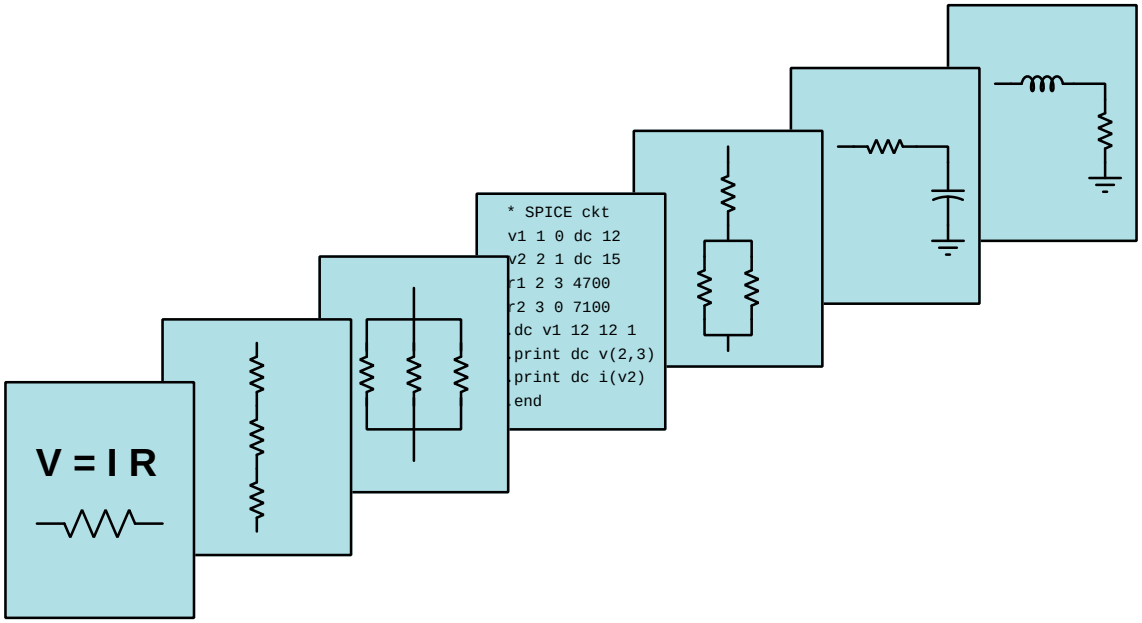
Modular Electronics Learning (ModEL) project



Bipolar Junction Transistors

© 2019-2026 by Tony R. Kuphaldt – under the terms and conditions of the Creative Commons Attribution 4.0 International Public License

Last update = 27 September 2026

This is a copyrighted work, but licensed under the Creative Commons Attribution 4.0 International Public License. A copy of this license is found in the last Appendix of this document. Alternatively, you may visit `http://creativecommons.org/licenses/by/4.0/` or send a letter to Creative Commons: 171 Second Street, Suite 300, San Francisco, California, 94105, USA. The terms and conditions of this license allow for free copying, distribution, and/or modification of all licensed works by the general public.

Contents

1 Introduction **3**
- 1.1 Recommendations for students . . . 3
- 1.2 Challenging concepts related to bipolar junction transistors . . . 5
- 1.3 Recommendations for instructors . . . 7

2 Case Tutorial **9**
- 2.1 Experiment: testing BJTs using diode-check multimeter mode . . . 10
- 2.2 Example: BJT collector current with varying base stimulus . . . 16
- 2.3 Example: BJT collector current with varying collector-emitter voltage . . . 17

3 Simplified Tutorial **19**

4 Full Tutorial **21**
- 4.1 Amplifier concepts . . . 22
- 4.2 Injected charge carriers . . . 29
- 4.3 BJT characteristics . . . 32
- 4.4 BJT construction . . . 39

5 Historical References **43**
- 5.1 John Bardeen's and Walter Brattain's transistor patent . . . 44
- 5.2 Early Bell Labs transistors . . . 48

6 Animations **49**
- 6.1 Animation of an NPN BJT turning on . . . 50
- 6.2 Animation of BJT curve-sketching . . . 65

7 Questions **101**
- 7.1 Conceptual reasoning . . . 105
  - 7.1.1 Reading outline and reflections . . . 106
  - 7.1.2 Foundational concepts . . . 107
  - 7.1.3 BJT patent . . . 109
  - 7.1.4 Controlling and controlled currents . . . 110
  - 7.1.5 Controlling voltages . . . 111
  - 7.1.6 Relative current magnitudes . . . 112
  - 7.1.7 Failed attempt at making a transistor . . . 113

7.1.8 Proper BJT types . . . 114
7.1.9 Common-emitter behavior . . . 115
7.1.10 Common-collector behavior . . . 116
7.1.11 Common-base behavior . . . 117
7.2 Quantitative reasoning . . . 118
7.2.1 Miscellaneous physical constants . . . 119
7.2.2 Introduction to spreadsheets . . . 120
7.2.3 DC bias calculations . . . 123
7.2.4 BJT circuit calculations . . . 124
7.3 Diagnostic reasoning . . . 129
7.3.1 Faults in test circuit . . . 130

8 Projects and Experiments **133**
8.1 Recommended practices . . . 133
8.1.1 Safety first! . . . 134
8.1.2 Other helpful tips . . . 136
8.1.3 Terminal blocks for circuit construction . . . 137
8.1.4 Conducting experiments . . . 141
8.1.5 Constructing projects . . . 145
8.2 Experiment: BJT terminal identification . . . 146
8.3 Experiment: demonstrate BJT amplification . . . 147
8.4 Experiment: measure beta ratio . . . 148
8.5 Experiment: demonstrate the Early effect . . . 149
8.6 Experiment: demonstrate a BJT used as a switch . . . 150
8.7 Experiment: measuring switching time . . . 151

A Problem-Solving Strategies **153**

B Instructional philosophy **155**
B.1 First principles of learning . . . 156
B.2 Proven strategies for instructors . . . 157
B.3 Proven strategies for students . . . 159
B.4 Design of these learning modules . . . 160

C Tools used **163**

D Creative Commons License **167**

E References **175**

F Version history **177**

Index **180**

Chapter 1

Introduction

1.1 Recommendations for students

This module explores the principles of operation, characteristics, and basic uses of bipolar transistors which form the basis of a great many modern electronic circuits.

Important concepts related to bipolar junction transistors (BJTs) include **amplification**, **active** versus **passive** devices, **gain**, **digital** versus **analog** signals, **semiconductors**, **doping**, **depletion regions**, **PN junctions**, forward and reverse **bias**, **beta ratio**, **Kirchhoff's Current Law**, **Conservation of Electric Charge**, **Joule's Law**, and the **Conservation of Energy**.

Here are some good questions to ask of yourself while studying this subject:

- How might an experiment be designed and conducted to measure the beta ratio for a bipolar junction transistor? What hypothesis (i.e. prediction) might you pose for that experiment, and what result(s) would either support or disprove that hypothesis?
- How might an experiment be designed and conducted to plot a family of characteristic curves for a bipolar junction transistor? What hypotheses (i.e. predictions) might you pose for that experiment, and what result(s) would either support or disprove those hypotheses?
- What are some practical applications of bipolar junction transistors?
- What differentiates an *active* from a *passive* component?
- What is amplification and how does it work?
- How is it possible for an amplifier to boost the power of a signal without violating the Conservation of Energy?
- How do we calculate gain for any particular device or system?
- Where might we find electromechanical relays used?
- What are vacuum tubes capable of that relays are not?

- Why are vacuum tubes now considered obsolete for most applications?
- How does electrical conduction differ in a semiconductive material compared to a conductive material?
- What role does the depletion region play in a semiconductor diode?
- How does a depletion region at a P-N junction differ from a true insulator?
- What role does the depletion region play in a bipolar junction transistor?
- What conditions are necessary to turn a BJT on?
- How does the analogy of sand and wind help explain charge carrier injection in a BJT?
- Compare the linear versus non-linear graphs showing transistor characteristics. How can two completely different graphs relate to the same component?
- What is the relationship of collector current to base current in an operating BJT?
- What is the relationship of collector current to collector-emitter voltage in an operating BJT?
- Under what conditions does a BJT act to regulate current, and why does it do so?
- How does the beta ratio relate base current, collector current, and emitter current to each other?
- How is power dissipation computed for a BJT?

1.2 Challenging concepts related to bipolar junction transistors

The following list cites concepts related to this module's topic that are easily misunderstood, along with suggestions for properly understanding them:

- **Electrical conduction in semiconductors** – what differentiates semiconductors from regular (Ohmic) conductors is the relative scarcity of free charge carriers, and how the concentration of charge carriers may be externally controlled in semiconductors by light and/or by strong electric fields. Current through a PN junction does not follow Ohm's Law (which is a great illustration of how Ohm's "Law" is not really a Law in the strict sense of the word, but rather a definition of resistance), but rather grows exponentially with applied voltage as described by the Shockley diode equation.

- **Necessary conditions for transistor operation** – bipolar junction transistors require a base current in the proper direction to inject minority carriers into the base layer (i.e. PNP transistors require hole injection from emitter to base, and so the emitter must be made positive and the base made negative, relative to each other; NPN transistors require electron injection from emitter to base, and so the emitter must be negative and the base positive.), and the collector-to-emitter voltage must be of the correct polarity (i.e. so that collector terminal current goes in the same direction as base terminal current, either *out* or *in*). Both currents join at the emitter terminal, making the emitter current the sum of the base and collector currents.

- **Transistor saturation** – a transistor is considered to be *saturated* in its "on" state whenever its conducted current is less than what the transistor is being stimulated to pass; in other words, a "saturated" transistor is turned on to a degree that is in excess of the load current it's actually conducting. This is particularly confusing for bipolar junction transistors because the collector-to-base current ratio Beta (β) rating assumes linear or "throttling" mode (also known as the "active" operating region for a BJT). By contrast, a "saturated" BJT is one where a super-abundance of charge carriers has been injected from the emitter region, and where collector current is limited by external load resistance and source voltage rather than being limited by a scarcity of free charge carriers within the transistor. This means the collector current predicted by Beta as a function of base current is actually a *maximum limit* and not necessarily the current seen at the collector given supply voltage and load resistance.

- **Schematic annotation as a problem-solving tool** – as an instructor you will find most students new to the study of electronics attempt to solve circuit-analysis problems by inspection, or by plugging given values into familiar equations until something resembling an answer emerges. Such efforts are usually fruitless because it is difficult to track the real-world context of every circuit parameter doing things this way. You will need to demonstrate, encourage, and at times *require* your students to apply the alternative strategy of first sketching a schematic diagram of the circuit and then annotating it with all known values. Once the given information gains context through this annotation, appropriate principles (e.g. Ohm's Law, KVL, KCL) are much easier to identify and then properly apply for that circuit. Then, as calculated values get annotated on the same schematic diagram the next logical steps to follow become easier and easier to identify. I recommend using different colors to label different

parameters (e.g. blue for voltage, red for current, black for component values, green for power). Students should have blank "scratch" paper and writing instruments with them at all times, even during discussions, so as to neatly organize information and to articulate and focus their thoughts! Despite introducing some new concepts in the study of transistors, all transistor circuits must still exhibit Kirchhoff's Laws as well as Ohm's Law for its resistors!

- **Relevance of foundational concepts** – this is not a new concept in itself, but rather an approach to learning and to problem-solving that many students struggle with. Students must be encouraged to continually apply trusted concepts such as Ohm's Law, Kirchhoff's Voltage Law, and Kirchhoff's Current Law to every new subject they encounter in Electronics, rather than to ignore these first principles in favor of memorizing new rules. Memorization without understanding is a recipe for failure. Logically reasoning from trusted laws and principles, however, strengthens recall and hones the ability to tackle novel problems. Within the subject of bipolar junction transistors, Ohm's Law, KVL, and KCL are ever-relevant and need only be augmented by a few new facts (e.g. charge carrier injection, depletion regions, beta, etc.) to fully equip oneself for analyzing most any BJT circuit.

- **Creative application of foundational concepts** – many students enter college-level study of electronics with an educational background stressing rote memorization at the expense of logical reasoning from trusted principles, and as such tend to find circuit analysis daunting where there is no single procedure or single formula always yielding the correct answer(s). The educational solution to this is for the instructor to *constantly* tie all new circuit concepts back to previously-learned principles and show the logical relationships between them, and then with every new circuit-analysis example challenge students to reason from those concepts on their own. When students construct a mental "toolbox" they may apply in a plurality of ways to any new challenge they encounter, they will be as free as any skilled mechanic who knows how and why each of their hand tools functions to freely select and use whichever of those tools is most appropriate at any given time. A word of caution for new instructors: *be very sparing in your direction on how to use those tools, instead engaging with students Socratically when they are "stuck" and cannot see a path to a solution!*

1.3 Recommendations for instructors

This section lists realistic student learning outcomes supported by the content of the module as well as suggested means of assessing (measuring) student learning. The outcomes state what learners should be able to do, and the assessments are specific challenges to prove students have learned.

- **Outcome** – Demonstrate effective technical reading and writing

  Assessment – Students present their outlines of this module's instructional chapters (e.g. Case Tutorial, Tutorial, Historical References, etc.) ideally as an entry to a larger Journal document chronicling their learning. These outlines should exhibit good-faith effort at summarizing major concepts explained in the text.

  Assessment – Students show how quantitative results were obtained by the author in the Tutorial chapter's examples.

  Assessment – Students present their "Conceptual Toolbox" Theory Journal entries listing important concepts applied in the Tutorial text(s). A "Conceptual Toolbox" serves as inspiration when solving problems, each concept suggesting a potential problem-solving action. Challenge students to collaboratively generate a Conceptual Toolbox for the day's topic during the class session and then use it when solving problems together, thus fostering the habit of reasoning from principles rather than recalling procedures!

- **Outcome** – Properly associate controlling stimuli to effects in transistors

  Assessment – Identify proper directions of current for bipolar junction transistors; e.g. pose problems in the form of the "Controlling and controlled currents" Conceptual Reasoning question.

  Assessment – Identify proper directions of applied voltage for bipolar junction transistors; e.g. pose problems in the form of the "Controlling voltages" Conceptual Reasoning question.

- **Outcome** – Sketch functional transistor switching circuits

  Assessment – Select appropriate transistor types and orientations to work within given switching circuits; e.g. pose problems in the form of the "Proper BJT types" Conceptual Reasoning question.

- **Outcome** – Calculate voltages and currents in a transistor circuit

  Assessment – Determine all voltages and currents within a bipolar junction transistor circuit given component ratings; e.g. pose problems in the form of the "BJT circuit calculations" Quantitative Reasoning question.

- **Outcome** – Diagnose a faulted transistor circuit

  Assessment – Predict the effect(s) of a single component failing either open or shorted in a bipolar junction transistor circuit; e.g. pose problems in the form of the "Faults in test circuit" Diagnostic Reasoning question.

- **Outcome** – Independent research

Assessment – Locate bipolar junction transistor datasheets and properly interpret some of the information contained in those documents including beta (h_{FE}) ranges, maximum collector current, maximum base-emitter voltage, maximum saturated collector-emitter voltage, maximum power dissipation, etc.

Chapter 2

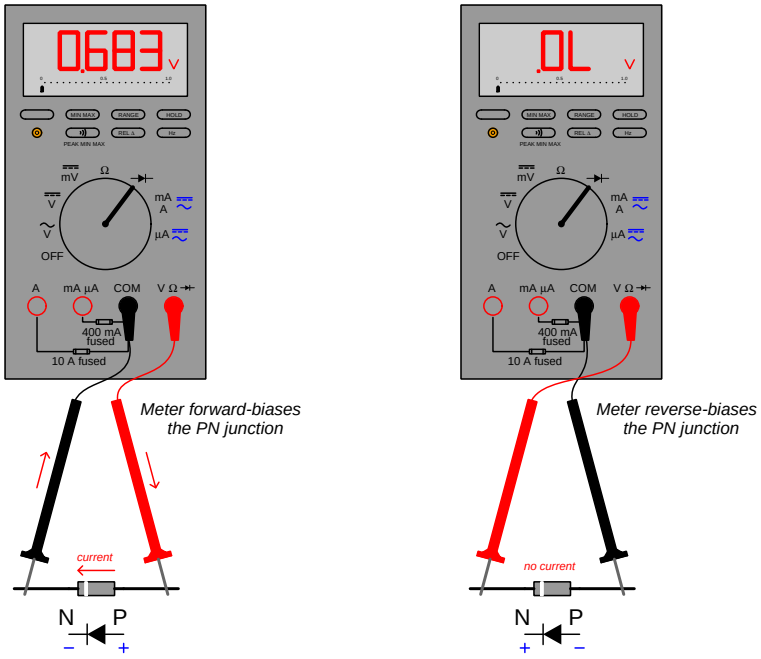
Case Tutorial

The idea behind a *Case Tutorial* is to explore new concepts by way of example. In this chapter you will read less presentation of theory compared to other Tutorial chapters, but by close observation and comparison of the given examples be able to discern patterns and principles much the same way as a scientific experimenter. Hopefully you will find these cases illuminating, and a good supplement to text-based tutorials.

These examples also serve well as challenges following your reading of the other Tutorial(s) in this module – can you explain *why* the circuits behave as they do?

2.1 Experiment: testing BJTs using diode-check multimeter mode

Modern digital multimeters typically offer a *diode-check* mode useful for testing the forward voltage drop of PN semiconductor junctions. This mode is especially useful in identifying the three terminals of any bipolar junction transistor, since a reverse-biased connection to a PN junction will measure "open" (sometimes "OL") on that multimeter mode while a forward-biased connection will register the forward voltage drop of that PN junction as the multimeter sends a small current through it:



The above illustration shows the testing of a standard rectifying diode using a multimeter in its diode-check mode. On the left we see the meter registering the forward voltage drop of the diode's PN junction at 0.683 Volts, and on the right we see the meter indicating the diode's "off' state as its PN junction is reverse-biased by the meter's applied stimulus.

Transistors are more complex than rectifying diodes, of course, but they do possess internal PN junctions just like diodes: one PN junction between the emitter and base terminals, and another between the collector and base terminals. If we use a multimeter's diode-test function to probe these PN junctions we should be able to identify the type of bipolar junction transistor (either NPN or PNP), and with close attention to detail we may even differentiate between the emitter and collector terminals.

Below is a table of data collected from a Fluke model 87-III multimeter set to the diode-check mode, while connecting to the following terminal pairs on a BJT:

Red lead to terminal	**Black lead to terminal**	**Diode-check reading**
1	2	OL
1	3	OL
2	1	0.623 Volts
2	3	OL
3	1	0.625 Volts
3	2	OL

The transistor tested in this case is a model TIP32G which is a PNP power transistor where terminal 1 is the base, terminal 2 the collector, and terminal 3 the emitter. We can identify both the PNP type of this transistor as well as the terminal identifications, though, just from these diode-check measurements knowing that the base-collector terminals connect across one PN junction and the base-emitter terminals connect across another. Note how the only two measurements that were not "OL" were from terminals 2 (+) to 1 (−) and from terminals 3 (+) to 1 (−). The fact that terminal 1 needed to be negative (−) in order to forward bias each of these two PN junctions tells us terminal 1 must be connected to the "N" layer of this transistor and that terminal 1 is the *only* "N" layer in it. From this we may deduce the transistor must be a PNP type and not an NPN type, as an NPN has *two* "N" type layers while the PNP type has only one.

Next we compare the two forward-biased voltage measurements with each other to determine which is larger, and in this case the measurement made between terminals 3 (+) and 1 (−) is the larger of the two. This tells us terminal 3 must be the emitter and terminal 2 the collector because in a bipolar junction transistor the emitter layer is more heavily doped than the collector, and that higher doping of the emitter layer causes the internal electric field across the emitter-base junction to be slightly stronger than that of the collector-base junction. Thus, the PN junction requiring more voltage drop to collapse the depletion region and enable conduction must be the emitter-base junction. This difference in voltage measurements is typically very subtle, in this particular case being only 2 milliVolts! For this reason one should be very careful to not make skin contact with the transistor during these tests, as body heat will warm up the PN junctions and cause their forward voltage drops to slightly change. It is also recommended that the forward voltage drop measurements are repeated in order to grant higher confidence that we know which one is greater.

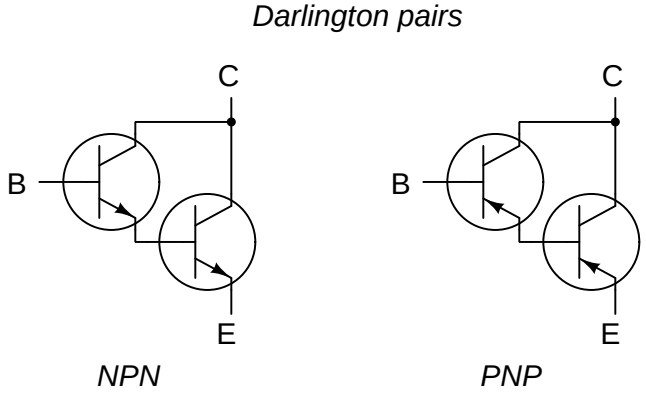
Below is another table of diode-test measurement data, this time made on a TIP31C power transistor:

Red lead to terminal	**Black lead to terminal**	**Diode-check reading**
1	2	0.633 Volts
1	3	0.635 Volts
2	1	OL
2	3	OL
3	1	OL
3	2	OL

Of course, we may simply look up this transistor's identity and terminal assignments in a datasheet knowing the part number, *but where is the sport in that?* As with the previous example our strategy will be to identify the two non-"OL" measurements and tell which terminal is common to both of them. In this case, that common terminal is 1, and in both cases terminal 1 must be positive (+) in order to yield numerical voltage values. Therefore, terminal 1 is the base of this transistor, and it connects to "P" type semiconductor material.

By process of elimination, terminals 2 and 3 must be collector and emitter. We can tell terminal 3 must be the emitter and terminal 2 the collector again by noting that the voltage drop between terminals 3 (−) and 1 (+) is slightly greater than that of terminals 2 (−) and 1 (+), owing to the fact that the emitter layer of a BJT is doped more heavily than the collector layer.

Next we will examine data collected from diode-test mode multimeter measurements made on a *Darlington pair* transistor IC, consisting of two transistors sharing a common collector terminal, with the emitter of one connected to the base of the other. Darlington pair assemblies enjoy extremely high beta (β) ratios, since the current gain of the first transistor multiplies with the current gain of the second.

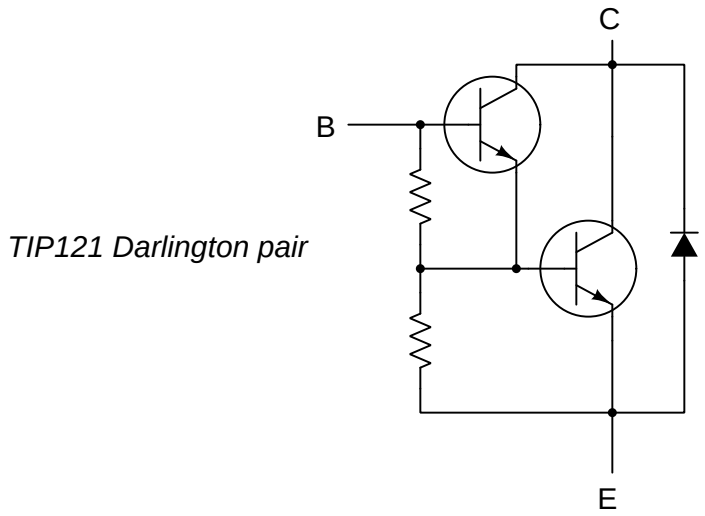


If we use a multimeter's diode-check function to test forward voltage drop on the internal PN junctions of a Darlington pair assembly, we would expect to find a single PN-junction voltage drop between base and collector, and *two* PN-junction voltage drops (in series) between base and emitter, making identification even easier than for a single BJT. Examine data in the following table measured on a model TIP121 Darlington pair:

Red lead to terminal	**Black lead to terminal**	**Diode-check reading**
1	2	0.665 Volts
1	3	0.807 Volts
2	1	OL
2	3	OL
3	1	2.450 Volts
3	2	0.590 Volts

Clearly, something strange is happening here that cannot be explained by the simple Darlington pair model of two bipolar junction transistors sharing a common collector terminal! The model TIP121 must therefore contain more than just a pair of BJTs.

Indeed, this is the case with the model TIP121. According to the datasheet, this IC contains not only two transistors but also one diode and two resistors:



TIP121 Darlington pair

If we also take plain resistance measurements using the same multimeter (set to "Ohms" mode rather than "diode-test" mode) on the same pairs of terminals on the TIP121, we obtain some very useful data because modern digital multimeters output less voltage in their "Ohms" modes so as to not fully forward-bias any PN junctions:

Red lead to terminal	**Black lead to terminal**	**Resistance reading**
1	2	3.63 MΩ
1	3	6.84 kΩ
2	1	OL
2	3	OL
3	1	6.84 kΩ
3	2	3.61 MΩ

Note that two of these resistance measurements are of identical, modest values and taken between the same terminals with opposite polarities (terminals 1 to 3, then terminals 3 to 1, each 6.84 kΩ). These two identical bilateral measurements must be across the two-resistor network inside the TIP121, which lie between the base and emitter terminals. From this we may deduce that the base and emitter terminals are 1 and 3, but we do not yet know which is which. Consulting the diode-test table for the TIP121 Darlington pair array we see that the test from terminal 1 (+) to terminal 3 (−) yielded 0.807 Volts while the test from terminal 3 (+) to terminal 1 (−) yielded a much greater reading of 2.450 Volts. The diode-test measurements were when the multimeter was able to output enough voltage to forward-bias PN junctions, while the resistance ("Ohms") measurements used voltages too low to completely collapse the depletion region of a PN junction. Therefore, while the

resistance measurements clearly show us which two terminals the internal resistors are connected between, the diode-test measurements between those same two terminals identify the PN junction polarities: in this case, terminal 1 must be the "P" and terminal 3 the "N", making terminal 1 the base and terminal 3 the emitter in an NPN Darlington pair.

The diode-test voltage measurement of 0.665 Volts between terminals 1 (+) and 2 (−) is the result of forward-biasing the base-collector PN junction of the first transistor in the TIP121 pair. The diode-test voltage measurement of 0.590 Volts between terminals 3 (+) and 2 (−) is the result of forward-biasing the protection diode stretching between the collector (cathode) and emitter (anode) terminals of the TIP121 Darlington pair assembly. The reason for the voltage measurement between terminals 1 (+) and 3 (−) being only 0.807 Volts rather than approximately double that of a single PN forward voltage drop (0.6 Volts or so) is because the two resistors are in parallel with the two series-connected base-emitter PN junction pairs, providing an alternate path for the meter's diode-test current to flow and thereby reducing the voltage drop.

All resistance measurements yielding mega-Ohm values are the result of forward-biased PN junctions passing *tiny* amounts of current with the multimeter in the "Ohms" mode where its output voltage is limited to less than the typical "on" forward voltage drop of a PN junction. This being only about 100 milliVolts or so, is not enough to *fully* forward-bias any semiconductor PN junction, but recall that the Shockley Diode Equation predicts an exponential relationship between applied forward voltage and junction current:

$$I = I_S \left(e^{\frac{qV}{nKT}} - 1 \right)$$

Where,

I = Forward-bias current through the diode, Amperes

I_S = Reverse-bias saturation current through the diode, Amperes

e = Euler's constant (≈ 2.71828)

V = Voltage applied to the PN junction externally, Volts

q = Elementary charge of an electron (1.602×10^{-19} Coulombs)

n = Ideality factor (1 for a perfect junction)

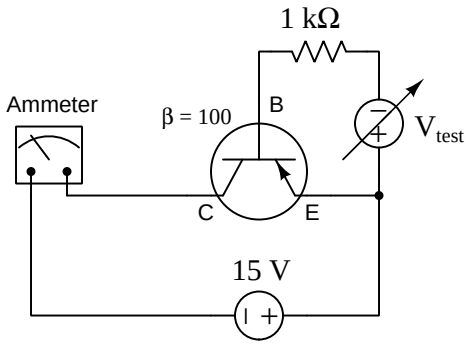
k = Boltzmann's constant ($1.3806504 \times 10^{-23}$ J / K)

T = Absolute temperature (Kelvin), 273.15 more than degrees Celsius

This means we are guaranteed to have at least *some* current pass through a PN junction that is forward-biased, but it will be vanishingly small if the applied voltage is significantly less than the customary 0.6 or 0.7 Volts. Such a tiny current resulting from forward-biasing a silicon PN junction with only 100 milliVolts DC is what causes the meter to register resistances in the mega-Ohm range.

2.2 Example: BJT collector current with varying base stimulus

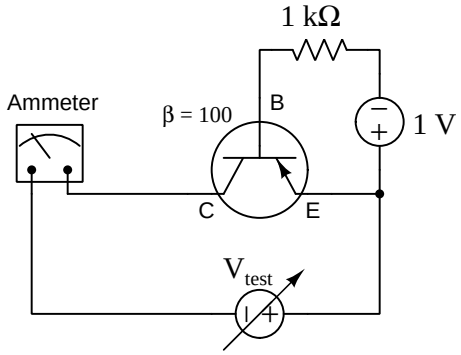
The following circuit is built and tested, with an ammeter sensing current through the transistor's collector terminal (C) while a variable "test" voltage source (V_{test}) is adjusted to provide control over base (B) current:



V_{test}	I_C
0 Volts	0 mA
0.5 Volts	0 mA
0.7 Volts	0 mA
1 Volt	30.0 mA
1.5 Volts	80.0 mA
2 Volts	130.0 mA
2.5 Volts	180.0 mA

2.3 Example: BJT collector current with varying collector-emitter voltage

The following circuit is built and tested, with an ammeter sensing current through the transistor's collector terminal (C) while a variable "test" voltage source (V_{test}) is adjusted to provide control over base (B) current:



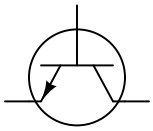
V_{test}	I_C
0 Volts	0 mA
1 Volt	30.0 mA
2 Volts	30.0 mA
3 Volts	30.0 mA
4 Volts	30.0 mA
5 Volts	30.0 mA
6 Volts	30.0 mA
7 Volts	30.0 mA

Chapter 3

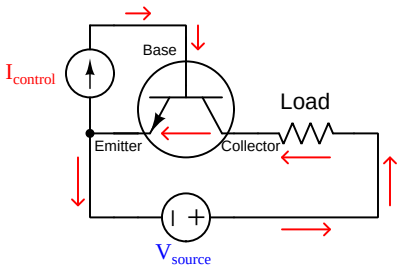
Simplified Tutorial

A *bipolar junction transistor* (*BJT*) is a semiconductor device that may be turned on (or made partially conductive) by introducing a current through its "base" terminal. Two types of BJT exist, *NPN* and *PNP*, referring to the layering of "P" and "N" doped semiconductor materials:

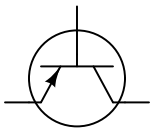
NPN schematic symbol



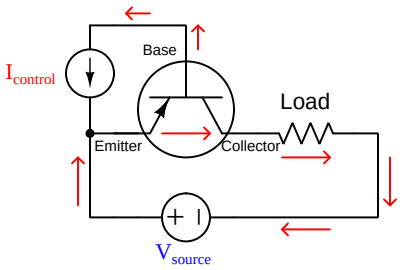
NPN transistor operation



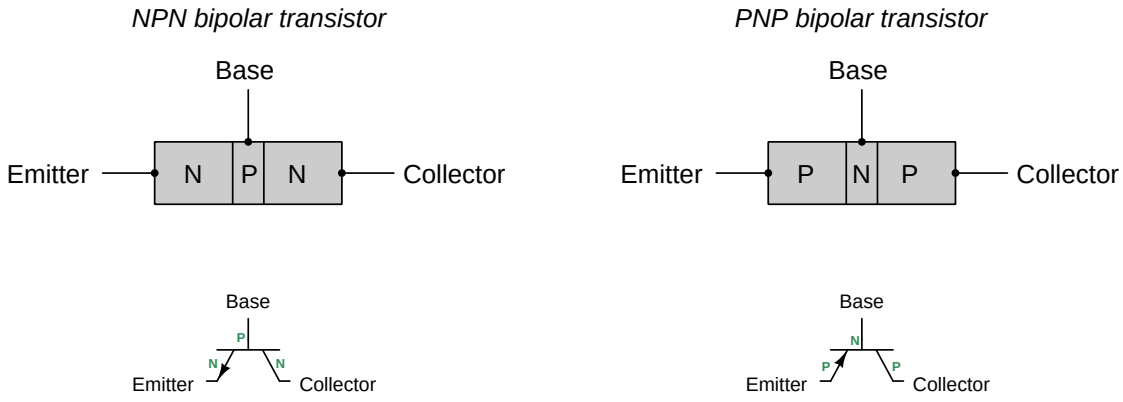
PNP schematic symbol



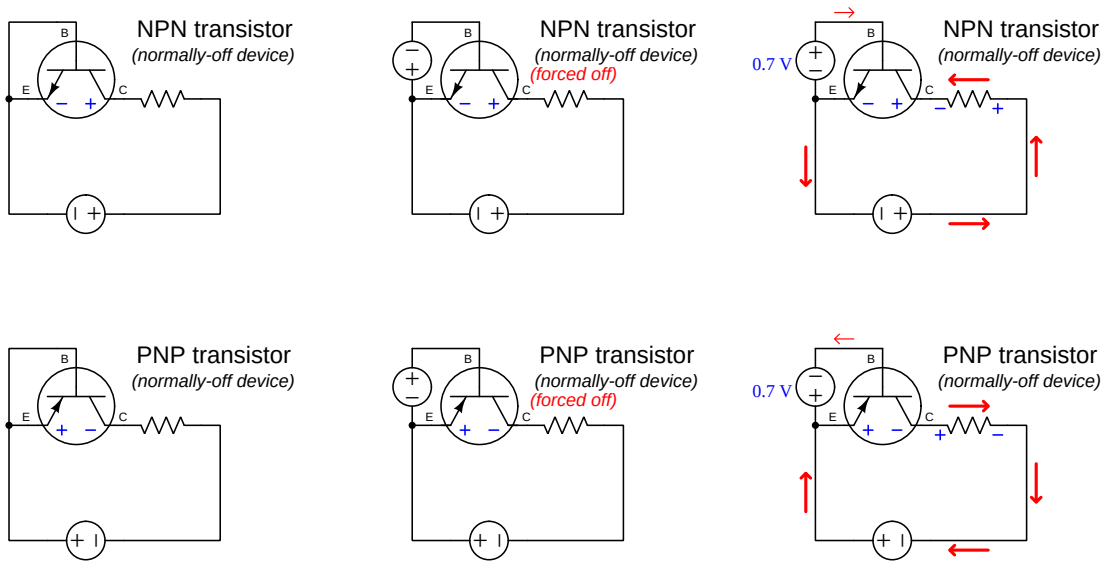
PNP transistor operation



All BJTs operate on the principle of *charge carrier injection*. The thin region of semiconductor material in the middle of the three-layer NPN or PNP "sandwich" is devoid of charge carriers *until* the base-emitter PN junction becomes forward-biased by an external source, at which point those injected charges make the transistor conductive between its emitter and source terminals.



Very little current is necessary through a BJT's base terminal to permit a much larger current to flow through the collector terminal, the ratio of collector current to base current being called "beta" ($\beta = \frac{I_C}{I_B}$), commonly 100 or more. The amount of voltage necessary across the base-emitter junction to forward-bias it is the same as for any other semiconductor PN junction, for silicon it is approximately 0.7 Volts.



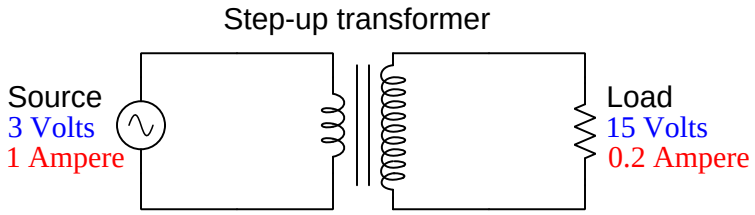
Emitter current is always the sum of collector and base currents, regardless of BJT type, as a consequence of Kirchhoff's Current Law. Current either enters the transistor through the emitter and exits through base and collector, or exits through the emitter and enters through base and collector. In all cases the base and and collector currents "mesh" together to match the direction of the emitter's arrow symbol.

Chapter 4

Full Tutorial

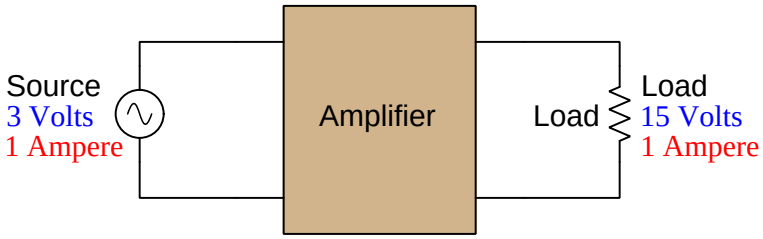
4.1 Amplifier concepts

An *amplifier* is any device allowing a small amount of energy to exert control over a larger amount of energy. It is easy to misunderstand this concept, and so we will begin by exploring what amplification is *not*. Consider a step-up transformer, where an AC source energizes a primary winding, producing a larger voltage at the secondary winding to energize a load:



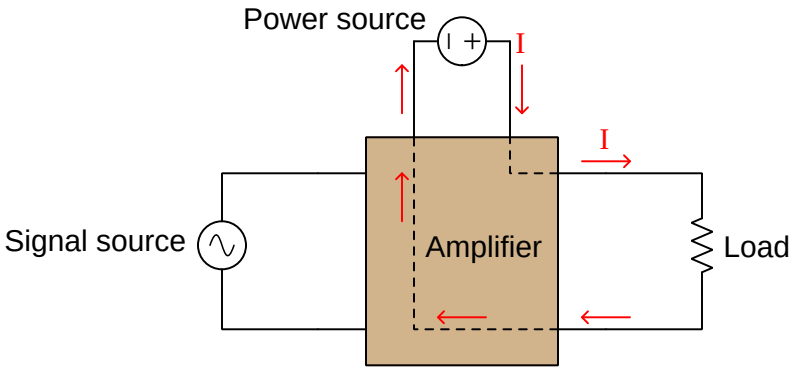
This transformer boosts voltage by a factor of five, but this magnification of voltage is *not* amplification because it comes at the expense of an equivalent reduction in current. Like a mechanical lever used to multiply either force or motion, multiplication of one always leads to division of the other by the same ratio so that energy in equals energy out (in this case, 3 Watts in = 3 Watts out).

An amplifier, then, must be capable of outputting more energy than it receives from its signal source. For example, a circuit boosting voltage while outputting the same current as it inputs would be a true amplifier:



How is this possible, though, without violating the Law of Energy Conservation?

The answer to this very good question is that amplifiers must *also* connect to an additional energy source providing the greater electrical power delivered to the load. The input signal merely controls the flow of electrical energy from source to load. Another way of stating this fact is to say that amplification is a form of *controlled conductivity*, where the input signal determines the degree of conductivity for some other energy source's output to flow to a load:



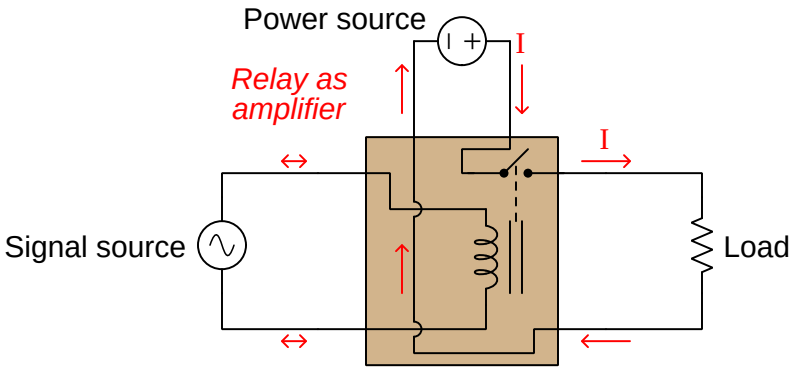
By the same definition, a lever used to magnify force is *not* an amplifying device because you never get more mechanical work out of the lever than what you put into it. However, a motor-driven crane is an amplifying device because a small amount of force and motion applied by a human operator to pushbuttons or levers on the crane *controls the flow of energy from the crane's motor to the load* to lift heavy objects over large distances. The specific component within the crane enabling the operator's weak hand motions to direct many kilowatts' worth of mechanical power to move the load could be a hydraulic valve, a clutch, or some other energy-controlling device. Its specific identity is not important, but its function is essential.

Terms used to distinguish energy-*controlling* devices from other devices, at least in the world of electric circuits, are *active* versus *passive*. Active devices are what make true amplification possible. Components such as resistors, inductors, capacitors, diodes, switches, fuses, and batteries are all passive: they may be sources, they may be loads, and they may alternate roles, or they may be neither, but they do not permit one electrical signal to *control conductivity* for another.

An important way to quantify amplification in any of its forms by a term called *gain*. Simply defined, gain is the ratio of controlled versus controlling quantities, usually symbolized by the variable A. As such, the value of A tells us how many times larger an amplifier's output signal is compared to its input signal. For example, the *power gain* of an amplifier is the ratio of its output signal power divided by the controlling input signal power: $A_P = \frac{P_{out}}{P_{in}}$. The voltage gain of an electronic amplifier is the ratio of output signal voltage to input signal voltage: $A_V = \frac{V_{out}}{V_{in}}$. The current gain of an electronic amplifier is the ratio of output signal current to input signal current: $A_I = \frac{I_{out}}{I_{in}}$. Similarly, we could rate the gain of a motorized crane in terms of force (i.e. the amount of lifting force output divided by the amount of force necessary to push the control lever): $A_F = \frac{F_{lift}}{F_{lever}}$. In every case, gain is a unitless quantity because it is always a ratio between quantities having identical units of measurement (e.g. Watts versus Watts, Volts versus Volts, etc.).

The first active circuit component was the *electromechanical relay* – a combination of solenoid coil and electrical switch. When a power source energizes the solenoid coil, its magnetic field attracts an iron armature which in turn actuates the switch. Relays may be considered active components because the amount of energy conveyed by the switch contact may be much greater than the amount of energy required to energize the coil.

An amplification circuit using a relay is shown in the following schematic diagram:



All energy delivered to the load originates from the power source, not the signal source. All the signal does is energize the relay's coil, to *control* energy sent to the load from the power source. The relay's switch contact then closes to allow the flow of electrical energy to the load, at whatever voltage and current levels the contact is able to handle, which is typically *thousands of times greater*[1] than the energy required to energize the coil.

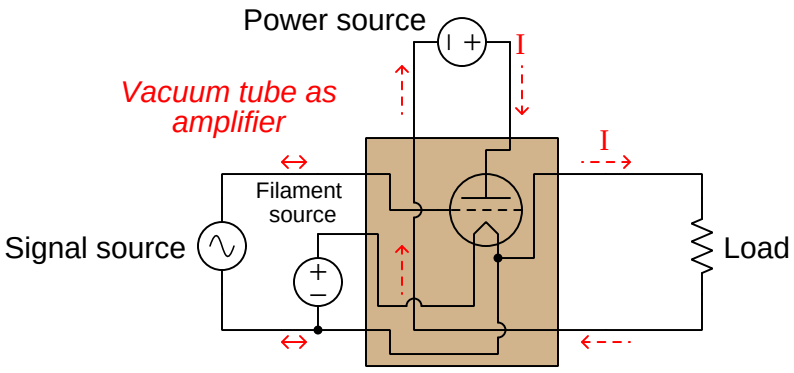
Electromechanical relays revolutionized electric circuits by enabling them to perform functions previously unimaginable. One of these functions is the ability of a fragile switch to exert control over a powerful load: the switch only needs to handle as much current when closed and voltage when open as is necessary to energize the relay coil, while the relay's much larger switch contacts handle the voltage and current requirements of the powerful load. *Telegraph* systems were one early application of relays, where they were used as "repeaters" to boost the power of a telegraph signal weakened by many kilometers of wire resistance, thereby extending the practical range of wired telegraphy. Another function permitted by relays is *automation*, where electrical systems become self-controlling. This is only possible where one electrical signal has the ability to control another, which is the very essence of relay function.

