Semiconductor Devices: Theory and Application

SEMICONDUCTOR DEVICES: THEORY AND APPLICATION

NSCC Edition

JAMES M. FIORE AND NSCC

Nova Scotia Community College

Nova Scotia

Semiconductor Devices: Theory and Application Copyright © 2023 by James M. Fiore is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License, except where otherwise noted.

CONTENTS

ABOUT THE BOOK

Semiconductor Devices: Theory and Application NSCC Edition is a condensed remixed version of the open textbook Semiconductor Devices – Theory and Application by James M. Fiore shared under a CC BY-NC-SA license.

HOW TO ACCESS AND USE

This is a free open textbook. It is available in the following formats at **no cost**:

- **Online web-book**. You can read this manual online on a computer or mobile device.
- **PDF**. You can download in two different PDF formats.
    - Digital PDF – optimized for digital use, reading on a computer or tablet.
    - Print PDF – optimized for printing.
- **eBook**. Download the EPUB file.
    - Most tablets and eReaders (including iPad and Kobo) and smartphones (including iPhone and Android) can read eBook files in EPUB format.

LAND ACKNOWLEDGEMENT

NSCC is located in Mi'kma'ki, the unceded territory and traditional homeland of the Mi'kmaq Nation. Our relationship is based on a series of peace and friendship treaties between the Mi'kmaq Nation and the Crown, dating back to 1725.

UNIT 1: PN JUNCTIONS AND DIODES

Learning Objectives

After completing this chapter, you should be able to:

- Describe and diagram the energy hill for a PN junction.
- Discuss the different kinds of diodes available and their uses: rectifier, Zener, LED, photodiode and varactor.
- Detail the device characteristics exhibited by different diode types.
- Graph the forward- and reverse-bias operation regions of diodes.
- Determine the effective resistance of a diode under specific conditions.
- Solve basic DC resistor-diode circuits for various system voltages and currents.

1.1: INTRODUCTION

Having investigated the characteristics of extrinsic N-type and P-type materials in the prior chapter, we shall continue by examining what happens when the these two materials are combined into a single device. It is critical to understand that when we combine P- and N-type materials, we do not do so through simple mechanical means. That is, we do not in some way solder, weld, bolt, friction-fit, glue or duct tape one type of material to another. Rather, we must maintain a single piece of mono-crystalline silicon, not a poly-crystalline amalgam of individual pieces. This can be achieved via a diffusion or ion implantation technique that is applied repeatedly to a single piece of silicon crystal. This will leave regions or zones in the crystal that are N-type or Ptype. In fact, it is quite possible to have a region of one type completely embedded within a region of the opposite type as we shall see in later chapters.

By creating a single zone of N material adjacent to a zone of P material, we wind up with the PN junction. The PN junction is arguably the fundamental building block of solid state semiconductor devices. PN junctions can be found in a variety of devices including bipolar junction transistors (BJTs) and junction field effect transistors (JFETs). The most basic device built from the PN junction is the diode. Diodes are designed for a wide variety of uses including rectifying, lighting (LEDs) and photodetection (photodiodes). We shall begin by examining the basic structure and operation of the PN junction. This will include a look at the many different kinds of diodes available. To assist with circuit analysis, a series of simplified models will be created and investigated. We shall use these models to solve a number of example circuits that feature the many diode variations available.

1.2 THE PN JUNCTION

If we were to create a region of N material abutting a region of P material in a single crystal, an interesting situation occurs. Assuming the crystal is not at absolute zero, the thermal energy in the system will cause some of the free electrons in the N material to "fall" into the excess holes of the adjoining P material. This will create a region that is devoid of charge carriers (remember, electrons are the majority charge carrier in N material while holes are the majority charge carrier in P material). In other words, the area where the N and P materials abut is depleted of available electrons and holes, and thus we refer to it as a depletion region. This is depicted in Figure 1.2.1. The excess electrons of the N material are denoted by minus signs while the excess holes of the P material are denoted with plus signs. At the interface, the free electrons have recombined with holes. When an electron recombines, it leaves behind a positive ion in the N material (shown here as a circled plus sign) and produces a negative ion in the P material (shown as a circled minus sign).

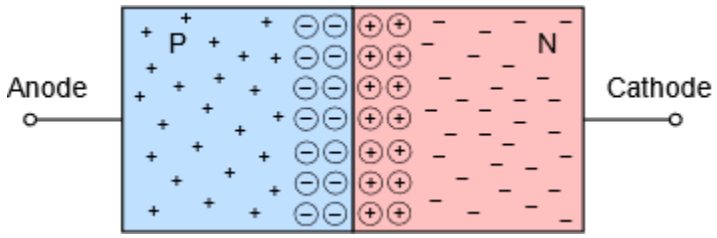


Figure 1.2.1: PN junction.

We now have a region depleted of charge carriers and this will have an effect on the ability to establish a flow of current through the device. We have, in essence, created an energy hill that will need to be overcome.

To understand the concept of the energy hill, recall that in the prior chapter it was discovered that doping an intrinsic crystal would shift the Fermi level. For N material, the Fermi level is shifted up, toward the conduction band. In contrast, for P material the Fermi level is shifted down, nearer to the valence band. When two dissimilar regions adjoin, as in the case here, the energy bands will adjust so that the Fermi levels are consistent. Effectively, this causes the bands of the P material to rise relative to the bands of the N material. The interface between the two appears as a hill, and this is the aforementioned depletion region. This situation is depicted graphically in Figure 1.2.2. Compare this energy diagram to the energy diagrams for N material and P material presented in the prior chapter. By simply aligning the Fermi levels, it should be clear how we arrive at the new energy diagram.

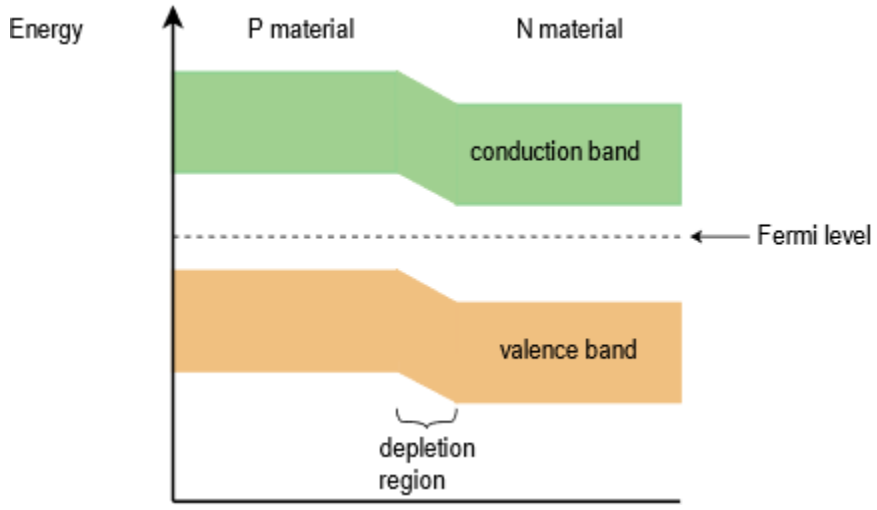


Figure 1.2.2: Energy bands in PN junction.

Now let's consider what happens if we were to connect this device to an external voltage source as shown in Figure 1.2.3. Obviously, there are two ways to orient the PN junction with respect to the voltage source. This version is termed forward-bias.

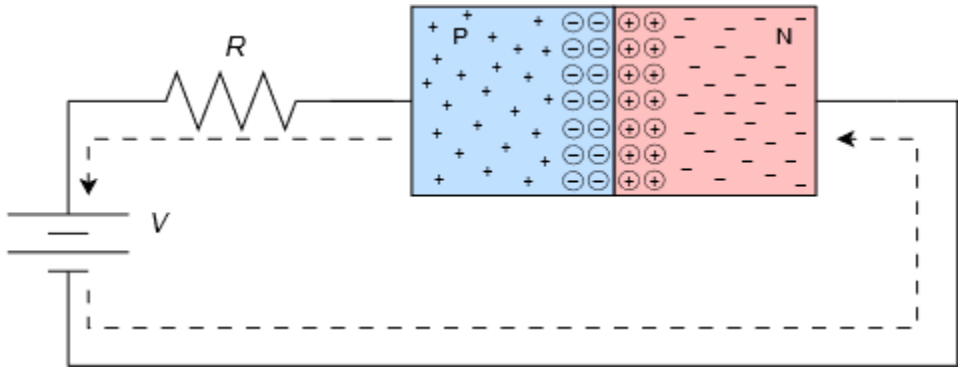


Figure 1.2.3: PN junction connected to external voltage source.

FORWARD-BIAS

The dotted line of Figure 1.2.3 shows the direction of electron flow (opposite the direction of conventional flow). First, electrons flow from the negative terminal of the battery toward the N material. In N material, the majority carriers are electrons and it is easy for these electrons to move through the N material. Upon entering the depletion region, if the supplied potential is high enough, the electrons can diffuse into the P material where there are a large number of lower energy holes. From here, the electrons can migrate through to the positive terminal of the source, completing the circuit (the resistor has been added to limit maximum current flow). The "trick" here is to assure that the supplied potential is large enough to overcome the effect of the depletion region. That is, a certain voltage will be dropped across the depletion region in order to achieve current flow. This required potential is called the barrier potential or forward voltage drop. The precise value depends on the material used. For silicon devices the barrier potential is usually estimated at around 0.7 volts. For germanium devices it is closer to 0.3 volts while LEDs may exhibit barrier potentials in the vicinity of 1.5 to 3 volts, partly depending on the color.

Another way of thinking about this is that the addition of the voltage source "flattens" the inherent energy hill of the junction. Once the applied forward-bias voltage is at least as big as the hill, current can flow easily.

REVERSE-BIAS

If the voltage source polarity is reversed in Figure 1.2.3, the behavior of the PN junction is altered radically. In this case, the electrons in the N material will be drawn toward the positive terminal of the source while the P material holes will be drawn toward the negative terminal, creating a small, short-lived current. This has the effect of widening the depletion region and once it reaches the supplied potential, the flow of current ceases. In essence, we have increased the size of the energy hill. Further increases in the source voltage only serve to make the situation worse. The depletion region simply expands to fill the void, so to speak. Ideally, the PN junction acts like an open circuit with an applied reverse-bias voltage.

This asymmetry in response to a supplied potential turns out to be extraordinarily useful. Perhaps the simplest of all semiconductor devices is the diode. In its basic form a diode is just a PN junction. It is a device that will allow current to pass easily in one direction but prevent current flow in the opposite direction.

SHOCKLEY EQUATION

We can quantify the behavior of the PN junction through the use of an equation derived by William Shockley.

$$I = I_S\left(\frac{V_D q}{e^{nkT}} - 1\right)$$

(1.2.1)

Where

I is the diode current,

I_S is the reverse saturation current,

V_D is the voltage across the diode,

q is the charge on an electron, 1.6E–19 coulombs,

n is the quality factor (typically between 1 and 2),

k is the Boltzmann constant, 1.38E–23 joules/kelvin,

T is the temperature in kelvin.

