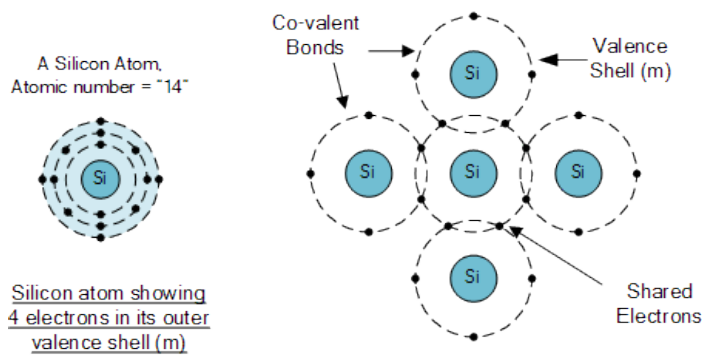
Lab 3: Introduction to Semiconductors

Semiconductors are one of the most important device categories in the modern world. From transistors in computer chips, to LED screen to sensors; much of the functionality of the 21st century depends on the properties and behavior of semiconductor devices.

In this lab you will be introduced to the two simplest semiconductors: an LED, which consists of a single semiconductor junction; and a BJT a Bipolar Junction Transistor, which consists of a pair of semiconductor junctions.

Semiconductor Basics

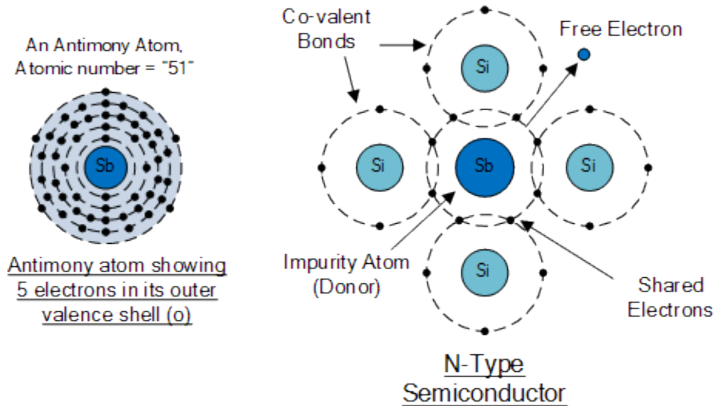
There are semiconductor devices, like LEDs and transistors and there are semiconductor materials like silicon. Semiconductor devices are made of semiconductor materials. However, the topic is more complicated than it may appear. Silicon does not conduct electricity. It is not a semi-conductor; it is an insulator. So why do we call it a semiconductor? Why is it used in these devices?



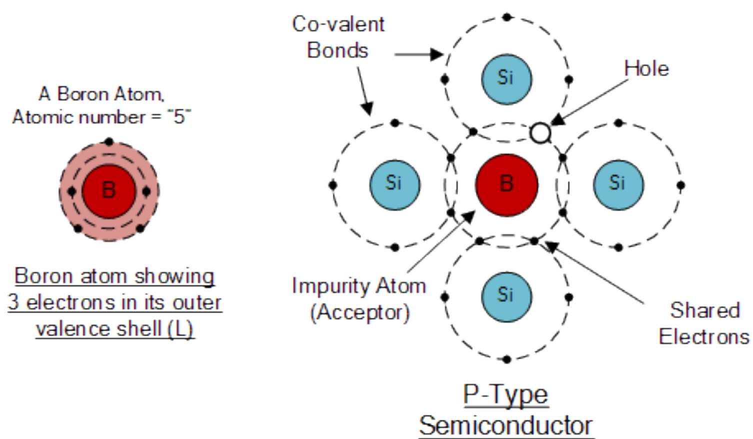
Silicon Crystal Lattice

In silicon, each atom has four valence electrons. A crystal structure is formed when each neighbouring silicon atom shares one electron in a covalent bond. In a pure silicon crystal, there are no excess or free electrons available to conduct electricity. It is therefore an insulator.

The crystal structure of a silicon crystal can be manipulated by adding in impurities or “doping”. This changes its properties. If an atom of antimony is placed into the structure it creates a free electron. Antimony has five electrons in its valence shell of electrons. It will share four of these valence electrons with its silicon neighbours, leaving one free electron available to conduct electricity. Since this kind of doping creates an excess of negative charge carriers it is called “N-Type”.

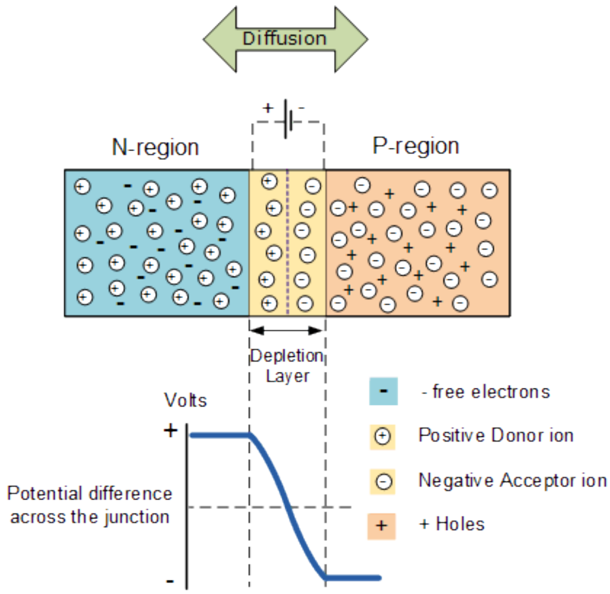


A similar effect is possible by adding a Boron impurity to a silicon crystal. A boron atom only has three atoms in its valence shell. It will make three covalent bonds with its silicon neighbours but that last bond is missing an electron. There is a hole. This hole will attract a nearby electron, but then the bond that the electron came from will be a hole… which will then attract another electron… which will leave another hole. We can consider these holes to be positive charge carriers. Since this type of doping creates excess positive charge carriers it is called “P-Type”.



PN Junction Diode

For a semiconductor to truly live up to its name, it needs to act like a conductor some of the time, and act like an insulator other times. So far, all doping does is turn silicon from an insulator to a conductor. The final step occurs when some “N-type” doped silicon and “P-type” doped silicon are brought together to form a PN junction.