While electromechanical relays made possible such technologies as long-distance telegraph networks, automatic motor control systems, automatic voltage regulators for generators, automatic protection systems for electrical power networks, and even the first digital electric computers, relays do have some significant limitations. Relay contacts are either open or closed, and as such they cannot achieve any states in-between those two. As such, relays are helpless to amplify anything but discrete on-off signals. Amplification of other signal types, such as *analog* signals where information is encoded in the variable magnitude of the signal, is beyond the reach of electromechanical relays.

[1] Thus, we could say that a relay's *power gain* is a unitless value ranging in the thousands.

The next innovation in active devices was the *vacuum tube* or *electron tube*, originally known as the *thermionic valve*. This was a derivative of the incandescent electric lamp, exploiting the fact that a glowing-hot metal surface emits free electrons into the space around it. Consisting of a glass envelope evacuated of air and containing a filament along with one or more metal grids and plates, vacuum tubes used the application of a small voltage between the grid and filament to control the flow of electrons passing from[2] the filament to the plate.

An amplification circuit using a vacuum tube is shown in the following schematic diagram:



Vacuum tubes once again revolutionized electric circuits by enabling them to perform functions previously unimaginable. In fact, this is the innovation that ushered in the era of *electronic* circuits, where the flow of electrons could be precisely controlled by a voltage signal. Unlike relays which are discrete in nature, vacuum tubes are capable of analog operation, gradually restricting the flow of electrons from full rated current to zero, and any level in-between. Vacuum tubes were also much *faster* than relays, allowing a time-varying signal to reproduce its same time-variance at the output. Note in the diagram shown above how the AC signal causes the DC power source to apply a *pulsed* DC signal to the load. It is as though the controlling signal causes an AC-like waveform to be "carved" out of the steady DC supply voltage.

The ability of a vacuum tube to amplify AC signals of arbitrary wave-shape led to the development of new forms of communications including *telephone*, *radio*, and *television*. Prior to vacuum tubes, the only practical form of long-distance electrical communication was the *telegraph*, using discrete on/off states to digitally represent letter and number characters. Vacuum tube electronic circuits also made possible *radar* to detect objects far out of human eyesight, and played a significant role in World War Two. Another significant military application of vacuum tubes was automatic gun direction systems, which when coupled to radar systems made possible automatic defense from aerial attacks.

For all their capabilities, vacuum tubes had significant limitations. Their requirement of a glowing-hot metal filament made them fairly energy-inefficient. They were physically fragile as well, owing to their glass envelope construction. Also, the high-vacuum condition required inside the glass envelope to permit electrons to freely flow between filament and plate was difficult to maintain. Over

[2] Vacuum tubes are one of the few applications where the *electron flow* convention of denoting current makes more sense than *conventional flow*. In the diagram shown here, the red current arrow still points in the direction of conventional flow, even though it is electrons flowing from the "pointed" filament structure to the flat plate structure inside the tube that comprises load current through the tube.

time, ambient air would leak inside the tube and render it useless. Finally, they were limited in current to milliAmperes, requiring high voltage to amplify signals of significant power.

These limitations spurred the invention of active devices based on electricity conducting through solid materials rather than through a vacuum. The most promising of these solid materials was a class of substances called *semiconductors*. A pure "semiconductor" may be defined as a material whose charge carriers inhabit energy states very nearly, but not quite, permitting charge carrier motion within the substance. This stands in contrast to conductors possessing an abundance of free charge carriers, and insulators where the charge carriers are fairly "locked" into place by a lack of available energy states.

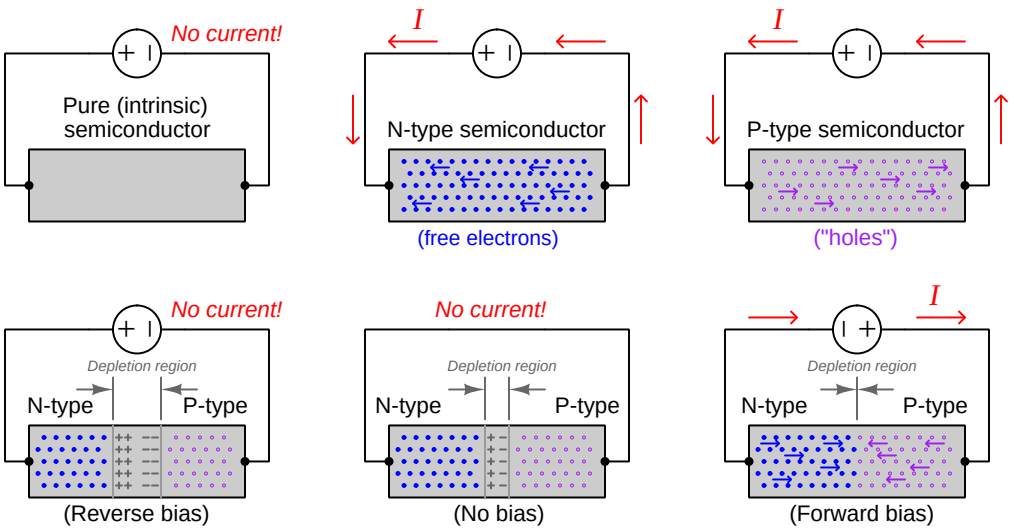
It is not as though a semiconductor is just a poor conductor, either. A conductor of poor quality contains mobile charge carriers, but its charge carriers collide with stationary atoms as they attempt to drift through the atomic lattice of that substance. These collisions convert the charge carriers' kinetic energy into heat, thus constituting electrical resistance. A semiconductor in its pure state, by contrast, contains no free charge carriers at room temperature – i.e. a pure semiconductor is an insulator – but just the right stimulus will *liberate* charge carriers, setting them free to drift through the material's volume. In other words, the phenomenon of *controllable conductivity* is what makes a semiconductor "semi"-conductive.

In a sense, the internal condition of a pure semiconductor is analogous to the internal condition of a vacuum tube: when cool, no free charge carriers exist within the vacuum space of the tube. When stimulated (e.g. heated), however, charge carriers jump off the heated filament and enter the vacuum where they are free to drift, a phenomenon known as *thermionic emission*. In fact, heat is one of the stimuli known to cause semiconductor materials to transition from insulator to conductor, just like a vacuum tube. Light[3] also liberates charge carriers within a pure semiconducting substance.

Through many years of research and experimentation it was discovered that electrical conductivity through a semiconductor could be controlled by carefully alloying ("doping") the semiconducting material with certain types of impurities and applying voltage across junctions of differently-doped semiconductor. The basis of this phenomenon is something called a *depletion region* that forms between differently-doped layers of semiconducting material. Semiconductors doped with impurities placing electrons at energy levels within reach of free motion within the material are called *N-type*. Semiconductors doped with impurities creating vacancies (called *holes*) for low-energy electrons to "hop" into and out of these vacancies from atom to atom are called *P-type*. Both N-type and P-type semiconductors are electrically conductive at room temperature and behave as resistances. Whenever an N-type and a P-type material meet, though, free electrons from the N-type rush to fill vacancies in the P-type, leaving a narrow region at the junction depleted of free charge carriers (i.e. insulating). This so-called *depletion region* may be expanded or collapsed by an applied voltage, making its conduction controllable by a small electrical signal: the key to amplification.

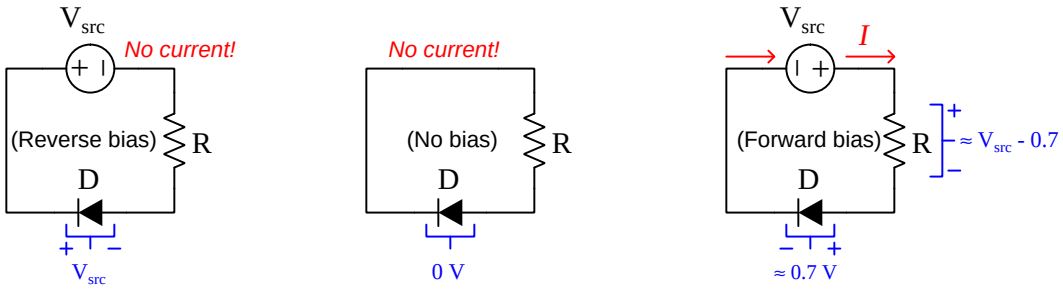
[3] Interestingly, light can also do the same in a vacuum tube, and this is known as the *photoelectric effect*. Not all colors of light will do this, but only those photonic frequencies high enough to meet the *work function* criterion, which is essentially the "escape velocity" energy needed to boost an electron free of a metal surface and into a surrounding vacuum.

A series of illustrations shows how electrical conduction through semiconductors works:



Any section of pure semiconductor, in the absence of heat and light, essentially acts as an insulator and conducts no current. Impurities added to the semiconductor making it either N-type or P-type permits charge carrier motion and thus the semiconductor becomes a conductor. At the junction point of N-type and P-type semiconductors a depletion region forms that is devoid of mobile charge carriers, much like a pure semiconductor, forming a narrow insulating region. However, this depletion region is elastic, and may be expanded or collapsed by the application of an external "bias" voltage.

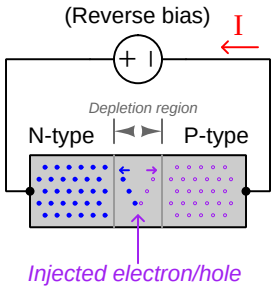
A single junction formed of N-type and P-type semiconductor alloys (called a *PN junction*) is the basis for solid-state *diodes*. When unbiased, a diode is electrically "open", and remains that way until forward-biased with sufficient voltage. The amount of voltage required to fully collapse the depletion region and make it conductive once again is fairly small, approximately 0.5 to 0.7 Volts for a PN junction made of silicon. It is for this reason that approximate calculations made in semiconductor circuits usually assume a 0.5 to 0.7 Volt forward voltage drop in a conducting diode:



Just like the relay and the vacuum tube, the advent of "solid state" semiconductor electronic devices made possible technologies that were previously unimaginable. Solid-state diodes made conversion of AC power to DC simple, reliable, and power-efficient. *Transistors* which are the solid-state equivalent of vacuum tube amplifiers made possible portable radios and televisions as well as the computer technology you are no doubt using to read this very document!

4.2 Injected charge carriers

At this point it is instructive to consider what would happen within a reverse-biased PN junction if somehow we were able to inject a charge carrier – either a free electron or a hole – into the widened depletion region. We know that the depletion region is so-named because it has been depleted of mobile charge carriers and is therefore non-conductive. However, it is important to realize that the depletion region is not an insulator in the strict sense of charge carriers being "locked" into place, rendered immobile by the lack of available energy states to inhabit. A depletion region is non-conducting simply because *no mobile charge carriers happen to exist there*[4]. If such charges were somehow placed there, they would move quite readily, being pulled along by the electric field imposed by the reverse-bias voltage source:



In fact, this is precisely how a *photodiode* functions. A PN junction is reverse-biased by an external source, and whenever sufficient light strikes the PN junction the energy of that photon[5] liberates an electron from a low-energy state creating both a free electron and a hole within the depletion region, both of which move in opposite directions with the impressed electric field to generate a current in the circuit.

The important lesson to take away from this "thought experiment" is that depletion regions are not non-conducting because they *cannot* support free charge carriers, but because they just *don't happen to contain any* free charge carriers at the moment. In other words, depletion regions are non-conducting by *circumstance* and not by *nature*[6]. A true insulator is a material unpermitting[7] of charge carrier motion.

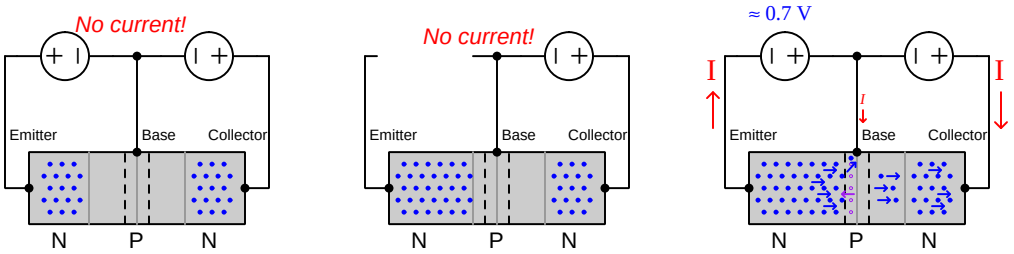
[4] An analogy appreciated by students is the lack of money in their bank accounts. Students are not broke because they are *destined* to be so; they are broke because they just *happen* to lack money at the present time. Once they obtain good-paying jobs (which is often their main purpose in pursuing a technical education), their bank accounts will contain cash and they will be free to experience *cash flow*. The only people destined to be poor are those who *cannot* maintain money in a bank account, whether by internal or external constraints.

[5] A "photon" is simply a particle of light.

[6] This is not unlike the empty space in a vacuum tube: a depletion region is electrically non-conductive for the same reason a perfect vacuum is electrically non-conductive – not because it *cannot* pass charge carriers but because there just happen to be no mobile charge carriers available. Any charge carrier placed within an otherwise perfect vacuum is free to drift with any applied electric field, and any charge carrier placed within a depletion region is similarly free to drift.

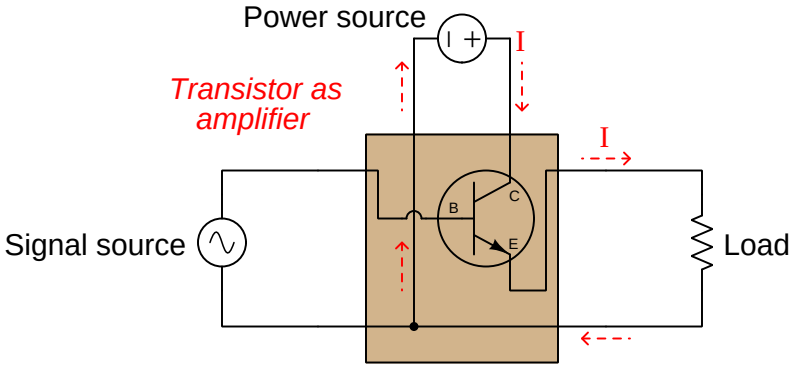
[7] We see this behavior in electrostatic experiments where electric charge carriers may be placed on the surface of an insulating substance, and those charge carriers refuse to move even when subjected to a strong electric field. Consider the case of a capacitor's dielectric layer, where imbalanced electrical charges exist across the thickness of the dielectric but those charge carriers are unable to cross that chasm to achieve lower energy states.

Bipolar junction transistors (BJTs) function by this phenomenon of charge carrier injection. A three-layer "sandwich" of oppositely-doped semiconductor materials (i.e. either PNP or NPN) is made where the middle layer is very thin. The two outer layers are called the *emitter* and the *collector*, while the middle layer is called the *base*. As with all PN-junctions, depletion regions form along each of the two junctions. An NPN-style transistor is shown here in three different modes of stimulus:



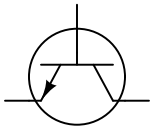
As this illustration shows, a DC voltage source provides a continuous reverse-bias electric field across the base-collector junction while the emitter-base junction is experimentally biased with another voltage source. When the emitter-base junction is either reverse-biased or has no bias at all, no conduction occurs through the transistor. However, once sufficient voltage is applied in forward-bias polarity to the emitter-base junction to collapse that depletion region, many of the electrons entering the very thin base region spill over into the base-collector depletion region where the strong reverse-bias electric field causes them to fall toward the collector.

The ratio of emitted electrons becoming "caught" by the collector versus exiting out the base terminal is typically about 100:1, representing a strong *amplification* of current from base terminal to collector terminal. The amount of controlling voltage necessary to collapse the emitter-base junction (approximately 0.7 Volts DC) is also significantly less than the amount of voltage that may be impressed across the base-collector junction, and so the combination of small base current and small emitter-base voltage versus large collector current and large base-collector voltage represents a very large *power gain* from input to output. Thus, the transistor meets the goal of analog electrical signal amplification without the need for fragile vacuum envelopes or power-hungry filaments:

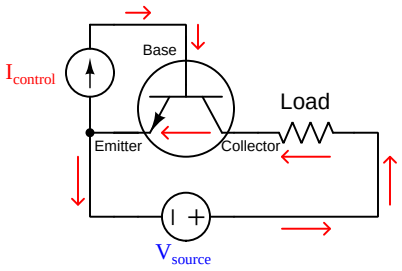


Shown below are diagrams of NPN and PNP transistors, respectively, being biased into a state of conduction by a small amount of base current labeled as $I_{control}$:

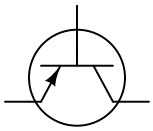
NPN schematic symbol



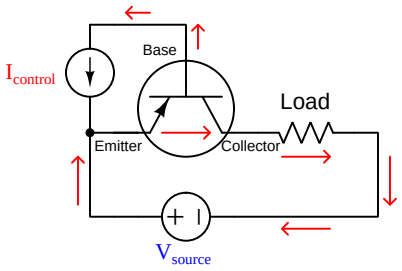
NPN transistor operation



PNP schematic symbol



PNP transistor operation

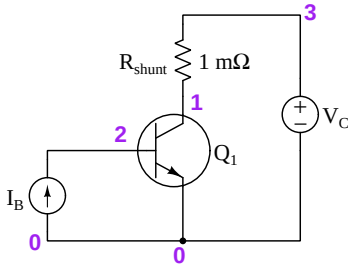


Neither transistor will permit current through its collector terminal unless and until this small controlling current passes through the base terminal. Also, note how for both transistors the emitter current is the sum of base and collector currents (by way of Kirchhoff's Current Law, if we consider the transistor itself to be a node where these three currents intersect).

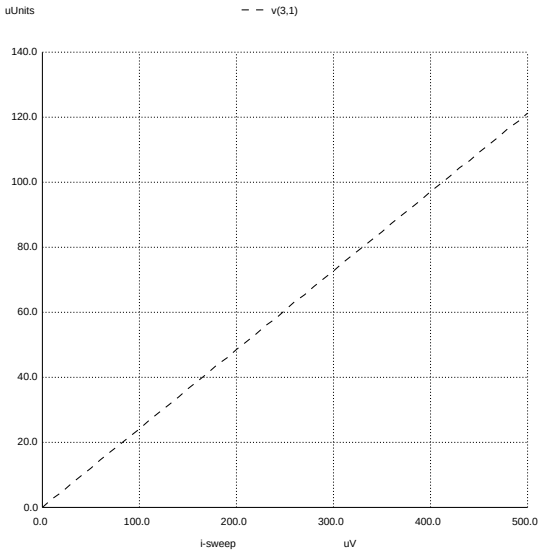
4.3 BJT characteristics

One of the most common bipolar junction transistor amplifier circuit configurations is the *common-emitter*, so-called because the emitter[8] terminal is shared in common by both the DC power supply and the input signal source. This circuit configuration serves well to experimentally explore transistor characteristics, as it allows us to vary the input signal and emitter-collector voltage independently of each other. A shunt resistor connected in series with the transistor's collector terminal allows us to take current measurements using a voltmeter.

If we hold the DC voltage source (V_C) at a constant voltage and vary the base current, we see that the transistor's collector current (as sensed by voltage dropped across R_{shunt} in this SPICE analysis[9]) increases linearly with base current:




```
* SPICE netlist
vc 3 0 dc 10
ib 0 2 dc
q1 1 2 0 realistic
rshunt 3 1 1e-3
.model realistic npn
+ is=15f vaf=90 bf=220
+ cjc=2p mjc=0.3 vjc=0.72
+ cje=5p mje=0.3 vje=0.72
+ nc=2 rc=0
.dc ib 0 0.5m 0.01m
.plot dc v(3,1)
.end
```

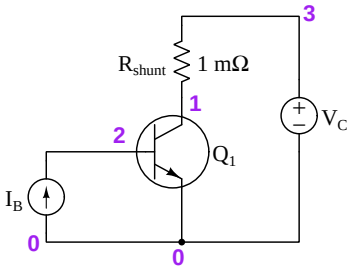


This behavior should not surprise us: with a constant DC bias between the emitter and collector terminals, current through the collector should be directly proportional to current through the base because it is the emitter-base forward-biasing that allows any charge carriers at all to be "caught" by the collector. The more we forward-bias the emitter-base junction, the more charge carriers get injected through the base into the collector. What we see here is linear *amplification* of the base current input signal.

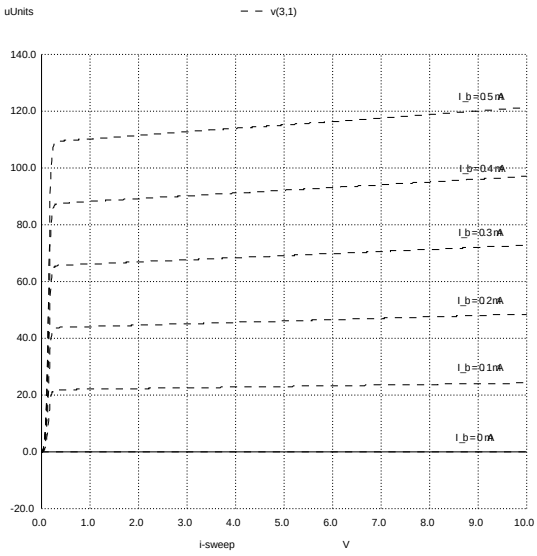
[8] The emitter terminal of a bipolar junction transistor is identified by the presence of an arrowhead. The collector terminal is on the opposite side of the transistor, and the base terminal is in the middle. The arrow's direction defines the P- versus N- doping of the layers: an outward-pointing arrow signifies an NPN transistor (as shown in the example diagram) while an inward-pointing arrow signifies a PNP transistor. This convention is in accord with standard diode labeling, where the arrowhead represents the "P" side and the line represents the "N" side. The only difference here is that the arrowhead's tip does not touch a perpendicular line segment as it does in the diode symbol.

[9] The `.dc` analysis card instructs SPICE to sweep I_b from 0 mA to 0.5 mA in 0.01 mA increments. Note the `.model` card in the SPICE deck, specifying operating parameters for a "realistic" NPN transistor. Model specifications are common in SPICE analyses, and are very useful because they allow the user to define individual characteristics of each device, or in this case a class of NPN transistors called "realistic".

If we set the base current value at constant levels while varying the DC voltage source (V_C), we notice very different behavior. The transistor's collector current rises rapidly as collector voltage increases from zero, then the collector current holds nearly steady as the supply voltage rises to its full value. It is as though collector current hits a limit and does not increase much beyond that limit as collector voltage rises. In the following SPICE analysis we set I_B to five different values (0 mA, 0.1 mA, 0.2 mA, 0.3 mA, 0.4 mA, and 0.5 mA) while varying V_C from 0 to 10 Volts, to achieve a "family" of five curves representing collector current as a function of collector voltage:




```
* SPICE netlist
vc 3 0 dc
ib 0 2 dc
q1 1 2 0 realistic
rshunt 3 1 1e-3
.model realistic npn
+ is=15f vaf=90 bf=220
+ cjc=2p mjc=0.3 vjc=0.72
+ cje=5p mje=0.3 vje=0.72
+ nc=2 rc=0
.dc vc 0 10 0.01 ib 0 0.5m 0.1m
.plot dc v(3,1)
.end
```



This is remarkable behavior for an electrical component, because it acts as a *current regulator*, the collector current regulation limit being a multiple of base current. Unlike resistances which follow Ohm's Law, transistor collector terminal current is most definitely *not* proportional to the voltage impressed across the collector and emitter terminals.

The reason for this behavior has everything to do with charge carrier injection through the transistor's base layer. It is important to know that Ohm's Law is true for a resistance because in normal conductors mobile charge carriers are plentiful. Even in poor conductors, there is usually an abundance of charge carriers able to move throughout the material, that motion impeded only by collisions between the moving charge carriers and stationary atoms in the substance. Greater voltage applied across a conductor, therefore, results in more charge carriers per second drifting through the material: those charge carriers simply collide more violently as drift velocity increases, but there is no theoretical limit to their drift rate. In a depleted semiconductor, however, mobile charge carriers are very scarce. In an ideal depletion region there are *zero* free charge carriers endemic to the substance – in an operating transistor, the only charge carriers available to pass through the collector are those injected into it from the base. With any given amount of base current, increasing emitter-collector voltage merely causes the individual charge carriers to drift faster, but it cannot cause the *total number* of charge carriers per second to increase because there simply are no more charge carriers to capture. Collector current hits a maximum value constrained by the rate of charge carrier injection, which is itself proportional to base current.

The following comparative analogy helps differentiate charge carrier motion through a normal conductor versus injected charge carrier motion through a depleted semiconductor region:

- Imagine a sandy desert plain on a windy day: wind picks up grains of sand from ground level and causes them to drift. If wind speed increases, the total number of sand grains moving past any point per second increases. This is analogous to current through a conductor, with sand grains playing the role of charge carriers and wind playing the role of applied voltage. If wind speed (voltage) increases, so does the rate of sand drift (current), because grains of sand are plentiful. This proportionality between voltage and current for any regular conductor of electricity is called *resistance*, and is defined as such in Ohm's Law ($R = \frac{V}{I}$).

- Now imagine a rocky plateau on a windy day, the ground devoid of any loose sand, analogous to the depletion region between P-type and N-type semiconductors. No matter how much wind there may be, there is no sand drift on the plateau because *no grains of sand are there to drift*. Now imagine standing on that rocky plateau holding a handful of sand, slowly letting grains of sand fall from your hand into the air. Any wind, of course, will cause the falling sand grains to drift away from you, but the speed of the wind has no bearing whatsoever on the number of sand grains per second drifting downwind – that is strictly controlled by the rate of sand falling from your grip. Higher wind speed may increase the *velocity* of each sand grain, but it cannot increase the *number of grains per second* drifting downwind because that is fixed by the rate at which you release them from your hand. Such is the case for charge carriers injected into a depleted region: within a region of material lacking native charge carriers of its own, the rate at which we inject new charge carriers into that region becomes a limiting factor for current through it regardless of applied voltage across that region.

Any homogeneous piece of P-doped or N-doped semiconductor acts as a plain conductor of electricity because the donor electrons or acceptor holes implanted into that material during the doping process gives it permanent placeholders for mobile charge carriers. This means that solely-P or solely-N semiconductor material exhibits greater current with greater applied voltage at all times, just like a resistance. However, in the charge-carrier-depleted region that arises at the interface of a P-type and N-type semiconductor, some external stimulus must act to create free charge carriers within, and so that depletion region won't conduct any electricity unless and until we stimulate it to do so.

As we saw previously, one way to inject mobile charge carriers into a depletion region is to shine light (of the correct color) on it, where the photons of light will knock valence-band electrons into the conduction band and leave holes behind. So far as the incident light persists, that depletion region will be electrically conductive. Bipolar junction transistors use a different method to inject mobile charge carriers into the base-collector depletion region, and that is by forward-biasing the emitter-base PN junction and making the base region thin enough that many of the charge carriers crossing from emitter to base happen to wander into the base-collector region where they get "swept" into the collector by the base-collector voltage.

This is how bipolar junction transistors regulate collector current: the forward-biasing voltage applied between base and emitter controls the rate at which minority charge carriers enter the depleted base region, while the collector-emitter voltage merely sweeps those charge carriers through to the collector. This makes collector current a strong function of base excitation and a very weak function[10] of collector voltage, so that from the perspective of source V_C the transistor is a current-regulating device. In other words, the maximum amount of collector current is *limited* by the amount of base current: reducing base current will force collector current to decrease proportionately, no matter how strong of a voltage exists between collector and emitter.

The ratio of maximum collector current to base current[11] is symbolized either as the Greek letter *beta* or by h_{FE}, and may reach values exceeding 100 for small transistors. Power transistors tend to exhibit lesser β ratios, but still vastly greater than one:

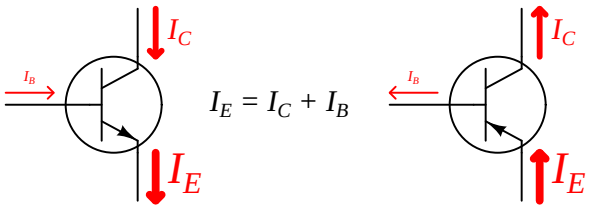
$$\beta = h_{FE} = \frac{I_C}{I_B}$$

It is this over-unity ratio that defines the bipolar junction transistor as an *amplifying* device, in that the amount of current in the *controlled* portion of the circuit (I_C) is much larger than the amount of current in the *controlling* portion of the circuit (I_B).

[10] Close inspection of the I_C-to-V_C graphs do reveal a slight positive slope on the "level" portion of each trace, indicating a slight increase in collector current with increasing collector voltage. This is called the *Early effect*, named in honor of James Early who was the first to explain its cause. Apparently, while working at Bell Laboratories with William Shockley and other pioneers of solid-state electronic devices, Dr. Early was sitting through a boring meeting and suddenly realized why an increased collector voltage would cause a slight increase in collector current: an increasing V_{CE} would cause the base-collector depletion region to widen (this is known as "base width modulation") and encroach more upon the base layer, thereby making it more likely for charge carriers to get swept up into the collector rather than exiting out the base terminal. If one projects the slopes of all the curves to the left, it will be found that all those lines intersect on the x-axis at some negative collector voltage value. This is called the *Early voltage*. To extend the grains-of-sand analogy, not all the grains of sand dropped from your hand will be fully carried by the wind. By chance, some of them will fall to the ground sooner than others and thereby travel a shorter distance. Increasing wind speed makes these stunted journeys less likely, resulting in a higher percentage of dropped sand grains being carried farther downwind. Therefore, the rate of sand grains passing by any given location downwind is a strong function of the rate at which you drop them (i.e. base current), and a weak function of wind speed (i.e. collector-emitter voltage).

[11] We could properly call this ratio the *current gain* of the transistor.

Although we define β as the ratio of maximum collector current to base current, this parameter is useful for relating either base or collector current to emitter current as well. The key to relating I_E and β is Kirchhoff's Current Law, shown in the following diagrams for both NPN- and PNP-type transistors:



Regardless how complex the circuit component may be internally, Kirchhoff's Current Law always applies because it is based on the fundamental principle of *Electric Charge Conservation*. Every charge carrier passing through the emitter *must* either pass through the base or through the collector.

Knowing the definition of β and how Kirchhoff's Current Law relates to transistor terminal currents, we may use algebraic substitution and manipulation to arrive at useful ratios between every pair of transistor terminal currents. Let's begin with the ratio between emitter current and base current:

$$\frac{I_E}{I_B} = \frac{I_C + I_B}{I_B} = \frac{I_C}{I_B} + \frac{I_B}{I_B} = \frac{I_C}{I_B} + 1$$

$$\frac{I_E}{I_B} = \beta + 1$$

This should make intuitive sense as well. If the ratio of maximum collector current to base current is β, and we know emitter current is slightly greater than collector current, then we would likewise expect the ratio of maximum emitter current to base current to be slightly more than β.

Next, let's explore the ratio between emitter and collector currents, which we know to be nearly equal to each other (with emitter current being slightly larger):

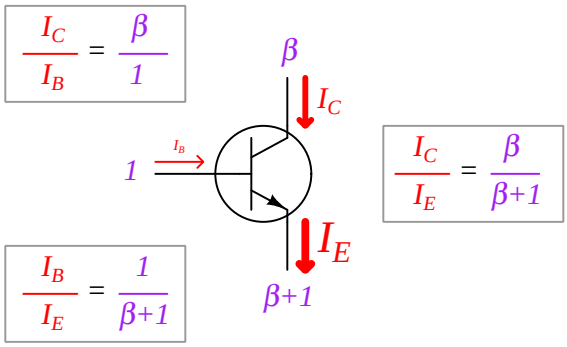
$$\frac{I_E}{I_C} = \frac{I_C + I_B}{I_C} = \frac{I_C}{I_C} + \frac{I_B}{I_C} = 1 + \frac{I_B}{I_C}$$

$$\frac{I_E}{I_C} = 1 + \frac{1}{\beta} = \frac{\beta}{\beta} + \frac{1}{\beta}$$

$$\frac{I_E}{I_C} = \frac{\beta + 1}{\beta}$$

Again, this agrees with our intuition. As emitter current is slightly greater than collector current, $\beta + 1$ is slightly greater than β. To summarize, the ratio between collector and base currents is β, the ratio between emitter and base currents is $\beta + 1$, and the ratio between emitter and collector currents is $\frac{\beta+1}{\beta}$.

A good way to remember these β-based formulae is to ponder a simple example of a transistor having a base current of 1 unit (i.e. 1 milliAmpere, 1 microAmpere, etc.), which of course would mean its collector carries β units of current and its emitter $\beta + 1$ units of current:



If you prefer, simple numerical values could be substituted for β and $\beta + 1$. If we were to assume a β ratio of one hundred, we would label the collector and emitter currents as 100 and 101 units, respectively. Then the ratio $\frac{I_C}{I_E}$ would then be $\frac{100}{101}$, the ratio $\frac{I_C}{I_B}$ would then be $\frac{100}{1}$, and the ratio $\frac{I_B}{I_E}$ would then be $\frac{1}{101}$. This is a good example of problem simplification by converting an algebraic problem into a numerical problem.

Calculating the amount of power dissipated by a bipolar junction transistor is a function of Joule's Law, of course, but it is complicated by the fact that transistors are three-terminal devices and as such have multiple currents and multiple terminal pairs between which to measure voltage. In order to understand how power dissipation works for a three-terminal device, it is helpful to review just what voltage, current, and power are and how they inter-relate.

Remember that *voltage* is defined as the amount of potential energy either gained or lost by an electric charge carrier as it moves from one location to another. We measure voltage in the unit of the *Volt*, but this unit is really a composite of more fundamental units: one Volt is equivalent to a Joule of potential energy per Coulomb of electric charge. *Current* is the rate at which charge carriers move in a circuit, and we measure it in the unit of the *Ampere*, but like the Volt this is also a composite unit: Coulombs of charge passing by per second of time.

When we multiply voltage by current to calculate power (Joule's Law), what we are really doing is canceling some of these fundamental units to arrive at a new composite unit. Below we see the familiar Joule's Law equation and below that we see it written in terms of units[12]:

$$VI = P$$

$$\left(\frac{[\text{Joules}]}{[\text{Coulomb}]}\right)\left(\frac{[\text{Coulombs}]}{[\text{second}]}\right) = \frac{[\text{Joules}]}{[\text{second}]}$$

Joules per Coulomb (Volts) multiplied by Coulombs per second (Amperes) equals Joules per second (Watts), which is a rate of energy transfer – the very definition of *power*.

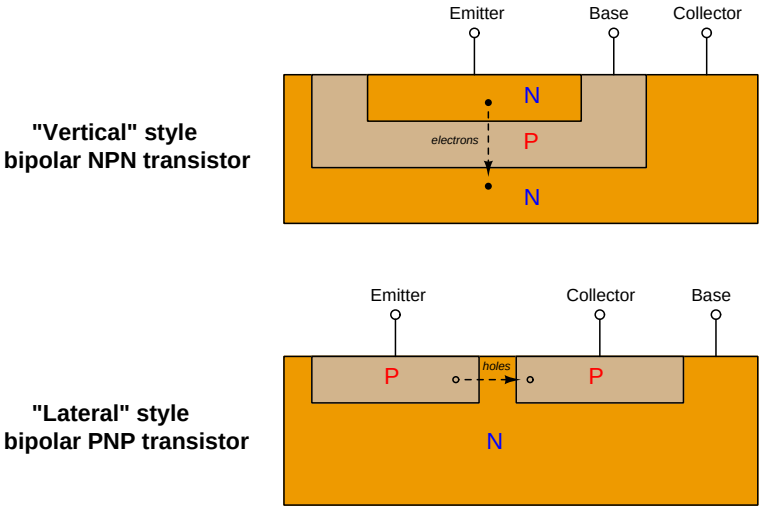
Turning our attention back to transistors, it is helpful to imagine the journey of charge carriers passing through a bipolar junction transistor in order to conceptualize power dissipation for this type of device. There are only two pathways for charge carriers to move through a BJT: between base and emitter (I_B), and between collector and emitter (I_C). Each pathway incurs its own energy-dissipation "penalty", expressed as V_{BE} and V_{CE}, respectively. So, if every Coulomb of charge carriers passing through the base terminal suffers a V_{BE} energy loss, and every Coulomb of charge carriers passing through the collector terminal suffers a V_{CE} energy loss, the Law of Energy Conservation demands their combined energy losses be equal the total dissipation of the transistor. Therefore, we may calculate total transistor power dissipation as the sum of the base-emitter power and the collector-emitter power:

$$P_{total} = I_B V_{BE} + I_C V_{CE}$$

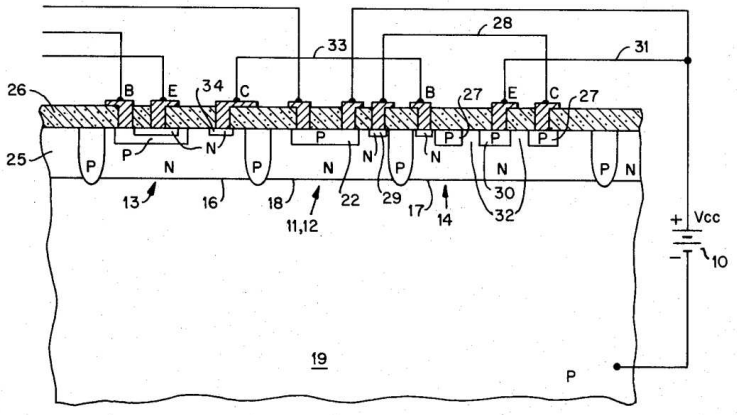
[12] This is called *dimensional analysis*, when we write an equation with units of measurement rather than standard variable symbols and then check to see that the units match.

4.4 BJT construction

Bipolar junction transistors are made by repeatedly doping and etching semiconductor material to form the three-layer "sandwich" structure of emitter-base-collector. Cross-sectional illustrations of so-called "Vertical NPN" (VNPN) and "Lateral PNP" (LPNP) transistors appear below:

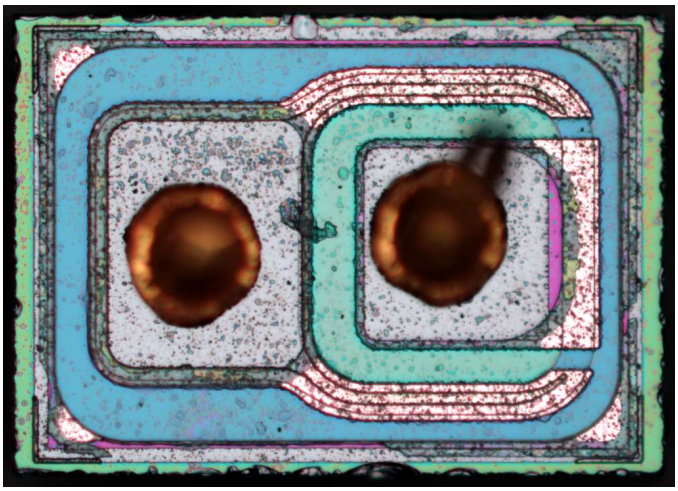


Lateral-style BJTs don't perform as well as vertical BJTs, but are found in integrated circuits where NPN and PNP transistors must co-exist on the same type of substrate as seen in the following illustration taken from United States Patent 3,829,709 (by Robert Maigret and Geoffrey King of Micro Components Corporation) figure 4. Here we see a VNPN at area 13, and an LPNP at area 14, the lowest active layer of both transistors being the same N-doped semiconductor material (16 and 17), layered atop P-type substrate (19):



A micro-photograph of a real bipolar transistor is shown below, courtesy of the creators of the website *zeptobars* (`https://zeptobars.com`) under the Creative Commons Attribution 3.0 Unported License:

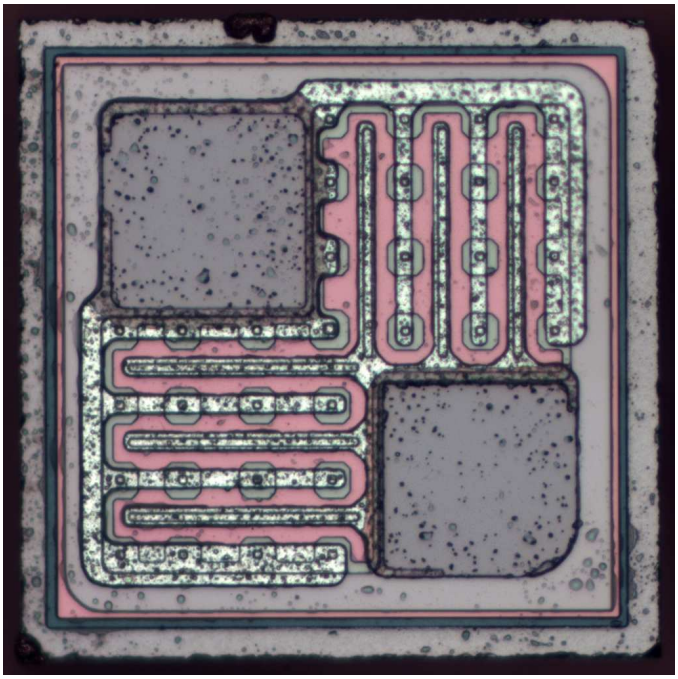
Infineon model BC858C NPN transistor, die size 337 μm by 238 μm
Image source: `s.zeptobars.com/Infineon-BC858CE-HD.jpg`



The two reddish-orange dots are the contact points between the metal bond wires and the silicon (green-colored emitter on right and blue-colored base on left). This view of the transistor's silicon die is from the top looking down, and so the collector bond wire is on the bottom of the die out of view.

A design strategy for minimizing contact resistance between bond wires and silicon is to arrange an array of metal "fingers" on the surface of the silicon die to maximize contact area and distribute current over the widest area of silicon possible. Another BJT micro-photograph appears below showcasing this design. Once again, the collector contact is not shown here as it is on the underside of the die and we are viewing the top of the die looking down. This is another photo from the website *zeptobars* (`https://zeptobars.com`) offered under the Creative Commons Attribution 3.0 Unported License:

OnSemi model MMBT2222A NPN transistor, die size 343 μm by 343 μm
Image source: `s.zeptobars.com/MMBT2222A-OnSemi-HD.jpg`



The light-colored "fingers" are the metal strips providing large-area contact between the bond wires (not shown) and the underlying silicon emitter (pink-colored) and base (grey-colored) regions.

Chapter 5

Historical References

This chapter is where you will find references to historical texts and technologies related to the module's topic.

Readers may wonder why historical references might be included in any modern lesson on a subject. Why dwell on old ideas and obsolete technologies? One answer to this question is that the initial discoveries and early applications of scientific principles typically present those principles in forms that are unusually easy to grasp. Anyone who first discovers a new principle must necessarily do so from a perspective of ignorance (i.e. if you truly *discover* something yourself, it means you must have come to that discovery with no prior knowledge of it and no hints from others knowledgeable in it), and in so doing the discoverer lacks any hindsight or advantage that might have otherwise come from a more advanced perspective. Thus, discoverers are forced to think and express themselves in less-advanced terms, and this often makes their explanations more readily accessible to others who, like the discoverer, comes to this idea with no prior knowledge. Furthermore, early discoverers often faced the daunting challenge of explaining their new and complex ideas to a naturally skeptical scientific community, and this pressure incentivized clear and compelling communication. As James Clerk Maxwell eloquently stated in the Preface to his book *A Treatise on Electricity and Magnetism* written in 1873,

> It is of great advantage to the student of any subject to read the original memoirs on that subject, for science is always most completely assimilated when it is in its nascent state . . . `[page xi]`

Furthermore, grasping the historical context of technological discoveries is important for understanding how science intersects with culture and civilization, which is ever important because new discoveries and new applications of existing discoveries will always continue to impact our lives. One will often find themselves impressed by the ingenuity of previous generations, and by the high degree of refinement to which now-obsolete technologies were once raised. There is much to learn and much inspiration to be drawn from the technological past, and to the inquisitive mind these historical references are treasures waiting to be (re)-discovered.