At 300 kelvin, q/kT is approximately 38.6. Consequently, for even very small forward (positive) voltages, the "–1" term can be ignored. Also, I_S is not a constant. It increases with temperature, approximately doubling for each 10 C° rise (more on this in a moment).

If we plot the Shockley equation using typical values for a silicon device, we arrive at the curve shown in Figure 1.2.4. This plots the junction current as a function of the forward (positive) device voltage. It is a representative curve only. While all silicon diodes will exhibit this same general shape, the precise value of current for a specific voltage will vary depending on the device design.

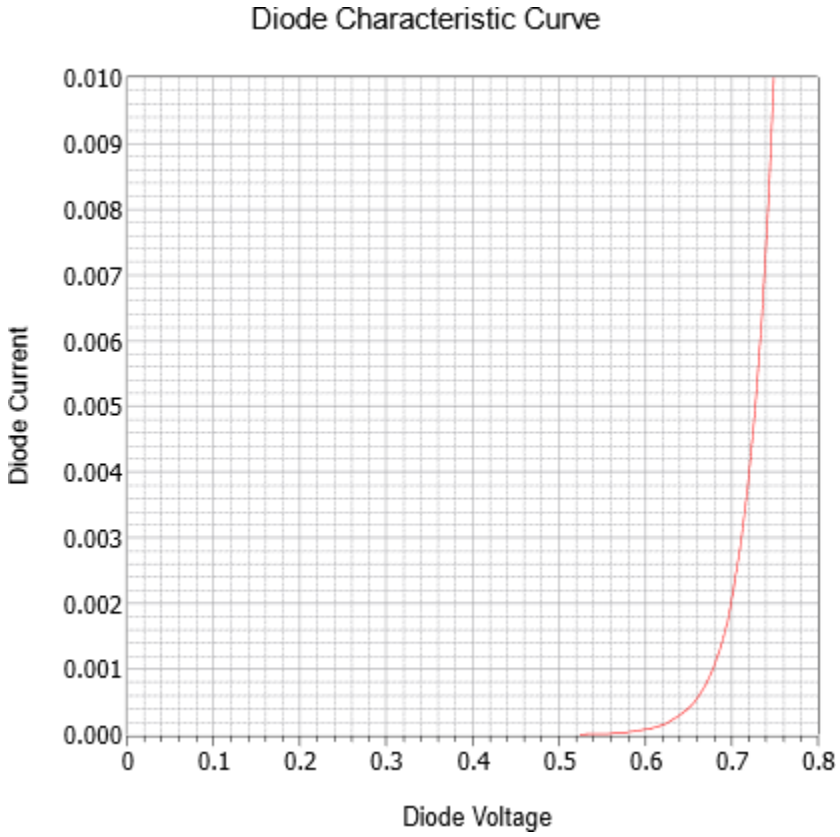


Figure 1.2.4: Characteristic curve of forward-biased silicon PN junction.

For potentials below about 0.5 volts, the current is virtually non-existent. Above this value, the current rises rapidly, becoming nearly vertical after approximately 0.7 volts. If the plot was recreated using a higher temperature, the effect would be to shift the curve to the left (i.e., a higher current for a given voltage).

If we were to alter the graph to use a logarithmic current scale rather than a linear scale, the graph of Figure 1.2.5 results. The resulting straight line plot shows clearly the logarithmic relationship between the diode's voltage and current.

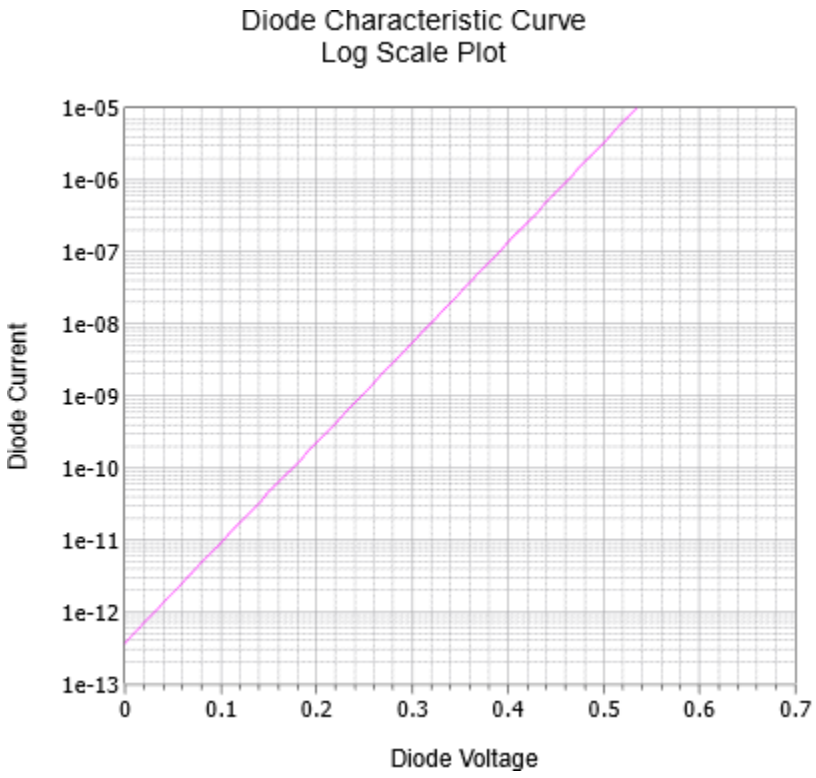


Figure 1.2.5: Characteristic curve of forward-biased silicon PN junction using log scale.

For negative voltages (reverse-bias) the Shockley equation predicts negligible diode current. This is true up to a point. The equation does not model the effects of breakdown. When the reverse voltage is large enough, the diode will start to conduct. This is shown in Figure 1.2.6. In the first quadrant we see the same general shape we found in Figure 1.2.4. V_F is the forward "knee" voltage (roughly 0.7 volts for silicon). I_R is the reverse saturation current (ideally zero but in reality a very small amount of current will flow). V_R is the reverse breakdown voltage. Note that the current increases rapidly once this reverse voltage is reached.

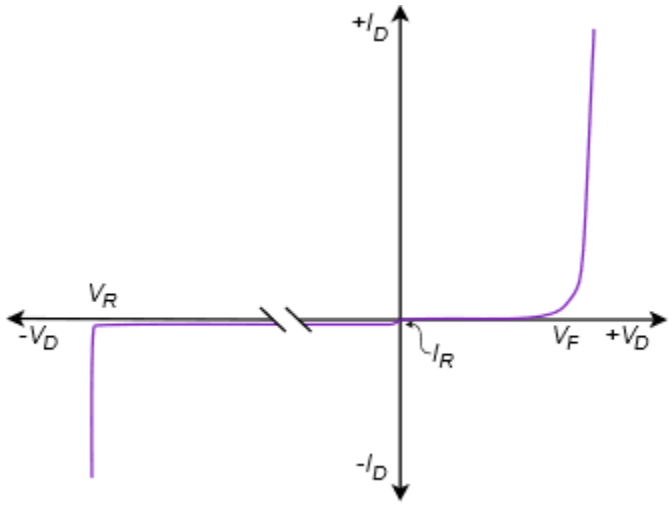


Figure 1.2.6: Simplified forward and reverse I-V curve for diode.

In general, diodes should not be operated in the breakdown region (the exception being Zener diodes). There are two mechanisms behind this phenomenon. The Zener effect, named after Clarence Zener,

predominates when the doping levels are high and produces breakdown voltages below roughly five or six volts. It is due to the production of a very high electric field across the depletion region which then results in the production of a high current through electron tunneling. In devices using lower levels of doping, avalanche dominates. In this instance, a high electric field accelerates the free electrons to the point where they can impact surrounding atoms and create new electron-hole pairs, thus creating new free electrons that can repeat the process, resulting in a rapid increase of current.

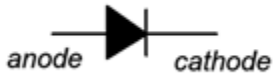


Figure 1.2.7: Diode schematic symbol (ANSI).

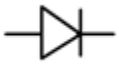


Figure 1.2.8: Alternate diode schematic symbol (IEC).

The schematic symbol for a basic switching or rectifying diode is shown in Figure 1.2.7. This is the ANSI standard which predominates in North America. The P material is the anode while the N material is the cathode[1]. As a general rule for semiconductor schematic symbols, arrows point toward N material. In this case, the arrow also points in the direction of easy conventional current flow. Figure 1.2.8 shows an alternate schematic symbol, the IEC international standard, the difference being that it is in outline form without the body being filled in.

Figure 1.2.9: DO-204 case.

Figure 1.2.10: DO-4 case.
Courtesy of Vishay
Intertechnology

When it comes to physical device packaging, small and medium current and power devices for through-hole mounting include the DO-35 and DO-204, with the model number stamped on the body. The typical size is comparable to a 1/4 to 1/8 watt resistor. As seen in Figure 1.2.9 the cathode end is denoted by a band, reminiscent of the bar on the schematic symbol. Surface mount packages are also available. Devices handling higher currents and powers often come in stud or bolt styles such

1. Cathode is often denoted by a k. This is likely due to the word's Greek root, kathodos.

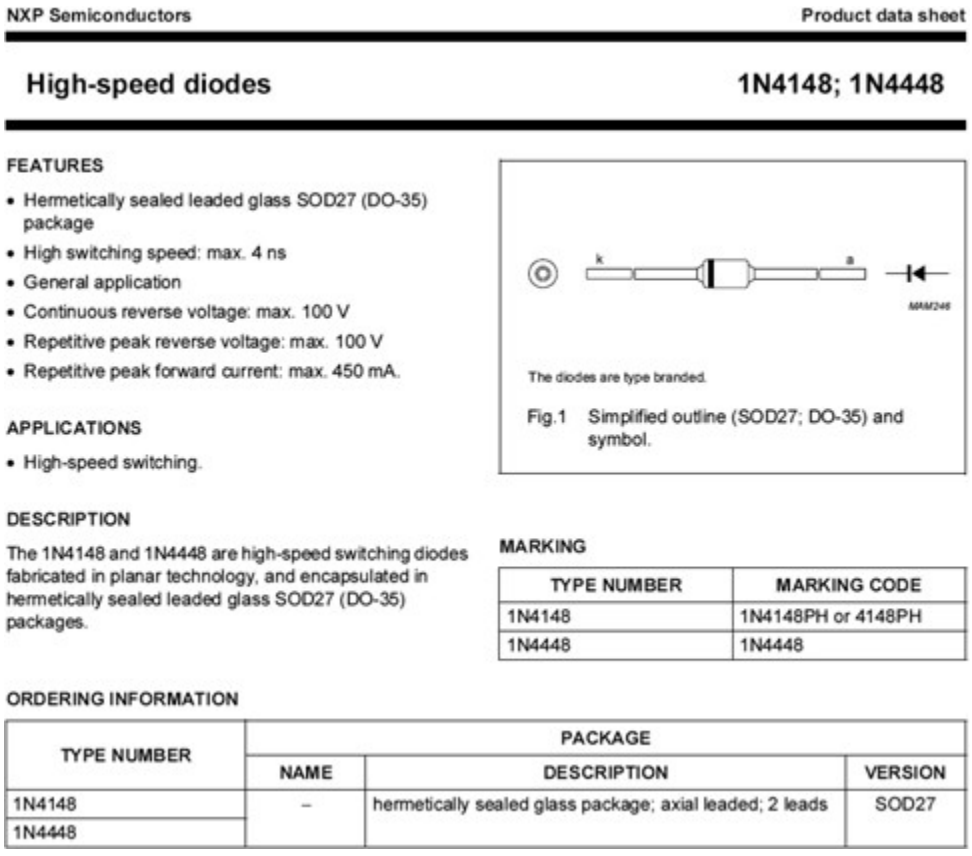
as the DO-4 shown in Figure 1.2.10. These packages facilitate mounting to a metal plate or heat sink to help dissipate the excess heat.

