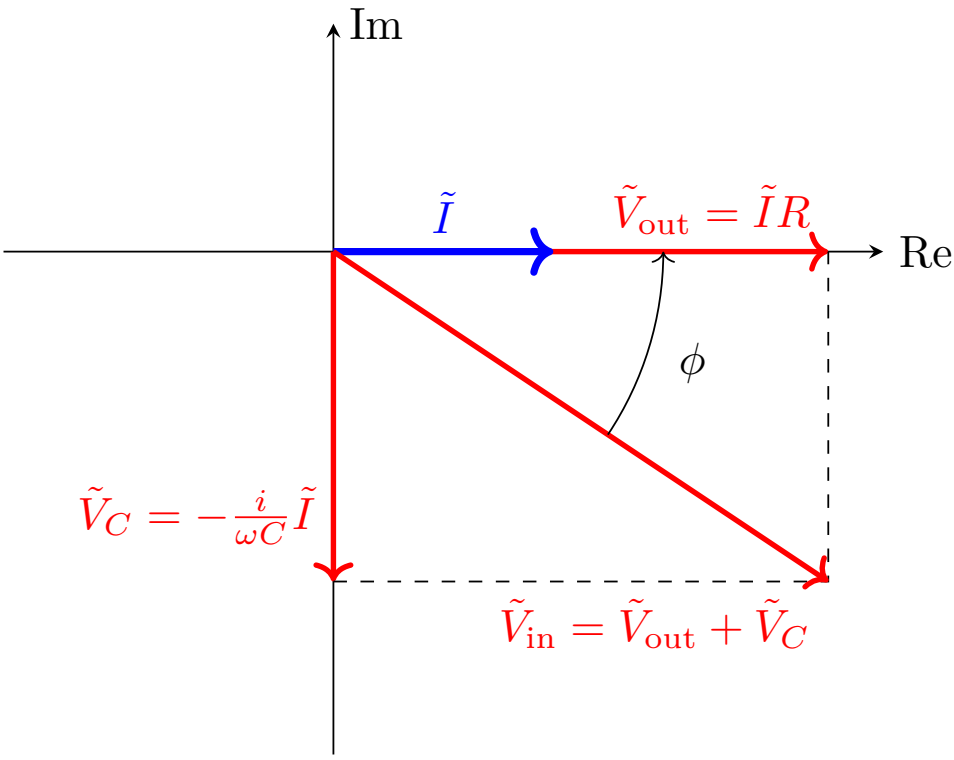


Figure 3.7.7

To construct the phasor diagram in Figure 3.7.7, we draw $\tilde{I}$ along the real axis. Then, $\tilde{V}_{\text{out}} = \tilde{V}_{\text{R}} = \tilde{I}R$ is also purely real. We can use Ohm's law to draw $\tilde{V}_C = i\tilde{I}/\omega C$ along the negative imaginary axis. Then, since Kirchhoff's Voltage Law requires $\tilde{V}_{\text{in}} - \tilde{V}_C - \tilde{V}_{\text{out}} = 0$, we use vector addition to draw $\tilde{V}_{\text{in}} = \tilde{V}_C + \tilde{V}_{\text{out}}$.

Based on our phasor diagram, we can already conclude the following:

1. $\tilde{V}_{\text{out}}$ leads $\tilde{V}_{\text{in}}$.
2. At high frequencies, $\tilde{V}_{\text{out}} \to \tilde{V}_{\text{in}}$ and $\phi \to 0$ because $\tilde{V}_C \ll \tilde{V}_{\text{out}}$.
3. At low frequencies, $\tilde{V}_{\text{out}} \to 0$ and $\phi \to 90°$ because $\tilde{V}_C \gg \tilde{V}_{\text{out}}$.

We can use the phasor diagram to derive more precise relationships. Using the Pythagorean theorem, we find that

$$V_{\text{in}_0} = \sqrt{I_0^2 \left(\frac{1}{\omega C}\right)^2 + I_0^2 R^2} = I_0 \sqrt{\left(\frac{1}{\omega C}\right)^2 + R^2}.$$

Also,

$$V_{\text{out}_0} = I_0 R.$$

Combining these results, we find that

$$G_V = \frac{V_{\text{out}_0}}{V_{\text{in}_0}} = \frac{R}{\sqrt{\left(\frac{1}{\omega C}\right)^2 + R^2}} = \frac{1}{\sqrt{\left(\frac{1}{\omega RC}\right)^2 + R^2}}$$
$$= \frac{1}{\sqrt{1 + (f_0/f)^2}}$$

where $f_0 = 1/2\pi RC$ as before. Likewise, we can find the phase between our signals using

$$|\phi| = \tan^{-1}\left(\frac{V_{C_0}}{V_{R_0}}\right) = \tan^{-1}\left(\frac{1}{\omega RC}\right) = \tan^{-1}\left(f_0/f\right).$$

These results are identical to what we had calculated previously using Kirchhoff's laws. ▲

Example 3.7.8 Phasor analysis of a resonant RLC circuit.

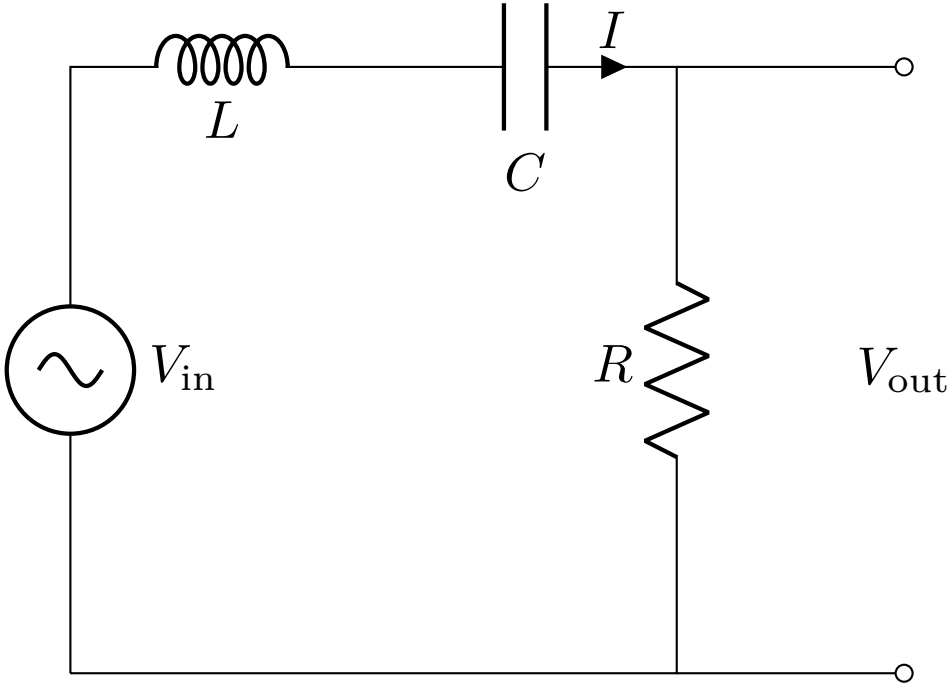


Figure 3.7.9 Resonant RLC circuit.

Find G_V and $\phi = \phi_{\text{out}} - \phi_{\text{in}}$ for the resonant RLC circuit shown in Figure 3.7.9.

Answer. The voltage gain is

$$\begin{aligned} G_v = \frac{V_{\text{out}_0}}{V_{\text{in}_0}} &= \frac{R}{\sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}} \\ &= \frac{1}{\sqrt{1 + \frac{L^2}{\omega^2 R^2}\left(\omega^2 - \omega_0^2\right)^2}}. \end{aligned}$$

The phase is

$$\begin{aligned} |\phi| &= \tan^{-1}\left(\frac{\omega L - \frac{1}{\omega C}}{R}\right) \\ &= \tan^{-1}\left(\frac{L}{\omega R}\left(\omega^2 - \omega_0^2\right)\right) \end{aligned}$$

and the resonant frequency is

$$\omega_0 = \frac{1}{\sqrt{LC}}$$

Solution.

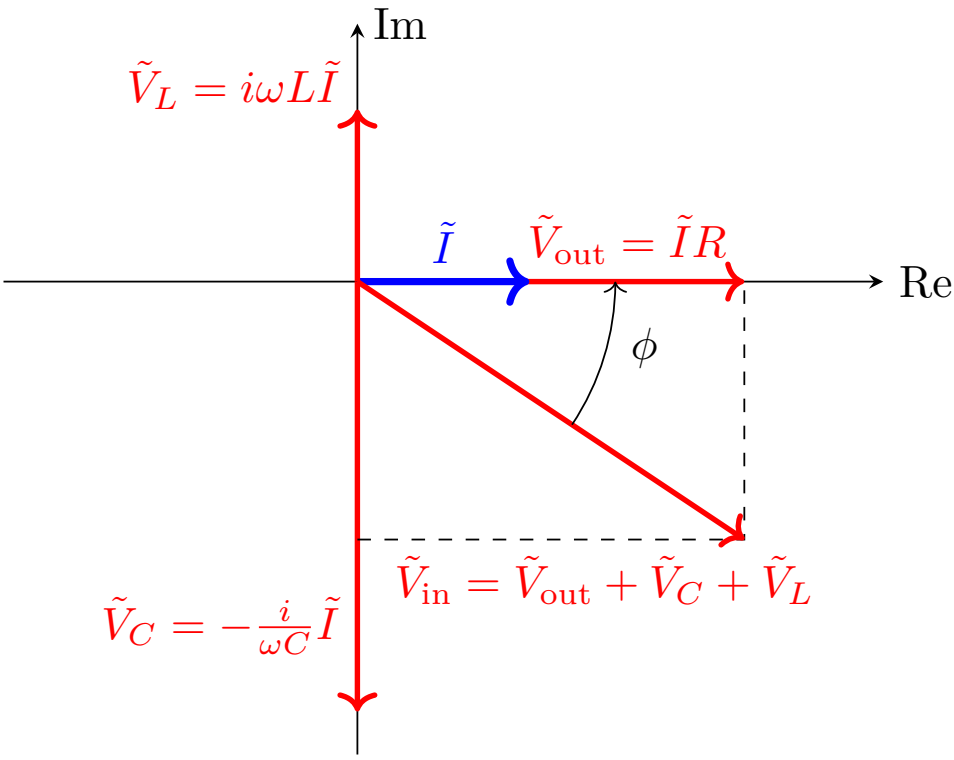


Figure 3.7.10 Phasor diagram.

Based on our phasor diagram, we can already conclude the following:

1. At high frequencies, $\tilde{V}_{\text{out}} \to 0$ and $\phi \to -90°$ because $\tilde{V}_C \ll \tilde{V}_{\text{out}}$ and $\tilde{V}_L \gg \tilde{V}_{\text{out}}$.
2. At low frequencies, $\tilde{V}_{\text{out}} \to 0$ and $\phi \to 90°$ because $\tilde{V}_C \gg \tilde{V}_{\text{out}}$ and $\tilde{V}_L \ll \tilde{V}_{\text{out}}$.
3. There exists a frequency $\omega = \omega_0$ where $\tilde{V}_C + \tilde{V}_L = 0$. At this frequency, $\tilde{V}_{\text{out}} = \tilde{V}_{\text{in}}$ and $\phi = 0°$. We call ω_0 the resonant frequency for this circuit.

We can be more quantitative with results gleaned from the phasor diagram. First, our resonance frequency can be determined by setting $\tilde{V}_C + \tilde{V}_L = 0$, giving us

$$\omega_0 L = \frac{1}{\omega_0 C}$$

which can be rearranged into

$$\omega_0 = \frac{1}{\sqrt{LC}}.$$

We can also find expressions for voltage gain and phase using trigonometry. Since the loop law gives

$$\begin{aligned} 0 &= \tilde{V}_{\text{in}} - \tilde{V}_L - \tilde{V}_C - \tilde{V}_{\text{out}} \\ \Rightarrow \tilde{V}_{\text{in}} &= \tilde{V}_L + \tilde{V}_C + \tilde{V}_{\text{out}} \\ &= i\left(\omega L - \frac{1}{\omega C}\right) I_0 + I_0 R, \end{aligned}$$

we can use Pythagorean's theorem to find V_{in_0}

$$\begin{aligned} V_{\text{in}_0} &= \sqrt{\left(\omega L - \frac{1}{\omega C}\right)^2 I_0^2 + I_0^2 R^2} \\ &= I_0 \sqrt{\left(\omega L - \frac{1}{\omega C}\right)^2 + R^2}. \end{aligned}$$

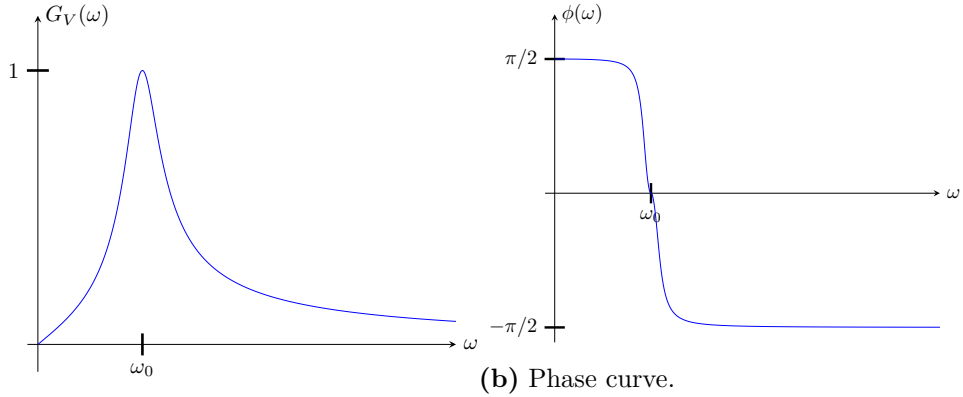
Solving this equation for I_0 in terms of V_{in_0} and inserting into $V_{\text{out}_0} = I_0 R$ gives

$$\begin{aligned} G_v = \frac{V_{\text{out}_0}}{V_{\text{in}_0}} &= \frac{R}{\sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}} \\ &= \frac{1}{\sqrt{1 + \frac{L^2}{\omega^2 R^2}\left(\omega^2 - \omega_0^2\right)^2}}. \end{aligned}$$

The phase is found using

$$\begin{aligned} |\phi| &= \tan^{-1}\left(\frac{\omega L - \frac{1}{\omega C}}{R}\right) \\ &= \tan^{-1}\left(\frac{L}{\omega R}\left(\omega^2 - \omega_0^2\right)\right). \end{aligned}$$

These results are plotted in Figure 3.7.11, demonstrating the same qualitative behaviors that we had deduced at the outset.



(a) Gain curve.

(b) Phase curve.

Figure 3.7.11 Resonant RLC circuit curves.

▲

Example 3.7.12 Phasor analysis of a parallel RLC circuit.

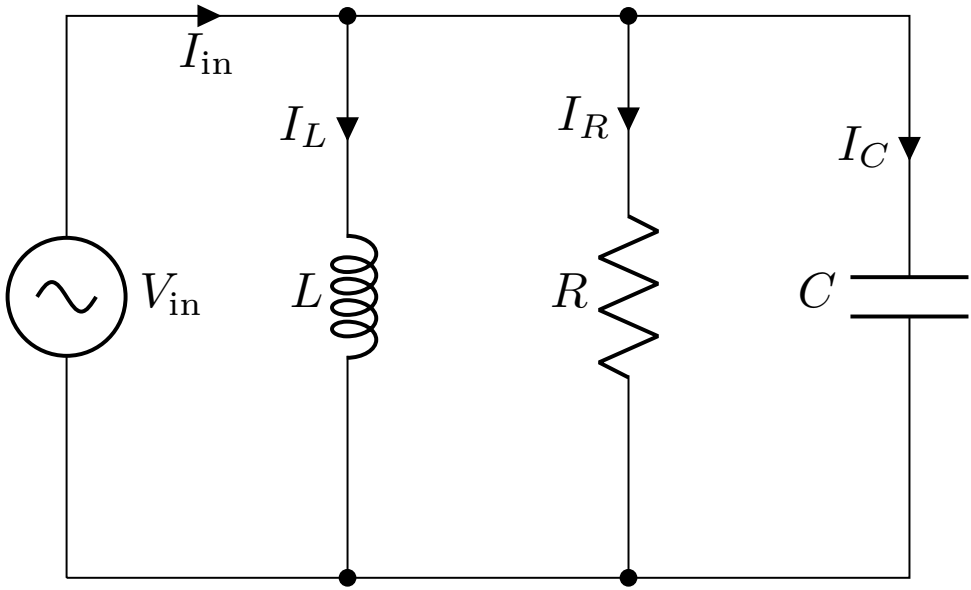


Figure 3.7.13 Parallel RLC circuit.

Examine Figure 3.7.13. Determine the current being drawn from the voltage source and the frequency at which the source current's magnitude is minimized.

Answer.

$$I_{\text{in}_0} = V_{\text{in}_0}\sqrt{\frac{1}{R^2} + \left(\omega C - \frac{1}{\omega L}\right)^2}$$

$$\phi = \tan^{-1}\left(\frac{1/R^2}{\omega C - 1/\omega L}\right) = \tan^{-1}\left(\frac{1}{R^2\left(\omega C - 1/\omega L\right)}\right)$$

Solution.

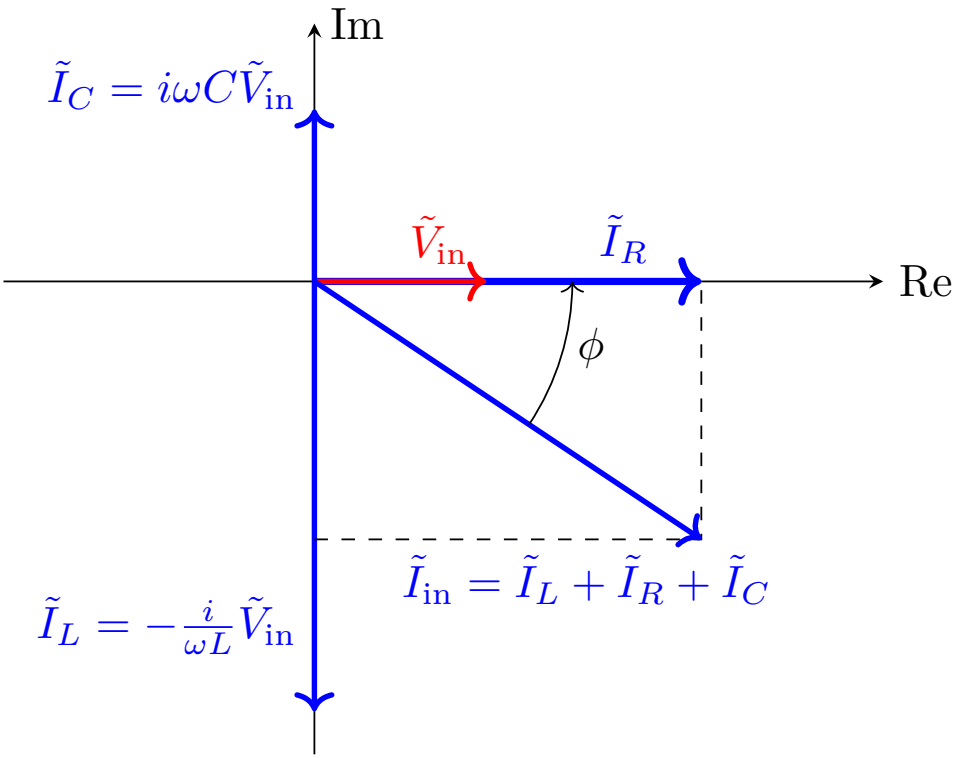


Figure 3.7.14 Phasor diagram.

In order to create the phasor diagram, we draw it at a time at which one convenient current phasor is aligned purely along the real axis. We will choose a time at which $\tilde{I}_R$ is purely real so that $\tilde{V}|_{\text{in}}$ is also purely real. We can then use Ohm's law to determine currents across the inductor and capacitor and draw them on the phasor diagram because the voltage across each of the three components is V_{in}.

Now, the junction law can be used to find $\tilde{I}_{\text{in}}$:

$$\tilde{I}_{\text{in}} = \tilde{I}_R + \tilde{I}_L + \tilde{I}_C.$$

Using Pythagorean's theorem,

$$\begin{aligned} I_{in_0}^2 &= I_{R_0}^2 + \left(I_{C_0} - I_{L_0}\right)^2 \\ &= \frac{V_{\text{in}_0}}{R^2} + \left(\omega C V_{\text{in}_0} - \frac{V_{\text{in}_0}}{\omega L}\right)^2 \\ &= V_{\text{in}_0}^2 \left[\frac{1}{R^2} + \left(\omega C - \frac{1}{\omega L}\right)^2\right] \end{aligned}$$

so that

$$I_{\text{in}_0} = V_{\text{in}_0} \sqrt{\frac{1}{R^2} + \left(\omega C - \frac{1}{\omega L}\right)^2}$$

and

$$\phi = \tan^{-1}\left(\frac{1/R^2}{\omega C - 1/\omega L}\right) = \tan^{-1}\left(\frac{1}{R^2\left(\omega C - 1/\omega L\right)}\right).$$

In the limits $\omega \to 0$ and $\omega \to \infty$, $\tilde{I}_{\text{in}} \to \infty$ since the inductor or capacitor short out in these limits, respectively. There are intermediate frequencies at which $\tilde{I}$ is finite. We can find the frequency ω_0 at which source current is minimized by solving for ω when $\mathrm{d}\tilde{I}_{\text{in}}/\mathrm{d}V = 0$. Current is also minimized when $\mathrm{d}\tilde{I}_{\text{in}}^2/\mathrm{d}V = 0$. So, we'll examine

$$0 = \frac{\mathrm{d}}{\mathrm{d}V}\tilde{I}_{\text{in}}^2 = 2\tilde{V}_{\text{in}}^2 \left(\frac{1}{\omega L} - \omega C\right)\left(-\frac{1}{\omega^2 L} - C\right).$$

Now, it is impossible for the second term in parentheses to equal zero. So, source current is minimized when

$$\frac{1}{\omega_0 L} = \omega_0 C$$

or

$$\omega_0 = \frac{1}{\sqrt{LC}}.$$

▲

3.8 Using Python in AC circuits

Just as with DC circuits, we can let Python do most of the mathematical manipulation for us when determining circuit behavior. As an example, we'll examine the behavior of voltage gain and phase $\phi = \phi_{\text{out}} - \phi_{\text{in}}$ for the circuit in Figure 3.8.1.

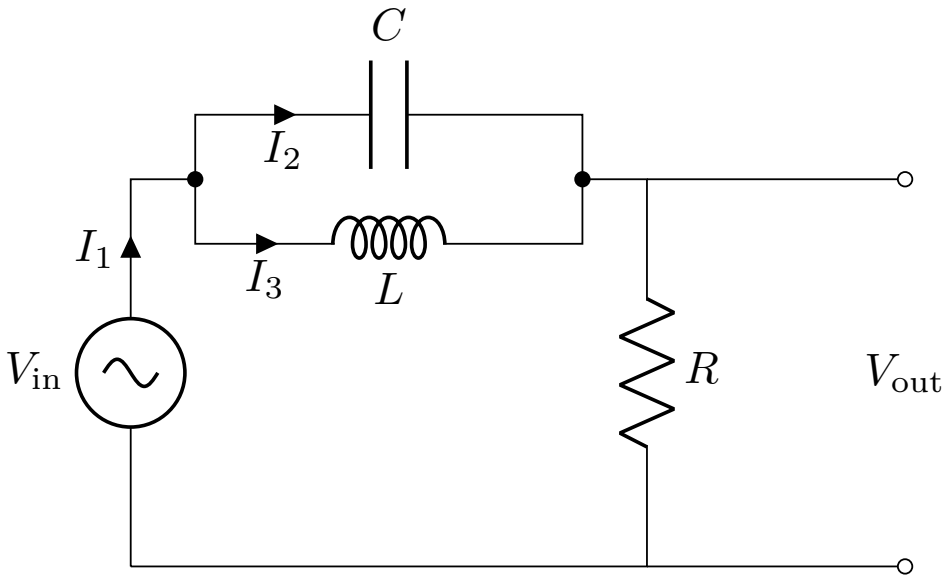


Figure 3.8.1 Bandgap RLC Circuit.

Using Kirchhoff's laws and the branch method, we find the following equations

$$\begin{aligned}\tilde{I}_1 - \tilde{I}_2 - \tilde{I}_3 &= 0\\ \tilde{V}_{\text{in}} - \frac{1}{i\omega C}\tilde{I}_2 - R\tilde{I}_1 &= 0\\ \tilde{V}_{\text{in}} - i\omega L\tilde{I}_3 - R\tilde{I}_1 &= 0.\end{aligned}$$

Rearranging these equations into

$$\begin{aligned}\tilde{I}_1 - \tilde{I}_2 - \tilde{I}_3 &= 0\\ R\tilde{I}_1 - \frac{i}{\omega C}\tilde{I}_2 &= \tilde{V}_{\text{in}}\\ R\tilde{I}_1 + i\omega L\tilde{I}_3 &= \tilde{V}_{\text{in}}\end{aligned}$$

facilitates efforts to recast this system of equations as a matrix equation

$$\begin{pmatrix} 1 & -1 & -1 \\ R_1 & -\frac{i}{\omega C} & 0 \\ R_1 & 0 & i\omega L \end{pmatrix} \begin{pmatrix} \tilde{I}_1 \\ \tilde{I}_2 \\ \tilde{I}_3 \end{pmatrix} = \begin{pmatrix} 0 \\ \tilde{V}_{\text{in}} \\ \tilde{V}_{\text{in}} \end{pmatrix}$$

which can be solved using Python:

```
            # Clear all variables that may be present from
                previous code executions
globals().clear()

import numpy as np
import scipy.linalg as linalg
import matplotlib.pyplot as plt

plt.close('all'); # Close all open figures

f = np.logspace(0,8,1000)
w_arr = 2*np.pi * f
Vin = 1 # Volts
R, C, L = [1000, 1e-9, 1e-3] # Ohms, Farads, Henrys

Gv = np.zeros(np.size(w_arr));
phi = np.zeros(np.size(w_arr));

for i in range(np.size(w_arr)):
    w = w_arr[i];
    A = np.array([[1, -1, -1],[-R, 1j/(w*C), 0],[0, -1j/(w*C),
        -1j*w*L]]);
    v = np.array([0, -Vin, 0]);

    x = linalg.solve(A,v);
    Gv[i] = np.abs(x[0]*R/Vin) # Voltage Gain where I1 = x[0]
    phi[i] = np.angle(x[0]*R) # Phase of output (assuming Vin is
        at zero phase)

fig, axs = plt.subplots(2,1,sharex=True)

axs[0].semilogx(f,Gv)
axs[1].semilogx(f,phi)
# set labels -- Note that $\\phi$ is used in the ylabel below in
    order to avoid deprication warning
plt.setp(axs[0], ylabel='Voltage␣Gain')
plt.setp(axs[1], xlabel='Frequency␣(Hz)',
    ylabel='$\\phi_{out}-\\phi_{in}$',
    yticks=[-np.pi/2,0,np.pi/2], yticklabels=['$-\\pi/2$', 0,
    '$\\pi/2$'])
plt.show()
```

Here, we have made use of the `abs` function in NumPy to find the voltage gain by measuring the magnitude of $\tilde{V}_{\text{out}}/\tilde{V}_{\text{in}}$, and we have used the `angle` function from

Numpy to find the phase of V_{out} at an instant in time when V_{in} is purely real.

Examining the results of our analysis, we see that this circuit has $G_v = 1$ and $\phi = 0$ at both high and low frequencies. There is a frequency at [PROVIDE EQUATION] at which $G_v = 0$ and $\phi = \pm\pi/2$. This circuit is called an RLC bandgap filter as it allows signals at most frequencies through unchanged, but has a band (defined by values of L and C) where frequency components are cut out. This can be especially useful if there is a source of noise with a single well-defined frequency contaminating a signal.

3.9 Input and Output Impedance

Let's revisit Thevenin's theorem and Norton's theorem. Both of these theorems remain valid when working with AC signals as long as we substitute impedances for resistances. Here, we'll focus our attention on the usefulness of Thevenin's theorem.

Thevenin's theorem states that the output voltage and current characteristics of a complicated circuit are identical to output characteristics of a simplified Thevenin-equivalent circuit (Figure 3.9.1) for proper choice of Thevenin voltage and impedance.

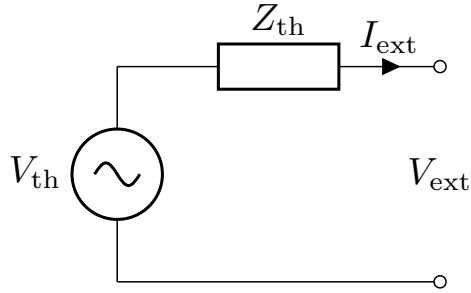


Figure 3.9.1

Here, we provide an example demonstrating the importance of Thevenin's theorem.

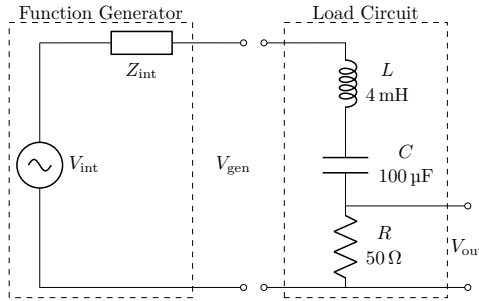


Figure 3.9.2 An LRC circuit connected to a function generator.

The complicated internal circuitry of a function generator will be equivalent to a single internal voltage source V_{int} and a single internal impedance Z_{int}. This internal impedance can be found on specification sheets for modern function generators, often labeled as **output impedance**. Most commonly, function generators have an output impedance of 50Ω, though some (like those built into prototyping boards) may have output impedances up to several hundred ohms.

The input impedance characterizes a circuit's output voltage and current characteristics based on the **load** that is applied across the output. A circuit's **input impedance** characterizes how hard the circuit is to drive and is equivalent to the load circuit's Z_{th}.

Let's use the example in Figure 3.9.2 to illustrate the meanings associated with the descriptions above. In this case, the load circuit is an inductor, a capacitor, and a resistor in series, so that

$$Z_{\text{load}} = Z_R + Z_C + Z_R = R + i\left(i\omega L - \frac{1}{\omega C}\right).$$

Now, if the load circuit is connected across the output of the function generator, we end up with a voltage-divider circuit as seen in Section 2.1. This means we can use that chapter's result while replacing resistances with impedances, so

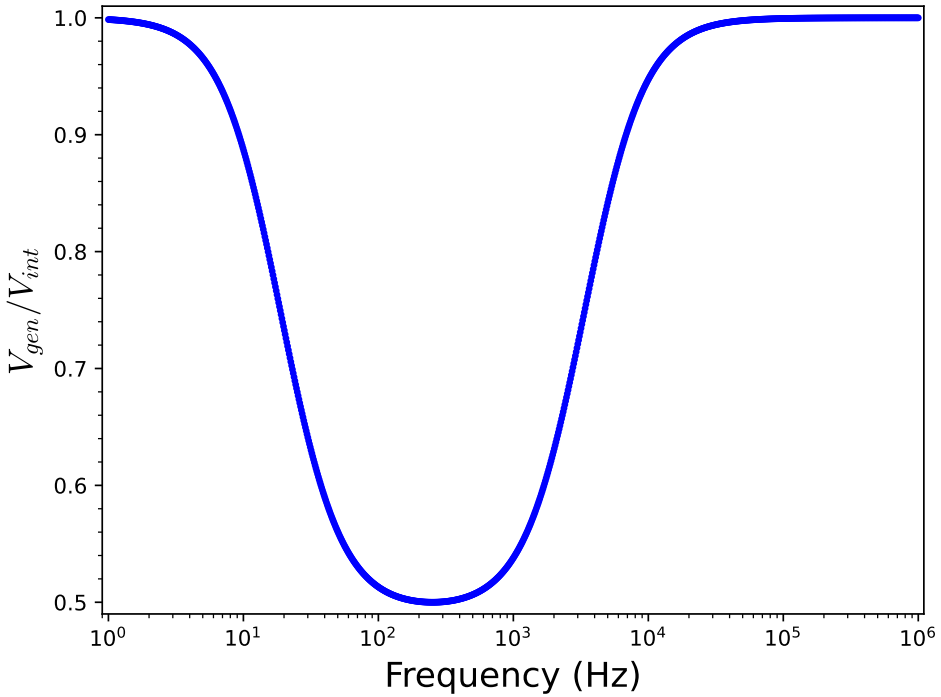
$$\tilde{V}_{\text{gen}} = \frac{Z_{\text{load}}}{Z_{\text{int}} + Z_{\text{load}}}\tilde{V}_{\text{int}}.$$

We find that Z_{load} is minimized when $Z_C + Z_L = 0$ which occurs when $\omega = \omega_0 = \frac{1}{\sqrt{LC}}$. We draw the following conclusions:

- For signal frequencies far from the resonant frequency ω_0, $|Z_{\text{load}}| \gg |Z_{\text{int}}|$ meaning that $\tilde{V}_{\text{supply}} \approx \tilde{V}_{\text{gen}}$.
- When $\omega \approx \omega_0$, we will find that $\tilde{V}_{\text{supply}}$ and $\tilde{V}_{\text{gen}}$ are very different if $Z_{\text{load}} \lesssim R$

We demonstrate this behavior by plotting the signal being produced at the output of the function generator V_{gen} as a function of frequency For the parameters specified in Figure 3.9.2.

```
import numpy as np
import scipy.linalg as linalg
import matplotlib.pyplot as plt

f = np.logspace(0,6,1000)
w = 2*np.pi * f
Vin = 1 # Volts
R, C, L = [50, 1e-4, 4e-3] # Ohms, Farads, Henrys
Z_load = 50 + 1j/(w*C) - 1j*w*L;
R_out = 50;
Gv = abs(Z_load / (R_out+Z_load));
phi = np.angle(Z_load / (R_out+Z_load));

plt.semilogx(f,Gv)
plt.xlabel('Frequency␣(Hz)');
plt.ylabel('$V_\\text{gen}␣/␣V_\\text{int}$');

plt.show()
```



Now, this has important consequences. First, extra care must be taken if one were to be studying voltage gain $V_{\text{out}}/V_{\text{gen}}$ and phase properties of the pictured load circuit. If one were to trust the signal amplitude displayed on the function generator for V_{gen} when calculating voltage gain values, these values would be grossly incorrect for frequencies near the resonance frequency because both V_{gen} and V_{in} vary with frequency.

These concepts also impact the design of oscilloscopes. If we want to use an oscilloscope to measure the voltage output of a function generator, we want the oscilloscope to have a very high impedance $\left(Z_{\text{osc}_{\text{input}}} \gg Z_{\text{int}}\right)$ so that the oscilloscope leaves the function generator output voltage V_{gen} unchanged across all frequencies.

So, input and output impedances for circuits is an important consideration when determining how a load circuit will impact the behavior of the driving circuit. We determine Thevenin equivalent circuit values for both the driving and load circuits in order to determine the input and output impedances of these circuits. Input and output impedances are also critical to consider when working at high frequencies (MHz and above). At these high frequencies, impedance matching (or matching input and output impedances of linked circuits) is needed to miminize signal reflections and maximize power transfer. This text will limit it's treatment to sub-MHz frequencies and will not treat these effects.

Chapter 4

Diodes

In this chapter, we're introduced to our first non-ohmic circuit element: the diode. These elements behave in ways that do not conform to Ohm's law.

Diodes are semiconductor-based circuit elements whose behaviors in circuits does not conform to Ohm's law. Before describing diode circuit behavior, we first need to understand what semiconductors are, what properties they exhibit, and how those behaviors combine to produce diode behavior.

4.1 Semiconductors

As the name suggets, semiconductors have conductivity between ocnductors and insulators. Semiondur properties have made many important electrical advances possible.

- diodes
- transistors
- operational amplifiers
- integrated circuit chips
- Light-emitting diodes (LEDs)
- photodiodes.

In order to understand semiconductors, we must examine the atomic structure of these materials and how they relate to molecular bonds. Let's start with Figure 4.1.1 which shows the energy levels for two isolated atoms.

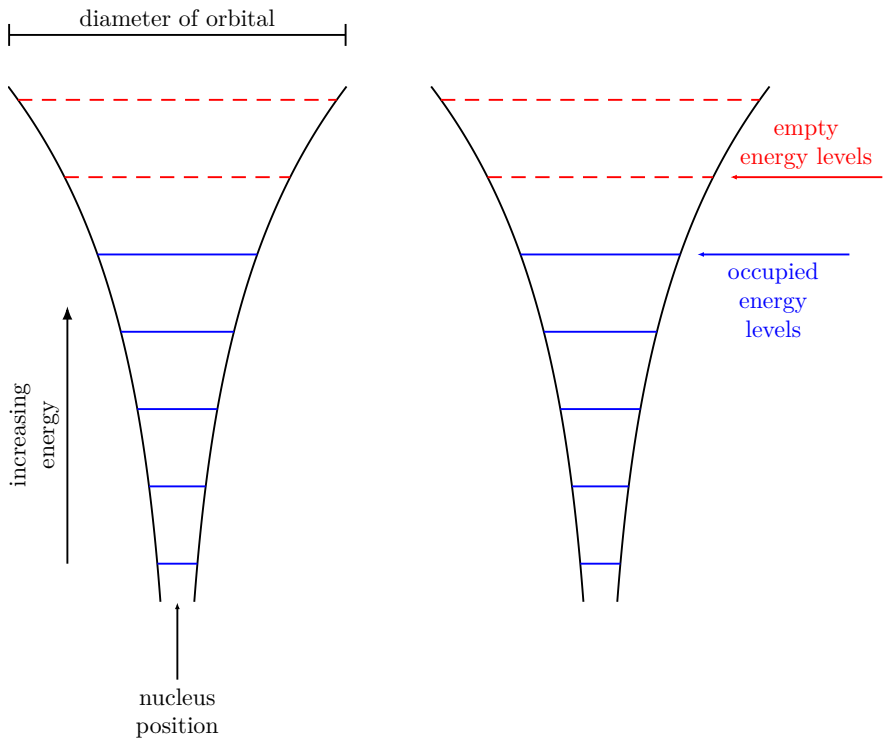


Figure 4.1.1

Quantum mechanics tells us that electrons can occupy discrete energy levels of an atom. Energy levels fill starting with the lowest and fill upward. Electrons occupying high energy levels are further from the nucleus and also have a larger spatial extent surrounding the nucleus compared to electrons in low energy levels, at least on average. Elecrons of course do not behave like macroscopic objects with definite position, but instead are represented by electron clouds. For the purposes of this class, we will not have to trouble ourselves with this level of detail.