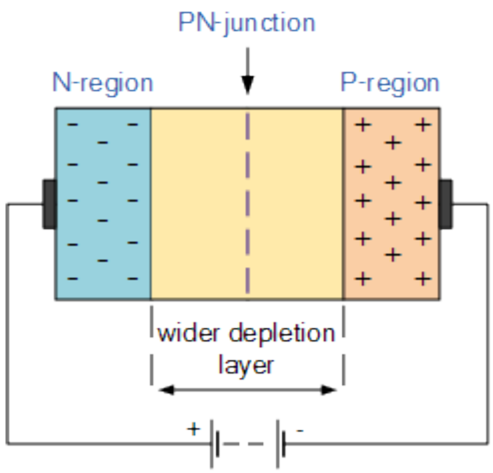
What is critical to understand is that while N and P doping create free charge carriers, each crystal structure is electrically neutral. Boron has one less valence electron, but it also has one less proton, so the total charges in the crystal are zero. The holes are free to move around, but the total charge is neutral. Similar to the N-Type doping. Antimony has an extra valence electron, but it also has an extra proton. The electron is free to move around, but the overall charge of the crystal is zero.

At this junction, some of the holes in the P-region will attract electrons from the N-region. When a hole from the P-region borrows a free electron from the N-region it breaks the charge neutrality. There is now an excess electron in the P-region, and it becomes negatively charged. Since the electron left a Antimony atom in the N-region it becomes positively charged. This charge imbalance creates a potential difference across the junction similar to the plates of a capacitor. In addition, each time an electron and a hole match up across the junction it reduces the number of excess charge carriers available to conduct electricity.

When P and N Type doped silicon are brought together into a junction, a depletion layer is created – a region between the doped materials that acts as an insulator. There are no free charge carriers available. For a current to flow, the electric potential created by the charge imbalance must be overcome.

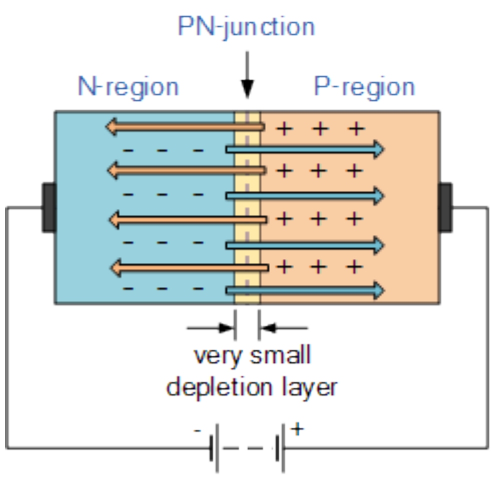
Applying an External Voltage to a PN Junction – Rectifier

Let's consider applying an external voltage to a PN junction using a battery or a power supply. Consider applying a positive external voltage to the N-region and a negative voltage to the P-region. This is called a "reverse" bias because we are applying a positive voltage to a region with an excess of negative charge carriers.



Reverse Biasing Voltage

The positive voltage will attract the negative charge carriers in the N-region pulling them away from the PN-junction. This increases the size of the depletion layer by taking away more excess charge carriers. The opposite happens on the P-region, where holes are attracted away from the junction. Charge will build up on either side of the device and no significant current will flow. Some of the charge carriers that migrated over to the holes in the P-region will flow back. This is called a reverse leakage current and on modern diodes is measured in picoamps (10^{-9} A). For practical purposes, the PN junction does not conduct electricity in this direction.



Forward Biasing Voltage

Now consider applying a negative voltage to the N-region. In this case, the negative charge carriers in the N-region have repelling forces on both sides. They are repelled from the external applied voltage and they are repelled from the charge imbalance on the far side of depletion region. The larger the external applied force, the further excess charge carriers are pushed into the depletion layer. There is a sudden change of behavior when the repelling force from the applied voltage exceeds the internal charge imbalance. The negative charge carriers that get through the depletion layer are then attracted to the positive voltage applied to P-region on far side. This completes the circuit and charge carriers can flow freely. The PN junction conducts electricity in this direction.

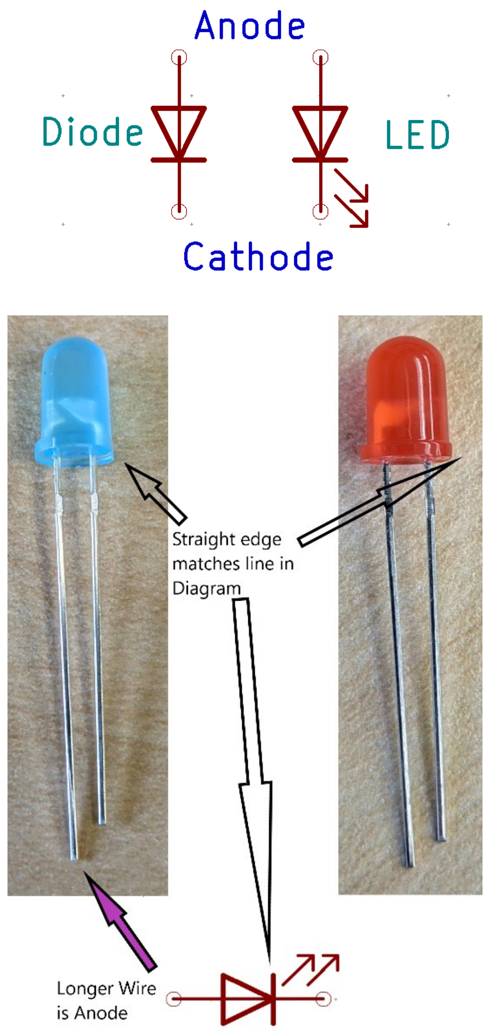
A quick note on vocabulary: While any positive voltage applied to the P-region can be called a forward biasing voltage, the specific applied voltage required to overcome the internal charge imbalance is called “the forward bias voltage”.

We now have a semiconductor. The PN junction conducts electricity in some conditions (forward biasing voltage), but not in others (reverse biasing voltage). The size of the depletion layer can be changed based on the materials used and the level of doping. The forward bias voltage can range from 0.3 V to 0.7 V in typical cases. However, larger or different ranges occur, sometimes due to intentional design decisions. A single PN junction device is called a “rectifier diode”. This is typically shortened to diode.