5.1 John Bardeen's and Walter Brattain's transistor patent

Two researchers working at Bell Laboratories in the late 1940's helped develop the world's first practical bipolar junction transistor. Their US Patent (number 2,524,035 granted in October of 1950) describes the operation of the device in some detail. The following text comes from page 1 of their patent, describing the purpose of their invention:

> This invention relates to a novel method of and means for translating electrical variations for such purposes as amplification, wave generation, and the like.
>
> The principal object of the invention is to amplify or otherwise translate electric signals or variations by use of compact, simple, and rugged apparatus of novel type.
>
> Another object is to provide a circuit element for use as an amplifier or the like which does not require a heated thermionic cathode for its operation, and which therefore is immediately operative when turned on. A related object is to provide such a circuit element which requires no evacuated or gas-filled envelope.

Note the explicit contrast to vacuum tube technology of the time, which required a heated filament ("thermionic cathode") which was slow to come up to operating temperature, and which also required a sealed envelope (i.e. tube) for proper operation. Clearly, the aim of this invention was to overcome these practical limitations.

The authors then go on to describe early attempts to design what we now know as a *field-effect transistor*, but which was met with failure:

> Attempts have been made in the past to convert solid rectifiers utilizing selenium, copper sulfide, or other semi-conductive materials into amplifiers by the direct expedient of embedding a grid-like electrode in a dielectric layer disposed between the cathode and the anode of the rectifier. The grid is supposed, by exerting an electric force at the surface of the cathode, to modify its emission and so, alter the cathode-anode current. As a practical matter it is impossible to embed a grid in a layer which is so thick as to insulate the grid from the other electrodes and yet so thin as to permit current to flow between them. It has also been proposed to pass a current from end to end of a strip of homogeneous isotropic semiconductive material and, by the application of a strong transverse electrostatic field, to control the resistance of the strip, and hence the current through it.
>
> So far as is known, all of such past devices are beyond human skill to fabricate with the fineness necessary to produce amplification. In any event they do not appear to have been commercially successful.

With this, the authors proceed to describe the behavior of different types of semiconducting material, as a prelude to a theoretical explanation of their transistor:

> It is well known that in semiconductors there are two types of carriers of electricity which differ in the signs of the effective mobile charges. The negative carriers are excess electrons which are free to move, and are denoted by the term conduction electrons or simply electrons. The positive carriers are missing or defect "electrons," and are denoted by the term "holes." The conductivity of a semiconductor is called excess or defect, or

N or P type, depending on whether the mobile charges normally present in excess in the material under equilibrium conditions are electrons (Negative carriers) or holes (Positive carriers).

When a metal electrode is placed in contact with a semiconductor and a potential difference is applied across the junction, the magnitude of the current which flows often depends on the sign as well as on the magnitude of the potential. A junction of this sort is called a rectifying contact. If the contact is made to an N-type semiconductor, the direction of easy current flow is that in which the semiconductor is negative with respect to the electrode. With a P-type semiconductor, the direction of easy flow is that in which the semiconductor is positive. A similar rectifying contact exists at the boundary between two semiconductors of opposite conductivity types.

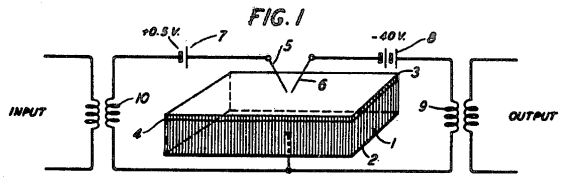
This boundary may separate two semiconductor materials of different constitutions, or it may separate zones or regions, within a body of semiconductor material which is chemically and stoichiometrically uniform, which exhibit different conductivity characteristics.

The present invention in one form utilizes a block of semiconductor material on which three electrodes are placed. One of these, termed the collector, makes rectifier contact with the body of the block. The other, termed the emitter, preferably makes rectifier contact with the body of the block also. The third electrode, which may be designated the base electrode, preferably makes a low resistance contact with the body of the block. When operated as an amplifier, the emitter is normally biased in the direction of easy current flow with respect to the body of the semiconductor block. The nature of the emitter electrode and of that portion of the semiconductor which is in the immediate neighborhood of the electrode contact is such that a substantial fraction of the current from this electrode is carried by charges whose signs are opposite to the signs of the mobile charges normally in excess in the body of the semiconductor. The collector is biased in the reverse, or high resistance direction relative to the body of the semiconductor. In the absence of the emitter, the current to the collector flows exclusively from the base electrode and is impeded by the high resistance of this collector contact. The sign of the collector bias potential is such as to attract the carriers of opposite sign which come from the emitter. The collector is so disposed in relation to the emitter that a large fraction of the emitter current enters the collector. The fraction depends in part on the geometrical disposition of the electrodes and in part on the bias potentials applied. As the emitter is biased in the direction of easy flow, the emitter current `[Page 1]`

is sensitive to small changes in potential between the emitter and the body of the semiconductor, or between the emitter and the base electrode. Application of a small voltage variation between the base electrode and emitter causes a relatively large change in the current entering the semiconductor from the emitter, and a correspondingly large change in the current to the collector. One effect of the change in emitter current is to modify the total current flowing to the collector, so that the overall change in collector current may be greater than the change in the emitter current. The collector circuit may contain a load of high impedance matched to the internal impedance of the collector, which, because of the high resistance rectifier contact of the collector, is high. As a

result, voltage amplification, current amplification, and power amplification of the input signal are obtained. [Page 2]

Theirs was a *point-contact* type of transistor, where the emitter and collector were formed by pointed wire ends touching a block of semiconducting material (germanium rather than silicon), as shown here in Figure 1 of the patent:



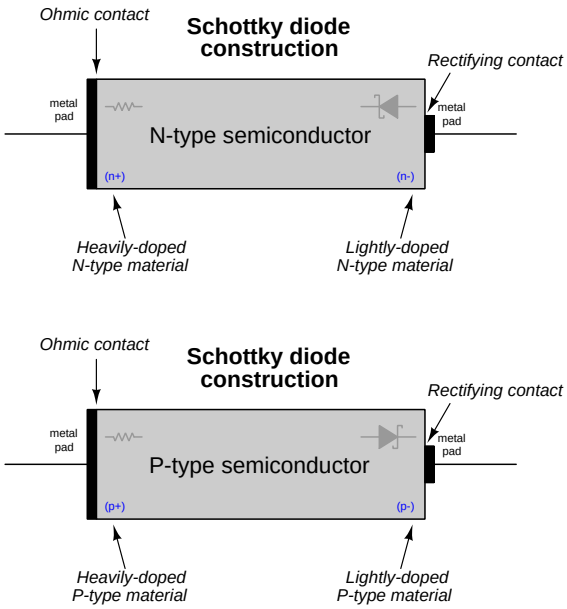
The emitter terminal is labeled by the numeral 5 while the collector is numeral 6. The transistor's base connection is on the bottom side of the semiconducting block. Figure 1 shows the transistor being applied in what we now call a *common-base* amplifier configuration with voltage source 7 used to properly bias the base-emitter junction for class-A operation and voltage source 8 providing power for the amplified output signal. Two transformers, 10 and 9, couple the AC input signal to the amplifier and the AC output signal from the amplifier, respectively. Based on the polarity of the bias source (+ to emitter, − to base) we can tell this transistor functions as a PNP type[1].

Amplification performance of three different prototype transistors is documented on page 5 (column 10) of the patent, showing impressive voltage and power gains for a prototype device:

Sample No	1	2	3
Input Res. (ohms)	640	500	1,000
Output Res. (ohms)	3×10^4	3×10^4	3×10^4
Input Voltage, A. C. R. M. S.	0.29	0.30	0.10
Output Voltage, A. C. R. M. S.	18	15	3.6
Voltage Gain	62	50	36
Power in (watts)	1.3×10^{-4}	1.8×10^{-4}	1.15×10^{-5}
Power out (watts)	100×10^{-4}	75×10^{-4}	42.5×10^{-5}
Power Gain	80	42	36
Input Bias D. C. (volts)	+0.2	+0.25	+0.2
Output Bias D. C. (volts)	−40	−20	−10

[1]This prototype transistor did not actually use a "sandwich" structure of P-N-P doped materials as is common with modern transistors. Instead, this transistor used thin metal wires contacting the germanium body, the metal-semiconductor junctions thus formed exhibiting diode ("rectifying") behavior. To their credit, the authors mention the use of "two semiconductors of opposite conductivity types" as a viable alternative to metal-semiconductor point-contact junctions.

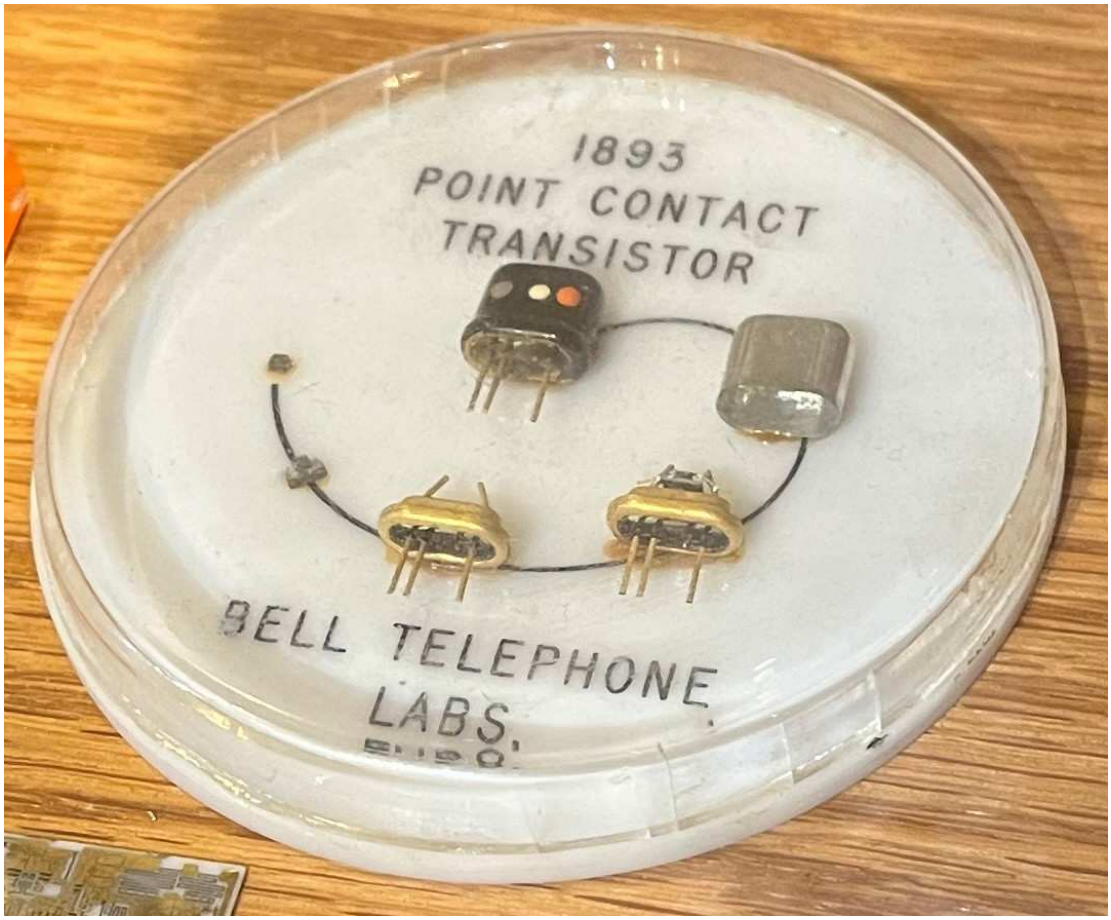
Incidentally, what the authors describe in their patent as "rectifying contact" versus "low resistance contact" connection points between the N-type semiconducting material and the metal terminals is precisely the distinction we see within Schottky diodes where the rectifying action takes place between a metal contact and the semiconducting block, but where a second metal contact must make a non-rectifying (i.e. non-diode) connection to that same block:



What distinguishes a rectifying metal-to-semiconductor junction versus an "ohmic" or "low resistance" or "non-rectifying" junction is the concentration of dopant impurities within the semiconducting material near the point of contact with the metal. As you can see, these different behaviors of metal-to-semiconductor junctions occur with both N-type and P-type semiconductor materials, the non-rectifying or "low resistance" junctions occurring where the dopant concentration is heavy and the "rectifying" junctions occurring where the dopant concentration is light.

5.2 Early Bell Labs transistors

The bipolar junction transistor (BJT) was invented at Bell Laboratories, and this same corporation was an early manufacturer of the same. One of their 1950's era transistor is seen in the following photograph preserved within a block of clear plastic[2]:



This display unit shows six different stages of the same transistor arranged along a spiral. At the far counter-clockwise end of this spiral is the fully assembled transistor as it would be found within a complete electronic circuit. At the far-clockwise end of the spiral we see the actual semiconductor substrate forming the transistor. Much of the assembled transistor's volume is empty space and metal framing.

[2] I cannot help but smile at the analogy to a prehistoric insect encased in fossilized amber!

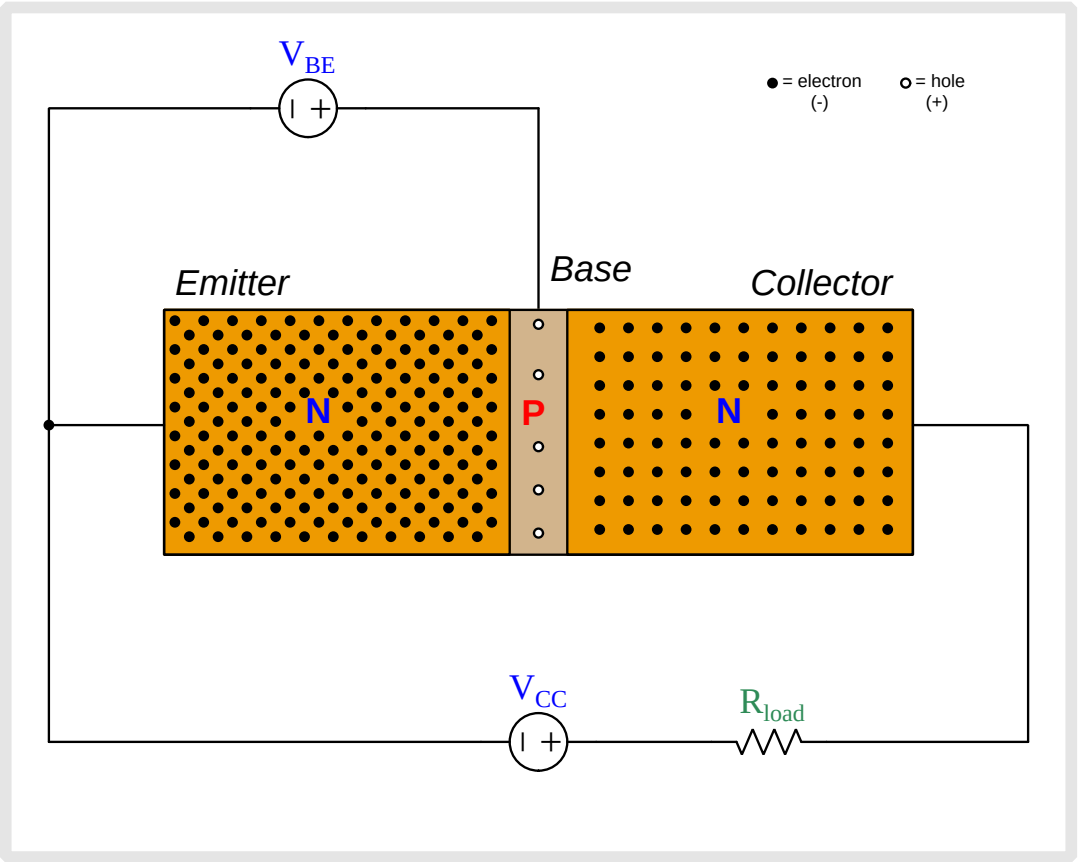
Chapter 6

Animations

Some concepts are much easier to grasp when seen in *action*. A simple yet effective form of animation suitable to an electronic document such as this is a "flip-book" animation where a set of pages in the document show successive frames of a simple animation. Such "flip-book" animations are designed to be viewed by paging forward (and/or back) with the document-reading software application, watching it frame-by-frame. Unlike video which may be difficult to pause at certain moments, "flip-book" animations lend themselves very well to individual frame viewing.

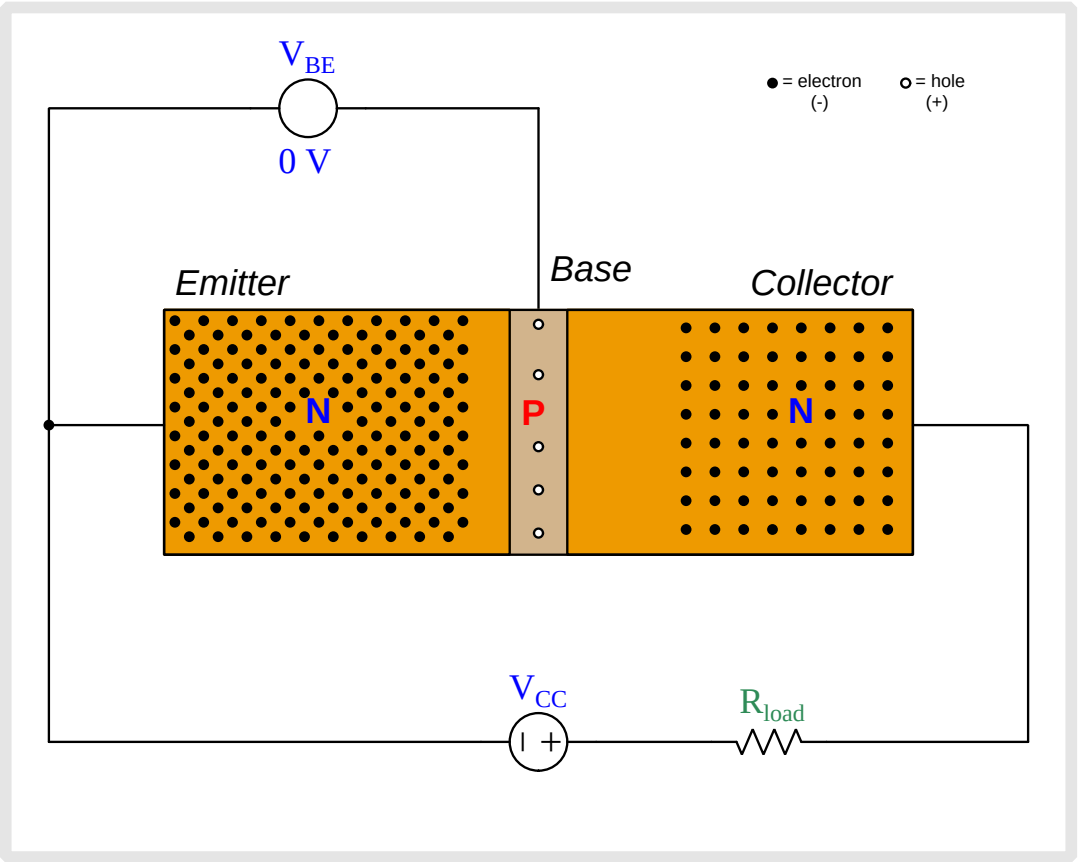
6.1 Animation of an NPN BJT turning on

Bipolar junction transistors are naturally "off" devices, having no continuous channel of conductive semiconductor material joining emitter and collector terminals. Instead the NPN or PNP sandwich interrupts that path with two series-opposed PN junctions. However, when the base-emitter PN junction is forward-biased by an external source, majority charge carriers from the emitter region are drawn into the thin and lightly-doped base region where some of them travel out the base but most carry past into the depletion region of the base-collector PN junction to be swept into the collector region. For an NPN transistor this means electrons from the N-type emitter region will be injected into the P-type base region (as "minority" type carriers there) before most of them get swept into the N-type collector region.

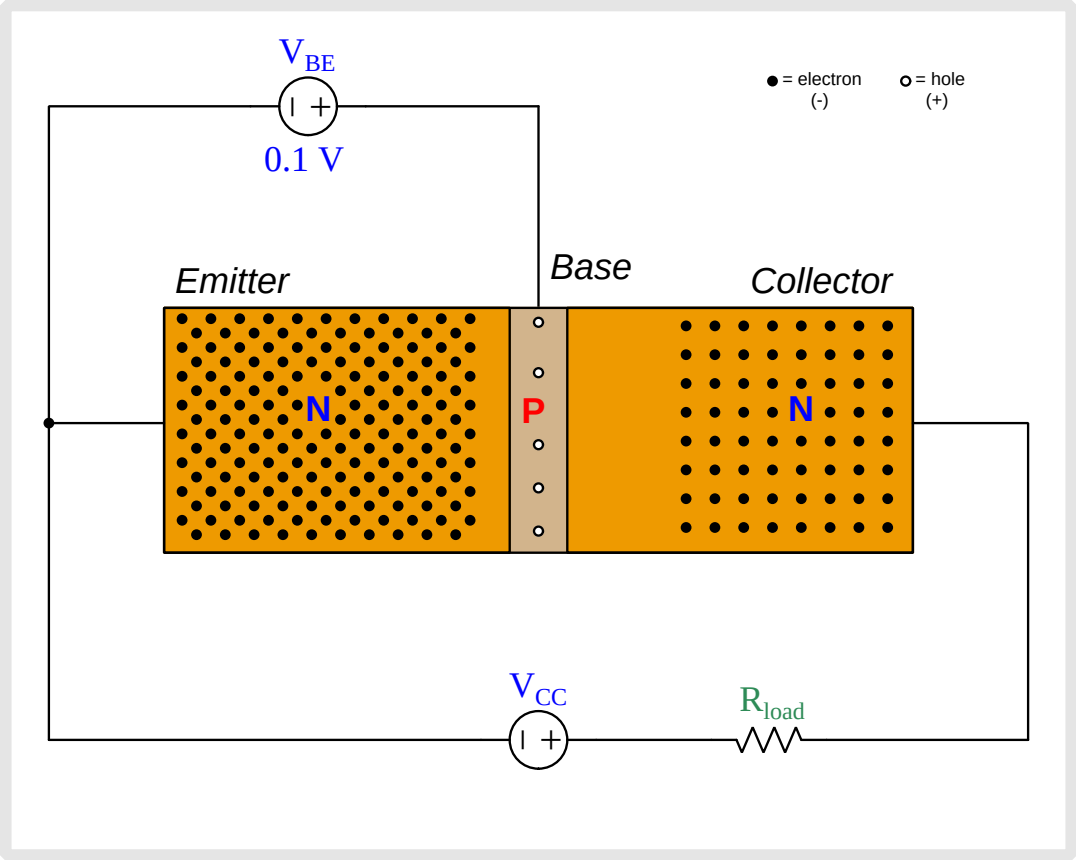


Above we see an NPN bipolar junction transistor connected to two sources. V_{BE} applies a controlling voltage between the base and emitter. V_{CC} is the voltage source which will drive current through the load resistor once the transistor begins to conduct.

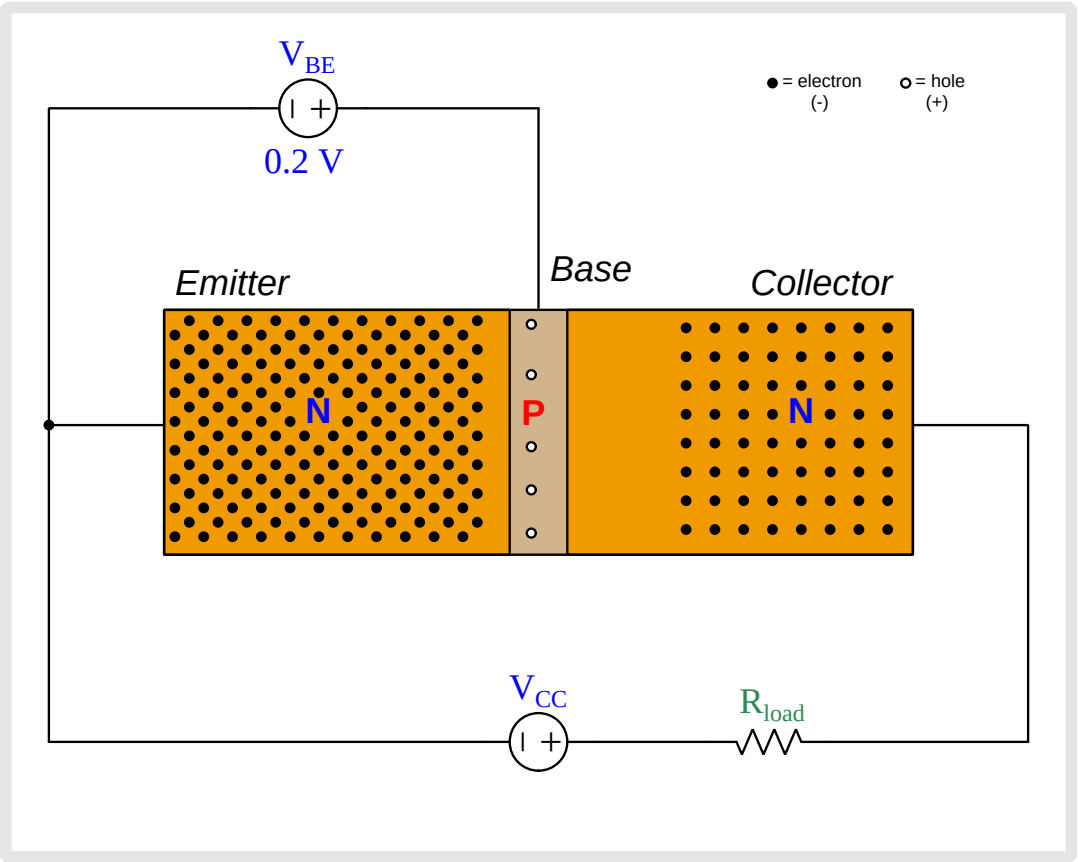
The N-doped emitter region is heavily doped and therefore contains many free electrons (negatively-charged) as mobile charge carriers. The collector region is also doped as an N-type semiconductor but the dopant concentration is less than that of the emitter. The P-doped base region is the least-doped of them all with just a few holes (positively-charged) as mobile charge carriers. Note that when load current begins to flow, its direction will be shown in *conventional flow notation* which is the same direction in which holes move but is opposite to the direction free electrons move.



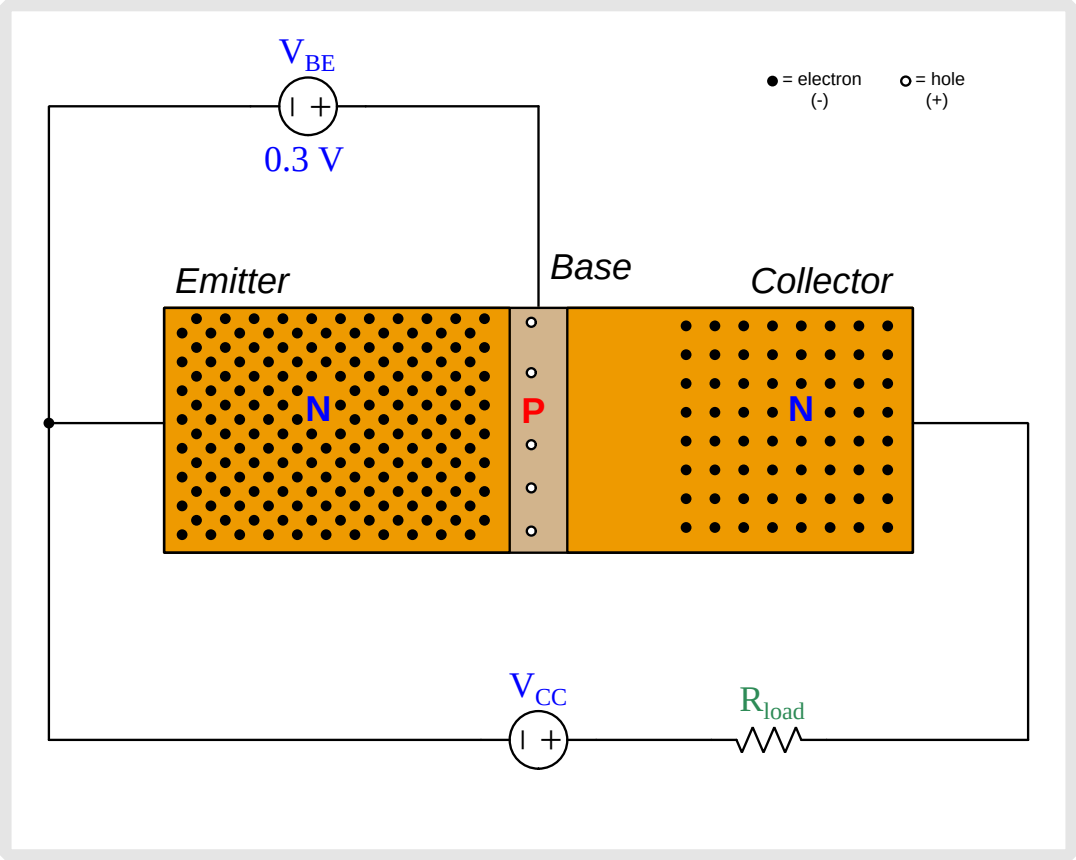
Note the depletion regions shown as areas with no free charge carriers, the base-collector depletion region being wider than the base-emitter depletion region due to the applied reverse-bias potential of V_{CC}.



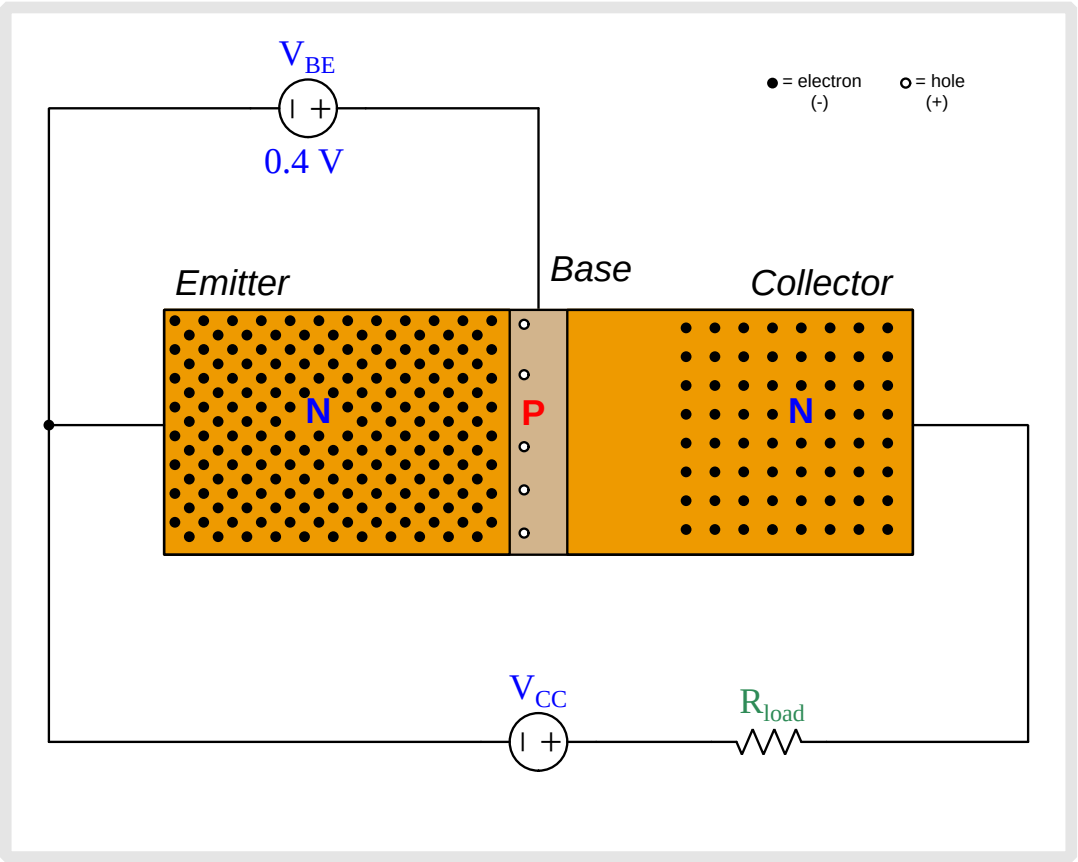
As we begin to apply a base-to-emitter voltage with the base positive and emitter negative, that applied potential difference drives the emitter and base regions' respective charge carriers toward each other, collapsing the base-emitter depletion region.



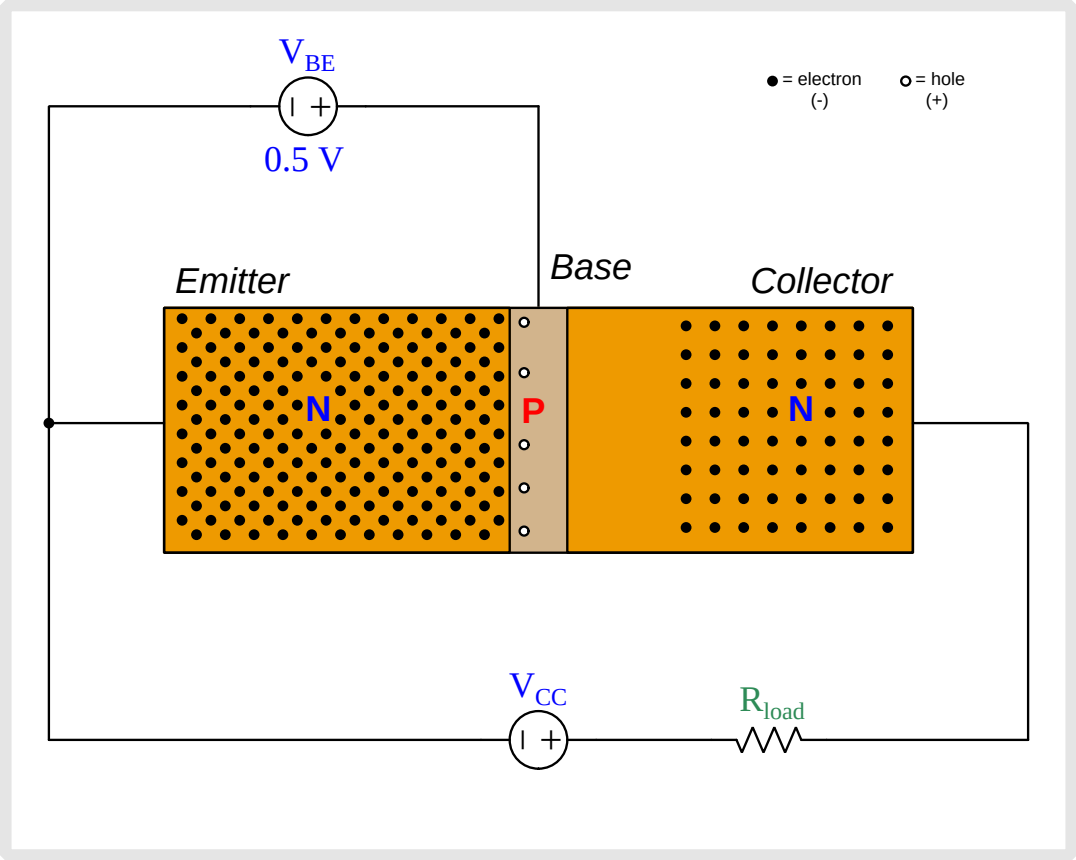
As we begin to apply a base-to-emitter voltage with the base positive and emitter negative, that applied potential difference drives the emitter and base regions' respective charge carriers toward each other, collapsing the base-emitter depletion region.



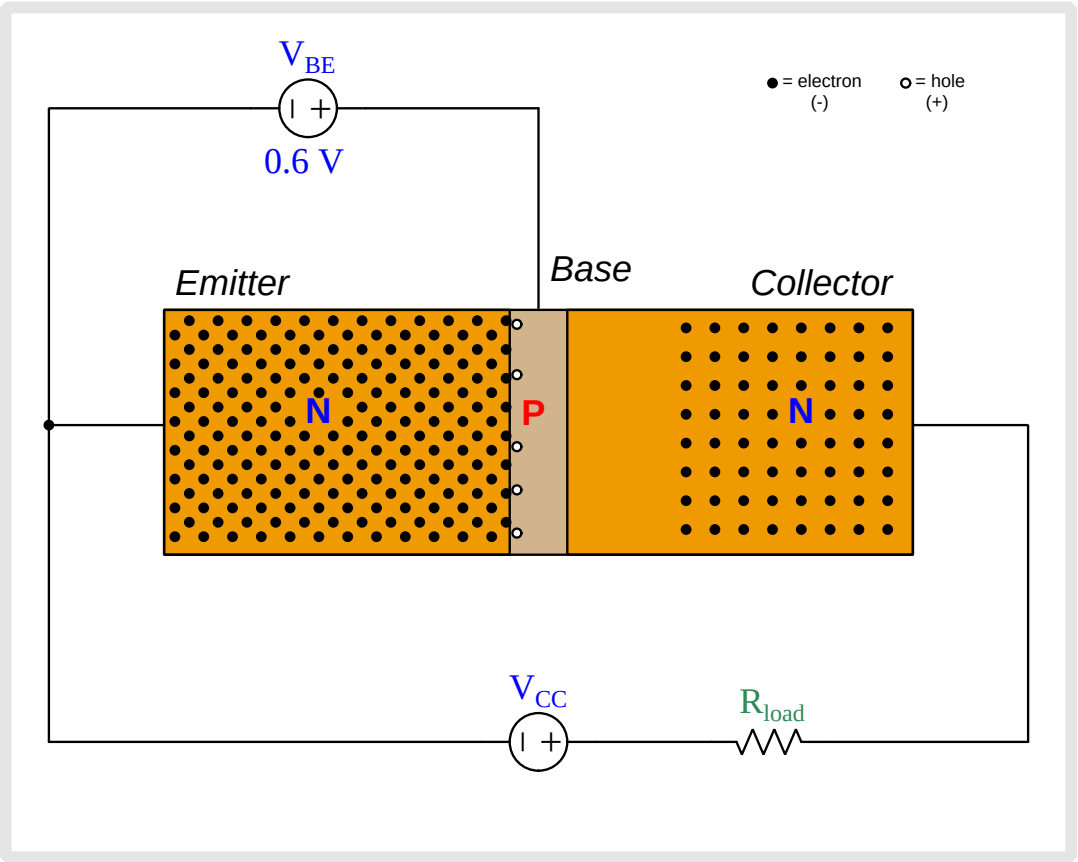
As we begin to apply a base-to-emitter voltage with the base positive and emitter negative, that applied potential difference drives the emitter and base regions' respective charge carriers toward each other, collapsing the base-emitter depletion region.



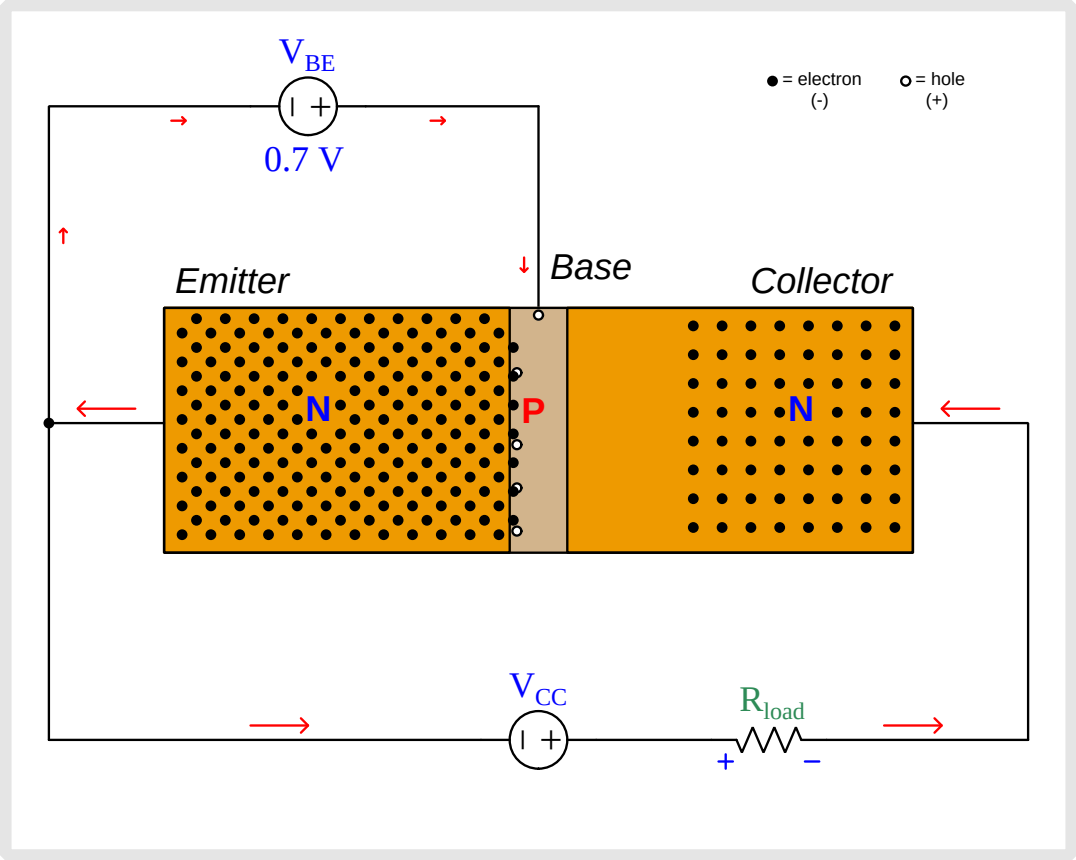
As we begin to apply a base-to-emitter voltage with the base positive and emitter negative, that applied potential difference drives the emitter and base regions' respective charge carriers toward each other, collapsing the base-emitter depletion region.