In the atom pictured in Figure 4.1.1, the energy levels represented by solid lines are occupied by electrons, with the term **valence electrons** used to identify electrons in the highest occupied energy level. The dashed lines represent empty energy levels.

When these two formerly isolated items are brought close to each other such that the valence electron orbitals start overlapping, the pair of atoms can reach a lower energy state by sharing some of their valence electrons. These shared energy

levels develop into bands rather than precise discrete levels. Fuller details about the physics of covalent bonds can be found in many Modern Physics or Quantum Mechanics textbooks but will be unnecessary for us. Figure 4.1.2 provides a rough picture of the elements important to us.

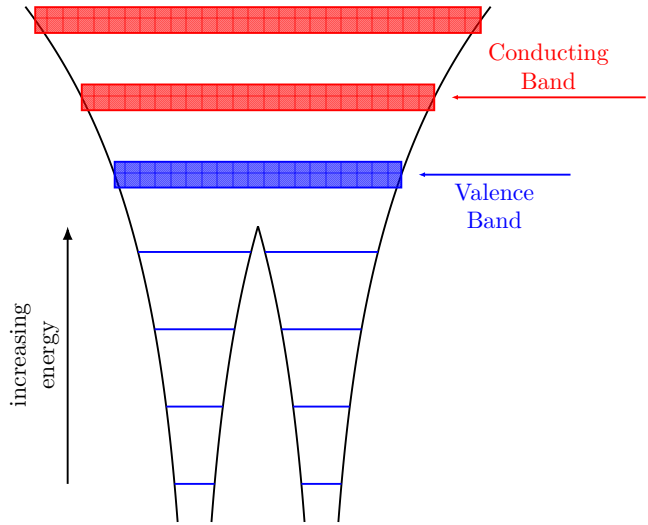


Figure 4.1.2

In this figure, the **valence band** represents the last occupied energy band at $T = 0$ Kelvin. The **conduction band** is the next energy band above the valence band, and the **band gap** is the energy difference between the top of the valence band and the bottom of the conduction band.

Now, these last three terms are critical to understanding the difference between conductors, semiconductors, and insulators, as illustrated in Figure 4.1.3.

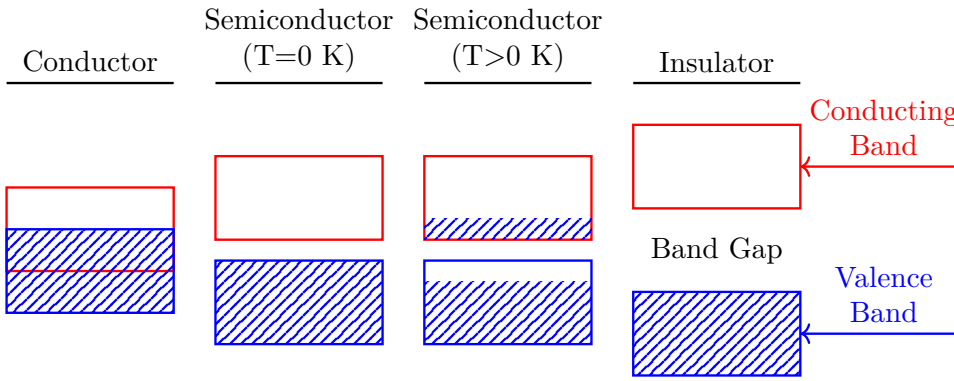


Figure 4.1.3

We see that conductors have overlapping valence and conduction bands. This means that the valence electrons, which are most loosely held by the nucleus due to their distance from the nucleus, can move relatively easily from one atom to the other using the empty orbital positions in the conduction band. In insulators, it is difficult to move valence electrons between atoms since those electrons would need to gain a large amount of energy to jump across the large bandgap from the valence to conduction band.

Semicondcuctors have non-overlapping valence and conduction bands, but with only a small bandgap between them. This means that at $T = 0$K, the semiconductor behaves like an insulator. But, at higher temperatures, there is enough thermal energy present that a small population of electrons that had been occupying the valence band instead make the jump into the conduction band and can now move freely.

One common semiconductor material is silicon $_{14}$Si which has an electron configuration that can be represented by $1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^2$. This means that Silicon has a full ground state (two electrons in the 1s orbital), a full second energy level (two electrons in the 2s orbital and six electrons in the 2p orbital), and a third energy level with electrons occupying half of the eight available orbital spots (two spots in the 3s orbital and six spots in the 3p orbital). Figure 4.1.4 shows a silicon crystal.

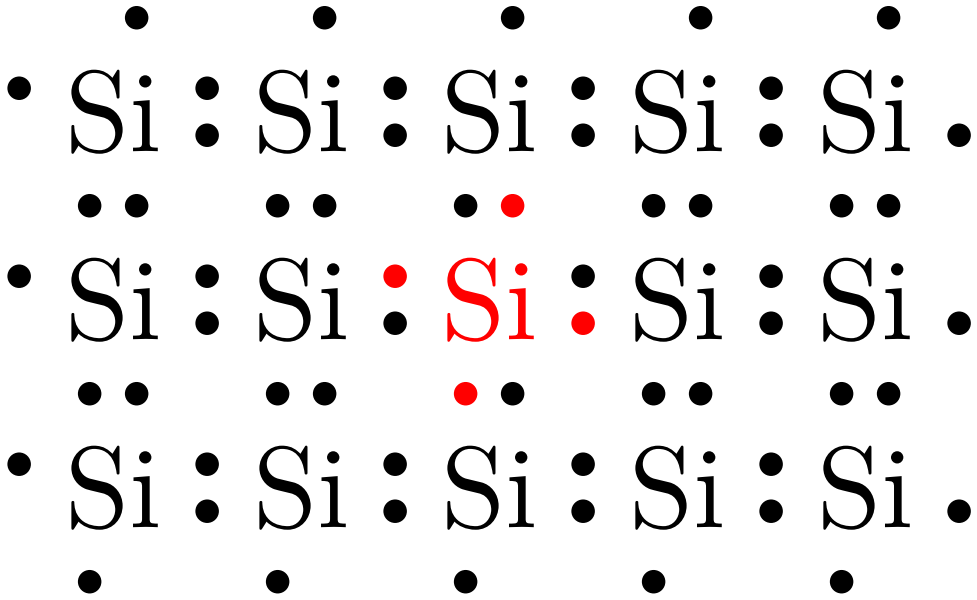
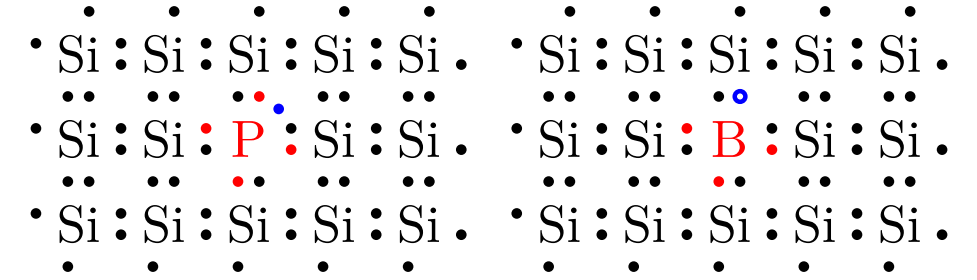


Figure 4.1.4 Silicon crystal. The shared electrons in covalent bonds fill the valence shell of the silicon crystal.

In this crystal, each Si atom is surrounded by eight electrons in the valence band, completely filling it. Each bond is formed by two electrons being shared between a pair of atoms, with each atom contributing one of the two shared electrons.

Now, on its own, silicon is a rather poor semiconductor. That said, the semiconductor properties of a Silicon crystal is improved significantly through doping. The term **doping** represents the process of mixing a small amount of an impurity into the silicon crystal.

There are two types of doping that are imporant to us here, and they are represented in Figure 4.1.5.



(a) n-type semiconductor. **(b)** p-type semiconductor.

Figure 4.1.5 Doped semiconductors.

In a **doped n-type semiconductor**, a small amount of phosphorus is added as an impurity into the silicon crystal. Phosphorus has five valence electrons in contrast to the four valence electrons that silicon has. This means that there are nine electrons trying to occupy the eight available spots in the valence band as shown in Figure 4.1.5(a). Consequently, each ninth phosphorus electron is bumped up into the shared conducting band and can move freely. This is called an n-type semiconductor because negatively-charged electrons are the **charge carrier**, meaning they are the moving charged particles that comprise any current flowing through the semiconductor.

In a **doped p-type semiconductor**, a small amount of Boron is added as an impurity into the silicon crystal. Phosphorus has three valence electrons in contrast to the four valence electrons that silicon has. This means that there are seven electrons trying to occupy the eight available spots in the valence band, leaving an open orbital spot in the valence bqand as shown in Figure 4.1.5(b). Consequently, the holes formed from the missing electron exist in the valence band and represent open spaces into which electrons can move. The reason this is called a p-type semiconductor is pictured in Figure 4.1.6

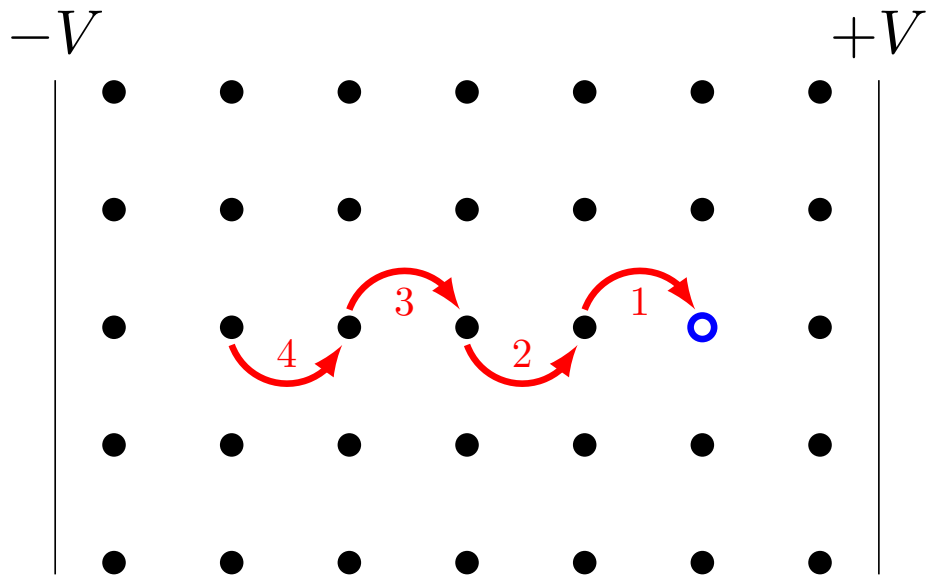


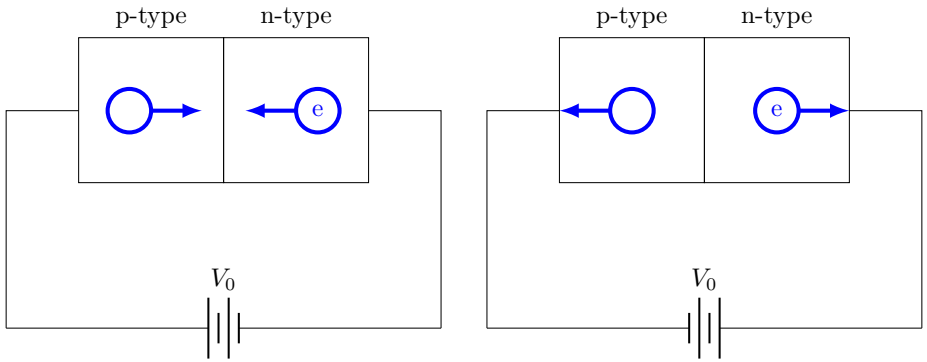
Figure 4.1.6

In this figure, a valence electron feels a force due to the applied voltages and moves to fill the hole (step 1), resulting in a new hole at the electron's former position. This process repeats (step 2) with a different electron moving into the hole location, again swapping places. This continues through steps 3 and 4. The

reality of the situation has four electrons each moving one space. it is convenient, instead, to treat the hole (the absence of an electron) as a single positive (relative to an electron) charge that moves four spaces to the left due to the applied voltage, just as a positive charge would move. As such, this hole (which we are treating akin to a positive charge) is considered the charge carrier and can move somewhat freely through the valence band of the semiconductor crystal.

4.2 Diode behavior

When we put p-type and n-type semiconductors together as in Figure 4.2.1, we get a **diode**. When a voltage difference is applied across the diode, interesting behaviors can be observed.



(a) Forward-biased. **(b)** Reverse biased.

Figure 4.2.1

When the voltage is applied across the diode as in Figure 4.2.1(a), we say the diode is **forward-biased**. Valence-band holes in the valence band of the p-type semiconductor and conduction-band electrons in the n-type semiconductor move toward the pn junction. At the junction, electrons in the n-type semiconductor fall across the pn junction into p-type semiconductor holes. Since the valence band is a lower energy level than the conduction band, this is accomplished easily. Electrons are also pulled out of the p-type into the wire, generating more valence-band holes. These electrons travel through the circuit and replenish the conduction-band electrons when they enter the n-type semiconductor. This process can proceed indefinitely, resulting in sustained current flow through the diode.

If the bias is reversed as in Figure 4.2.1(b), we say the diode is **reverse-biased**. In this case, valence-band holes and conduction-band electrons move away from the pn junction in their respective semiconductors. While electrons can leave the n-type semiconductor, travel through the circuit, and fill the holes in the p-type

semiconductor, this will quickly come to a halt because no new holes are formed at the p-n boundary.

Using the behaviors observed above, we find that diodes act similar to one-directional wires. The behavior is captured in Figure 4.2.2.

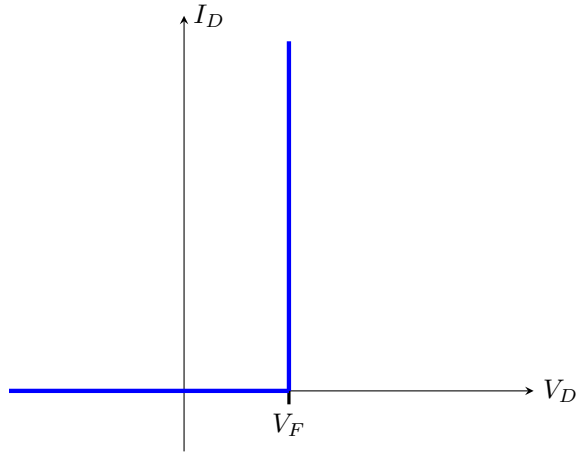


Figure 4.2.2

In this figure, V_D is the voltage drop across the diode from the p-type to n-type diode terminals and V_F is called the **forward turn-on voltage**. When $V_D < V_F$, the diode is 'off' meaning no current flows through the diode and it acts as a circuit break. When $V_D = V_F$, the diode is 'on' meaning the diode starts acting like a wire, albiet with a voltage drop V_F in the direction of current. The value of V_F varies based on the semiconductor material that the diode is made from, with $V_F \approx 0.6$V for Silicon, $V_F \approx 0.3$V for Germanium, and varied values for other materials. For the purposes of this text, we will assume that diodes are silicon-based with $V_F = 0.6$V.

When using the behavior shown in Figure 4.2.2 to analyze circuits containing diodes, the voltage drop across the diode will remain V_F as long as the diode is on regardless of the size of the current through the diode. Likewise, a diode that is off will allow no current regardless of the voltage change applied across the diode. Thus, a diode acts much like a unidirectional wire, allowing current to flow in one direction but not the other.

4.2.1 Origin of the forward turn-on voltage

We've seen that much of the cartoon behavior of a diode shown in Figure 4.2.2 can be understood qualitatively through an examination of Figure 4.2.1, though this examination failed to predict the forward turn-on voltage V_F. In Figure 4.2.2, the diode fails to conduct current while $V_D < V_F$. In order to understand the origins

of V_F, we will examine a more sophisticated treatment of charge carrier motions at the pn-junction.

A diode is made by joining p-type and n-type semiconductors together, but the typical reality is that a single semiconductor crystal is grown and the **p-n junction** refers to the region in the crystal where doping has been changed from one impurity to another. This means that the boundary between the semiconductor types is not as clearly defined as we're treating it to be. Nonethetless, our phenomenological exploration of diode behavior does not require the level of precision that a more complex junction model would provide, so we will continue treating the pn-junction as a sharp boundary.

Figure 4.2.3 represents a diode and takes a close look at the region around the p-n junction. In this figure, the 'e' symbols represent electrons in the conducting band of the n-type semiconductor and the circles represent holes in the valence band of the p-type semiconductor. Charge carriers distribute themselves uniformly throughout most of each semiconductor.

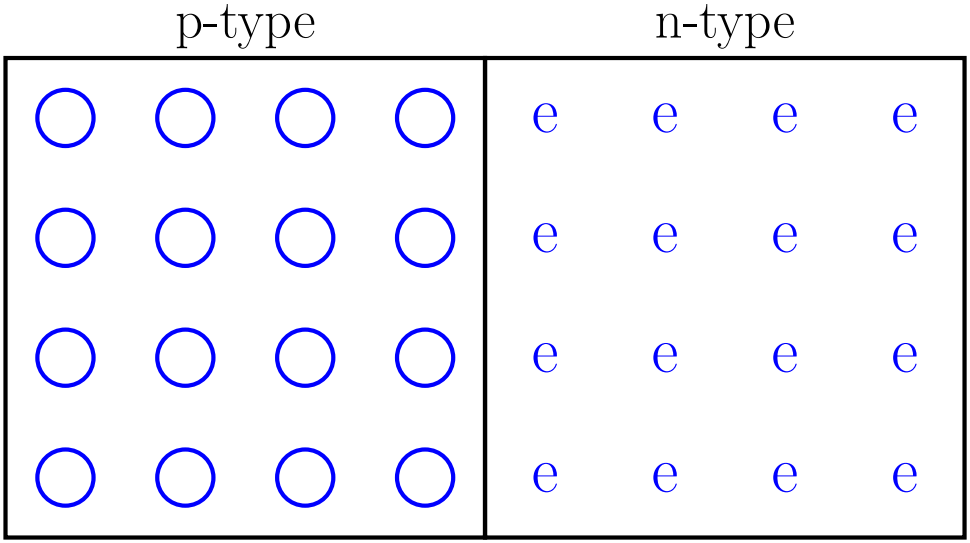


Figure 4.2.3 Charge carriers distribute uniformly within each semiconductor due to thermal motions.

Near the p-n junction boundary in the n-type semiconductor, thermal motions will cause some electrons to fall from the n-type conducting band into electron holes near the junction in the p-type valence band. These electrons lose energy as they fall from the conduction band to the valence band and become trapped on the p-type side of the p-n junction. Consequently, holes appear in the n-type conduction band in spaces vacated by the falling electrons. A **depletion layer** is formed around the

boundary as shown in Figure 4.2.4.

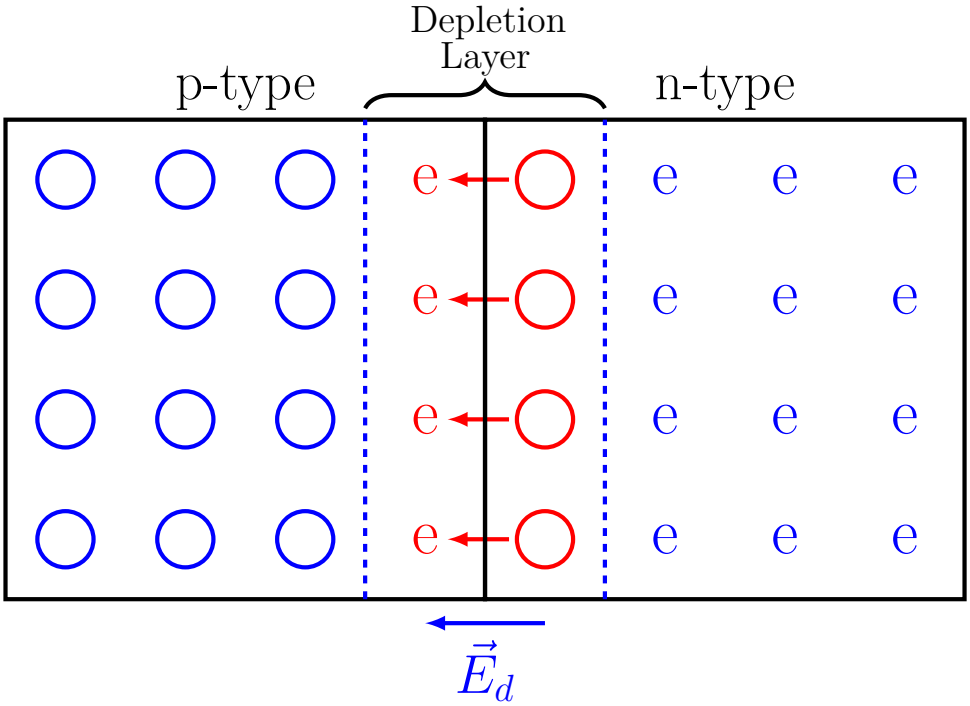


Figure 4.2.4 In equilibrium, a depletion layer forms around the pn-junction. An electric field $\vec{E}_d$ forms in this depletion layer pointing from the n-type to p-type semiconductor.

This depletion layer is so named because mobile charge carriers have been depleted in this layer. Electrons that have fallen into the valence band holes have no other holes around them in which to move. Likewise, the holes that appear in the n-type portion of the depletion layer are surrounded by other holes in the conduction band. The electrons that have fallen into the p-type valence band result in a negative charge imbalance (since boron has one fewer proton than Silicon) and the new n-type conduction band holes likewise result in a positive charge imbalance, producing an electric field $\vec{E}_d$ directed across the depletion layer from the n-type to p-type semiconductor. This depletion layer electric field will repel any further charge carriers that would otherwise enter the depletion layer region, leaving the depletion layer without mobile charge carriers in the equilibrium state of this diode.

In order for our diode to conduct current when forward biased, an electric field must be applied externally that will overcome the repulsive effects of $\vec{E}_d$ and allow new charge carriers to enter the depletion layer. Assuming a depletion layer thick-

ness Δx, the potential at the n-type boundary of the depletion layer will be higher than the potential at the p-type boundary of the depletion layer by an amount $V_d = \left|\vec{E}_d\right| \Delta x$. Since the diode's semiconductor regions outside of the depletion layer are highly conductive, any bias voltage applied externally across the diode will appear unchanged across the depletion layer. This means that a bias voltage $|V_{\text{bias}}| \geq |V_d|$ is required for current to be conducted through the diode, as shown in Figure 4.2.2.

4.2.2 Behavior of real diodes

We can also be a bit more precise and realistic in our description of diode behavior. Instead of the behavior described in Figure 4.2.2, we can instead relate the current through the diode I_D in the forward direction to the voltage applied across the diode V_D:

$$I_D = I_0 \left(e^{eV_D/nk_BT} - 1\right) = I_0 \left(e^{V_D/nV_T} - 1\right) \tag{4.2.1}$$

where I_0 is the saturation current, k_B is Boltzmann's constant, and T is temperature. The value of n is dependent on the materials used to construct the diode. For silicon, $n \approx 1$. The diode behavior is graphically represented in Figure 4.2.5.

```
# Clear all variables that may be present from previous code run
    on this page.
globals().clear()

import numpy as np
import scipy.constants as sc
import matplotlib.pyplot as plt

plt.close('all') # Close all open figures

# Define input parameters
VD = np.linspace(-1,.65,10000);
I0, T, n = [3e-12, 293, 1];
VT = sc.k * T / sc.e;

# Calculate diode current
I_D = I0*(np.exp(VD/(n*VT))-1);

# Plot diode I-V characteristic curve
plt.plot(VD,I_D);
plt.xlabel('$V_D$␣(V)');
plt.ylabel('$I_D$␣(A)');
plt.show()
```

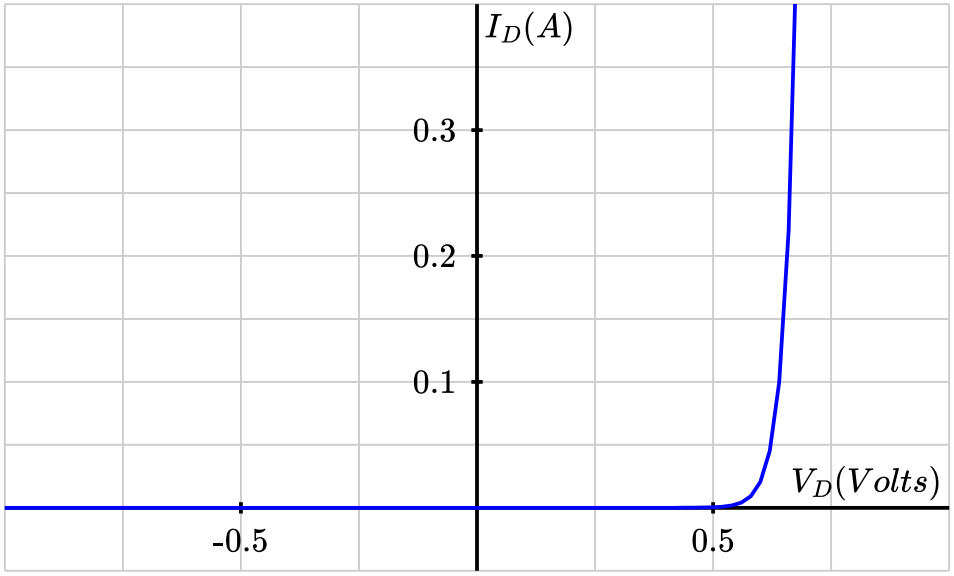


Figure 4.2.5 More realistic diode behavior.

The saturation current is the current that can bleed through the diode even when reverse-biased. This current is typically very small and can be found in the specifications for a real diode. For the purposes of this text, I will often assume that $I_0 = 1 \times 10^{-12}$A. I will also typically assume room temperature $T = 293$K.

4.3 Diode Types

Diodes come in many different types, each built to accomplish some task. [include circuit symbol for standard diodes. below, include symbols for special diodes.]

For instance, **Zener diodes** are designed to allow operation both when forward biased and reverse biased. [FIGURE TO COME].

Light-emitting diodes (LEDs) are the basis of many light bulbs and other light indicators. LEDs are designed so that when the diode is forward biased, electrons and holes are pushed toward each other at the p-n junction. When an electron falls across the p-n junction from the conduction band to the lower-energy valence band (a process called recombination), the electron releases energy in the form of light. The color and intensity of this light depends on the materials used to construct the diode. LEDs are very energy-efficient, especially compared to incandescent light bulbs.

Photodiodes are a bit like LEDs operating in reverse. When light hits a photodiode surface, energy is depositied into atoms. This energy kicks an electron from the valence band in the p-type semiconductor into the conduction band of the n-type semiconductor. Electrons are accelerated away from the depletion layer by $\vec{E}_d$, driving current from the n-type to p-type semiconductor, which is the opposite direction of current in standard diodes. Photodiodes are at the heart of solar panel technology as well as many light sensors.

4.4 Analysis of diode circuits

4.4.1 Analytic circuit treatment

How do we analyze circuits containing diodes? The simplest way is to use the cartoonish behavior demonstrated in Figure 4.2.2. Let's use this behavior to analyze the circuit in Figure 4.4.1.

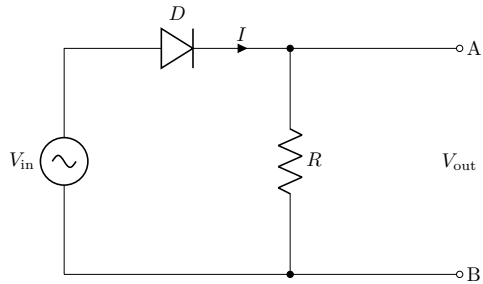


Figure 4.4.1

As the input voltage varies, it is not immediately obvious under what conditions the diode is on or off. Thus, our analysis will consist of us guessing at the diode state, analyzing the circuit, and looking for self-consistent solutions (indicating we guessed correctly) or contradictory statements (indicating we guessed incorrectly).

Let's start out guessing that the diode is off, requiring that the diode current $I_D = 0$. It follows from Ohm's law that there is no voltage change across the resistor R, thus $V_{\text{out}} = 0$. Kirchhoff's Voltage law requires $V_D = V_{in}$ which is a self-consistent solution for all times when $V_D < V_F$. For any times when $V_D > V_F$, we have a statement that contradicts our known diode behavior which states that $V_D = V_F$ when the diode is on.

For time periods where $V_D > V_F$ in the above analysis, we will instead assume that the diode is on meaning that the voltage change across the diode will lock at $V_D = V_F$. Under this condition, Kirchhoff's loop law requires that $V_{\text{out}} = V_{\text{in}} - V_F$. These results are not contradictory to any other circuit behaviors so must be correct. The final result of our circuit analysis for all times is found by splicing together the

time intervals for the diode's off and on behaviors as demonstrated in the following code and is plotted in Figure 4.4.2.

```
                # Clear all variables that may be present from
                    previous code run on this page.
globals().clear()

import numpy as np
import matplotlib.pyplot as plt

plt.close('all') # Close all open figures

# Define input parameters
VF = 0.6 # Forward turn-on voltage for diode
A = 5 # volts; amplitude for sinusoidal signal
w = 20 # rad/s; frequency of sinusoidal input voltage
t0 = 0 # initial time
t1 = 1 # final time
nt = 500 # number of points in time array

# Create time and V_in arrays
t = np.linspace(t0,t1,nt);
Vin = A*np.sin(w*t);

# Arrays resulting from analytical work
VD = Vin.copy() # V_D = V_in when the diode is off
VD[VD>VF] = VF # Find all times when VD>VF from the previous
    step. Diode should instead be on.
Vout = Vin - VD # V_out from KVL for all times.

plt.figure(1)
plt.plot(t,Vin, 'b', label='$V_{in}$')
plt.plot(t,Vout,'r-.', label='$V_{out}$')
plt.plot(t,VD,'k--',label='$V_D$')
plt.legend()
plt.xlabel('Time␣(s)')
plt.ylabel('Voltage␣(V)')
plt.show()
```

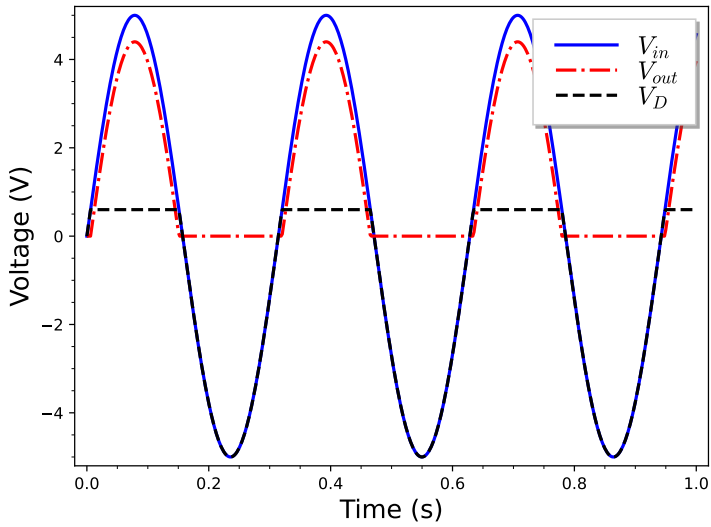


Figure 4.4.2

To summarize, the method of analysis above makes use of the cartoon diode behavior and requires guesses at whether each diode is off or on for a given instantaneous value of the input voltage. The remaining circuit elements are then analyzed under the assumption that the initial guesses are correct. In order to determine the veracity of the initial guesses, one looks for inconsistencies or impossibilities that arise from the rest of the circuit analysis. If an impossible situation occurs, then one knows that the initial assumptions were incorrect. In this scenario, one adjusts the initial guesses and re-analyzes the circuit, repeating until no inconsistencies occur. Stated differently, the steps to follow are:

1. Check the conditions in the circuit and see if it is possible to achieve a forward voltage drop V_F across the diode. If it is completely impossible, then the diode will be off and you can treat it as a circuit break. If it is possible that the diode may be forward biased, then continue on to the next steps.

2. For each diode in your circuit, guess whether the diode is on or off. For each diode that is on, let the voltage drop $(\Delta V)_D = V_F$ be locked. For each diode that is off, treat the diode as a circuit break meaning no current can pass through it.

3. Analyze the rest of the circuit elements and search for inconsistencies.

4. If any results of your analysis conflict with your guesses about diodes that are on/off, then go back and change your guesses until there are no inconsistencies.

Example 4.4.3 Diode example #2.

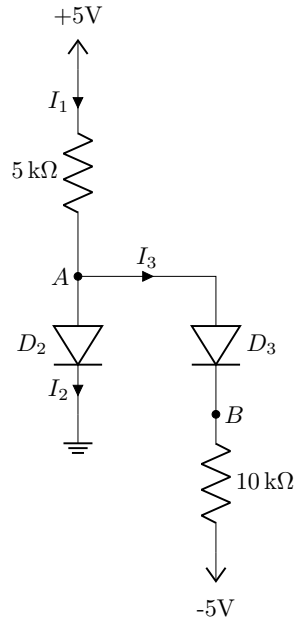


Figure 4.4.4 Diode example #2.

Find the voltage V_B in Figure 4.4.4 assuming $V_F = 0.6$V.

Answer.

$$V_B = 0$$

Solution. Let's examine this case-by-case.

- Case 1: Both diodes are off. Then, $I_2 = I_3 = 0$ so that $I_1 = 0$ as well. So, $V_A = +5$V. This is inconsistent with our assumption, since this results in a bias that is both large enough and in the correct direction to turn on diode $D2$. Thus, the assumptions in this case cannot be true.

- Case 2: Both diodes are on. Then, $V_A = +0.6$V. Using Ohm's law, $I_1 = 0.88$mA. Also, we'd see that $V_B = 0$ due to the diode drop from point A to point B. Given this information, Ohm's law can again be applied to tell us that $I_3 = 0.5$mA. The KCL is then used to determine that $I_2 = I_1 - I_3 = 0.38$mA. There is no inconsistency here, meaning that we were correct in our assumptions and $V_B = 0$. Though it is unnecessary (there can be only one correct answer!), we will examine the remaining cases.

- Case 3: Diode 2 is off while diode 3 is on. In this case, our circuit acts as a voltage divider, with $10.0\text{V} - 0.6\text{V} = 9.4\text{V}$ across the two resistors. So,

$$V_A = -5.0\text{V} + \frac{2}{3}(9.4\text{V}) + 0.6\text{V} = 1.87\text{V}$$

  which is more than enough to turn on diode two. Thus, this case cannot be valid.

- Case 4: Diode 2 is on while diode 3 is off. In this case, $V_A = 0.7\text{V}$. This is identical to V_A in case 2 meaning that diode 3 should activate. This conflicts with our assumption that diode 3 is off. Thus, this case cannot be valid.

▲

Example 4.4.5 Diode example #3.

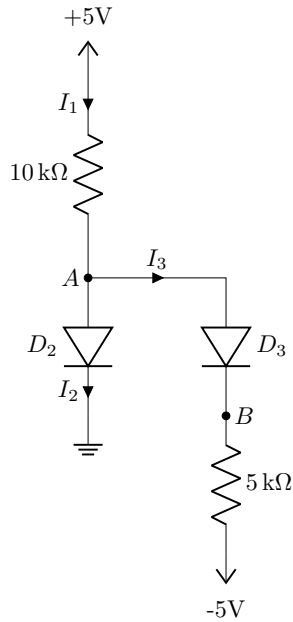


Figure 4.4.6 Diode example #3.

Find the voltage V_B in Figure 4.4.6 assuming $V_F = 0.6\text{V}$. Note that this is identical to the previous problem except that the two resistors have been swapped.

Answer.

$$V_B = -1.87\text{V}$$

Solution. Let's examine this case-by-case.

- Case 1: Both diodes are off. Then, $I_2 = I_3 = 0$ so that $I_1 = 0$ as well. So, $V_A = +5\text{V}$. This is inconsistent with our assumption, since this results in a bias that is both large enough and in the correct direction to turn on diode $D2$. Thus, the assumptions in this case cannot be true.

- Case 2: Both diodes are on. Then, $V_A = +0.6\text{V}$. Using Ohm's law, $I_1 = 0.44\text{mA}$. Also, we'd see that $V_B = 0$ due to the diode drop from point A to point B. Given this information, Ohm's law can again be applied to tell us that $I_3 = 1.0\text{mA}$. The KCL is then used to determine that $I_2 = I_1 - I_3 = -0.56\text{mA}$. This is impossible since we've assumed ideal diodes which do not allow current to pass in the reverse direction.

- Case 3: Diode 2 is off while diode 3 is on. In this case, our circuit acts as a voltage divider, with $10.0\text{V} - 0.6\text{V} = 9.4\text{V}$ across the two resistors. So,

$$V_B = -5.0\text{V} + \frac{1}{3}(9.4\text{V}) + 0.6\text{V} = -1.87\text{V}$$

  and $V_A = V_B + 0.6\text{V} = -1.27\text{V}$ which would not turn on D_2. Thus, everything is consistent and this must be the valid scenario.

- Case 4: Diode 2 is on while diode 3 is off. In this case, $V_A = 0.7\text{V}$. This is identical to V_A in case 2 meaning that diode 3 should activate. This conflicts with our assumption that diode 3 is off. Thus, this case cannot be valid.