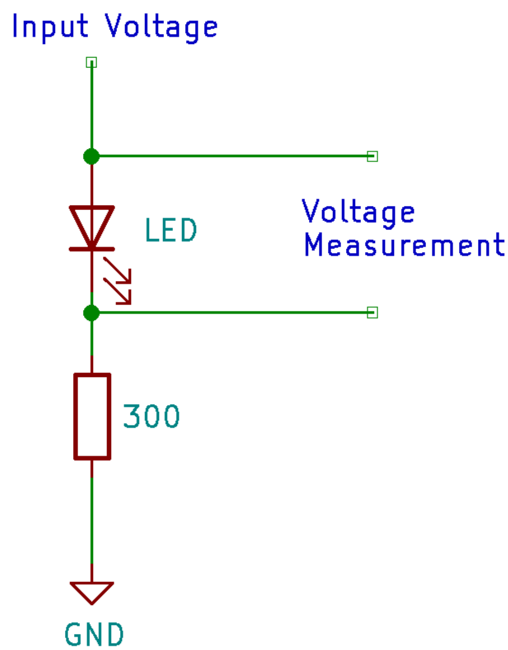
Diodes have many practical purposes that make use of this unidirectional electrical conductivity. They are often used to protect electrical components from reverse voltage spikes. They are also used to “rectify” or change AC signals to DC signals. In the next exercise you will study the forward bias voltage of a Light Emitting Diode (LED). The quantum mechanics description of how an LED uses a PN junction to create light is beyond the scope of this course, but the general principle of PN junctions and biases still apply.

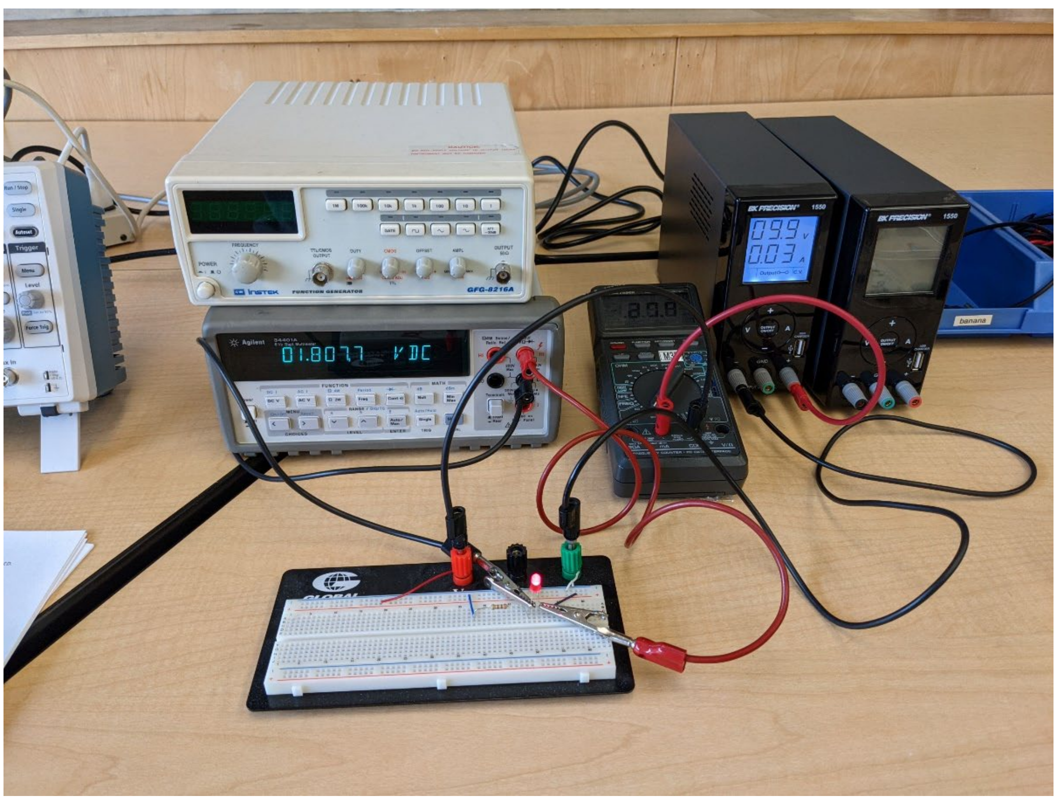
Exercise 1: LED Forward Bias

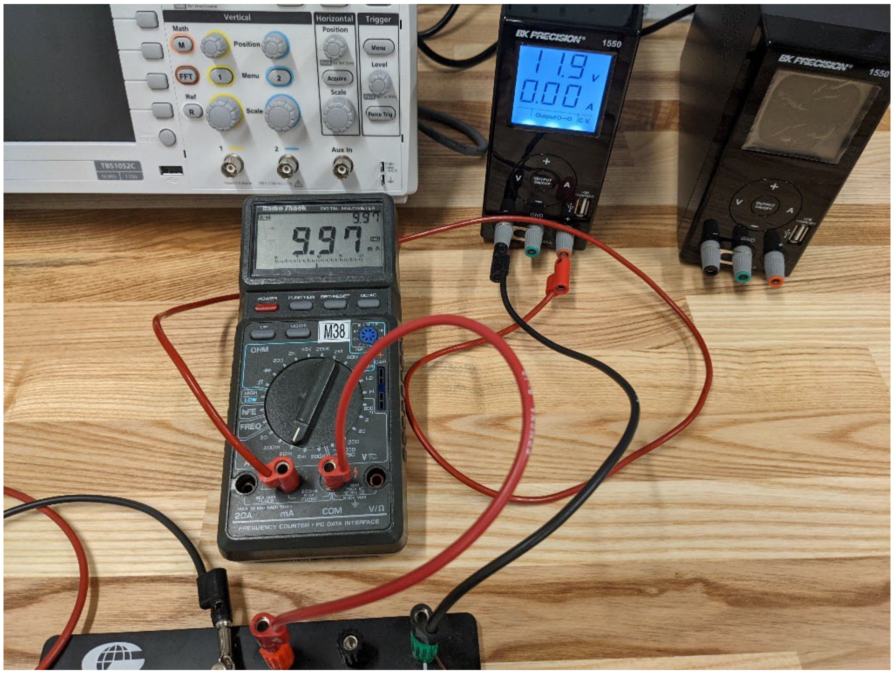
Since an LED is a device with a directionality, you must learn how to identify which wire coming out of the LED is connected to the N-region and which is connected to the P-region. Since the very first rectifiers were vacuum tubes, we call the wire connected to the P-region the "anode" and to the N-region the "cathode".