1.3 DIODE DATA SHEET INTERPRETATION

A data sheet for the popular 1N4148 switching diode is shown in Figure 1.3.1. The 1N4148 is designed for high speed operation required in high frequency signal applications but also finds use in a variety of general purpose applications that do not require very high current or power handling.

Some of the key features include a four nanosecond switching speed, a maximum reverse voltage of 100 volts and a 450 milliamp maximum forward current (with short single pulses as high as four amps being possible). Power dissipation is 500 milliwatts.

Referring to Figure 1.3.1*b*, the variation in reverse current with regard to temperature is obvious. This also verifies the "doubles every 10 C° " rule-of-thumb.



NXP Semiconductors — Product data sheet

High-speed diodes — **1N4148; 1N4448**

FEATURES

- Hermetically sealed leaded glass SOD27 (DO-35) package
- High switching speed: max. 4 ns
- General application
- Continuous reverse voltage: max. 100 V
- Repetitive peak reverse voltage: max. 100 V
- Repetitive peak forward current: max. 450 mA.

APPLICATIONS

- High-speed switching.

The diodes are type branded.

Fig.1 Simplified outline (SOD27; DO-35) and symbol.

DESCRIPTION

The 1N4148 and 1N4448 are high-speed switching diodes fabricated in planar technology, and encapsulated in hermetically sealed leaded glass SOD27 (DO-35) packages.

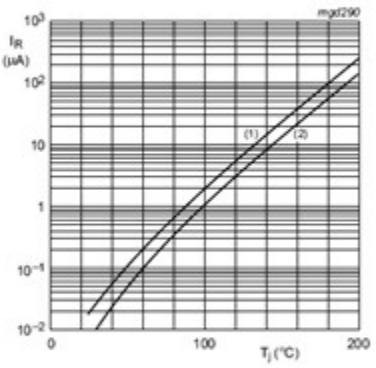
MARKING

TYPE NUMBER	MARKING CODE
1N4148	1N4148PH or 4148PH
1N4448	1N4448

ORDERING INFORMATION

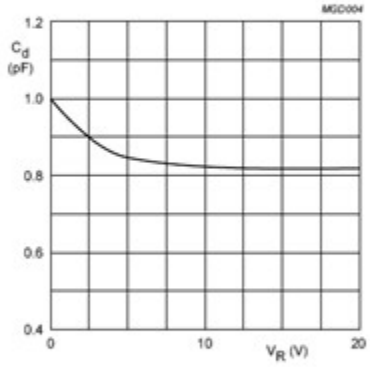
TYPE NUMBER	PACKAGE		
	NAME	DESCRIPTION	VERSION
1N4148	–	hermetically sealed glass package; axial leaded; 2 leads	SOD27
1N4448			

Figure 1.3.1a: 1N4148 data sheet. Courtesy of NXP Semiconductors.



(1) V_R = 75 V; typical values.
(2) V_R = 20 V; typical values.

Fig.5 Reverse current as a function of junction temperature.



f = 1 MHz; T_j = 25 °C.

Fig.6 Diode capacitance as a function of reverse voltage; typical values.

Figure 1.3.1b: 1N4148 data sheet (continued).

LIMITING VALUES

In accordance with the Absolute Maximum Rating System (IEC 60134).

SYMBOL	PARAMETER	CONDITIONS	MIN.	MAX.	UNIT
V_{RRM}	repetitive peak reverse voltage		–	100	V
V_R	continuous reverse voltage		–	100	V
I_F	continuous forward current	see Fig.2; note 1	–	200	mA
I_{FRM}	repetitive peak forward current		–	450	mA
I_{FSM}	non-repetitive peak forward current	square wave; T_j = 25 °C prior to surge; see Fig.4			
		t = 1 μs	–	4	A
		t = 1 ms	–	1	A
		t = 1 s	–	0.5	A
P_{tot}	total power dissipation	T_{amb} = 25 °C; note 1	–	500	mW
T_{stg}	storage temperature		−65	+200	°C
T_j	junction temperature		–	200	°C

Note

1. Device mounted on an FR4 printed-circuit board; lead length 10 mm.

ELECTRICAL CHARACTERISTICS

T_j = 25 °C unless otherwise specified.

SYMBOL	PARAMETER	CONDITIONS	MIN.	MAX.	UNIT
V_F	forward voltage	see Fig.3			
	1N4148	I_F = 10 mA	–	1	V
	1N4448	I_F = 5 mA	0.62	0.72	V
		I_F = 100 mA	–	1	V
I_R	reverse current	V_R = 20 V; see Fig.5		25	nA
		V_R = 20 V; T_j = 150 °C; see Fig.5	–	50	μA
I_R	reverse current; 1N4448	V_R = 20 V; T_j = 100 °C; see Fig.5	–	3	μA
C_d	diode capacitance	f = 1 MHz; V_R = 0 V; see Fig.6	–	4	pF
t_{rr}	reverse recovery time	when switched from I_F = 10 mA to I_R = 60 mA; R_L = 100 Ω; measured at I_R = 1 mA; see Fig.7	–	4	ns
V_{fr}	forward recovery voltage	when switched from I_F = 50 mA; t_r = 20 ns; see Fig.8	–	2.5	V

THERMAL CHARACTERISTICS

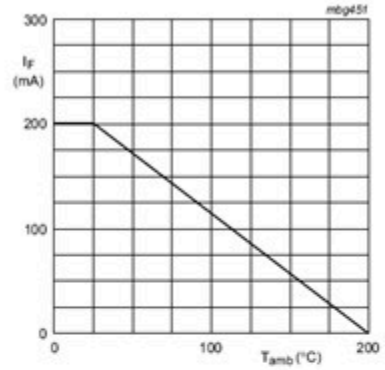
SYMBOL	PARAMETER	CONDITIONS	VALUE	UNIT
$R_{th(j-tp)}$	thermal resistance from junction to tie-point	lead length 10 mm	240	K/W
$R_{th(j-a)}$	thermal resistance from junction to ambient	lead length 10 mm; note 1	350	K/W

Note

1. Device mounted on a printed-circuit board without metallization pad.

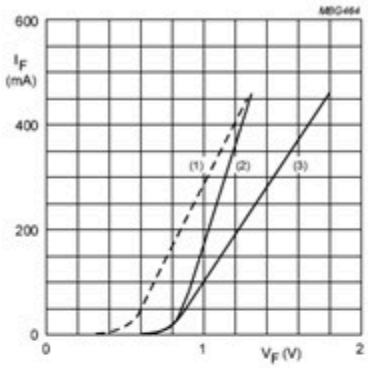
Figure 1.3.1c: 1N4148 data sheet (continued).

GRAPHICAL DATA



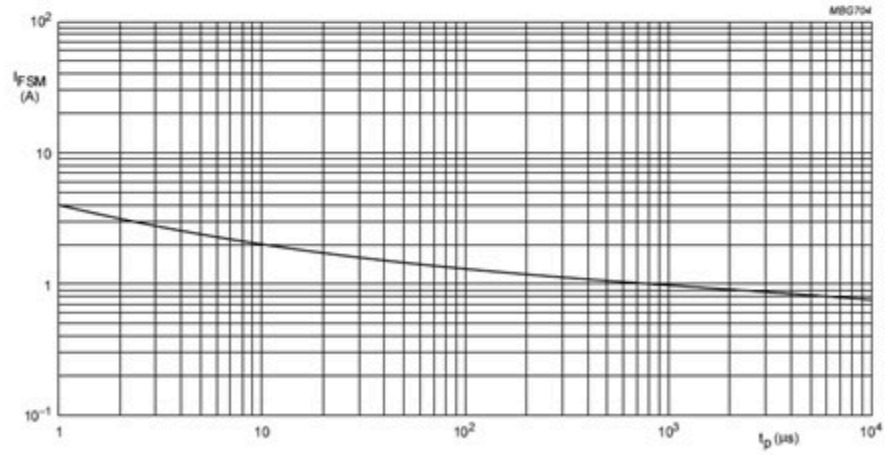
Device mounted on an FR4 printed-circuit board; lead length 10 mm.

Fig.2 Maximum permissible continuous forward current as a function of ambient temperature.



(1) T_j = 175 °C; typical values.
(2) T_j = 25 °C; typical values.
(3) T_j = 25 °C; maximum values.

Fig.3 Forward current as a function of forward voltage.



Based on square wave currents.
T_j = 25 °C prior to surge.

Fig.4 Maximum permissible non-repetitive peak forward current as a function of pulse duration.

Figure 1.3.1d: 1N4148 data sheet (continued).

[Accessible version of Data Tables in Figures 1.3.1]

Power derating and permissible pulse amplitudes can be seen in Figure 1.3.1*d*. Finally, note the variation in the forward voltage curves due to temperature. As stated previously, for a given current, an increase in temperature results in a lower forward voltage. Also, at room temperature, we see a knee voltage of approximately 0.7 volts.

1.4 DIODE CIRCUIT MODELS

One thing is very clear from the characteristic curve of the diode: It is not a linear bilateral[1] device, quite unlike a resistor. Consequently, we cannot use the superposition technique to solve diode circuits unless we have a priori knowledge about it, that is, whether or not it is forward- or reverse-biased. For example, we can imagine a circuit comprised of two voltage sources, resistors and a diode. By itself, one of the voltage sources might forward-bias the diode while the other would reverse-bias it. Obviously, a diode cannot be both forward and reverse-biased at the same time.

A second problem we face with circuit analysis is the added complexity of the Shockley equation. For speed and ease of computation we find it useful to model the diode with simpler circuit elements. Three diode models are shown in Figure 1.4.1.

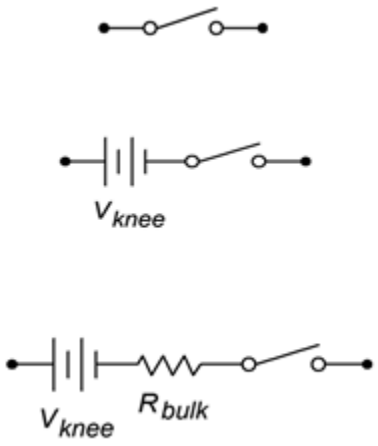


Figure 1.4.1: Simplified diode models. Top to bottom: first, second and third approximations, increasing in accuracy.