▲

4.4.2 Python circuit treatment

Let's revisit the circuit in Figure 4.4.1. This time, we will use an improved mathematical representation of diode behavior from (4.2.1), reproduced here as

$$I_D = I_0 \left(e^{V_D/nV_T} - 1 \right).$$

Assuming we're using a silicon-based diode, we can assume $I_0 \approx 3 \times 10^{12}\text{A}$, $V_T = k_B T/e$, room temperature is $T \approx 293\text{K}$, and $n \approx 1$. The KVL can be used to write

$$\tilde{V}_{\text{in}} - V_D - I_0 R \left(e^{V_D/nV_T} - 1 \right) = 0.$$

This is a nonlinear equation that cannot easily be solved analytically for V_D, so we'll use computational methods to estimate V_D. Specifically, we'll use Newton's Method which is described in Section C.1. In order to use this method, we'll let

$$f(V_D) = \tilde{V}_{\text{in}} - V_D - I_0 R \left(e^{V_D/nV_T} - 1 \right)$$

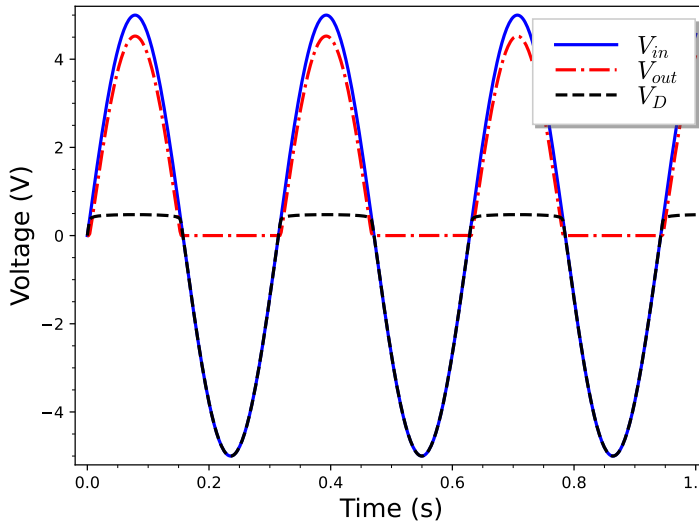
and then calculate the first derivative

$$f'(V_D) = -1 - \frac{I_0 R}{V_T} e^{V_D/V_T}.$$

Using these equations, we implement Newton's Method (Section C.1) using the following code:

```
# Clear all variables that may be present from previous code run
    on this page.
globals().clear()

import numpy as np
import scipy.constants as sc
import matplotlib.pyplot as plt

plt.close('all') # Close all open figures

t = np.linspace(0,1,500);
Vin_array = 5*np.sin(20*t);
Vout = np.zeros(np.size(t));
Varr = np.zeros(np.size(t));

R, T, I0 = [1e4, 293, 3.0e-12] # in Ohms, Kelvin, Amps
VT = sc.k * T / sc.e;

target = 1.0e-8 # User-defined tolerance for final answer

for i1 in range(np.size(t)):
    Vin = Vin_array[i1]

    # Initial guesses for the solution
    V = 3.0

    # Main loop
    error = 1.0
    while error > target:
        f = Vin - V - I0*R*(np.exp(sc.e*V/(sc.k*T))-1);
        fp = -1-I0*R*sc.e/(sc.k*T)*np.exp(sc.e*V/(sc.k*T))
        DeltaV = f/fp
        V -= DeltaV
        error = DeltaV
    Vout[i1]=Vin-V
    Varr[i1]=V

plt.figure(1)
```

```
plt.plot(t,Vin_array, 'b', label='$V_{in}$')
plt.plot(t,Vout,'r-.', label='$V_{out}$')
plt.plot(t,Varr,'k--',label='$V_D$')
plt.legend()
plt.xlabel('Time␣(s)')
plt.ylabel('Voltage␣(V)')
plt.show()
```



Example 4.4.7 Diode Example #4. Examine the circuit in Figure 4.4.8. Plot the time behavior of $I_1(t)$, $I_2(t)$, $I_3(t)$, $V_{\text{in}}(t)$, $V_{\text{out}}(t)$, and the voltage across the diode $V_D(t)$.

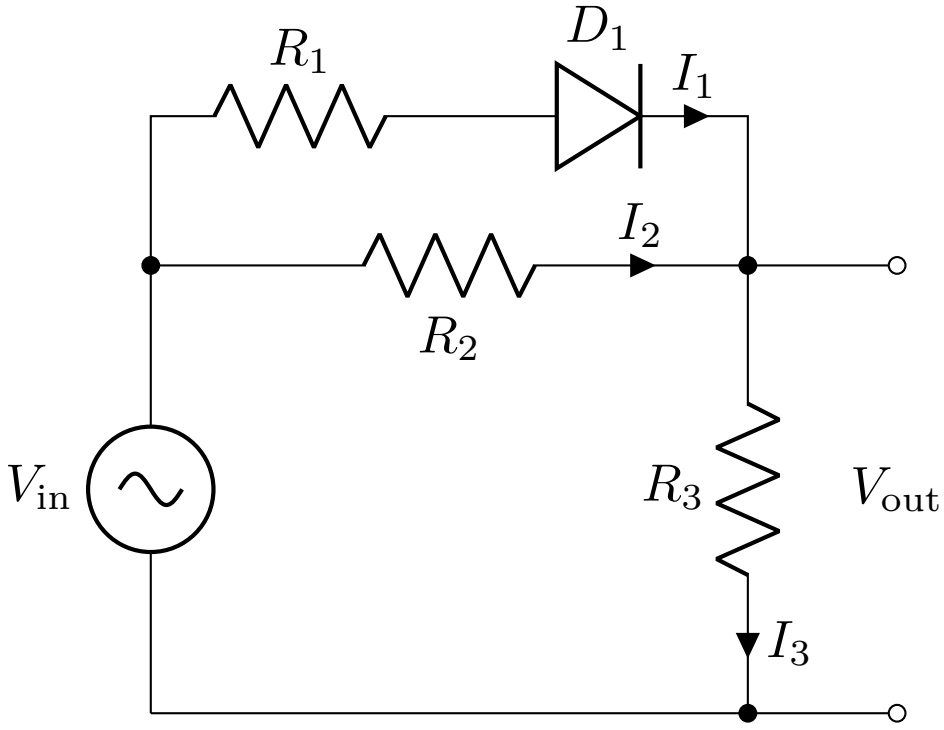


Figure 4.4.8

Solution. We can write down an equation for KCL

$$I_3 = I_1 + I_2.$$

Using Ohm's law,

$$\begin{aligned} I_1 &= I_D \\ I_2 &= \frac{V_{\text{in}} - V_2}{R_2} \\ I_3 &= \frac{V_2}{R_3} \end{aligned}$$

where V_2 is the voltage at the top side of V_{out} (at the junction between R_2, R_3, and D_1). Combining these equations, we find

$$I_D + \frac{V_{\text{in}} - V_2}{R_2} - \frac{V_2}{R_3} = 0.$$

We also get an additional equation by examining the circuit branch containing the diode:

$$V_2 = V_{\text{in}} - I_D R_1 - V_D.$$

Plugging into the previous equation, we thus find that

$$I_D + \frac{I_D R_1 + V_D}{R_2} - \frac{V_{\text{in}} - I_D R_1 - V_D}{R_3} = 0.$$

If we now group all of the I_D terms, we find

$$\left(1 + \frac{R_1}{R_2} + \frac{R_1}{R_3}\right) I_D + \left(\frac{1}{R_2} + \frac{1}{R_3}\right) V_D - \frac{V_{\text{in}}}{R_3} = 0.$$

We can then use Newton's method to solve for the currents letting

$$f(V_D) = \left(1 + \frac{R_1}{R_2} + \frac{R_1}{R_3}\right) I_0 \left(e^{V_D/V_T} - 1\right) + \left(\frac{1}{R_2} + \frac{1}{R_3}\right) V_D - \frac{V_{\text{in}}}{R_3}$$

$$f'(V_D) = \left(1 + \frac{R_1}{R_2} + \frac{R_1}{R_3}\right) \frac{I_0}{V_T} e^{V_D/V_T} + \left(\frac{1}{R_2} + \frac{1}{R_3}\right)$$

```
# Clear all variables that may be present from previous code run
    on this page.
globals().clear()

import numpy as np
import scipy.constants as sc
import matplotlib.pyplot as plt

plt.close('all') # Close all open figures

# Initialize arrays
t = np.linspace(0,1,500)
Vin_array = 5*np.sin(20*t)
Vout = np.zeros(np.size(t))
VD = np.zeros(np.size(t))

I0, R1, R2, R3, T = [1e-13, 1e3, 2e3, 3e3, 293] # Circuit
    parameters
target = 1.0e-8 # Target tolerance for Newton's method

for i1 in range(np.size(t)):
    Vin = Vin_array[i1]

    # Initial guesses for the solution
    V = 1.0

    # Main loop
    error = 1.0
    while error > target:
        f =
            (1+R1/R2+R1/R3)*I0*(np.exp(sc.e*V/(sc.k*T))-1)+V*(1/R2+1/R3)-
        fp =
            (1+R1/R2+R1/R3)*sc.e*I0/(sc.k*T)*np.exp(sc.e*V/(sc.k*T))
            + (1/R2+1/R3)
        DeltaV = f/fp
        V -= DeltaV
        error = DeltaV
    VD[i1]=V

ID = I0*(np.exp(sc.e*VD/(sc.k*T))-1) # Diode current;
    exponential diode equation
Vout = Vin_array - ID*R1 - VD # Kirchhoff loop law

# Plot voltage changes across components
plt.figure(1)
plt.subplot(2,1,1)
plt.plot(t,Vin_array,'b',label='$V_{in}$')
plt.plot(t,Vout,'r-.',label='$V_{out}$')
plt.plot(t,VD,'k--',label='$V_D$')
plt.xlabel('Time␣(s)')
plt.ylabel('Voltage␣(V)')
plt.legend()

# Plot circuit currents
```

4.5 Full-Wave Rectifier

Let's look at a full-wave rectifier circuit pictured in Figure 4.5.1.

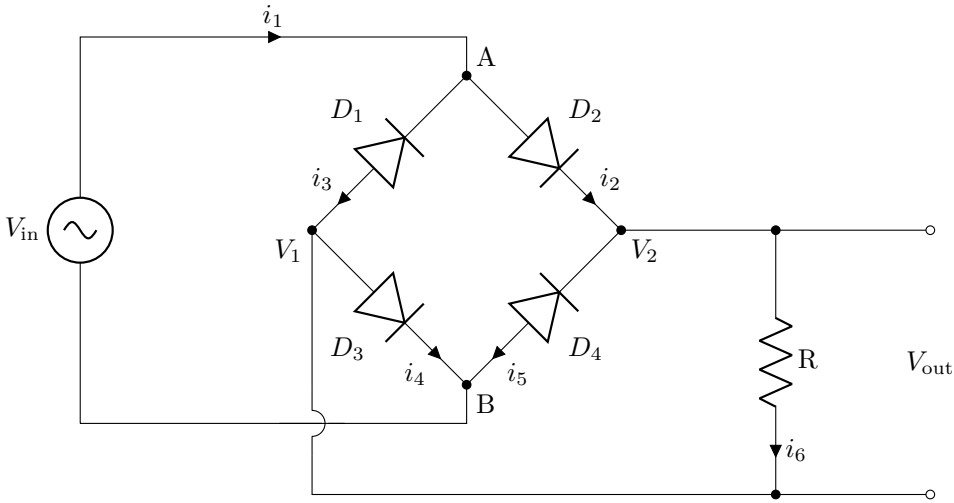


Figure 4.5.1

Let's examine this circuit analytically through several cases.

1. Times when $V_{\text{in}} > 2V_F$:

   In this case, $V_A = V_{\text{in}} > V_1,\ V_2,\ V_B$. This means that diode D_2 will conduct current i_2 while $i_3 = 0$. Since $V_2 > V_B$, diode D_4 will remain off meaning that current will flow across the resistor R, dropping voltage in the process. Then, since $V_A > V_1$, we confirm that diode D_1 remains off. Consequently, diode D_3 will conduct current i_4. Since there was a voltage drop across R and D_3, we find that $V_B < V_2$, confirming our assertion that D_4 is off. This current path is represented in Figure 4.5.2

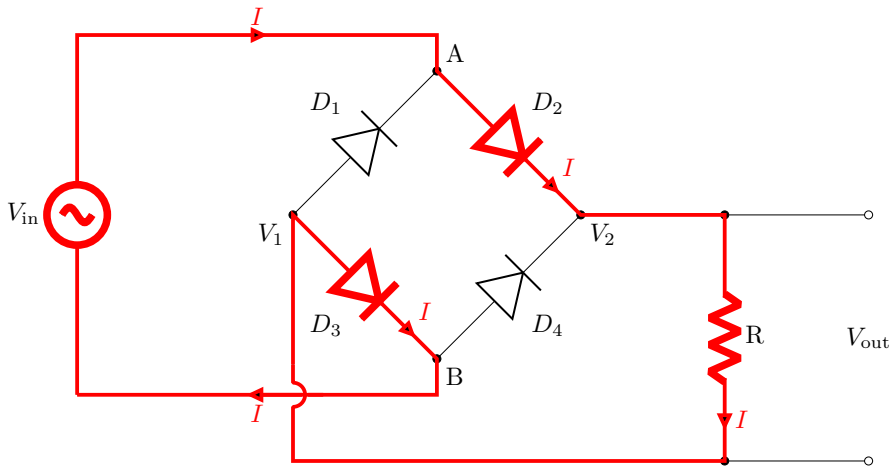


Figure 4.5.2

2. In a similar way, we find that Figure 4.5.3 shows the current path when $V_{\text{in}} < -2V_F$.

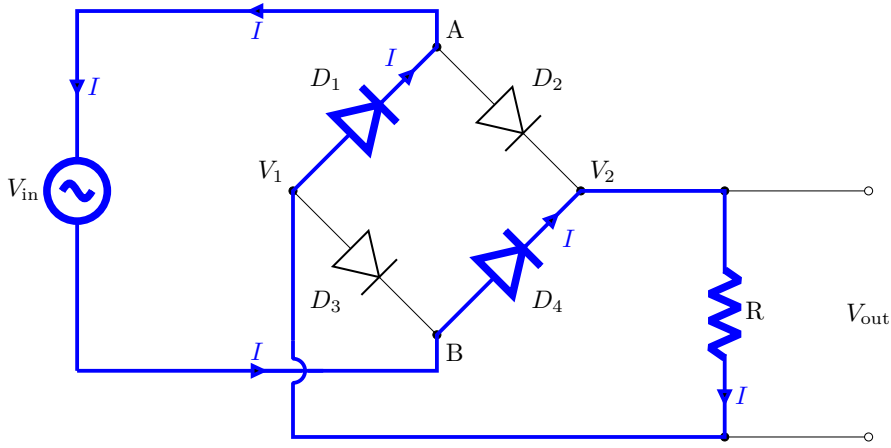
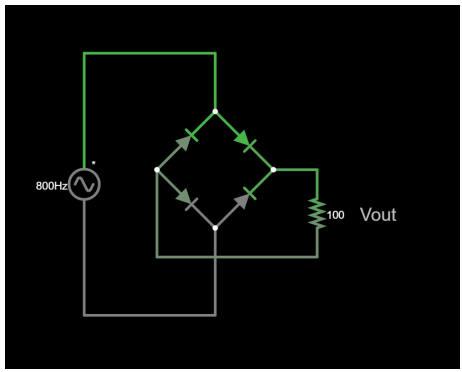


Figure 4.5.3

An animation showing this behavior more comprehensively is shown in Figure 4.5.4.



Standalone
In Context

Figure 4.5.4 Animation showing behavior of the diode full-wave bridge rectifier. Both V_{in} and V_{out} are shown underneath the circuit; place the mouse over the input signal or the resistor to highlight the appropriate curve in the oscilloscope display. Color represents voltage in the schematic, with green representing positive voltage and red representing negative voltage. Moving dots show the current path. Simulation speed can be adjusted by clicking the two vertical lines in the upper right of the figure, revealing sliders.

Example 4.5.5 Solution Using Newton's method in Python. We find KCL equations for the junctions labeled V_1 and V_2

$$0 = I_3 + I_6 - I_4 \tag{4.5.1}$$
$$0 = I_2 - I_5 - I_6. \tag{4.5.2}$$

We can also write down expressions for the currents by examining voltage changes across components

$$I_2 = I_0 \left(e^{(V_{\text{in}} - V_2)/V_T} - 1 \right) \tag{4.5.3}$$
$$I_3 = -I_0 \left(e^{(V_1 - V_{\text{in}})/V_T} - 1 \right) \tag{4.5.4}$$
$$I_4 = I_0 \left(e^{V_1/V_T} - 1 \right) \tag{4.5.5}$$
$$I_5 = -I_0 \left(e^{-V_2/V_T} - 1 \right) \tag{4.5.6}$$
$$I_6 = \frac{V_2 - V_1}{R}. \tag{4.5.7}$$

Combining (4.5.1)-(4.5.2) with (4.5.3)-(4.5.7), we find that

$$0 = -I_0 \left(e^{(V_1 - V_{\text{in}})/V_T} + e^{V_1/V_T} - 2 \right) + \frac{V_2 - V_1}{R}$$
$$0 = I_0 \left(e^{-V_2/V_T} + e^{(V_{\text{in}} - V_2)/V_T} - 2 \right) - \frac{V_2 - V_1}{R}.$$

Using Newton's method in Section C.1, we'll set

$$\vec{f}(\vec{V}) = \begin{pmatrix} -I_0\left(e^{(V_1-V_{\text{in}})/V_T} + e^{V_1/V_T} - 2\right) + \frac{V_2-V_1}{R} \\ I_0\left(e^{-V_2/V_T} + e^{(V_{\text{in}}-V_2)/V_T} - 2\right) - \frac{V_2-V_1}{R} \end{pmatrix}$$

where

$$\vec{V} = \begin{pmatrix} V_1 \\ V_2 \end{pmatrix}.$$

Given $\vec{f}(\vec{V})$, we then calculate

$$\nabla\vec{f}\left(\vec{V}\right) = \begin{pmatrix} -\frac{I_0}{V_T}\left(e^{(V_1-V_{\text{in}})/V_T} + e^{V_1/V_T}\right) - \frac{1}{R} & \frac{1}{R} \\ \frac{1}{R} & -\frac{I_0}{V_T}\left(e^{-V_2/V_T} + e^{(V_{\text{in}}-V_2)/V_T}\right) - \frac{1}{R} \end{pmatrix}.$$

Then, Newton's method can be used in Python to estimate the circuit behavior governed by

$$\nabla\vec{f}\left(\vec{V}\right) \cdot \Delta\vec{V} = \vec{f}\left(\vec{V}\right)$$

to solve for $\Delta\vec{x}$ and find our new root estimate

$$\vec{x}' = \vec{x} - \Delta\vec{x},$$

iterating to some desired tolerance.

```
# Clear all variables that may be present from previous code run
    on this page.
globals().clear()

import numpy as np
from scipy.constants import k
import matplotlib.pyplot as plt

plt.close('all') # Close all open figures

# Initialize arrays
t = np.linspace(0,1,500)
Vin_array = 10*np.sin(20*t)
Vout = np.zeros(np.size(t))

# Circuit parameters and target precision for roots
R, VT, I0 = [100, 0.05, 3.0e-9]
target = 1.0e-2

# Perform Newton's Method for every time value
for i1 in range(np.size(t)):

    Vin = Vin_array[i1]

    # Initial guesses for the solution
    V1 = 0.4
    V2 = abs(Vin)

    # Main loop
    error = 1.0
    #for i1 in range(100):
    while error > target: # perform iterations until our
        estimated root gets close enough to real root
        # c1 and c2 are defined to make entry of the f and A
            arrays easier
        c1 = -I0*(np.exp((V1-Vin)/VT)+np.exp(V1/VT))
        c2 = I0*(np.exp(-V2/VT)+np.exp((Vin-V2)/VT))

        f = np.array([c1+2*I0+(V2-V1)/R, c2-2*I0-(V2-V1)/R]) # f
            from system of equations
        A = np.array([[c1/VT-1/R, 1/R],[1/R, -c2/VT-1/R]]) # f'
            for system of equations
        DeltaV = np.linalg.solve(A,f) # distance from our guess
            to our estimated root
        error = np.linalg.norm(DeltaV)
        V1 -= DeltaV[0]
        V2 -= DeltaV[1]

    Vout[i1]=V2-V1
plt.plot(t,Vin_array,label='$V_{in}$')
plt.plot(t,Vout,label='$V_{out}$')
plt.xlabel('Time␣(s)')
plt.ylabel('Voltage␣(V)')
plt.legend()
```

▲

4.6 Useful Diode Circuits

Let's take our full-wave rectifier circuit and add a capacitor across the output as in Figure 4.6.1.

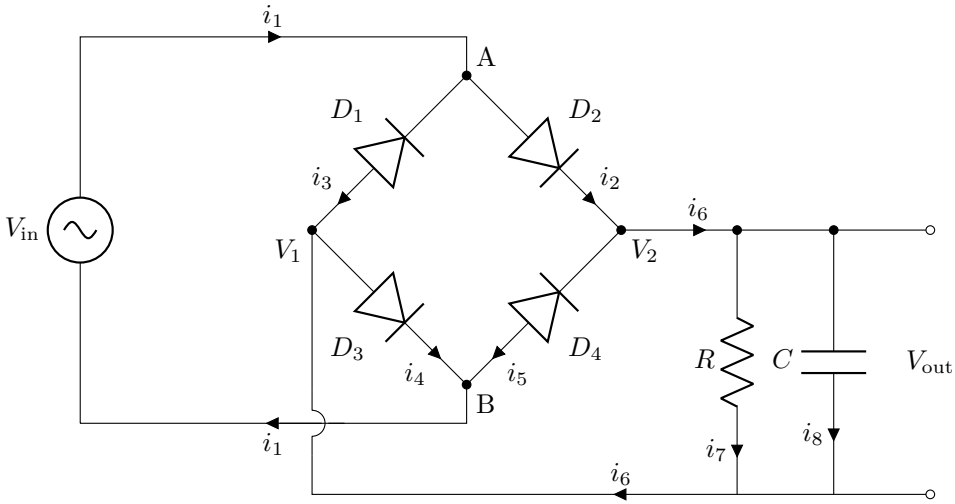


Figure 4.6.1

Chapter 5

Transistors

In this chapter, we move away from **passive electronic components** and transition to **active electronic components**. Unlike passive components, active electronic components require an external power source as part of its operation. This external power source allows active components to amplify voltage and current, allowing for signal gains greater than one and resulting in power amplification.

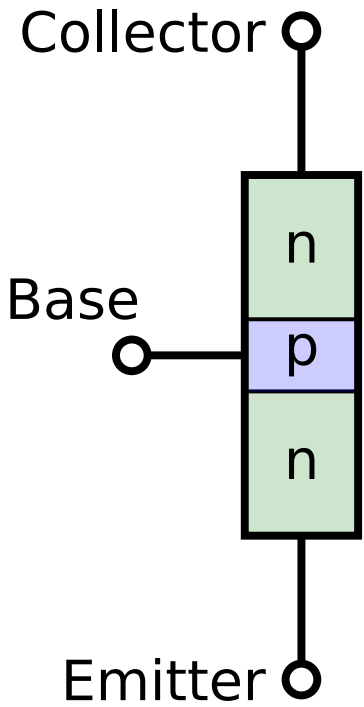
The first active component that we will examine is the transistor. Invented in 1947, transistors are arguably one of the most important inventions of the 20th century. Transistors provide better reliability, efficiency, and speed than vacuum tubes while having a much smaller size. Since 1970, the size of transistors has reduced from from 10^4nm to approximately 3nm (smaller than a human DNA strand).[1] This dramatic reduction in transistor size has been mirrored by a similar reduction in manufacturing cost, two of the main factors behind Moore's law [CITE AND DEFINE] and the greatly-increased prevalance of small but powerful consumer electronics (such as smartphones, smart watches, and notebook computers).

There are two major classifications for transistors: the Bipolar Junction Transistor (BJT) and the Field-Effect Transistor (FET). We will direct our focus to BJTs in this text, using them to illustrate the basics of transistor behavior and operation. With the background provided here, one will be prepared to learn about FETs from other texts if desired.

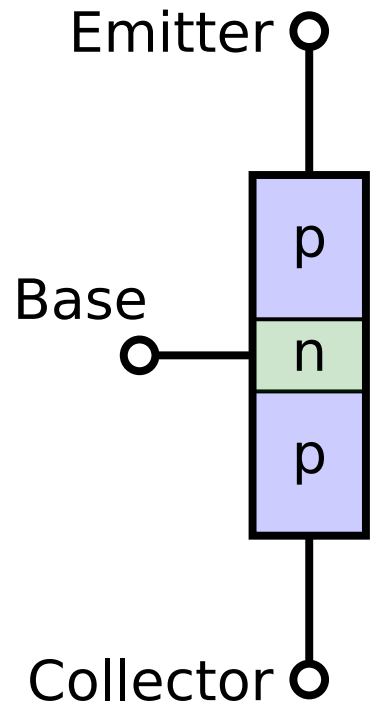
5.1 Introduction to transistors

There are two types of BJTs, the npn transistor and the pnp transistor, pictured in Figure 5.1.1.

[1] https://spectrum.ieee.org/transistor-density, https://www.kla.com/wp-content/uploads/Transistor-75-year-timeline_Final.pdf



(a) npn transistor.



(b) pnp transistor.

Figure 5.1.1

5.2 Transistor behavior

5.3 Transistor switching

Chapter 6

Operational Amplifiers

In the last chapter, we examined single-transistor circuits that provide voltage amplification, current amplification, switching circuits, and signal buffering. These behaviors can be improved even further through the design of more complex circuits using multiple transistors, resistors, capacitors, and other components. Given the small size that can be achieved in transistor construction, these complex multi-transistor circuits can be built on a single small silicon chip. The ability to create silicon chips containing specialized multi-transistor circuit revolutionized electronics development.

The operational amplifier (or op-amp) is one example of an **integrated circuit** in which a single small chip component contains many discrete circuit elements (such as transistors, resistors, capacitors, and diodes) connected together into a circuit. Thus, an integrated circuit chip will be much smaller than an equivalent circuit that is built with discrete components, and will often perform much better since the chip's internal components are precisely designed to work in concert with each other.

We will start this chapter by introducing op-amps in more detail along with the rules governing op-amp behavior. We will then examine several useful circuits that feature op-amps.

6.1 The Op-amp Chip

The operational amplifier is an integrated circuit that has been designed to act as a voltage amplifier. The 741 op-amp was one of the first low-cost, high-performance, general-purpose operational amplifiers. Figure 6.1.1 shows a circuit schematic for the 741 op-amp, revealing 24 transistors, 11 resistors, and one capacitor.

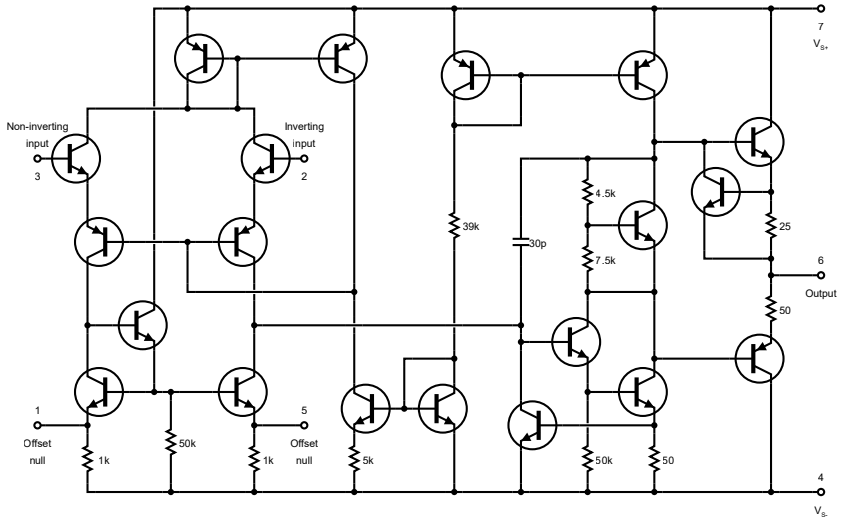


Figure 6.1.1 A schematic diagram showing the voltage amplifier circuit design for the general-purpose 741 op-amp. (Source: User Omegatron, 741 op-amp schematic, October 2011, Wikimedia Commons. Public Domain.)

The smallest 741 op-amp chips are manufactured using the SOIC-8 package which measures approximately 4mm×5mm×1.75mm, though other more modern models can be as small as 0.88mm×0.58mm×0.33mm.[TEMP LINK] A look at the internal layout of a real physical 741 op-amp can be seen in this linked blog post, which also provides an interactive chip viewer that allows the reader to view the op-amp chip's physical layout and click on physical components to have the corresponding element in the circuit schematic highlighted. This image shows that the sole capacitor occupies a rather large percentage of the chip's area.

In circuit schematics, the ideal op-amp is represented by the symbol shown in Figure 6.1.2, where V_- and V_+ are called the **inverting input** and **non-inverting input** respectively. The amplifier output is given as V_{out}. Op-amps are active circuit components and require external voltage sources for operation, just as transistors require V_{CC} and V_{EE} as supply voltages. These two supply voltages are marked as V_{s+} and V_{s-}, though these connections are sometimes left out in circuit diagrams.

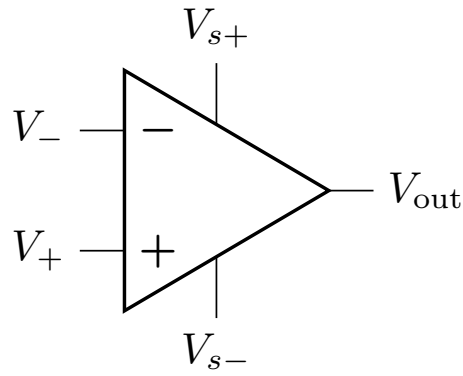
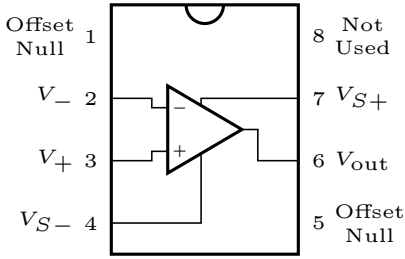
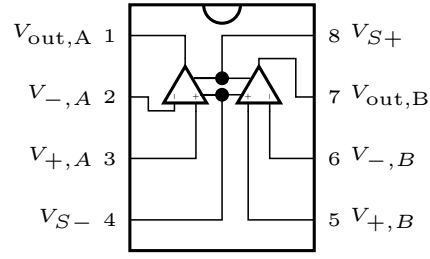


Figure 6.1.2 Schematic symbol for an ideal op-amp. The inverting and non-inverting inputs are specified as V_- and V_+. The two voltages V_{s+} and V_{s-} are supplied by an external power source.

Figure 6.1.3 shows a top-view pin diagram for two general-purpose op-amps. The half-moon shape allows one to identify pin 1, with remaining pins being counted in a counter-clockwise manner.



(a) Pin diagram, LM741 (PDIP package)



(b) Pin diagram, LM358 (PDIP package)

Figure 6.1.3 Sample top-view pin diagrams for two general-purpose op-amp chips, the LM741 and the LM358. This particular pin diagram is specific to the general purpose LM741 op-amp chip. Other op-amp chips may have different pin diagrams.

6.2 Ideal Op-amp Behavior

Previously, we have seen that transistors are very useful in buffering AC signals. An emitter-follower circuit placed between an AC voltage source and a load circuit increases the input impedance of the load, effectively lightening the load and making it easier for the AC voltage source to drive the load circuit. While the common-emitter amplifier circuit added the additional benefit of AC voltage gain, it did so

by weakening its buffering capabilities. More than that, increases in AC voltage gain also lead to degradation of bandwidth performance and circuit stability.

The operational amplifier has been designed to serve as a differential voltage amplifier, but has been designed to do so while maintaining and vastly improving upon signal buffering, amplification, stability, and bandwidth behaviors when compared to the transistor circuits that we have studied. The behavior of ideal op-amps are governed by four rules:

1. The output voltage reflects the difference between the noninverting and inverting inputs multiplied by a constant A_o so that $v_{\text{out}} = A_0 (V_+ - V_-)$.

2. The **open-loop gain** is assumed to be $A_o = \infty$. The open-loop gain represents the op-amp's built-in voltage gain when no feedback is present. We will discuss positive and negative feedback in the coming sections.

3. Input impedance is assumed to be very high, $Z_{\text{in}} = \infty$.

4. The output impedance is assumed to be very low, $Z_{\text{out}} = 0$.

The symbols above have been previously defined in Figure 6.1.2. These ideal op-amp rules often work quite well in describing the behavior of real op-amps, though real op-amps have open-loop gain $A_o = 10^4 - 10^6$, input impedance $Z_{\text{in}} = 10^6\Omega - 10^{12}\Omega$, and open-loop output impedance $Z_{\text{out}} = 10\Omega - 100\Omega$, though Z_{out} becomes much smaller (often less than one ohm) when negative feedback (discussed later) is employed. The op-amp output voltage V_{out} is constrained to values between the two supply voltages V_{S+} and V_{S-}.

On its own, an op-amp can be used for several purposes. We've already seen that an op-amp without feedback converts time-varying signals into square waves. This behavior can also be used as a comparator, where the output voltage can provide a very rapid comparison between input signals, with V_{out} saturating at either the upper or lower supply voltage depending on whether V_+ is greater than or less than V_-. Op-amps can also be used in switching applications in which one op-amp input is set at a threshold value and then the op-amp output value depends sensitively on whether the other op-amp input exceeds this threshold setting.

Like transistors, op-amp behaviors can be modified based on how they are incorporated into circuits containing other discrete circuit components. Before we discuss these more advanced applications, let's apply these four ideal op-amp rules to the examples below.

Example 6.2.1 Non-inverting open-loop amplifier. Determine V_{out} for the circuit pictured in Figure 6.2.2, assuming V_{in} is a sinusoidal signal.

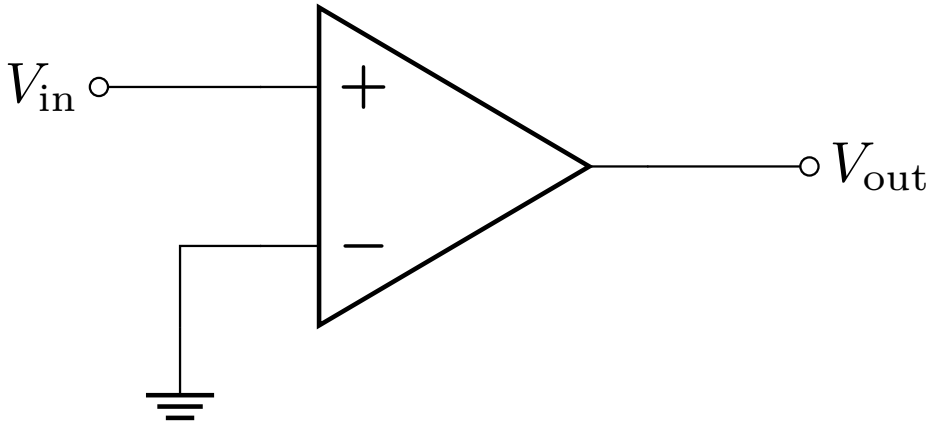


Figure 6.2.2 Non-inverting open-loop amplifier.

Solution. Using the ideal op-amp rules, $V_{\text{out}} = A_0 (V_+ - V_-) = A_0 V_{\text{in}}$ since $V_+ = V_{\text{in}}$ and $V_- = 0$. Therefore, the voltage gain is $G_V = A_0 = \infty$. Since the gain is considered to be infinite, V_{out} will be a square wave with the same phase and frequency as V_{in} and an amplitude limited by the op-amp supply voltages V_{S+}, V_{S-}. The results are plotted in Figure 6.2.3.

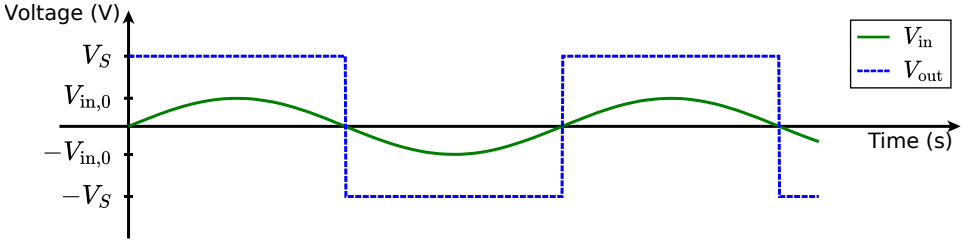


Figure 6.2.3 Input and output voltages for the non-inverting open-loop amplifier.

▲

Example 6.2.4 Inverting open-loop amplifier. Determine V_{out} for the circuit pictured in Figure 6.2.5 assuming V_{in} is a sinusoidal signal.

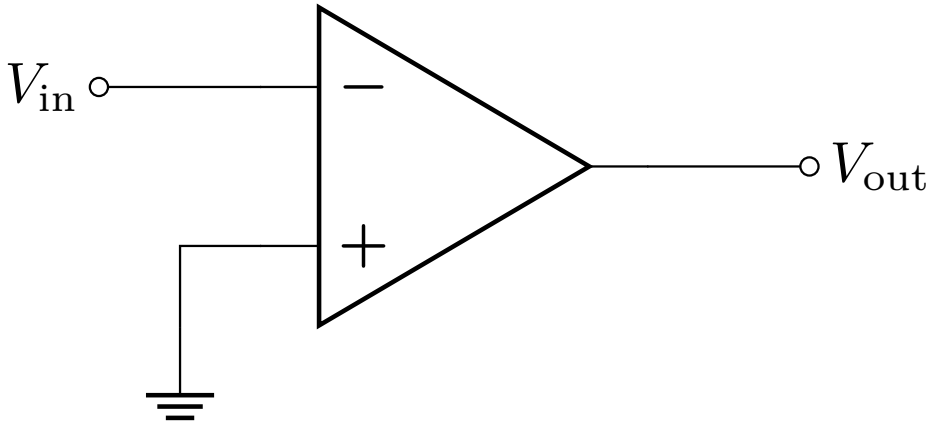


Figure 6.2.5 Non-inverting open-loop amplifier.