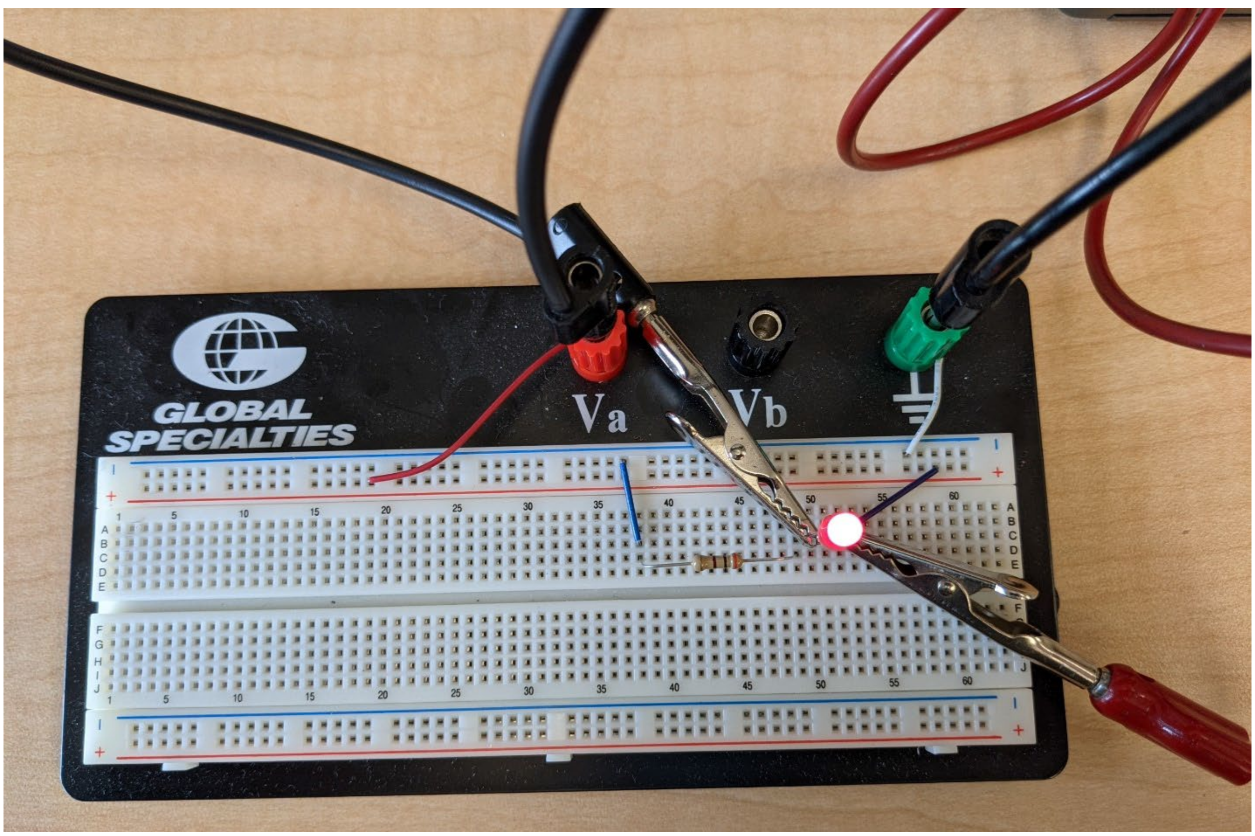


Step 1: Build the following circuit using the red LED. Set up a Radio Shack digital multimeter to measure the current flow through the LED. Use an Agilent multimeter to measure the voltage across the LED. The power supply input voltage should be initially set to 0 V.









Question 1: [2 points]

What is a good current limit to set on the power supply?
What is the current limit stated in the LED data sheet? What is maximum power output of the resistor?

Step 2: Turn on the power supply and take measurements of current and voltage drop across the LED. Take 10 measurements with the power supply set between 0 – 10 V. Then take 5 – 10 more points around the forward bias voltage.

Note: The lab's power supply model has a minimum voltage above the lowest range for typical forward biases mentioned in the introduction; however, the LEDs used for the lab have a forward bias that you will be able to capture with our equipment.

Question 2: [3 points]

What is the forward bias voltage of the red LED? [1 point]

Make a graph of the voltage across the LED vs the current through it for your answer. [2 points]

Step 3: Replace the red LED with a blue LED. Repeat the measurements of current and voltage drop across the LED.

Question 3: [1 point]

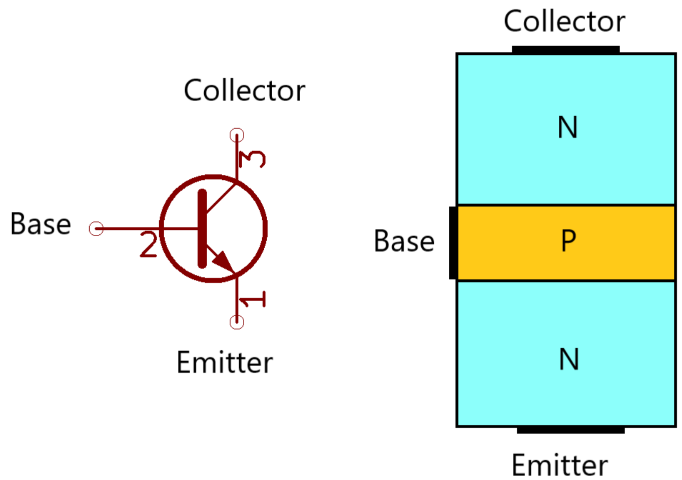
What is the forward bias voltage of the blue LED?

Question 4: [1 point]

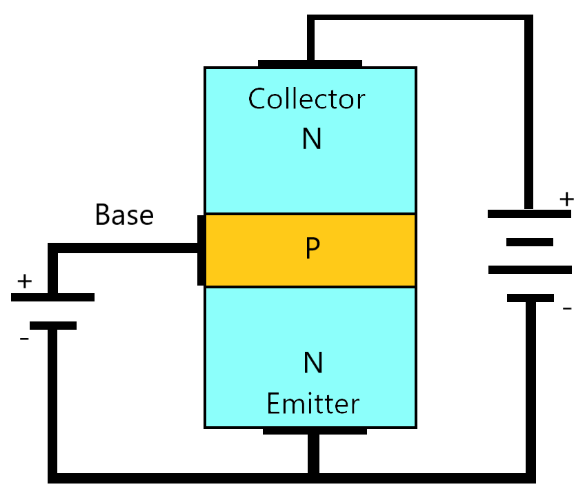
Based on your knowledge of semiconductors, speculate why different colour LED would have different forward bias voltages?