The first approximation is the simplest of the three. It treats the diode as a simple dependent switch: the switch is closed if the diode is forward-biased and open if it is reverse-biased. The second approximation adds the effect of the forward voltage. V_{knee} is the "turn-on" potential required to overcome the energy hill. It would be 0.7 volts for a silicon device. The third approximation is the most accurate of the three. A close look at the characteristic curve of Figure 1.2.4 shows that once the knee voltage is reached, the curve does not transition to a perfect vertical line. Instead, there remains some positive, non-infinite slope. That is, the voltage continues to increase, although modestly, with further increases in current. We can approximate this effect as a small resistive value, R_{bulk} . The three corresponding I-V plots are shown in Figure 1.4.2. Compare these to Figure 2.2.6 and note the increasing accuracy.

1. The I-V plot is not a straight line (linear) and the forward and reverse quadrants are not identical (bilateral).

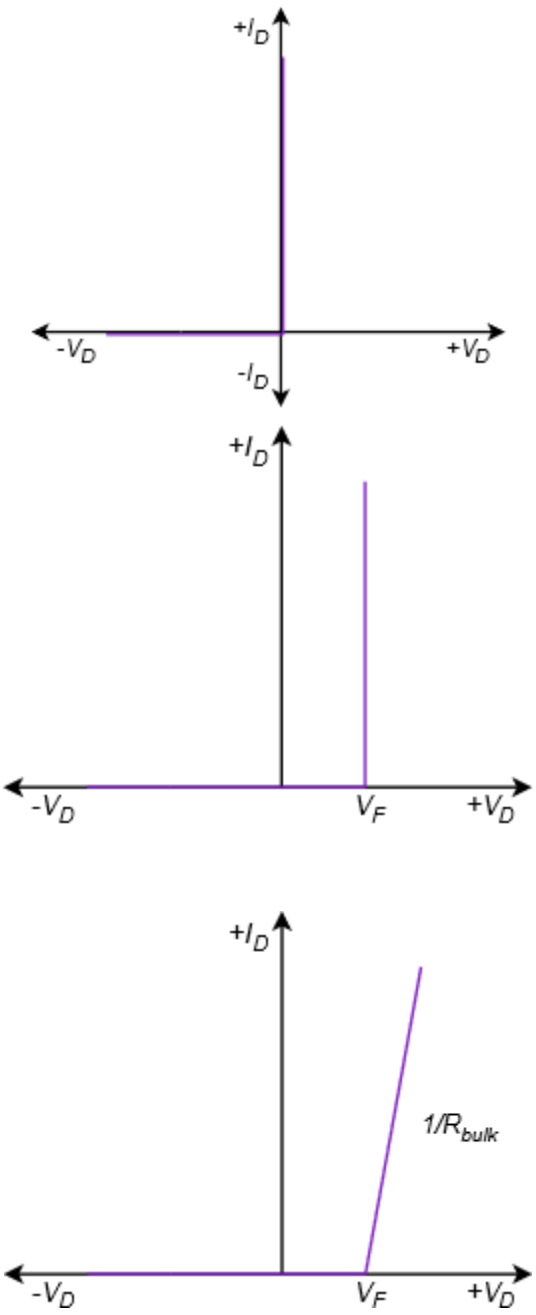


Figure 1.4.2 : I-V curves for simplified diode models. Top to bottom: first, second and third approximations.

In many applications the second approximation will yield sufficiently accurate results and we will tend to make greatest use of it. Just remember that these are behavioral models; don't think that there are literally 0.7 volt sources or little resistors in the diodes.

It should be noted that R_{bulk} does not represent the "diode resistance" per se, rather, it models a minimum value. There really is no such thing as a singular diode resistance. We can, however, talk about the effective resistance of a diode in a particular circuit in both DC and AC terms.

The key to understanding this concept is to remember that resistance is a linear function, a straight line on a I-V graph. Therefore, we need to find a straight line "fit" for the diode curve. Two possibilities are shown in Figure 1.4.3.

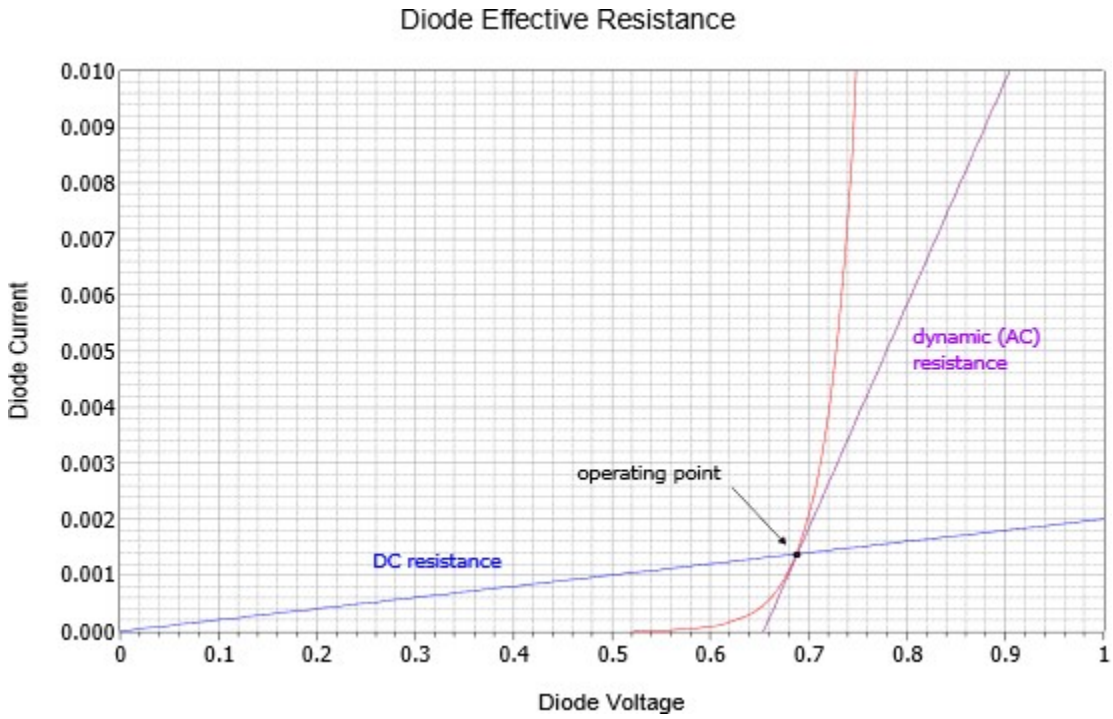


Figure 1.4.3: Diode effective resistance for DC and AC

The red curve is the diode's characteristic curve (arbitrary current values are shown). For some particular DC circuit, a specific current will flow through the diode which will produce a particular voltage, denoted on the graph as the operating point. If we simply compute the ratio of that voltage to the driving current, we wind up with a resistance. This is the effective DC resistance of the diode under these circuit conditions and is represented by the blue line. That is, the reciprocal of the slope of the blue line is the effective DC resistance. Obviously, if we shift the operating point along the red diode curve, the slope of the intersecting blue line changes and therefore we arrive at a new DC resistance. The higher the current, the lower the effective DC resistance.

Instead of just DC, the diode might see a combination of DC and AC signals. Visualize this as adding a small AC variation on top of the DC. We can imagine the operating point moving along the red diode curve, back and forth about the operating point. If we divide the small AC voltage variation by its associated AC current variation, we wind up with the AC equivalent resistance, also known as the dynamic resistance. Graphically, we can think of this as finding the slope of a line that is tangent to the operating point (the purple line). This will in fact, be an average value across the AC variation. It should also be apparent that the effective AC resistance must be smaller than its DC counterpart because the AC approximation (purple line) must be steeper than the DC approximation (blue line)[2]. It is time for a few illustrative examples.

2. The dynamic resistance of a PN junction may be approximated as 26 mV/◇◇◇◇◇◇◇◇◇◇. This will be shown in an upcoming chapter.

Example 1.4.1

Consider the resistor-diode circuit of Figure 1.4.4. Assume the voltage source is 12 volts and the resistor is 2 kΩ. Further, assume the diode is silicon and its bulk resistance is 10 Ω. Using the three diode approximations, compute the circulating current.

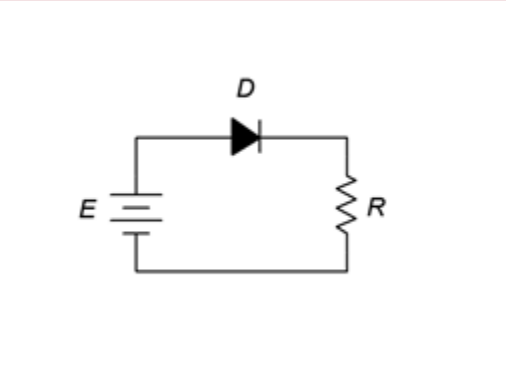


Figure 1.4.4: Schematic for Example 1.4.1 .

First, note that the diode is forward-biased. This must be the case because there is a single voltage that is larger than the knee voltage and its positive terminal is attached to the diode's anode. No matter which approximation we use, Kirchhoff's voltage law (KVL) must be true so it will be a matter of summing the available voltage drops versus resistance(s).

Using the first approximation:

Here we assume the diode is a closed switch. Consequently all of the source voltage must drop across the single resistor.

$$I = \frac{E}{R}$$

$$I = \frac{12V}{2k\Omega}$$

$$I = 6mA$$

Using the second approximation:

In this instance we include the knee voltage.

$$I = \frac{E - V_{knee}}{R}$$

$$I = \frac{12V - 0.7V}{2k\Omega}$$

$$I = 5.65mA$$

Using the third approximation:

The most accurate of the three, we include both the knee voltage and bulk resistance.

$$I = \frac{E - V_{knee}}{R + R_{bulk}}$$

$$I = \frac{12V - 0.7V}{2k\Omega + 10\Omega}$$

$$I = 5.622mA$$

In this particular case the difference between the second and third approximations is less than 1%. It is also worth noting that the third approximation predicts a diode voltage of slightly more than 0.7 volts (approximately 0.756 volts) due to the additional potential across the bulk resistance.

Example 1.4.2

Determine the circulating current for the circuit in Figure 1.4.5. Also find the diode and resistor voltages. Assume the power supply is 20 volts, the diode is silicon and the resistor is 2 kΩ.

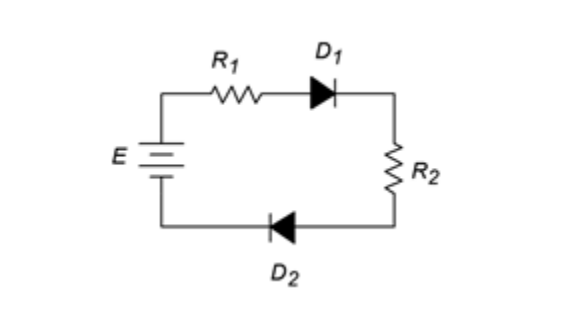


Figure 1.4.5: Schematic for Example 1.4.2.

This problem is deceptively easy. Note that the positive terminal of the source is connected to the cathode. As there are no other sources in the circuit, the diode must be reverse-biased. The model for a reverse-biased diode is an open switch and the circulating current in an open circuit is zero. Therefore, the resistor voltage must also be zero and values for the knee voltage and bulk resistance are not needed. In order to satisfy KVL, the diode voltage will equal the source of 20 volts (+ to – from cathode to anode).