Solution. Using the ideal op-amp rules, $V_{\text{out}} = A_0 (V_+ - V_-) = -A_0 V_{\text{in}}$ since $V_- = V_{\text{in}}$ and $V_+ = 0$. Therefore, the voltage gain is $G_V = -A_0 = -\infty$. Since the gain is considered to be infinite, V_{out} will be a square wave with the same phase and frequency as V_{in} and an amplitude limited by the op-amp supply voltages V_{S+}, V_{S-}. The results are plotted in Figure 6.2.6.

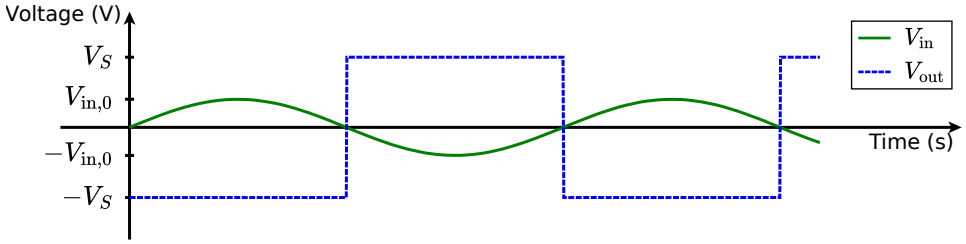


Figure 6.2.6 Input and output voltages for the inverting open-loop amplifier.

▲

6.3 Negative Feedback

In the previous section, we saw that operational amplifiers are quite effective at taking sinusoidal V_{in} signals (or any time varying signal) and producing a square wave output according to $V_{\text{out}} = \text{sign}(V_{\text{in}})$. Based on these examples alone, it may seem like op-amps have limited usefulness.

Op-amps are frequently used with **negative feedback** in which a conducting path is placed between V_{out} and the inverting input V_- as shown in Figure 6.3.1.

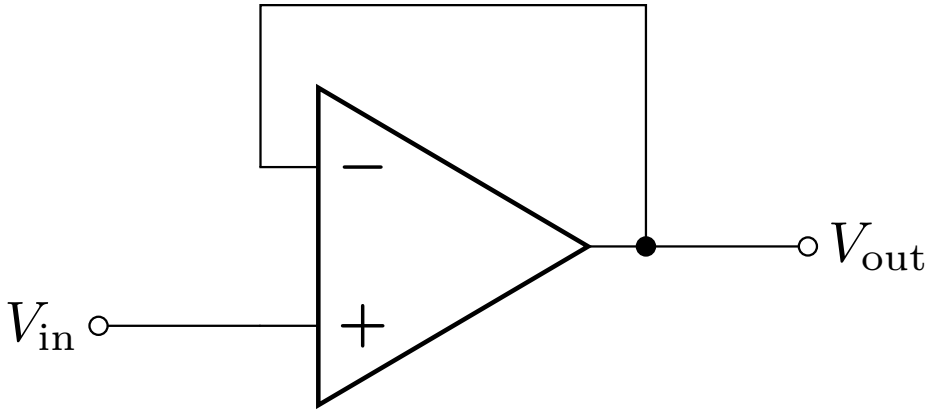


Figure 6.3.1 A circuit employing negative feedback on an op-amp has a conducting path placed that connects V_{out} and the inverting input V_-.

We gain the following two **golden rules of op-amps** that must be satisfied when analyzing circuits containing op-amps with negative feedback:

1. The output of the op-amp will do whatever it needs to do to ensure that $V_+ = V_-$.

2. No current flows into or out of the op-amp inputs V_+ and V_-, though there may be a lot of current that can flow to or from the op-amp output.

The first golden rule results from the behavior that we examined in the previous section and can be illustrated by looking at three cases, and we'll assume that $|V_+| < |V_s|$ and $|V_-| < |V_s|$. In case 1, $V_+ > V_-$, resulting in $V_{\text{out}} = +V_S$. Negative feedback would then cause $V_- = +V_s > V_+$, switching us from case 1 to case 2. In case 2, $V_+ < V_-$, resulting in $v_{\text{out}} = -V_s$. Negative feedback would then cause $V_- = -V_s < V_+$, returning us to case 1. Instead of an unstable output where V_{out} oscillates uncontrollably between $\pm V_s$, the op-amp instead settles into case 3 where $V_+ = V_-$. In reality, there must be a very small difference between the two op-amp inputs, otherwise $V_{\text{out}} = 0$ would result in case 3. The voltage difference between op-amp inputs in this case will typically be on the order of microvolts and thus will be ignored.

Since $V_+ = V_-$ for an op-amp with negative feedback, the circuit in Figure 6.3.1 turns out to be a buffer amplifier with voltage gain $G_V = 1$ since we must have $V_{\text{out}} = V_{\text{in}}$. This means that this op-amp circuit provides a benefit similar to that provided by the BJT emitter-follower circuit, except better. There is no voltage drop between V_{in} at the npn base and V_{out} at the npn emitter. Additionally, the input and output impedances of this op-amp circuit are much enhanced over the emitter-follower circuit.

Further examples with negative feedback. Op-amp circuits with negative feedback can provide voltage gains other than one and infinity through the placement of additional resistors. Let's analyze the **inverting amplifier** circuit in Figure 6.3.2 to see how this works.

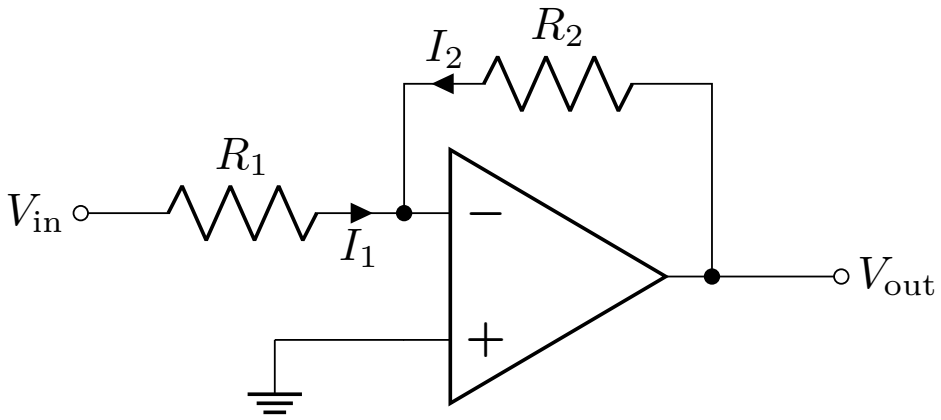


Figure 6.3.2 This circuit will use negative feedback to produce a voltage gain $G_V = -R_2/R_1$.

Applying the first golden rule, we find that $V_- = V_+ = 0$, meaning V_- is a **virtual ground** since the op-amp rules require $V_- = 0$ Volts but V_- is not itself connected to the real circuit ground. Next, recognize that

$$I_1 = \frac{V_{\text{in}} - V_-}{R_1} = \frac{V_{\text{in}}}{R_1}$$

and

$$I_2 = \frac{V_{\text{out}} - V_-}{R_2} = \frac{V_{\text{out}}}{R_2}.$$

Since the golden rules require that no current can enter or leave an op-amp input, Kirchhoff's junction law requires that $I_1 + I_2 = 0$ or $I_2 = -I_1$. Plugging in our expressions for the currents, we find that

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

which results in a voltage gain

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

Like the common-emitter amplifier, this circuit applies a negative gain to our input signal, and the magnitude of the gain is dependent on the chosen values of external

resistors. This op-amp inverting amplifier does not suffer from the same stability issues as the common-emitter amplifier, and this op-amp circuit maintains much higher input impedance and much lower output impedance than seen with the common-emitter amplifier.

Example 6.3.3 Non-inverting amplifier. Determine G_V for the circuit pictured in Figure 6.3.4.

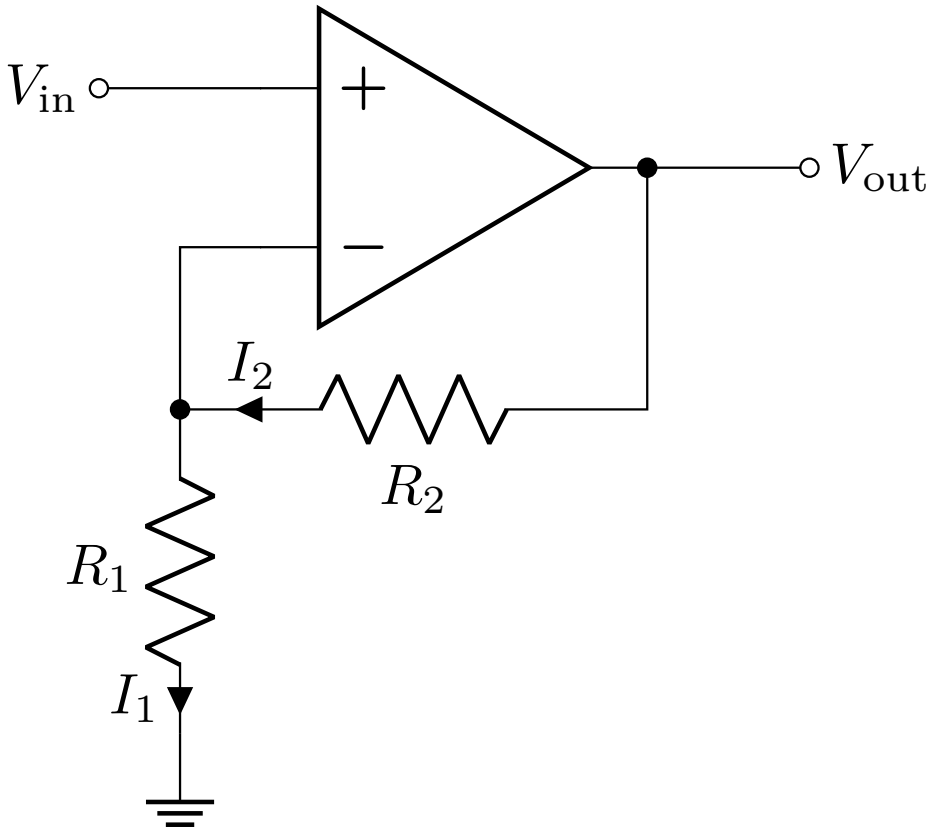


Figure 6.3.4 Non-inverting amplifier.

Solution. As before, the golden rules require that $V_+ = V_-$. This time, instead of a virtual ground we find that $V_- = V_{\text{in}}$. Since no current can enter/leave the op-amp inputs, we must have $I_1 = I_2$. Using Ohm's law,

$$I_1 = V_{\text{in}}/R_1$$

so that

$$V_{\text{out}} - V_{\text{in}} = I_1 R_2.$$

Combining these results,

$$V_{\text{out}} = V_{\text{in}} + \frac{R_2}{R_1} V_{\text{in}} = \left(1 + \frac{R_2}{R_1}\right) V_{\text{in}}.$$

Thus, the voltage gain for this circuit is given by

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

▲

Example 6.3.5 Weighted summing amplifier. Determine V_{out} for the circuit pictured in Figure 6.3.6.

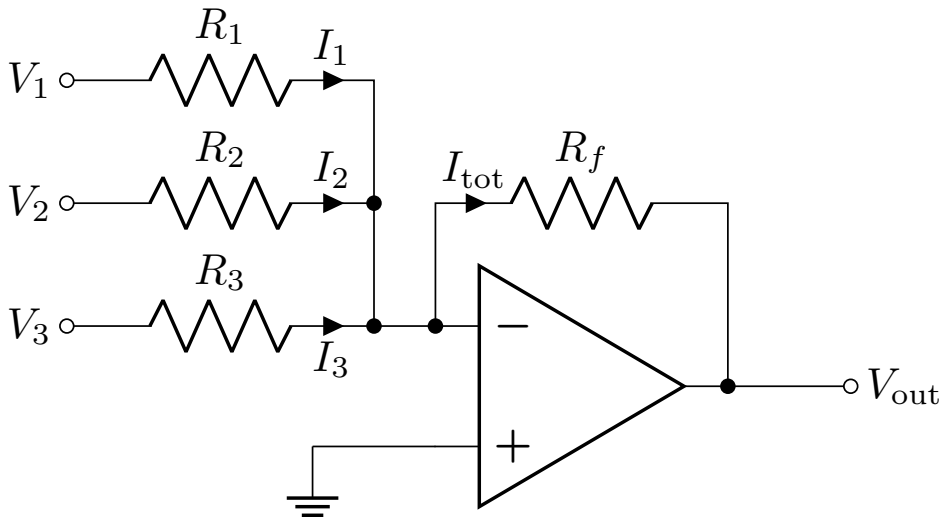


Figure 6.3.6 Weighted summing amplifier.

Solution. The golden rules require that $V_+ = V_-$, meaning that V_+ is grounded and V_- is a virtual ground. We can then use Ohm's law three times to find

$$I_1 = \frac{V_1}{R_1}, \qquad I_2 = \frac{V_2}{R_2}, \qquad I_3 = \frac{V_3}{R_3}.$$

When Kirchhoff's junction law is applied at the V_- junction, we find

$$\begin{aligned} I_{\text{tot}} &= I_1 + I_2 + I_3 \\ &= \frac{V_1}{R_1} + \frac{V_2}{R_2} + \frac{V_3}{R_3}. \end{aligned}$$

Since $V_- = 0$, one further application of Ohm's law gives

$$\begin{aligned} V_{\text{out}} - \cancelto{0}{V_-} &= -I_{\text{tot}} R_f \\ &= -R_f \left(\frac{V_1}{R_1} + \frac{V_2}{R_2} + \frac{V_3}{R_3} \right) \end{aligned}$$

based on the chosen I_{tot} direction in Figure 6.3.6. In the special case where $R_1 = R_2 = R_3$, this results in

$$V_{\text{out}} = -\frac{R_f}{R_1} \left(V_1 + V_2 + V_3 \right).$$

▲

Example 6.3.7 Differentiator. Determine the behavior of V_{out} for the circuit pictured in Figure 6.3.8.

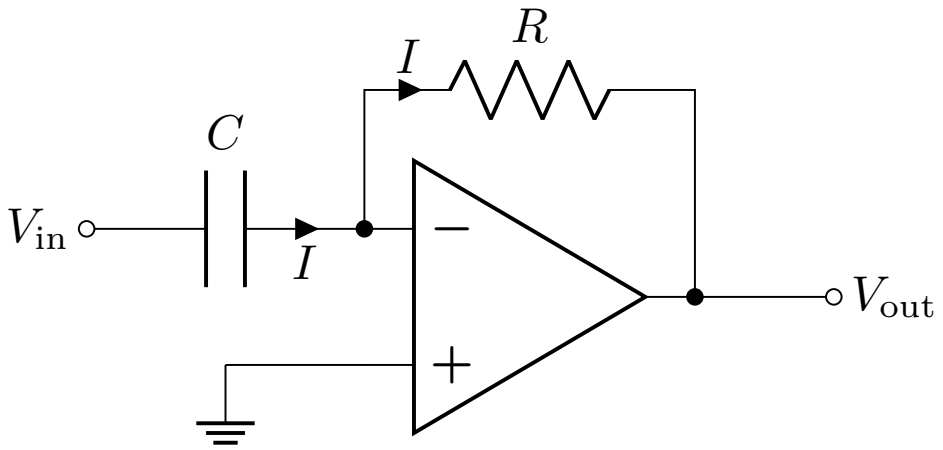


Figure 6.3.8 Differentiator.

Solution. We find that V_- is a virtual ground since V_+ is connected to ground. Also, the currents through the capacitor and resistor are identical since no current can enter/leave the op-amp inputs.

We have seen in introductory physics that $\Delta V_C = Q/C$ where $\Delta V_C = V_{\text{in}} - V_-$. Therefore, $V_{\text{in}} = Q/C$. We can use Ohm's law to show $0 - V_{\text{out}} = IR$. In introductory physics, we also learned that $I = \mathrm{d}Q/\mathrm{d}t$. Combining these expressions, we find that

$$I = C \frac{\mathrm{d}V_{\text{in}}}{\mathrm{d}t}$$

and thus

$$V_{\text{out}} = -RC\frac{\mathrm{d}V_{\text{in}}}{\mathrm{d}t}.$$

This result reveals this circuit's ability to produce an output voltage V_{out} that is a constant multiplying the time derivative of V_{in}. ▲

Frequency response of negative feedback amplifiers. ADD DISCUSSION OF THE FREQUENCY RESPONSE OF NEGATIVE FEEDBACK AMPLIFIERS AND DISCUSS THE EQUATION GOVERNING THE PLOT IN Figure 6.3.9.

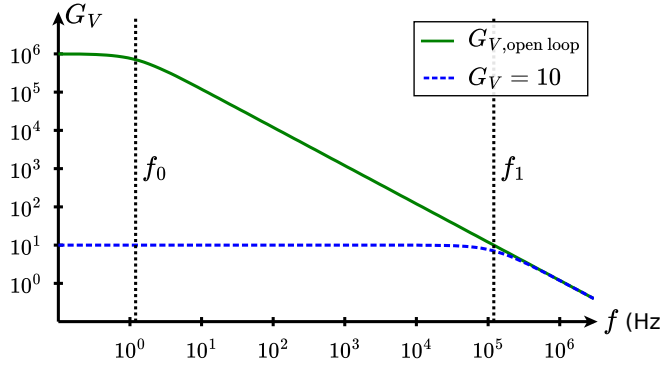


Figure 6.3.9 The frequency response of amplifiers using negative feedback depends on the circuit gain selected through design. As gain drops, bandwidth increases.

6.4 Positive Feedback

The next external modification we'll examine is the application of **positive feedback** to op-amps. When positive feedback is applied to an op-amp, a conducting path is placed between V_{out} and the non-inverting input V_+ as shown in Figure 6.4.1.

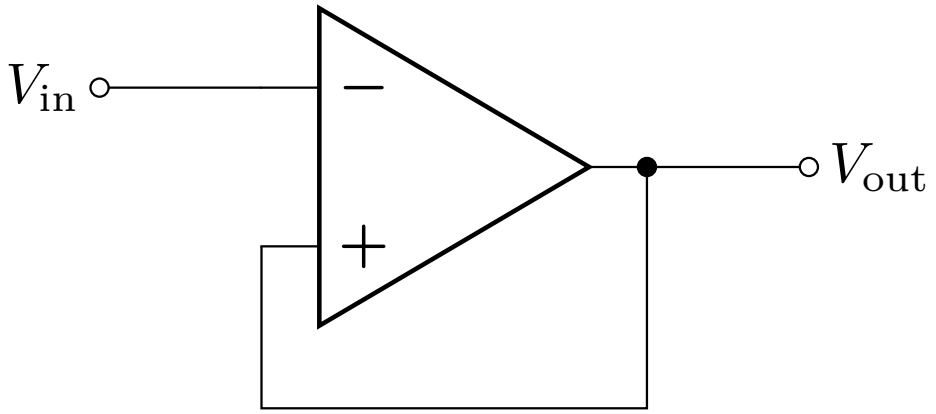


Figure 6.4.1 A circuit employing positive feedback on an op-amp has a conducting path placed that connects V_{out} and the inverting input V_{-}.

In this configuration, any difference between the input voltages causes V_{out} to 'latch', at either the positive or negative output state . This is an example of **hysteresis**, meaning that the circuit responds to some change at the inputs but then fails to return to the original output state when the inputs return to their original states.

We can modify the above hysteresis behavior by introducing resistors into our positive-feedback circuit in a configuration called a **Schmitt trigger**. The Schmitt trigger circuit is shown in Figure 6.4.2.

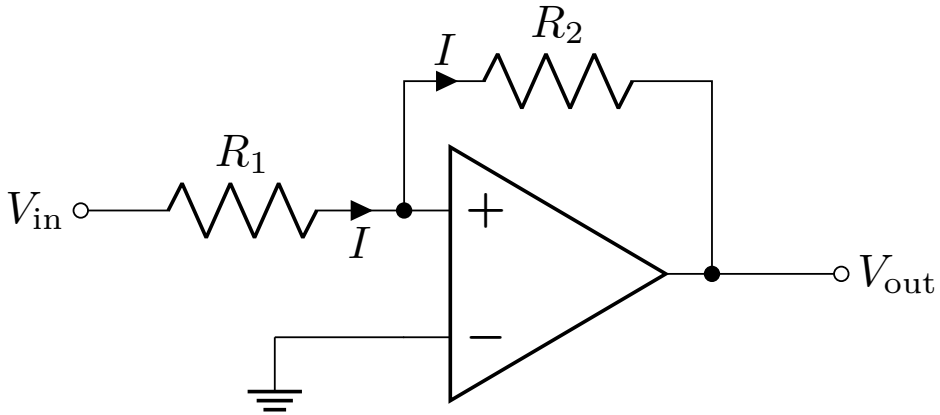


Figure 6.4.2 A Schmitt trigger circuit, employing an op-amp with positive feedback.

In this circuit, V_{-} is connected to ground resulting in a virtual ground at V_{+}. Since current cannot flow into or out of the op-amp inputs, the same current I must

be passing through each of R_1 and R_2. An application of Ohm's law across the two series resistors gives

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

Since op-amp behavior will cause V_{out} to depend directly on V_+, let's find an expression for this parameter. Using Ohm's law on R_1, we find that

$$\begin{aligned} V_+ &= V_{\text{in}} - IR_1 \\ &= V_{\text{in}} - \frac{V_{\text{in}} - V_{\text{out}}}{R_1 + R_2} \\ &= \frac{R_2}{R_1 + R_2} V_{\text{in}} + \frac{R_1}{R_1 + R_2} V_{\text{out}}. \end{aligned}$$

Now, the state of V_{out} will be $+V_s$ or $-V_s$ depending on whether $V_+ > 0$ or $V_+ < 0$ with the transition occurring at any crossing of $V_+ = 0$. We can use our V_+ expression above to re-express in terms of V_{in} the condition for changing the state of V_{out}. After a little algebra, we find that this new condition is given by

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

So, if $V_{\text{out}} = +V_S$ initially, then the output state will only change once V_{in} crosses the threshold voltage $-\frac{R_1}{R_2} V_S = -V_T$. Likewise, if in some second time period we find $V_{\text{out}} = -V_S$ initially, the output state will only change once V_{in} crosses the threshold voltage $+\frac{R_1}{R_2} V_S = +V_T$. Combined, V_{out} changes from $\mp V_S \to \pm V_S$ when $V_{\text{in}} = \pm V_T$

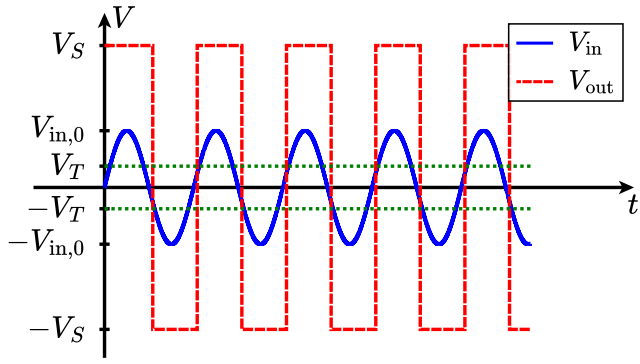


Figure 6.4.3 This plot shows the response of V_{out} based on when V_{in} satisfies the necessary condition related to the threshold voltages $\pm V_T$.

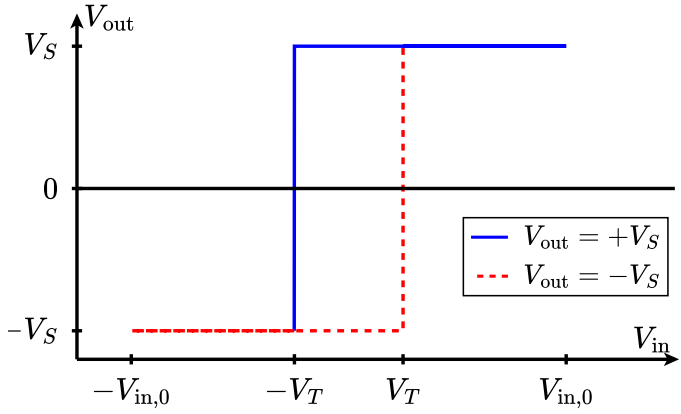


Figure 6.4.4 The hysteresis associated with the Schmitt trigger is demonstrated more clearly in a plot of V_{out} vs V_{in}. The value of V_{out} in the region $-V_T < V_{\text{in}} < +V_T$ is dependent on the output state at the previous time point.

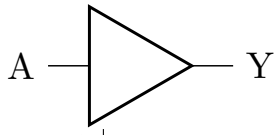
6.5 More Op-amp Circuit Examples

Chapter 7

Digital Electronics

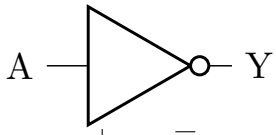
7.1 Digital gates and truth tables

Digital circuits are built using a variety of **gates** that accept digital inputs with values of 0 or 1 and produce a digital output based on a specific gate's operation. We characterize the behavior of digital gates using **truth tables**. Figure 7.1.1 shows the digital gates that will be the building blocks for our digital circuits and the truth tables that describe their behaviors.



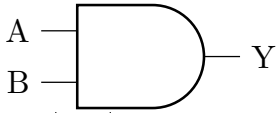
A	$Y = A$
0	0
1	1

(a) Buffer gate



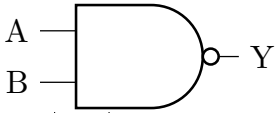
A	$Y = \overline{A}$
0	1
1	0

(b) NOT gate



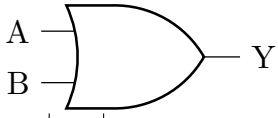
A	B	$Y = A \cdot B$
0	0	0
0	1	0
1	0	0
1	1	1

(c) AND gate



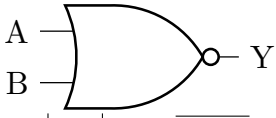
A	B	$Y = \overline{A \cdot B}$
0	0	1
0	1	1
1	0	1
1	1	0

(d) NAND gate



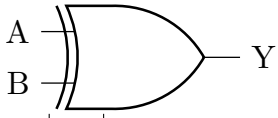
A	B	$Y = A + B$
0	0	0
0	1	1
1	0	1
1	1	1

(e) OR gate



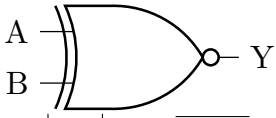
A	B	$Y = \overline{A + B}$
0	0	1
0	1	0
1	0	0
1	1	0

(f) NOR gate



A	B	$Y = A \oplus B$
0	0	0
0	1	1
1	0	1
1	1	0

(g) XOR gate

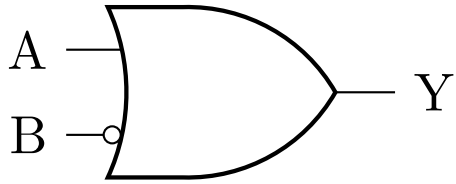


A	B	$Y = \overline{A \oplus B}$
0	0	1
0	1	0
1	0	0
1	1	1

(h) XNOR gate

Figure 7.1.1

Take note of the symbols used to represent logic functions in our expressions for output Y above: the dot '$\cdot$' represents AND, '$+$' represents OR, '$\oplus$' represents XOR, and a bar over a symbol represents NOT. These symbols will be used when we discuss **Boolean algebra** in the next section. Also note how a circle on a gate output on a circuit symbol above results in the inversion of the original gate output. This circle symbol can be thought of as shorthand for the placement of a NOT gate, and this functionality can be used on any gate port as seen in Figure 7.1.2.



A	B	$\overline{B}$	$Y = A + \overline{B}$
0	0	1	1
0	1	0	0
1	0	1	1
1	1	0	1

Figure 7.1.2 A demonstration showing use of a circle symbol in place of a NOT gate for an OR gate input. This gate operates as a standard OR gate for inputs A and $\overline{B}$.

Two of the gate types shown in Figure 7.1.1 hold special prominence. The NAND and NOR gates are known as **universal gates**, which means that the behavior of any logic function (Buffer, NOT, AND, OR, XOR, NAND, NOR, or XNOR) can be reproduced using a combination of several universal gates of a single type. For example, XNOR functionality can be reproduced solely using NAND gates that are connected together properly. This property causes universal gates to be very popular in circuit design as it allows manufacturers to minimize the number of gate types used in a circuit which can often result in a reduction in manufacturing costs.

Example 7.1.3 Creating a NOT gate using only NAND gates. Use truth tables to show that NOT functionality can be replicated by the circuit shown in Figure 7.1.4 which uses only one NAND gate.

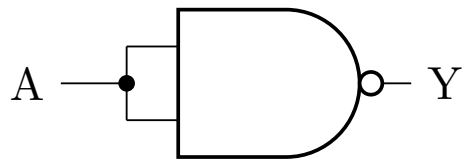


Figure 7.1.4 NAND circuit providing NOT functionality.

Solution. The NAND truth table can be used, letting $B = A$, to show

A	$Y = \overline{A \cdot A}$
0	1
1	0

which is identical to the NOT truth table. ▲

Example 7.1.5 Creating an AND gate using NAND gates. Use truth tables to show that AND functionality can be replicated by the circuit shown in Figure 7.1.6 which uses two NAND gates.

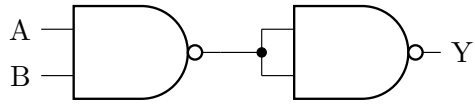


Figure 7.1.6 NAND circuit providing AND functionality.

Solution. First, note that the right-most NAND gate is in a configuration to replicate the NOT gate according to Example 7.1.3. Thus, the output Y is just the inverse of the standard NAND output

A	B	$\overline{A \cdot B}$	$Y = \overline{\overline{A \cdot B}}$
0	0	1	0
0	1	1	0
1	0	1	0
1	1	0	1

which is identical to the AND truth table. ▲

7.2 Boolean algebra

The analysis of digital circuits and their behaviors can be conducted in two ways: 1) the method of exhaustion, and 2) Boolean algebra. The **method of exhaustion** is a brute-force method in which each gate in a circuit is evaluated in turn based on the inputs values it receives. We used this method in the previous section as we were evaluating circuits constructed solely with NAND gates. **Boolean algebra** is a mathematical formalism used for two-state variables and the operations that act on these variables. This formalism allows us to analyze circuits containing multiple gates through equation manipulation.

Boolean algebra, like standard algebra, is governed by rules that determine how mathematical operations interact. The following Boolean identities provide these rules by which we must adhere.

List 7.2.1 Boolean identities

1. $A + B = B + A$
2. $A \cdot B = B \cdot A$
3. $A + (B + C) = (A + B) + C$

4. $A \cdot (B \cdot C) = (A \cdot B) \cdot C$
5. $A \cdot (B + C) = A \cdot B + A \cdot C$
6. $(A + B) \cdot (C + D) = A \cdot C + A \cdot D + B \cdot C + B \cdot D$
7. $\overline{1} = 0$
8. $\overline{0} = 1$
9. $A \cdot 0 = 0$
10. $A \cdot 1 = A$
11. $A + 0 = A$
12. $A + 1 = 1$
13. $A + A = A$
14. $A \cdot A = A$
15. $\overline{\overline{A}} = A$
16. $A + \overline{A} = 1$
17. $A \cdot \overline{A} = 0$
18. $\overline{A + B} = \overline{A} \cdot \overline{B}$
19. $\overline{A \cdot B} = \overline{A} + \overline{B}$
20. $A + \overline{A} \cdot B = A + B$
21. $\overline{A} + A \cdot B = \overline{A} + B$
22. $A \oplus B = \overline{A} \cdot B + A \cdot \overline{B} = (A + B) \cdot (\overline{A \cdot B})$
23. $\overline{A \oplus B} = A \cdot B + \overline{A} \cdot \overline{B}$

Let's look at a few examples demonstrating the use of Boolean algebra.

Example 7.2.2 Boolean algebra: NOT gate from NAND gates. Use Boolean algebra to build a NOT gate from one NAND gate.

Answer. We want the result $\overline{A}$ for input A using only NAND gates. Starting with our desired result, we find that

$$\overline{A} = \overline{A \cdot A} \qquad \text{(Identity 14)}.$$

So, NOT-gate functionality can be produced using the circuit in Figure 7.1.4. ▲

Example 7.2.3 Boolean algebra: AND gate from NAND gates. Use Boolean algebra to make an AND gate from two NAND gates.

Answer. We want the result $A \cdot B$. Starting with our desired result, we find that

$$\begin{aligned} A \cdot B &= \overline{\overline{A \cdot B}} && \text{(identity 15)} \\ &= \overline{\left(\overline{A \cdot B}\right) \cdot \left(\overline{A \cdot B}\right)} && \text{(identity 14).} \end{aligned}$$

So, AND-gate functionality can be produced using the circuit in Figure 7.1.6. ▲

Example 7.2.4 Boolean algebra: OR gate from NAND gates. Use Boolean algebra to make an OR gate from three NAND gates.

Answer. We want the result $A + B$ and want expressions containing AND-like functions instead of OR-like functions. DeMorgan's theorems (identities 18 and 19) are thus going to be instrumental to our effort. Starting with our desired result, we find that

$$\begin{aligned} A + B &= \overline{\overline{A + B}} && \text{(identity 15)} \\ &= \overline{\overline{A} \cdot \overline{B}} && \text{(identity 18)} \\ &= \overline{\left(\overline{A \cdot A}\right) \cdot \left(\overline{B \cdot B}\right)} && \text{(identity 14)} \end{aligned}$$

which describes the circuit in Figure 7.2.5.

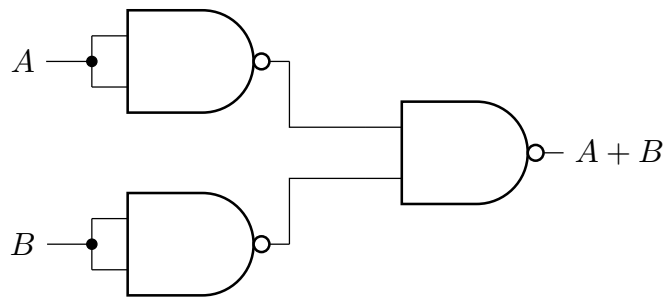


Figure 7.2.5 NAND circuit providing AND functionality.

▲

Example 7.2.6 Boolean algebra: Circuit simplification. Content not yet available. ▲

7.3 Karnaugh Maps

Often, one may have a particular circuit behavior in mind but not know what logic gate configuration will achieve the desired result. A **Karnaugh maps** is a tool that can be used to determine the simplest circuit comprised of AND and OR gates that will produce our desired behavior. Once we have our desired circuit design using AND and OR gates, Boolean algebra can be used to explore alternative circuits that will demonstrate the same behaviors while using other types of logic gates.

There are two types of Karnaugh maps: the Sum-of-Products (SOP) Karnaugh maps and the Product-of-Sums (POS) Karnaugh maps. While we will devote most of our focus to the SOP Karnaugh maps, we will briefly treat POS Karnaugh maps as well.

7.3.1 Karnaugh Maps (Sum-of-Products)

Because Karnaugh maps assist in the design of a circuit that will demonstrate a particular desired behavior, our procedure begins with specifying our desired behavior in the form of a truth table. For this example, we will use the truth table in Table 7.3.1.

Table 7.3.1

A	B	C	Q
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1

This truth table is then used in constructing a Karnaugh map according to the following steps:

1. Divide your inputs into two groups. One group will have possible input values listed along the top row and the other group will have possible input values listed down the first column. Each group must contain either one or two inputs. For this example, I will group inputs A and B into one group (top row) and input C into the second group (left column).

2. For each group of inputs, list the possible combinations of input values in Gray code along their associated row or column of the map. **Gray code** requires an ordering in which only one input changes values between neighboring elements.

So, the AB grouping will list the possible input value combinations in the following order: 00, 01, 11, 10. The C group will have entries of 0 and 1.

3. All other entries in the map will be filled in according to the circuit outputs displayed in Table 7.3.1. For example, the map location representing the intersection between $AB = 01$ and $C = 0$ will contain the value of 1 since the $ABC = 010$ entry of the truth table contains an output value of 1.

The resulting Karnaugh map can be seen in Figure 7.3.2.

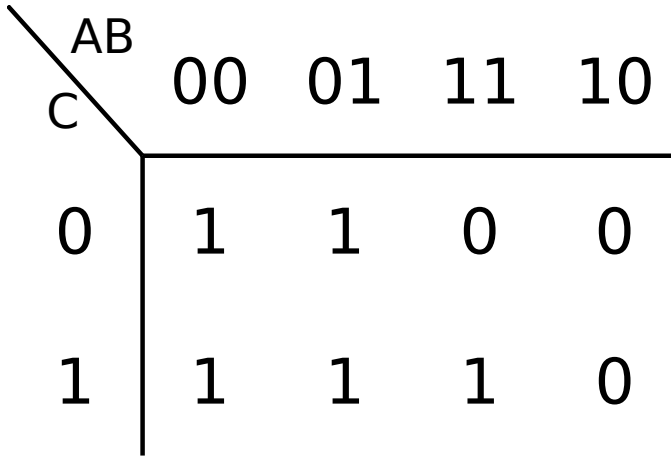


Figure 7.3.2

Examine the Karnaugh map and draw rectangles around groups of 1s, ensuring that each group contains a number of elements that is equal to an integer power of 2 and ensuring that every 1 that appears in the Karnaugh map is contained in at least one group. This results in the Karnaugh map in Figure 7.3.3.