The NPN Bipolar Junction

The next semiconductor device we will be studying is the dual junction NPN. Initially, a NPN junction was formally called a triode because it had 3 connections. It is now known as a transistor. The three connections to the device are labeled the emitter, collector, and base. You can see the location of these connections in the figure below.



The general principles you learned about PN junctions in diodes can be applied to the junctions in a transistor. There will be two depletion layers at each junction in a transistor. Any voltage applied across the collector and emitter will put one PN junction in a forward bias and the other in a reverse bias. For example, a positive voltage applied to the collector will place the top NP junction in a reverse bias, while placing the bottom PN junction in a forward bias. Because of this geometry, a transistor acts as a insulator for any voltage placed on the collector or emitter connections.



The functionality of the transistor is revealed when you consider placing a positive voltage on the base. Now the PN junction between the base and emitter is forward biased. Negative charge carriers from the N emitter region can cross the depletion layer and current can flow from base to emitter.

If there is a simultaneous voltage across the collector and emitter, this will also force negative charge carriers up from the N emitter region. There is a limit to the amount of charge carriers that can flow through the base. The number of charge carriers from the collector voltage can vastly exceed the number of charge carriers from the base. This buildup of charge carriers creates an electric potential. If this potential becomes large enough, it can override the reverse bias of the N collector region. Negative charge carriers can move past the depletion layer and are then attracted by the positive voltage at the collector. Current can flow from the collector to the emitter.

The base connector of the transistor can manipulate the depletion layer and charge balance in the NPN junction. This allows voltages at the base to control conductivity between the collector and emitter. If a high enough voltage is supplied to forward bias the base, current will flow from the collector to emitter as well.

Exercise 2: NPN Forward Bias

Before building the circuit, you will need to know which wire on the transistor corresponds to the emitter, base, and collector. This information is normally available on the transistor's data sheet. However, the transistors we are using are labeled right on the device itself. Each wire is labelled by the first letter in its name, i.e., E for emitter.

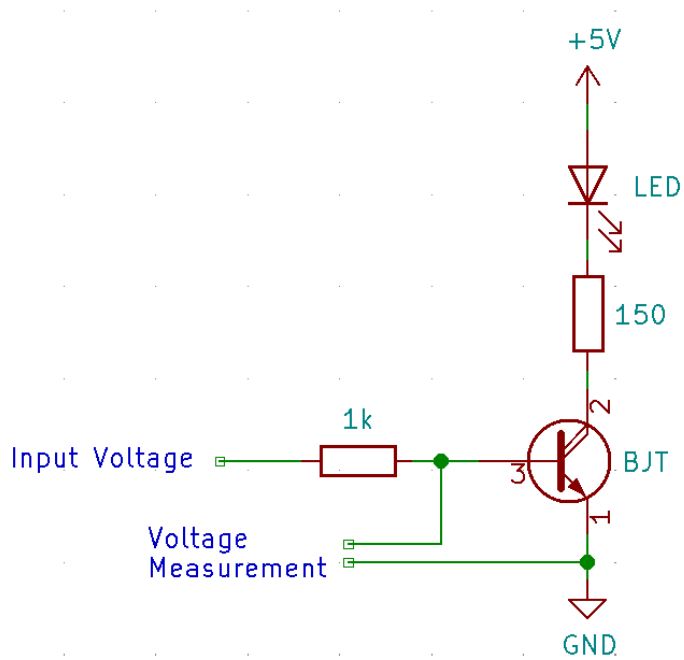
Step 1: Build the following circuit. You will need to use two power supplies, two Radio Shack multimeters and the Agilent multimeter.

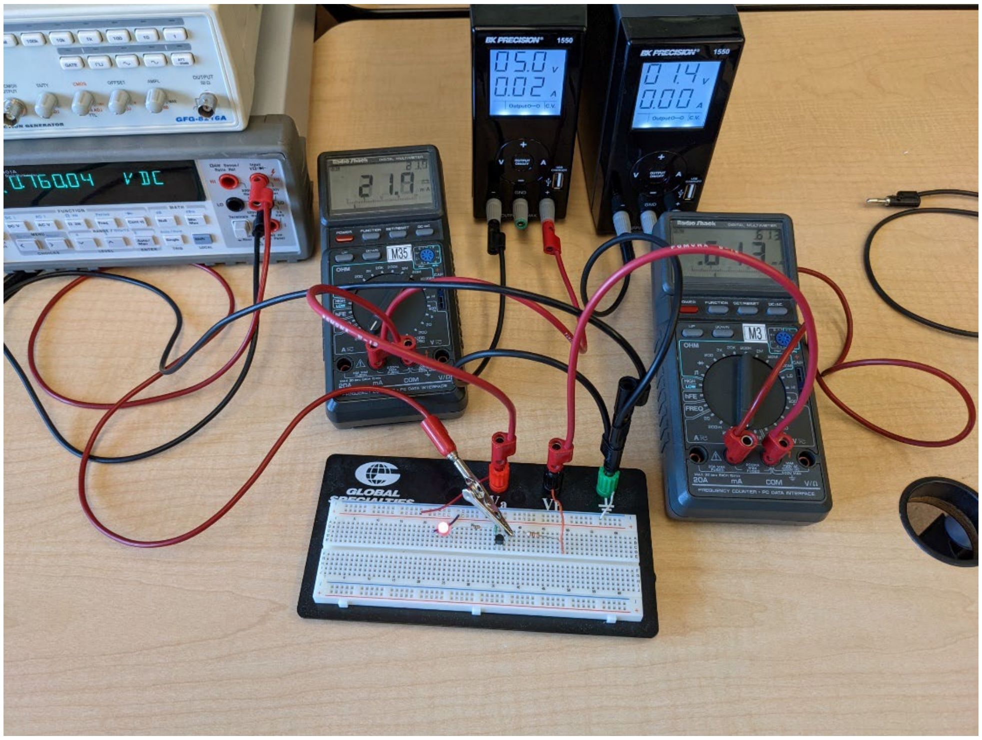
Put the Radio Shack multimeters is series with the output of the power supplies to measure their current.