The only time this would not be the case is if the reverse breakdown voltage of the diode is less than the 20 volt source. In that case the diode voltage would equal the breakdown voltage with the remainder of the source voltage dropping across the resistor.

Example 1.4.3

Determine the circulating current for the circuit in Figure 1.4.6. Also find the diode and resistor voltages. Assume the power supply is 9 volts, the diodes are silicon and R_1 = 1 kΩ, R_2 = 2 kΩ.

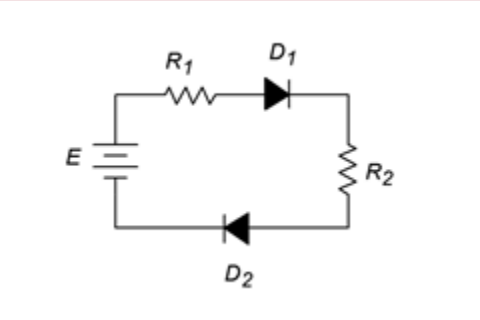


Figure 1.4.6: Schematic for Example 1.4.3.

According to KVL, the applied source must equal the sum of the voltage drops across the resistors and diodes as this is a single loop. Both diodes are forward-biased (conventional current entering the anodes).

$$I = \frac{E - V_{knee1} - _{k\ nee2}}{R_1 + R_2}$$

$$I = \frac{9V - 0.7V - 0.7V}{1k\Omega + 2k\Omega}$$

$$I = 2.533mA$$

Note that if either diode was reversed, there would be no current flow and all of the source potential would drop across the reversed diode.

Example 1.4.4

Determine the source current and resistor voltages for the circuit in Figure 1.4.7. Also find the resistor voltages if the diode polarity is reversed. Assume the power supply is 10 volts, the diode is silicon and the resistors are 1 kΩ each.

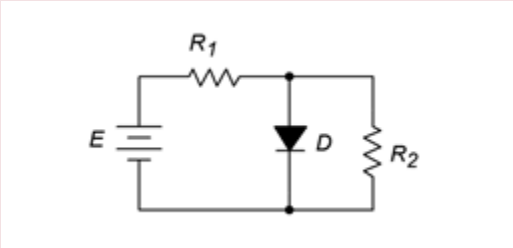


Figure 1.4.7: Schematic for Example 1.4.4 .

As D and R_2 are in parallel they must have the same voltage drop. Also, the diode is forward-biased. Therefore, the voltage across R_2 must be approximately 0.7 volts, leaving 9.3 volts to drop across R1. The current through R1 is the source current.

$$I = \frac{E - V_D}{R_1}$$

$$I = \frac{10V - 0.7V}{1k\Omega}$$

$$I = 9.3mA$$

COMPUTER SIMULATION

To verify our results, Example 1.4.4 is simulated. The circuit is captured as shown in Figure 1.4.8*a*. This particular example is shown in Multisim although any decent quality simulator will do. The very common 1N4148 switching diode is used here. Another popular choice would be the 1N914 switching diode or a 1N400X series rectifier.

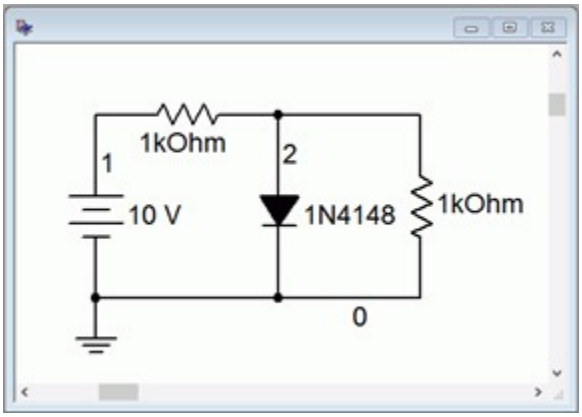


Figure 1.4.8a: The circuit of Example 1.4.4 in Multisim.

Next, a DC Operating Point analysis is performed. The results are shown in Figure 1.4.8*b*. Note that the diode potential is just under the 0.7 volt approximation. From this we can deduce that the voltage

drop across the first resistor must be slightly more than 9.3 volts, producing a current slightly more than 9.3 mA.

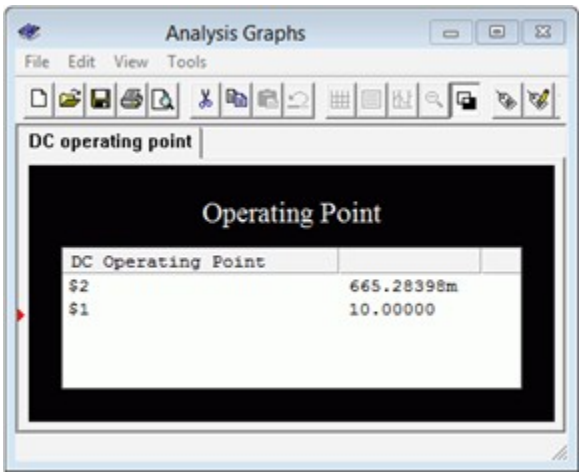


Figure 1.4.8b: DC Operating Point simulation results for the circuit of Example 1.4.4.

Finally, Figure 1.4.8*c* shows the results when the diode is reversed in the circuit. The second resistor (node 3 to ground) shows 5 volts as expected. Therefore, the first resistor must also be dropping 5 volts.

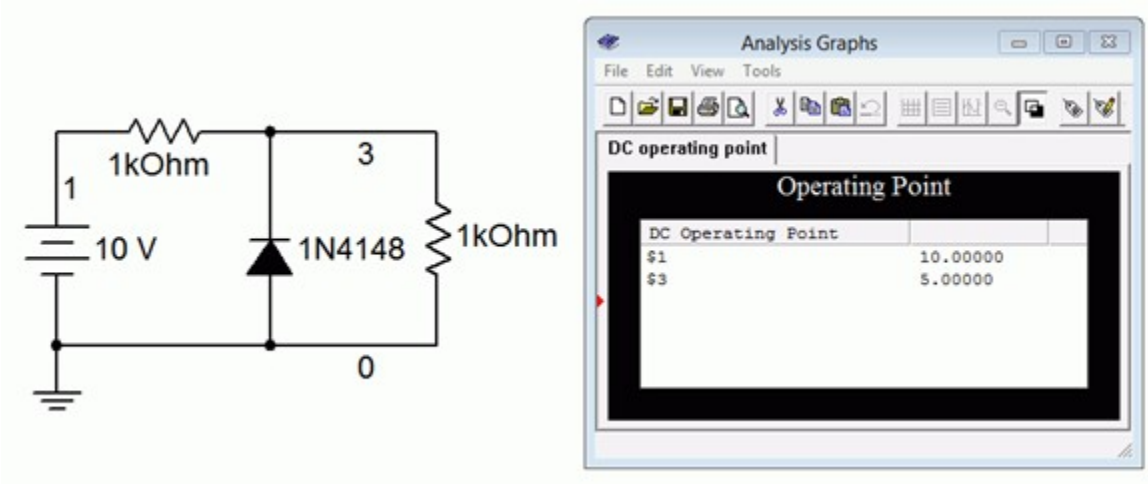


Figure 1.4.8c: Simulation of Example 1.4.4 using reversed diode orientation.

Before moving on to another topic, let's take a look at a somewhat more involved example using multiple diodes.

Example 1.4.5

Determine the diode and resistor voltages for the circuit in Figure 1.4.9. Assume the diodes are silicon.

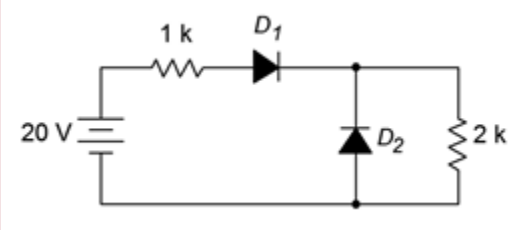


Figure 1.4.9: Schematic for Example1.4.5 .

The first thing to notice is that D_1 is forward-biased while D_2 is reverse-biased. Therefore, the 20 volt source must equal the drop across D_1 and the two resistors. D_2 will take on whatever the drop across the 2 kΩ works out to as they are in parallel.

$$I = \frac{E - V_D 1}{R_1 + R_2}$$

$$I = \frac{20V - 0.7V}{1k\Omega + 2k\Omega}$$

$$I = 6.433mA$$

Note that virtually no current flows down through D2 as it is reverse-biased. Using Ohm's law, the drop across the first resistor is 6.433 volts and for the second resistor, 12.867 volts.

1.5 OTHER TYPES OF DIODES

Diodes have been designed to exploit different aspects of PN junctions. Besides the basic use as a switching or rectifying device, diodes are available for voltage regulation, variable capacitance, illumination and light sensing. The schematic symbols for a number of popular diode types are shown in Figure 1.5.1. Note the similarities of the symbols. The "bar" portion represents the cathode for all of them.

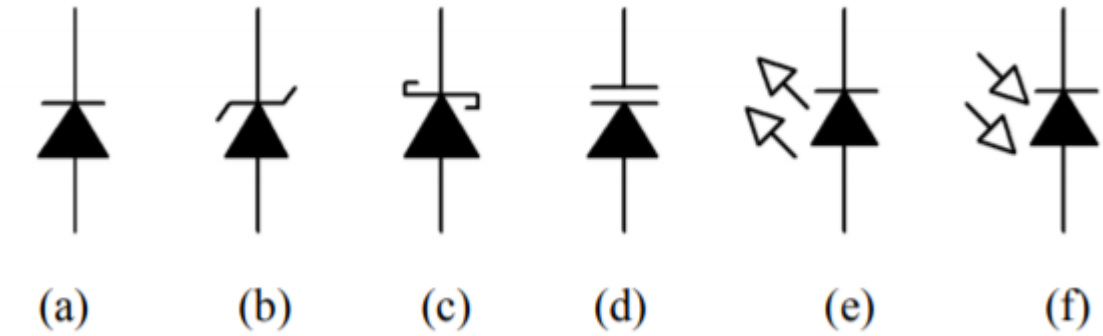


Figure 1.5.1: Diode schematic symbols: a) switching or rectifying b) Zener c) Schottky d) varactor e) LED f) photodiode

ZENER DIODE

The Zener diode behaves like an ordinary signal diode when forward-biased. Normally, though, Zeners are used in a reverse-bias condition. From our previous discussion, recall that if the reverse potential is high enough, a diode can go into breakdown, causing a rapid increase current. This was caused by either of two effects, Zener conduction or avalanche. The Zener diode takes advantage of this in order to produce a stable voltage.[1] Zeners are specified by their reverse potential (generically referred to as the "Zener voltage") and are designed to handle larger currents and powers than the average signal diode. Zener voltages are standardized in much the same manner as resistors so values such as 3.9 volts, 5.1 volts and 6.8 volts are to be expected. The Zener voltage is measured at I_{ZT} , the Zener test current. A lower current may not fully push the diode into conduction resulting in a lower than expected diode potential.

1. Although they are called Zener diodes, they rely on either the Zener or avalanche effects, depending on the magnitude of the voltage.

Figure 1.5.2 : Zener diode schematic symbol.