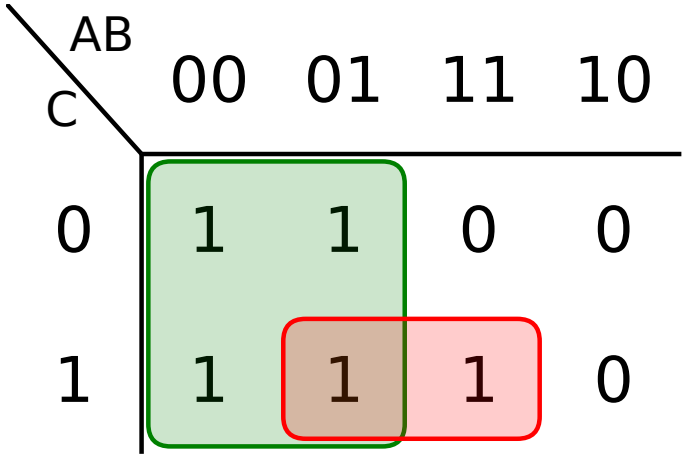


Figure 7.3.3

For each group of 1s, determine which circuit inputs remain constant for all elements in that group. These inputs will be joined together by the AND function with the caveat that an input will be inverted if it remains constant with a value of zero for a given grouping. In our example, the 2x2 grouping has input $A = 0$ as the only input value common to all elements in the group, so this grouping is representative of the Boolean expression $\overline{A}$. The 1x2 grouping has both $B = 1$ and $C = 1$ common to all elements in the group, so this grouping is representative of the Boolean expression $B \cdot C$. The Boolean expression representing the full circuit functionality is then given by joining our AND groups together with ORs, resulting in $Q = \overline{A} + B \cdot C$.

When identifying groups of 1s, larger groupings will result in a simpler final circuit. In our example, we could have used the alternative groupings shown in Figure 7.3.4

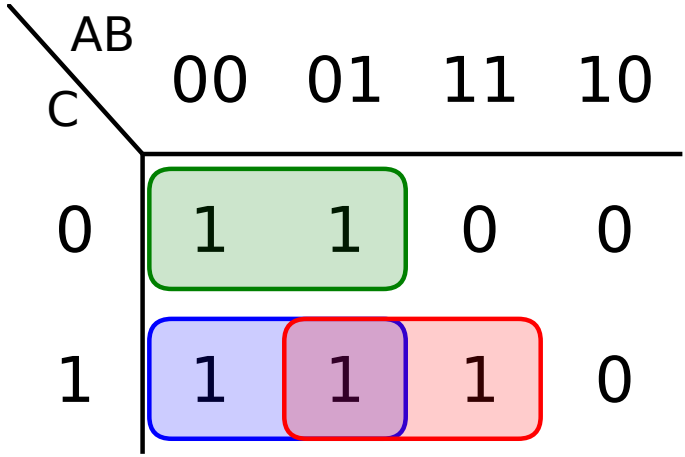


Figure 7.3.4

This results in one group with $A = 0$ and $C = 0$, one group with $A = 0$ and $C = 1$, and one group with $B = 1$ and $C = 1$. The final Boolean expression would then be given by $Q = \overline{A} \cdot \overline{C} + \overline{A} \cdot C + B \cdot C$.

SEE COURSE NOTES FOR WORKED EXAMPLES.

7.3.2 Karnaugh Maps (Product-of-Sums)

In development. Include the following:

1. Use DeMorgan's theorem to connect SOP to POS methods
2. Description of approach
3. Example or two
4. When to use POS instead of SOP

7.4 Binary numbers

Digital computation uses the binary, or base-2, number system. In the binary number system, we will refer to **bits** rather than **digits** which are used in the decimal (base-10) system. Each digit in a base-10 number can assume any one of ten values (0-9) which is then multiplied by 10^n where n is the number of digits to the left of the decimal point. For example, decimal number 1092 can be understood as $1 \times 10^3 + 0 \times 10^2 + 9 \times 10^1 + 2 \times 10^0$. Likewise, each bit in a binary number can take on one of two possible values (0 and 1) with these values multiplying a power of 2. So, the binary number 1011 can be understood as $1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 1 \times 2^0$.

This means that the binary number 1011 represents 11 in decimal. A binary number comprised of n bits has 2^n possible values. Table 7.4.1 provides a list of all fifteen 4-bit positive integers with decimal counterparts.

Table 7.4.1 Binary to decimal equivalents for every 4-bit positive integer.

Binary	Decimal
0000	00
0001	01
0010	02
0011	03
0100	04
0101	05
0110	06
0111	07
1000	08
1001	09
1010	10
1011	11
1100	12
1101	13
1110	14
1111	15

Decimal integers can be subdivided by powers of ten through the use of the decimal point. Similarly, binary integers can be subdivided by powers of two through the use of the binary point. For example, the binary number 11.01 represents $1 \times 2^1 + 1 \times 2^0 + 0 \times 2^{-1} + 1 \times 2^{-2}$ which is equivalent to the decimal number 3.25. In other words, bits to the left of the binary point represent positive powers of two while bits to the right of the binary point represent negative powers of two.

7.5 Adders

In this section, we will discuss **adders** which are circuit elements that perform the addition operation in binary arithmetic. The first circuit element that we'll examine is the **half-adder**, which is shown in Figure 7.5.1.

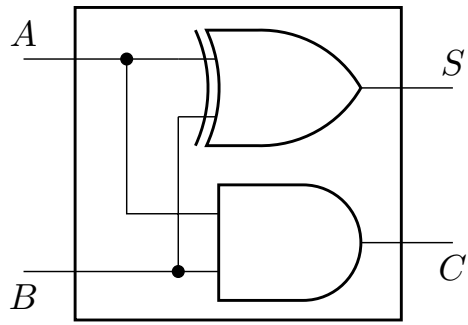


Figure 7.5.1 The half-adder.

A half-adder chip component takes two inputs A and B and produces two outputs: the **sum** $S = A \oplus B$ and the **carry** $C = A \cdot B$. The half-adder truth table is given by Table 7.5.2.

Table 7.5.2 Half-adder truth table

A	B	S	C
0	0	0	0
0	1	1	0
1	0	1	0
1	1	0	1

This behavior characterizes binary addition, where S is the binary sum of A and B. In the event that $A = 1$ and $B = 1$, we are forced to 'carry' a value of one over into the next highest bit using $C = 1$ because S cannot take the value of 2 in binary. In words, this means that A plus B equals S and we carry C to the next bit.

To handle addition for multi-bit numbers, our adder must be able to produce output values for S and C that account for any values 'carried' from the next lowest bit in addition to the values of A and B as demonstrated by the truth table in Table 7.5.3.

Table 7.5.3 Full-adder truth table for the nth bit.

C_{n-1}	A_n	B_n	S_n	C_n
0	0	0	0	0
0	0	1	1	0
0	1	0	1	0
0	1	1	0	1
1	0	0	1	0
1	0	1	0	1
1	1	0	0	1
1	1	1	1	1

Here, A_n, B_n, S_n, and C_n are defined for the nth bit identically to the half-adder. The C_{n-1} input did not appear in the half-adder and represents the act of 'carrying' overflow values from the $(n-1)$th bit to the nth bit. Table 7.5.3 is the truth table associated with the **full-adder** circuit element. The full-adder circuit element can be built using the logic gate arrangement in Figure 7.5.4.

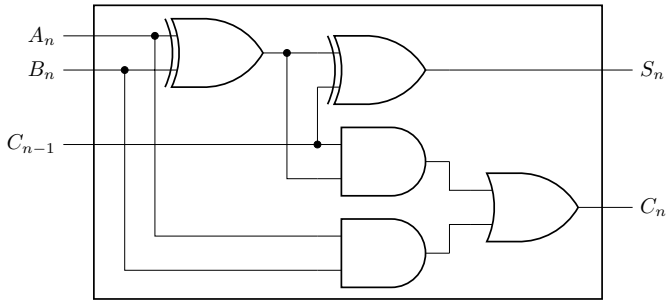


Figure 7.5.4 The full-adder.

The adders described above can be used to perform addition on multi-bit numbers, with one full-adder used for each bit. The addition of two 3-bit numbers

$$\underline{A_2}\ \underline{A_1}\ \underline{A_0} + \underline{B_2}\ \underline{B_1}\ \underline{B_0} = \underline{C_2}\ \underline{S_2}\ \underline{S_1}\ \underline{S_0}$$

will be computed by the circuit in Figure 7.5.5.

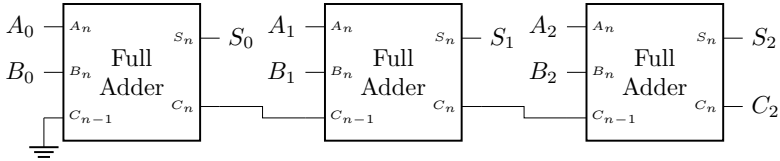


Figure 7.5.5 Circuit schematic for the 3-bit adder circuit.

The full-adder associated with the lowest bit position can of course be replaced by a half-adder and achieve the same functionality.

7.6 Latches

So far, we've limited our treatment of digital circuits to those with outputs that are fully determined by input values at any given instant. In this section, we'll study **sequential logic**, a branch of digial electronics in which circuit outputs depend on input values at that instant in addition to the history of input values. Though the 'memory' that we'll observe in sequential logic circuits is rudimentary compared to that demonstrated by random access memory (RAM) and solid-state drives (SSDs)

in computers, the circuits we'll examine are the fundamental building blocks that make the more advanced applications possible.

7.6.1 SR Cross-NOR Latch

The first sequential logic circuit element that we'll examine is called a **latch**. Figure 7.6.1(a) shows the **SR cross-NOR latch** with inputs S, R and outputs Q, $\overline{Q}$.

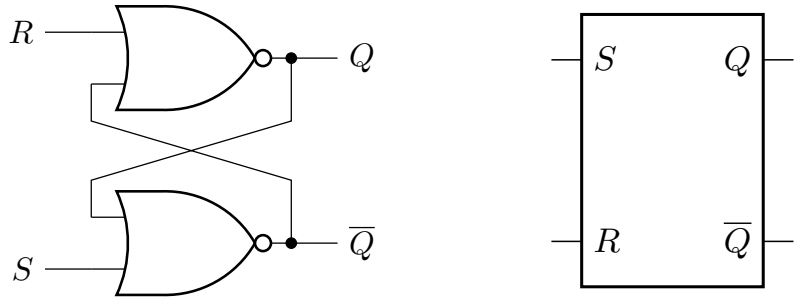


Figure 7.6.1

The SR cross-NOR latch is designed such that $Q = 1$ when $S = 1$ and $R = 0$; we'll call this the 'Set' state. Likewise, the latch is designed so that $Q = 0$ when $S = 0$ and $R = 1$; this is called the 'Reset' state. The output labeled $\overline{Q}$ is in the state opposite to output Q in both of these cases, the desired result implied by the naming convention used for these outputs. Analysis of the internal NOR gates for these two input states requires that the outputs satisfy

$$\begin{aligned} Q &= \overline{R + \overline{Q}} \\ \overline{Q} &= \overline{S + Q}. \end{aligned} \tag{7.6.1}$$

Examples 7.6.2–7.6.3 show that our latch design does indeed satisfy the desired behaviors for the Set and Reset input states.

Example 7.6.2 Analysis of Set and Reset states for SR cross-NOR latch. Determine the output values for the SR cross-NOR latch in the Set state.

Solution. In the Set state, $S = 1$ and $R = 0$. Since $S = 1$, the bottom NOR gate must produce $\overline{Q} = 0$ regardless of the value of Q. Then, since $R = 0$ and $\overline{Q} = 0$, the top NOR gate produces $Q = 1$. This is a self-consistent solution describing the latch behavior. ▲

Example 7.6.3 Analysis of Set and Reset states for SR cross-NOR latch. Determine the output values for the SR cross-NOR latch in the Reset state.

Solution. In the Reset state, $S = 0$ and $R = 1$. Since $R = 1$, the top NOR gate must produce $Q = 0$ regardless of the value of $\overline{Q}$. Then, since $S = 0$ and $Q = 0$, the bottom NOR gate produces $\overline{Q} = 1$. This is a self-consistent solution describing the latch behavior. ▲

The third input state that we'll consider is $S = 1$ and $R = 1$. Just as for the two cases above, the output state for this case is fully determined by the input state values, resulting in $Q = 0$ and $\overline{Q} = 0$. This state will be called the 'Forbidden' state and should be avoided in standard operation because the outputs are not inverses of each other as the naming convention indicates they should be.

The final input that we must examine is the case where $S = 0$ and $R = 0$. Unlike the cases above, the feedback of output values to the NOR gate inputs must be considered. Combining (7.6.1) with $S = R = 0$ and Boolean Identity 11 (from List 7.2.1), we find that $Q = \overline{\overline{Q_0}} = Q_0$ and $\overline{Q} = \overline{Q_0}$ where the subscript zero represents the output states prior to entering the 'Hold' state. Though these look similar to identity statements, we must be careful in our interpretation. Logic gates take a finite amount of time to respond and update output values due to changes in input values. If we examine the NOR gate inputs at time $t = t_0$ and assume a logic gate response time δt, our analysis would determine output values at $t_1 = t_0 + \delta t$. Thus, when $S(t_0) = 0$ and $R(t_0) = 0$, we find

$$Q(t_1) = \overline{R(t_0) + \overline{Q(t_0)}} = Q(t_0)$$
$$\overline{Q(t_1)} = \overline{S(t_0) + Q(t_0)} = \overline{Q(t_0)}.$$

We'll call this the 'Hold' state since both Q and $\overline{Q}$ remain unchanged with whatever values they held before entering the Hold state.

The truth table in Table 7.6.4 collects our results for the SR cross-NOR latch from above. The behavior of this latch can alternatively be demonstrated in the **timing diagram** in Figure 7.6.5, showing the time response of the outputs based on input changes in time.

Table 7.6.4 SR cross-NOR latch truth table.

S	R	Q	$\overline{Q}$	**Condition**
0	0	Q_0	$\overline{Q_0}$	Hold
0	1	0	1	Reset
1	0	1	0	Set
1	1	0	0	Forbidden

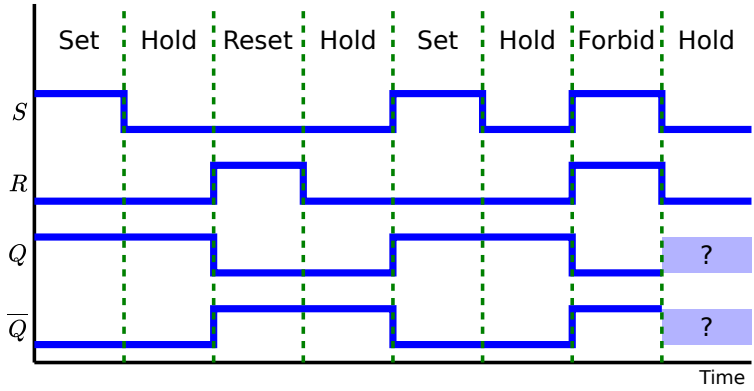
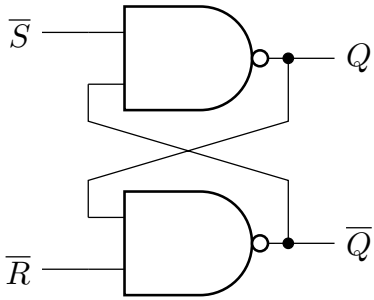


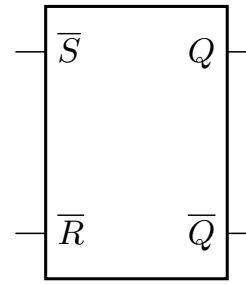
Figure 7.6.5 Example timing diagram for the SR cross-NOR latch. Note that the final state has undetermined outputs. Outputs Q and $\overline{Q}$ must be inverses of each other in this final state, but a race will determine which latch output reaches a HI state first with the other output renaming low. This unpredictability is the reason the 'forbidden' state should be avoided for our uses.

7.6.2 SR Cross-NAND Latch

The **SR cross-NAND latch** shown in Figure 7.6.6 is a variation on the latch described above. In contrast to the previous latch, the SR cross-NAND latch tracks which input was most recently LO.



(a) SR cross-NAND latch circuit element.



(b)

$\overline{S}$	$\overline{R}$	Q	$\overline{Q}$	**Condition**
0	0	1	1	Forbidden
0	1	1	0	Set
1	0	0	1	Reset
1	1	Q_0	$\overline{Q_0}$	Hold

(c) SR cross-NAND latch truth table.

Figure 7.6.6

One application that makes use of the SR cross-NAND latch is the **switch debouncer**. When using a mechanical switch to turn current off and on through a load (as in Figure 7.6.7(a)), the connection at the instant of physical contact closing the switch can be be flaky and/or unstable, with the switch 'bouncing' back and forth between open and closed several times before a solid connection is made. The instability in the mechanical switch's connection leads to the current rapidly oscillating between off and on through the load. This undesirable behavior is rectified through the use of a switch debouncer (as shown in Figure 7.6.7(b)). Connecting the switch to $\overline{S}$ causes the latch inputs take values $\overline{S} = 0$ and $\overline{R} = 1$, resulting in the latch Set state and allowing current to pass through the load. Likewise, when the switch is connected to $\overline{R}$, the latch enters the Reset state and no current flows through the load. During the switch transition between $\overline{S}$ and $\overline{R}$, both $\overline{S} = \overline{R} = +5\text{V}$ causing the switch to enter the Hold state. Therefore, any flakiness that may occur as the switch's physical connection is made or broken does not result in any flakiness in the current through the load. Instead, the load current depends on the most recent connection only and maintains the output behavior even if the physical connection is unstable or flaky.

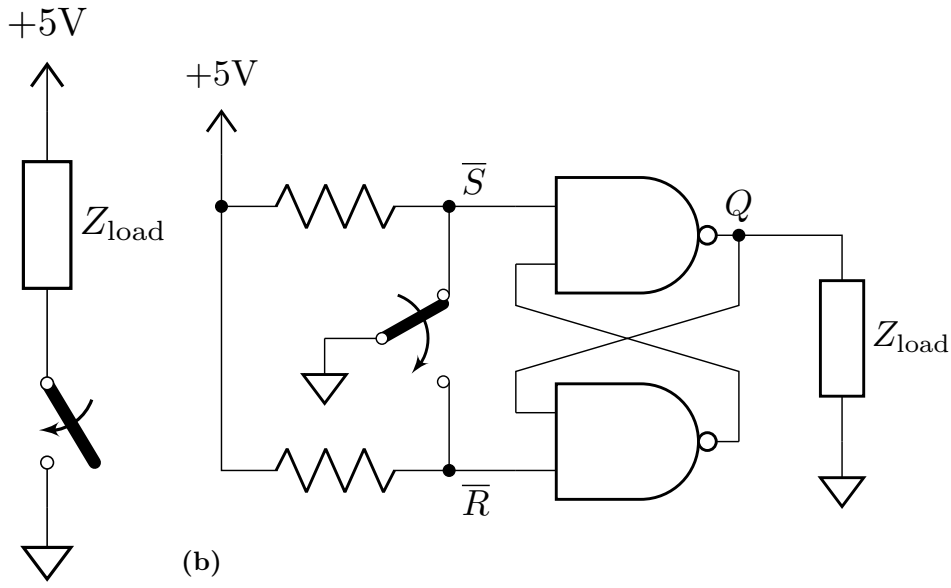
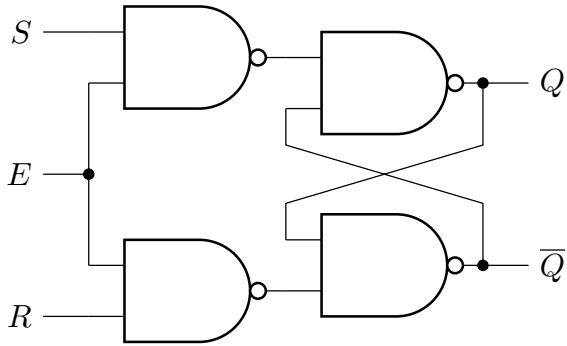


Figure 7.6.7 A switch debouncer is used to reduce the 'flakines' of mechanical switches by entering the Hold state during the transition between switch positions.

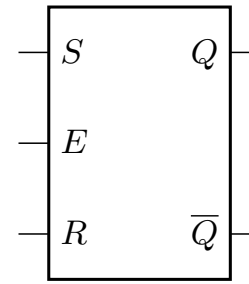
APPLICATION: LATCHED TEMPERATURE OR LIGHT ALARM.

7.6.3 Gated latches

There are latch applications where it may be beneficial to quickly enable or disable latch input control over the latch outputs. Figure 7.6.8 shows the circuit and truth table for a **gated** SR latch. When $E = 1$, the latch is **enabled** meaning that the latch outputs are determined by the latch inputs. When $E = 0$, the inputs are disabled and the latch enters the 'Hold' state independent of the values of S and R. The implementation of the gating functionality shown in Figure 7.6.8(a) results in a latch with **active high** inputs. This means that the latch, when enabled, has outputs that respond to inputs that are high. So, the Set state requires $E = 1$, $S = 1$, and $R = 0$. This behavior is in contrast to the un-gated cross-NAND SR latch which has **active low** inputs.



(a) Gated SR latch.



(b)

E	S	R	Q	$\overline{Q}$	**Condition**
0	×	×	Q_0	$\overline{Q_0}$	Hold
1	0	0	Q_0	$\overline{Q_0}$	Hold
1	0	1	0	1	Reset
1	1	0	1	0	Set
1	1	1	0	0	Forbidden

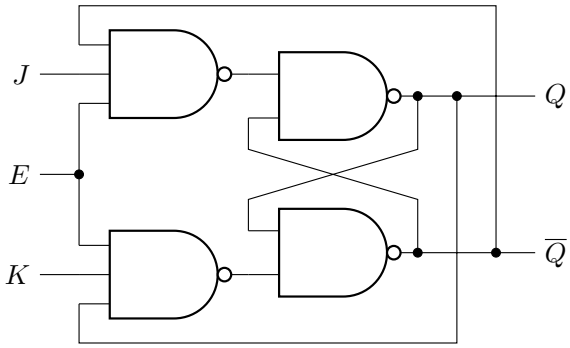
(c) Clocked SR latch truth table.

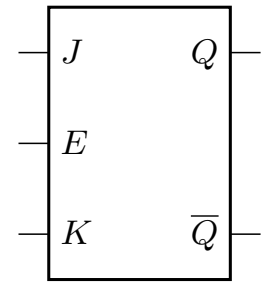
Figure 7.6.8

PROVIDE EXAMPLE APPLICATIONS

All of the SR latch variants above suffer from the same issue, namely the accessibility of the state that we've labeled 'forbidden'. It would be very beneficial to have a latch circuit for which the 'forbidden' state were inaccessible. We will examine two possible solutions to this problem that are each modifications to our original cross-NAND SR latch.

The **JK latch** uses feedback of the latch outputs back to the latch inputs to convert the 'forbidden' state (when both inputs are HI) into a 'toggle' state in which Q and $\overline{Q}$ swap values. Figure 7.6.9 provides the circuit diagram and truth table for the JK latch.





(b)

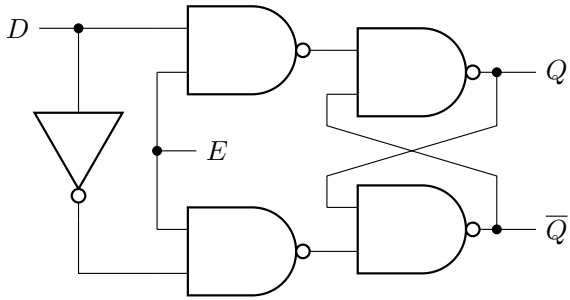
(a) Circuit diagram for the JK latch.

E	J	K	Q	$\overline{Q}$	**Condition**
0	×	×	Q_0	$\overline{Q_0}$	Hold
1	0	0	Q_0	$\overline{Q_0}$	Hold
1	0	1	0	1	Reset
1	1	0	1	0	Set
1	1	1	$\overline{Q_0}$	Q_0	Toggle

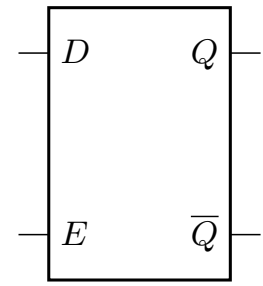
(c) Truth table for the JK latch.

Figure 7.6.9

While the JK latch uses feedback to repurpose the 'forbidden' state that appeared with SR latches, the **D latch** is designed instead to make it impossible to access that forbidden state at all. Figure 7.6.10 shows a modified SR latch version of the D latch with its truth table.



(a) Circuit diagram for the D latch.



(b)

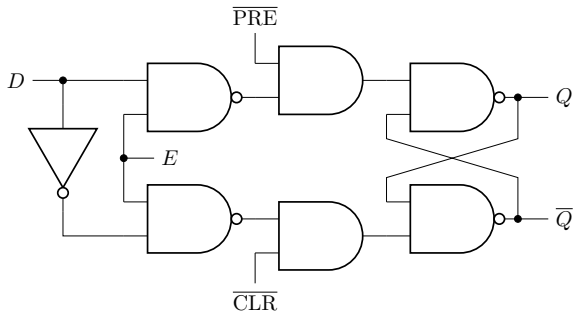
E	D	Q	$\overline{Q}$	**Condition**
0	×	Q_0	$\overline{Q_0}$	Hold
1	0	0	1	Reset
1	1	1	0	Set

(c) Truth table for the D latch.

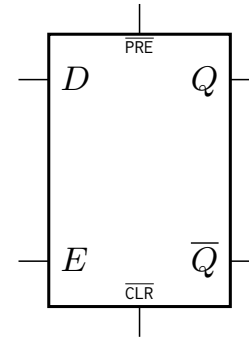
Figure 7.6.10

7.6.4 Preset and Clear

There are applications in which it is beneficial to place a latch into a specific state while overriding all gate and input states. Many latch circuits incorporate **Preset** and **Clear** inputs for just this purpose. The modified D latch circuit in Figure 7.6.11 provides an example of this functionality with *active low* Preset and Clear inputs.



(a) Circuit diagram for the D latch.



(b)

$\overline{\text{PRE}}$	$\overline{\text{CLR}}$	E	D	Q	$\overline{Q}$	**Condition**
0	0	×	×	1	1	Forbidden
0	1	×	×	1	0	Preset
1	0	×	×	0	1	Clear
1	1	0	×	Q_0	$\overline{Q_0}$	Hold
1	1	1	0	0	1	Reset
1	1	1	1	1	0	Set

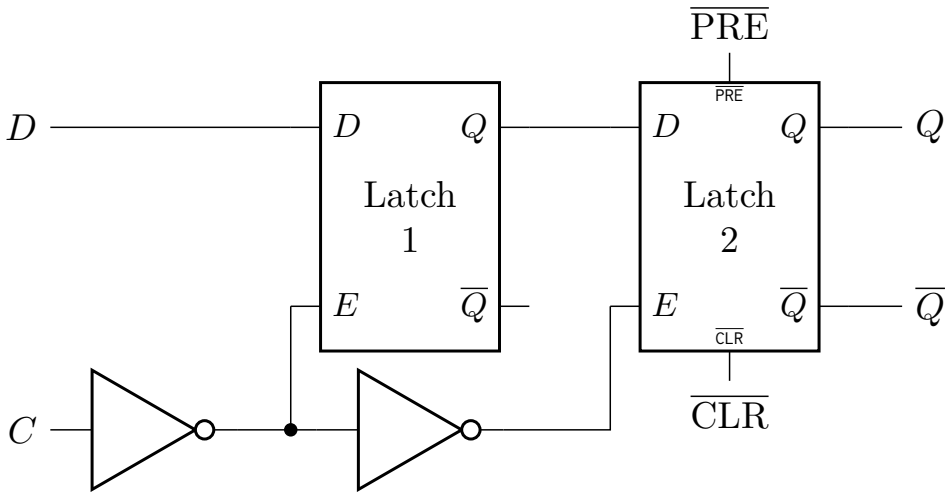
(c) Truth table for the D latch.

Figure 7.6.11

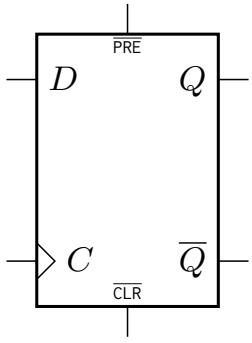
7.7 Flip-flops

In the last section, we discussed latches which are **level-triggered** circuits with output levels that depend on the input levels at any particular instant of time. **Flip-flops**, the topic for this section, are **edge-triggered** devices with output levels that change only at the instant at which a **clock** signal (CLK) transitions between high and low in a particular direction.

Our first D flip-flop circuit design is shown in Figure 7.7.1 along with its circuit symbol and truth table.



(a) Circuit schematic.



(b) Circuit symbol.

C	D	Q	$\overline{Q}$	**Condition**
0	×	Q	$\overline{Q}$	Hold
1	×	Q	$\overline{Q}$	Hold
↑	0	0	1	Reset
↑	1	1	0	Set

(c) tbl-digital-flipflop-D

Figure 7.7.1 D flip-flop that triggers on the rising edge of the CLK signal. (a) shows the circuit schematic, (b) shows the circuit symbol where the wedge shape on the CLK input distinguishes itself from the symbol for a D latch, and Table 7.7.1(c) shows the truth table for this flip-flop.

This circuit acts much like the D latch from the previous section except that the outputs are only allowed to change at instants when the CLK signal transitions from LOW to HIGH (Table 7.7.1(c)). The timing diagram in Figure 7.7.2 demonstrates the response of the flip-flop output to a time-varying CLK and D input signal.

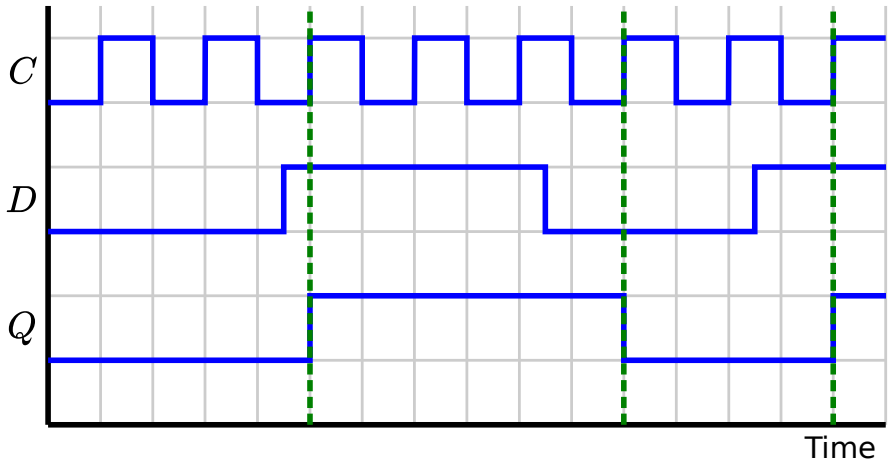


Figure 7.7.2 Timing diagram for the circuit pictured in Figure 7.7.1(a).

Example 7.7.3 Intermediate stages of rising-edge D flip-flop.. Use the behavior of D latches to show that the circuit in Figure 7.7.2 does indeed produce the timing diagram shown in timing-digital-flipflop-D. To accomplish this, assume C and D as shown in the original timing diagram and also assume that $Q_1 = 0$ at $t = 0$. Produce lines showing the time behavior of EN_1, Q_1, EN_2, and finally $Q = Q_2$. Explain the rationale used to draw each signal.

Answer.

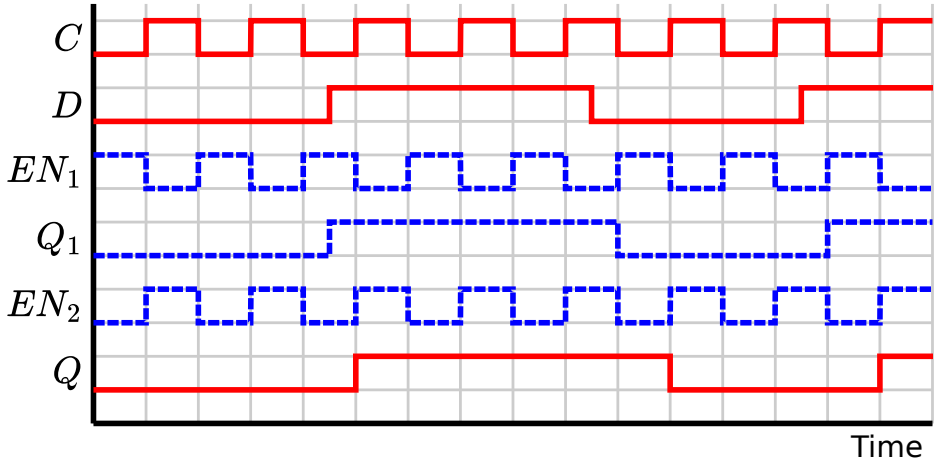
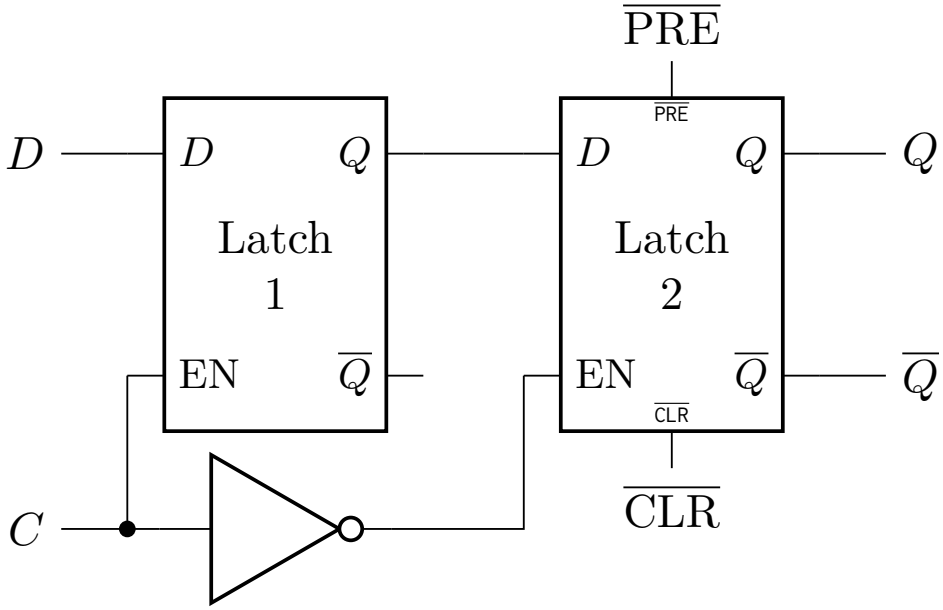


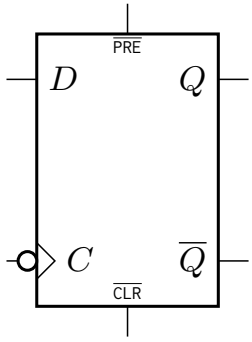
Figure 7.7.4 Timing diagram for the circuit pictured in Figure 7.7.1(a).

Solution. To come. ▲

Removal of one NOT gate from the flip-flop circuit above produces a D flip-flop that triggers on falling edge of the CLK signal instead of the rising edge. The circuit schematic, circuit symbol, and truth table are shown in Figure 7.7.5.



(a) Circuit schematic.



(b) Circuit symbol.

C	D	Q	$\overline{Q}$	**Condition**
0	×	Q	$\overline{Q}$	Hold
1	×	Q	$\overline{Q}$	Hold
↓	0	0	1	Reset
↓	1	1	0	Set

(c) tbl-digital-flipflop-D

Figure 7.7.5 D flip-flop that triggers on the falling edge of the CLK signal. (a) shows the circuit schematic, (b) shows the circuit symbol where the addition of a bubble on the CLK input indicates the falling-edge response, and Table 7.7.5(c) shows the truth table for this flip-flop.

7.8 Shift Registers and Counters

Having established the behavior of flip-flops, we move on to applications that make use of these circuit elements. In particular, we will examine shift registers and the counters that are based on them.

7.8.1 Shift Registers

Let's examine a circuit that is called a **shift register**. An example of a shift register circuit is provided in Figure 7.8.1

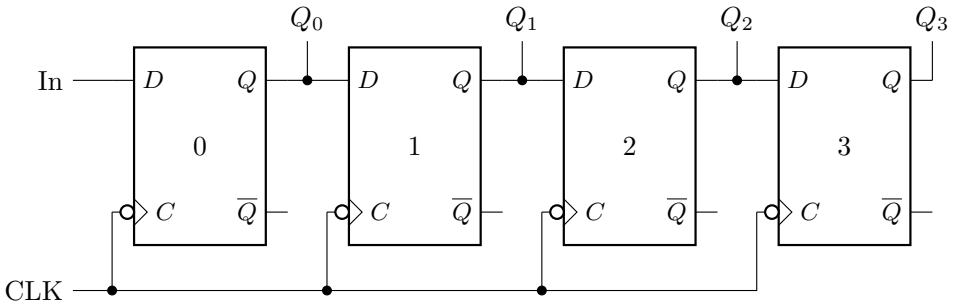


Figure 7.8.1 Shift register circuit constructed using four D flip-flops that are each triggered by the falling edge on the clock signal C.

In this circuit, the input signal (In) is assumed to be controlled by some external source (another circuit or a person) and can change between high and low at any time. Output Q_0 reproduces an approximation to this input signal, though only allows changes between low and high on the negative edge of the clock signal CLK. In a sense, the 'resolution' with which the flip-flop reproduces the input signal is limited by the CLK frequency which is why the pulse lengths in Q_0 are not identical to those in the input signal and is why the circled input signal transitions do not appear in Q_0. Outputs Q_1, Q_2, and Q_3 are all identical to Q_0 except that each output is time delayed by $\Delta t = 1/f_{\text{CLK}}$ from the output before it. Figure 7.8.2 shows this behavior where circuit outputs at some instant are not dependent on the input at that instant. Instead, Q_1, Q_2, and Q_3 depend on the state of Q_0 progressively further in the past and Q_0 is a sampled version of the input signal. *Memory* in this context is represented by circuit output behavior that is based on past rather than present input signal behavior.