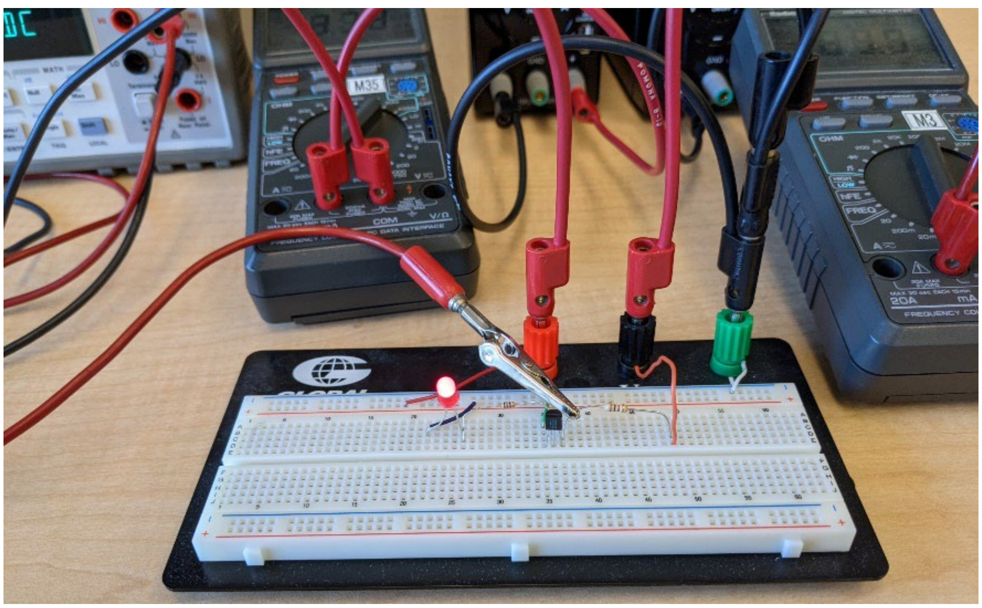
Connect the negative terminals of both power supplies to the ground binding post and then connect this to the "–" rail. Each positive terminal should be given its own binding post.

Connect one of these binding posts to the "+" rail. This power supply will be set to +5 V. The other power supply will be the input to the base of the transistor.

Measure the voltage across base and emitter of the transistor with the Agilent multimeter.







Step 2: Turn on both power supplies and all multimeters. Take measurements of the voltage at the base and the current from each power supply at 10 points between 0.5 V and 2 V. Make note of which power supply provides current to the base or the emitter.

Question 5: [3 points]

Make a graph of current into the base and the voltage across the base and emitter. [2 points]

What is the forward bias voltage of the PN base – emitter junction? [1 point]

Question 6: [2 points]

The current gain of the transistor is given by

$$\beta = \frac{I_c}{I_B}$$

Calculate β for the points you measured above. Make a plot – what does this show? [1 point]
Is β constant or is it dependent on forward bias at the base? [1 point]

Note: the forward bias is less of a single value in some cases and more of a region of behaviour.

Question 7: [3 points]

A transistor behavior is described in 3 broad areas: cutoff, active, and saturation.

In the cutoff region, the voltage applied to the base is not sufficient to forward bias the base-emitter PN junction. There is no (or very low) current moving through the collector to emitter. The transistor can be seen as a switch and the switch is off.

In the saturation region, the voltage applied to the base has fully forward biased the base-emitter PN junction. Further increases in voltage and current to the base does not increase the current moving through the collector to emitter. The transistor can be seen as a switch and the switch is on.

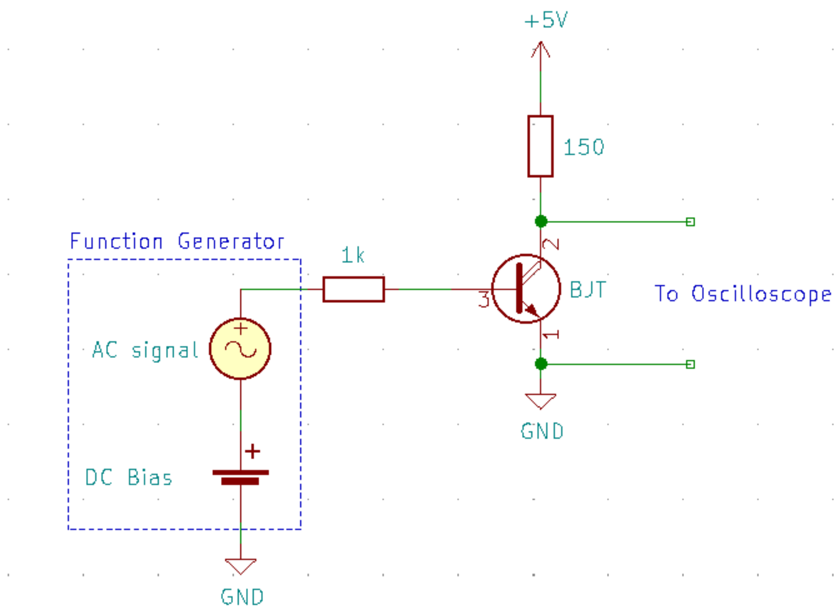
The active region is somewhere in between. The voltage to the base-emitter PN junction is not fully forward biased. Small increases in current and voltage to the base result in large increases in current through the collector to emitter.

Based on the data you collected, what voltages applied to the base of your transistor result in the different behaviour regions?

Exercise 3: AC signals – Transistor as Amplifier

Let's examine the use of a smaller voltage and current at the base of the transistor to control a larger voltage at the collector. Specifically, how a transistor can amplify a small AC signal at the base when operating in the active behavior region.

So far you have looked at circuit's behavior with either a DC or AC voltage input. In this exercise you will need a small AC signal that is superimposed on a DC voltage. The purpose of the DC signal is to bias the base-emitter junction into the active region. We are then interested in measuring the AC signal at the base (the input) and at the collector (the output) to see what impact the transistor has on the AC component of the signal.