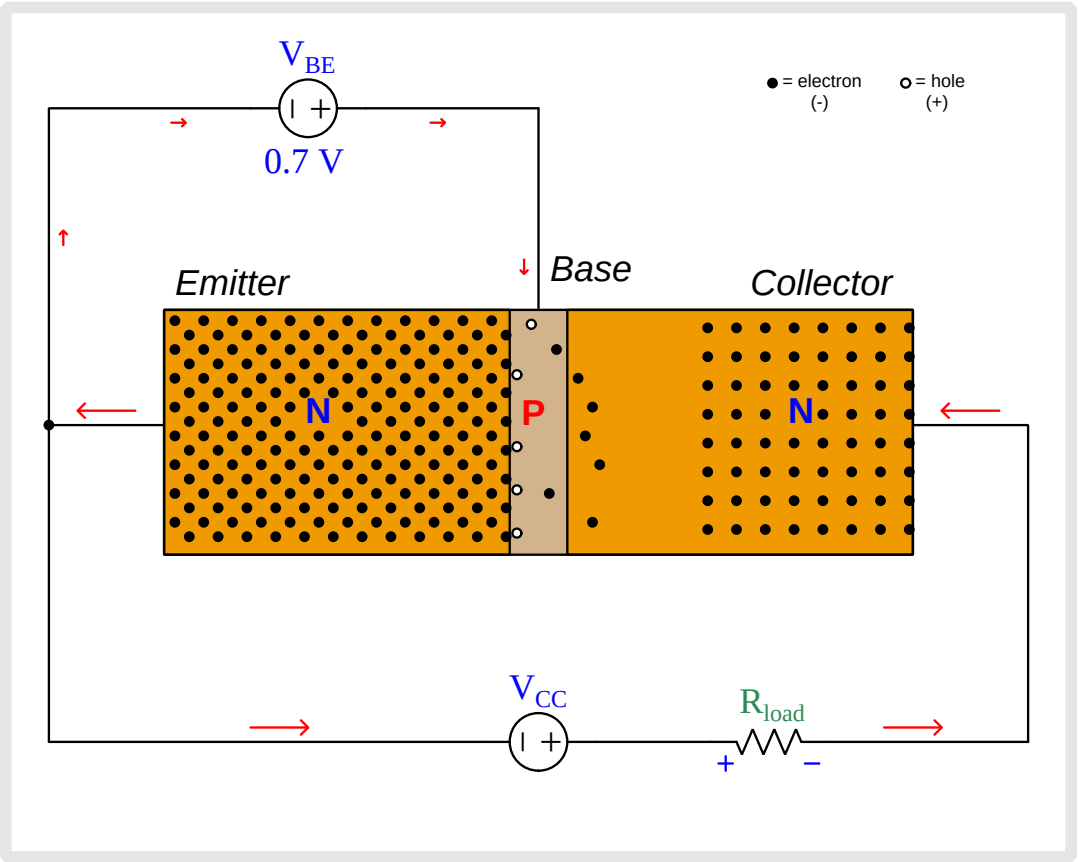
As we begin to apply a base-to-emitter voltage with the base positive and emitter negative, that applied potential difference drives the emitter and base regions' respective charge carriers toward each other, collapsing the base-emitter depletion region.



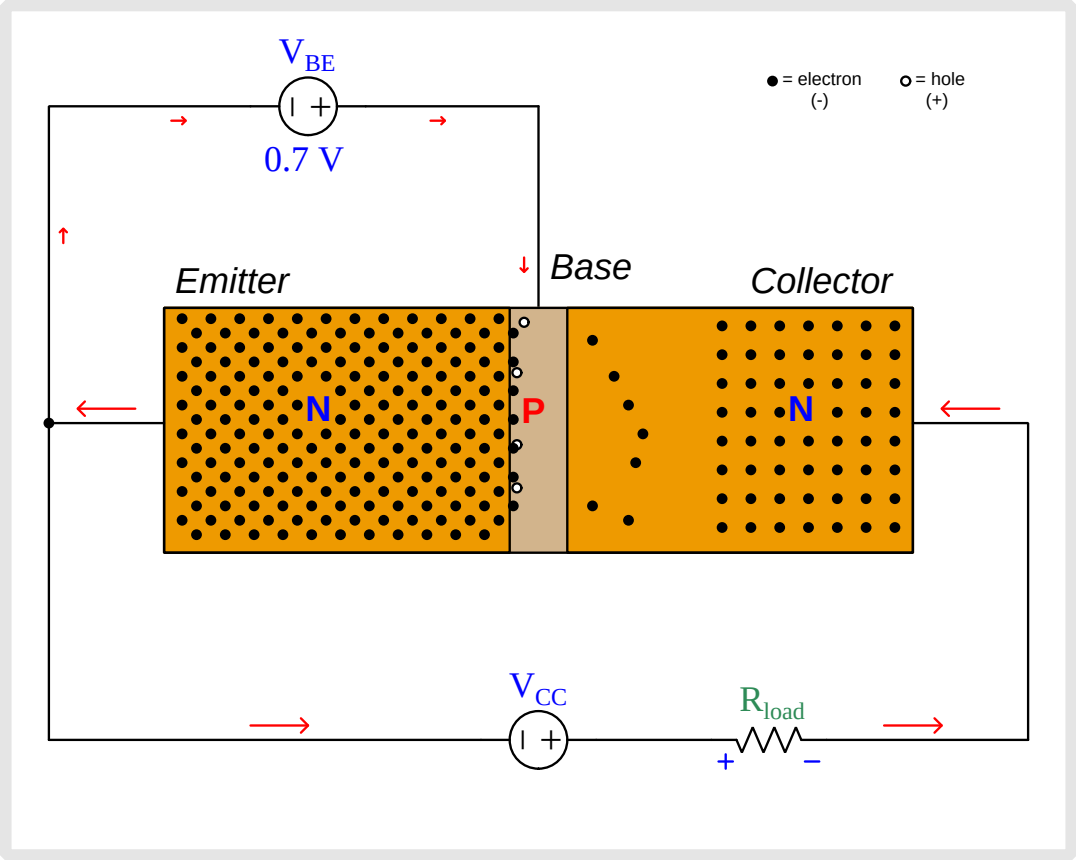
As we begin to apply a base-to-emitter voltage with the base positive and emitter negative, that applied potential difference drives the emitter and base regions' respective charge carriers toward each other, collapsing the base-emitter depletion region.



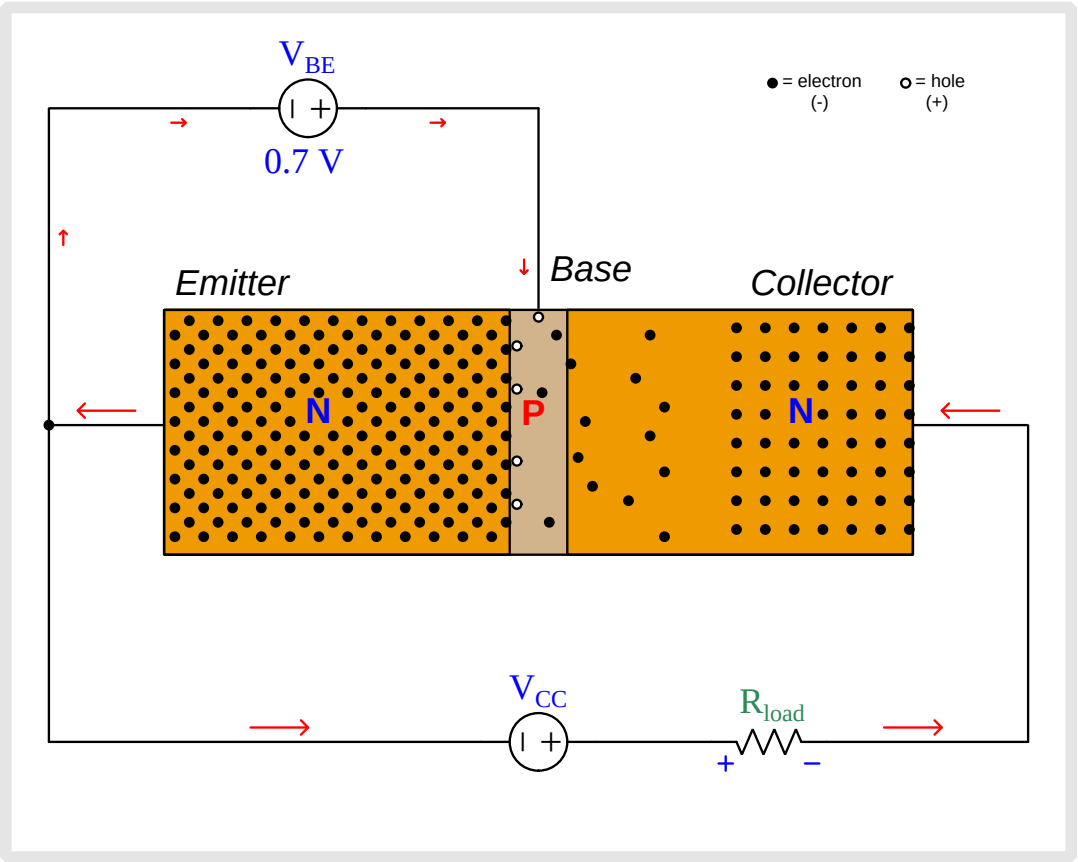
Now there is enough base-to-emitter voltage to fully collapse the base-emitter depletion region which allows base current to flow through the V_{BE} source. New holes enter into the base region from the base terminal to replenish any holes that have "re-combined" with electrons from the emitter region. Most of the emitter-injected electrons do *not* combine with holes, though, because the lightly-doped base region provides few holes to combine with. Therefore, most of the electrons injected into the base region from the emitter region begin to drift toward the collector.



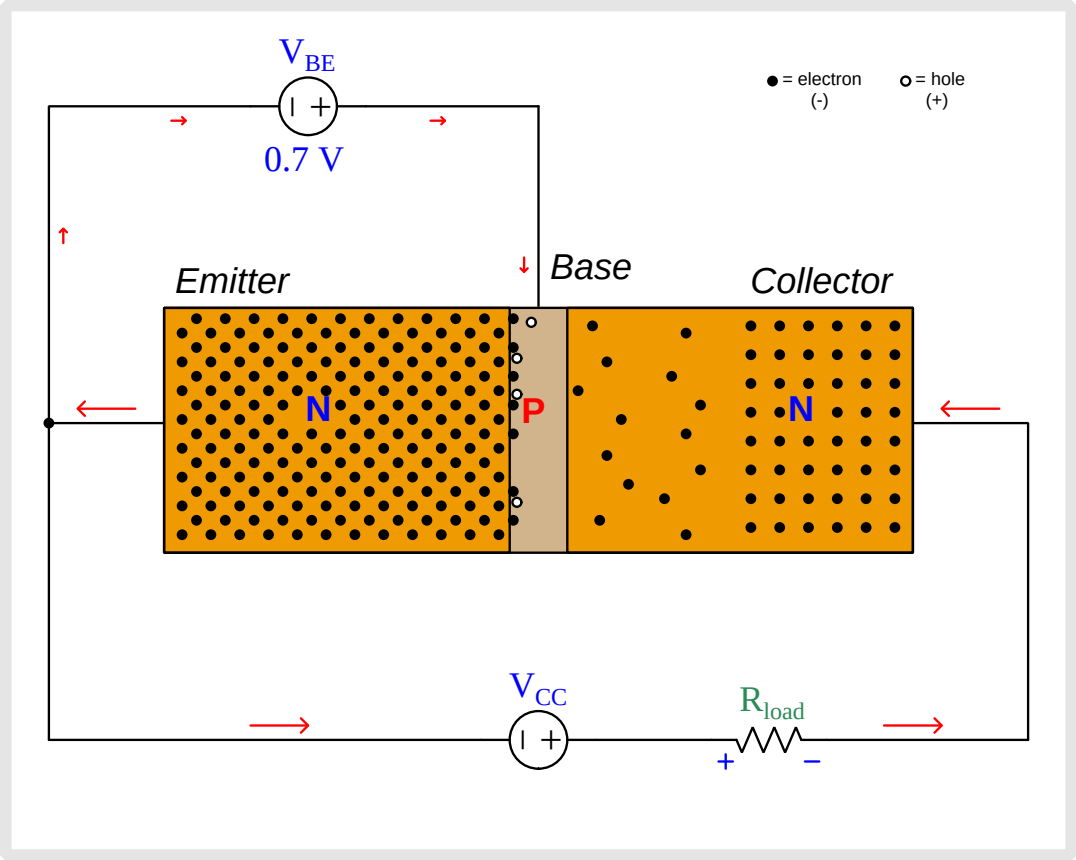
The electric field imposed by V_{CC} across the base-collector PN junction quickly "sweeps" electrons toward the collector terminal, constituting current through the load resistor. A small number of electrons injected into the base "recombined" with holes there to constitute base current rather than collector current.



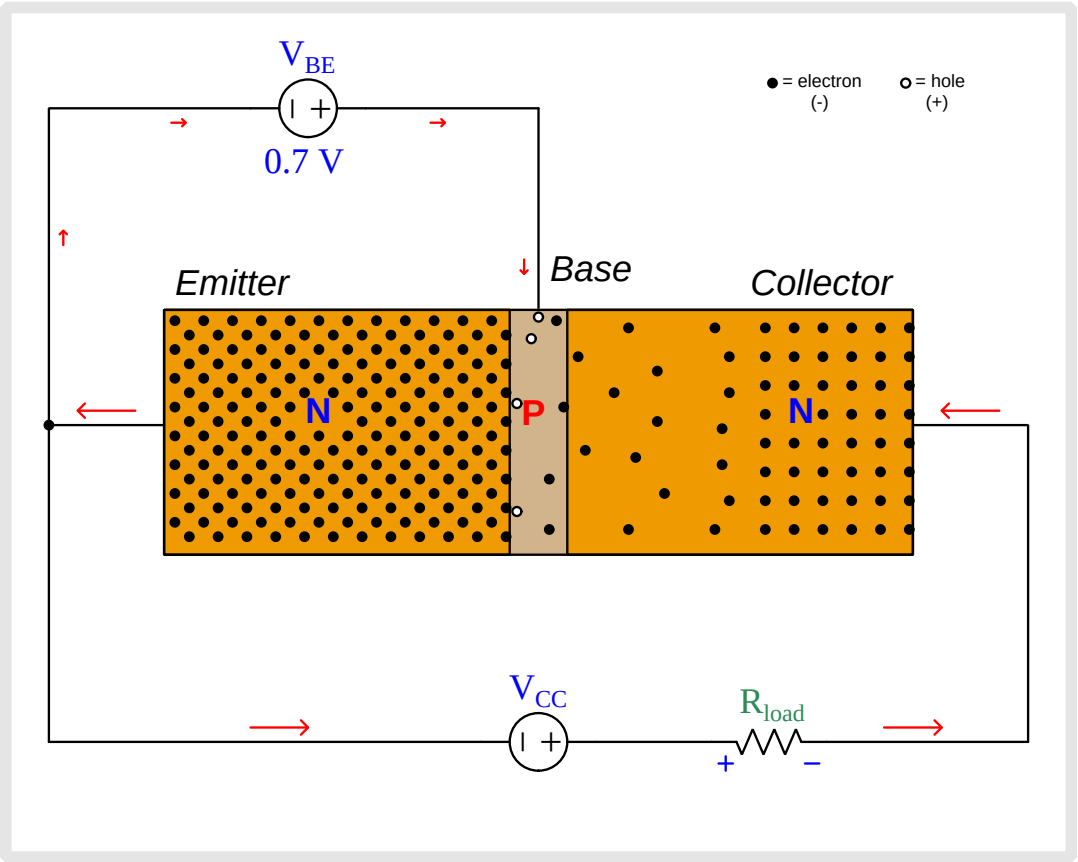
The electric field imposed by V_{CC} across the base-collector PN junction quickly "sweeps" electrons toward the collector terminal, constituting current through the load resistor. A small number of electrons injected into the base "recombined" with holes there to constitute base current rather than collector current.



The electric field imposed by V_{CC} across the base-collector PN junction quickly "sweeps" electrons toward the collector terminal, constituting current through the load resistor. A small number of electrons injected into the base "recombined" with holes there to constitute base current rather than collector current.



The electric field imposed by V_{CC} across the base-collector PN junction quickly "sweeps" electrons toward the collector terminal, constituting current through the load resistor. A small number of electrons injected into the base "recombined" with holes there to constitute base current rather than collector current.

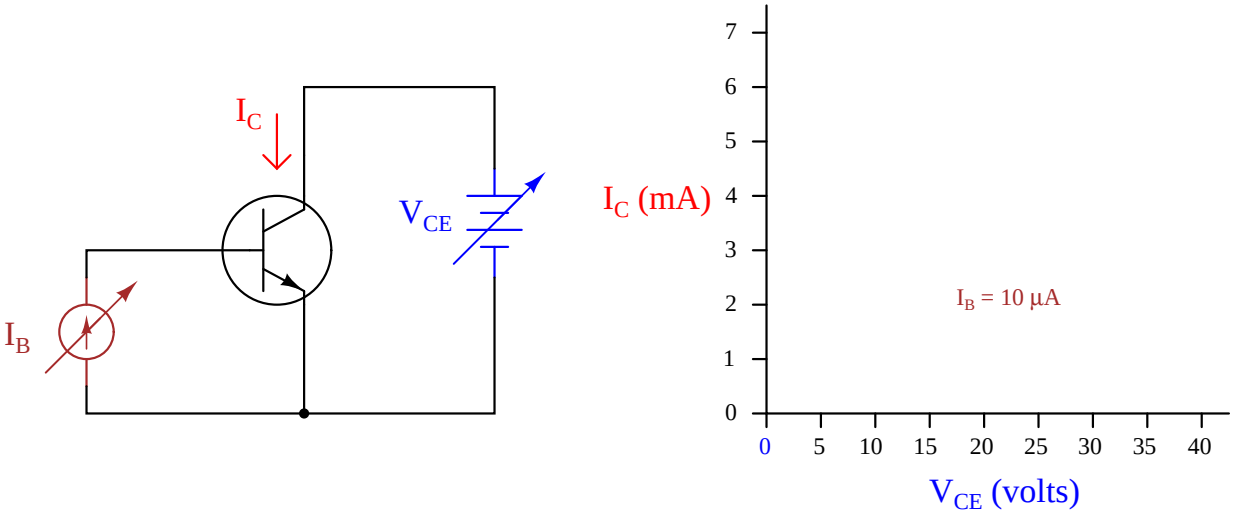


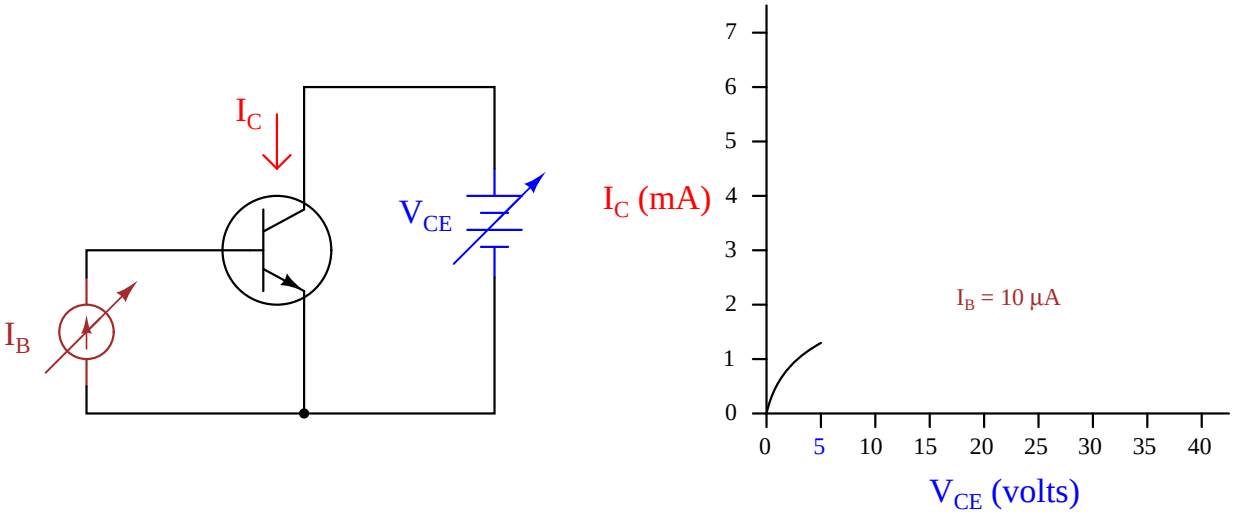
The electric field imposed by V_{CC} across the base-collector PN junction quickly "sweeps" electrons toward the collector terminal, constituting current through the load resistor. A small number of electrons injected into the base "recombined" with holes there to constitute base current rather than collector current.

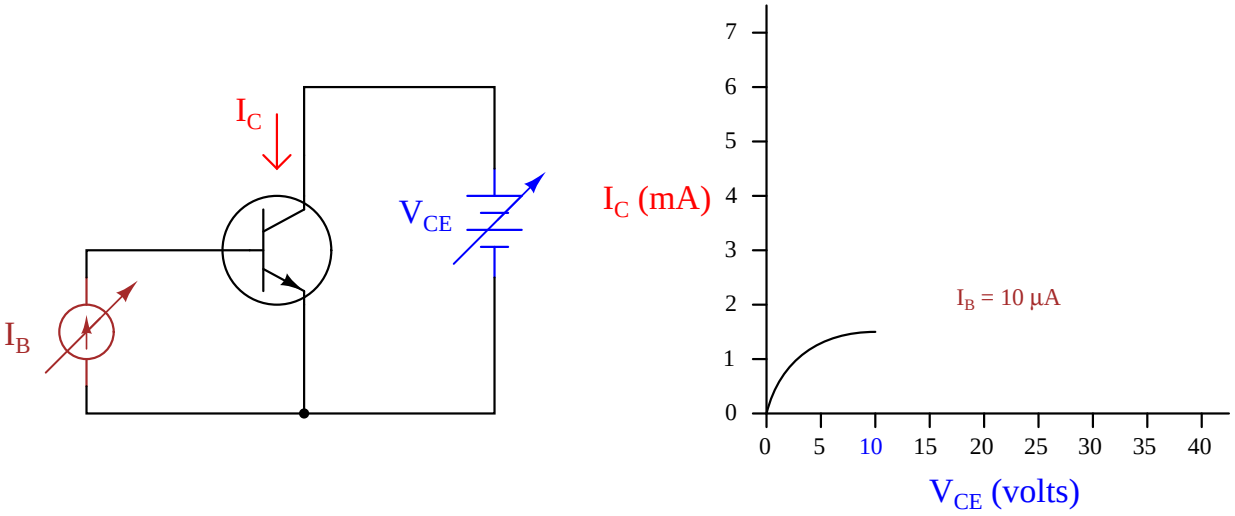
6.2 Animation of BJT curve-sketching

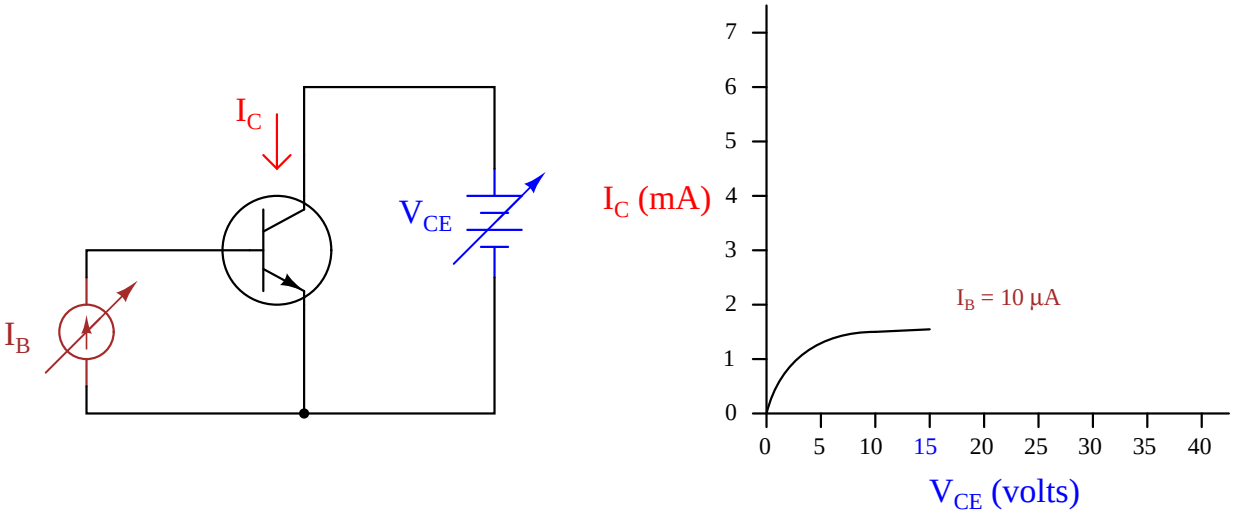
The following animation shows how a "family" of curves may be generated to characterize the behavior of a bipolar junction transistor.

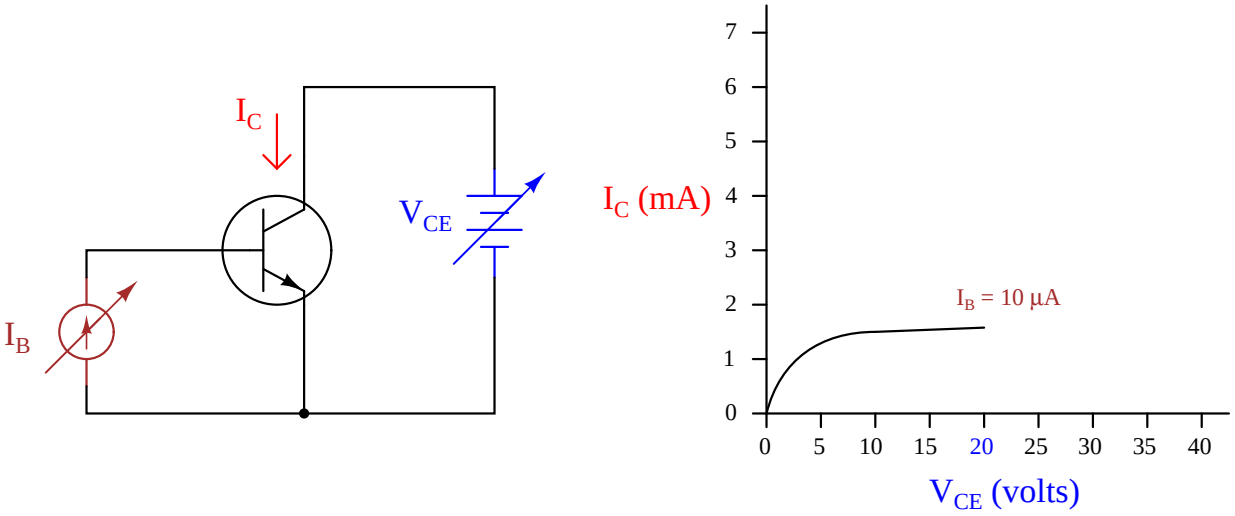
Pay close attention to which parameters are varied as each curve is sketched. The base current is stepped up by 10 μA increments, and then held constant while collector-emitted voltage (V_{CE}) is swept from 0 Volts to 40 Volts. The resulting collector current (I_C) is then plotted as a function of V_{CE}. Each new curve represents the transistor's response to collector-emitter voltage for a new (constant) value of base current.

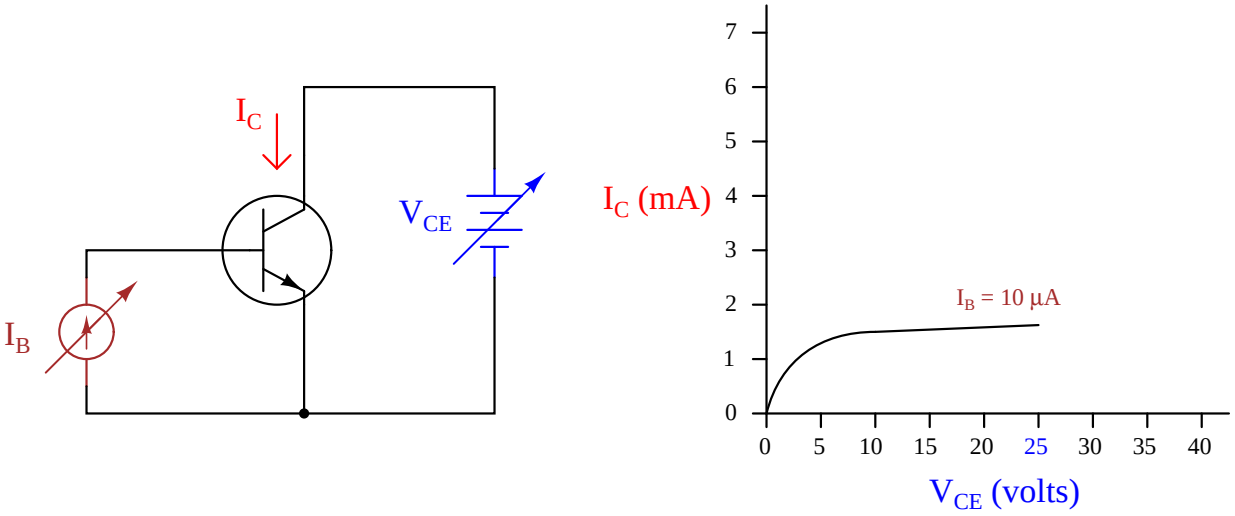

I_C
V_CE
I_B
I_C (mA)
7
6
5
4
3
2
1
0
I_B = 10 µA
0
5
10
15
20
25
30
35
40
V_CE (volts)

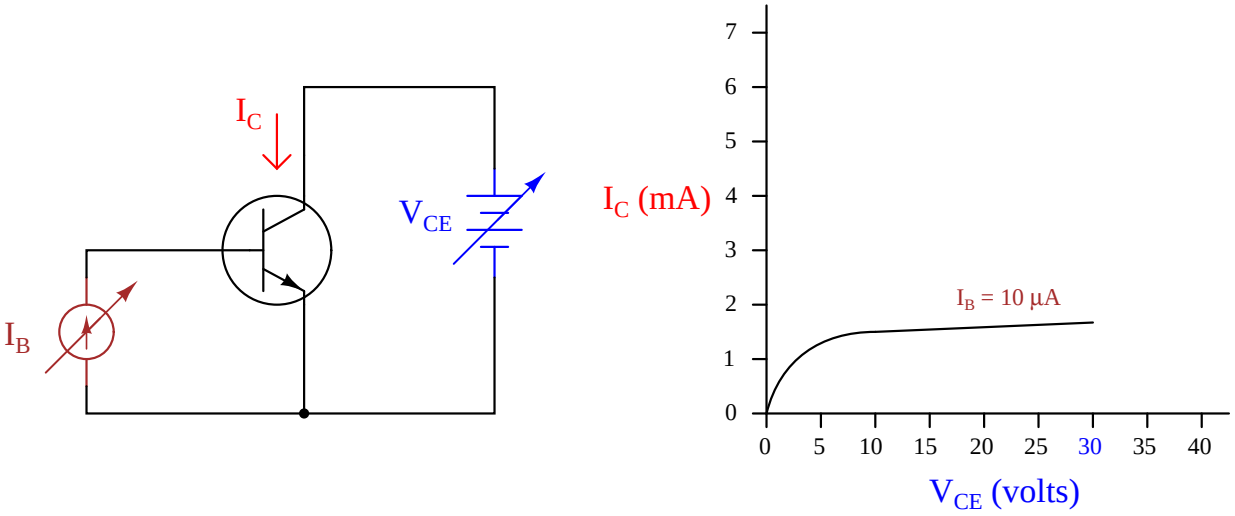

I_C
V_CE
I_B
I_C (mA)
7
6
5
4
3
2
1
0
I_B = 10 µA
0
5
10
15
20
25
30
35
40
V_CE (volts)

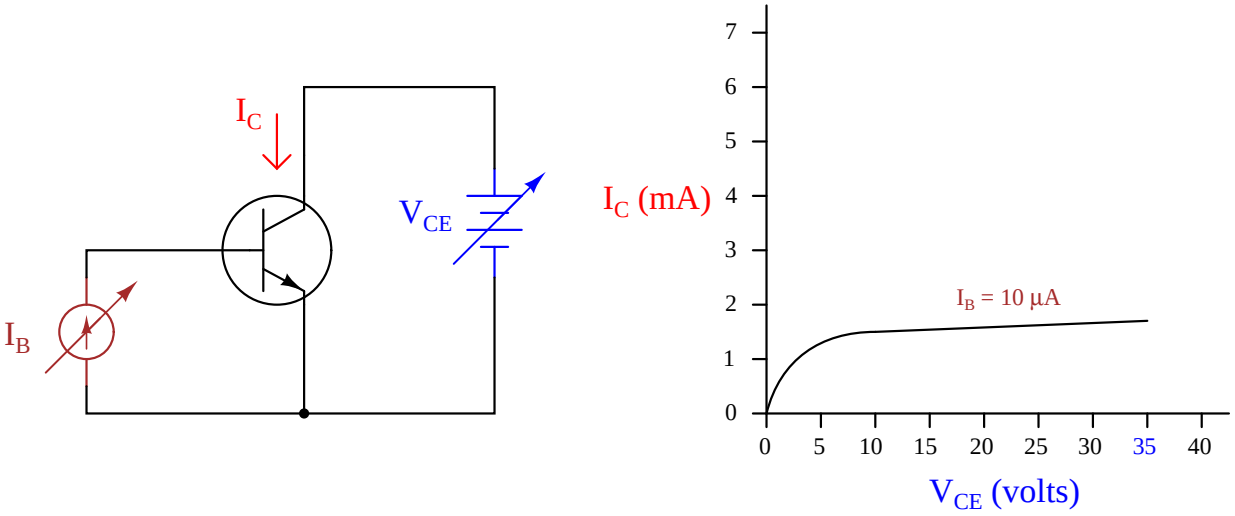

I_C
V_CE
I_B
I_C (mA)
7
6
5
4
3
2
1
0
I_B = 10 µA
0
5
10
15
20
25
30
35
40
V_CE (volts)

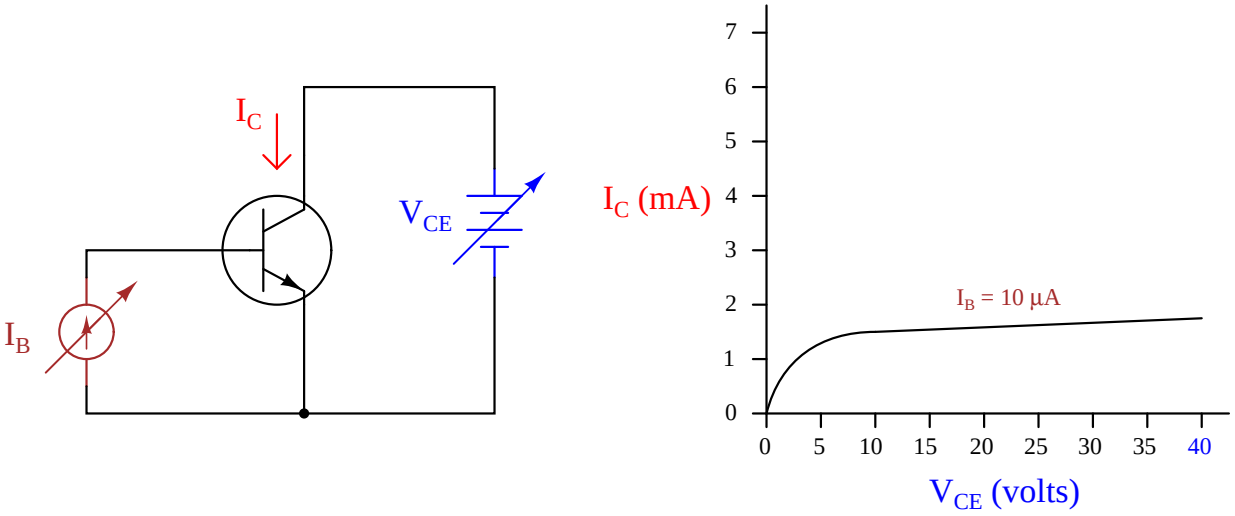

I_C
V_CE
I_B
I_C (mA)
7
6
5
4
3
2
1
0
I_B = 10 µA
0
5
10
15
20
25
30
35
40
V_CE (volts)

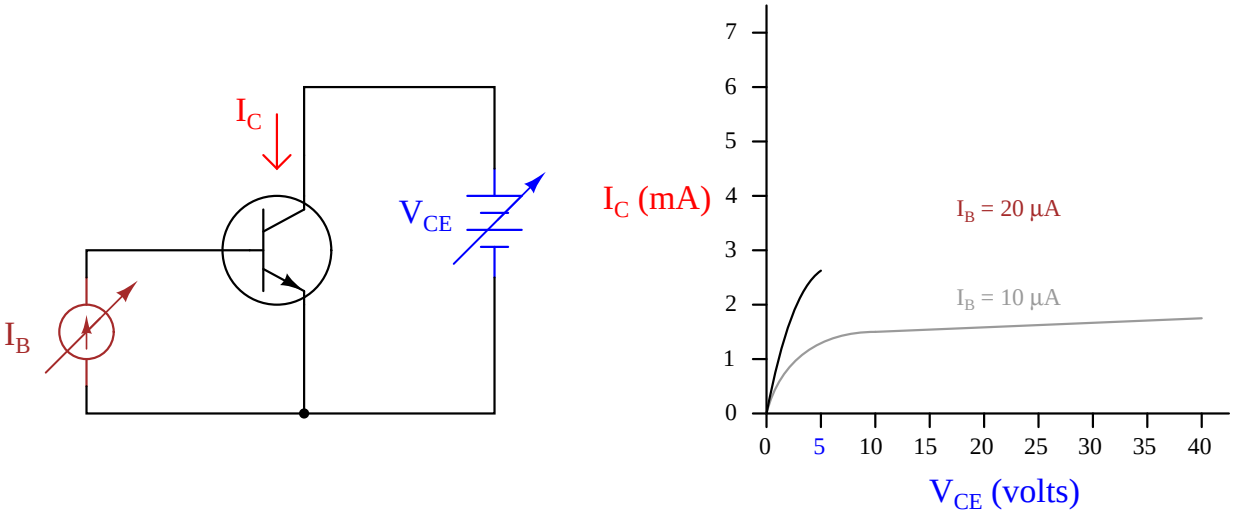

I_C
V_CE
I_B
I_C (mA)
7
6
5
4
3
2
1
0
I_B = 10 µA
0
5
10
15
20
25
30
35
40
V_CE (volts)

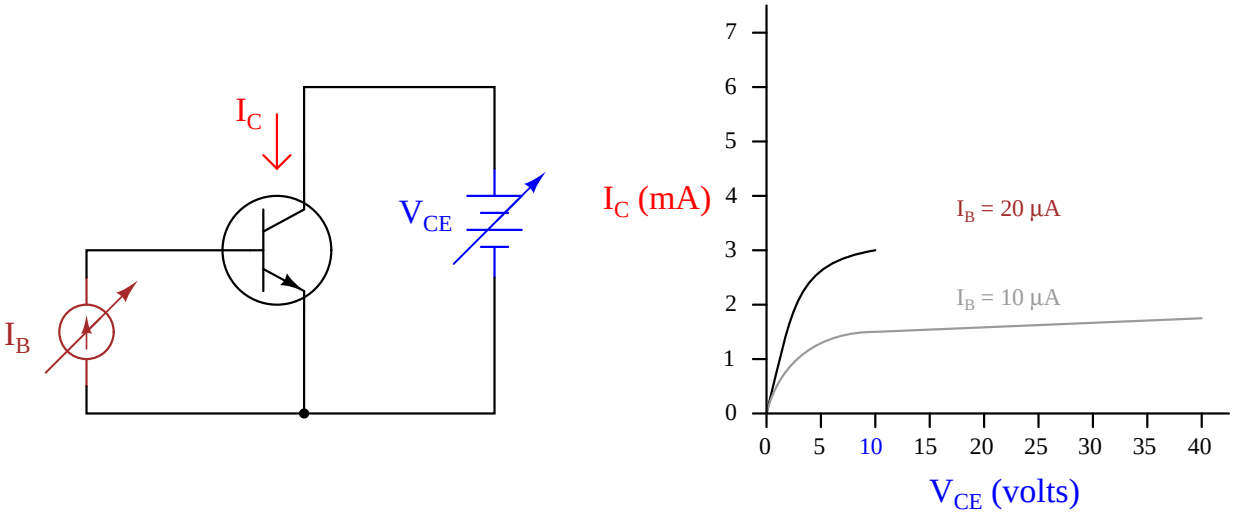

I_C
V_CE
I_B
I_C (mA)
7
6
5
4
3
2
1
0
I_B = 10 µA
0
5
10
15
20
25
30
35
40
V_CE (volts)

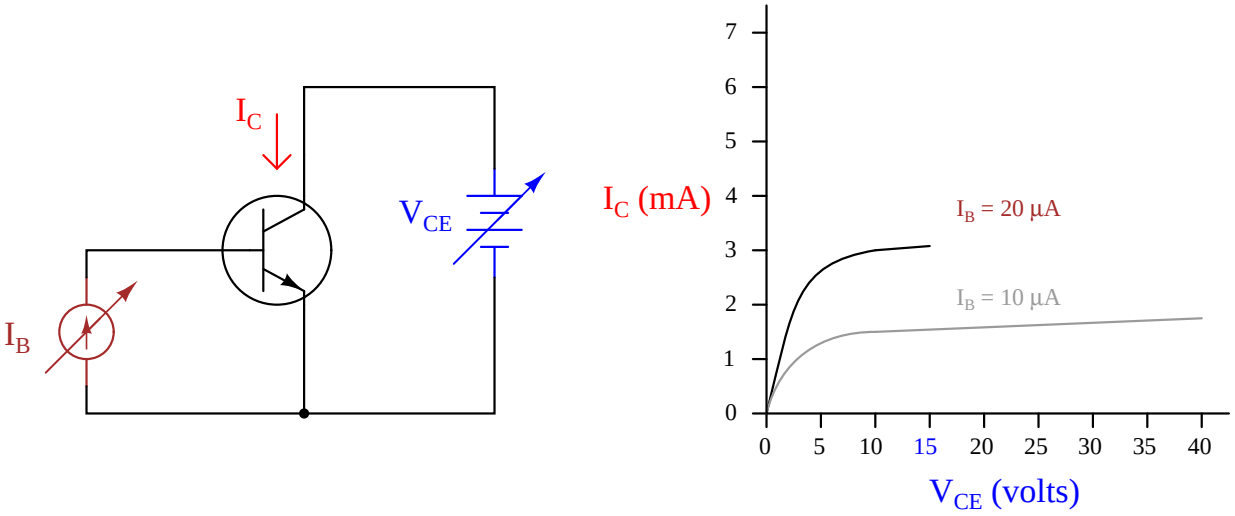

I_C
V_CE
I_B
I_C (mA)
7
6
5
4
3
2
1
0
I_B = 10 µA
0
5
10
15
20
25
30
35
40
V_CE (volts)

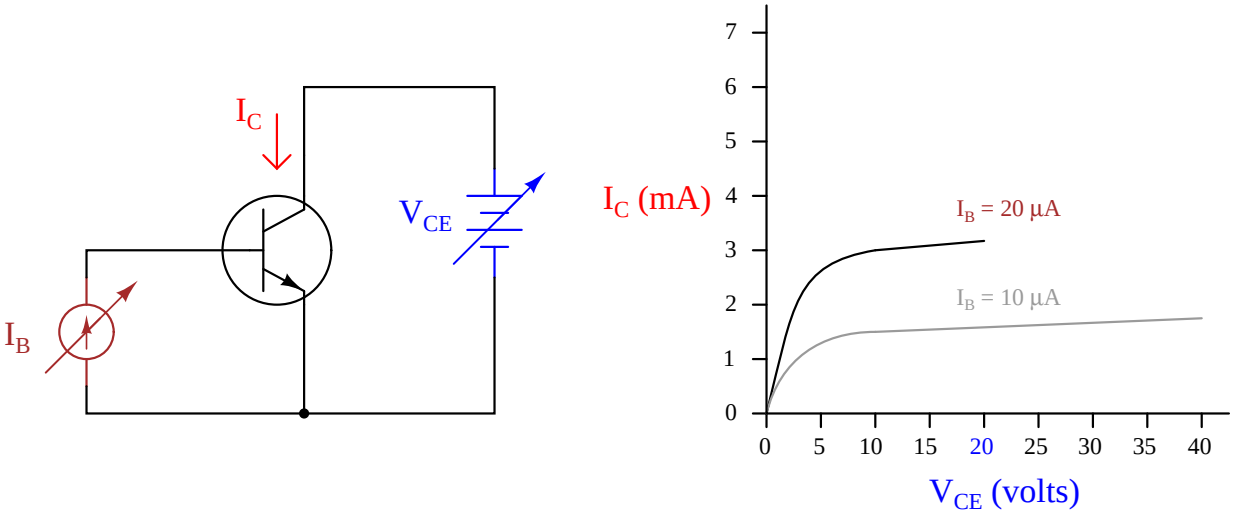

I_C
V_CE
I_B
I_C (mA)
7
6
5
4
3
2
1
0
I_B = 10 µA
0
5
10
15
20
25
30
35
40
V_CE (volts)