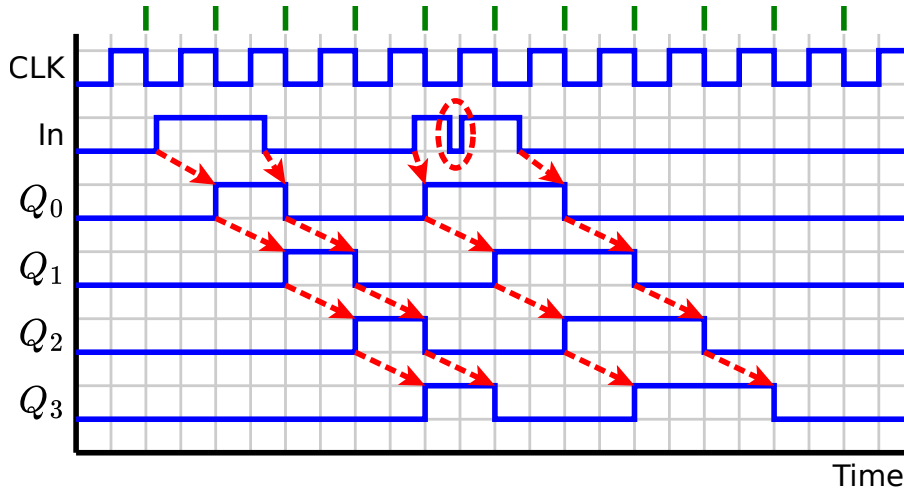


Figure 7.8.2 This shift register timing diagram reflects the behavior of the shift register circuit in Figure 7.8.1. The output Q_0 represents the sampling of the input signal based on the negative edge triggering of flip-flop 1. Outputs Q_1, Q_2, and Q_3 are identical to Q_0 except they are time delayed by one clock period relative to the previous output signal.

7.8.2 Counters

We can use the memory functionality demonstrated by flip-flop circuits above to produce a circuit that *counts*. Let's first start with the flip-flop configuration that is called a **mod-2 counter**. This circuit, shown in Figure 7.8.3, takes a square-wave clock input and produces a square-wave output with half of the clock's frequency.

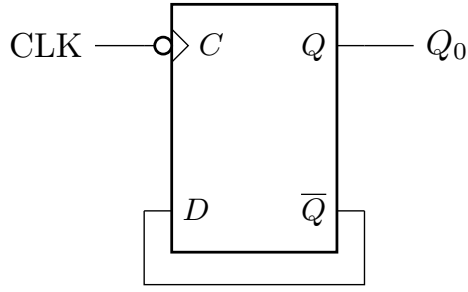


Figure 7.8.3 Mod-2 counter circuit.

The circuit shown in Figure 7.8.4 uses four mod-2 counters in a cascading configuration. The output of one mod-2 counter serves as the input to the next mod-2 counter in the chain, resulting in square-wave outputs that have frequencies that get

progressively smaller by a factor of two for each subsequent flip-flop. This circuit is called an **asynchronous 4-bit counter**. It counts from 0 to 15 in binary before resetting to 0, with Q_i representing the value of the i-th bit of a 4-bit number as a function of time.

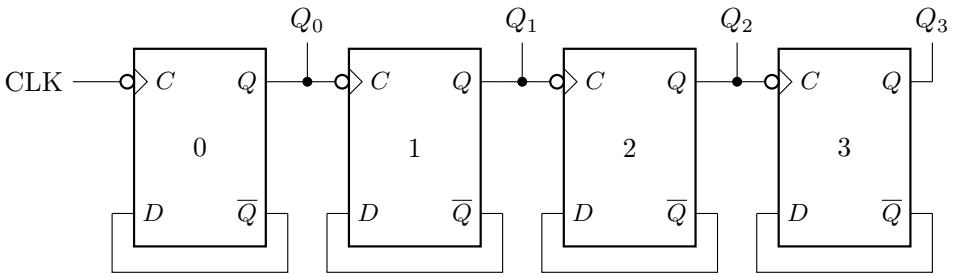


Figure 7.8.4

The term **asynchronous** is used because the four flip-flops in the circuit are not triggered at the same time. Instead, each flip-flop is triggered by changes in the output state of the flip-flop appearing to its left. Because a flip-flop takes a finite time to change its output upon triggering, this leads to a **ripple effect** in which a small delay in flip-flop triggering propagates left to right through the counter circuit.

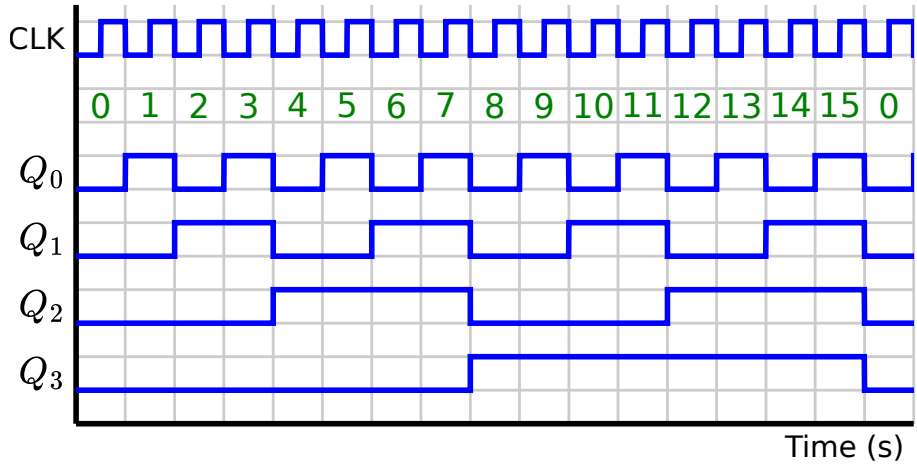


Figure 7.8.5 Timing diagram for the asynchronous 4-bit counter.

The Preset and Clear flip-flop inputs can be used to modify the asynchronous 4-bit counter circuit in Figure 7.8.4 to truncate the count and cause the counter to reset before it reaches 15. For example, an **asynchronous decade counter** will count from 0-9 before resetting back to 0. In order to truncate the count, the circuit

must reset to the 4-bit binary number 0000 instead of proceeding to 1010 (the binary equivalent to the digit 10). So, the circuit must send a clear signal to all four flip-flops immediately upon the outputs reaching $Q_3 = 1$, $Q_2 = 0$, $Q_1 = 1$, and $Q_0 = 0$. This condition can be simplified as this output is the first time at which both $Q_3 = 1$ and $Q_1 = 1$ simultaneously. Since the CLR inputs are *active low*, a single NAND gate provides this truncation when the gate has inputs Q_1,Q_3 and an output that connects to all four flip-flop CLR inputs. Figure 7.8.6 shows the asynchronous decade counter circuit while Figure 7.8.7 shows the counting functionality.

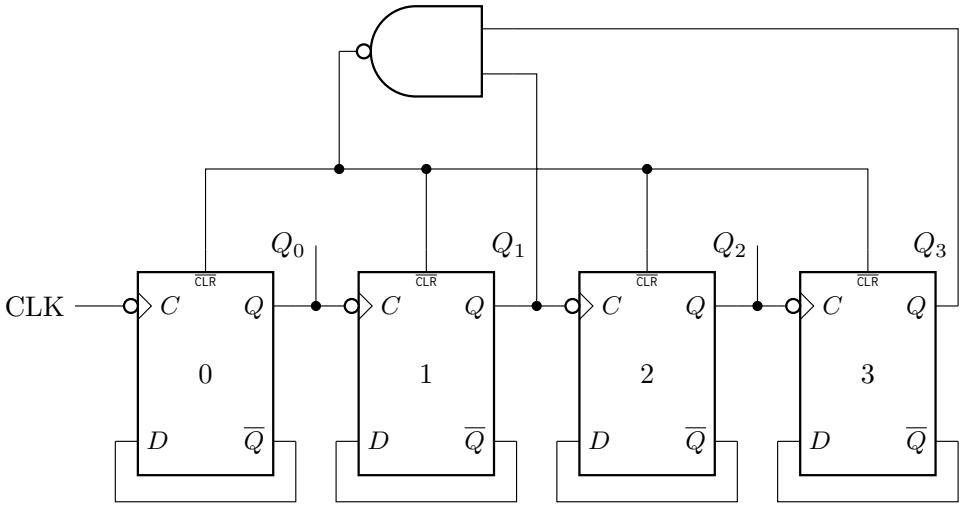


Figure 7.8.6

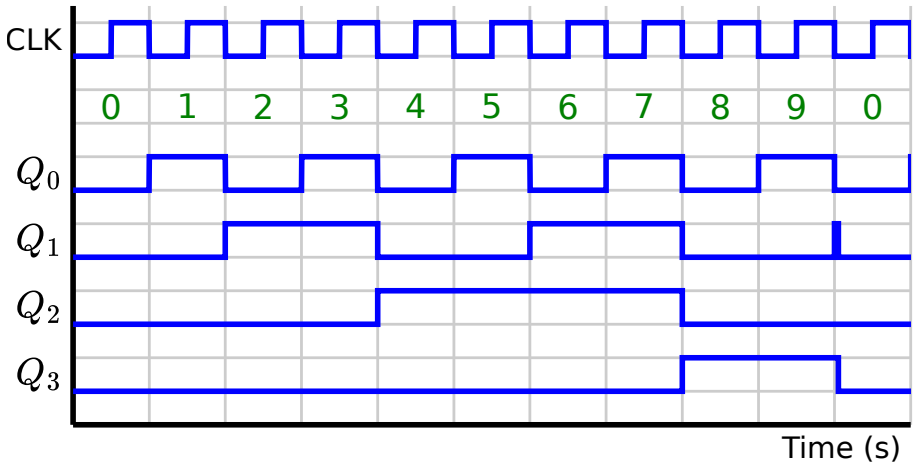


Figure 7.8.7 Timing diagram for the asynchronous decade counter.

Example 7.8.8 Divide-by-5 and Divide-by-100 counters. Assign as homework questions. ▲

7.8.3 Seven-segment LED Display

Chapter 8

Oscillators

Part II

Laboratory

Chapter 9

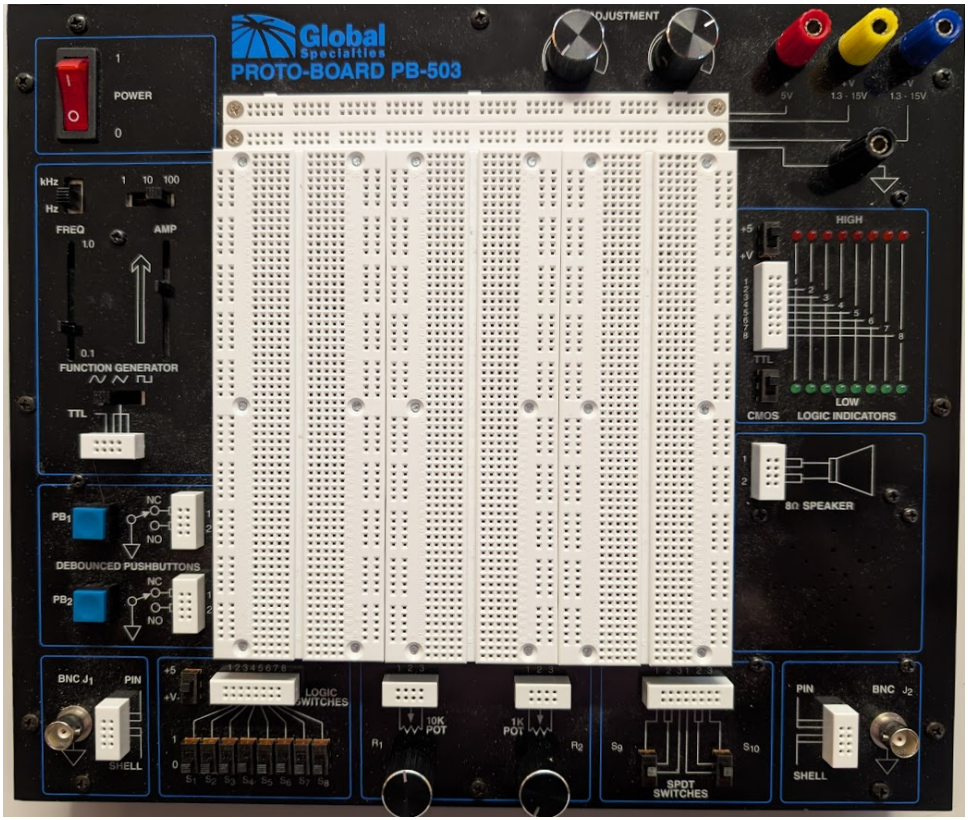
Introduction to Equipment in the Laboratory

This document provides descriptions of common equipment that may be used in the laboratory component of an undergraduate course in Electronics for physicists. In particular, this will provide the base knowledge necessary for one to successfully conduct the experiments that follow. Please refer back to this chapter frequently when conducting experiments.

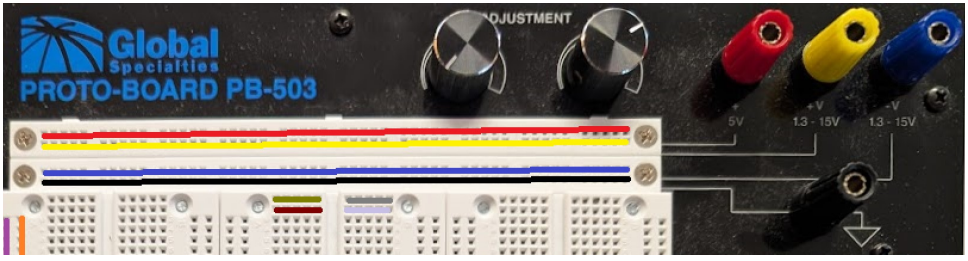
9.1 Prototyping Board (or Breadboard)

Prototyping boards are rows of connectors wired together behind a plastic face. While circuits in your electronics are soldered (often on printed circuitboards), prototyping boards have holes that allow component/wire leads to be pushed in to form a solderless connection. This makes circuit assembly, disassembly, and modification much easier and faster while developing a circuit. Once a final circuit design has been determined, a more permanent construction may be performed.

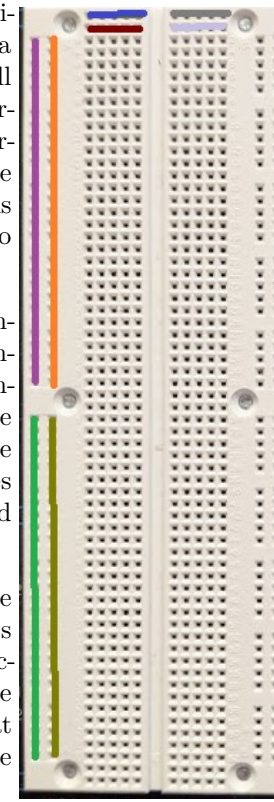
The prototyping boards that we'll use look like this:



Along the top of this Proto-Board, we have four rows of connectors. The holes in the top row are all connected to +5V, the holes in the second row are connected to +V (variable between +1.3V to +15V), the holes in the third row are connected to –V (variable between -1.3V to -15V), and the holes in the fourth row are connected to the board's ground. These connections are hard-wired and cannot be changed.



The rest of the board is broken up into segments that look like the figure to the right. The connections are set up such that:



- The holes that are grouped into a single long vertical column are connected to each other, but with a break halfway down each column. For instance, all holes covered by the purple line are connected internally,as are holes covered by the green line and orange line, though holes covered by different colors are not connected internally. One use for these columns is to distribute voltages from the top voltage strip to other parts of the circuit.
- There are also a series of horizontal groupings containing five holes in a row. These five holes (for instance, the holes covered by the blue line) are connected to each other internally. Likewise, the same is true for the holes covered by the brick red line and by the lavender line and gray line. Again, holes covered by different colored lines are not connected internally.
- Using these various rows and columns, we are able to make connections between electrical components (such as resistors, capacitors, transistors, semiconductor chips, and wires). While it takes a little practice to get used to this design, you will quickly find that these boards allow great flexibility as you prototype your circuits.

9.2 Digital Multimeter (DMM)

The handheld digital multimeter (or DMM) is commonly used to measure

- voltage (DC and AC),
- current (DC and AC),
- resistance,
- continuity, and
- other lesser-used quantities such as frequency.

In voltage and current modes, the internal design of the DMM is such that it has negligible effect on the circuit being measured. That said, you should always spare a moment to consider whether the input impedance in each mode is of the appropriate size to minimize the effect on the measured circuit.

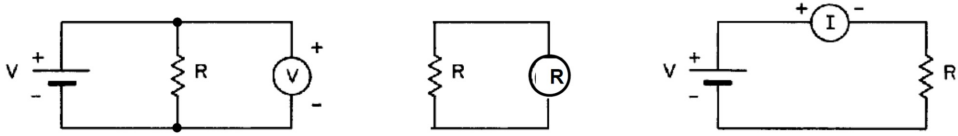
(a) Measuring Voltage or Resistance (Switch dial to Ω for resistance)

(b) Measuring Currents known to be lower than 600mA

(c) Measuring Currents known to be lower than 10A

The use of the DMM varies based on what quantity is being measured. Note that the meter must be connected:

- in parallel to measure voltage or resistance, and
- in series to measure current.



9.3 Oscilloscopes (BK Precision 2190E or 2542B)

The oscilloscope is an extremely versatile instrument. It can be used to measure both steady and time-dependent voltages, frequency, time duration, phase difference, and harmonic distortion. Some oscilloscopes automatically test transistors, perform spectral analysis, integrate, differentiate, sum, subtract, filter, and store electrical signals. Anyone working in an experimental science should be able to use an oscilloscope with considerable confidence.

The oscilloscopes that we are using in lab (BK Precision 2542B or 2190E) are general-purpose, two-channel digital storage instruments capable of displaying and capturing electrical signals with frequencies up to 100 MHz. Its front panel can be divided into a number of sections:

- The color LCD display shows the signal on one or more of the two input channels, using a different color for each channel.
- A set of buttons along the right hand side of the display are used to make menu item selections.
- A set of buttons and knobs on the right half of the instrument control some of the advanced oscilloscope functions, including:
    - a TRIGGER section that permits one to control the starting point for the displayed waveform,
    - a VERTICAL section, accepting up to two input signals and containing controls to vary the voltage resolution (among other things),
    - a HORIZONTAL or 'time base' section that controls the time resolution of the display.

In its most commonly used mode of operation, an oscilloscope creates a visual display by capturing and successively overlaying identical windows of time, each window beginning at the same point (relative to the 'trigger point') on the signal. The time base of the scope determines how much time is represented by each horizontal division of the LCD display. The triggering function determines when each time window is to begin; proper triggering is necessary for stable, unambiguous display of the desired waveform. Finally, the vertical section of the scope controls the amplitude of the information displayed. It determines for each of the two channels how many volts are represented by each vertical division on the display.

Hints regarding oscilloscope controls.

- To change the horizontal (time base) scale:
    1. The Time/div knob changes the time per division (sec/div are color coded by input channel and displayed on the screen).
    2. Position knob changes location of the trigger (labeled as an arrow along the top of the screen).
- To change the vertical (voltage base) scale:
    1. Scale know changes the voltage per division (V/div are color coded by input channel and displayed on the screen).
    2. Position knob changes location of the ground (labeled as a color-coded arrow on the left of the screen).
- To access the parameters for each channel, press Ch. 1 (or 2,3,4) button twice and chose from:

1. Input coupling (AC/DC/Ground)
2. Invert signal
3. Probe setup
4. Bandwidth limit
5. Probe
6. Digital Filter
7. Volts/scale
8. Invert

- To adjust the trigger:

  1. Trigger knob changes the voltage level of the trigger.
  2. Trigger MENU button
     (a) Select which channel to trigger off of (trigger level arrow changes color to show which channel is being triggered)
     (b) Slope: can trigger off a rising edge or falling edge
     (c) Trigger mode: Auto, Normal, other

- To select the ACQUIRE mode:

  1. Run/Stop: sets the scope to continuously acquire or freeze after the last trigger.
  2. Single sequence: draws the signal only once after a single trigger. Will not trigger on successive events.
  3. Autoset: scope tries to choose overall best scope settings (sometimes useful when you can't see anything, but generally used as a last resort because it can also mislead you.

- Measure menu:

  ◦ You can select various measurements and which channel to measure. Be very careful with automatic measurements (again these can be deceptive). Generally use cursors to make accurate measurements. Note that the automatic mode searches for the very top and bottom of the signals (when measuring voltages) which may not provide an accurate value for the actual amplitude as noise on the signal may artificially inflate the max/min values.

- Cursor menu:

  1. Choose time (vertical bars) or voltage (horizontal bars) measurement.

2. Choose which channel to measure (measurement cursors may be the same color as the channel they are measuring).
3. Choose manual adjustment (adjust each cursor separately) or fixed (adjust both cursors simultaneously).
4. Choose which cursor to adjust (Cursor A/Cursor B).
5. Use the knob in the upper-left corner of the controls to adjust the cursor positions.
6. The position relative to ground (or trigger zero time for time measurements) is displayed on the screen with the '@' symbol.
7. The relative distance between the cursors is displayed on the screen with the 'Δ' symbol.

For further details, please read A Shortish Guide to Using an Oscilloscope, posted to the course webpage.

9.4 Function Generator

9.5 Electronics components

9.5.1 Resistor Color Codes

Resistors are marked with (typically) four colored bands. These bands provide a code that can be used to determine the fvalue of the resistance. The meanings of the colors and band placements are provided in Figure 9.5.1.

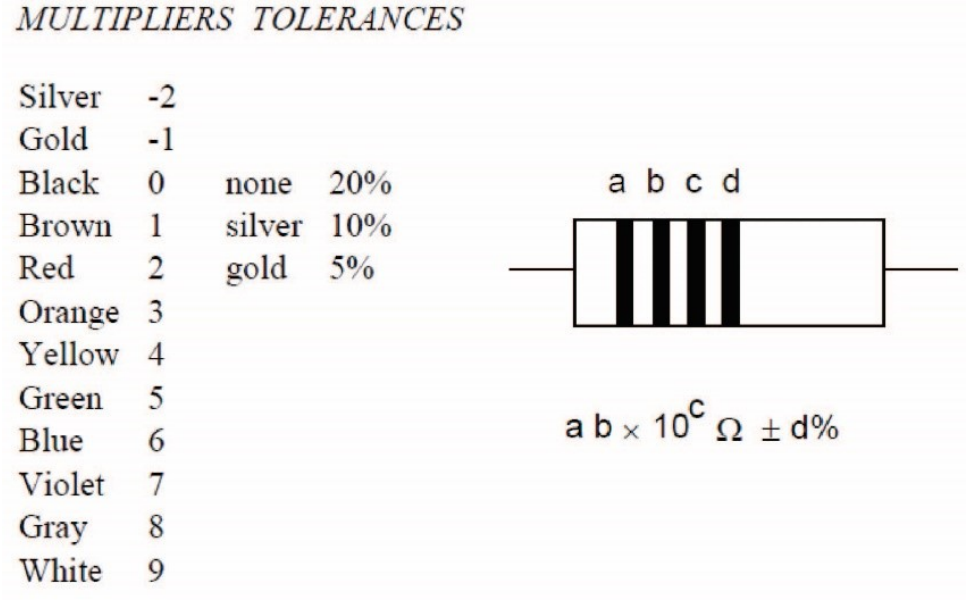


Figure 9.5.1 [GENERATE OWN FIGURE] Resistor color codes.

Thus, a resistors with a Green-Violet-Orange-Silver code has a resistance of $57 \times 10^3 \Omega = 57\text{k}\Omega \pm 10\%$.

Figure 9.5.2 Resistor color code practice. put into example.

Figure 9.5.3 Resistor color code practice. put into example.

9.5.2 Potentiometers

Read about potentiometers here for now: Digikey Potentiometer Explanation

Potentiometers (or pots) typically have three terminals: Two are the ends of a fixedc resistor and one is the 'wiper', as pictured in Figure 9.5.4.

Figure 9.5.4 Potentiometer

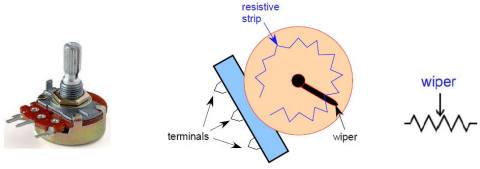


Figure 9.5.5 [GENERATE OWN IMAGE] a) Physical design, b) what's inside, c) schematic symbol

This design allows the potentiometer to be used as a variable resistor, where the resistance varies between the wiper terminal and each of the fixed terminals as the knob is rotated. Before using aany potentiometer, ensure that you are sure of

the role tqaken by each of the terminals, as there can be variation in the terminal layouts.

9.5.3 Switches

The SPST (single pull, single throw) switch has two terminals, while the SPDT (single pull, double throw) switch has three terminals as shown in Figure 9.5.6.

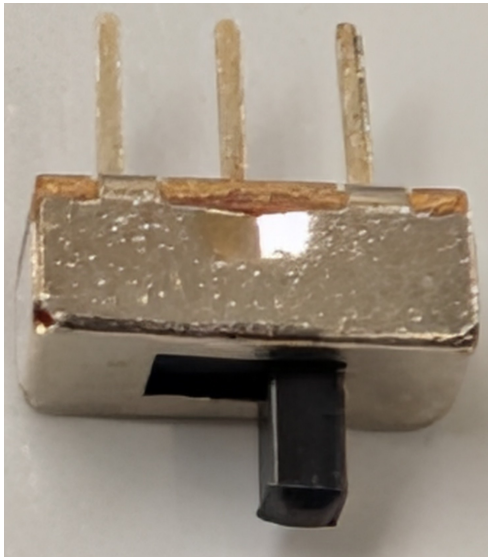


Figure 9.5.6 Single-pull double-throw (SPDT) switch.

Figure 9.5.7 [GENERATE OWN FIGURE]

The SPST switch is meant to flip between a closed circuit and an open circuit, depending on the switch position. The SPDT is meant to flip between connection A and connection B, depending on the switch position. Always verify the switch operation before incorporating it into your circuit. Many prototyping boards have built-in switches available, as shown in Figure 9.5.8.

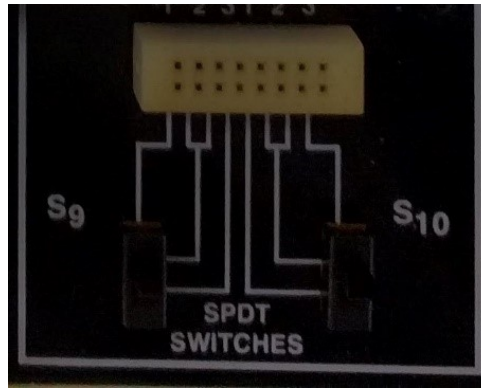


Figure 9.5.8 [GENERATE OWN IMAGE]

9.5.4 Wires, Cables, and Connectors

Banana Plugs (Figure 9.5.9) are typically at the ends of an insulated wire containing a single conducting path. The cable color is irrelevant to the behavior of the wires, though some thought should be used when choosing colors to assist with wiring organization. Alligator clips can be attached to banana wires, allowing them to clip to a component.

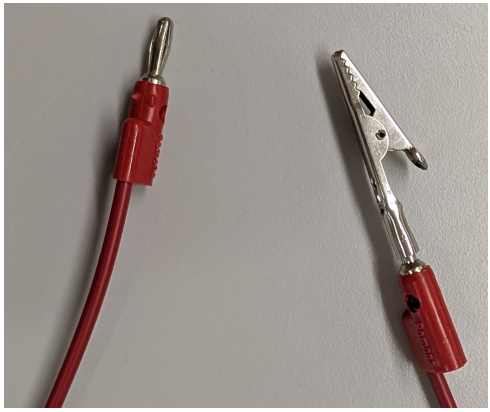


Figure 9.5.9 Cable with banana connectors. There is an alligator clip placed on one of the male banana connectors.

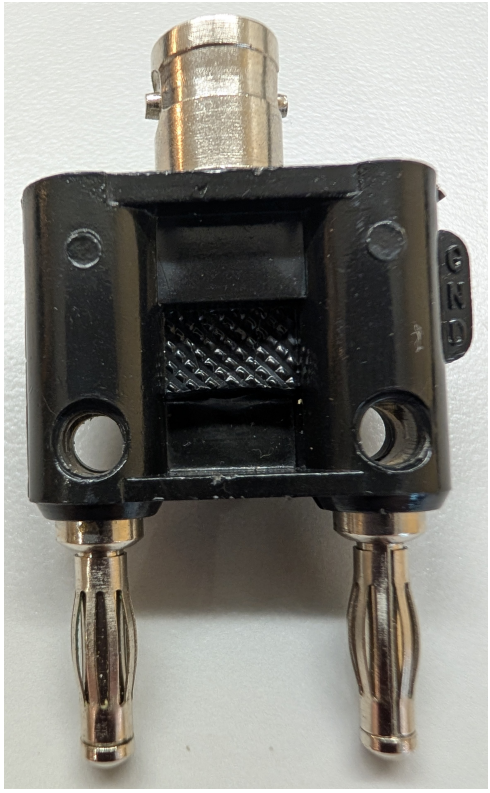


Figure 9.5.10 Male BNC to male banana adapter.

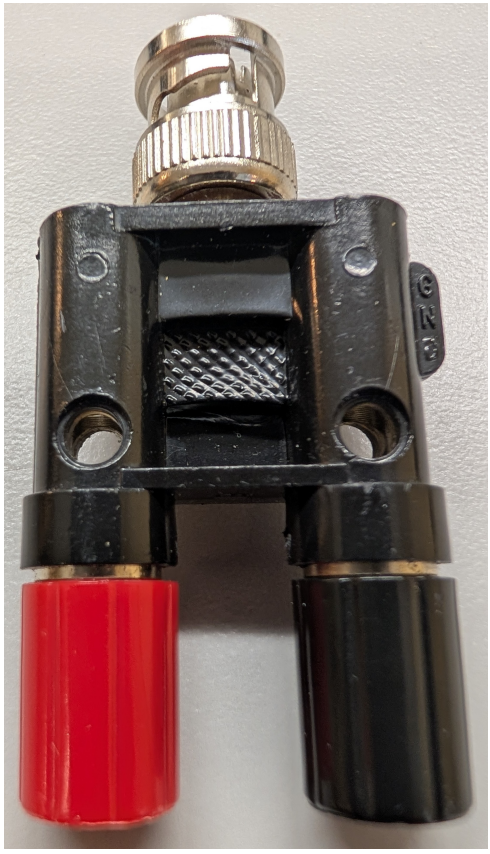


Figure 9.5.11 Male BNC to femail banana adapter.

Figure 9.5.12 BNC tee

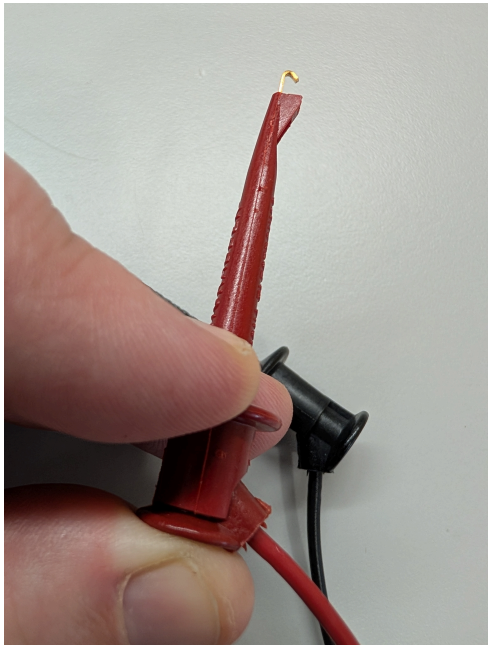


Figure 9.5.13 Adapter that converts female BNC connector to clip leads.

Chapter 10

Data Analysis with Python

10.1 Loading Data Files

Here, we are going to import data from a spreadsheet file, bringing it into Python as a NumPy array.

Example 10.1.1 Import from .csv file using NumPy. If you are starting with data in an .xlsx file, you must first convert it into a .csv file using the 'Save As' feature, selecting 'CSV (Comma Delimited)' format. Then, import your data into Python using code like the following:

```
import numpy as np

# Read filename.csv file into a pandas dataframe. Skip the
    first two lines as they are string headers.
data = np.loadtxt('filename.csv', skiprows=2, delimiter=',');
R = data[:,0];
V = data[:,1];
```

▲

Example 10.1.2 Import from .xlsx file using Pandas package.

```
import pandas as pd
import numpy as np

# Read filename.xlsx file into a pandas dataframe
df = pd.read_excel('filename.xlsx');

# Create a pandas sequence using the 0th column of data. Rows 0
    and 1 contain header info.
df_R = df.iloc[2:,0];

# Create a pandas sequence using the 1st column of data. Rows 0
    and 1 contain header info.
df_V = df.iloc[2:,1];

# Convert pandas sequences to numpy arrays
R = df_R.to_numpy();
V = df_V.to_numpy();
```

▲

10.2 Producing Line Plots

Once data is accessible in Python, data analysis and visualization can be performed. As an example, let's assume that we've conducted an experiment on the circuit shown in Figure 10.2.1.

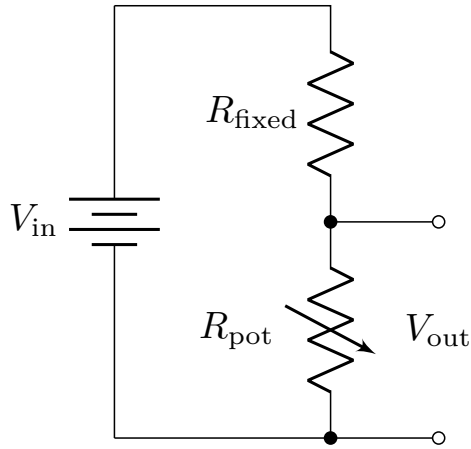


Figure 10.2.1

Here, R_{pot} is a potentiometer (or variable resistor) that the experimenter controls. In this experiment, $V_{\text{in}} = 5.0\text{V}$, $R_{\text{fixed}} = 5\text{k}\Omega$, and we measured V_{out} as

the potentiometer resistance R_{pot} was varied between $0 - 10\text{k}\Omega$. The code below produces a plot in which the data are represented as closed circles. A solid line represents the theoretical prediction

$$V_{\text{out}} = \frac{R_{\text{pot}}}{R_{\text{fixed}} + R_{\text{pot}}} V_{\text{in}},$$

which can be derived using content from Chapter 2.

```
globals().clear() # Clear all variables from previous code runs

import numpy as np
import matplotlib.pyplot as plt

plt.close('all') # Close all open figures

# Data points
Rpot = np.array([1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000,
    9000, 10000]);
Vout = np.array([0.815, 1.332, 1.880, 2.238, 2.614, 2.777,
    2.881, 3.289, 2.974, 3.096]);

# Theory curves
Vin = 5.0 # Volts
Rfixed = 5000 # Ohms
Rth = np.arange(10000);
Vth = Vin*Rth/(Rfixed+Rth);

# Plotting routine
plt.plot(Rpot, Vout, 'o', label='Data');
plt.plot(Rth, Vth, label='Theory');
plt.xlim(0,10500);
plt.ylim(0,3.5);
plt.xlabel(r'Resistance␣($\Omega$)')
plt.ylabel(r'$V_\text{out}$␣(Volts)')
plt.legend();
plt.show();
```

10.3 Fitting a Curve to Data

Frequently, curve fitting will be used to extract results from experimental data. In this section, we demonstrate use of `scipy.optimize.curve_fit`, the curve fitting function provided within SciPy.

Let's illustrate this process through an example. Assume that we have conducted an experiment in which we investigated the behavior of a voltage divider circuit shown in Figure 10.2.1 where V_{in} and R_{fixed} are treated as being unknown. We collect measurements of V_{out} as we vary and measure R_{pot}. A curve fitting process can then be employed to find V_{in} and R_{fixed} from fit parameters, where

$$V_{\text{out}} = \frac{R_{\text{pot}}}{R + R_{\text{pot}}} \tag{10.3.1}$$

is our fitting function. Here, an executable routine is provided that uses the `curve_fit` function from the `scipy.optimize` module to perform a curve fit to data, with a code break-down with explanation following. When examining the results provided by this curve fitting routine, note that values of $V_{\text{in}} = 5.0\text{V}$ and $R_{\text{fixed}} = 5\text{k}\Omega$ were used when generating the synthetic data in this example.

```
# FULL ROUTINE

# Clear all variables from previous code runs
globals().clear()

# Import functions for use in this program
import numpy as np
import scipy.optimize as opt
import matplotlib.pyplot as plt

# Close all open figures
plt.close('all')

# Define our fitting function
def func(R, Vin, Rfix):
    out = Vin * R/(Rfix + R)
    return out

# Create arrays containing data from experiment
Rpot = np.array([1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000,
    9000, 10000]);
Vout = np.array([0.815, 1.332, 1.880, 2.238, 2.614, 2.777,
    2.881, 3.289, 2.974, 3.096]);

# Perform curve fitting
fit, fit_cov = opt.curve_fit(func, Rpot, Vout);

# Calculate standard deviation on fit parameters (associated
    with data scatter)
sigma=np.sqrt(np.diag(fit_cov));
```

```
# Print values and uncertainties for fit parameters
print('Vin={0:.2f}V +/- {1:.2f}V and Rfixed={2:.2f}ohms +/-
    {3:.2f}ohms'.format(fit[0],sigma[0],fit[1], sigma[1]))

# Create arrays containing values for the best fit curve (useful
    for plotting best fit curve)
Rpot_fit=np.linspace(0,10000, 100);
Vout_fit=fit[0]*Rpot_fit/(fit[1]+Rpot_fit);

# Plot experiment data and best fit curve on the same axes
plt.plot(Rpot, Vout, 'o', label='Data');
plt.plot(Rpot_fit, Vout_fit, label='Fit')
plt.xlim(0,10500);
plt.ylim(0,3.5);
plt.xlabel(r'Resistance ($\Omega$)')
plt.ylabel(r'$V_\text{out}$')
plt.legend();
plt.show();
```

Breakdown of curve fitting routine. Let's take a closer look at the pieces that go into this Python routine. We start by initializing our variable and function space.

```
# Clear all variables from previous code runs
globals().clear()

# Import functions for use in this program
import numpy as np
import scipy.optimize as opt
import matplotlib.pyplot as plt

# Close all open figures
plt.close('all')
```

We ensure that all variables in the variable-space are cleared out and don't carry over from other code boxes in the textbook. We then import functions from NumPy, SciPy, and MatPlotLib.pyplot packages that we'll wish to use. Finally, we close all plot windows that may exist from other instances of running code in this textbook.