Instead of modeling the Zener as an open switch when reverse-biased, instead we model it as an open when its voltage is less than the rated voltage, and as a voltage source equal to the rated value if its voltage tries to exceed that value. When analyzing Zener- based circuits, the first thing to do is determine if the diode is forward-biased. If it is then treat it like an ordinary switching diode. If, on the other hand, it is reverse-biased then treat it like an open switch. If the resulting diode voltage is greater than the Zener voltage then recompute the circuit but this time mentally replace the Zener with a voltage source equal to the Zener voltage. Our next example will illustrate this method.

Example 1.5.1

Determine the circulating current for the circuit in Figure 1.5.3. Also find the diode and resistor voltages. Assume the power supply is 9 volts, the Zener voltage is 5.1 volts and the resistor is 3.3 kΩ.

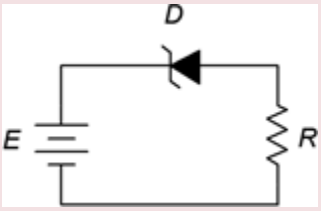


Figure 1.5.3: Schematic for Example 1.5.1.

The diode is reverse-biased. If we treat it as an open then it would drop the entire source voltage, or 9 volts. This is greater than the Zener potential so the device must be in Zener conduction. This means that conventional current will flow relatively easily in a clockwise direction. The voltage across the diode will equal the rated value of 5.1 volts, + to – from cathode to anode. By KVL the resistor drop must be 9 V – 5.1 V, or 3.9 volts.

$$I = \frac{EV_{Zener}}{R}$$

$$I = 9V - 5.1V3.3k\Omega$$

$$I = 1.182mA$$

If the diode was flipped in orientation then it would be forward-biased and show the expected 0.7 volts with 8.3 volts across the resistor.

COMPUTER SIMULATION

A Zener diode circuit is simulated as shown in Figure 1.5.4.

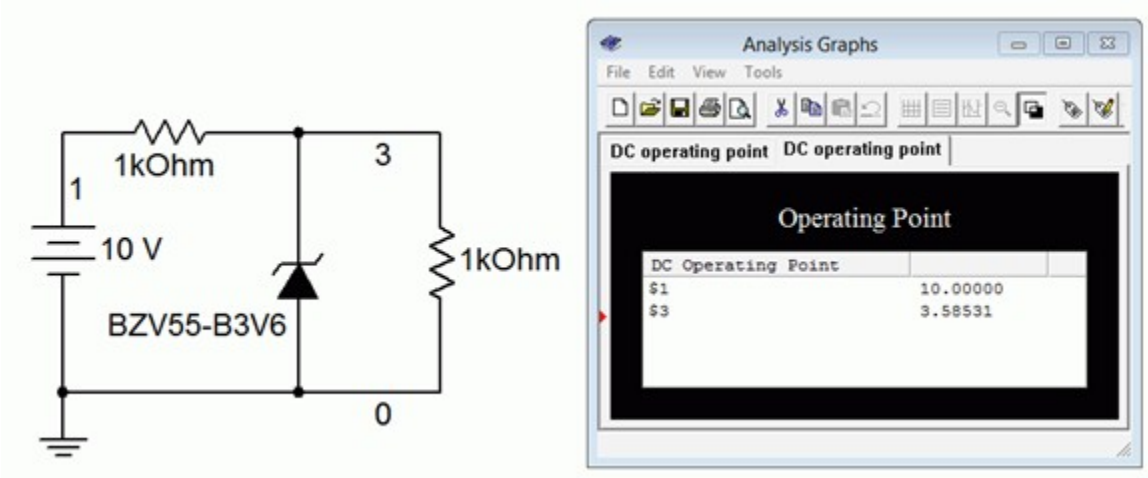


Figure 1.5.4: Zener diode circuit simulation.

Without the Zener, the two resistors would simply split the supply voltage equally, each receiving 5 volts. If an ordinary diode was used it would be reverse-biased and act as an open. The resulting voltages would be the same. In this case, however, the Zener is activated at 3.6 volts (it is typical to include the Zener voltage as part of the model number with the letter "V" replacing the decimal point when needed). Therefore we see roughly 3.6 volts across the Zener and the parallel second resistor (node 3 to ground). The precise value of voltage will depend on the magnitude of the diode current. If the simulation is rerun with a higher voltage source, the increased current will produce a slightly higher voltage at node 3. This is because the breakdown curve is not infinitely steep once it goes past the rated Zener voltage. The effect is similar to that of R_{bulk} in a forward-biased diode. On a data sheet this value is referred to as the differential resistance, or R_{dif} .

An excellent use of the Zener is to limit or regulate a voltage. When a Zener is placed in parallel with other components we can ensure that those components will not see a potential higher than the rated Zener voltage. We will take a much closer look at this in the next chapter.

LIGHT EMITTING DIODE (LED) AND PHOTODIODE

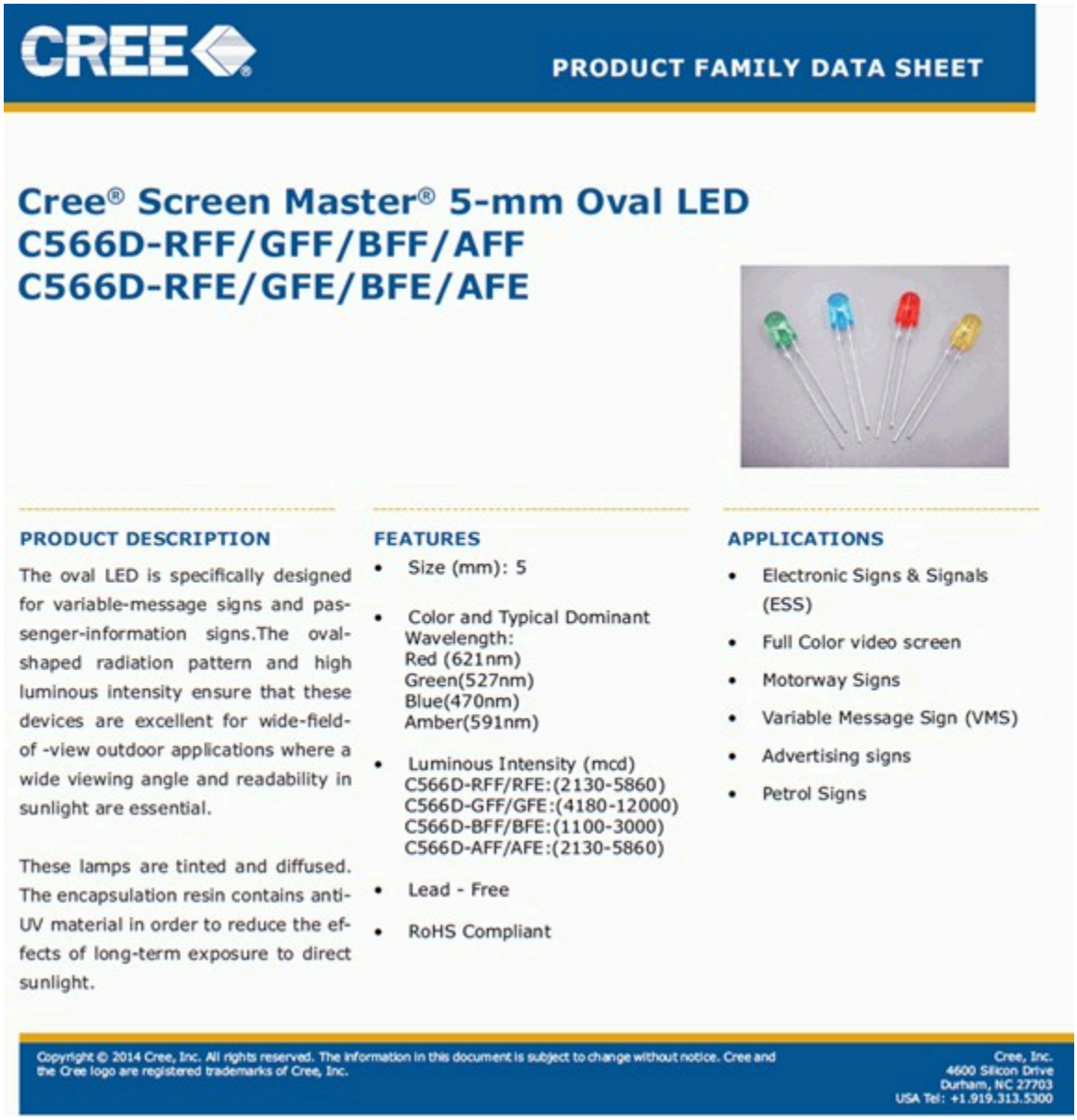
The light emitting diode (LED) and photodiode are complements. While the LED produces light with an electrical input, the photodiode produces a current when exposed to light. Both devices can operate within the human visible spectrum and can also be designed to operate at wavelengths outside this range, in the infrared (IR) and ultraviolet (UV). In fact, most TV remote controls rely on IR emitter/detector pairs for communication.[2]

2. There are advantages to using the infrared over the visible spectrum for this application. It tends to be less sensitive to room lighting conditions and there are no potentially annoying visible flashes of light coming from the remote.

Figure 1.5.5: Light emitting diode (LED) schematic symbol.

The LED has displaced traditional incandescent (filament-based) light sources in many applications due to its high efficiency in turning an electrical energy input into a light output. They are small, physically robust, operate relatively cool and are available in a number of different colors. The schematic symbol is shown in Figure 1.5.5. The basic idea behind its operation is fairly simple. In a forward-biased PN junction, when free electrons recombine and "fall" into lower energy valence holes they must give up this energy differential in some manner. In most diodes, this energy is emitted as heat. In LEDs, the energy transition is designed such that it is radiated at shorter wavelengths (i.e., visible light). In order to achieve this, LEDs are not formed just using silicon as in a typical switching diode. Instead, somewhat more exotic materials are used. From an analysis or design standpoint, the important thing to remember is that the forward voltage tends to be noticeably higher than silicon's 0.7 volt drop. The precise value will depend on the material, which in turn effects the color. A generic red LED will likely exhibit a forward drop of around 1.8 volts or so. Other colors tend to be somewhat higher as we move through the rainbow, ending with blue and UV LEDs (and also high brightness versions) up around 3 to 4 volts. In a lab it is easy to determine the approximate forward drop of a given diode by connecting it in series with a voltage source and current limiting resistor. The supply is increased until the desired brightness is achieved and then the diode drop can be measured with a DMM. When reverse-biased the LED behaves like a switching diode, that is, it looks like an open switch. Unlike switching and rectifying diodes, LED maximum reverse potentials tend to be relatively low, perhaps just a few volts.

A datasheet for the Cree C566D series LED is presented in Figure 1.5.6. Notice that the colors are specified in terms of wavelength (in nanometers) and luminous intensity (brightness) is given in millicandella (mcd).