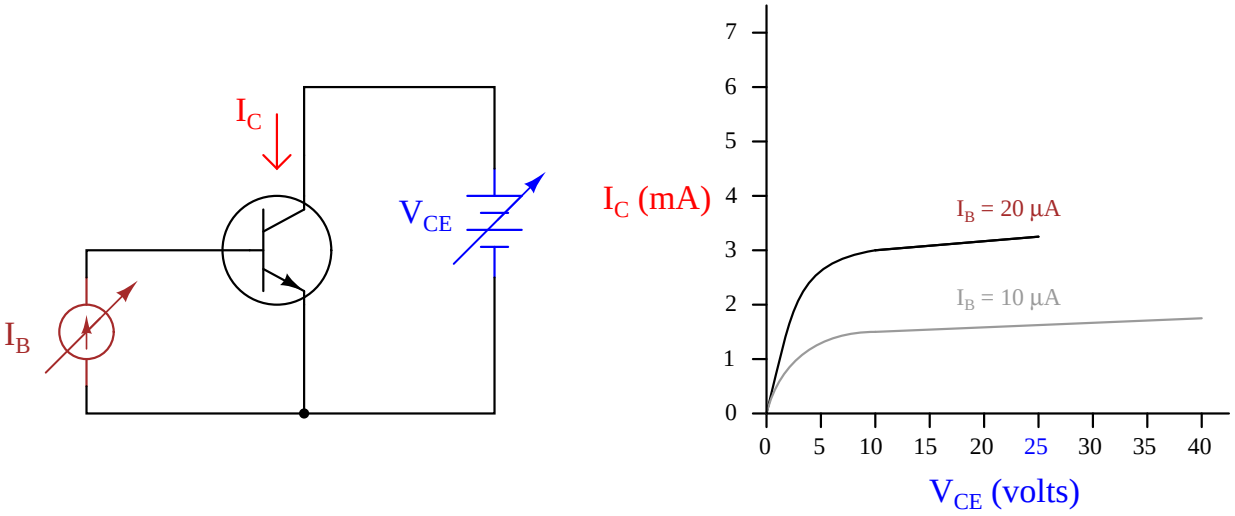

I_C
V_CE
I_B
I_C (mA)
7
6
5
4
3
2
1
0
I_B = 10 µA
0
5
10
15
20
25
30
35
40
V_CE (volts)

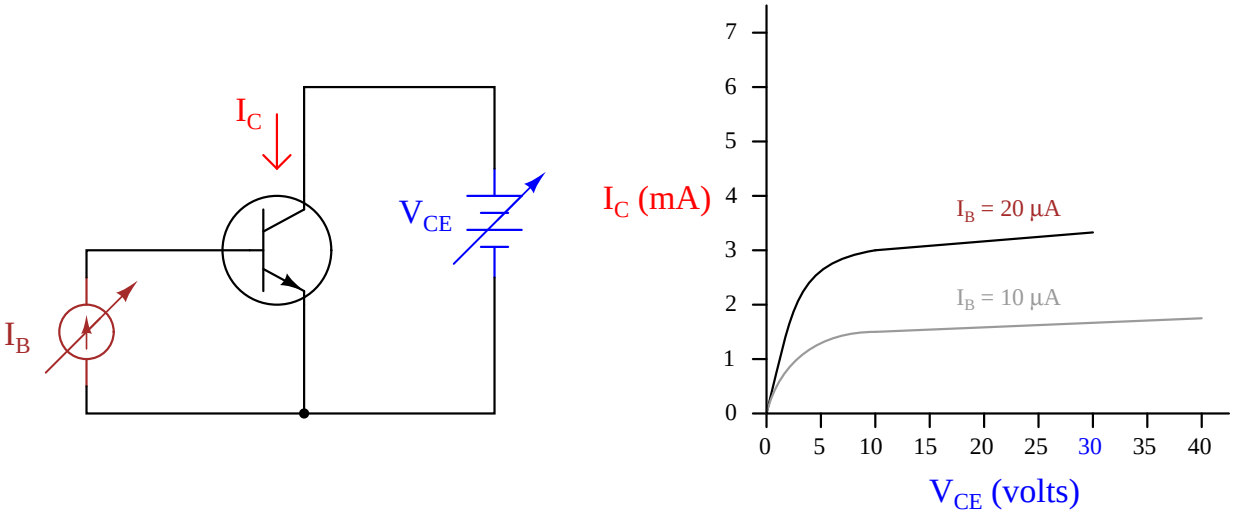

I_C
V_CE
I_B
I_C (mA)
7
6
5
4
3
2
1
0
I_B = 20 µA
I_B = 10 µA
0
5
10
15
20
25
30
35
40
V_CE (volts)

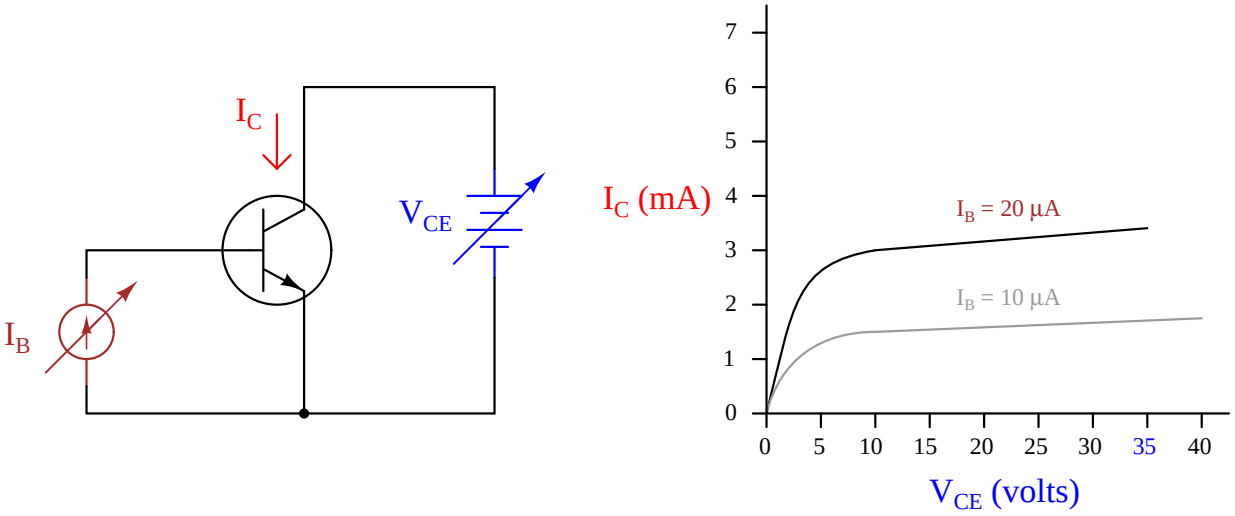

I_C
V_CE
I_B
I_C (mA)
7
6
5
4
3
2
1
0
I_B = 20 µA
I_B = 10 µA
0
5
10
15
20
25
30
35
40
V_CE (volts)

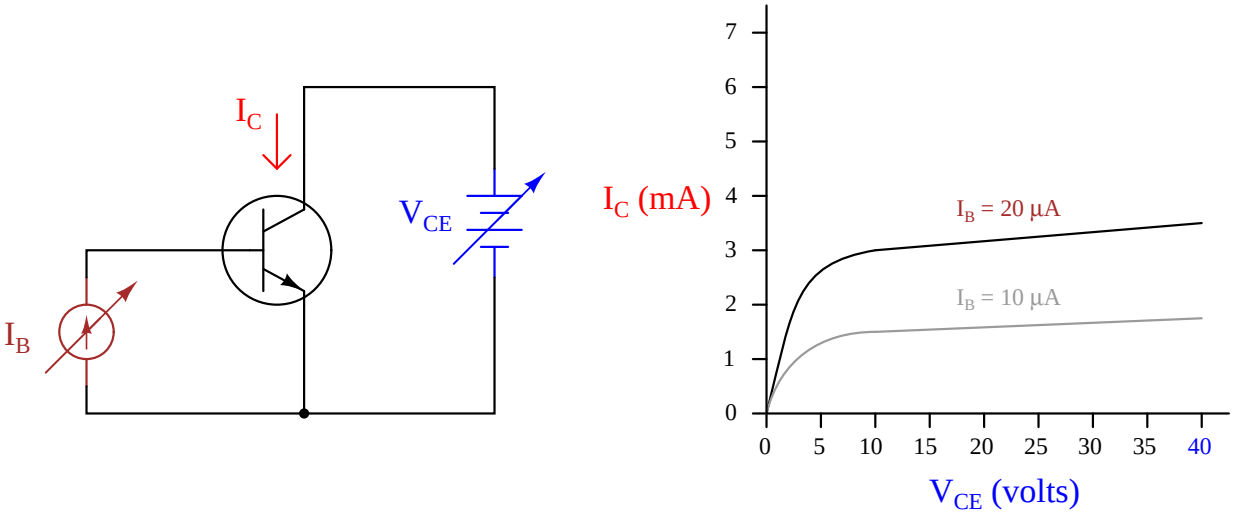

I_C
V_CE
I_B
I_C (mA)
7
6
5
4
3
2
1
0
I_B = 20 µA
I_B = 10 µA
0
5
10
15
20
25
30
35
40
V_CE (volts)

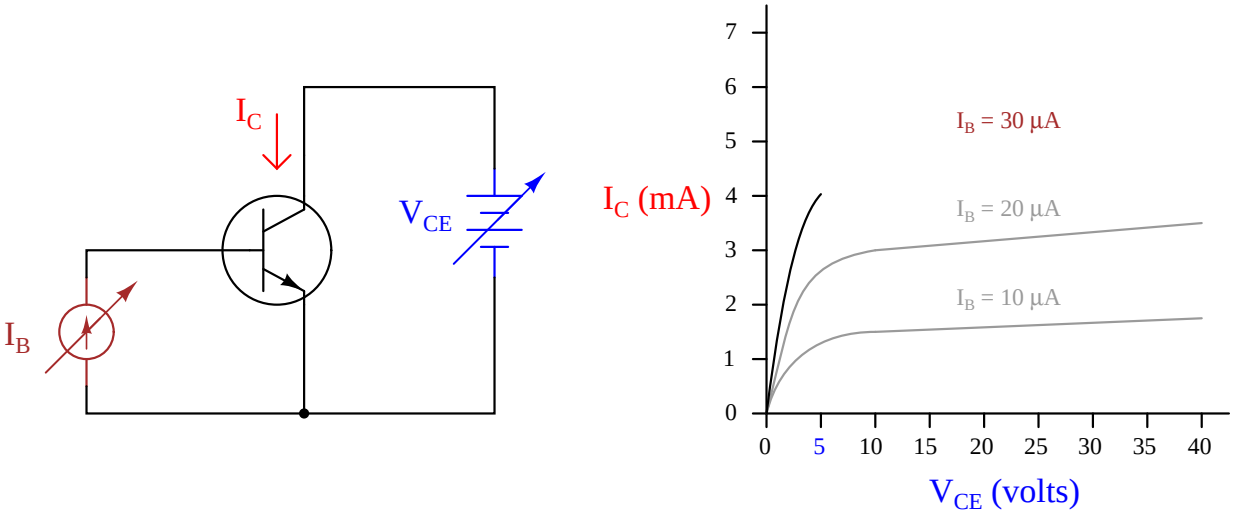

I_C
V_CE
I_B
I_C (mA)
7
6
5
4
3
2
1
0
I_B = 20 µA
I_B = 10 µA
0
5
10
15
20
25
30
35
40
V_CE (volts)

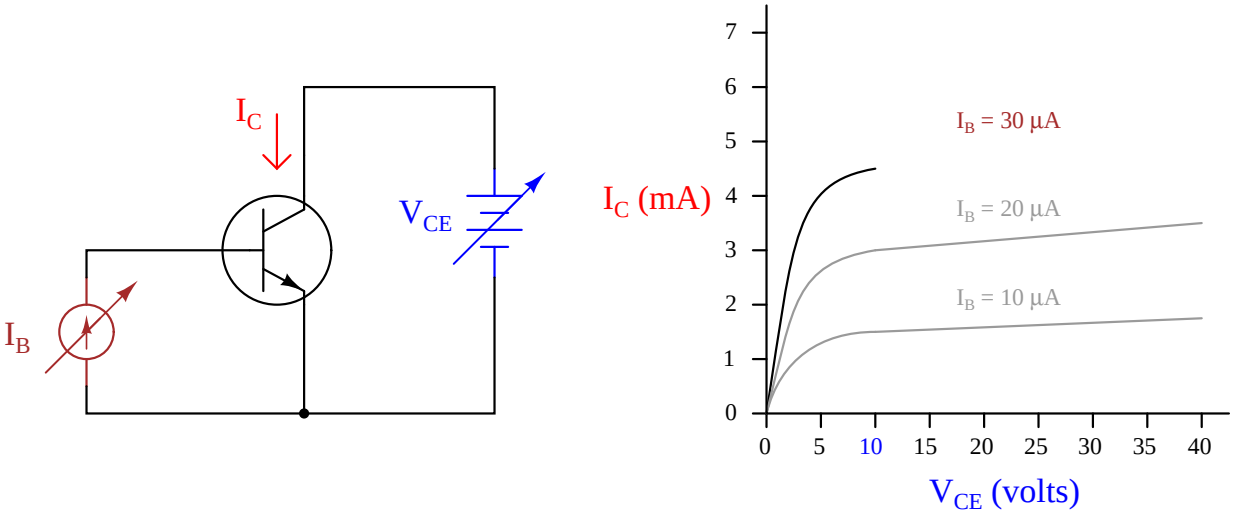

I_C
V_CE
I_B
I_C (mA)
7
6
5
4
3
2
1
0
I_B = 20 µA
I_B = 10 µA
0
5
10
15
20
25
30
35
40
V_CE (volts)

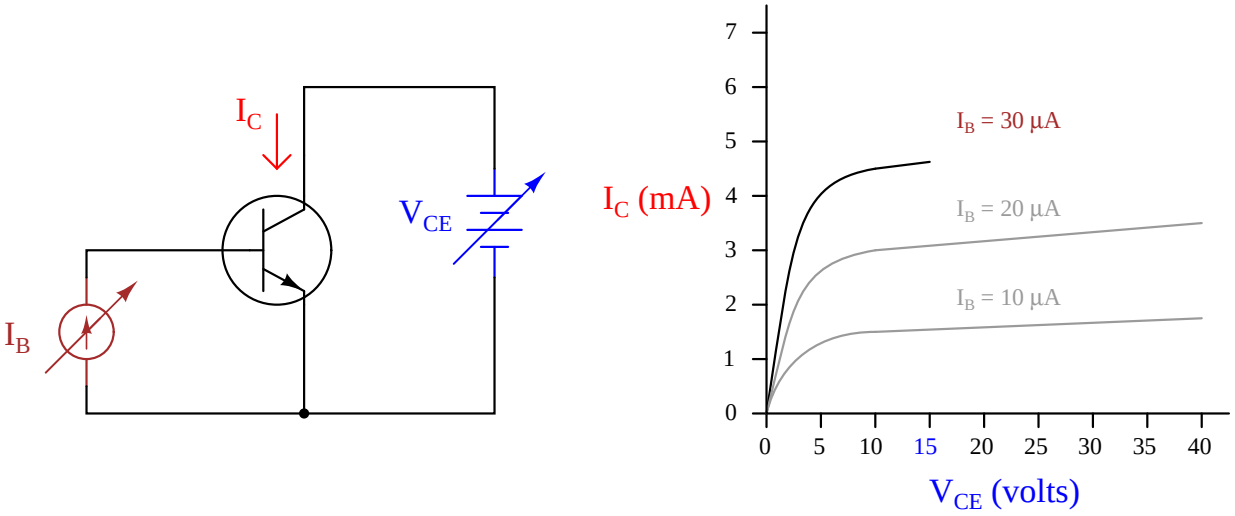

I_C
V_CE
I_B
I_C (mA)
7
6
5
4
3
2
1
0
I_B = 20 µA
I_B = 10 µA
0 5 10 15 20 25 30 35 40
V_CE (volts)

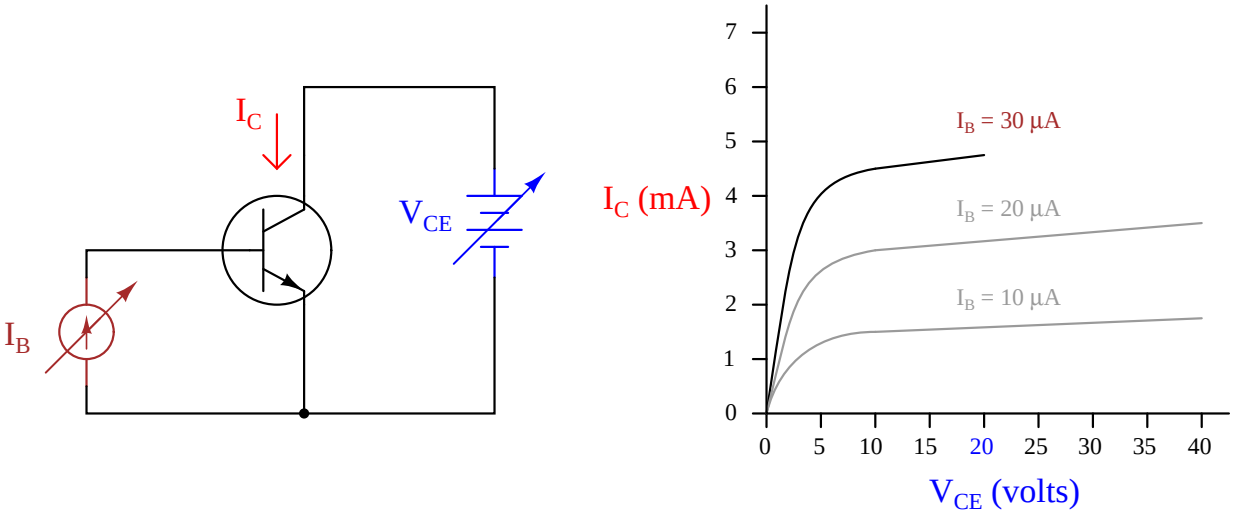

I_C
V_CE
I_B
I_C (mA)
7
6
5
4
3
2
1
0
I_B = 20 µA
I_B = 10 µA
0
5
10
15
20
25
30
35
40
V_CE (volts)

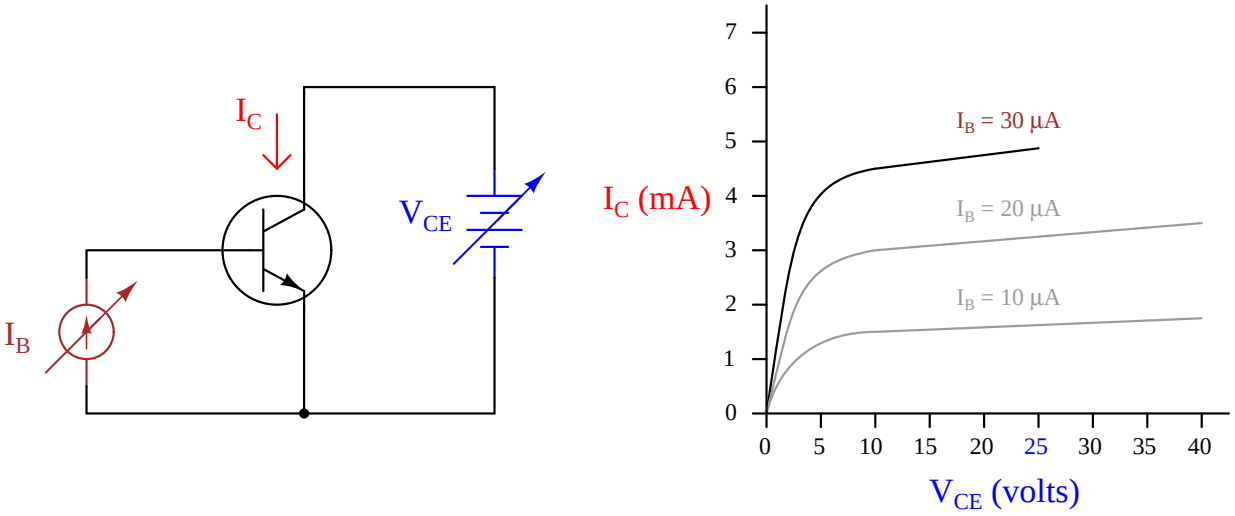

I_C
V_CE
I_B
I_C (mA)
7
6
5
4
3
2
1
0
I_B = 20 µA
I_B = 10 µA
0 5 10 15 20 25 30 35 40
V_CE (volts)

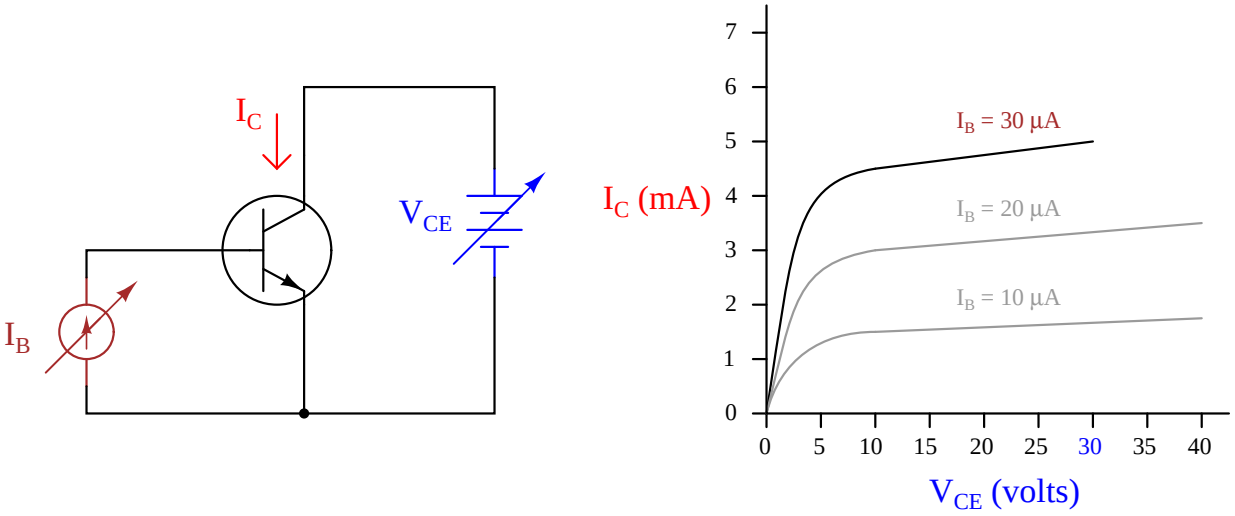

I_C
V_CE
I_B
I_C (mA)
7
6
5
4
3
2
1
0
I_B = 30 µA
I_B = 20 µA
I_B = 10 µA
0
5
10
15
20
25
30
35
40
V_CE (volts)

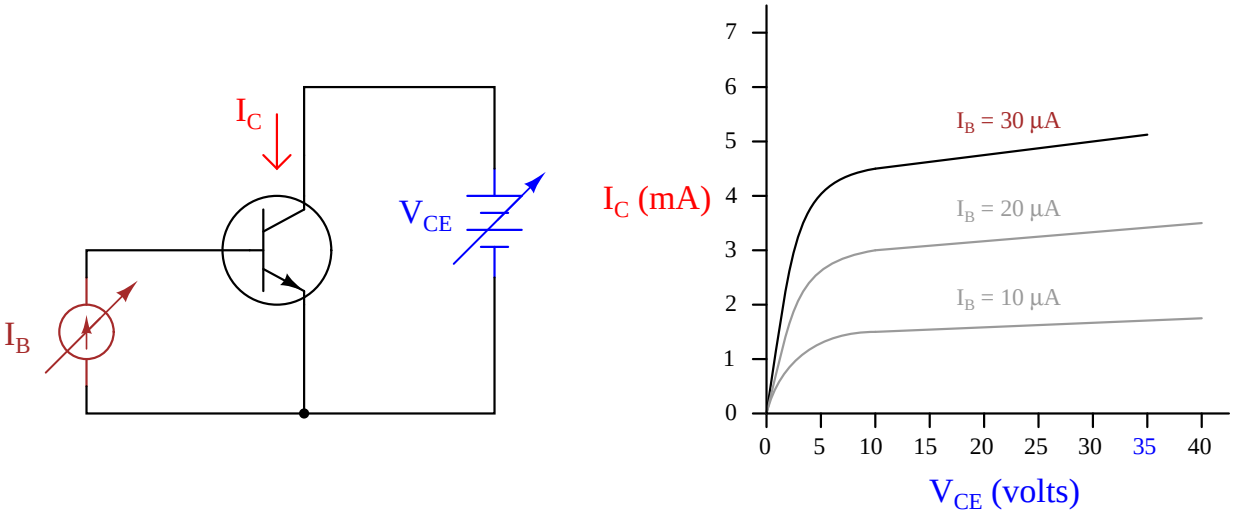

I_C
V_CE
I_B
I_C (mA)
I_B = 30 µA
I_B = 20 µA
I_B = 10 µA
0
1
2
3
4
5
6
7
0
5
10
15
20
25
30
35
40
V_CE (volts)

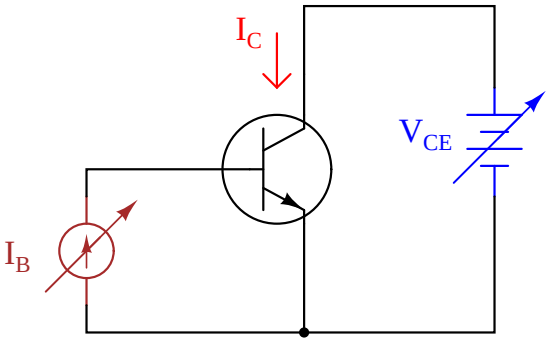

I_C
V_CE
I_B
I_C (mA)
7
6
5
4
3
2
1
0
I_B = 30 µA
I_B = 20 µA
I_B = 10 µA
0 5 10 15 20 25 30 35 40
V_CE (volts)

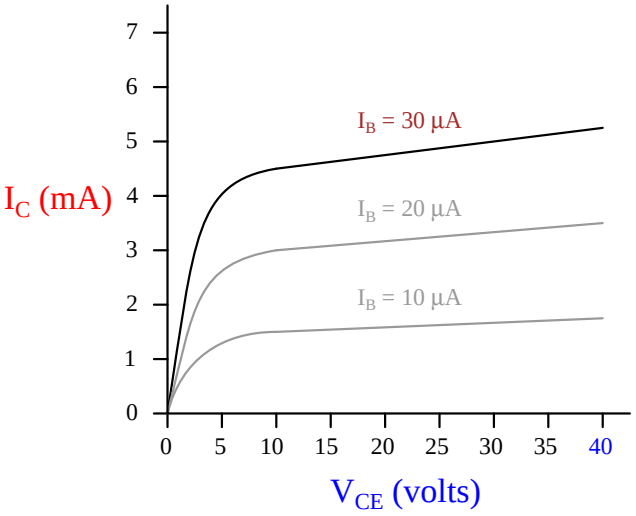

I_C
V_CE
I_B
I_C (mA)
7
6
5
4
3
2
1
0
I_B = 30 µA
I_B = 20 µA
I_B = 10 µA
0 5 10 15 20 25 30 35 40
V_CE (volts)

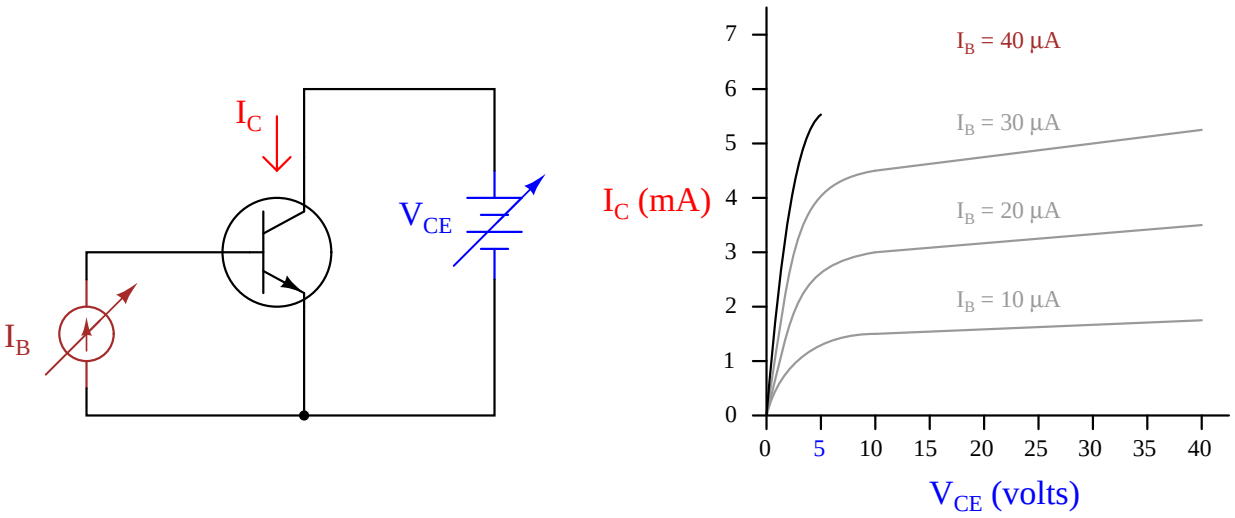

I_C
V_CE
I_B
I_C (mA)
7
6
5
4
3
2
1
0
I_B = 30 µA
I_B = 20 µA
I_B = 10 µA
0
5
10
15
20
25
30
35
40
V_CE (volts)

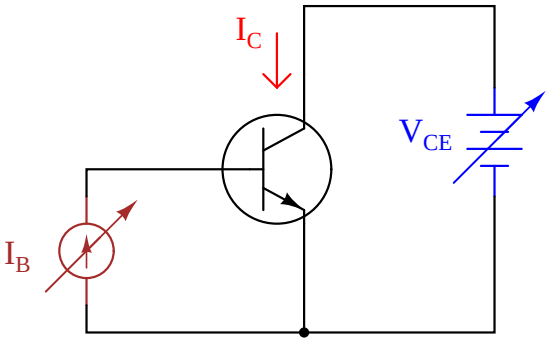

I_C
V_CE
I_B
I_C (mA)
I_B = 30 µA
I_B = 20 µA
I_B = 10 µA
0
1
2
3
4
5
6
7
0
5
10
15
20
25
30
35
40
V_CE (volts)

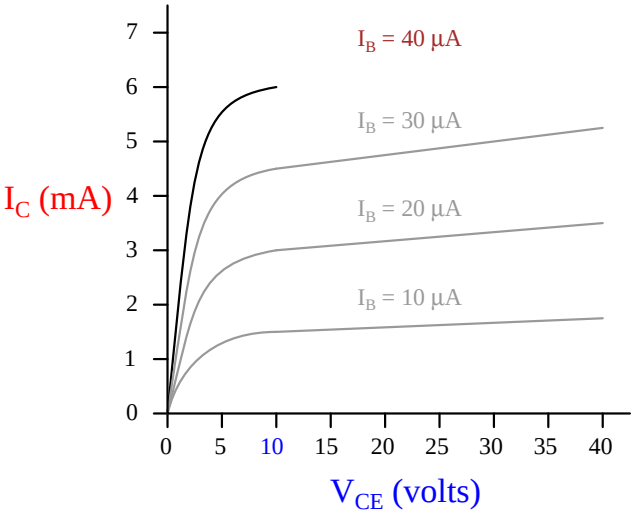

I_C
V_CE
I_B
I_C (mA)
7
6
5
4
3
2
1
0
I_B = 30 µA
I_B = 20 µA
I_B = 10 µA
0
5
10
15
20
25
30
35
40
V_CE (volts)

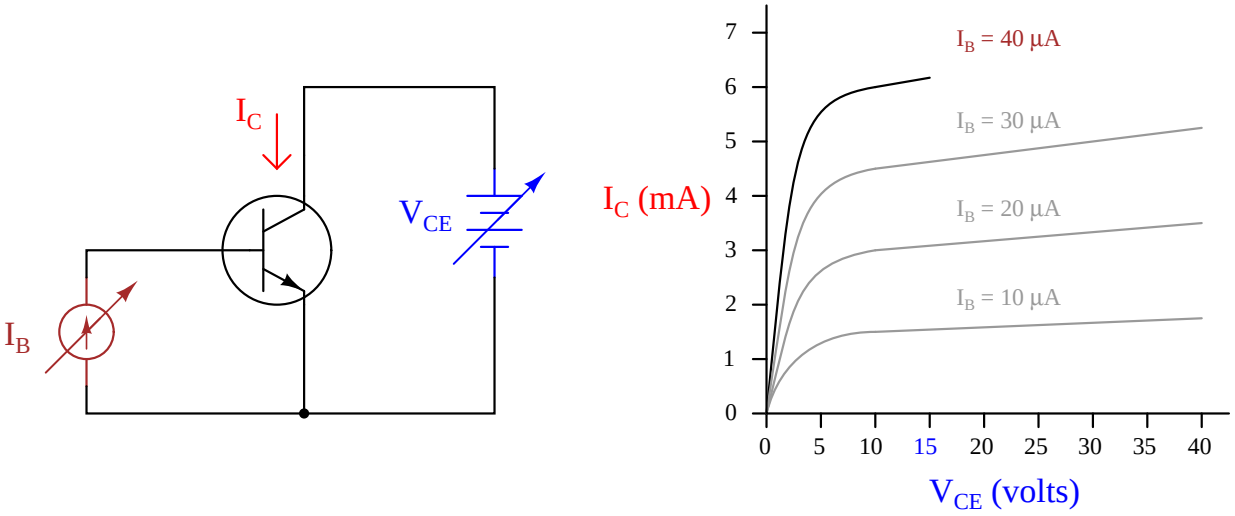

I_C
V_CE
I_B

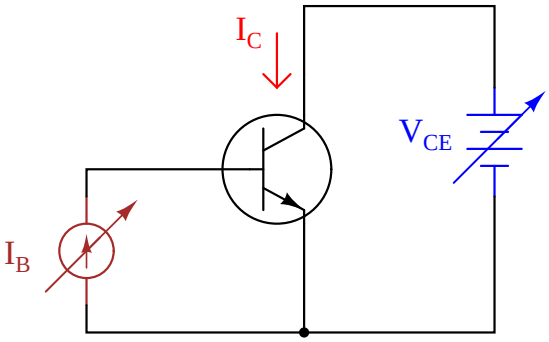

I_C (mA)
7
6
5
4
3
2
1
0
I_B = 30 µA
I_B = 20 µA
I_B = 10 µA
0 5 10 15 20 25 30 35 40
V_CE (volts)

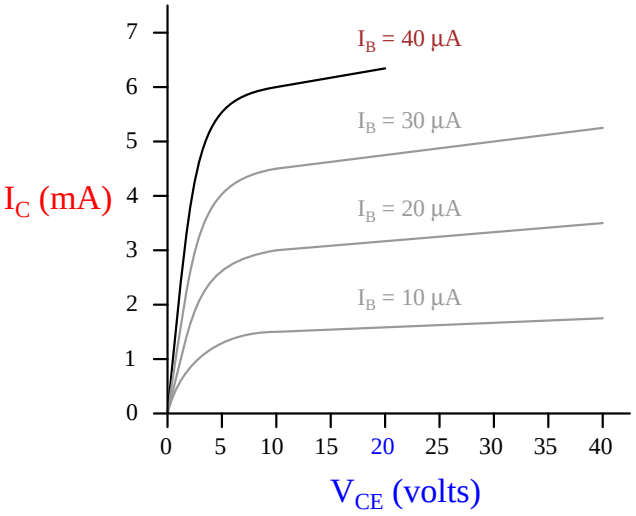

I_C
V_CE
I_B
I_C (mA)
I_B = 40 µA
I_B = 30 µA
I_B = 20 µA
I_B = 10 µA
0
1
2
3
4
5
6
7
0
5
10
15
20
25
30
35
40
V_CE (volts)

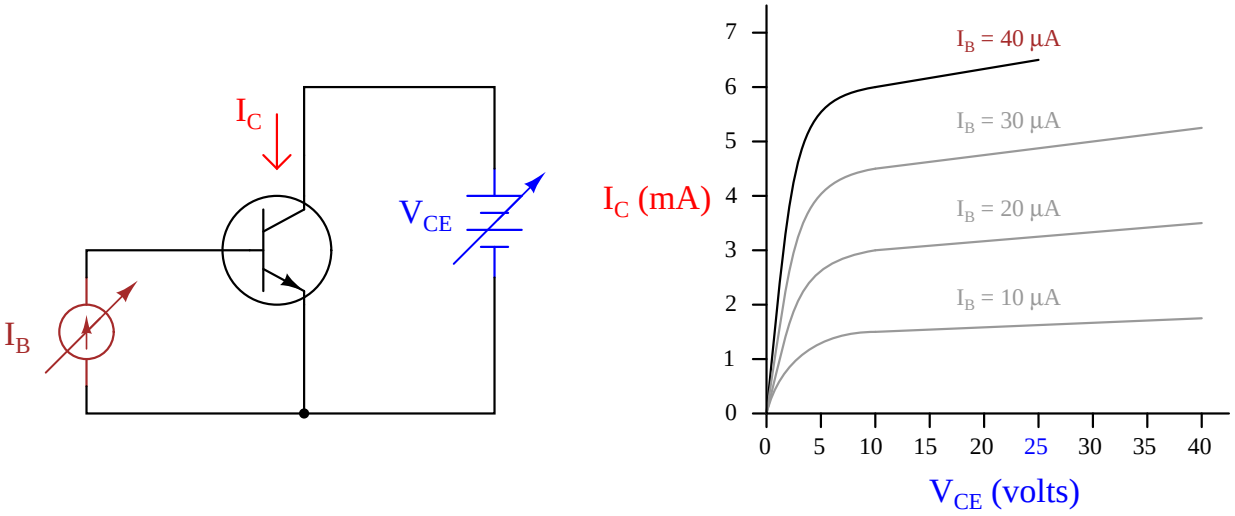

I_C
V_CE
I_B

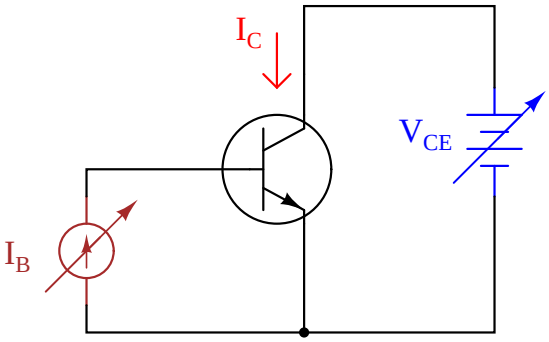

I_B = 40 µA
I_B = 30 µA
I_B = 20 µA
I_B = 10 µA
I_C (mA)
0
1
2
3
4
5
6
7
0
5
10
15
20
25
30
35
40
V_CE (volts)

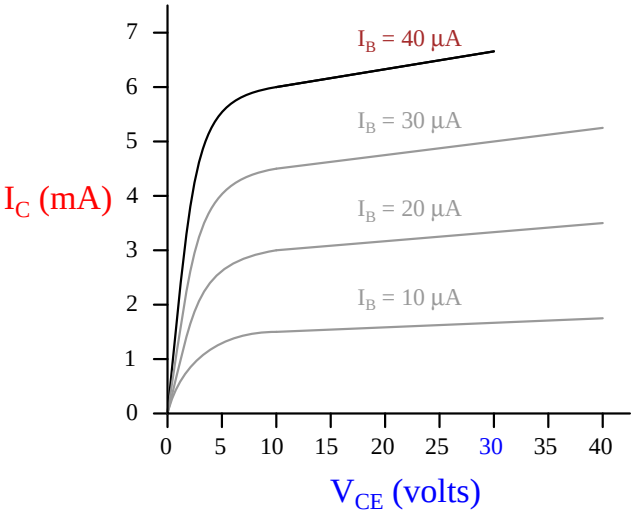

I_C
V_CE
I_B
I_C (mA)
I_B = 40 µA
I_B = 30 µA
I_B = 20 µA
I_B = 10 µA
0
1
2
3
4
5
6
7
0
5
10
15
20
25
30
35
40
V_CE (volts)

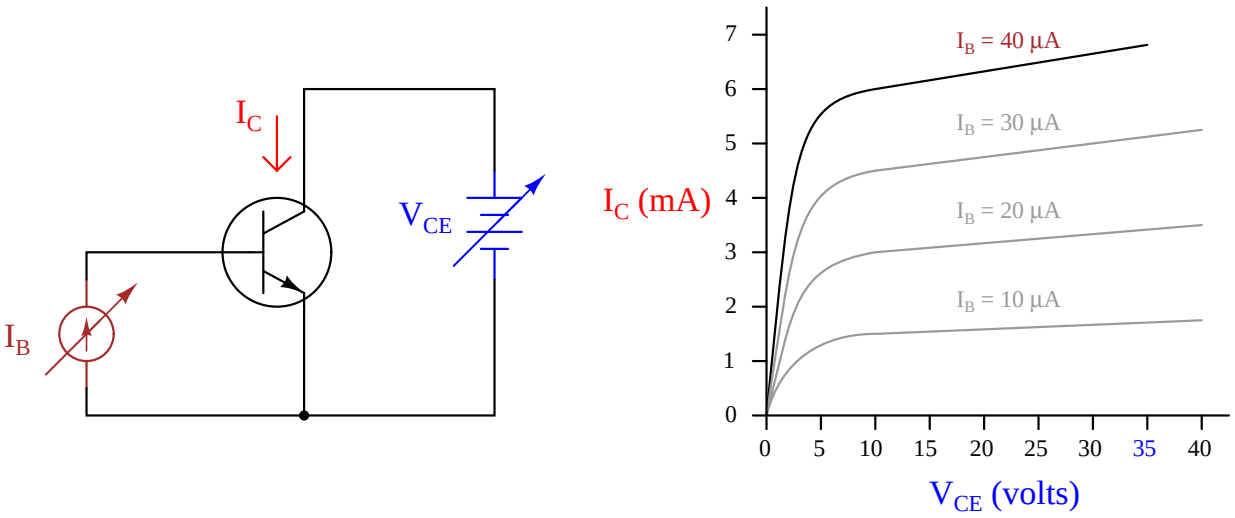

I_C
V_CE
I_B

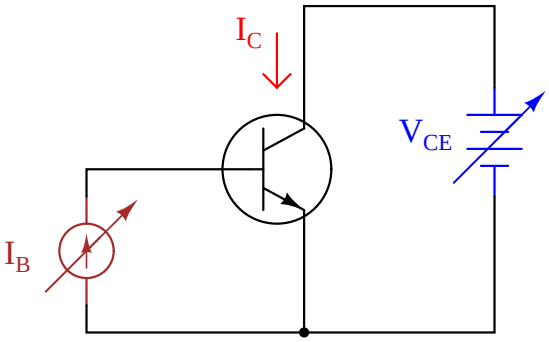

I_C (mA)
I_B = 40 µA
I_B = 30 µA
I_B = 20 µA
I_B = 10 µA
0
1
2
3
4
5
6
7
0
5
10
15
20
25
30
35
40
V_CE (volts)

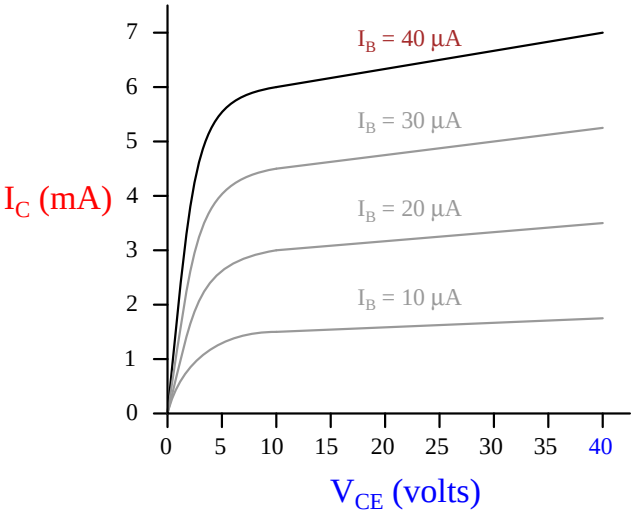

I_C
V_CE
I_B
I_C (mA)
I_B = 40 µA
I_B = 30 µA
I_B = 20 µA
I_B = 10 µA
0
1
2
3
4
5
6
7
0
5
10
15
20
25
30
35
40
V_CE (volts)

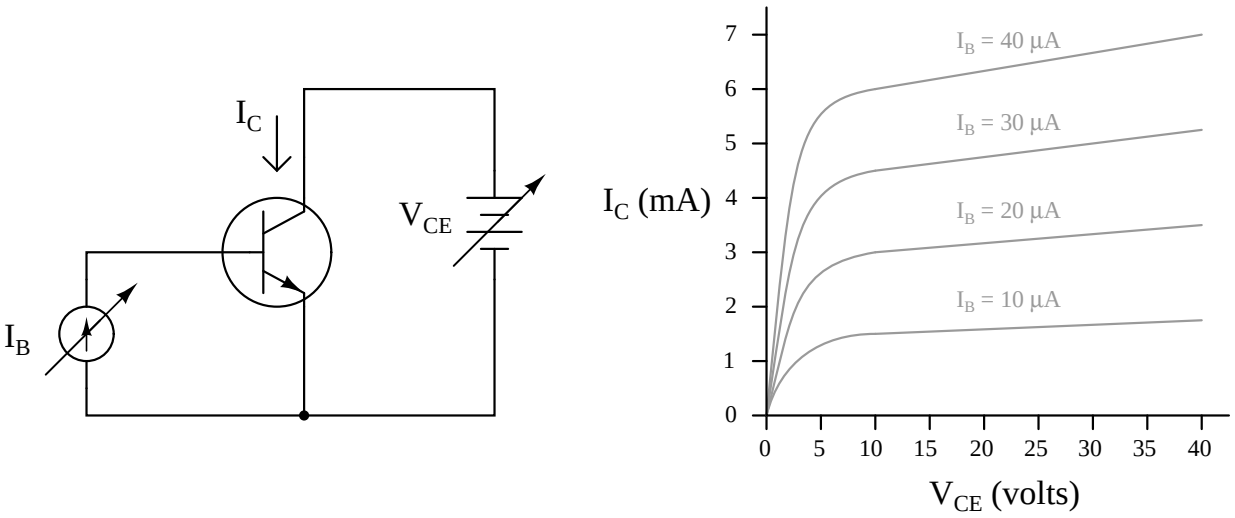

I_C
V_CE
I_B


I_C (mA)
I_B = 40 µA
I_B = 30 µA
I_B = 20 µA
I_B = 10 µA
0
1
2
3
4
5
6
7
0
5
10
15
20
25
30
35
40
V_CE (volts)


I_C
V_CE
I_B
I_C (mA)
I_B = 40 µA
I_B = 30 µA
I_B = 20 µA
I_B = 10 µA
0
1
2
3
4
5
6
7
0
5
10
15
20
25
30
35
40
V_CE (volts)


I_C
V_CE
I_B


I_C (mA)
I_B = 40 µA
I_B = 30 µA
I_B = 20 µA
I_B = 10 µA
0
1
2
3
4
5
6
7
0
5
10
15
20
25
30
35
40
V_CE (volts)


I_C
V_CE
I_B
I_C (mA)
I_B = 40 µA
I_B = 30 µA
I_B = 20 µA
I_B = 10 µA
0
1
2
3
4
5
6
7
0
5
10
15
20
25
30
35
40
V_CE (volts)

Chapter 7

Questions

This learning module, along with all others in the ModEL collection, is designed to be used in an inverted instructional environment where students independently read[1] the tutorials and attempt to answer questions on their own *prior* to the instructor's interaction with them. In place of lecture[2], the instructor engages with students in Socratic-style dialogue, probing and challenging their understanding of the subject matter through inquiry.