The `curve_fit` function that we intend to use must be given information about the functional form of our fitting equation. This information will be provided by a function `func` that we define based on Equation (10.3.1). Once defined, `func` will be an input provided to the `curve_fit` function.

```
# Define our fitting function
def func(R, Vin, Rfix):
    out = Vin * R/(Rfix + R)
    return out
```

The first line of our function definition always begins with `def`, followed by the name of the function we're defining and its arguments. The `curve_fit` function expects the first argument of `func` to represent our data's independent variable (often the horizontal axis of a plot). In this example, `R` will represent R_{pot}. Any additional arguments specified for `func` represent fitting parameters that will be determined by the curve fitting process. In this example, these fit parameters `Vin` and `Rfix` will represent V_{in} and R_{fixed} respectively. The next line(s) contain the actual functional form for our fitting function which will be dictated by Equation (10.3.1) in terms of the argument names above. The `out` variable represents our data's dependent variable (often the vertical axis of a plot) and represents V_{out} in this example. The final line of the function definition for `func` will return to `curve_fit` a value of `out` based on the argument values that `curve_fit` provided to `func`.

The next lines of code represent the introduction of experimental data $(R_{\text{pot}}, V_{\text{out}})$ into the routine. There are only ten data points in this example, so it is not too onerous to enter the data by hand directly into the routine. For larger data sets, it is instead recommended that data import routines similar to those shown in Section 10.1 be used.

```
# Create arrays containing data from experiment
Rpot = np.array([1000, 2000, 3000, 4000, 5000, 6000, 7000,
    8000, 9000, 10000]);
Vout = np.array([0.815, 1.332, 1.880, 2.238, 2.614, 2.777,
    2.881, 3.289, 2.974, 3.096]);
```

We are now ready to perform a curve fit to our data. The `curve_fit` function takes as inputs the name of the fitting function, the variable name for the independent variable, and the variable name for the dependent variable. Further inputs and options are available but optional, and can be found in official documentation for SciPy.

```
# Perform curve fitting
fit, fit_cov = opt.curve_fit(func, Rpot, Vout);
```

Here, we provide the name of our defined function `func`, the name of our independent variable data `Rpot`, and the name of our dependent variable data `Vout`. The `curve_fit` function will then return its estimate for values of the fit parameters as an array of values in `fit` and something called the covariance as an array of values in `fit_cov`. A full description of the meaning of the covariance matrix can be found in data analysis or statistics texts. For our purposes, the covariance can be used to

provide estimates of the one standard deviation uncertainty (arising from scatter in the data) for each of the fit parameters stored in `fit`.

```
# Calculate standard deviation on fit parameters (associated
    with data scatter)
sigma=np.sqrt(np.diag(fit_cov));
```

Estimates for the fit parameters and data-scatter uncertainties are printed for the user.

```
# Print values and uncertainties for fit parameters
print('Vin={0:.2f}V␣+/-␣{1:.2f}V␣and␣Rfixed={2:.2f}ohms␣+/-␣
    {3:.2f}ohms'.format(fit[0],sigma[0],fit[1], sigma[1]))
```

In this print command, the structure of the string that will be displayed is provided first, followed by values taken from variables present in our program. As an example, the portion of the command reading `Rfixed={2:.2f}ohms` indicates that the spot enclosed by the curly brackets should be replaced by the second element (since the number before the colon is a '2') in the `format` attribute (`fit[1]` in this case). The ':.2f' indicates that the number placed in this spot should be a floating point number (because of the 'f') that is truncated two digits after the decimal point (because of the ':.2'). To better demonstrate this functionality, an alternative set of print commands, shown here, could have been used to split the statement of results into two lines of text.

```
# Print values and uncertainties for fit parameters
print('Vin={0:.2f}V␣+/-␣{1:.2f}V␣'.format(fit[0],sigma[0]))
print('Rfixed={0:.2f}ohms␣+/-␣{1:.2f}ohms'.format(fit[1],
    sigma[1]))
```

To produce a theory curve based on these fit parameters, we first generate an array of values for R_{pot} associated with our theoretical prediction, and then evaluate a theoretical prediction for V_{out} for each of these R_{pot} values.

```
# Create arrays containing values for the best fit curve
Rpot_fit=np.linspace(0,10000, 100);
Vout_fit=fit[0]*Rpot_fit/(fit[1]+Rpot_fit);
```

Here, `Rpot_fit` is an array containing 100 values spaced linearly in the range (0,10000). These values are plugged into Equation (10.3.1) to generate an array of values for `Vout_fit` .

The remainder of the code contains plotting commands along the lines of those discussed in Section 10.2.

```
# Plot experiment data and best fit curve on the same axes
plt.plot(Rpot, Vout, 'o', label='Data');
plt.plot(Rpot_fit, Vout_fit, label='Fit')
plt.xlim(0,10500);
```

```
plt.ylim(0,3.5);
plt.xlabel(r'Resistance␣($\Omega$)')
plt.ylabel(r'$V_\text{out}$')
plt.legend();
plt.show();
```

10.4 Propagating Uncertainties Through Curve Fits

When conducting experiments, any data collected will have some uncertainty associated with it. These uncertainties will naturally give rise to uncertainties in any subsequent curve fitting results. Here, we will explain the Monte Carlo technique for estimating uncertainties for fit parameters.

The Monte Carlo technique uses repeated curve fits on synthetic data to generate a distribution of values for each fit parameter. Fit parameter uncertainties can be found from these distributions. The synthetic data that is used in these repeated curve fits are generated using the original data with their associated uncertainties.

We will use [THIS SECTION REMAINS IN DEVELOPMENT]

Chapter 11

Lab Activities

In this chapter, laboratory exercises are provided to accompany the content from Part I.

11.1 Lab: Voltage Divider

Objective: Through the course of this lab, you will familiarize yourself with the operation of a prototyping board and its various components (breadboard, voltage sources, potentiometers, and switches) and with resistor color codes. You will build a DC voltage divider circuit using this prototyping board and will use a handheld digital multimeter (DMM) to measure properties of circuit behavior.

Equipment: Prototyping board, digital multimeter (DMM) with banana cables and alligator clips, 1kΩ resistor, wires.

THEORY. Using content from class, derive an expression for the ratio $V_{\text{out}}/V_{\text{in}}$ for the voltage divider circuit in Figure 11.1.1.

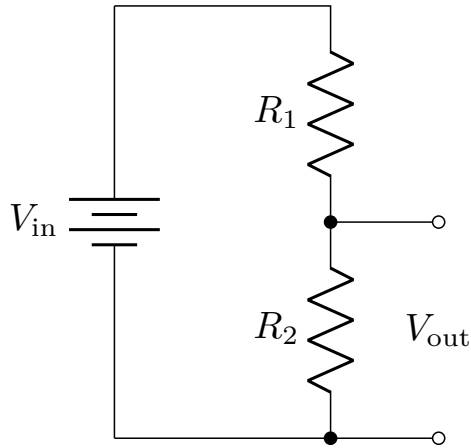


Figure 11.1.1

INTRODUCTION TO EQUIPMENT. For the following steps, explain in your lab notebook the behavior of the various components that you test.

1. With the prototyping board turned off, use the DMM in resistance mode to verify the internal connections of the breadboard. In other words, verify that the resistance between holes is 0Ω if they are supposed to be connected and 0.L (essentially, out of range) if the holes are not supposed to be connected.

2. With the prototyping board still turned off, set up the DMM to verify the +5V row (with respect to the board's ground). Once the DMM is physically connected and set to voltage mode, only then turn the prototyping board on and verify the expected voltage behavior. Specifically, confirm that the 5V row is indeed 5V above the board common.

3. Turn the prototyping board off, change the DMM connections, and verify the other power strips. ALWAYS TURN OFF THE POWER WHEN CHANGING CONNECTIONS!

4. Use the DMM to verify the behavior of both on-board potentiometers. Do behaviors deviate from expected behaviors in any way? Describe how the available terminals correspond to parts of the potentiometers' designs.

5. Use the DMM to verify the SPDT switch behavior.

6. You've been given five different resistors. Use the color codes to determine their specified resistance values, and then verify these values using your DMM. Provide all of these predictions, results, and corrections in your notebook. Remember, this isn't about getting things right immediately. The lab notebook is a journal of your activities. Record your mistakes, corrections, and notes about what you learned.

Discuss your findings with your professor prior to moving on to the next section.

EXPERIMENT. In this section, you will construct the DC voltage divider pictured below using the prototyping board, one fixed resistor, the 10kΩ on-board potentiometer, and a SPDT switch.

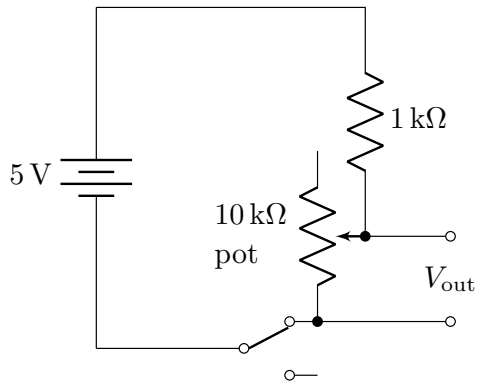


Figure 11.1.2

1. With the prototyping board turned off, construct your circuit, remembering the following:

    - Make use of the long columns of holes to distribute voltages to convenient locations on your board.

- Remember the internal wiring of the prototyping boards. Don't accidentally 'short' your resistor or potentiometer.
- Don't use long wires when short wires will do. Try to keep your circuit organized and use efficient placement of wires. A neat and organized circuit is much easier to troubleshoot!

Verify your circuit and discuss the next steps with your professor before moving on.

2. You must measure the ratio $V_{\text{out}}/V_{\text{in}}$ as a function of the variable resistance R. Use the DMM to first measure the actual V_{in} value (with the prototyping board turned on). Then, with the prototyping board off again, move the DMM connections to measure the variable resistance. Please note the following:

   - When measuring the resistance of the variable resistor, ensure that the multimeter dial is correctly set and that the SPDT is in the position that creates a break in the circuit.
   - When measuring V_{out}, ensure that the multimeter dial is correctly set and that the SPDT is in the position that creates a closed connection in this segment of the circuit.
   - You will need to repeat the previous two steps, measuring potentiometer resistance and V_{out} for each of *ten* potentiometer resistance values. Ensure that your ten potentiometer resistances are chosen to provide adequate coverage of the full potentiometer range.
   - Record your results in a spreadsheet table and ensure this table is included in your lab notebook.

3. Use Python to graph the results of your measurements ($V_{\text{out}}/V_{\text{in}}$ vs R). Use symbols (not a line) to represent your data points. On the same axes, generate a theoretical plot based on the predicted behavior of this circuit and represent this theory curve using a solid line plot. Also provide a legend.

   Question: Are there any deviations from expected behavior at low or high resistances? If so, discuss this in your lab notebook.

4. Extension: In your lab notebook, discuss how you can make a voltage divider and conduct this experiment using only a single potentiometer (without the fixed resistor that we used in our circuit). Include a circuit diagram. Also discuss common places in your everyday life where you may find this type of circuit used.

5. Ensure that all required parts are included in your lab notebook. See for details.

11.2 Lab: Oscilloscope Primer

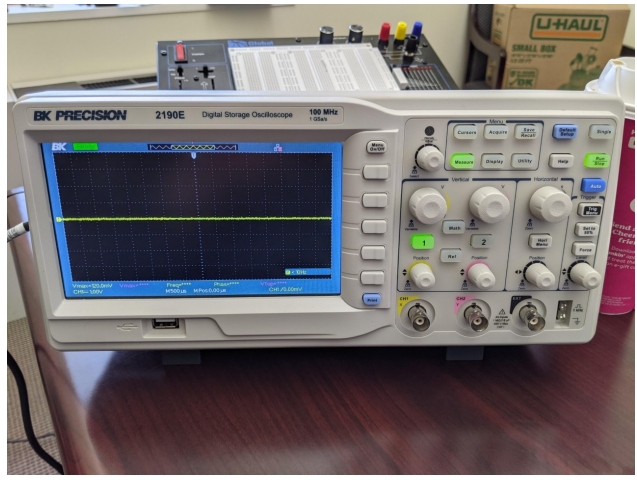
Objective: Through the course of this lab, you will familiarize yourself with the operation of a the BK Precision 2190E Digital Oscilloscope (or scope). You will use the scope to measure an AC voltage signal from the internal function generator of the scope, and from an external function generator. Our goal will be to understand how to display a signal on the scope, measure frequency and amplitude values, and trigger the scope properly by adjusting the triggering threshold and other trigger settings.

Equipment: Function generator, oscilloscope, oscilloscope probe, BNC coaxial cable.

Lab Notebook: In your theory section of the report, you should discuss any functions of the oscilloscope that we use today and their purpose. In the data section, you should include the signal data and photos asked for throughout the lab. You do not need a procedure, results or conclusion section for this lab. Otherwise, please follow the lab notebook guidelines provided in .

THE OSCILLOSCOPE INTERFACE. The oscilloscope interface seems quite complicated at first, but in order to measure one signal vs. time we will only need a few functions.

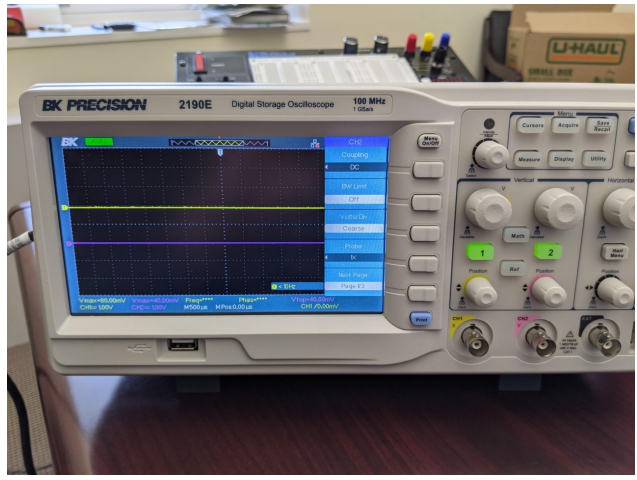
1. Turn on your scope using the power button on the top of the case. Wait for it to start and then hit the "Default Setup" button- it is blue in the upper right of the front face. You should see the following display:



2. The scope has the ability to display two signals at once. Turn on the channel two display by hitting the "2" button. A pink line, representing the voltage

plugged into the channel 2 BNC input connector, should now appear on top of Channel 1's yellow line. (Both voltages are zero at the moment as nothing is plugged into either channel.)

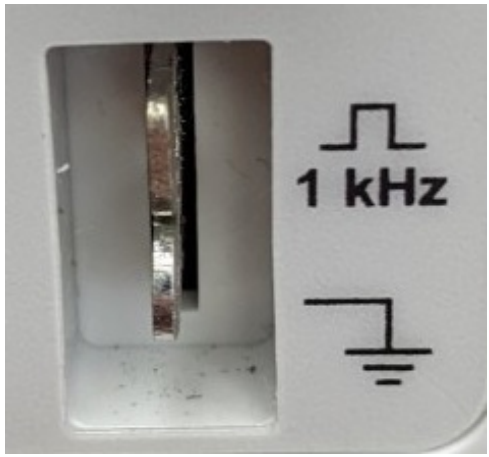
3. You can use the ***small knobs*** below the "1" and "2" buttons to vertically displace the signal lines on the display. Try it and see if you can move channel 1 upwards and channel two downwards as shown below:



4. Turn off channel 2 by hitting the "2" button. Re-center signal 1 on the x-axis. Turn off the channel 2 menu that appeared on the right of the screen by hitting the "Menu On/Off" button.

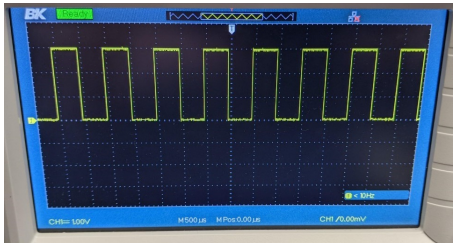
To explore more of the functions of the scope we will need to look at a signal.

MEASURING THE BUILT-IN TEST SIGNAL. The oscilloscope has a built in test signal. It is a 1KHz square wave, and can be measured by connecting an **oscilloscope probe** to the two metal clips in the lower right of the front face.



1. Find your oscilloscope probe. Plug it into the **Channel 1** BNC connector on the oscilloscope. ***Set the attenuation switch on the probe to 1X.*** (10X means the voltage measured in attenuated by 10- useful for high voltage projects.)

2. The probe has two clips. The side **alligator clip** must be connected to ground. The center **hook clip** is connected to the point in a circuit where you want to measure the voltage. Find the metal leads for the test signal (shown above in the figure). Connect the oscilloscope probe to the test signal clips, clipping the probe ground to the test signal ground and the probe's center hook to the test signal clip that is marked as a 1 kHz square wave. Channel 1 should display the test signal as shown below:



3. Take a picture of your channel 1 display for your lab notebook.

4. You should see the oscilloscope's vertical and horizontal scales on the display screen. The CH1 = 1.00V marker means that the vertical grid spacing (between horizontal dotted lines) represents a change of 1.00 V. The M = 500 s marker means there is a 500 s spacing between each vertical dotted line. ***What is the amplitude of this square wave? Include this value in your report.***

5. Try to change the vertical (voltage) scale by using the large CH 1 knob on the front panel. Also try to change the horizontal (time) scale by using the large knob marked 'Horizontal' on the front panel.

6. Change the horizontal scale to display the wave on a 100 s timescale and a 2V vertical scale. ***Take a picture of the output to include in your report.***

7. Return the waveform to its default view. You can do this either by reversing your changes made above or, even easier, by just hitting the **Default Setup** button again.

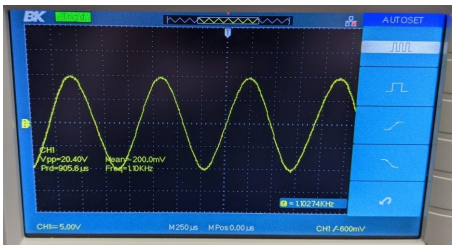
MEASURING A SINE WAVE. Now we will use an external function generator to produce an AC sine wave signal for us to measure.

1. With the function generator turned off, disconnect the oscilloscope probe.

2. Use a BNC cable to connect the function generator ***output*** to the oscilloscope Channel 1 input.

3. Turn on the function generator. We want a 1V amplitude 1kHz sine wave signal to measure. Press the 'Amp' menu button and type in 1.0 and then press V_{pp}. Then, click **Freq** and type 1.0 and click the **kHz** button. At this point, there is still no signal being produced. Press the **Output** button to start producing the desired signal. The **Output** button should now be glowing, indicating that the function generator output is active.

4. The display should show a sine wave, but probably too zoomed in to see the whole wave. Hit the **Auto** button to autoscale the display to the signal. The following image shows the correctly auto-scaled sine wave. Click the 'Measure' button on the oscilloscope and select parameters that you might want to measure for each channel, including the Peak-to-Peak Amplitude (V_{pp}) and frequency of the signal. ***What are the values of these parameters? Record values in your notebook along with an image of the oscilloscope display.***



5. To make sure you understand the different settings of the function generator and scope, do the following:

(a) Set the function generator to display a 500Hz signal with a 20V pk-pk amplitude

(b) Display the signal on the scope with a 10V vertical scale and a 1ms timescale.

(c) Take a picture of this output and include with your report.

6. Note that you can use the **position** knob to move the signal up and down on the screen. This is purely moving the 0V reference position on the screen. The position of this reference point is indicated by the yellow arrow with a '1' in it.

7. Also note that you can adjust the vertical **Zoom**, or scaling for each channel independently. As you adjust the CH1 knob, you can see that the value 'CH1=____V' changes in the bottom left corner. This number indicates the change in voltage vertically across one of the small boxes in the grid displayed on screen for the Channel 1 signal. You can also adjust the horizontal scaling in the same way, with the 'M____ms' number providing the change in time horizontally across one of the small boxes in your grid. You will often need to adjust these appropriately to see your desired signal. Note: You should always try to maximize the vertical size of each signal that you are trying to measure to get the most precise measurements using the built-in 'measure' feature and using the cursors (explained below). If you have a periodic signal, it is often desirable to have 3-5 periods displayed in order to use the built-in 'measure' feature.

USING THE CURSORS. With complicated signals, the auto measured output for amplitude and frequency cannot necessarily be trusted. In these cases, it's better to do manual measurements with the oscilloscope's cursor functionality.

1. Set up the function generator to produce a 1kHz, 20V pk-pk sine wave.

2. Display this signal on the scope, auto-scaled. Hit the **Run/Stop** button to freeze the signal for better analysis. The button should light up red when data acquisition is stopped. (Don't forget it's stopped and that you're no longer measuring real time data!)

3. To manually measure the period of the wave, we will use vertical cursors. Hit the **Cursors** button on the top center of the scope control panel. In the menu that appears, change the mode to **Manual**. Then change the type to **Time** and ensure that the source is **CH1**. You should see two dotted lines appear on the display.

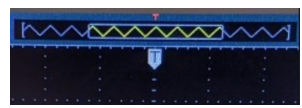
4. You can move the position of the lines by selecting cursor A or B in the screen menu and then turning the small knob in the upper left of the control panel. Try to move cursor A onto a peak of the wave. Do the same for cursor B.

   NOTE: If you can only see one cursor line on the screen, it might be that the other is off the edge of the display. Try selecting the invisible cursor and moving the knob until it appears in view. Sometimes you might need to zoom out in time or voltage to find the cursors.

5. The cursors appear with readouts that tell you the time location for each individual cursor (with respect to the trigger, which will be discussed next), the time interval between the two cursors, and the associated frequency. Do these values make sense? ***Record these values and take a photo of the screen for your report.***

6. Switch the cursor type to Voltage to get horizontal cursors. Place them so that you can measure the pk-pk voltage amplitude. ***Record the pk-pk voltage, and take a photo for your report.***

Triggering. The oscilloscope needs to know where the wave starts in order to display a nice, steady display output. This process of finding the start of the waveform is called **triggering**. The trigger determines the location of your signal with respect to the scope's definition of t=0s. The blue T in the image below indicates the time location of the trigger.



Using the trigger menu and cursor wheel, we can set the trigger level at whatever voltage threshold we want. When the experimental signal crosses the trigger level that we set, the oscilloscope calls that time t=0s and all other times are measured with respect to that time. The oscilloscope can be told to trigger off of CH1, CH2, or even other external signals. It can also be told to trigger when the signal rises through the trigger level or falls through the trigger level.

For a simple AC sine wave, triggering is pretty easy to set up with the Auto-scale feature, but for other signals you may need to manually adjust the trigger settings.

1. Make sure the scope is in RUN mode, not STOP mode.

2. With the AC sine signal displayed, find the **Trig Menu** button on the control panel (on the right, under Auto). Hit this button to bring up the trigger settings menu.

3. In this menu, we say how we want the scope to find the beginning of the signal. By default, Type is set to **Edge**, Source is set to **CH1**, and Slope is set to

Up. This means we are telling the scope to search for an edge with a rising slope on channel 1, and place everything such that that edge is at t=0s.

4. Try to set the slope type to **Down**. Discuss in your report how the display changes.

5. We also need to tell the scope how big of a voltage actually counts as a signal. This is especially important if we have a noisy experiment. We don't want to accidentally trigger on noise instead of our real signal. This function is controlled by the **Trigger Level** (or just **Level**) knob on the control panel (lower right).

6. Try adjusting the trigger level and you should see the level line appear on the display. Any voltage signal with a peak below this line will not be considered a signal. Try moving the level to above the sine wave peak. Describe what happens in your report.

7. Adjust the level knob to manually regain good triggering. As a rule of thumb, it is good to have the trigger level around 30% below the peak of your signal.

The location of the trigger's t=0s can be adjusted left or right on the screen using the **Horizontal Position** knob. ***A word of advice: if the trigger time is offset from the middle of the screen and you change the time resolution on the scope, the trigger indicator may be outside of the window of view on the display. This can sometimes cause confusion if you do not realize that you are looking at a portion of the signal well after (or before) the trigger time.***

In more advanced contexts one can trigger all oscilloscope channels off of Channel 1, Channel 2, or off of a third external trigger. We will discuss these options later on in the course, as needed.

11.3 Lab: RC High-Pass Filter

Objective: To use the oscilloscope to explore the attenuation, phase-shifting, and transient behavior of an RC high pass filter circuit. This experiment involves resistance, capacitance, filtering, attenuation, half-power frequencies, time constants, phase shifts, and oscilloscope and function generator operation.

Equipment: Proto-board, digital oscilloscope, oscilloscope probes, function generator, resistor, capacitor.

RC HIGH-PASS FILTER THEORY. The aim of this experiment is to test the reliability of the theoretical predictions for a RC high pass filter circuit like the one shown in Figure 11.3.1.

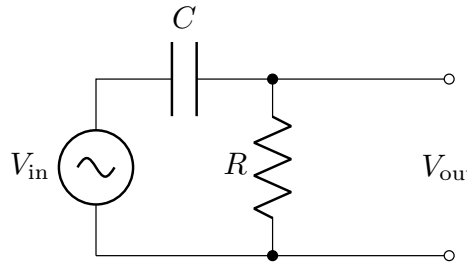


Figure 11.3.1 RC high-pass filter circuit. Use $C = 47\text{nF}$ and $R = 2\text{k}\Omega$.

The prediction for the ratio of the amplitudes of the output voltage to the input voltage is

$$G_V = \frac{V_{\text{out}}}{V_{\text{in}}} = \frac{1}{\sqrt{1 + (f_0/f)^2}}.$$

The half-power frequency is given by $f_0 = 1/2RC$. The theoretical prediction for the phase shift between output and input voltages is

$$\phi = \phi_{\text{out}} - \phi_{\text{in}} = \tan^{-1}(f_0/f).$$

RC CIRCUIT EXPERIMENTS.

1. ***Make sure all equipment is turned off*** before proceeding to construct the high pass filter circuit shown in Fig. 1 on your prototyping board.

2. Attach a BNC tee to split the function generator output.

3. Connect one side of the BNC tee to the circuit input using clip connectors, and connect the other side of the tee to the CH1 input of the oscilloscope.

4. Connect the circuit output to the CH2 input on the scope. Use oscilloscope probes to make connections between the oscilloscope and the circuit. Make sure both probes are set to '×1' attenuation.

5. Turn on the oscilloscope and set it to display CH1 and CH2.

6. ***Have your professor check your circuit before turning on the prototyping board and function generator.***

7. Program the function generator to produce a sine wave with a peak-to-peak amplitude of approximately 2.0 volts and a frequency of 100 Hz.

8. Trigger the oscilloscope on CH1 (AUTO, Rising Slope). Set the trigger level appropriately based on the CH1 amplitude.

9. Proceed to acquire data that lets you test the theoretical predictions given by the two equations above.

   Your data should consist of pairs of voltage values (V_{in},V_{out}) as well as phase shifts acquired at 24 frequencies (100 Hz, 200 Hz, ..., 900 Hz, 1 kHz, 2 kHz, ... , 9 kHz, 10 kHz, 20 kHz, ... 50 kHz). Don't forget to record your frequencies.

   Tip: You are allowed to use the 'Measure' functionality to measure V_{in}, V_{out}, f_{CH1}, and the phase ϕ between channels 1 and 2. Given that this functionality can sometimes provide poor results, you should manually check your measurements occasionally using the cursors (for instance, manually check at 100 Hz, 1 kHz, and 10 kHz).

   Tip: To measure the phase shift, we recommend 'zooming' in on the zero-crossings and measuring the time delay between the signals on the two channels (using the cursor function on the oscilloscope). From the frequency (or period) and the time delay, you can calculate the phase shift (in radians or degrees).

   Tip: When advancing to a new frequency, find the frequency you want and wait for the signal to stabilize before recording values. If conducting a manual measurement of parameters using cursors, you may want to press RUN/STOP to temporarily freeze the oscilloscope. This allows you to have a stable display while you are writing down your results or using your cursors to find experimental values. When complete, press RUN/STOP again to unfreeze the oscilloscope.

   Q: Explain how you determined the phase shift manually from experimental measurements using cursors. Provide an example of your work in your lab notebook, including a picture of your oscilloscope display.

10. Experimentally determine the value of f_0 (to the nearest Hz) and proceed to record values for $(V_{\text{out}}, V_{\text{in}}, \phi)$ at $f = f_0$. Include this data point as part of the data set that you recorded earlier.

    Hint: What value should ϕ have when $f = f_0$?

    Q: Explain the process you used to find f_0.

    Q: Include in your notebook an image of your oscilloscope display at your best estimate of f_0?

    Q: What is your experimentally-determined value of f_0?

11. Plot $V_{\text{out}}/V_{\text{in}}$ vs f using symbols (not a line). Decide whether to plot your data using linear axes or a semi-log plot (with a logarithmic frequency axis).

    Q: How did you decide whether to use linear or semi-log plot axes? Think about the information you are trying to present and which option best highlights the behavior.

12. Proceed to perform a curve fit to your experimental data (based on the theoretical gain equation) with f_0 as a fit parameter. (See Section 10.3 for guidance.)

    Q: What is the fitted value of f_0.

13. Overplot a theoretical curve (as a line) using the fitted value of f_0. Make sure that you provide a legend on the plot.

14. Plot phase shift vs frequency using the same type of axes (linear or semi-log) that you used in step 11. Proceed to perform a curve fit to your experimental data (based on the theoretical phase equation) with f_0 as a fit parameter. (See Section 10.3 for guidance.) Overplot your curve fit as a line.

    Q: What is the fitted value of f_0.

Note on the lab notebook discussion section: Include all of the standard topics in your discussion section, but also make sure to address the following points:

- You found f_0 in multiple different ways (theory, experiment, curve fit). As part of your discussion, compare your f_0 values and discuss possible reasons for any discrepancies.
- Examine the visual agreemenet between data and curve fit plots and discuss the validity of our theoretical expressions.

11.4 Lab: RLC Resonant Circuit

Objective: In your lab notebook, write a few sentences stating your objective in conducting this laboratory exrercise. Consider the following questions:

- What kind of circuit am I examining?
- What is new about this circuit as compared to circuits previously studied?
- What theoretical predictions do you have regarding circuit behavior?
- What effects do you expect to observe?

Your answers should be specific to the type of circuit you are examining, but do not discuss specific component values.

Equipment: Proto-board, digital oscilloscope, resistor, capacitor, inductor.

RLC RESONANT CIRCUIT THEORY. You will be studying properties of the circuit pictured in Figure 11.4.1.

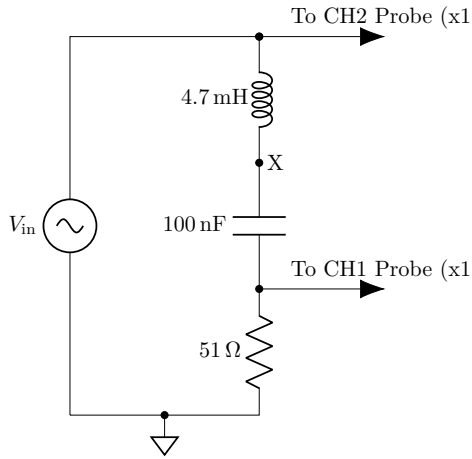


Figure 11.4.1

We developed the theory for this circuit in Example 3.7.8. In your lab notebook, summarize the main theoretical results/circuit behaviors that you expect to observe today. To guide you, here are a few questions you may wish to address:

- What voltage gain and phase behaviors do you expect to observe as $\omega \to 0$?
- What voltage gain and phase behaviors do you expect to observe as $\omega \to \infty$?

- At what frequency f_0 do you expect the highest voltage gain? Express your answer as a function of variables L and C and show your work. What do you expect voltage gain and phase to be at frequency f_0?

These are just a few examples. Are there any other important qualities that could be tested experimentally?

RLC CIRCUIT EXPERIMENTS. In your lab notebook, describe a procedure that you intend to follow to make measurements of voltage gain G_V and the phase difference between output and input voltages $\phi = \phi_{\text{out}} - \phi_{\text{in}}$. Here are a few questions you may wish to consider:

- How will you confirm the values associated with your circuit components? ***Note:*** You will be unable to verify the inductance L, but you should be able to measure values for other components.
- In this experiment, you will need to measure V_{in}, V_{out}, and $\phi = \phi_{\text{out}} - \phi_{\text{in}}$ for many values of frequency increasing from 200~Hz.
- What equipment do you intend to use as a voltage source?
- What equipment do you intend to use to measure the desired quantities?
- If using the oscilloscope, describe the functionality that will be used (e.g. cursors, 'MEASURE' functions, etc.). Why did you make your choice?
- Plan to display your results on a plot with a logarithmic frequency axis and a linear Gain and phase axis using the `semilogx` command instead of the `plot` command.
- At how many frequencies will you take data? How are you going to determine measurement spacing?
    - When determining your answers to this question, consider that you likely don't need a high density of data points in regions where you expect little variation in G_V and ϕ, but you will want a higher density of data points in regions where you expect high variation in these quantities.
    - To determine frequency ranges associated with high and low variation in gain and phase, it can be helpful to perform a quick informal scan of experimental parameters to find the experimental resonant frequency while not actually recording gain and phase measurements. With this information, you can better determine the distribution of frequencies at which to record data.
    - Since you will be using a logarithmic horizontal axis, your frequency selections should take this into account.

- How will you identify when you are at or near the resonant frequency?
- When comparing to theoretical predictions, use the equations that result from the analysis in Example 3.7.8 to generate theoretical curves for G_V and ϕ vs frequency, plotting the theoretical curves as a different colored line on the same axes as the experimental data in each case.

Build your circuit on the prototyping board. When laying out your circuit, it can be useful to build it in a way that is clean and similar to the circuit diagram when possible. See Figure 11.4.2 for an illustration of how one may do this.

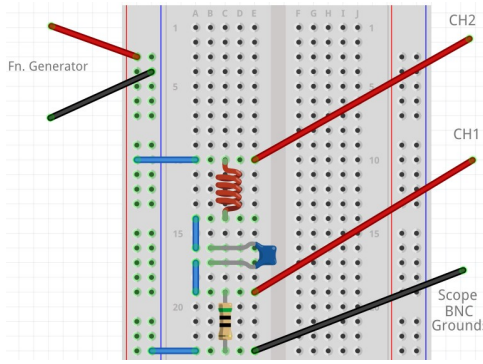


Figure 11.4.2 Suggested prototyping board layout created using Fritzing. This deliberate placement of CH1 and CH2 connections means that the oscilloscopes phase measurement for `CH1-CH2` will correspond to a measurement of $\phi_{\text{out}} - \phi_{\text{in}}$ as we desire.

As you proceed to conduct your experiment, discuss any changes you have to make to your procedure and the reason for those changes. Additionally, if you observe some behavior that you were not expecting or that confuses you, make a note of it in your notebook.

Reminder on oscilloscope operation: Before recording any measurement values into your notebook, you must first

1. center each signal vertically on the oscilloscope display and adjust the vertical scaling on each channel so that the signals are magnified as much as possible on the screen.
2. adjust the timebase so that there are 3-5 cycles displayed on the oscilloscope screen.

You should ensure these optimal settings are satisfied ***for each frequency*** at which you take data.

Tip: Your starting input voltage should be sinusoidal with amplitude $V_{\text{in},0} = 10\text{V}_{\text{pp}}$ and an initial frequency of $f = 200\text{Hz}$.

After completing the experiment above, return to the resonance frequency that you observed and do the following:

- Make sure you always turn off the circuit power when moving probes or changing components!
- Move the oscilloscope Channel 2 probe to the node marked X on the circuit diagram.
- Make sure that both channels have the same VOLTS/DIV settings.
- Press the MATH button (in Channel 1 and 2 controls area) and explore the various calculations that can be accomplished using the functions in this menu.
- Use options in the MATH menu to subtract the Channel 1 signal from the Channel 2 signal.
    - ***Q:*** What are you measuring when you do this?
    - ***Q:*** What do you observe?
    - ***Q:*** Take a photo displaying the `MATH` signal as well as your V_{out} signal and place it in your notebook. What is the peak-to-peak amplitude of the `MATH` signal? What is its phase relative to V_{out}?
- Turn off your circuit. Switch the positions of L and C in your circuit (leaving the Channel 2 probe where it is) and repeat the previous bullet points.
    - ***Q:*** What are you measuring now?
    - ***Q:*** What do you observe?
    - ***Q:*** Take a photo displaying the `MATH` signal as well as your V_{out} signal and place it in your notebook. What is the peak-to-peak amplitude of the `MATH` signal? What is its phase relative to V_{out}?
- ***Q:*** Do your observations make sense to you? How do you expect your observed MATH signals to compare to V_{in} and V_R?

Once you have completed data collection and analysis, write a discussion section that summarizes your results and compares them to your initial theory.

- What options can you think of to compare theory to experiment in this lab?
- What option(s) did you choose for your comparison of theory to experiment? Why did you make the choice that you did?
- Address any discrepancies that you may have seen along with any issues you may have run into. Speculate regarding solutions/explanations related to these discrepancies/issues. This lab has been designed such that there ***will*** be discrepancies between observation and our theory.

- Part of this is because there may be a discrepancy between the quoted inductance value L and the actual inductance of your physical component.
- We must also confront the idea that real inductors do not necessarily behave as nicely as the ideal inductors that we discuss in class. The behavior of a real inductor can be better approximated by an inductance L_{inductor} in series with an internal resistance R_{inductor}, and then that combination in parallel with a parasitic capacitance C_{inductor}.

Include this discussion in your lab notebook.