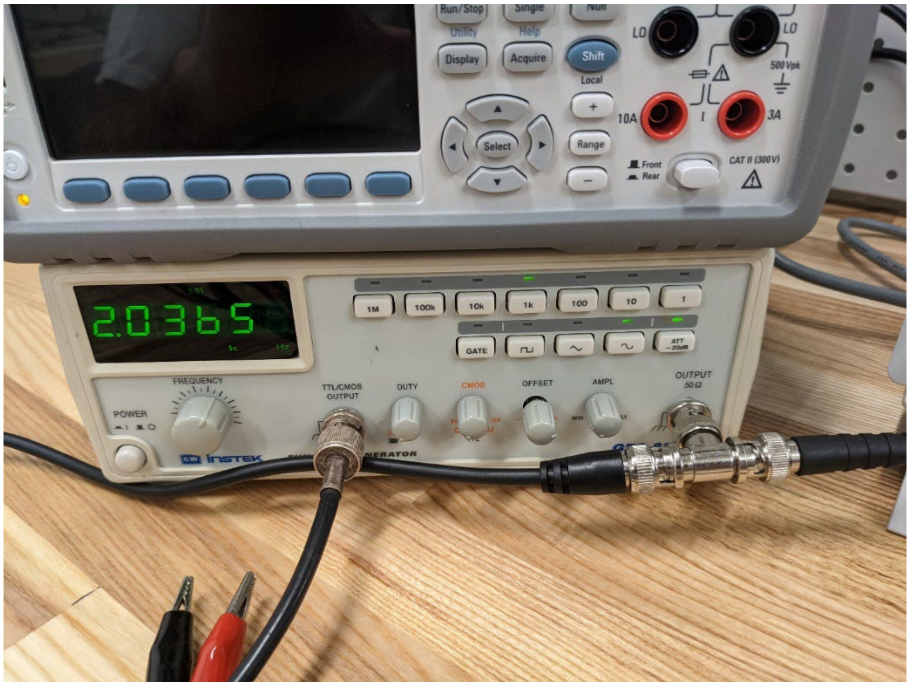


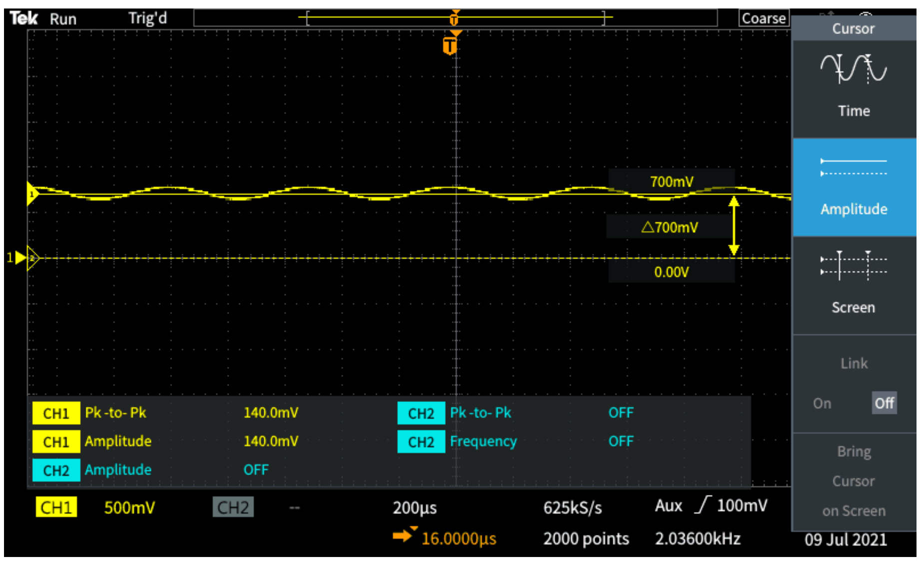
Step 1: The function generator can provide this kind of signal. The DC component is provided by the "offset" knob on the front of the function generator.

Set the function generator to a 2 KHz sine wave and activate the −20 dB button. Reduce the amplitude to the lowest possible setting. Turn on the oscilloscope and observe the output of the function generator in channel 1.

Note: Open the channel 1 menu and make sure the "coupling" is set to "DC". Make sure the attenuation is set to 1×.

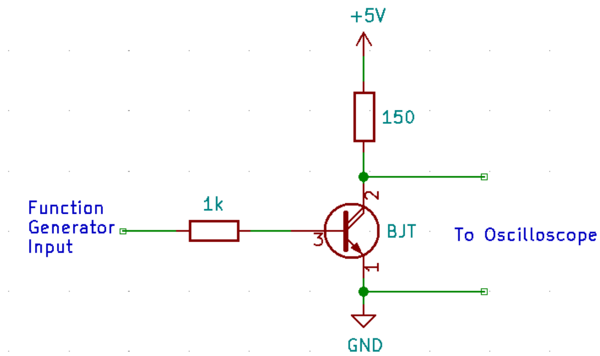
You should only see a small AC signal. Pull out the DC offset knob to activate this mode on your function generator. Notice that as you turn the knob the position of the AC signal moves up and down the screen. This is the DC bias or "offset".

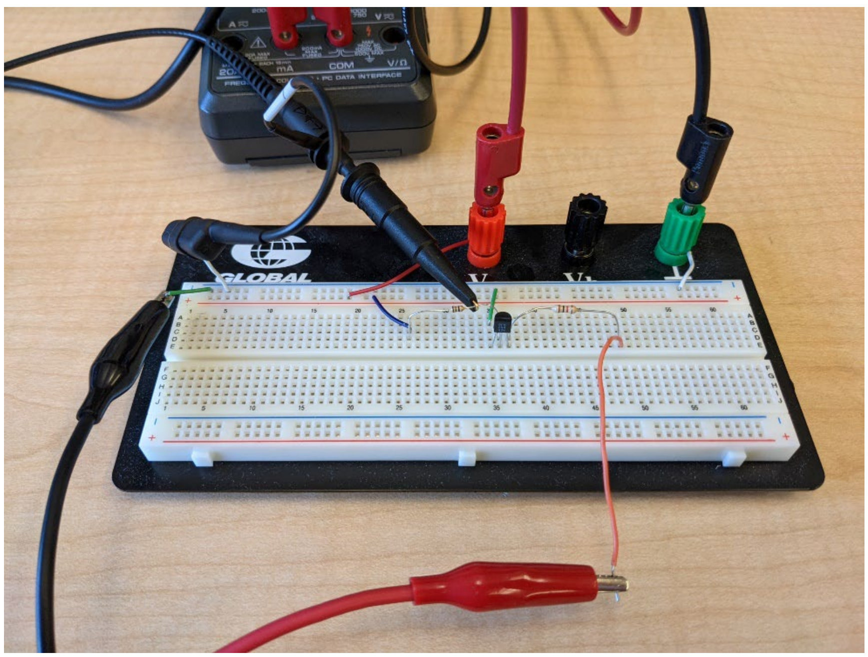




Next, set this DC bias such that the AC signal is within the active behaviour region you determined in the last exercise. To do this, use the cursors on the oscilloscope. Place one cursor at 0 V and the other such that it cuts through the middle of the sine wave. Adjust the offset so that it is in the active region of your transistor. In the photo above, the offset is set to 700 mV.