Answers are not provided for questions within this chapter, and this is by design. Solved problems may be found in the Tutorial and Derivation chapters, instead. The goal here is *independence*, and this requires students to be challenged in ways where others cannot think for them. Remember that you always have the tools of *experimentation* and *computer simulation* (e.g. SPICE) to explore concepts!

The following lists contain ideas for Socratic-style questions and challenges. Upon inspection, one will notice a strong theme of *metacognition* within these statements: they are designed to foster a regular habit of examining one's own thoughts as a means toward clearer thinking. As such these sample questions are useful both for instructor-led discussions as well as for self-study.

[1] Technical reading is an essential academic skill for any technical practitioner to possess for the simple reason that the most comprehensive, accurate, and useful information to be found for developing technical competence is in textual form. Technical careers in general are characterized by the need for continuous learning to remain current with standards and technology, and therefore any technical practitioner who cannot read well is handicapped in their professional development. An excellent resource for educators on improving students' reading prowess through intentional effort and strategy is the book textitReading For Understanding – How Reading Apprenticeship Improves Disciplinary Learning in Secondary and College Classrooms by Ruth Schoenbach, Cynthia Greenleaf, and Lynn Murphy.

[2] Lecture is popular as a teaching method because it is easy to implement: any reasonably articulate subject matter expert can talk to students, even with little preparation. However, it is also quite problematic. A good lecture always makes complicated concepts seem easier than they are, which is bad for students because it instills a false sense of confidence in their own understanding; reading and re-articulation requires more cognitive effort and serves to verify comprehension. A culture of teaching-by-lecture fosters a debilitating dependence upon direct personal instruction, whereas the challenges of modern life demand independent and critical thought made possible only by gathering information and perspectives from afar. Information presented in a lecture is ephemeral, easily lost to failures of memory and dictation; text is forever, and may be referenced at any time.

General challenges following tutorial reading

- Summarize as much of the text as you can in one paragraph of your own words. A helpful strategy is to explain ideas as you would for an intelligent child: as simple as you can without compromising too much accuracy.
- Simplify a particular section of the text, for example a paragraph or even a single sentence, so as to capture the same fundamental idea in fewer words.
- Where did the text make the most sense to you? What was it about the text's presentation that made it clear?
- Identify where it might be easy for someone to misunderstand the text, and explain why you think it could be confusing.
- Identify any new concept(s) presented in the text, and explain in your own words.
- Identify any familiar concept(s) such as physical laws or principles applied or referenced in the text.
- Devise a proof of concept experiment demonstrating an important principle, physical law, or technical innovation represented in the text.
- Devise an experiment to disprove a plausible misconception.
- Did the text reveal any misconceptions you might have harbored? If so, describe the misconception(s) and the reason(s) why you now know them to be incorrect.
- Describe any useful problem-solving strategies applied in the text.
- Devise a question of your own to challenge a reader's comprehension of the text.

General follow-up challenges for assigned problems

- Identify where any fundamental laws or principles apply to the solution of this problem, especially before applying any mathematical techniques.
- Devise a thought experiment to explore the characteristics of the problem scenario, applying known laws and principles to mentally model its behavior.
- Describe in detail your own strategy for solving this problem. How did you identify and organized the given information? Did you sketch any diagrams to help frame the problem?
- Is there more than one way to solve this problem? Which method seems best to you?
- Show the work you did in solving this problem, even if the solution is incomplete or incorrect.
- What would you say was the most challenging part of this problem, and why was it so?
- Was any important information missing from the problem which you had to research or recall?
- Was there any extraneous information presented within this problem? If so, what was it and why did it not matter?
- Examine someone else's solution to identify where they applied fundamental laws or principles.
- Simplify the problem from its given form and show how to solve this simpler version of it. Examples include eliminating certain variables or conditions, altering values to simpler (usually whole) numbers, applying a limiting case (i.e. altering a variable to some extreme or ultimate value).
- For quantitative problems, identify the real-world meaning of all intermediate calculations: their units of measurement, where they fit into the scenario at hand. Annotate any diagrams or illustrations with these calculated values.
- For quantitative problems, try approaching it qualitatively instead, thinking in terms of "increase" and "decrease" rather than definite values.
- For qualitative problems, try approaching it quantitatively instead, proposing simple numerical values for the variables.
- Were there any assumptions you made while solving this problem? Would your solution change if one of those assumptions were altered?
- Identify where it would be easy for someone to go astray in attempting to solve this problem.
- Formulate your own problem based on what you learned solving this one.

General follow-up challenges for experiments or projects

- In what way(s) was this experiment or project easy to complete?
- Identify some of the challenges you faced in completing this experiment or project.

- Show how thorough documentation assisted in the completion of this experiment or project.
- Which fundamental laws or principles are key to this system's function?
- Identify any way(s) in which one might obtain false or otherwise misleading measurements from test equipment in this system.
- What will happen if (component X) fails (open/shorted/etc.)?
- What would have to occur to make this system unsafe?

7.1 Conceptual reasoning

These questions are designed to stimulate your analytic and synthetic thinking[3]. In a Socratic discussion with your instructor, the goal is for these questions to prompt an extended dialogue where assumptions are revealed, conclusions are tested, and understanding is sharpened. Your instructor may also pose additional questions based on those assigned, in order to further probe and refine your conceptual understanding.

Questions that follow are presented to challenge and probe your understanding of various concepts presented in the tutorial. These questions are intended to serve as a guide for the Socratic dialogue between yourself and the instructor. Your instructor's task is to ensure you have a sound grasp of these concepts, and the questions contained in this document are merely a means to this end. Your instructor may, at his or her discretion, alter or substitute questions for the benefit of tailoring the discussion to each student's needs. The only absolute requirement is that each student is challenged and assessed at a level equal to or greater than that represented by the documented questions.

It is far more important that you convey your *reasoning* than it is to simply convey a correct answer. For this reason, you should refrain from researching other information sources to answer questions. What matters here is that *you* are doing the thinking. If the answer is incorrect, your instructor will work with you to correct it through proper reasoning. A correct answer without an adequate explanation of how you derived that answer is unacceptable, as it does not aid the learning or assessment process.

You will note a conspicuous lack of answers given for these conceptual questions. Unlike standard textbooks where answers to every other question are given somewhere toward the back of the book, here in these learning modules students must rely on other means to check their work. The best way by far is to debate the answers with fellow students and also with the instructor during the Socratic dialogue sessions intended to be used with these learning modules. Reasoning through challenging questions with other people is an excellent tool for developing strong reasoning skills.

Another means of checking your conceptual answers, where applicable, is to use circuit simulation software to explore the effects of changes made to circuits. For example, if one of these conceptual questions challenges you to predict the effects of altering some component parameter in a circuit, you may check the validity of your work by simulating that same parameter change within software and seeing if the results agree.

[3] *Analytical* thinking involves the "disassembly" of an idea into its constituent parts, analogous to dissection. *Synthetic* thinking involves the "assembly" of a new idea comprised of multiple concepts, analogous to construction. Both activities are high-level cognitive skills, extremely important for effective problem-solving, necessitating frequent challenge and regular practice to fully develop.

7.1.1 Reading outline and reflections

"Reading maketh a full man; conference a ready man; and writing an exact man" – Francis Bacon

Francis Bacon's advice is a blueprint for effective education: reading provides the learner with knowledge, writing focuses the learner's thoughts, and critical dialogue equips the learner to confidently communicate and apply their learning. Independent acquisition and application of knowledge is a powerful skill, well worth the effort to cultivate. To this end, students should read these educational resources closely, journal their own reflections on the reading, and discuss in detail their findings with classmates and instructor(s). You should be able to do all of the following after reading any instructional text:

☑ Briefly SUMMARIZE THE TEXT in the form of a journal entry documenting your learning as you progress through the course of study. Share this summary in dialogue with your classmates and instructor. Journaling is an excellent self-test of thorough reading because you cannot clearly express what you have not read or did not comprehend.

☑ Demonstrate ACTIVE READING STRATEGIES, including verbalizing your impressions as you read, simplifying long passages to convey the same ideas using fewer words, annotating text and illustrations with your own interpretations, working through mathematical examples shown in the text, cross-referencing passages with relevant illustrations and/or other passages, identifying problem-solving strategies applied by the author, etc. Technical reading is a special case of problem-solving, and so these strategies work precisely because they help solve any problem: paying attention to your own thoughts (metacognition), eliminating unnecessary complexities, identifying what makes sense, paying close attention to details, drawing connections between separated facts, and noting the successful strategies of others.

☑ Identify IMPORTANT THEMES, especially GENERAL LAWS and PRINCIPLES, expounded in the text and express them in the simplest of terms as though you were teaching an intelligent child. This emphasizes connections between related topics and develops your ability to communicate complex ideas to anyone.

☑ Form YOUR OWN QUESTIONS based on the reading, and then pose them to your instructor and classmates for their consideration. Anticipate both correct and incorrect answers, the incorrect answer(s) assuming one or more plausible misconceptions. This helps you view the subject from different perspectives to grasp it more fully.

☑ Devise EXPERIMENTS to test claims presented in the reading, or to disprove misconceptions. Predict possible outcomes of these experiments, and evaluate their meanings: what result(s) would confirm, and what would constitute disproof? Running mental simulations and evaluating results is essential to scientific and diagnostic reasoning.

☑ Specifically identify any points you found CONFUSING. The reason for doing this is to help diagnose misconceptions and overcome barriers to learning.

7.1.2 Foundational concepts

Correct analysis and diagnosis of electric circuits begins with a proper understanding of some basic concepts. The following is a list of some important concepts referenced in this module's full tutorial. Define each of them in your own words, and be prepared to illustrate each of these concepts with a description of a practical example and/or a live demonstration.

Energy

Conservation of Energy

Thought experiments as a problem-solving strategy

Kirchhoff's Current Law

Amplification

Transformer

Active versus Passive devices

Gain

Relay

Analog signal

Discrete signal

Thermionic emission

Vacuum tube

Doping

N-type semiconductor

P-type semiconductor

Depletion region

Diode

Forward-biasing

Reverse-biasing

Forward voltage drop

Photodiode

Majority versus Minority carriers

Current regulation

Beta

Dimensional analysis

7.1.3 BJT patent

Read the John Bardeen's and Walter Brattain's Transistor Patent section of the Historical References chapter, and answer the following questions:

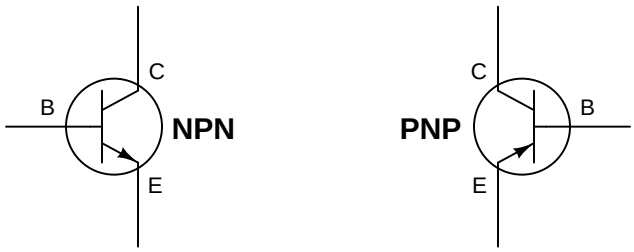
- What do you suppose the authors mean by the word *translate*?
- Identify some of the lesser-used semiconducting materials mentioned by the authors.
- What type of semiconductor (N- or P-type) is the block of semiconducting material made of in this transistor, and how can we tell?
- What is the significance of matching the impedance (i.e. resistance) of a load to the internal resistance of the collector terminal in the transistor amplifier circuit?
- Examining the construction of this prototype transistor, why do you think the "base" terminal is so named?
- What does the term *gain* mean in the context of this device?

Challenges

- What are some advantages this transistor enjoyed over vacuum tubes of that era?

7.1.4 Controlling and controlled currents

Bipolar junction transistor (BJT) function is usually considered in terms of currents: a relatively small current through one of the transistor's terminals exerts control over a much larger current. Draw the directions of all currents for these two transistors (one NPN and one PNP), clearly identifying which of the currents is *doing* the control, and which of the currents is *being* controlled, for each of these transistors:



Also, apply the analogy described in the Full Tutorial of sand grains being dropped onto a windy plateau to these two transistors:

- Which transistor quantity is analogous to the rate of sand being dropped from your hand?
- Which transistor quantity is analogous to the rate of sand being swept downwind?
- Which transistor quantity is analogous to the wind itself?
- Which transistor quantity is analogous to dropped grains of sand that get stuck on your shoes and don't drift downwind?
- Explain what we would have to do (electrically) to the transistor to simulate the analogous condition of wind picking up speed but having the same number of sand grains per second being carried downwind?

Challenges

- Which type of electric charge carrier (electrons or holes) is being injected into the base region of a PNP transistor, and from which direction?
- Which type of electric charge carrier (electrons or holes) is being injected into the base region of an NPN transistor, and from which direction?

7.1.5 Controlling voltages

Draw the polarities (+ and −) of the applied voltages necessary to turn both these transistors on:

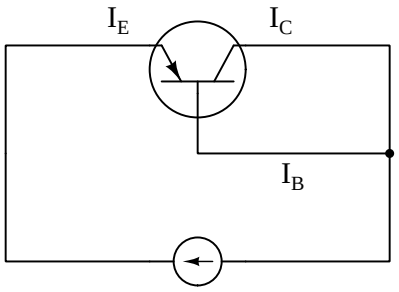
Also, draw the direction of the *controlled* current (flowing between collector and emitter) that will result from a power source properly connected between these terminals.

Challenges

- Which type of electric charge carrier (electrons or holes) is being injected into the base region of a PNP transistor, and from which direction?
- Which type of electric charge carrier (electrons or holes) is being injected into the base region of an NPN transistor, and from which direction?

7.1.6 Relative current magnitudes

Compare the relative magnitudes of each current in this bipolar transistor circuit:



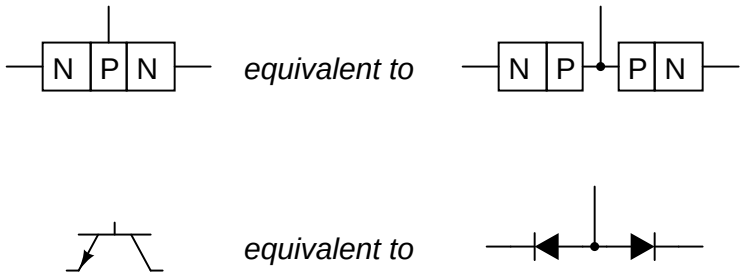
Which current is the smallest and which is the largest? Are there any two currents that are closer in magnitude than with the third? If so, which currents are they?

Challenges

- What would happen if the base terminal of the transistor became disconnected?
- What would happen if the emitter terminal of the transistor became disconnected?
- What would happen if the collector terminal of the transistor became disconnected?

7.1.7 Failed attempt at making a transistor

A beginning electronics student is just learning about transistors, and reads in the textbook that a bipolar transistor (either NPN or PNP) can be thought of as two diodes connected back-to-back as such:



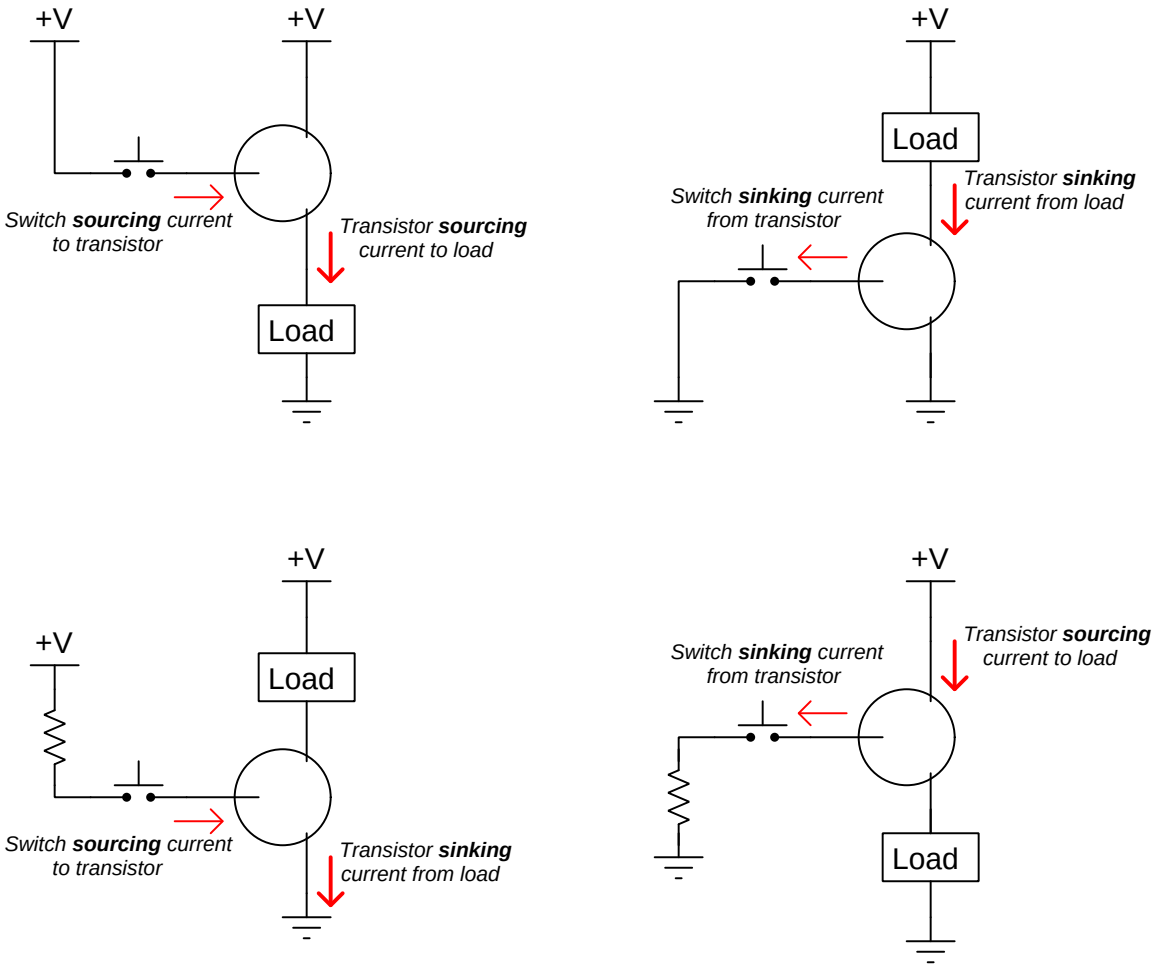
Acting on this idea, the student proceeds to connect two 1N4001 rectifying diodes back-to-back and try to use it as a transistor. This idea does not work: although the diode pair reads the same patterns of continuity as a transistor would, it does not amplify. Explain why.

Challenges

- What do you suppose is *really* meant by the textbook's statement of bipolar transistors being equivalent to back-to-back diodes, if two diodes connected back-to-back do not exhibit amplifying behavior? Is this a completely wrong statement, or is there some truth to it?

7.1.8 Proper BJT types

Choose the right type of bipolar junction transistor for each of these switching applications (i.e. either NPN or PNP), drawing the correct transistor symbol inside each circle:



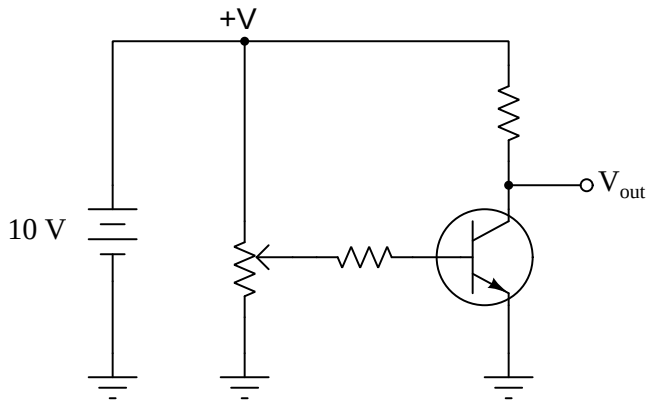
Also, explain why resistors are necessary in some of these circuits but not others.

Challenges

- Define the terms “source” and “sink” as they apply to these circuits.

7.1.9 Common-emitter behavior

Describe what the output voltage of this transistor circuit will do (measured with reference to ground), if the potentiometer wiper begins at the full-down position (common with ground), and is slowly moved in the *upward* direction (closer to +V):

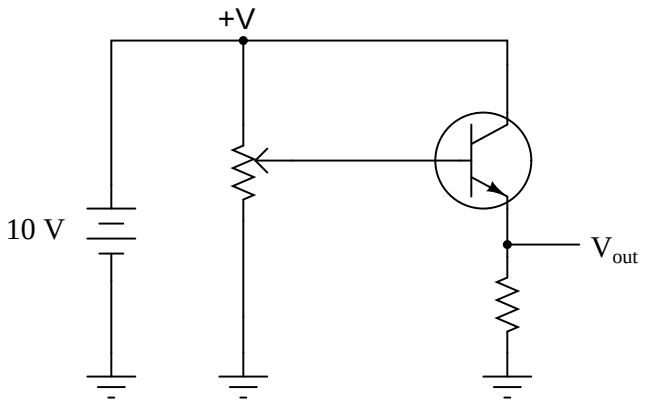


Challenges

- Explain how the problem-solving technique of *limiting cases* might be applied to this circuit.
- Based on this result, would you be inclined to call this amplifier an *inverting* or a *noninverting* circuit?
- Sketch a qualitative graph expressing V_{out} as a function of V_{pot} (i.e. voltage between the potentiometer wiper and ground).

7.1.10 Common-collector behavior

Describe what the output voltage of this transistor circuit will do (measured with reference to ground), if the potentiometer wiper begins at the full-down position (common with ground), and is slowly moved in the *upward* direction (closer to +V):

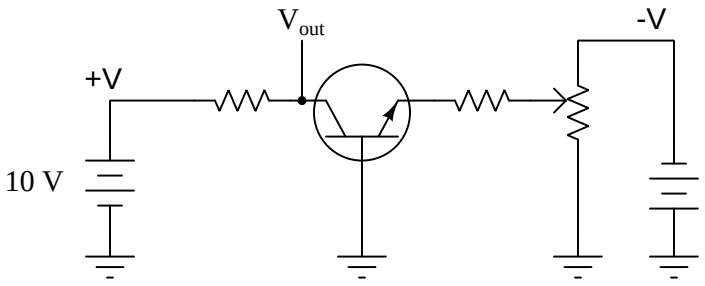


Challenges

- Explain how the problem-solving technique of *limiting cases* might be applied to this circuit.
- Based on this result, would you be inclined to call this amplifier an *inverting* or a *noninverting* circuit?
- Sketch a qualitative graph expressing V_{out} as a function of V_{pot} (i.e. voltage between the potentiometer wiper and ground).

7.1.11 Common-base behavior

Describe what the output voltage of this transistor circuit will do (measured with reference to ground), if the potentiometer wiper begins at the full-down position (common with ground), and is slowly moved in the *upward* direction (closer to $-V$):



Challenges

- Explain how the problem-solving technique of *limiting cases* might be applied to this circuit.
- Based on this result, would you be inclined to call this amplifier an *inverting* or a *noninverting* circuit?
- Sketch a qualitative graph expressing V_{out} as a function of V_{pot} (i.e. voltage between the potentiometer wiper and ground).

7.2 Quantitative reasoning

These questions are designed to stimulate your computational thinking. In a Socratic discussion with your instructor, the goal is for these questions to reveal your mathematical approach(es) to problem-solving so that good technique and sound reasoning may be reinforced. Your instructor may also pose additional questions based on those assigned, in order to observe your problem-solving firsthand.

Mental arithmetic and estimations are strongly encouraged for all calculations, because without these abilities you will be unable to readily detect errors caused by calculator misuse (e.g. keystroke errors).

You will note a conspicuous lack of answers given for these quantitative questions. Unlike standard textbooks where answers to every other question are given somewhere toward the back of the book, here in these learning modules students must rely on other means to check their work. My advice is to use circuit simulation software such as SPICE to check the correctness of quantitative answers. Refer to those learning modules within this collection focusing on SPICE to see worked examples which you may use directly as practice problems for your own study, and/or as templates you may modify to run your own analyses and generate your own practice problems.

Completely worked example problems found in the Tutorial may also serve as "test cases[4]" for gaining proficiency in the use of circuit simulation software, and then once that proficiency is gained you will never need to rely[5] on an answer key!

[4] In other words, set up the circuit simulation software to analyze the same circuit examples found in the Tutorial. If the simulated results match the answers shown in the Tutorial, it confirms the simulation has properly run. If the simulated results disagree with the Tutorial's answers, something has been set up incorrectly in the simulation software. Using every Tutorial as practice in this way will quickly develop proficiency in the use of circuit simulation software.

[5] This approach is perfectly in keeping with the instructional philosophy of these learning modules: *teaching students to be self-sufficient thinkers*. Answer keys can be useful, but it is even more useful to your long-term success to have a set of tools on hand for checking your own work, because once you have left school and are on your own, there will no longer be "answer keys" available for the problems you will have to solve.

7.2.1 Miscellaneous physical constants

Note: constants shown in **bold** type are *exact*, not approximations. Values inside of parentheses show one standard deviation (σ) of uncertainty in the final digits: for example, the magnetic permeability of free space value given as $1.25663706212(19) \times 10^{-6}$ H/m represents a center value (i.e. the location parameter) of $1.25663706212 \times 10^{-6}$ Henrys per meter with one standard deviation of uncertainty equal to $0.0000000000019 \times 10^{-6}$ Henrys per meter.

Avogadro's number (N_A) = **6.02214076** $\times 10^{23}$ **per mole** (mol^{-1})

Boltzmann's constant (k) = **1.380649** $\times 10^{-23}$ **Joules per Kelvin** (J/K)

Electronic charge (e) = **1.602176634** $\times 10^{-19}$ **Coulomb** (C)

Faraday constant (F) = **96,485.33212...** $\times 10^{4}$ **Coulombs per mole** (C/mol)

Magnetic permeability of free space (μ_0) = $1.25663706212(19) \times 10^{-6}$ Henrys per meter (H/m)

Electric permittivity of free space (ϵ_0) = $8.8541878128(13) \times 10^{-12}$ Farads per meter (F/m)

Characteristic impedance of free space (Z_0) = 376.730313668(57) Ohms (Ω)

Gravitational constant (G) = $6.67430(15) \times 10^{-11}$ cubic meters per kilogram-seconds squared ($m^3/kg\text{-}s^2$)

Molar gas constant (R) = **8.314462618... Joules per mole-Kelvin** (J/mol-K) = 0.08205746(14) liters-atmospheres per mole-Kelvin

Planck constant (h) = **6.62607015** $\times 10^{-34}$ **joule-seconds** (J-s)

Stefan-Boltzmann constant (σ) = **5.670374419...** $\times 10^{-8}$ **Watts per square meter-Kelvin**^{4} ($W/m^2 \cdot K^4$)

Speed of light in a vacuum (c) = **299,792,458 meters per second** (m/s) = 186282.4 miles per second (mi/s)

Note: All constants taken from NIST data "Fundamental Physical Constants – Complete Listing", from `http://physics.nist.gov/constants`, National Institute of Standards and Technology (NIST), 2018 CODATA Adjustment.

7.2.2 Introduction to spreadsheets

A powerful computational tool you are encouraged to use in your work is a *spreadsheet*. Available on most personal computers (e.g. Microsoft Excel), *spreadsheet* software performs numerical calculations based on number values and formulae entered into cells of a grid. This grid is typically arranged as lettered columns and numbered rows, with each cell of the grid identified by its column/row coordinates (e.g. cell `B3`, cell `A8`). Each cell may contain a string of text, a number value, or a mathematical formula. The spreadsheet automatically updates the results of all mathematical formulae whenever the entered number values are changed. This means it is possible to set up a spreadsheet to perform a series of calculations on entered data, and those calculations will be re-done by the computer any time the data points are edited in any way.

For example, the following spreadsheet calculates average speed based on entered values of distance traveled and time elapsed:

	A	B	C	D
1	Distance traveled	46.9	Kilometers	
2	Time elapsed	1.18	Hours	
3	Average speed	= B1 / B2	km/h	
4				
5				

Text labels contained in cells `A1` through `A3` and cells `C1` through `C3` exist solely for readability and are not involved in any calculations. Cell `B1` contains a sample distance value while cell `B2` contains a sample time value. The formula for computing speed is contained in cell `B3`. Note how this formula begins with an "equals" symbol (`=`), references the values for distance and speed by lettered column and numbered row coordinates (`B1` and `B2`), and uses a forward slash symbol for division (`/`). The coordinates `B1` and `B2` function as *variables*[6] would in an algebraic formula.

When this spreadsheet is executed, the numerical value `39.74576` will appear in cell `B3` rather than the formula `= B1 / B2`, because 39.74576 is the computed speed value given 46.9 kilometers traveled over a period of 1.18 hours. If a different numerical value for distance is entered into cell `B1` or a different value for time is entered into cell `B2`, cell `B3`'s value will automatically update. All you need to do is set up the given values and any formulae into the spreadsheet, and the computer will do all the calculations for you.

Cell `B3` may be referenced by other formulae in the spreadsheet if desired, since it is a variable just like the given values contained in `B1` and `B2`. This means it is possible to set up an entire chain of calculations, one dependent on the result of another, in order to arrive at a final value. The arrangement of the given data and formulae need not follow any pattern on the grid, which means you may place them anywhere.

[6] Spreadsheets may also provide means to attach text labels to cells for use as variable names (Microsoft Excel simply calls these labels "names"), but for simple spreadsheets such as those shown here it's usually easier just to use the standard coordinate naming for each cell.

Common[7] arithmetic operations available for your use in a spreadsheet include the following:

- Addition (`+`)
- Subtraction (`-`)
- Multiplication (`*`)
- Division (`/`)
- Powers (`^`)
- Square roots (`sqrt()`)
- Logarithms (`ln()` , `log10()`)

Parentheses may be used to ensure[8] proper order of operations within a complex formula. Consider this example of a spreadsheet implementing the *quadratic formula*, used to solve for roots of a polynomial expression in the form of $ax^2 + bx + c$:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

	A	B
1	x_1	= (-B4 + sqrt((B4^2) - (4*B3*B5))) / (2*B3)
2	x_2	= (-B4 - sqrt((B4^2) - (4*B3*B5))) / (2*B3)
3	a =	9
4	b =	5
5	c =	-2

This example is configured to compute roots[9] of the polynomial $9x^2 + 5x - 2$ because the values of 9, 5, and -2 have been inserted into cells `B3`, `B4`, and `B5`, respectively. Once this spreadsheet has been built, though, it may be used to calculate the roots of *any* second-degree polynomial expression simply by entering the new a, b, and c coefficients into cells `B3` through `B5`. The numerical values appearing in cells `B1` and `B2` will be automatically updated by the computer immediately following any changes made to the coefficients.

[7] Modern spreadsheet software offers a bewildering array of mathematical functions you may use in your computations. I recommend you consult the documentation for your particular spreadsheet for information on operations other than those listed here.

[8] Spreadsheet programs, like text-based programming languages, are designed to follow standard order of operations by default. However, my personal preference is to use parentheses even where strictly unnecessary just to make it clear to any other person viewing the formula what the intended order of operations is.

[9] Reviewing some algebra here, a *root* is a value for x that yields an overall value of zero for the polynomial. For this polynomial ($9x^2 + 5x - 2$) the two roots happen to be $x = 0.269381$ and $x = -0.82494$, with these values displayed in cells `B1` and `B2`, respectively upon execution of the spreadsheet.

Alternatively, one could break up the long quadratic formula into smaller pieces like this:

$$y = \sqrt{b^2 - 4ac} \qquad z = 2a$$

$$x = \frac{-b \pm y}{z}$$

	A	B	C
1	x_1	= (-B4 + C1) / C2	= sqrt((B4^2) - (4*B3*B5))
2	x_2	= (-B4 - C1) / C2	= 2*B3
3	a =	9	
4	b =	5	
5	c =	-2	

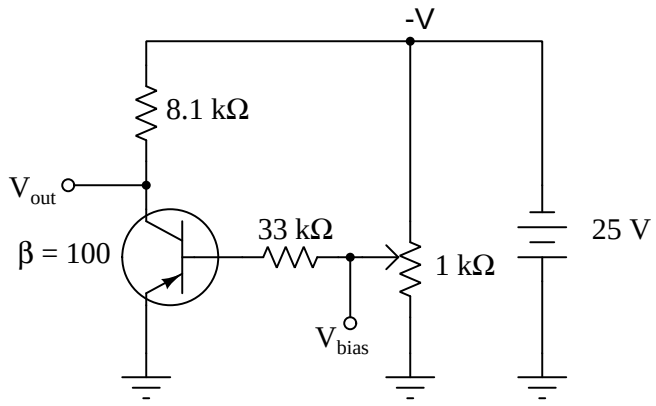
Note how the square-root term (y) is calculated in cell C1, and the denominator term (z) in cell C2. This makes the two final formulae (in cells B1 and B2) simpler to interpret. The positioning of all these cells on the grid is completely arbitrary[10] – all that matters is that they properly reference each other in the formulae.

Spreadsheets are particularly useful for situations where the same set of calculations representing a circuit or other system must be repeated for different initial conditions. The power of a spreadsheet is that it automates what would otherwise be a tedious set of calculations. One specific application of this is to simulate the effects of various components within a circuit failing with abnormal values (e.g. a shorted resistor simulated by making its value nearly zero; an open resistor simulated by making its value extremely large). Another application is analyzing the behavior of a circuit design given new components that are out of specification, and/or aging components experiencing drift over time.

[10] My personal preference is to locate all the "given" data in the upper-left cells of the spreadsheet grid (each data point flanked by a sensible name in the cell to the left and units of measurement in the cell to the right as illustrated in the first distance/time spreadsheet example), sometimes coloring them in order to clearly distinguish which cells contain entered data versus which cells contain computed results from formulae. I like to place all formulae in cells below the given data, and try to arrange them in logical order so that anyone examining my spreadsheet will be able to figure out *how* I constructed a solution. This is a general principle I believe all computer programmers should follow: *document and arrange your code to make it easy for other people to learn from it.*

7.2.3 DC bias calculations

Calculate collector current, base current, and output voltage (referenced to ground) if V_{bias} is set to -1.3 Volts. Assume this PNP transistor has a beta of 100 and a emitter-base forward voltage drop of 0.7 Volts:



- $I_C =$
- $I_B =$
- $V_{out} =$

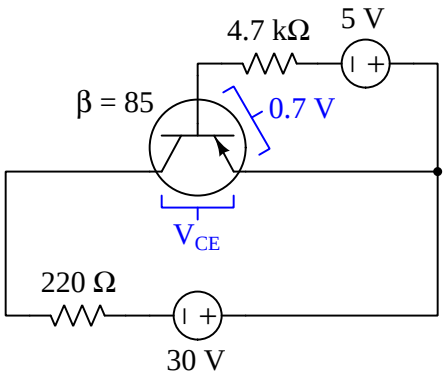
Challenges

- What is it about the transistor that establishes 0.7 Volts as the default V_{BE}?
- Explain how annotation of this schematic diagram is helpful to the analysis of the circuit.
- How would these values change if the DC power supply voltage were increased and the potentiometer re-adjusted to maintain the same -1.3 Volt V_{bias}?

7.2.4 BJT circuit calculations

Circuit #1

Calculate the following parameters in this transistor circuit:

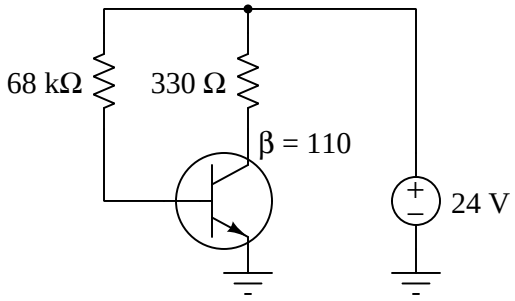


- V_{CE} =
- I_B =
- I_C =
- I_E =
- P_Q =
- V_{CB} =

Also, identify how some of the foundational concepts you've studied apply to this circuit: **behavior of sources vs. loads**, **Ohm's Law**, **Joule's Law**, **Kirchhoff's Voltage Law**, **Kirchhoff's Current Law**, **Beta**. Feel free to include any other foundational concepts not listed here.

Circuit #2

Calculate the following parameters in this transistor circuit:

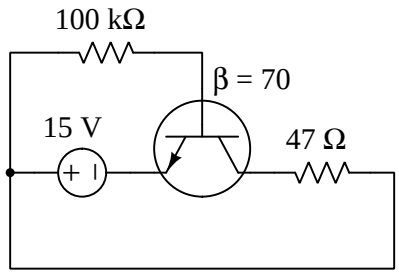


- $V_{CE} =$
- $I_B =$
- $I_C =$
- $I_E =$
- $P_Q =$
- $V_{CB} =$

Also, identify how some of the foundational concepts you've studied apply to this circuit: **behavior of sources vs. loads**, **Ohm's Law**, **Joule's Law**, **Kirchhoff's Voltage Law**, **Kirchhoff's Current Law**, **Beta**. Feel free to include any other foundational concepts not listed here.

Circuit #3

Calculate the following parameters in this transistor circuit:

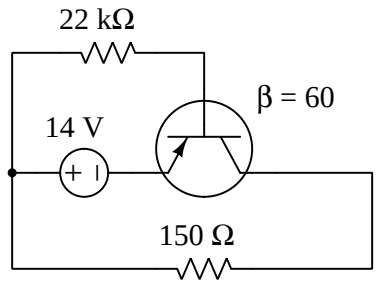


- $V_{CE} =$
- $I_B =$
- $I_C =$
- $I_E =$
- $P_Q =$
- $V_{CB} =$

Also, identify how some of the foundational concepts you've studied apply to this circuit: **behavior of sources vs. loads**, **Ohm's Law**, **Joule's Law**, **Kirchhoff's Voltage Law**, **Kirchhoff's Current Law**, **Beta**. Feel free to include any other foundational concepts not listed here.

Circuit #4

Calculate the following parameters in this transistor circuit:

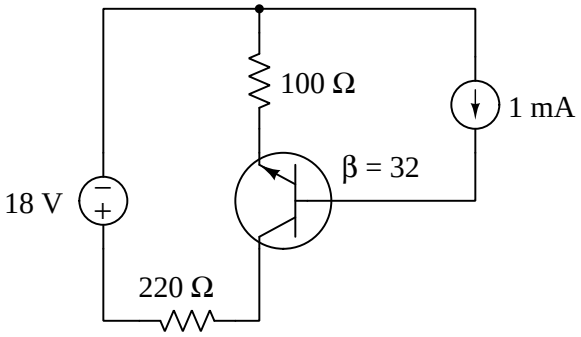


- $V_{CE} =$
- $I_B =$
- $I_C =$
- $I_E =$
- $P_Q =$
- $V_{CB} =$

Also, identify how some of the foundational concepts you've studied apply to this circuit: **behavior of sources vs. loads**, **Ohm's Law**, **Joule's Law**, **Kirchhoff's Voltage Law**, **Kirchhoff's Current Law**, **Beta**. Feel free to include any other foundational concepts not listed here.

Circuit #5

Calculate the following parameters in this transistor circuit:



- $V_{CE} =$
- $I_B =$
- $I_C =$
- $I_E =$
- $P_Q =$
- $V_{CB} =$

Also, identify how some of the foundational concepts you've studied apply to this circuit: **behavior of sources vs. loads**, **Ohm's Law**, **Joule's Law**, **Kirchhoff's Voltage Law**, **Kirchhoff's Current Law**, **Beta**. Feel free to include any other foundational concepts not listed here.

Challenges

- What will be the effects if beta were to increase?