11.5 Lab: I-V Characteristic for Diodes

Objective: In your lab notebook, write a few sentences stating your objective in conducting this laboratory exrercise. Consider the following questions:

- What kind of circuit(s) or components are you exploring?
- What is new about this circuit as compared to circuits previously studied?
- What theoretical predictions do you have regarding circuit behavior?
- What effects do you expect to observe?

Your answers should be specific to the type of circuit you are examining, but do not discuss specific component values.

Equipment: Proto-board, digital oscilloscope, DMM, 1N4001 silicon diode, 10kΩ resistor.

Introduction. To date we have used only linear circuit elements in lab: resistors, capacitors, and inductors. In this experiment, we take up the diode, a nonlinear device. Diodes provide signal conditioning, voltage regulation, and the conversion of AC voltages and currents into DC. In this experiment, you will measure a silicon diode's "I-V characteristic" using the circuit pictured in Figure 11.5.1.

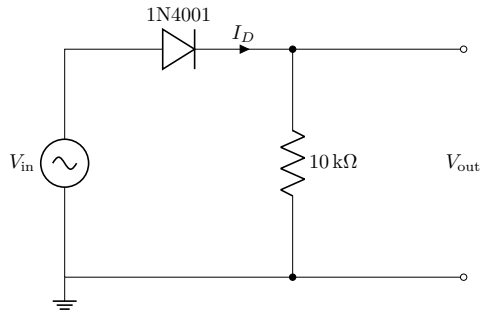


Figure 11.5.1

Before conducting the experiment, discuss the following topics in your lab notebook:

- How does a single diode behave in a circuit?
- What should the I-V characteristic look like? Provide a rough sketch.

EXPERIMENT. In this experiment, you need to collect many data points to construct the I-V characteristic for your diode. You can either do this using a DC supply for your input, collecting point-by-point the diode voltage drop V_D and associated diode current I_D as you vary the DC supply. This process is very slow. So, we will instead use a sinusoidal voltage source as V_{in} and collect time traces for V_D and I_D. We will export this data to a computer and then use Python to generate the I-V characteristic.

1. Unplug your prototyping board. You will not need to plug it in for this experiment.

2. Obtain a 1N4001 silicon diode and test its function with the multimeter set to its diode-testing mode. In this mode, the DMM checks the effective resistance of a diode by attempting to force a positive current out through its '+' or red lead, through the external diode, and back into its '-' or black lead. When the DMM displays a small result such as 0.5V, one may infer that the red lead is connected to the anode (or p-type semiconductor) and the black lead to the cathode (or n-type semiconductor) of the diode, and that the diode is conducting in its 'easy' direction. If, instead, a large off-scale indication of '.OL' is displayed, then the diode is reverse-biased. Note how the diode is marked to indicate which end is the anode.

3. Proceed to construct the circuit shown in Figure 11.5.1.

4. Attach the oscilloscope probes so that CH1 measures V_{in} and CH2 measures V_{out} in Figure 11.5.1.

   Q: What will you need to do in Python to extract V_D from your data?

   Q: What will you need to do in Python to extract I_D from your data?

5. Connect your waveform generator to your circuit and set it to produce a 60Hz sinusoidal signal that varies between $\pm$5V.

6. Save your data to a USB flash drive. The oscilloscope can export data in several formats. For our purposes it is best to save and export the actual digitally sampled data points (as opposed to an image file of the display) so that we can plot, manipulate, and analyze the data using a program like Excel or LoggerPro. This requires that we save the data in channels one and two separately.

   (a) Press the RUN/STOP button to capture the data and to make sure that you are saving the data you want.

   (b) Insert your flash drive into the USB port on the front of the oscilloscope.

(c) Press the SAVE/LOAD button and, from the menu options figure out how to save the CH1 and CH2 data (in separate files) to a spreadsheet format file (such as a .csv file). Make note of the filename for each channel. The instructions depend on the oscilloscope type. Try one of the following:

- SAVE/RECALL → TYPE (CSV), press 'Confirm' and follow the rest of the menu options to save a new file.
- SAVE/RECALL → EXTERNAL STORAGE → NEW → NEW FILE → SAVE AS CSV → SAVE.

7. Turn off your waveform generator and oscilloscope.

8. Load your data into a spreadsheet file (which you can upload to your lab notebook as a file; no printout is necessary). Examine how the data is formatted and determine the meaning of each column of data.

9. Write a program in Python to load your data, extract V_D and I_D from the data you collected, and plot your I-V characteristic. Make sure to label your axes and include units.

   P: Include your I-V characteristic in your lab notebook.

   Q: Discuss similarities and differences between your measured I-V characteristic and the 'cartoon' behavior displayed in Figure 4.2.2.

   Q: Discuss similarities and differences between your measured I-V characteristic and the more realistic behavior shown in Figure 4.2.5.

10. Add to your Python program above the ability to produce a second plot, this time of V_{in}, V_{out}, and V_D vs t on the same axes using different colors for each signal. Ensure that your axes are labeled (with units!) and that you include a legend.

    P: Include your plot of input, output, and diode voltage signals in your notebook.

    Note: This result illustrates the ability of diodes to 'rectify' oscillating voltages, converting symmetric AC signals into signals that, on average, have a DC component. The present circuit is called a **half-wave rectifier**. In this case, the diode acts to turn off current flow for one half of each input cycle. In the next experiment, you will construct a full-wave rectifier that effectively outputs the absolute value of the input voltage.

11. Produce a third plot in Python that shows I_D vs t. Ensure that the time range shown matches the range in the previous voltage vs time plots.

    P: Include your plot of I_D vs t in your notebook.

Provide a discussion in your lab notebook based on guidelines provided in

11.6 Lab: AC to DC Conversion Using Diode Bridge Rectifier

Objective: In your lab notebook, write a few sentences stating your objective in conducting this laboratory exercise. Consider the following questions:

- What kind of circuit(s) or components are you exploring?
- What is new about this circuit as compared to circuits previously studied? Ignore the differential amplifier aspect of the circuit setup for the time being.
- What theoretical predictions do you have regarding circuit behavior?
- What effects do you expect to observe?

Your answers should be specific to the type of circuit you are examining, but do not discuss specific component values.

Equipment: Proto-board, digital oscilloscope, DMM, INA117P differential amplifier chip, 4 x 1N4001 silicon diodes, 3.3kΩ resistor, five capacitors in the range from 1μF to 500μF.

Introduction. The operation of many electronic devices depend on a DC voltage source, but the electricity delivered to households in the United States is 60-Hz AC voltage. Today, we'll explore one method of converting AC voltage to DC voltage using the **diode bridge full-wave rectifier** with a smoothing capacitor as pictured in Figure 11.6.1.

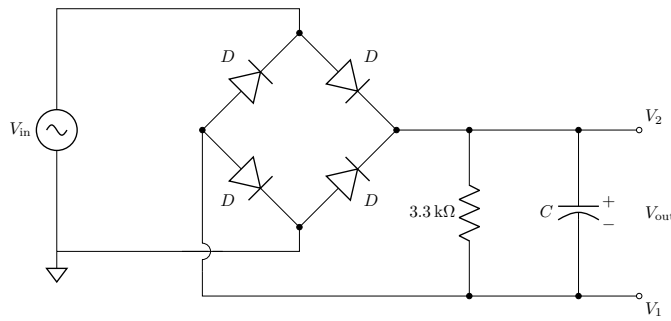


Figure 11.6.1

This circuit converts an AC input voltage to a nearly-constant output voltage, though a small 'ripple' will remain. Larger values of smoothing capacitance will reduce the size of this ripple and thus will improve the quality of the DC voltage output. High-capacitance capacitors are typically more expensive and physically larger than small capacitors, so there is incentive to keep the capacitance as small

as possible. Thus, devices that require DC voltages often specify a maximum ripple size that they can tolerate.

Your goal in this experiment is to predict the optimal smoothing capacitance one would use to ensure a 5% ripple on V_{out}. The ripple size depends on the smoothing capacitance (which you will determine), the input signal frequency (which we will assume to be 120-Hz) and the load resistance provided by the device across the output (here, represented by the 3.3-kΩ resistor).

Before conducting the experiment, be sure to review the behavior of the unsmoothed diode bridge full-wave rectifier circuit in Section 4.5.

NOTE: One downside of the circuit shown in Figure 11.6.1 is that neither V_{out} terminal is connected to ground. This means that we cannot simply connect a single oscilloscope probe across V_{out} since the ground provided by the oscilloscope and the ground provided by the function generator would conflict. We will use a differential amplifier chip based on operational amplifiers that takes V_1 and V_2 as inputs and produces a single output signal $V_{\text{out}} = V_2 - V_1$ referenced to ground. The details of this differential amplifier's operation and the behavior of operational amplifiers in general will be treated later.

EXPERIMENT. Once you've reviewed the above information, complete the following steps.

1. Use your DMM to determine exact values for your resistor and your capacitors.

   Warning: The capacitors that you are using are called 'electrolytic capacitors' and are polarized, meaning that they only work when the negative lead (often marked with a negative sign) is at a lower voltage than the positive lead. When measuring the capacitance with the DMM, connect the COM to the negative capacitor lead.

2. Plug in your prototyping board but keep it turned off for the moment.

3. Set up a DMM to measure the variable V+ (yellow) voltage with respect to Ground (black). Pay attention to which of the four top rows correspond to +V and Ground. Turn on the prototyping board.

4. Adjust the appropriate prototyping board knob so that the +V strip is +15V relative to Ground. Repeat the process above to set -V to be -15V relative to Ground. Do not touch these knobs for the rest of this lab session.

5. Turn off your prototyping board.

6. Construct the circuit shown in Figure 11.6.1 on your prototyping board, but leave the smoothing capacitor out of the circuit. The input voltage V_{in} will be provided by your function generator.

7. Connect the ±15V supplies and Grounds to the INA117P differential amplifier chip as shown in Figure 11.6.2. Note that your instructor should have already inserted the INA117P chip into your prototyping board in the correct orientation.

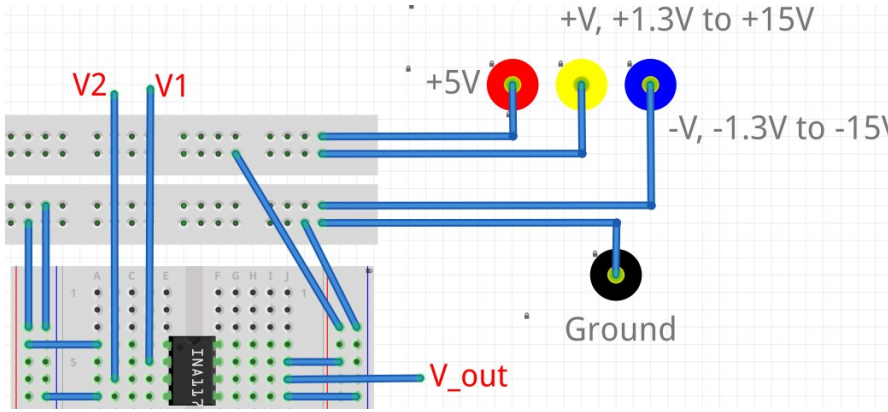


Figure 11.6.2 Differential amplifier setup.

8. Connect points V_1 and V_2 from your diode bridge rectifier circuit to the differential amplifier chip as shown in Figure 11.6.2

9. Set up the oscilloscope to measure V_{in} on CH1 and V_{out} ***from the differential amplifier*** on CH2. Turn on the oscilloscope.

10. Turn on the function generator (but leave the OUTPUT disabled). Set the generator to produce a 120-Hz sinusoidal voltage with a peak amplitude of 5-V. Then, enable the OUTPUT.

11. Adjust the CH1 vertical scale so that the input signal is as large as possible without cutting off any of the signal. Adjust the horizontal timebase such that you can see 3-5 periods of the sinusoidal signal.

12. Use the CURSOR functionality to measure the amplitude of V_{in}.

    Q: Record your V_{in} amplitude in your lab notebook.

13. Enter the CH2 menu and choose DC coupling. Adjust the CH2 vertical scale to be identical to the CH1 vertical scale.

    P: Insert into your lab notebook a photograph of the oscilloscope screen displaying V_{in} and V_{out}.

14. Repeat the following steps for each smoothing capacitor you've been given:

    (a) Disable the function generator output and turn off the prototyping board.

(b) Add the next largest smoothing capacitor to the circuit. If the smoothing capacitor is currently absent from the circuit, add the smallest capacitor.

Warning: Pay attention to the capacitor polarity in the circuit schematic. If the capacitor is wired in backwards, it may be destroyed (sometimes resulting in a puff of smoke or small flame). Remember that $V_2 > V_1$ ***always in this circuit.***

(c) Enable the function generator output and turn on the prototyping board.

(d) Enter the CH2 menu and choose DC coupling. Adjust the CH2 vertical scale to be identical to the CH1 vertical scale.

P: Insert into your lab notebook a photograph of the oscilloscope screen displaying V_{in} and V_{out}.

(e) Enter the CH2 menu and choose AC coupling. This will remove the DC offset and allow you to vertically magnify the display of the AC ripple piece of V_{out}.

(f) Adjust the CH2 vertical scale to maximize the displayed size of the AC ripple signal and use the oscilloscope's ***cursor functionality*** to measure the peak-to-peak amplitude of the AC ripple.

Q: Record the ripple size and smoothing capacitance in your lab notebook.

P: Insert into your lab notebook a photograph of the oscilloscope screen displaying V_{in} and V_{out}.

15. Disable the function generator OUTPUT and turn off your oscilloscope.

16. Use Python to perform the following analysis:

(a) Perform a curve fit to your $(\Delta V)_{\text{ripple}}$ vs smoothing capacitance data. Assume a function of the form $(\Delta V)_{\text{ripple}} = AC^B$ where A and B are fitting parameters and C is your capacitance variable.

(b) Plot ripple voltage size vs smoothing capacitance and use symbols to display your data points. Overplot a line representing your curve fit.

P: Include your plot in your lab notebook.

Q: Provide your fit parameters in your lab notebook.

17. Predict the minimum smoothing capacitance one would require to ensure a 5% ripple on V_{out}.

Q: Include this result in your lab notebook.

Provide a discussion in your lab notebook based on standard lab notebook guidelines.

11.7 Lab: Operational Amplifiers

Objective: In your lab notebook, write a few sentences stating your objective in conducting this laboratory exercise. Consider the following questions:

- What kind of circuit(s) or components are you exploring?
- What is new about this circuit as compared to circuits previously studied? Ignore the differential amplifier aspect of the circuit setup for the time being.
- What theoretical predictions do you have regarding circuit behavior?
- What effects do you expect to observe?

Your answers should be specific to the type of circuit you are examining, but do not discuss specific component values.

Equipment: Proto-board, function generator, digital oscilloscope, oscilloscope probes, DMM, LM358 op-amp, resistors (1k, 5.1k, 10k), 1 nF capacitor.

Introduction and theory. Operational amplifiers are employed in a broad spectrum of analog circuits designed to amplify, filter, buffer, sum, integrate, differentiate, and so on. Such circuits require few external discrete components beyond the op amps themselves, and they require very little sophistication on the part of the user. All op amp circuits fall in either the **inverting** or the **non-inverting** configuration. An example of an inverting configuration is shown in Figure 11.7.1.

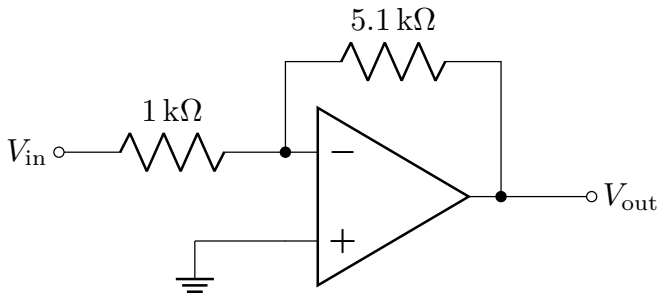


Figure 11.7.1 Inverting amplifier

In the first part of this lab, you will investigate the gain and bandwidth of an inverting amplifier. The gain should remain flat over a wide range of frequencies and should agree with the theoretical-predicted gain. At high frequencies, the gain will start decreasing in a very predictable way. In the second part of this experiment, you will investigate the behavior of a differentiator, verifying its behavior for several input signals.

In this introduction/theory section, include generalized circuit diagrams for the two circuits you will be examining. Then proceed to discuss the relevant theoretical predictions that you hope to experimentally verify.

Experiment: Inverting amplifier. Construct the inverting amplifier circuit using a general purpose LM358 op amp. The pin diagram for the LM358 (8-pin DIP) is shown in Figure 11.7.2. Note that you are given a top-down view of the chip; take careful note of the 'divot' location on the chip. On your prototyping board, the op-amp should be placed across the 'gap' that exists between neighboring five-hole rows, with op amp pins 1-4 on one side of the gap and pins 5-8 on the other side of the gap. Use pins 1-5 for this experiment, as they contain the two supply voltages as well as the inputs and output for op-amp A. You will not use op-amp B (pins 5-7) today. ***You should use*** ± 15 ***Volts for*** $V_{S\pm}$**.**

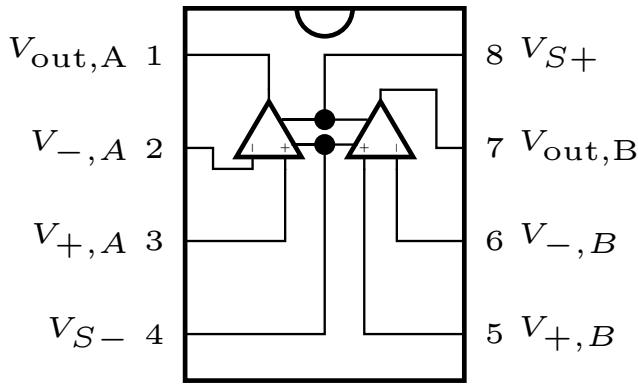


Figure 11.7.2 Pin diagram for the LM358 general-purpose operational amplifier (PDIP package).

Proceed to measure the gain of your amplifier using a small (2.0V ***peak-to-peak*** or so) sinusoidal input signal at frequencies ranging from 10 Hz to 2 MHz. Determine appropriate frequencies within this range that will allow you to convincingly demonstrate the expected bandwidth properties for this op-amp through a plot of gain vs frequency with frequency represented on a logarithmic axis. Consider the following when designing your procedure:

- How are you going to space your frequencies out in the range given above?
  - Must these be evenly spaced?
  - Consider taking fewer data points in ranges where the gain is expected to be relatively constant, but take more densely-sampled data points where you expect your gain to be changing quickly with frequency.

- The Gain Bandwidth Product (GBWP) for the LM358 op-amp has a typical value of 700 kHz.
    - In what approximate frequency range do you expect to see your gain changing fastest for the amplifier gain present in your circuit?
    - Is there a way to quickly check this experimentally before you start taking real data?

Your results should include a plot of Gain vs frequency where frequency is plotted on a logarithmic axis.

Experiment: Differentiator. Assemble the differentiator circuit shown in Figure 11.7.3. Note the added presence of the 1 kΩ input resistor; it is present to lessen the noise sensitivity of the differentiator.

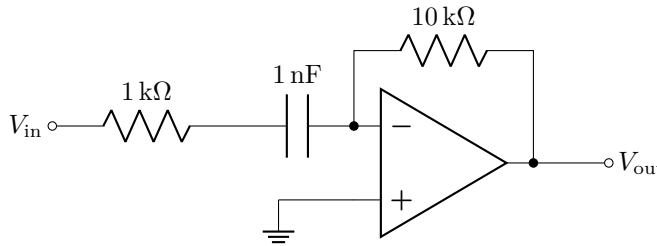


Figure 11.7.3 Inverting amplifier

Examine the behavior of this circuit for an input sine-wave signal (1 Volt peak-to-peak, 4 kHz). Include clear screenshots/photos of the observed behavior, making sure your volts/div scope settings are optimal and that both V_{in} and V_{out} are displayed.

Repeat your observations for a triangle wave input signal. Then, repeat observations for both square wave and triangle wave signals for f=20 kHz and f=5 Hz, including screenshots for each.

Discuss your observations. In particular, be specific about what you would expect to see on V_{out} for each V_{in} that you use and discuss how your observations conform or conflict with your expectations.

11.8 Lab: Digital Circuits - 4-Bit Adder

Objective: In your lab notebook, write a few sentences stating your objective in conducting this laboratory exercise. Consider the following questions:

- What kind of circuit(s) or components are you exploring?
- What is new about this circuit as compared to circuits previously studied?
- What theoretical predictions do you have regarding circuit behavior?
- What effects do you expect to observe?

Your answers should be specific to the type of circuit you are examining, but do not discuss specific component values.

Equipment: Proto-board, 74LS00 Quad 2-Input NAND Gates, 74LS86 Quad 2-Input XOR Gates.

Introduction and theory. Today we will be looking at an example where we use a digital circuit for computation. While the circuit we examine will be far from our advanced modern computing capabilities, it nonetheless provides insight into some of the foundational building blocks that computing is built upon. We will be using two chips in this circuit, each containing either four NAND gates or four XOR gates. The 74LS00 and 74LS86 pinouts are provided in Figure 11.8.1.

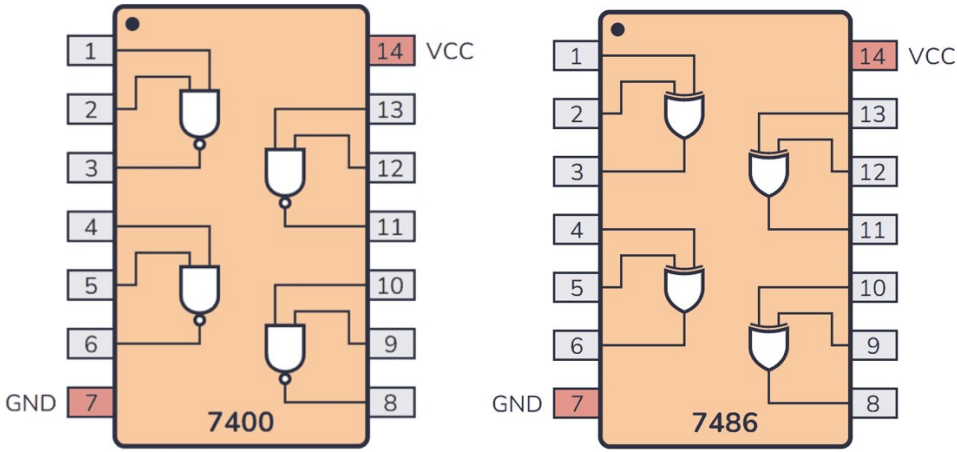


Figure 11.8.1

Digital computation uses the binary, or base-2, number system. In the binary number system, we will refer to **bits** rather than the **digits** used in a base-10 system. Each digit in a base-10 number can take on any one of 10 values (0-9) which is then

multiplied by 10^n where n is the number of digits to the left of the decimal point. For example, base-10 number 1092 can be understood as $1\times10^3+0\times10^2+9\times10^1+2\times10^0$. Likewise, each bit in a binary number can take on one of two possible values (0 and 1) with these values multiplying a power of 2. So, the binary number 1011 can be understood as $1\times2^3+0\times2^2+1\times2^1+1\times2^0$. This means that the binary number 1011 represents 11 in base-10.

The circuit we'll be examining today is called the digital full-adder.

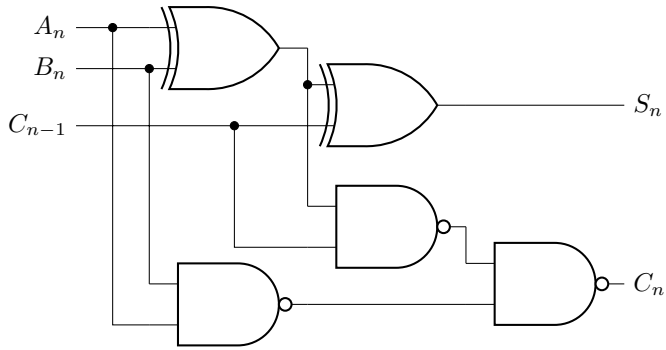


Figure 11.8.2 Circuit schematic for the full-adder circuit. ***Note:*** Locations where lines cross in the schematic are only connected if a junction dot appears. There is no connection where wires cross if there is no dot present.

On its own, the full-adder can be used to add two 1-bit numbers represented by A_n and B_n, providing resulting 'S'um S_n. If both A_n and B_n are equal to 1, the sum should be equal to 2, but the number 2 is not an allowed value for a bit. So, S_n will be 1 and we will track the extra that doesn't fit in S_n and place the leftover 1 into the 'C'arry output C_n. This C_n result gets carried over into the next higher bit.

Calculations with larger numbers can be completed using multiple full-adders chained together. The circuit in Figure 11.8.3 accepts two 3-bit numbers as inputs and provides an output that is a 4-bit number

$$\underline{A_2}\ \underline{A_1}\ \underline{A_0} + \underline{B_2}\ \underline{B_1}\ \underline{B_0} = \underline{C_2}\ \underline{S_2}\ \underline{S_1}\ \underline{S_0}.$$

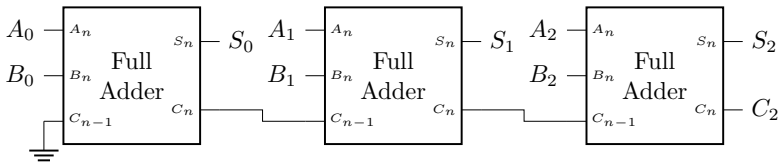


Figure 11.8.3 Circuit schematic for the full-adder circuit. ***Note:*** Locations where lines cross in the schematic are only connected if a junction dot appears. There is no connection where wires cross if there is no dot present.

Experiment: Build and test a full-adder logic circuit.

1. Build a full-adder using three NAND gates and two XOR gates as shown in Figure 11.8.2. Here are a few tips to help you design the circuit layout on the prototyping board:
   - Place the 74LS00 and 74LS86 chips across the central vertical gap with a 4-6 row vertical gap between them. The 74LS86 should be above the 74LS00. This may have already been completed by your instructor.
   - Counting from the left side of the prototyping board, connect the sixth vertical distribution strip (column of vertically-connected holes) to ground and connect the seventh vertical distribution strip to +5 Volts.
   - Connect each of the third, fourth, and fifth vertical distribution strips to a separate logic switch (labeled S_1 through S_8). ***Note:*** Ensure that the +5 option is selected (rather than the +V option) in the Logic Switch section of the prototyping board. You will use switch S_1 for input A_n, S_2 for input B_n, and S_3 for input C_{n-1}.
   - Use XOR gates 1 and 2 on the 74LS86 chip. Use NAND gates 1, 2, and 3 on the 74LS00 for the three NAND gates in order from left to right as they appear in Figure 11.8.2
   - Connect your S_n and C_n outputs to pins 1 and 2 in the Logic Indicators section of the prototyping board. Ensure that the +5 and CMOS options are selected with the switches in this section.
2. Set all circuit inputs to 0 with the logic switches and turn on the board. You should see the first two LOW (green) LEDs light up indicating that outputs S_n and C_n are 0 (LOW) when all inputs are low.
3. Proceed to test all possible input combinations and record the output values, recording your results in the following truth table.

A_n	B_n	C_{n-1}	S_n	C_n
0	0	0		
0	0	1		
0	1	0		
0	1	1		
1	0	0		
1	0	1		
1	1	0		
1	1	1		

4. Confirm with your instructor that your circuit is working properly.

Experiment: Add two 4-bit numbers. Combine the full-adders from multiple groups into a circuit that will compute the base-10 sum $13 + 6$, providing an answer in binary. Check to make sure that this binary representation converts to the correct base-10 answer. Use Figure 11.8.3 as a guide.

Note: Do *NOT* disassemble your circuits. I'm asking you to use long wires to connect together circuits from multiple prototyping boards. Your instructor will assist with this process.

Appendices

Appendix A

Introduction to Complex Numbers

Appendix B

Matrices

Properties of Matrices. A matrix is an array of numbers

$$\begin{pmatrix} 1 & 3 & 4 \\ 4 & 1 & 9 \end{pmatrix}.$$

In order to use matrices to solve systems of equations, we must understand how to multiply matrices using the dot product.

Suppose we have matrix $\mathbf{A}$ that has the form

$$\mathbf{A} = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \\ A_{41} & A_{42} & A_{43} \end{pmatrix}$$

where A_{ij} is the matrix element in the i-th row and j-th column. Now, if we multiply a matrix by a scalar, we find

$$u\mathbf{A} = \begin{pmatrix} uA_{11} & uA_{12} & uA_{13} \\ uA_{21} & uA_{22} & uA_{23} \\ uA_{31} & uA_{32} & uA_{33} \\ uA_{41} & uA_{42} & uA_{43} \end{pmatrix}.$$

Let's now say that we want to multiply matrices $\mathbf{A}$ and $\mathbf{B}$ to find $\mathbf{C} = \mathbf{A} \cdot \mathbf{B}$ where

$$\mathbf{B} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \\ B_{31} & B_{32} \end{pmatrix}.$$

First off, $\mathbf{A} \cdot \mathbf{B}$ is only defined if $\mathbf{A}$ has a number of columns that is equal to the number of rows of $\mathbf{B}$. In other words, $\mathbf{A} \cdot \mathbf{B}$ is only defined for matrices where $\mathbf{A}$ is an $\ell \times m$ matrix and $\mathbf{B}$ is an $m \times n$ matrix. Now we have

$$\mathbf{C} = \mathbf{A} \cdot \mathbf{B} = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \\ A_{41} & A_{42} & A_{43} \end{pmatrix} \cdot \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \\ B_{31} & B_{32} \end{pmatrix}.$$

The elements of $\mathbf{C}$ are given by

$$C_{ik} = \sum_{j=1}^{n} A_{ij} B_{jk}$$

where n is the number of columns of $\mathbf{A}$ and equivalently the number of rows in $\mathbf{B}$. In other words,

$$C_{11} = A_{11}B_{11} + A_{12}B_{21} + A_{13}B_{31}$$

and

$$C_{42} = A_{41}B_{12} + A_{42}B_{22} + A_{43}B_{32}.$$

In other words, in order to find C_{ik}, you can multiply, element-by-element, the cells in row i of $\mathbf{A}$ with the cells in column k of $\mathbf{B}$. The resulting matrix $\mathbf{C} = \mathbf{A} \cdot \mathbf{B}$ will have the same number of rows as $\mathbf{A}$ and same number of columns as $\mathbf{B}$.

The **identity matrix**, typically labeled $\mathbf{I}$ is another concept that will be necessary to understand. The identity matrix is defined such that

$$\mathbf{I} \cdot \mathbf{A} = \mathbf{A}$$

for any matrix $\mathbf{A}$. In order to have this property, the identity matrix must be a square matrix with dimensions $n \times n$ (same number of rows and columns) with values of 1 on the diagonal and 0 everywhere else:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The identity matrix can have any value of n in order to make the desired matrix multiplication work.

Now, let's say that we have an equation

$$\mathbf{C} = \mathbf{A} \cdot \mathbf{B}$$

and let's further assume that we have prior knowledge of both $\mathbf{C}$ and $\mathbf{A}$. How do we find $\mathbf{B}$? We cannot perform division with matrices. Instead, we use the concept

of an **inverse matrix**. Let's define $\mathbf{A}^{-1}$ to be the inverse matrix of $\mathbf{A}$ such that $\mathbf{A}^{-1} \cdot \mathbf{A} = \mathbf{A} \cdot \mathbf{A}^{-1} = \mathbf{I}$. Then, we can multiply both sides of our equation by $\mathbf{A}^{-1}$ to get

$$\mathbf{A}^{-1} \cdot \mathbf{C} = \mathbf{A}^{-1} \cdot \mathbf{A} \cdot \mathbf{B} = \mathbf{I} \cdot \mathbf{B} = \mathbf{B}.$$

You can learn how to find $\mathbf{A}^{-1}$ given knowledge of $\mathbf{A}$ in a linear algebra course. For our purposes, we will just rely on Python to invert matrices.

Connecting to systems of equations. Coming soon.

Appendix C

Numerical Methods

C.1 Newton's Method

FINDING THE ROOT OF ONE NONLINEAR EQUATION. Newton's method is a computational algorithm that can be used to find the roots of an equation. Let's illustrate with an example.

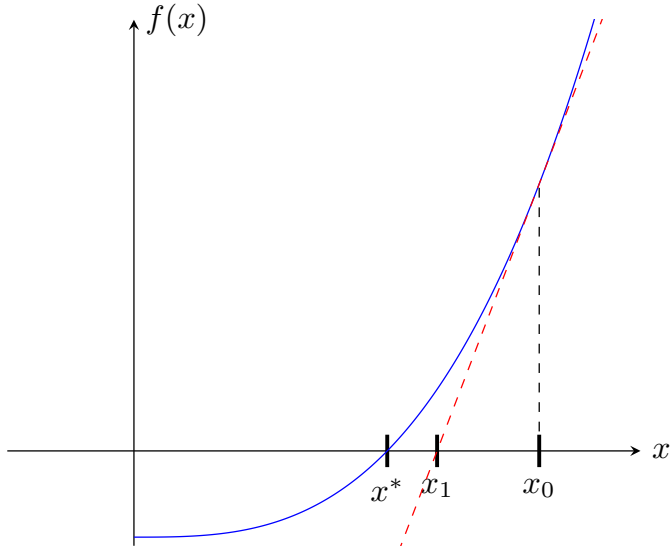


Figure C.1.1

In Figure C.1.1, the solid blue line represents the function $f(x)$. Assume that

x^* is an unknown root of $f(x)$ such that $f(x^*) = 0$. We may be able to guess at an approximate value of the root $x = x_0$, but this is unlikely to be very precise. Examining Figure C.1.1, we can see that we can find a better approximation of the function's root by using a straight-line (dashed red curve) with the same slope as our function at the location of our initial guess,

$$f'(x_0) = \frac{f(x_0) - 0}{x_0 - x_1}$$

which can be rearranged to find our new estimate

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$

We'll call this new approximated root x_1. It's very possible that this new guess still lacks the precision that we desire, so we can repeat the process again, iterating as many times as we desire using

$$x_n = x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})}$$

until $|x_n - x_{n-1}|$ is less than some user-defined tolerance that defines the desired precision, as the change in the estimate to our root from one iteration to the next should become smaller as our estimate approaches the real value of the function's root. Note that this method requires knowledge of the functional form of our function $f(x)$ as well as knowledge of or the ability to calculate the analytic form of $f'(x)$.

We can be a little more rigorous with the theory underlying Newton's method. Let's assume that we have a function $f(x)$ that has a root x^* such that $f(x^*) = 0$. We can expand our function as a Taylor series around some initial guess for the root, x_g, giving us

$$f(x) = f(x_g) + (x - x_g)f'(x_g) + \frac{1}{2}(x - x_g)^2 f''(x_g) + \cdots.$$

Evaluating this expansion at the root $x = x^*$ gives us

$$0 = f(x^*) = f(x_g) + (x^* - x_g)f'(x_g) + \frac{1}{2}(x^* - x_g)^2 f''(x_g) + \cdots.$$

If we divide both sides by $f'(x_g)$ and assume that all terms that are order two or higher in $(x^* - x_g)$ are small, then

$$0 = (x^*_{\text{est}} - x_g) + \frac{f(x_g)}{f'(x_g)}$$

or

$$x^*_{\text{est}} = x_g - \frac{f(x_g)}{f'(x_g)}$$

where x^*_{est} is our estimate of the root based on one iteration of this method. Now, the closer x_g is to the real root x^*, the better our estimate x^*_{est} will be. So, once we have x^*_{est}, we can use this as our new guess for another iteration, giving us a result identical to what we found above,

$$x_n = x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})},$$

which we iterate until $|x_n - x_{n-1}|$ is less than some user-defined tolerance that defines the desired precision.

A word of caution: Newton's method is not foolproof and it can fail sometimes. Please take a look at other computational methods textbooks or webpages to investigate these failure modes. For the purposes of this text, try several different initial guesses if your first choice leads to a failure of the method.

FINDING THE ROOTS OF A SYSTEM OF NONLINEAR EQUATIONS. Write your equations in the following form:

$$\begin{aligned} f_1(x_1, ..., x_N) &= 0 \\ &\vdots \\ f_N(x_1, ..., x_N) &= 0 \end{aligned}$$

with roots $x^*_1, ..., x^*_N$. The Taylor expansion for f_i can be written as

$$f_i(x^*_1, \ldots, x^*_N) = f_i(x_1, \ldots, x_N) + \sum_j \left(x^*_j - x_j\right) \frac{\partial f_i}{\partial x_j} + \mathcal{O}(\text{small}) = 0$$

so that

$$\sum_j \left(x_j - x^*_j\right) \frac{\partial f_i}{\partial x_j} + \mathcal{O}(\text{small}) = f_i(x_1, \ldots, x_N).$$

Each i value can represent the i-th equation in a system of equations. This system can be written in vector form as

$$\vec{f}(\vec{x}^*) = \vec{f}(\vec{x}) + \nabla \vec{f} \cdot (\vec{x}^* - \vec{x}) + \mathcal{O}(\text{small})$$

where $\nabla \vec{f}$ is the $N \times N$ matrix with elements $\partial f_i / \partial x_j$. Since $\vec{x}^*$ is a root of the equations, we have $\vec{f}(\vec{x}^*) = 0$ so that

$$\nabla \vec{f} \cdot \Delta \vec{x} + \mathcal{O}(\text{small}) = \vec{f}(\vec{x})$$

where $\Delta\vec{x} = \vec{x} - \vec{x}^*$. Thus, we can get a good estimate of our root, $\vec{x}'$, using

$$\nabla\vec{f} \cdot \Delta\vec{x} = \vec{f}(\vec{x})$$

where $\Delta\vec{x} = \vec{x} - \vec{x}'$ now. This is a set of ordinary linear simultaneous equations of the form $\vec{A}\vec{x} = \vec{v}$ which can be solved using the `linalg.solve` package in NumPy. Once we've solved for $\Delta\vec{x}$, our new estimate $\vec{x}'$ of the root is

$$\vec{x}' = \vec{x} - \Delta\vec{x}.$$

Colophon

This book was authored in PreTeXt.