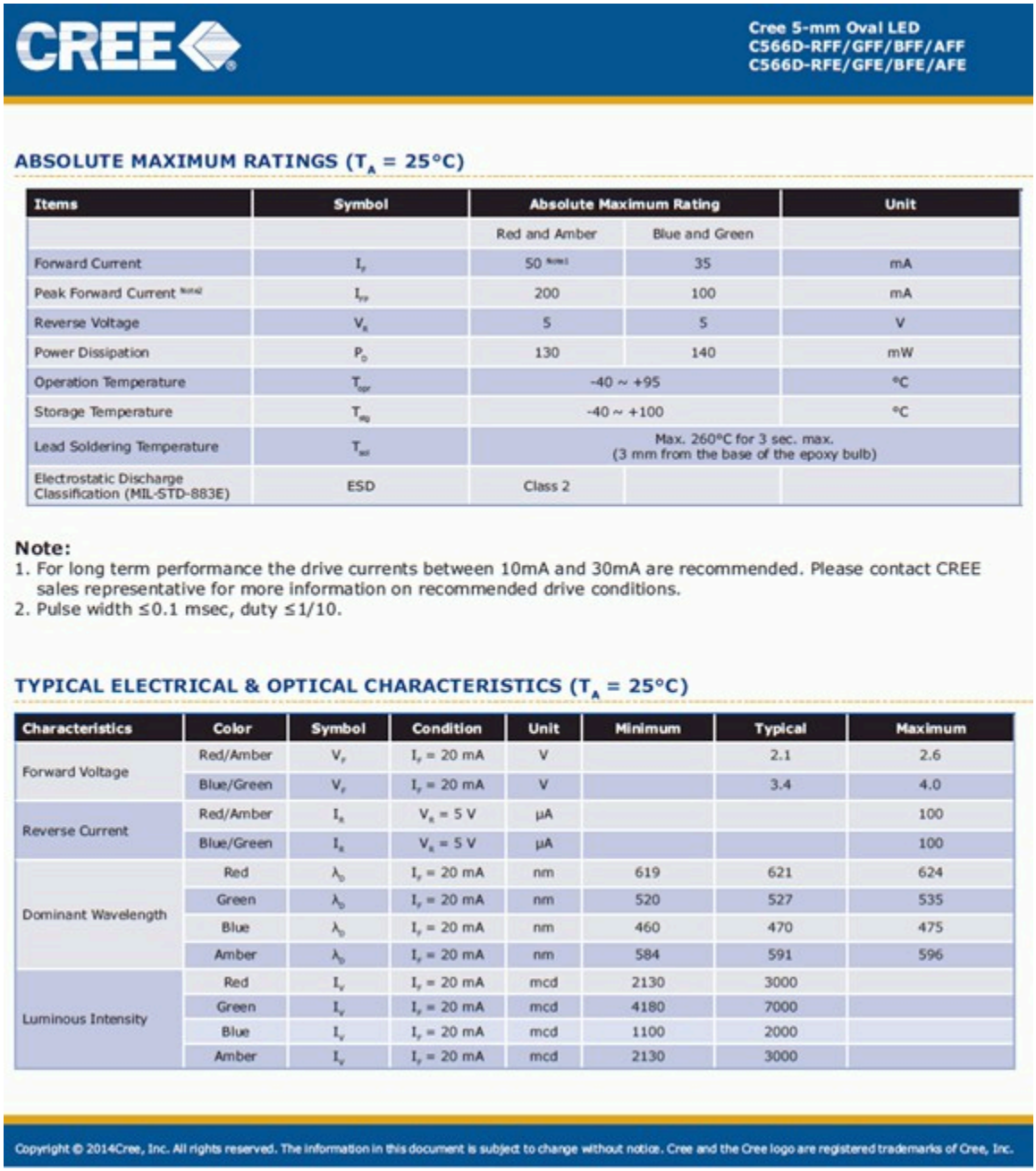
CREE

PRODUCT FAMILY DATA SHEET

Cree® Screen Master® 5-mm Oval LED
C566D-RFF/GFF/BFF/AFF
C566D-RFE/GFE/BFE/AFE

Figure 1.5.6a: LED datasheet. Image Credit: Cree

These devices are RoHS compliant which stands for Restriction of Hazardous Substances. It is an EU directive that limits usage of toxic materials such as lead, cadmium and mercury. Continuing, we find the device maximum ratings:



CREE

Cree 5-mm Oval LED
C566D-RFF/GFF/BFF/AFF
C566D-RFE/GFE/BFE/AFE

ABSOLUTE MAXIMUM RATINGS (T_A = 25°C)

Items	Symbol	Absolute Maximum Rating		Unit
		Red and Amber	Blue and Green	
Forward Current	I_F	50	35	mA
Peak Forward Current	I_{FP}	200	100	mA
Reverse Voltage	V_R	5	5	V
Power Dissipation	P_D	130	140	mW
Operation Temperature	T_{opr}	-40 ~ +95		°C
Storage Temperature	T_{stg}	-40 ~ +100		°C
Lead Soldering Temperature	T_{sol}	Max. 260°C for 3 sec. max. (3 mm from the base of the epoxy bulb)		
Electrostatic Discharge Classification (MIL-STD-883E)	ESD	Class 2		

Note:
1. For long term performance the drive currents between 10mA and 30mA are recommended. Please contact CREE sales representative for more information on recommended drive conditions.
2. Pulse width ≤0.1 msec, duty ≤1/10.

TYPICAL ELECTRICAL & OPTICAL CHARACTERISTICS (T_A = 25°C)

Characteristics	Color	Symbol	Condition	Unit	Minimum	Typical	Maximum
Forward Voltage	Red/Amber	V_F	I_F = 20 mA	V		2.1	2.6
	Blue/Green	V_F	I_F = 20 mA	V		3.4	4.0
Reverse Current	Red/Amber	I_R	V_R = 5 V	μA			100
	Blue/Green	I_R	V_R = 5 V	μA			100
Dominant Wavelength	Red	λ_D	I_F = 20 mA	nm	619	621	624
	Green	λ_D	I_F = 20 mA	nm	520	527	535
	Blue	λ_D	I_F = 20 mA	nm	460	470	475
	Amber	λ_D	I_F = 20 mA	nm	584	591	596
Luminous Intensity	Red	I_V	I_F = 20 mA	mcd	2130	3000	
	Green	I_V	I_F = 20 mA	mcd	4180	7000	
	Blue	I_V	I_F = 20 mA	mcd	1100	2000	
	Amber	I_V	I_F = 20 mA	mcd	2130	3000	

Copyright © 2014Cree, Inc. All rights reserved. The information in this document is subject to change without notice. Cree and the Cree logo are registered trademarks of Cree, Inc.

Figure 1.5.6b: LED datasheet (continued).

Note the different values between the various colors. The forward current is specified as 50 mA for red/amber with 35 mA for green/blue. Nominal operating currents are between 10 and 30 mA. Reverse voltage is 5 volts, typical for many LEDs although much lower than the average switching diode. Forward voltage is typically 2.1 volts for the red end of the spectrum and, as expected, 3.4 volts for the green/blue end. The expected luminous intensities also vary with color. Further, it should be noted that LEDs do not produce "pure color" light in the manner of a laser. Rather, they produce a range of wavelengths clustered in a specific area. The wavelength that produces the highest output in

this area is referred to as the peak or dominant wavelength. Human vision covers the range of roughly 400 nanometers (violet) to 700 nanometers (red).[3]

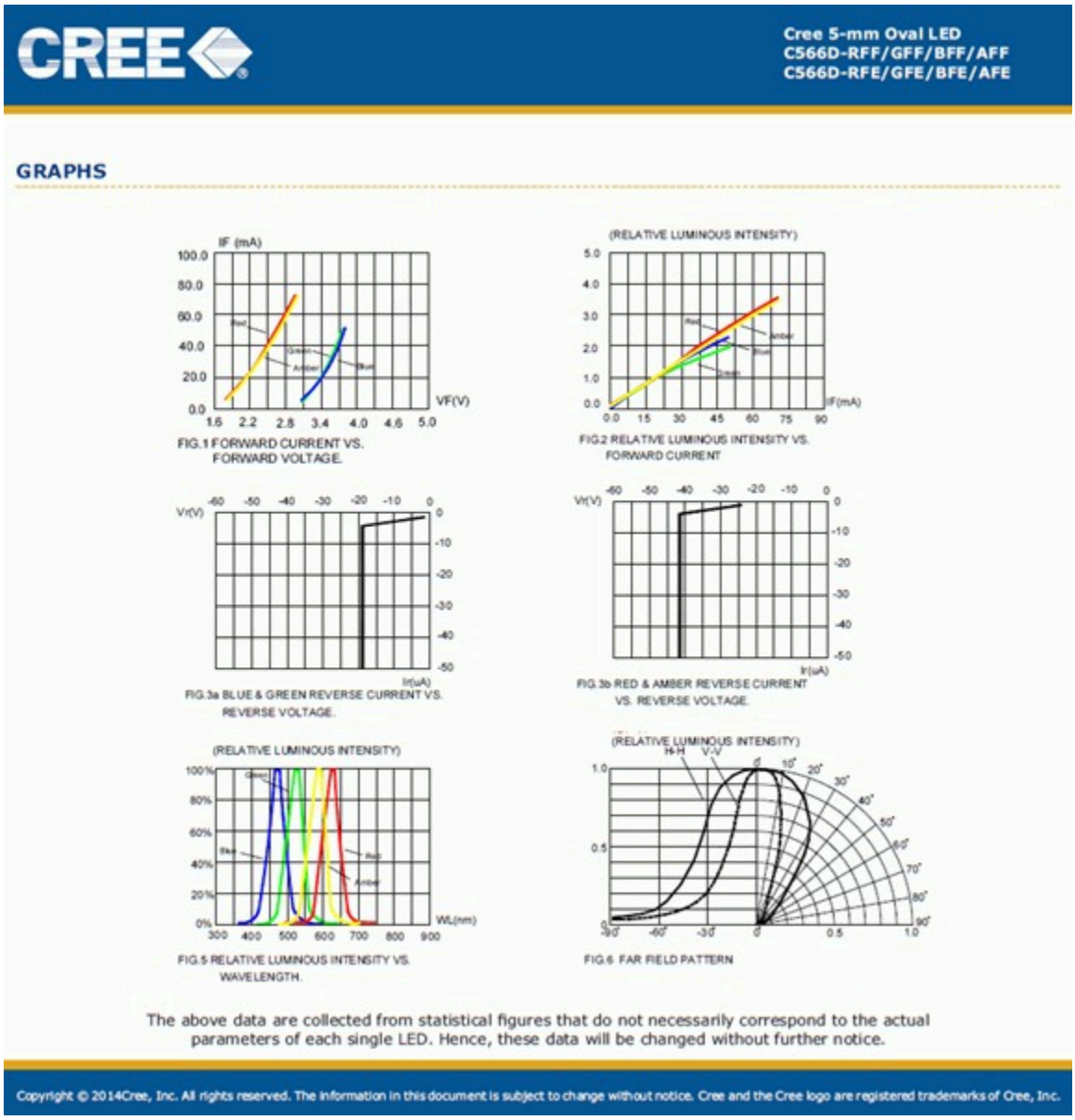


Figure 1.5.6c: LED datasheet (continued).