7.3 Diagnostic reasoning

These questions are designed to stimulate your deductive and inductive thinking, where you must apply general principles to specific scenarios (deductive) and also derive conclusions about the failed circuit from specific details (inductive). In a Socratic discussion with your instructor, the goal is for these questions to reinforce your recall and use of general circuit principles and also challenge your ability to integrate multiple symptoms into a sensible explanation of what's wrong in a circuit. Your instructor may also pose additional questions based on those assigned, in order to further challenge and sharpen your diagnostic abilities.

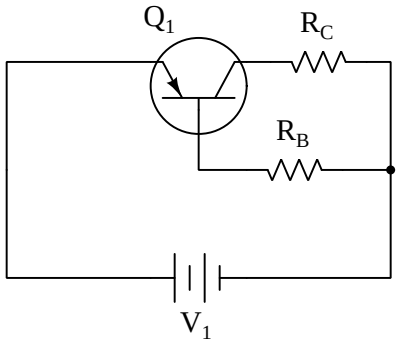
As always, your goal is to fully *explain* your analysis of each problem. Simply obtaining a correct answer is not good enough – you must also demonstrate sound reasoning in order to successfully complete the assignment. Your instructor's responsibility is to probe and challenge your understanding of the relevant principles and analytical processes in order to ensure you have a strong foundation upon which to build further understanding.

You will note a conspicuous lack of answers given for these diagnostic questions. Unlike standard textbooks where answers to every other question are given somewhere toward the back of the book, here in these learning modules students must rely on other means to check their work. The best way by far is to debate the answers with fellow students and also with the instructor during the Socratic dialogue sessions intended to be used with these learning modules. Reasoning through challenging questions with other people is an excellent tool for developing strong reasoning skills.

Another means of checking your diagnostic answers, where applicable, is to use circuit simulation software to explore the effects of faults placed in circuits. For example, if one of these diagnostic questions requires that you predict the effect of an open or a short in a circuit, you may check the validity of your work by simulating that same fault (substituting a very high resistance in place of that component for an open, and substituting a very low resistance for a short) within software and seeing if the results agree.

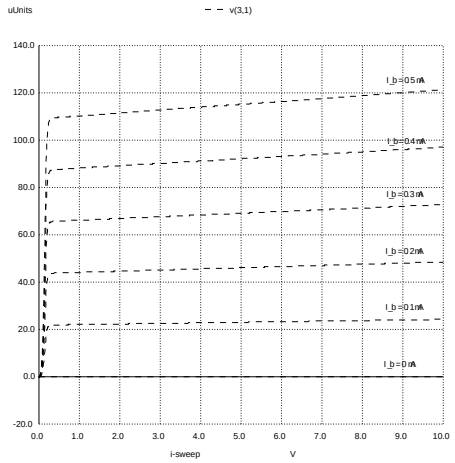
7.3.1 Faults in test circuit

Predict how all three transistor currents (I_B, I_C, and I_E) will be affected as a result of the following faults. Consider each fault independently (i.e. one at a time, no coincidental faults):



- Base resistor R_B fails open:
- Collector resistor R_C fails open:
- Solder bridge (short) past base resistor R_B:
- Solder bridge (short) past collector resistor R_C:

Additionally, determine what each of these faults would do to the operating point of a BJT as marked on a family of characteristic curves shown below. In this graph the horizontal axis represents collector-emitter voltage (V_{CE}), the vertical axis represents collector current (I_C), and each one of the curves is labeled with a fixed base current value (I_B):



- Base resistor R_B fails open:
- Collector resistor R_C fails open:
- Solder bridge (short) past base resistor R_B:
- Solder bridge (short) past collector resistor R_C:

Challenges

- Which current in this (healthy) circuit is the greatest?
- Which current in this (healthy) circuit is the least?
- Modify this circuit to use an NPN transistor rather than a PNP.

Chapter 8

Projects and Experiments

The following project and experiment descriptions outline things you can build to help you understand circuits. With any real-world project or experiment there exists the potential for physical harm. *Electricity can be very dangerous in certain circumstances, and you should follow proper safety precautions at all times!*

8.1 Recommended practices

This section outlines some recommended practices for all circuits you design and construct.

8.1.1 Safety first!

Electricity, when passed through the human body, causes uncomfortable sensations and in large enough measures[1] will cause muscles to involuntarily contract. The overriding of your nervous system by the passage of electrical current through your body is particularly dangerous in regard to your heart, which is a vital muscle. Very large amounts of current can produce serious internal burns in addition to all the other effects.

Cardio-pulmonary resuscitation (CPR) is the standard first-aid for any victim of electrical shock. This is a very good skill to acquire if you intend to work with others on dangerous electrical circuits. You should never perform tests or work on such circuits unless someone else is present who is proficient in CPR.

As a general rule, any voltage in excess of 30 Volts poses a definitive electric shock hazard, because beyond this level human skin does not have enough resistance to safely limit current through the body. "Live" work of any kind with circuits over 30 volts should be avoided, and if unavoidable should only be done using electrically insulated tools and other protective equipment (e.g. insulating shoes and gloves). If you are unsure of the hazards, or feel unsafe at any time, stop all work and distance yourself from the circuit!

A policy I strongly recommend for students learning about electricity is to *never come into electrical contact*[2] *with an energized conductor, no matter what the circuit's voltage*[3] *level!* Enforcing this policy may seem ridiculous when the circuit in question is powered by a single battery smaller than the palm of your hand, but it is precisely this instilled habit which will save a person from bodily harm when working with more dangerous circuits. Experience has taught me that students who learn early on to be careless with safe circuits have a tendency to be careless later with dangerous circuits!

In addition to the electrical hazards of shock and burns, the construction of projects and running of experiments often poses other hazards such as working with hand and power tools, potential

[1] Professor Charles Dalziel published a research paper in 1961 called "The Deleterious Effects of Electric Shock" detailing the results of electric shock experiments with both human and animal subjects. The threshold of perception for human subjects holding a conductor in their hand was in the range of 1 milliampere of current (less than this for alternating current, and generally less for female subjects than for male). Loss of muscular control was exhibited by half of Dalziel's subjects at less than 10 milliamperes alternating current. Extreme pain, difficulty breathing, and loss of all muscular control occurred for over 99% of his subjects at direct currents less than 100 milliamperes and alternating currents less than 30 milliamperes. In summary, it doesn't require much electric current to induce painful and even life-threatening effects in the human body! Your first and best protection against electric shock is maintaining an insulating barrier between your body and the circuit in question, such that current from that circuit will be unable to flow through your body.

[2] By "electrical contact" I mean either directly touching an energized conductor with any part of your body, or indirectly touching it through a conductive tool. The only physical contact you should ever make with an energized conductor is via an electrically insulated tool, for example a screwdriver with an electrically insulated handle, or an insulated test probe for some instrument.

[3] Another reason for consistently enforcing this policy, even on low-voltage circuits, is due to the dangers that even some low-voltage circuits harbor. A single 12 Volt automobile battery, for example, can cause a surprising amount of damage if short-circuited simply due to the high current levels (i.e. very low internal resistance) it is capable of, even though the voltage level is too low to cause a shock through the skin. Mechanics wearing metal rings, for example, are at risk from severe burns if their rings happen to short-circuit such a battery! Furthermore, even when working on circuits that are simply too low-power (low voltage and low current) to cause any bodily harm, touching them while energized can pose a threat to the circuit components themselves. In summary, it's generally wise (and *always* a good habit to build) to "power down" *any* circuit before making contact between it and your body.

contact with high temperatures, potential chemical exposure, etc. You should never proceed with a project or experiment if you are unaware of proper tool use or lack basic protective measures (e.g. personal protective equipment such as safety glasses) against such hazards.

Some other safety-related practices should be followed as well:

- All power conductors extending outward from the project must be *firmly* strain-relieved (e.g. "cord grips" used on line power cords), so that an accidental tug or drop will not compromise circuit integrity.
- All electrical connections must be sound and appropriately made (e.g. soldered wire joints rather than twisted-and-taped; terminal blocks rather than solderless breadboards for high-current or high-voltage circuits). Use "touch-safe" terminal connections with recessed metal parts to minimize risk of accidental contact.
- Always provide overcurrent protection in any circuit you build. *Always.* This may be in the form of a fuse, a circuit breaker, and/or an electronically current-limited power supply.
- Always ensure circuit conductors are rated for more current than the overcurrent protection limit. *Always.* A fuse does no good if the wire or printed circuit board trace will "blow" before it does!
- Always bond metal enclosures to Earth ground for any line-powered circuit. *Always.* Ensuring an equipotential state between the enclosure and Earth by making the enclosure electrically common with Earth ground ensures no electric shock can occur simply by one's body bridging between the Earth and the enclosure.
- Avoid building a high-energy circuit when a low-energy circuit will suffice. For example, I always recommend beginning students power their first DC resistor circuits using small batteries rather than with line-powered DC power supplies. The intrinsic energy limitations of a dry-cell battery make accidents highly unlikely.
- Use line power receptacles that are GFCI (Ground Fault Current Interrupting) to help avoid electric shock from making accidental contact with a "hot" line conductor.
- Always wear eye protection when working with tools or live systems having the potential to eject material into the air. Examples of such activities include soldering, drilling, grinding, cutting, wire stripping, working on or near energized circuits, etc.
- Always use a step-stool or stepladder to reach high places. Never stand on something not designed to support a human load.
- When in doubt, *ask an expert.* If anything even seems remotely unsafe to you, do not proceed without consulting a trusted person fully knowledgeable in electrical safety.

8.1.2 Other helpful tips

Experience has shown the following practices to be very helpful, especially when students make their own component selections, to ensure the circuits will be well-behaved:

- Avoid resistor values less than 1 kΩ or greater than 100 kΩ, unless such values are definitely necessary[4]. Resistances below 1 kΩ may draw excessive current if directly connected to a voltage source of significant magnitude, and may also complicate the task of accurately measuring current since any ammeter's non-zero resistance inserted in series with a low-value circuit resistor will significantly alter the total resistance and thereby skew the measurement. Resistances above 100 kΩ may complicate the task of measuring voltage since any voltmeter's finite resistance connected in parallel with a high-value circuit resistor will significantly alter the total resistance and thereby skew the measurement. Similarly, AC circuit impedance values should be between 1 kΩ and 100 kΩ, and for all the same reasons.

- Ensure all electrical connections are low-resistance and physically rugged. For this reason, one should avoid *compression splices* (e.g. "butt" connectors), solderless breadboards[5], and wires that are simply twisted together.

- Build your circuit with **testing** in mind. For example, provide convenient connection points for test equipment (e.g. multimeters, oscilloscopes, signal generators, logic probes).

- Design permanent projects with **maintenance** in mind. The more convenient you make maintenance tasks, the more likely they will get done.

- **Always document and save your work**. Circuits lacking schematic diagrams are more difficult to troubleshoot than documented circuits. Similarly, circuit construction is simpler when a schematic diagram precedes construction. Experimental results are easier to interpret when comprehensively recorded. Consider modern videorecording technology for this purpose where appropriate.

- **Record your steps** when troubleshooting. **Talk to yourself** when solving problems. These simple steps clarify thought and simplify identification of errors.

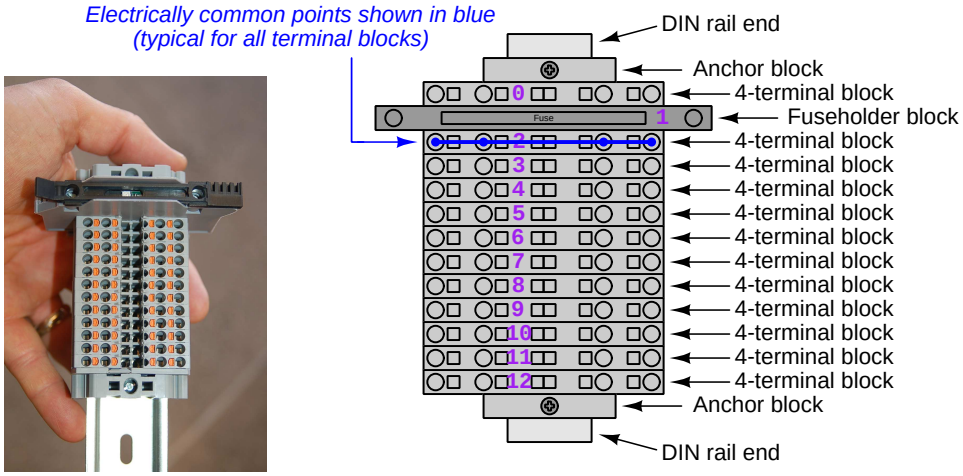
[4] An example of a necessary resistor value much less than 1 kΩ is a *shunt resistor* used to produce a small voltage drop for the purpose of sensing current in a circuit. Such shunt resistors must be low-value in order not to impose an undue load on the rest of the circuit. An example of a necessary resistor value much greater than 100 kΩ is an electrostatic *drain resistor* used to dissipate stored electric charges from body capacitance for the sake of preventing damage to sensitive semiconductor components, while also preventing a path for current that could be dangerous to the person (i.e. shock).

[5] Admittedly, solderless breadboards are very useful for constructing complex electronic circuits with many components, especially DIP-style integrated circuits (ICs), but they tend to give trouble with connection integrity after frequent use. An alternative for projects using low counts of ICs is to solder IC sockets into prototype printed circuit boards (PCBs) and run wires from the soldered pins of the IC sockets to terminal blocks where reliable temporary connections may be made.

8.1.3 Terminal blocks for circuit construction

Terminal blocks are the standard means for making electric circuit connections in industrial systems, and are also quite useful for student-built experiments, providing a robust alternative to solderless breadboards[6]. Terminal blocks provide highly reliable connections rated for significant voltage and current magnitudes, and they force the builder to think carefully about component placement which is an important mental practice. Terminal blocks designed to mount on standard 35 mm DIN rail[7] come in many types and capabilities.

Every student of electricity should build their own terminal block array consisting of several terminal blocks where each block has at least 4 connection points all electrically common to each other[8] and at least one terminal block that is a fuse holder for overcurrent protection. A pair of anchoring blocks hold all terminal blocks securely on the DIN rail, preventing them from sliding off the rail. Each of the terminals should bear a unique number starting from 0. An example is shown in the following photograph and illustration:



Screwless terminal blocks (using internal spring clips to clamp wire and component lead ends) are preferred over screw-based terminal blocks, as they reduce assembly and disassembly time, and also minimize repetitive wrist stress from twisting screwdrivers. Some screwless terminal blocks require the use of a special tool to release the spring clip, while others such as the Phoenix Contact model 3211019 blocks shown in the photograph provide buttons[9] for this task which may be pressed using the tip of any suitable tool such as a small screwdriver or multimeter test probe tip.

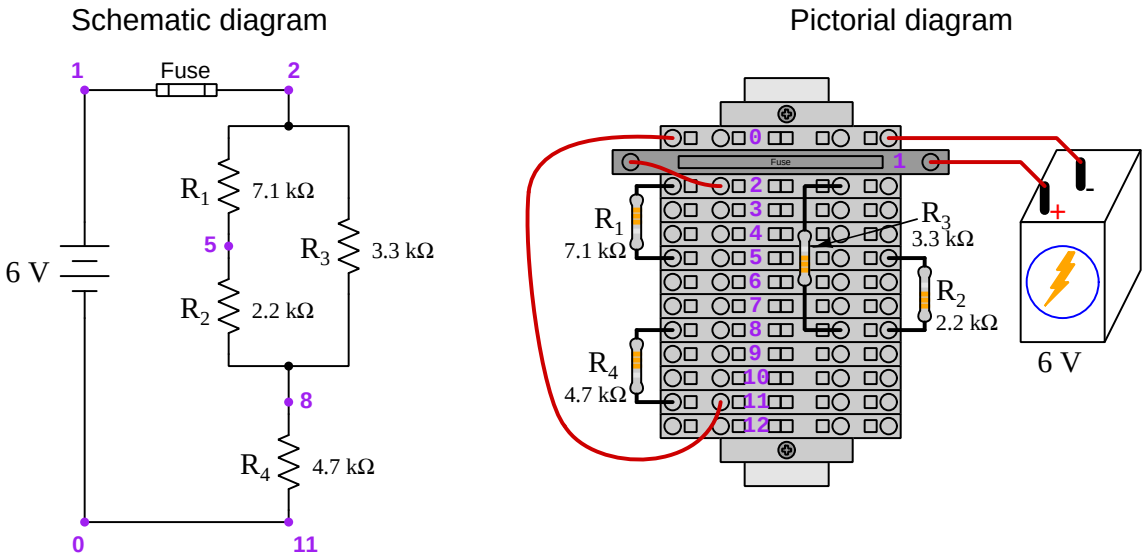
[6] Solderless breadboard are preferable for complicated electronic circuits with multiple integrated "chip" components, but for simpler circuits I find terminal blocks much more practical. An alternative to solderless breadboards for "chip" circuits is to solder chip sockets onto a PCB and then use wires to connect the socket pins to terminal blocks. This also accommodates *surface-mount* components, which solderless breadboards do not.

[7] DIN rail is a metal rail designed to serve as a mounting point for a wide range of electrical and electronic devices such as terminal blocks, fuses, circuit breakers, relay sockets, power supplies, data acquisition hardware, etc.

[8] Sometimes referred to as *equipotential*, *same-potential*, or *potential distribution* terminal blocks.

[9] The small orange-colored squares seen in the above photograph are buttons for this purpose, and may be actuated by pressing with any tool of suitable size.

The following example shows how such a terminal block array might be used to construct a series-parallel resistor circuit consisting of four resistors and a battery:



Numbering on the terminal blocks provides a very natural translation to SPICE[10] netlists, where component connections are identified by terminal number:

```
* Series-parallel resistor circuit
v1 1 0 dc 6
r1 2 5 7100
r2 5 8 2200
r3 2 8 3300
r4 8 11 4700
rfuse 1 2 0.01
rwire 0 11 0.01
.op
.end
```

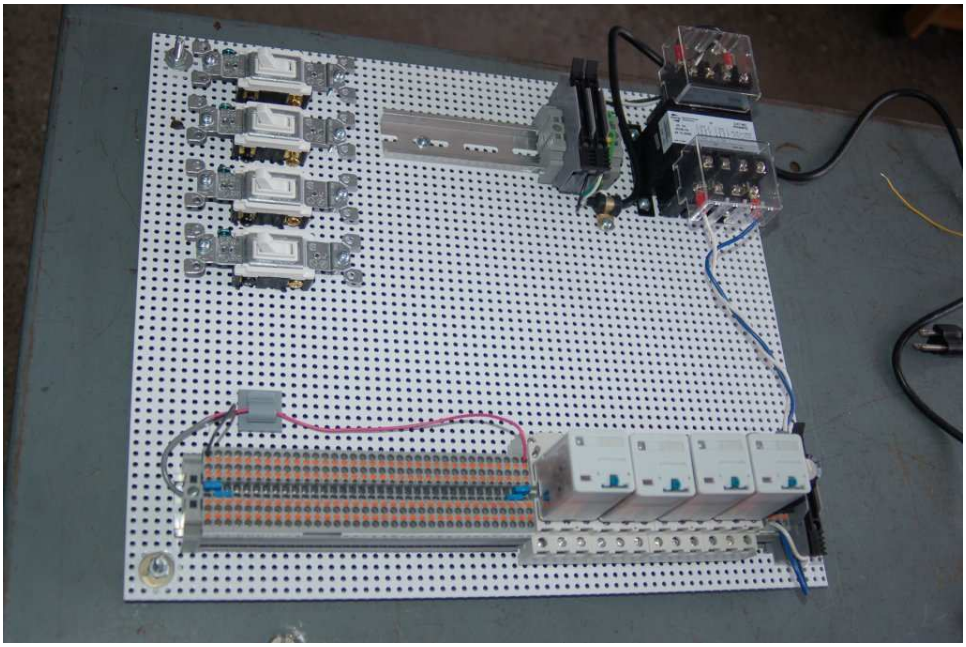
Note the use of two resistances `rfuse` and `rwire` to describe fuse resistance between terminals 1 and 2 and "jumper" wire resistance between terminals 0 and 11, respectively. Being resistances, SPICE requires a resistance value for each, and here we see they have both been set to an arbitrarily low value of 0.01 Ohm appropriate to their real-life resistance values.

Listing all components and wires along with their numbered terminals happens to be a useful documentation method for any circuit built on terminal blocks, independent of SPICE. Such a

[10] SPICE is computer software designed to analyze electrical and electronic circuits. Circuits are described for the computer in the form of *netlists* which are text files listing each component type, connection node numbers, and component values.

"wiring sequence" may be thought of as a *non-graphical description* of an electric circuit, and is exceptionally easy to follow.

An example of a more elaborate terminal block array is shown in the following photograph, with terminal blocks and "ice-cube" style electromechanical relays mounted to DIN rail, which is turn mounted to a perforated subpanel[11]. This "terminal block board" hosts an array of thirty five undedicated terminal block sections, four SPDT toggle switches, four DPDT "ice-cube" relays, a step-down control power transformer, bridge rectifier and filtering capacitor, and several fuses for overcurrent protection:



Four plastic-bottomed "feet" support the subpanel above the benchtop surface, and an unused section of DIN rail stands ready to accept other components. Safety features include electrical bonding of the AC line power cord's ground to the metal subpanel (and all metal DIN rails), mechanical strain relief for the power cord to isolate any cord tension from wire connections, clear plastic finger guards covering the transformer's screw terminals, as well as fused overcurrent protection for the 120 Volt AC line power and the transformer's 12 Volt AC output. The perforated holes happen to be on $\frac{1}{4}$ inch centers with a diameter suitable for tapping with 6-32 machine screw threads, their presence making it very easy to attach other sections of DIN rail, printed circuit boards, or specialized electrical components directly to the grounded metal subpanel. Such a "terminal block board" is an inexpensive[12] yet highly flexible means to construct physically robust circuits using industrial wiring practices.

[11] An electrical *subpanel* is a thin metal plate intended for mounting inside an electrical enclosure. Components are attached to the subpanel, and the subpanel in turn bolts inside the enclosure. Subpanels allow circuit construction outside the confines of the enclosure, which speeds assembly. In this particular usage there is no enclosure, as the subpanel is intended to be used as an open platform for the convenient construction of circuits on a benchtop by students. In essence, this is a modern version of the traditional *breadboard* which was literally a wooden board such as might be used for cutting loaves of bread, but which early electrical and electronic hobbyists used as platforms for the construction of circuits.

[12] At the time of this writing (2019) the cost to build this board is approximately $250 US dollars.

8.1.4 Conducting experiments

An *experiment* is an exploratory act, a test performed for the purpose of assessing some proposition or principle. Experiments are the foundation of the *scientific method*, a process by which careful observation helps guard against errors of speculation. All good experiments begin with an *hypothesis*, defined by the American Heritage Dictionary of the English Language as:

> An assertion subject to verification or proof, as (a) A proposition stated as a basis for argument or reasoning. (b) A premise from which a conclusion is drawn. (c) A conjecture that accounts, within a theory or ideational framework, for a set of facts and that can be used as a basis for further investigation.

Stated plainly, an hypothesis is an *educated guess* about cause and effect. The correctness of this initial guess matters little, because any well-designed experiment will reveal the truth of the matter. In fact, *incorrect* hypotheses are often the most valuable because the experiments they engender lead us to surprising discoveries. One of the beautiful aspects of science is that it is more focused on the process of *learning* than about the status of *being correct*[13]. In order for an hypothesis to be valid, it must be testable[14], which means it must be a claim possible to refute given the right data. Hypotheses impossible to critique are useless.

Once an hypothesis has been formulated, an experiment must be designed to test that hypothesis. A well-designed experiment requires careful regulation of all relevant variables, both for personal safety and for prompting the hypothesized results. If the effects of one particular variable are to be tested, the experiment must be run multiple times with different values of (only) that particular variable. The experiment set up with the "baseline" variable set is called the *control*, while the experiment set up with different value(s) is called the *test* or *experimental*.

For some hypotheses a viable alternative to a physical experiment is a *computer-simulated experiment* or even a *thought experiment*. Simulations performed on a computer test the hypothesis against the physical laws encoded within the computer simulation software, and are particularly useful for students learning new principles for which simulation software is readily available[15].

[13] Science is more about clarifying our view of the universe through a systematic process of error detection than it is about proving oneself to be right. Some scient*ists* may happen to have large egos – and this may have more to do with the ways in which large-scale scientific research is *funded* than anything else – but scientific *method* itself is devoid of ego, and if embraced as a practical philosophy is quite an effective stimulant for humility. Within the education system, scientific method is particularly valuable for helping students break free of the crippling fear of *being wrong*. So much emphasis is placed in formal education on assessing correct retention of facts that many students are fearful of saying or doing anything that might be perceived as a mistake, and of course making mistakes (i.e. having one's hypotheses disproven by experiment) is an indispensable tool for learning. Introducing science in the classroom – *real* science characterized by individuals forming actual hypotheses and testing those hypotheses by experiment – helps students become self-directed learners.

[14] This is the principle of *falsifiability*: that a scientific statement has value only insofar as it is liable to disproof given the requisite experimental evidence. Any claim that is unfalsifiable – that is, a claim which can *never* be disproven by any evidence whatsoever – could be completely wrong and we could never know it.

[15] A very pertinent example of this is learning how to analyze electric circuits using simulation software such as SPICE. A typical experimental cycle would proceed as follows: (1) Find or invent a circuit to analyze; (2) Apply your analytical knowledge to that circuit, predicting all voltages, currents, powers, etc. relevant to the concepts you are striving to master; (3) Run a simulation on that circuit, collecting "data" from the computer when complete; (4) Evaluate whether or not your hypotheses (i.e. predicted voltages, currents, etc.) agree with the computer-generated results; (5) If so, your analyses are (provisionally) correct – if not, examine your analyses and the computer simulation again to determine the source of error; (6) Repeat this process as many times as necessary until you achieve mastery.

Thought experiments are useful for detecting inconsistencies within your own understanding of some subject, rather than testing your understanding against physical reality.

Here are some general guidelines for conducting experiments:

- The clearer and more specific the hypothesis, the better. Vague or unfalsifiable hypotheses are useless because they will fit *any* experimental results, and therefore the experiment cannot teach you anything about the hypothesis.
- Collect as much data (i.e. information, measurements, sensory experiences) generated by an experiment as is practical. This includes the time and date of the experiment, too!
- *Never* discard or modify data gathered from an experiment. If you have reason to believe the data is unreliable, write notes to that effect, but never throw away data just because you think it is untrustworthy. It is quite possible that even "bad" data holds useful information, and that someone else may be able to uncover its value even if you do not.
- Prioritize *quantitative* data over *qualitative* data wherever practical. Quantitative data is more specific than qualitative, less prone to subjective interpretation on the part of the experimenter, and amenable to an arsenal of analytical methods (e.g. statistics).
- Guard against your own bias(es) by making your experimental results available to others. This allows other people to scrutinize your experimental design and collected data, for the purpose of detecting and correcting errors you may have missed. Document your experiment such that others may independently replicate it.
- Always be looking for sources of error. No physical measurement is perfect, and so it is impossible to achieve *exact* values for any variable. Quantify the amount of uncertainty (i.e. the "tolerance" of errors) whenever possible, and be sure your hypothesis does not depend on precision better than this!
- Always remember that scientific confirmation is provisional – no number of "successful" experiments will prove an hypothesis true for all time, but a single experiment can disprove it. Put into simpler terms, *truth is elusive but error is within reach.*
- Remember that scientific method is about *learning*, first and foremost. An unfortunate consequence of scientific triumph in modern society is that science is often viewed by non-practitioners as an unerring source of truth, when in fact science is an ongoing process of challenging existing ideas to probe for errors and oversights. This is why it is perfectly acceptable to have a failed hypothesis, and why the only truly failed experiment is one where nothing was learned.

The following is an example of a well-planned and executed experiment, in this case a physical experiment demonstrating Ohm's Law.

```
Planning Time/Date = 09:30 on 12 February 2019

HYPOTHESIS: the current through any resistor should be exactly proportional
to the voltage impressed across it.

PROCEDURE: connect a resistor rated 1 k Ohm and 1/4 Watt to a variable-voltage
DC power supply.  Use an ammeter in series to measure resistor current and
a voltmeter in parallel to measure resistor voltage.

RISKS AND MITIGATION: excessive power dissipation may harm the resistor and/
or pose a burn hazard, while excessive voltage poses an electric shock hazard.
30 Volts is a safe maximum voltage for laboratory practices, and according to
Joule's Law a 1000 Ohm resistor will dissipate 0.25 Watts at 15.81 Volts
(P = V^2 / R), so I will remain below 15 Volts just to be safe.

Experiment Time/Date = 10:15 on 12 February 2019

DATA COLLECTED:
  (Voltage)    (Current)         (Voltage)   (Current)
   0.000 V   =  0.000 mA          8.100    =  7.812 mA
   2.700 V   =  2.603 mA          10.00 V  =  9.643 mA
   5.400 V   =  5.206 mA          14.00 V  =  13.49 mA

Analysis Time/Date = 10:57 on 12 February 2019

ANALYSIS: current definitely increases with voltage, and although I expected
exactly one milliAmpere per Volt the actual current was usually less than
that.  The voltage/current ratios ranged from a low of 1036.87 (at 8.1 Volts)
to a high of 1037.81 (at 14 Volts), but this represents a variance of only
-0.0365% to +0.0541% from the average, indicating a very consistent
proportionality -- results consistent with Ohm's Law.

ERROR SOURCES: one major source of error is the resistor's value itself.  I
did not measure it, but simply assumed color bands of brown-black-red meant
exactly 1000 Ohms.  Based on the data I think the true resistance is closer
to 1037 Ohms.  Another possible explanation is multimeter calibration error.
However, neither explains the small positive and negative variances from the
average.  This might be due to electrical noise, a good test being to repeat
the same experiment to see if the variances are the same or different.  Noise
should generate slightly different results every time.
```

The following is an example of a well-planned and executed *virtual* experiment, in this case demonstrating Ohm's Law using a computer (SPICE) simulation.

```
Planning Time/Date = 12:32 on 14 February 2019

HYPOTHESIS: for any given resistor, the current through that resistor should be
exactly proportional to the voltage impressed across it.

PROCEDURE: write a SPICE netlist with a single DC voltage source and single
1000 Ohm resistor, then use NGSPICE version 26 to perform a "sweep" analysis
from 0 Volts to 25 Volts in 5 Volt increments.

    * SPICE circuit
    v1 1 0 dc
    r1 1 0 1000
    .dc v1 0 25 5
    .print dc v(1) i(v1)
    .end

RISKS AND MITIGATION: none.

DATA COLLECTED:
       DC transfer characteristic  Thu Feb 14 13:05:08  2019
    ------------------------------------------------------------
    Index   v-sweep         v(1)            v1#branch
    ------------------------------------------------------------
    0       0.000000e+00    0.000000e+00    0.000000e+00
    1       5.000000e+00    5.000000e+00    -5.00000e-03
    2       1.000000e+01    1.000000e+01    -1.00000e-02
    3       1.500000e+01    1.500000e+01    -1.50000e-02
    4       2.000000e+01    2.000000e+01    -2.00000e-02
    5       2.500000e+01    2.500000e+01    -2.50000e-02

Analysis Time/Date = 13:06 on 14 February 2019

ANALYSIS: perfect agreement between data and hypothesis -- current is precisely
1/1000 of the applied voltage for all values.  Anything other than perfect
agreement would have probably meant my netlist was incorrect.  The negative
current values surprised me, but it seems this is just how SPICE interprets
normal current through a DC voltage source.

ERROR SOURCES: none.
```

As gratuitous as it may seem to perform experiments on a physical law as well-established as Ohm's Law, even the examples listed previously demonstrate opportunity for real learning. In the physical experiment example, the student should identify and explain why their data does not perfectly agree with the hypothesis, and this leads them naturally to consider sources of error. In the computer-simulated experiment, the student is struck by SPICE's convention of denoting regular current through a DC voltage source as being *negative* in sign, and this is also useful knowledge for future simulations. Scientific experiments are most interesting when things *do not* go as planned!

Aside from verifying well-established physical laws, simple experiments are extremely useful as educational tools for a wide range of purposes, including:

- Component familiarization (e.g. *Which terminals of this switch connect to the NO versus NC contacts?*)
- System testing (e.g. *How heavy of a load can my AC-DC power supply source before the semiconductor components reach their thermal limits?*)
- Learning programming languages (e.g. *Let's try to set up an "up" counter function in this PLC!*)

Above all, the priority here is to inculcate the habit of hypothesizing, running experiments, and analyzing the results. This experimental cycle not only serves as an excellent method for self-directed learning, but it also works exceptionally well for troubleshooting faults in complex systems, and for these reasons should be a part of every technician's and every engineer's education.

8.1.5 Constructing projects

Designing, constructing, and testing projects is a very effective means of practical education. Within a formal educational setting, projects are generally chosen (or at least vetted) by an instructor to ensure they may be reasonably completed within the allotted time of a course or program of study, and that they sufficiently challenge the student to learn certain important principles. In a self-directed environment, projects are just as useful as a learning tool but there is some risk of unwittingly choosing a project beyond one's abilities, which can lead to frustration.

Here are some general guidelines for managing projects:

- Define your goal(s) before beginning a project: what do you wish to achieve in building it? What, exactly, should the completed project *do*?
- Analyze your project prior to construction. Document it in appropriate forms (e.g. schematic diagrams), predict its functionality, anticipate all associated risks. In other words, *plan ahead*.
- Set a reasonable budget for your project, and stay within it.
- Identify any deadlines, and set reasonable goals to meet those deadlines.
- Beware of *scope creep*: the tendency to modify the project's goals before it is complete.
- Document your progress! An easy way to do this is to use photography or videography: take photos and/or videos of your project as it progresses. Document failures as well as successes, because both are equally valuable from the perspective of learning.

8.2 Experiment: BJT terminal identification

Conduct an experiment to identify the three terminals of a BJT using nothing but an ohmmeter.

EXPERIMENT CHECKLIST:

- Prior to experimentation:
  - ☑ Write an hypothesis (i.e. a detailed description of what you expect will happen) unambiguous enough that it could be disproven given the right data.
  - ☑ Write a procedure to test the hypothesis, complete with adequate controls and documentation (e.g. schematic diagrams, programming code).
  - ☑ Identify any risks (e.g. shock hazard, component damage) and write a mitigation plan based on best practices and component ratings.
- During experimentation:
  - ☑ Safe practices followed at all times (e.g. no contact with energized circuit).
  - ☑ Correct equipment usage according to manufacturer's recommendations.
  - ☑ All data collected, ideally quantitative with full precision (i.e. no rounding).
- After each experimental run:
  - ☑ If the results fail to match the hypothesis, identify the error(s), correct the hypothesis and/or revise the procedure, and re-run the experiment.
  - ☑ Identify any uncontrolled sources of error in the experiment.
- After all experimental re-runs:
  - ☑ Save all data for future reference.
  - ☑ Write an analysis of experimental results and lessons learned.

Challenges

- Science is an *iterative* process, and for this reason is never complete. Following the results of your experiment, what would you propose for your *next* hypothesis and *next* experimental procedure? Hint: if your experiment produced any unexpected results, exploring those unexpected results is often a very good basis for the next experiment!

8.3 Experiment: demonstrate BJT amplification

Conduct an experiment to demonstrate the *amplifying* nature of a bipolar junction transistor.

EXPERIMENT CHECKLIST:

- Prior to experimentation:
  - ☑ Write an hypothesis (i.e. a detailed description of what you expect will happen) unambiguous enough that it could be disproven given the right data.
  - ☑ Write a procedure to test the hypothesis, complete with adequate controls and documentation (e.g. schematic diagrams, programming code).
  - ☑ Identify any risks (e.g. shock hazard, component damage) and write a mitigation plan based on best practices and component ratings.
- During experimentation:
  - ☑ Safe practices followed at all times (e.g. no contact with energized circuit).
  - ☑ Correct equipment usage according to manufacturer's recommendations.
  - ☑ All data collected, ideally quantitative with full precision (i.e. no rounding).
- After each experimental run:
  - ☑ If the results fail to match the hypothesis, identify the error(s), correct the hypothesis and/or revise the procedure, and re-run the experiment.
  - ☑ Identify any uncontrolled sources of error in the experiment.
- After all experimental re-runs:
  - ☑ Save all data for future reference.
  - ☑ Write an analysis of experimental results and lessons learned.

Challenges

- Science is an *iterative* process, and for this reason is never complete. Following the results of your experiment, what would you propose for your *next* hypothesis and *next* experimental procedure? Hint: if your experiment produced any unexpected results, exploring those unexpected results is often a very good basis for the next experiment!

8.4 Experiment: measure beta ratio

Conduct an experiment to quantify the beta ratio of a particular bipolar junction transistor.

EXPERIMENT CHECKLIST:

- Prior to experimentation:
  - ☑ Write an hypothesis (i.e. a detailed description of what you expect will happen) unambiguous enough that it could be disproven given the right data.
  - ☑ Write a procedure to test the hypothesis, complete with adequate controls and documentation (e.g. schematic diagrams, programming code).
  - ☑ Identify any risks (e.g. shock hazard, component damage) and write a mitigation plan based on best practices and component ratings.
- During experimentation:
  - ☑ Safe practices followed at all times (e.g. no contact with energized circuit).
  - ☑ Correct equipment usage according to manufacturer's recommendations.
  - ☑ All data collected, ideally quantitative with full precision (i.e. no rounding).
- After each experimental run:
  - ☑ If the results fail to match the hypothesis, identify the error(s), correct the hypothesis and/or revise the procedure, and re-run the experiment.
  - ☑ Identify any uncontrolled sources of error in the experiment.
- After all experimental re-runs:
  - ☑ Save all data for future reference.
  - ☑ Write an analysis of experimental results and lessons learned.

Challenges

- Science is an *iterative* process, and for this reason is never complete. Following the results of your experiment, what would you propose for your *next* hypothesis and *next* experimental procedure? Hint: if your experiment produced any unexpected results, exploring those unexpected results is often a very good basis for the next experiment!

8.5 Experiment: demonstrate the Early effect

Conduct an experiment to demonstrate the *Early effect* which is based on modulation of effective base width during transistor operation.

EXPERIMENT CHECKLIST:

- Prior to experimentation:
  - ☑ Write an hypothesis (i.e. a detailed description of what you expect will happen) unambiguous enough that it could be disproven given the right data.
  - ☑ Write a procedure to test the hypothesis, complete with adequate controls and documentation (e.g. schematic diagrams, programming code).
  - ☑ Identify any risks (e.g. shock hazard, component damage) and write a mitigation plan based on best practices and component ratings.
- During experimentation:
  - ☑ Safe practices followed at all times (e.g. no contact with energized circuit).
  - ☑ Correct equipment usage according to manufacturer's recommendations.
  - ☑ All data collected, ideally quantitative with full precision (i.e. no rounding).
- After each experimental run:
  - ☑ If the results fail to match the hypothesis, identify the error(s), correct the hypothesis and/or revise the procedure, and re-run the experiment.
  - ☑ Identify any uncontrolled sources of error in the experiment.
- After all experimental re-runs:
  - ☑ Save all data for future reference.
  - ☑ Write an analysis of experimental results and lessons learned.

Challenges

- Science is an *iterative* process, and for this reason is never complete. Following the results of your experiment, what would you propose for your *next* hypothesis and *next* experimental procedure? Hint: if your experiment produced any unexpected results, exploring those unexpected results is often a very good basis for the next experiment!

8.6 Experiment: demonstrate a BJT used as a switch

Conduct an experiment to demonstrate how a bipolar junction transistor may be used as an on-off switch for a load.

EXPERIMENT CHECKLIST:

- Prior to experimentation:
  - ☑ Write an hypothesis (i.e. a detailed description of what you expect will happen) unambiguous enough that it could be disproven given the right data.
  - ☑ Write a procedure to test the hypothesis, complete with adequate controls and documentation (e.g. schematic diagrams, programming code).
  - ☑ Identify any risks (e.g. shock hazard, component damage) and write a mitigation plan based on best practices and component ratings.
- During experimentation:
  - ☑ Safe practices followed at all times (e.g. no contact with energized circuit).
  - ☑ Correct equipment usage according to manufacturer's recommendations.
  - ☑ All data collected, ideally quantitative with full precision (i.e. no rounding).
- After each experimental run:
  - ☑ If the results fail to match the hypothesis, identify the error(s), correct the hypothesis and/or revise the procedure, and re-run the experiment.
  - ☑ Identify any uncontrolled sources of error in the experiment.
- After all experimental re-runs:
  - ☑ Save all data for future reference.
  - ☑ Write an analysis of experimental results and lessons learned.

Challenges

- Science is an *iterative* process, and for this reason is never complete. Following the results of your experiment, what would you propose for your *next* hypothesis and *next* experimental procedure? Hint: if your experiment produced any unexpected results, exploring those unexpected results is often a very good basis for the next experiment!

8.7 Experiment: measuring switching time

Conduct an experiment to measure how *quickly* a BJT is able to turn on and turn off.

EXPERIMENT CHECKLIST:

- Prior to experimentation:
  - ☑ Write an hypothesis (i.e. a detailed description of what you expect will happen) unambiguous enough that it could be disproven given the right data.
  - ☑ Write a procedure to test the hypothesis, complete with adequate controls and documentation (e.g. schematic diagrams, programming code).
  - ☑ Identify any risks (e.g. shock hazard, component damage) and write a mitigation plan based on best practices and component ratings.
- During experimentation:
  - ☑ Safe practices followed at all times (e.g. no contact with energized circuit).
  - ☑ Correct equipment usage according to manufacturer's recommendations.
  - ☑ All data collected, ideally quantitative with full precision (i.e. no rounding).
- After each experimental run:
  - ☑ If the results fail to match the hypothesis, identify the error(s), correct the hypothesis and/or revise the procedure, and re-run the experiment.
  - ☑ Identify any uncontrolled sources of error in the experiment.
- After all experimental re-runs:
  - ☑ Save all data for future reference.
  - ☑ Write an analysis of experimental results and lessons learned.

Challenges

- Science is an *iterative* process, and for this reason is never complete. Following the results of your experiment, what would you propose for your *next* hypothesis and *next* experimental procedure? Hint: if your experiment produced any unexpected results, exploring those unexpected results is often a very good basis for the next experiment!

Appendix A

Problem-Solving Strategies

The ability to solve complex problems is arguably one of the most valuable skills one can possess, and this skill is particularly important in any science-based discipline.

- Study principles, not procedures. Don't be satisfied with merely knowing how to compute solutions – learn *why* those solutions work.
- Identify what it is you need to solve, identify all relevant data, identify all units of measurement, identify any general principles or formulae linking the given information to the solution, and then identify any "missing pieces" to a solution. Annotate all diagrams with this data.
- Sketch a diagram to help visualize the problem. When building a real system, always devise a plan for that system and analyze its function *before* constructing it.
- Follow the units of measurement and meaning of every calculation. If you are ever performing mathematical calculations as part of a problem-solving procedure, and you find yourself unable to apply each and every intermediate result to some aspect of the problem, it means you don't understand what you are doing. Properly done, every mathematical result should have practical meaning for the problem, and not just be an abstract number. You should be able to identify the proper units of measurement for each and every calculated result, and show where that result fits into the problem.
- Perform "thought experiments" to explore the effects of different conditions for theoretical problems. When troubleshooting real systems, perform *diagnostic tests* rather than visually inspecting for faults, the best diagnostic test being the one giving you the most information about the nature and/or location of the fault with the fewest steps.
- Simplify the problem until the solution becomes obvious, and then use that obvious case as a model to follow in solving the more complex version of the problem.
- Check for exceptions to see if your solution is incorrect or incomplete. A good solution will work for *all* known conditions and criteria. A good example of this is the process of testing scientific hypotheses: the task of a scientist is not to find support for a new idea, but rather to *challenge* that new idea to see if it holds up under a battery of tests. The philosophical

principle of *reductio ad absurdum* (i.e. disproving a general idea by finding a specific case where it fails) is useful here.