Step 2: Build the following circuit. You no longer need any of the multimeters. The power supply used to supply the voltage to the base of the transistor from the last exercise can be disconnected.





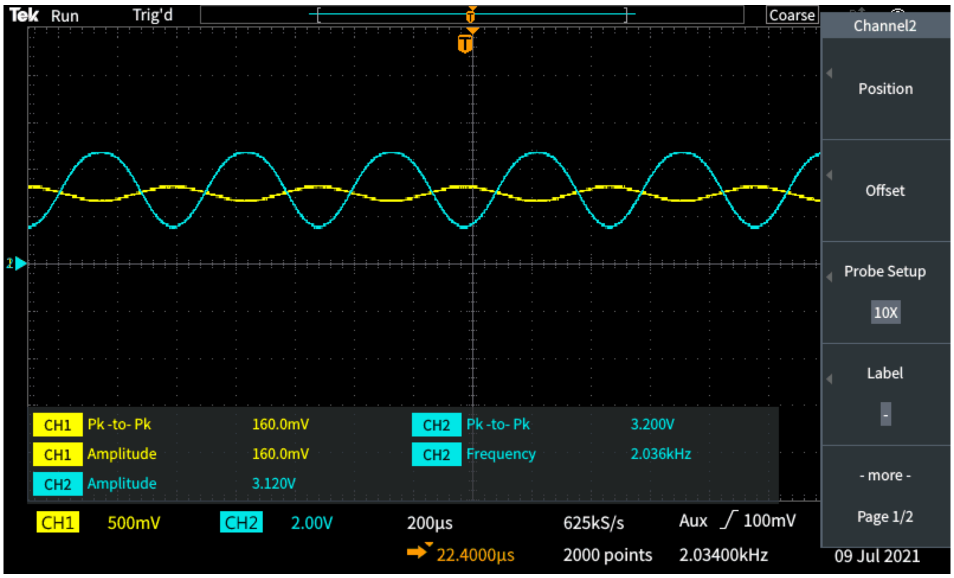
Step 3: Measure the voltage across the collector and ground using channel 2 of the oscilloscope.

(Check the channel settings and make sure coupling is set to "DC" and attenuation is set to 10×.)

Turn on the 5 V power supply. You are looking for an undistorted sine wave in both channel 1 and channel 2. You will need to adjust the setting of your function generator. First, adjust the "offset". Then, adjust the amplitude.

The goal is to have the largest amplitude in the output (channel 2) without any distortion.

Both the input and output should show a DC offset.



Question 8: [2 point]

Use the measurement tools of the oscilloscope to measure the pk-pk amplitude of the input from the function generator and output from the transistor. Determine the amplifier's gain using the equation:

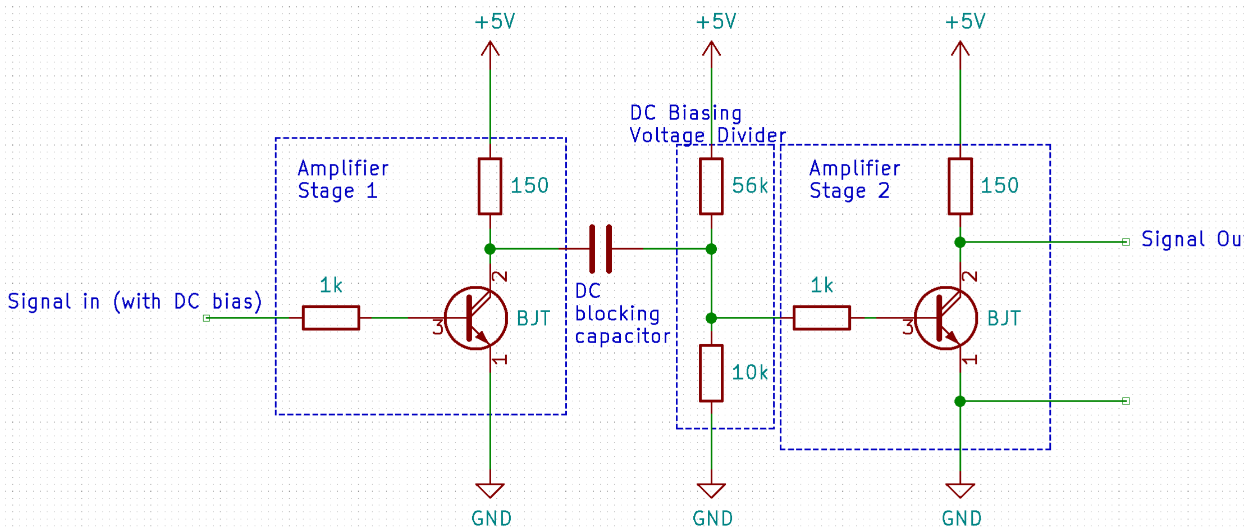
$$\alpha = \frac{V_{output}}{V_{input}}$$

Question 9: [3 point]

If you increase the amplitude of the input signal you will always reach a point where the sine wave gets distorted. From your knowledge of behavior regions of the transistor why does this occur?

Multistage Amplifiers

You have now seen how a transistor can be used to amply small voltage signals. Sometimes the voltage output from a sensor can be very small and needs to be amplified by several orders of magnitude: one hundred thousand times or even a million times! One transistor is not enough to achieve this. However, it is possible to chain many transistors together to achieve large gains.



The circuit above is an example of a multistage amplifier based on the circuit we studied in the last exercise. The theoretical gain of this amplifier would be the square of whatever you calculated earlier or α^2. Multistage amplifiers are very difficult to design and build. The tricky part is the amplification from stage 1 can cause the transistor stage 2 to saturate. This creates distortion of the signal.

Designing proper DC biasing between stages is complicated. You can see an example of how to achieve DC biasing using a capacitor and voltage divider in the circuit above, however, building these multistage amplifiers is beyond the scope of this course.

In practice you rarely see multistage amplifiers built from individual transistors and components. Instead, these circuits are built into integrated circuits (ICs) called operational amplifiers (op amps). The issues related to properly biasing the stages of the amplifier are handled by the chip, making the overall circuit design easier and much smaller.

You will learn more about op amps in the next lab.