- Work "backward" from a hypothetical solution to a new set of given conditions.
- Add quantities to problems that are qualitative in nature, because sometimes a little math helps illuminate the scenario.
- Sketch graphs illustrating how variables relate to each other. These may be quantitative (i.e. with realistic number values) or qualitative (i.e. simply showing increases and decreases).
- Treat quantitative problems as qualitative in order to discern the relative magnitudes and/or directions of change of the relevant variables. For example, try determining what happens if a certain variable were to increase or decrease before attempting to precisely calculate quantities: how will each of the dependent variables respond, by increasing, decreasing, or remaining the same as before?
- Consider limiting cases. This works especially well for qualitative problems where you need to determine which direction a variable will change. Take the given condition and magnify that condition to an extreme degree as a way of simplifying the direction of the system's response.
- Check your work. This means regularly testing your conclusions to see if they make sense. This does *not* mean repeating the same steps originally used to obtain the conclusion(s), but rather to use some other means to check validity. Simply repeating procedures often leads to *repeating the same errors* if any were made, which is why alternative paths are better.

Appendix B

Instructional philosophy

B.1 First principles of learning

- **Anyone can learn anything** given appropriate time, effort, resources, challenges, encouragement, and expectations. Dedicating time and investing effort are the student's responsibility; providing resources, challenges, and encouragement are the teacher's responsibility; high expectations are a responsibility shared by both student and teacher.

- **Transfer is not automatic.** The human mind has a natural tendency to compartmentalize information, which means the process of taking knowledge learned in one context and applying it to another usually does not come easy and therefore should never be taken for granted.

- **Learning is iterative.** The human mind rarely learns anything perfectly on the first attempt. Anticipate mistakes and plan for multiple tries to achieve full understanding, using the lessons of those mistakes as feedback to guide future attempts.

- **Information is absorbed, but understanding is created.** Facts and procedures may be memorized easily enough by repeated exposure, but the ability to reliably apply principles to novel scenarios only comes through intense personal effort. This effort is fundamentally creative in nature: casting new concepts in one's own words, running mental experiments to perceive patterns and trends, etc. These acts of making knowledge "one's own" need not be perfect in order to be effective, as the value lies in the activity and not necessarily the finished product. Depth of understanding and permanence of memory are both directly related to mental effort during the learning process – i.e. *easy learning is shallow and fleeting.*

- **Education trumps training.** There is no such thing as an entirely isolated subject, as all fields of knowledge are connected. Training is narrowly-focused and task-oriented. Education is broad-based and principle-oriented. When preparing for a life-long technical career, education beats training every time.

- **Character matters.** Poor habits are more destructive than deficits of knowledge or skill. This is especially true in collective endeavors, where a team's ability to function depends on trust between its members. Simply put, no one wants an untrustworthy person on their team. An essential component of education then, is character development.

- **People learn to be responsible by bearing responsibility.** An irresponsible person is someone who has never *had* to be responsible for anything that mattered enough to them. Just as anyone can learn anything, anyone can become responsible if the personal cost of irresponsibility becomes high enough.

- **What gets measured, gets done.** Accurate and relevant assessment of learning is key to ensuring all students learn. Therefore, it is imperative to measure what matters.

- **Failure is nothing to fear.** Every human being fails, and fails in multiple ways at multiple times. Eventual success only happens when we don't stop trying.

B.2 Proven strategies for instructors

- Assume every student is capable of learning anything they desire given the proper conditions. Treat them as capable adults by granting real responsibility and avoiding artificial incentives such as merit or demerit points.
- Create a consistent culture of high expectations across the entire program of study. Demonstrate and encourage patience, persistence, and a healthy sense of self-skepticism. Anticipate and de-stigmatize error. Teach respect for the capabilities of others as well as respect for one's own fallibility.
- Replace lecture with "inverted" instruction, where students first encounter new concepts through reading and then spend class time in Socratic dialogue with the instructor exploring those concepts and solving problems individually. There is a world of difference between observing someone solve a problem versus actually solving a problem yourself, and so the point of this form of instruction is to place students in a position where they *cannot* passively observe.
- Require students to read extensively, write about what they learn, and dialogue with you and their peers to sharpen their understanding. Apply Francis Bacon's advice that "reading maketh a full man; conference a ready man; and writing an exact man". These are complementary activities helping students expand their confidence and abilities.
- Use artificial intelligence (AI) to challenge student understanding rather than merely provide information. Find productive ways for AI to critique students' clarity of thought and of expression, for example by employing AI as a Socratic-style interlocutor or as a reviewer of students' journals. Properly applied, AI has the ability to expand student access to critical review well outside the bounds of their instructor's reach.
- Build frequent and rapid feedback into the learning process so that students know at all times how well they are learning, to identify problems early and fix them before they grow. Model the intellectual habit of self-assessing and self-correcting your own understanding (i.e. a cognitive *feedback loop*), encouraging students to do the same.
- Use "mastery" as the standard for every assessment, which means the exam or experiment or project must be done with 100% competence in order to pass. Provide students with multiple opportunity for re-tries (different versions of the assessment every time).
- Require students to devise their own hypotheses and procedures on all experiments, so that the process is truly a scientific one. Have students assess their proposed experimental procedures for risk and devise mitigations for those risks. Let nothing be pre-designed about students' experiments other than a stated task (i.e. what principle the experiment shall test) at the start and a set of demonstrable knowledge and skill objectives at the end.
- Have students build as much of their lab equipment as possible: building power sources, building test assemblies[1], and building complete working systems (no kits!). In order to provide

[1] In the program I teach, every student builds their own "Development Board" consisting of a metal chassis with DIN rail, terminal blocks, and an AC-DC power supply of their own making which functions as a portable lab environment they can use at school as well as take home.

this same "ground-up" experience for every new student, this means either previous students take their creations with them, or the systems get disassembled in preparation for the new students, or the systems grow and evolve with each new student group.

- Incorporate external accountability for you and for your students, continuously improving the curriculum and your instructional methods based on proven results. Have students regularly network with active professionals through participation in advisory committee meetings, service projects, tours, jobshadows, internships, etc. Practical suggestions include requiring students to design and build projects for external clients (e.g. community groups, businesses, different departments within the institution), and also requiring students attend all technical advisory committee meetings and to dialogue with the industry representatives at those meetings.
- Repeatedly explore difficult-to-learn concepts across multiple courses, so that students have multiple opportunities to build their understanding.
- Relate all new concepts, whenever possible, to previous concepts and to relevant physical laws. Challenge each and every student, every day, to *reason* from concept to concept and to explain the logical connections between. Challenge students to verify their conclusions by multiple approaches (e.g. double-checking their work using different methods). Ask *"Why?"* often.
- Maintain detailed records on each student's performance and share these records privately with them. These records should include academic performance as well as professionally relevant behavioral tendencies.
- Hold mandatory "check-in" meetings between all program faculty and each new student during their first term. Offer these to all other students as an option, except for any students continuing to manifest unprofessional behaviors, poor academic performance, or who have some other need for a face-to-face meeting with faculty.
- Address problems while they are small, before they grow larger. This is equally true for tutoring technical concepts as it is for helping students build professional habits.
- Build rigorous quality control into the curriculum to ensure every student masters every important concept, and that the mastery is retained over time. This includes (1) review questions added to every exam to re-assess knowledge taught in previous terms, (2) cumulative exams at the end of every term to re-assess all important concepts back to the very beginning of the program, and (3) review assessments in practical (hands-on) coursework to ensure critically-important skills were indeed taught and are still retained. What you will find by doing this is that it actually boosts retention of students by ensuring that important knowledge gets taught and is retained over long spans of time. In the absence of such quality control, student learning and retention tends to be spotty and this contributes to drop-out and failure rates later in their education.
- Finally, *never rush learning.* Education is not a race. Give your students ample time to digest complex ideas, as you continually remind yourself of just how long it took you to achieve mastery! Long-term retention and the consistently correct application of concepts are always the result of *focused effort over long periods of time* which means there are no shortcuts to learning.

B.3 Proven strategies for students

The single most important piece of advice I have for any student of any subject is to take responsibility for your own development in all areas of life including mental development. Expecting others in your life to entirely guide your own development is a recipe for disappointment. This is just as true for students enrolled in formal learning institutions as it is for auto-didacts pursuing learning entirely on their own. Learning to think in new ways is key to being able to gainfully use information, to make informed decisions about your life, and to best serve those you care about. With this in mind, I offer the following advice to students:

- **Approach all learning as valuable**. No matter what course you take, no matter who you learn from, no matter the subject, there is something useful in every learning experience. If you don't see the value of every new experience, you are not looking closely enough!
- **Continually challenge yourself**. Let other people take shortcuts and find easy answers to easy problems. The purpose of education is to stretch your mind, in order to shape it into a more powerful tool. This doesn't come by taking the path of least resistance. An excellent analogy for an empowering education is productive physical exercise: becoming stronger, more flexible, and more persistent only comes through intense personal effort.
- **Master the use of language**. This includes reading extensively, writing every day, listening closely, and speaking articulately. To a great extent language channels and empowers thought, so the better you are at wielding language the better you will be at grasping abstract concepts and articulating them not only for your benefit but for others as well.
- **Do not limit yourself to the resources given to you**. Read books that are not on the reading list. Run experiments that aren't assigned to you. Form study groups outside of class. Take an entrepreneurial approach to your own education, as though it were a business you were building for your future benefit.
- **Express and share what you learn**. Take every opportunity to teach what you have learned to others, as this will not only help them but will also strengthen your own understanding[2].
- Realize that **no one can give you understanding**, just as no one can give you physical fitness. These both must be *built*.
- **Above all, recognize that learning is hard work, and that a certain level of frustration is unavoidable**. There are times when you will struggle to grasp some of these concepts, and that struggle is a natural thing. Take heart that it will yield with persistent and varied[3] effort, and never give up! That concepts don't immediately come to you is not a sign of something wrong, but rather of something right: that you have found a worthy challenge!

[2] On a personal note, I was surprised to learn just how much my own understanding of electronics and related subjects was strengthened by becoming a teacher. When you are tasked every day with helping other people grasp complex topics, it catalyzes your own learning by giving you powerful incentives to study, to articulate your thoughts, and to reflect deeply on the process of learning.

[3] As the old saying goes, "Insanity is trying the same thing over and over again, expecting different results." If you find yourself stumped by something in the text, you should attempt a different approach. Alter the thought experiment, change the mathematical parameters, do whatever you can to see the problem in a slightly different light, and then the solution will often present itself more readily.

B.4 Design of these learning modules

"The unexamined circuit is not worth energizing" – Socrates (if he had taught electricity)

These learning modules, although useful for self-study, were designed to be used in a formal learning environment where a subject-matter expert challenges students to digest the content and exercise their critical thinking abilities in the answering of questions and in the construction and testing of working circuits. Every effort has been made to embed the following instructional and assessment philosophies within:

- The first goal of education is to enhance clear and independent thought, in order that every student reach their fullest potential in a highly complex and inter-dependent world. Robust reasoning is *always* more important than particulars of any subject matter, because its application is universal.
- Literacy is fundamental to independent learning and thought because text continues to be the most efficient way to communicate complex ideas over space and time. Those who cannot read with ease are limited in their ability to acquire knowledge and perspective.
- Articulate communication is fundamental to work that is complex and interdisciplinary.
- Faulty assumptions and poor reasoning are best corrected through challenge, not presentation. The rhetorical technique of *reductio ad absurdum* (disproving an assertion by exposing an absurdity) works well to discipline student's minds, not only to correct the problem at hand but also to learn how to detect and correct future errors.
- Important principles should be repeatedly explored and widely applied throughout a course of study, not only to reinforce their importance and help ensure their mastery, but also to showcase the interconnectedness and utility of knowledge.

These learning modules were expressly designed to be used in an "inverted" teaching environment[4] where students first read the introductory and tutorial chapters on their own, then individually attempt to answer the questions and construct working circuits according to the experiment and project guidelines. The instructor never lectures, but instead meets regularly with each individual student to review their progress, answer questions, identify misconceptions, and challenge the student to new depths of understanding through further questioning. Regular meetings between instructor and student should resemble a Socratic[5] dialogue, where questions serve as scalpels to dissect topics and expose assumptions. The student passes each module only after consistently demonstrating their ability to logically analyze and correctly apply all major concepts in each question or project/experiment. The instructor must be vigilant in probing each student's understanding to ensure they are truly *reasoning* and not just *memorizing*. This is why "Challenge" points appear throughout, as prompts for students to think deeper about topics and as starting points for instructor queries. Sometimes these challenge points require additional knowledge that hasn't been covered in the series to answer in full. This is okay, as the major purpose of the Challenges is to stimulate analysis and synthesis on the part of each student.

The instructor must possess enough mastery of the subject matter and awareness of students' reasoning to generate their own follow-up questions to practically any student response. Even completely correct answers given by the student should be challenged by the instructor for the purpose of having students practice articulating their thoughts and defending their reasoning. Conceptual errors committed by the student should be exposed and corrected not by direct instruction, but rather by reducing the errors to an absurdity[6] through well-chosen questions and thought experiments posed by the instructor. Becoming proficient at this style of instruction requires time and dedication, but the positive effects on critical thinking for both student and instructor are spectacular.

An inspection of these learning modules reveals certain unique characteristics. One of these is a bias toward thorough explanations in the tutorial chapters. Without a live instructor to explain concepts and applications to students, the text itself must fulfill this role. This philosophy results in lengthier explanations than what you might typically find in a textbook, each step of the reasoning process fully explained, including footnotes addressing common questions and concerns students raise while learning these concepts. Each tutorial seeks to not only explain each major concept in sufficient detail, but also to explain the logic of each concept and how each may be developed

[4] In a traditional teaching environment, students first encounter new information via *lecture* from an expert, and then independently apply that information via *homework*. In an "inverted" course of study, students first encounter new information via *homework*, and then independently apply that information under the scrutiny of an expert. The expert's role in lecture is to simply *explain*, but the expert's role in an inverted session is to *challenge*, *critique*, and if necessary *explain* where gaps in understanding still exist.

[5] Socrates is a figure in ancient Greek philosophy famous for his unflinching style of questioning. Although he authored no texts, he appears as a character in Plato's many writings. The essence of Socratic philosophy is to leave no question unexamined and no point of view unchallenged. While purists may argue a topic such as electric circuits is too narrow for a true Socratic-style dialogue, I would argue that the essential thought processes involved with scientific reasoning on *any* topic are not far removed from the Socratic ideal, and that students of electricity and electronics would do very well to challenge assumptions, pose thought experiments, identify fallacies, and otherwise employ the arsenal of critical thinking skills modeled by Socrates.

[6] This rhetorical technique is known by the Latin phrase *reductio ad absurdum*. The concept is to expose errors by counter-example, since only one solid counter-example is necessary to disprove a universal claim. As an example of this, consider the common misconception among beginning students of electricity that voltage cannot exist without current. One way to apply *reductio ad absurdum* to this statement is to ask how much current passes through a fully-charged battery connected to nothing (i.e. a clear example of voltage existing without current).

from "first principles". Again, this reflects the goal of developing clear and independent thought in students' minds, by showing how clear and logical thought was used to forge each concept. Students benefit from witnessing a model of clear thinking in action, and these tutorials strive to be just that.

Another feature of these learning modules is that they do not treat topics in isolation. Rather, important concepts are introduced early in the series, and appear repeatedly as stepping-stones toward other concepts in subsequent modules. This helps to avoid the "compartmentalization" of knowledge, demonstrating the inter-connectedness of concepts and simultaneously reinforcing them. Each module is fairly complete in itself, reserving the beginning of its tutorial to a review of foundational concepts.

To high standards of education,

Tony R. Kuphaldt

Appendix C

Tools used

I am indebted to the developers of many open-source software applications in the creation of these learning modules. The following is a list of these applications with some commentary on each.

You will notice a theme common to many of these applications: a bias toward *code*. Although I am by no means an expert programmer in any computer language, I understand and appreciate the flexibility offered by code-based applications where the user (you) enters commands into a plain ASCII text file, which the software then reads and processes to create the final output. Code-based computer applications are by their very nature *extensible*, while WYSIWYG (What You See Is What You Get) applications are generally limited to whatever user interface the developer makes for you.

The `GNU/Linux` computer operating system

> There is so much to be said about Linus Torvalds' `Linux` and Richard Stallman's `GNU` project. First, to credit just these two individuals is to fail to do justice to the *mob* of passionate volunteers who contributed to make this amazing software a reality. I first learned of `Linux` back in 1996, and have been using this operating system on my personal computers almost exclusively since then. It is *free*, it is completely *configurable*, and it permits the continued use of highly efficient `Unix` applications and scripting languages (e.g. shell scripts, Makefiles, `sed`, `awk`) developed over many decades. `Linux` not only provided me with a powerful computing platform, but its open design served to inspire my life's work of creating open-source educational resources.

Bram Moolenaar's `Vim` text editor

> Writing code for any code-based computer application requires a *text editor*, which may be thought of as a word processor strictly limited to outputting plain-ASCII text files. Many good text editors exist, and one's choice of text editor seems to be a deeply personal matter within the programming world. I prefer `Vim` because it operates very similarly to `vi` which is ubiquitous on `Unix`/`Linux` operating systems, and because it may be entirely operated via keyboard (i.e. no mouse required) which makes it fast to use.

Donald Knuth's TEX typesetting system

Developed in the late 1970's and early 1980's by computer scientist extraordinaire Donald Knuth to typeset his multi-volume magnum opus *The Art of Computer Programming*, this software allows the production of formatted text for screen-viewing or paper printing, all by writing plain-text code to describe how the formatted text is supposed to appear. TEX is not just a markup language for documents, but it is also a Turing-complete programming language in and of itself, allowing useful algorithms to be created to control the production of documents. Simply put, TEX *is a programmer's approach to word processing*. Since TEX is controlled by code written in a plain-text file, this means anyone may read that plain-text file to see exactly how the document was created. This openness afforded by the code-based nature of TEX makes it relatively easy to learn how other people have created their own TEX documents. By contrast, examining a beautiful document created in a conventional WYSIWYG word processor such as Microsoft `Word` suggests nothing to the reader about *how* that document was created, or what the user might do to create something similar. As Mr. Knuth himself once quipped, conventional word processing applications should be called WYSIAYG (What You See Is *All* You Get).

Leslie Lamport's LATEX extensions to TEX

Like all true programming languages, TEX is inherently extensible. So, years after the release of TEX to the public, Leslie Lamport decided to create a massive extension allowing easier compilation of book-length documents. The result was LATEX, which is the markup language used to create all ModEL module documents. You could say that TEX is to LATEX as `C` is to `C++`. This means it is permissible to use any and all TEX commands within LATEX source code, and it all still works. Some of the features offered by LATEX that would be challenging to implement in TEX include automatic index and table-of-content creation.

Tim Edwards' `Xcircuit` drafting program

This wonderful program is what I use to create all the schematic diagrams and illustrations (but not photographic images or mathematical plots) throughout the ModEL project. It natively outputs PostScript format which is a true vector graphic format (this is why the images do not pixellate when you zoom in for a closer view), and it is so simple to use that I have never had to read the manual! Object libraries are easy to create for `Xcircuit`, being plain-text files using PostScript programming conventions. Over the years I have collected a large set of object libraries useful for drawing electrical and electronic schematics, pictorial diagrams, and other technical illustrations.

`Gimp` graphic image manipulation program

Essentially an open-source clone of Adobe's `PhotoShop`, I use `Gimp` to resize, crop, and convert file formats for all of the photographic images appearing in the ModEL modules. Although `Gimp` does offer its own scripting language (called `Script-Fu`), I have never had occasion to use it. Thus, my utilization of `Gimp` to merely crop, resize, and convert graphic images is akin to using a sword to slice bread.

`SPICE` circuit simulation program

`SPICE` is to circuit analysis as TEX is to document creation: it is a form of markup language designed to describe a certain object to be processed in plain-ASCII text. When the plain-text "source file" is compiled by the software, it outputs the final result. More modern circuit analysis tools certainly exist, but I prefer `SPICE` for the following reasons: it is *free*, it is *fast*, it is *reliable*, and it is a fantastic tool for *teaching* students of electricity and electronics how to write simple code. I happen to use rather old versions of `SPICE`, version 2g6 being my "go to" application when I only require text-based output. `NGSPICE` (version 26), which is based on Berkeley `SPICE` version 3f5, is used when I require graphical output for such things as time-domain waveforms and Bode plots. In all `SPICE` example netlists I strive to use coding conventions compatible with all `SPICE` versions.

Andrew D. Hwang's `ePiX` mathematical visualization programming library

This amazing project is a `C++` library you may link to any `C/C++` code for the purpose of generating PostScript graphic images of mathematical functions. As a completely free and open-source project, it does all the plotting I would otherwise use a Computer Algebra System (CAS) such as `Mathematica` or `Maple` to do. It should be said that `ePiX` is *not* a Computer Algebra System like `Mathematica` or `Maple`, but merely a mathematical *visualization* tool. In other words, it won't determine integrals for you (you'll have to implement that in your own `C/C++` code!), but it can graph the results, and it does so beautifully. What I really admire about `ePiX` is that it is a `C++` programming library, which means it builds on the existing power and toolset available with that programming language. Mr. Hwang could have probably developed his own stand-alone application for mathematical plotting, but by creating a `C++` library to do the same thing he accomplished something much greater.

`gnuplot` mathematical visualization software

Another open-source tool for mathematical visualization is `gnuplot`. Interestingly, this tool is *not* part of Richard Stallman's GNU project, its name being a coincidence. For this reason the authors prefer "gnu" *not* be capitalized at all to avoid confusion. This is a much "lighter-weight" alternative to a spreadsheet for plotting tabular data, and the fact that it easily outputs directly to an X11 console or a file in a number of different graphical formats (including PostScript) is very helpful. I typically set my `gnuplot` output format to default (X11 on my Linux PC) for quick viewing while I'm developing a visualization, then switch to PostScript file export once the visual is ready to include in the document(s) I'm writing. As with my use of `Gimp` to do rudimentary image editing, my use of `gnuplot` only scratches the surface of its capabilities, but the important points are that it's *free* and that it *works well*.

`Python` programming language

Both Python and C++ find extensive use in these modules as instructional aids and exercises, but I'm listing Python here as a *tool* for myself because I use it almost daily as a *calculator*. If you open a Python interpreter console and type `from math import *` you can type mathematical expressions and have it return results just as you would on a hand calculator. Complex-number (i.e. *phasor*) arithmetic is similarly supported if you include the complex-math library (`from cmath import *`). Examples of this are shown in the Programming References chapter (if included) in each module. Of course, being a fully-featured programming language, Python also supports conditionals, loops, and other structures useful for calculation of quantities. Also, running in a console environment where all entries and returned values show as text in a chronologically-ordered list makes it easy to copy-and-paste those calculations to document exactly how they were performed.

Appendix D

Creative Commons License

Creative Commons Attribution 4.0 International Public License

By exercising the Licensed Rights (defined below), You accept and agree to be bound by the terms and conditions of this Creative Commons Attribution 4.0 International Public License ("Public License"). To the extent this Public License may be interpreted as a contract, You are granted the Licensed Rights in consideration of Your acceptance of these terms and conditions, and the Licensor grants You such rights in consideration of benefits the Licensor receives from making the Licensed Material available under these terms and conditions.

Section 1 – Definitions.

a. **Adapted Material** means material subject to Copyright and Similar Rights that is derived from or based upon the Licensed Material and in which the Licensed Material is translated, altered, arranged, transformed, or otherwise modified in a manner requiring permission under the Copyright and Similar Rights held by the Licensor. For purposes of this Public License, where the Licensed Material is a musical work, performance, or sound recording, Adapted Material is always produced where the Licensed Material is synched in timed relation with a moving image.

b. **Adapter's License** means the license You apply to Your Copyright and Similar Rights in Your contributions to Adapted Material in accordance with the terms and conditions of this Public License.

c. **Copyright and Similar Rights** means copyright and/or similar rights closely related to copyright including, without limitation, performance, broadcast, sound recording, and Sui Generis Database Rights, without regard to how the rights are labeled or categorized. For purposes of this Public License, the rights specified in Section 2(b)(1)-(2) are not Copyright and Similar Rights.

d. **Effective Technological Measures** means those measures that, in the absence of proper authority, may not be circumvented under laws fulfilling obligations under Article 11 of the WIPO Copyright Treaty adopted on December 20, 1996, and/or similar international agreements.

e. **Exceptions and Limitations** means fair use, fair dealing, and/or any other exception or

limitation to Copyright and Similar Rights that applies to Your use of the Licensed Material.

f. **Licensed Material** means the artistic or literary work, database, or other material to which the Licensor applied this Public License.

g. **Licensed Rights** means the rights granted to You subject to the terms and conditions of this Public License, which are limited to all Copyright and Similar Rights that apply to Your use of the Licensed Material and that the Licensor has authority to license.

h. **Licensor** means the individual(s) or entity(ies) granting rights under this Public License.

i. **Share** means to provide material to the public by any means or process that requires permission under the Licensed Rights, such as reproduction, public display, public performance, distribution, dissemination, communication, or importation, and to make material available to the public including in ways that members of the public may access the material from a place and at a time individually chosen by them.

j. **Sui Generis Database Rights** means rights other than copyright resulting from Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, as amended and/or succeeded, as well as other essentially equivalent rights anywhere in the world.

k. **You** means the individual or entity exercising the Licensed Rights under this Public License. **Your** has a corresponding meaning.

Section 2 – Scope.

a. License grant.

1. Subject to the terms and conditions of this Public License, the Licensor hereby grants You a worldwide, royalty-free, non-sublicensable, non-exclusive, irrevocable license to exercise the Licensed Rights in the Licensed Material to:

A. reproduce and Share the Licensed Material, in whole or in part; and

B. produce, reproduce, and Share Adapted Material.

2. Exceptions and Limitations. For the avoidance of doubt, where Exceptions and Limitations apply to Your use, this Public License does not apply, and You do not need to comply with its terms and conditions.

3. Term. The term of this Public License is specified in Section 6(a).

4. Media and formats; technical modifications allowed. The Licensor authorizes You to exercise the Licensed Rights in all media and formats whether now known or hereafter created, and to make technical modifications necessary to do so. The Licensor waives and/or agrees not to assert any right or authority to forbid You from making technical modifications necessary to exercise the Licensed Rights, including technical modifications necessary to circumvent Effective Technological Measures.

For purposes of this Public License, simply making modifications authorized by this Section 2(a)(4) never produces Adapted Material.

5. Downstream recipients.

A. Offer from the Licensor – Licensed Material. Every recipient of the Licensed Material automatically receives an offer from the Licensor to exercise the Licensed Rights under the terms and conditions of this Public License.

B. No downstream restrictions. You may not offer or impose any additional or different terms or conditions on, or apply any Effective Technological Measures to, the Licensed Material if doing so restricts exercise of the Licensed Rights by any recipient of the Licensed Material.

6. No endorsement. Nothing in this Public License constitutes or may be construed as permission to assert or imply that You are, or that Your use of the Licensed Material is, connected with, or sponsored, endorsed, or granted official status by, the Licensor or others designated to receive attribution as provided in Section 3(a)(1)(A)(i).

b. Other rights.

1. Moral rights, such as the right of integrity, are not licensed under this Public License, nor are publicity, privacy, and/or other similar personality rights; however, to the extent possible, the Licensor waives and/or agrees not to assert any such rights held by the Licensor to the limited extent necessary to allow You to exercise the Licensed Rights, but not otherwise.

2. Patent and trademark rights are not licensed under this Public License.

3. To the extent possible, the Licensor waives any right to collect royalties from You for the exercise of the Licensed Rights, whether directly or through a collecting society under any voluntary or waivable statutory or compulsory licensing scheme. In all other cases the Licensor expressly reserves any right to collect such royalties.

Section 3 – License Conditions.

Your exercise of the Licensed Rights is expressly made subject to the following conditions.

a. Attribution.

1. If You Share the Licensed Material (including in modified form), You must:

A. retain the following if it is supplied by the Licensor with the Licensed Material:

i. identification of the creator(s) of the Licensed Material and any others designated to receive attribution, in any reasonable manner requested by the Licensor (including by pseudonym if designated);

ii. a copyright notice;

iii. a notice that refers to this Public License;

iv. a notice that refers to the disclaimer of warranties;

v. a URI or hyperlink to the Licensed Material to the extent reasonably practicable;

B. indicate if You modified the Licensed Material and retain an indication of any previous modifications; and

C. indicate the Licensed Material is licensed under this Public License, and include the text of, or the URI or hyperlink to, this Public License.

2. You may satisfy the conditions in Section 3(a)(1) in any reasonable manner based on the medium, means, and context in which You Share the Licensed Material. For example, it may be reasonable to satisfy the conditions by providing a URI or hyperlink to a resource that includes the required information.

3. If requested by the Licensor, You must remove any of the information required by Section 3(a)(1)(A) to the extent reasonably practicable.

4. If You Share Adapted Material You produce, the Adapter's License You apply must not prevent recipients of the Adapted Material from complying with this Public License.

Section 4 – Sui Generis Database Rights.

Where the Licensed Rights include Sui Generis Database Rights that apply to Your use of the Licensed Material:

a. for the avoidance of doubt, Section 2(a)(1) grants You the right to extract, reuse, reproduce, and Share all or a substantial portion of the contents of the database;

b. if You include all or a substantial portion of the database contents in a database in which You have Sui Generis Database Rights, then the database in which You have Sui Generis Database Rights (but not its individual contents) is Adapted Material; and

c. You must comply with the conditions in Section 3(a) if You Share all or a substantial portion of the contents of the database.

For the avoidance of doubt, this Section 4 supplements and does not replace Your obligations under this Public License where the Licensed Rights include other Copyright and Similar Rights.

Section 5 – Disclaimer of Warranties and Limitation of Liability.

a. Unless otherwise separately undertaken by the Licensor, to the extent possible, the Licensor offers the Licensed Material as-is and as-available, and makes no representations or warranties of any kind concerning the Licensed Material, whether express, implied, statutory, or other. This includes, without limitation, warranties of title, merchantability, fitness for a particular purpose, non-infringement, absence of latent or other defects, accuracy, or the presence or absence of errors,

whether or not known or discoverable. Where disclaimers of warranties are not allowed in full or in part, this disclaimer may not apply to You.

b. To the extent possible, in no event will the Licensor be liable to You on any legal theory (including, without limitation, negligence) or otherwise for any direct, special, indirect, incidental, consequential, punitive, exemplary, or other losses, costs, expenses, or damages arising out of this Public License or use of the Licensed Material, even if the Licensor has been advised of the possibility of such losses, costs, expenses, or damages. Where a limitation of liability is not allowed in full or in part, this limitation may not apply to You.

c. The disclaimer of warranties and limitation of liability provided above shall be interpreted in a manner that, to the extent possible, most closely approximates an absolute disclaimer and waiver of all liability.

Section 6 – Term and Termination.

a. This Public License applies for the term of the Copyright and Similar Rights licensed here. However, if You fail to comply with this Public License, then Your rights under this Public License terminate automatically.

b. Where Your right to use the Licensed Material has terminated under Section 6(a), it reinstates:

1. automatically as of the date the violation is cured, provided it is cured within 30 days of Your discovery of the violation; or

2. upon express reinstatement by the Licensor.

For the avoidance of doubt, this Section 6(b) does not affect any right the Licensor may have to seek remedies for Your violations of this Public License.

c. For the avoidance of doubt, the Licensor may also offer the Licensed Material under separate terms or conditions or stop distributing the Licensed Material at any time; however, doing so will not terminate this Public License.

d. Sections 1, 5, 6, 7, and 8 survive termination of this Public License.

Section 7 – Other Terms and Conditions.

a. The Licensor shall not be bound by any additional or different terms or conditions communicated by You unless expressly agreed.

b. Any arrangements, understandings, or agreements regarding the Licensed Material not stated herein are separate from and independent of the terms and conditions of this Public License.

Section 8 – Interpretation.

a. For the avoidance of doubt, this Public License does not, and shall not be interpreted to, reduce, limit, restrict, or impose conditions on any use of the Licensed Material that could lawfully

be made without permission under this Public License.

b. To the extent possible, if any provision of this Public License is deemed unenforceable, it shall be automatically reformed to the minimum extent necessary to make it enforceable. If the provision cannot be reformed, it shall be severed from this Public License without affecting the enforceability of the remaining terms and conditions.

c. No term or condition of this Public License will be waived and no failure to comply consented to unless expressly agreed to by the Licensor.

d. Nothing in this Public License constitutes or may be interpreted as a limitation upon, or waiver of, any privileges and immunities that apply to the Licensor or You, including from the legal processes of any jurisdiction or authority.

Creative Commons is not a party to its public licenses. Notwithstanding, Creative Commons may elect to apply one of its public licenses to material it publishes and in those instances will be considered the "Licensor." Except for the limited purpose of indicating that material is shared under a Creative Commons public license or as otherwise permitted by the Creative Commons policies published at `creativecommons.org/policies`, Creative Commons does not authorize the use of the trademark "Creative Commons" or any other trademark or logo of Creative Commons without its prior written consent including, without limitation, in connection with any unauthorized modifications to any of its public licenses or any other arrangements, understandings, or agreements concerning use of licensed material. For the avoidance of doubt, this paragraph does not form part of the public licenses.

Creative Commons may be contacted at `creativecommons.org`.

Appendix E

References

Bardeen, John and Brattain, Walter, *US Patent 2,524,035*, "Three-Electrode Circuit Element Utilizing Semiconductive Materials", application 17 June 1948, patent granted 3 October 1950.

Corning, John J., *Transistor Circuit Analysis and Design*, Prentice-Hall, Inc., New Jersey, 1965.

Dr. James Early, an oral history conducted in 1999 by David Hochfelder, IEEE History Center, Hoboken, NJ, USA.

Maigret, Robert and King, Geoffrey, *US Patent 3,829,709*, "Supply Reversal Protection Circuit", application 31 August 1973, patent granted 13 August 1974.

Pierret, Robert F. *Semiconductor Device Fundamentals*, Addison-Wesley Publishing Company, 1996.

Shockley, William, "Transistor Technology Evokes New Physics", Nobel prize lecture, 11 December 1956.

Appendix F

Version history

This is a list showing all significant additions, corrections, and other edits made to this learning module. Each entry is referenced by calendar date in reverse chronological order (newest version first), which appears on the front cover of every learning module for easy reference. Any contributors to this open-source document are listed here as well.

27 September 2026 – minor edits to the Tutorial chapter, and also added some new Challenge questions to the "Relative current magnitudes" Conceptual Reasoning question.

8 April 2026 – typographical error corrections courtesy of Wyatt Gleason.

26-27 October 2025 – added and edited some instructor notes, as well as some prompts within the "BJT circuit calculations" Quantitative Reasoning question. Also corrected some typographical errors courtesy of Wyatt Gleason.

13-22 October 2025 – added a new Tutorial section showing how planar transistor construction works, and also corrected a typographical error.

4 June 2025 – added a new Case Tutorial section showing how to test a bipolar junction transistor using a multimeter's "diode-check" mode.

1 April 2025 – added some comments about operating modes for transistors (e.g. active, cutoff, saturated) to the Introduction chapter. Also elaborated on rectifying versus non-rectifying metal-to-semiconductor contacts in the Historical References chapter.

30 March 2025 – minor edits made to the Full Tutorial chapter for clarity. Also added a sub-question to the "Controlling and controlled currents" Conceptual Reasoning question. Also added another bullet-point to the "Challenging concepts related to bipolar junction transistors" section of the Introduction chapter.

14 March 2025 – added a new Historical References section showing a photograph of an early BJT manufactured by Bell Laboratories.

2 November 2024 – added a new Animations chapter with an animation showing the injection of charge carriers in an NPN transistor with applied base-emitter voltage.

28 October 2024 – added a characteristic curve graph for a BJT to the "Faults in test circuit" Diagnostic Reasoning question, along with prompting questions and instructor-version answers.

11 October 2024 – divided the Introduction chapter into sections, one with recommendations for students, one with a listing of challenging concepts, and one with recommendations for instructors.

10 April 2024 – added another page to the Tutorial discussing beta.

1 April 2024 – made minor edits to the sand/wind analogy for charge carrier drift in conductors versus depletion regions and also corrected a typographical error.

4-17 August 2023 – added more example circuits to the "BJT circuit calculations" Quantitative Reasoning question, and re-wrote portions of the Simplified Tutorial (as well as image_5289) for clearer reading.

28 November 2022 – placed questions at the top of the itemized list in the Introduction chapter prompting students to devise experiments related to the tutorial content.

31 October 2022 – added some Challenge questions to the "DC bias calculations" Quantitative Reasoning question, and also added some explanatory text to the Historical Reference section on "John Bardeen's and Walter Brattain's transistor patent".

27 October 2022 – minor edits to the Full Tutorial, elaborating on how beta (β) establishes a *maximum* collector current due to the action of injected charge carriers. Also added some more questions to the Introduction chapter.

10 August 2022 – added a summary page to the end of the "Injected charge carriers" section of the Full Tutorial showing proper BJT biasing. Also, renamed the "Transistor characteristics" section to "BJT characteristics".

5 August 2022 – fixed a labeling error in image_5290 where PNP transistors were wrongly labeled as being NPN.

4-5 April 2022 – added more questions to the Introduction chapter, made minor edits to the Full Tutorial text, and added some Challenge questions.

11 November 2021 – added questions.

8 November 2021 – divided the Full Tutorial into sections.

21 July 2021 – added some Foundational concept application to the "BJT circuit calculations" Quantitative Reasoning question.

19 May 2021 – minor edits to the Full Tutorial.

11 May 2021 – added a Simplified Tutorial.

9 May 2021 – commented out or deleted empty chapters.

27 April 2021 – minor edit to the Tutorial.

23 April 2021 – minor addition to the "BJT circuit calculations" Quantitative Reasoning question.

11 January 2021 – edited capitalization on Historical References section (the Bardeen/Brattain patent).

3 November 2020 – minor additions to instructor notes for the "DC bias calculations" question.

17 September 2020 – minor additions to the Introduction chapter.

31 August 2020 – significantly edited the Introduction chapter to make it more suitable as a pre-study guide and to provide cues useful to instructors leading "inverted" teaching sessions.

24 May 2020 – added more content to the Tutorial on power dissipation.

28 April 2020 – capitalized all instances of the word "Volt".

23 April 2020 – added to the Tutorial section, showing how beta may be used to calculate any transistor terminal current given any other.

8 April 2020 – added Case Tutorial section.

29 March 2020 – added Quantitative Reasoning question on simple transistor circuit calculations.

27 January 2020 – added Foundational Concepts to the list in the Conceptual Reasoning section.

23 January 2020 – added Foundational Concepts to the list in the Conceptual Reasoning section.

1 December 2019 – added Animation showing how characteristic curves are created for a BJT.

20 October 2019 – added footnote calling beta the "current gain" of the transistor.

2 October 2019 – added reference to John Corning's excellent text, *Transistor Circuit Analysis and Design*. Also added more Experiments.

1 October 2019 – added an Historical References chapter, and included within it some US patents. Also added two Experiments.

29 September 2019 – took some of the existing Tutorial and packaged it into a review file (review_07.latex) for inclusion in other transistor modules (e.g. field-effect transistors), since this information is a common prelude to any type of amplifying component. Also made minor clarifying edits to the Introduction and Tutorial chapters.

13 January 2019 – renamed Simplified Tutorial to Tutorial, since I haven't written a Full Tutorial yet.

16 June 2019 – minor edits to diagnostic questions, replacing "no multiple faults" with "no coincidental faults".

14 June 2019 – numerous minor edits to the Simplified Tutorial, including some typographical error corrections.

11 June 2019 – added more content to the Simplified Tutorial.

8 June 2019 – document first created.

Index

Active device, 23
Adding quantities to a qualitative problem, 154
Alloying, 26
Amplification, 22
Analog, 24
Annotating diagrams, 153

Bias voltage, 27
Breadboard, solderless, 136, 137
Breadboard, traditional, 140

Cardio-Pulmonary Resuscitation, 134
Checking for exceptions, 154
Checking your work, 154
Code, computer, 163
Conductor, 26
Conservation of Electric Charge, 36
Conservation of Energy, 22, 38
CPR, 134
Current, 38

Dalziel, Charles, 134
Darlington pair, 13
Depletion region, 26, 27
Dimensional analysis, 38, 153
DIN rail, 137
Diode, 27
Diode equation, 5
DIP, 136
Discrete, 24
Doping, 26

Early effect, 35
Early voltage, 35
Early, James, 35
Edwards, Tim, 164
Electric shock, 134
Electrically common points, 135
Electron tube, 25
Enclosure, electrical, 140
Equipotential points, 135, 137
Experiment, 141
Experimental guidelines, 142

Gain, 23, 24, 35
Graph values to solve a problem, 154
Greenleaf, Cynthia, 101

Holes, 26
How to teach with these modules, 161
Hwang, Andrew D., 165

IC, 136
Identify given data, 153
Identify relevant principles, 153
Insulator, 26
Intermediate results, 153
Inverted instruction, 161

Joule's Law, 38

Kirchhoff's Current Law, 20, 31, 36
Knuth, Donald, 164

Lamport, Leslie, 164
Lateral transistor, 39
Limiting cases, 154
Low resistance contact, 47
LPNP, 39

Maxwell, James Clerk, 43
Metacognition, 106
Moolenaar, Bram, 163
Murphy, Lynn, 101

Non-rectifying contact, 47

Ohm's Law, 5, 34
Ohmic contact, 47
Open-source, 163

Passive device, 23
Photodiode, 29
Photoelectric effect, 26
Photon, 29
Potential distribution, 137
Power, 38
Problem-solving: annotate diagrams, 153
Problem-solving: check for exceptions, 154
Problem-solving: checking work, 154
Problem-solving: dimensional analysis, 153
Problem-solving: graph values, 154
Problem-solving: identify given data, 153
Problem-solving: identify relevant principles, 153
Problem-solving: interpret intermediate results, 153
Problem-solving: limiting cases, 154
Problem-solving: qualitative to quantitative, 154
Problem-solving: quantifying a qualitative problem, 37
Problem-solving: quantitative to qualitative, 154
Problem-solving: reductio ad absurdum, 154
Problem-solving: simplify the system, 153
Problem-solving: thought experiment, 29, 142, 153
Problem-solving: track units of measurement, 153
Problem-solving: visually represent the system, 153
Problem-solving: work in reverse, 154
Project management guidelines, 145

Qualitatively approaching a quantitative problem, 154
Quantifying a qualitative problem, 37

Radar, 25
Radio, 25
Reading Apprenticeship, 101
Rectifying contact, 47
Reductio ad absurdum, 154, 160, 161

Safety, electrical, 134
Schoenbach, Ruth, 101
Scientific method, 106, 141
Scope creep, 145
Semiconductor, 26
Shockley diode equation, 5
Shockley, William, 5
Shunt resistor, 32, 136
Simplifying a system, 153
Socrates, 160
Socratic dialogue, 161
Solderless breadboard, 136, 137
SPICE, 101, 142
SPICE netlist, 138
SPICE, computer analysis, 32
Stallman, Richard, 163
Subpanel, 140
Surface mount, 137

Telegraph, 25
Telephone, 25
Television, 25
Terminal block, 135–138, 140
Thermionic emission, 26
Thermionic valve, 25
Thought experiment, 29, 142, 153
Torvalds, Linus, 163
Tube, vacuum, 25

Units of measurement, 153

Vacuum tube, 25
Valve, thermionic, 25
Vertical transistor, 39
Visualizing a system, 153
VNPN, 39
Voltage, 38

Wiring sequence, 139
Work function, 26
Work in reverse to solve a problem, 154
WYSIWYG, 163, 164