Figure 1.5.6*c* presents pertinent graphical data. We observe a roughly linear increase in luminous intensity with increasing current. Also, note the difference in the reverse voltage/current plots between blue/green and red/amber. Of particular interest is the final graph which shows the beam pattern or beam angle. You can think of this in terms of how narrow or broad the illumination pattern is. When comparing different model LEDs it is useful to remember that on-axis brightness can be increased by narrowing the angle. This graph is split in half using two different ways of showing the

3. It is interesting to observe that the human visual system operates over a frequency range of less than 2:1 while the human auditory system operates over a frequency range of about 1000:1 (20 hertz to 20,000 hertz). If human hearing had a range equivalent to that of our sight, we'd hear less than a full octave of pitches in total. In other words, do-re-mi-fa-sol-la-ti-do would end at ti and anything beyond would be inaudible. In such a case one thing is certain: piano keyboards would be much shorter.

data. On the left side we have a linear graph depicting the relative brightness as we move off of the center axis (zero degrees). On the right side we see a polar plot version of the same data.

Examples 1.5.2

Determine the circulating current for the circuit in Figure 1.5.7. Assume the power supply is 5 volts, the LED forward voltage is 2.1 volts and the resistor is 330 Ω.

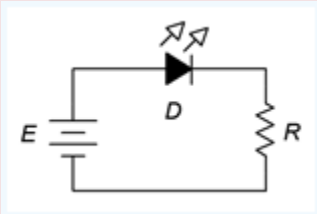


Figure 1.5.7: Schematic for Example 1.5.2.

The LED is forward-biased and as the source is greater than the LED potential, it should light. Using KVL,

$$I = \frac{E - V_{LED}}{R}$$

$$\frac{5V - 2.1V}{330\Omega}$$

$$I = 8.788mA$$

This should result in a relatively bright LED. The resistor can be used to effectively program the brightness by changing the current level (a smaller resistance yields a higher current and therefore a brighter LED). Given the 2.1 volt forward potential, it is likely that this is an amber or yellow LED. If a different color had been used, say a 1.6 volt red or 3.2 volt blue, there would be a change in current and most likely a change in brightness. The change in brightness might not perfectly echo the change in current because the conversion efficiency for the two diodes may not be the same (refer to Figure 1.5.7b to compare the luminous intensities at 20 mA for different wavelengths).Figure 1.5.8: Dual LED application.

An interesting circuit using two differently colored LEDs is shown in Figure 1.5.8. An AC source is used to drive the LEDs. Only one of the two will be forward-biased at any given time. For positive source voltages D_1 will be on while D_2 will be off. For negative source voltages the opposite will be true. Resistor R serves to limit current for both of them. Assume D_1 is red and D_2 is blue. Further, suppose the source frequency is relative low, say 1 hertz. For the positive half cycle (.5 seconds) the red LED will light and for the negative half cycle the blue LED will light. This alternating pattern continues for as long as the source is applied but a curious thing happens as we increase the frequency. At first, the blink rate will increase with the red and blue flickering back and forth faster and faster. At some point, perhaps around 30 hertz or so, it will appear as though both LEDs are lit continuously. This is because the human visual sense will tend to integrate the rapid motion and we effectively see the "average" intensity. In fact, this "on-off" trick is often used in digital circuits to control the

brightness of LEDs or the speed of motors. Bi-color LEDs are available in a single package. Using a common lead and two control leads (one for each color), it is possible to achieve color mixing.

The logical inverse of the LED is the photodiode, the schematic symbol of which is illustrated in Figure 1.5.9. The photodiode includes some manner of port that allows light to hit the junction. A sufficiently energetic photon of light can knock lose an electron. This creates an electron-hole pair which results in current flow. As more light energy is added to the system, an increasing current or voltage will result.[4]

Figure 1.5.9: Photodiode schematic symbol.

Photodiodes can operate in one of two modes. The first mode is photovoltaic mode. It uses zero-bias (that is, no external bias potential). In this mode, the photodiode operates as a voltage source. This is the mode used by photovoltaic solar cells. They can be thought of as very large photodiodes. The second mode of operation is photoconductive mode. This mode requires reverse-biasing the diode with an external potential. In this mode the diode acts more like a current source. The advantage is that the response is faster than photovoltaic mode. The downside is that noise and dark current are worse. Dark current is the current produced even when no light is shining on the photodiode. Ideally this would be zero. A large dark current reduces the effective dynamic range of the device.

SCHOTTKY AND VARACTOR DIODES

The Schottky diode is a special purpose device. It is named after Walter Schottky, a German physicist. Unlike other diodes that rely on a semiconductor-tosemiconductor junction, the schottky diode is comprised of a semiconductor-tometal contact. The Schottky diode exhibits two major advantages over traditional diodes. First, they have very fast switching times, perhaps orders of magnitude of improvement. Second, they exhibit relatively low turn on voltages. Instead of the 0.6 to 0.7 volts seen with a silicon junction diode, a Schottky diode may turn on with as little as 0.2 or 0.3 volts. Consequently, Schottky diodes are used when very fast switching speed and/or minimizing forward voltage drops are important. Examples include shunting diodes in switch mode power supplies and RF detector circuits. Its schematic symbol is shown in Figure 1.5.10.

Figure 1.5.10: Schottky diode schematic symbol.

4. As a side note, depending on their construction some LEDs can be used as crude photodiodes. Although they are not optimized for this use it can be entertaining to shine a light on an LED and watch it produce a voltage.

The varactor diode is another special purpose device. Its schematic symbol is illustrated in Figure 1.5.11. It is used as an electrically controlled capacitance (note that the schematic symbol appears as a hybrid of normal diode and capacitor symbols).

Figure 1.5.11: Varactor diode schematic symbol.

Varactors are used in reverse-bias mode. The key to understanding their operation is to consider the structure of a diode, comparing it to the construction of a capacitor. Consider the depletion region to be the dielectric of a capacitor with the anode and cathode being the capacitor plates. Consequently, all junction diodes exhibit some capacitance. Normally, designers try to minimize this effect but it is exploited with varactors. As noted in our earlier discussion, increasing the reverse-bias potential on a diode causes its depletion region to widen. All else being equal, increasing the plate separation of a capacitor decreases its capacitance. Thus, by increasing the reverse-bias potential, we increase the effective plate spacing and decrease the diode junction capacitance. We now have a capacitance the value of which is determined by a DC bias voltage. This capacitance can be used as part of electronic tuning circuits for applications such as oscillators and filters. Compared to fixed capacitors the values tend to be small, in the tens to hundreds of picofarads, but it is sufficient for much radio frequency work. The advantages over mechanically adjustable capacitors are manifold, including small size, high reliability, low cost and the ability to rapidly change the capacitance.[5]

5. The mechanical version would require a rotary-style adjustable capacitor connected to some form of small motor or solenoid to move the capacitor plates. While this can work at lower frequencies, if rapid changes are needed the resulting friction-generated heat may cause this contraption to burst into flames. Generally speaking, this is not something we want our circuits to do.

1.6 SUMMARY

In this chapter we have examined the structure and functioning of the PN junction. A PN junction produces a depletion region which is an area devoid of free charges. This leads to an energy hill or barrier voltage, the precise value of which depends on the material used as well as other factors such as temperature. The PN junction is the basis for most diodes. Its current-voltage characteristic is described by the Shockley equation and shows a logarithmic characteristic (i.e., the voltage is proportional to the log of the current).

The terminals of a diode are identified as the anode (P material) and the cathode (N material). If a positive potential which is greater than the barrier voltage is applied from anode to cathode, the diode will conduct current. If the polarity is reversed, the diode will not conduct. Therefore a simple model of the diode is a polarity sensitive switch. Improved models include the forward barrier voltage and the bulk resistance of the diode. Another refinement includes the effect of reverse breakdown, that is, the tendency of a diode to suddenly begin conducting if the reverse-bias potential is large enough. For ordinary diodes, the reverse potential should not be allowed to reach breakdown.

Besides the common switching and rectifying diodes, other types are also available. These include the Zener which is normally used in reverse-bias mode. It is commonly used to set or limit a specific voltage. In forward-bias, a Zener behaves like an ordinary diode. LEDs produce light from an electrical input. Their forward potentials tend to be in the neighborhood of a few volts. The photodiode is the complement of the LED and produces a current or voltage that scales with incident light. The Schottky diode is notable for its fast switching speeds and low barrier potential. Finally, the varactor is used as an electrically controlled capacitance. It is used in reverse-bias mode.

1.6.1: Review Questions

1. What is a depletion region?
2. Draw and explain the energy diagram for a PN junction, including the Fermi level.
3. Describe and compare the three diode models.
4. Explain the difference between the effective DC resistance and AC resistance of a diode.
5. List some of the practical differences between switching diodes, Zener diodes and LEDs.

1.7 EXERCISES

Assume diodes are silicon unless stated otherwise

ANALYSIS PROBLEMS

1. For the circuit of Figure 1.7.1 determine the circulating current if the supply is 6 volts and the resistor is 10 kΩ.

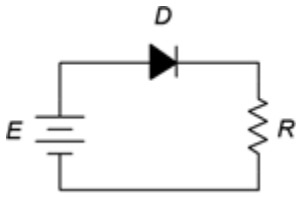


Figure 1.7.1

2. Repeat Problem 1 if the diode is inserted in the opposite orientation.
3. Given the circuit of Figure 1.7.2, determine the voltage drops across the resistors. The source is 12 volts, R_1 = 4.7 k and R_2 = 3.3 k.

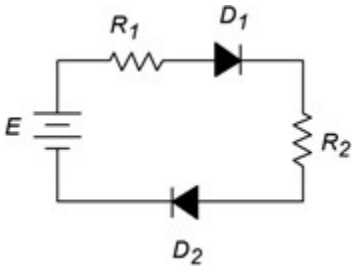


Figure 1.7.2

4. In Figure 1.7.3 determine the voltage drops across the resistors.

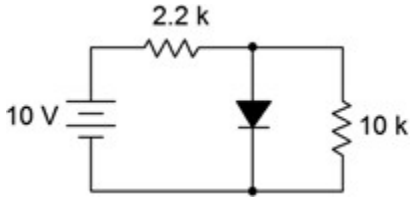


Figure 1.7.3

5. Determine the LED current in Figure 1.7.4. Assume the LED barrier is 2.1 volts, the source is 5 volts and the resistor is 330 Ω.

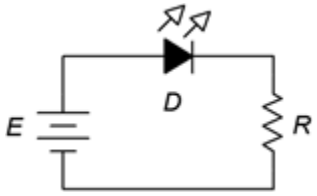


Figure 1.7.4

6. Repeat Problem 5 if the LED is inserted in reverse orientation.
7. Determine the resistor currents in Figure 1.7.5. The source is 15 volts, R_1 = 8.2 k and R_2 = 3.9 k.

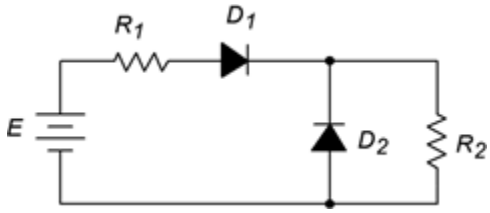


Figure 1.7.5

8. For the circuit of Figure 1.7.6, determine the resistor voltage. The source is 9 volts, the Zener potential is 5.1 volts and the resistor is 1 k.

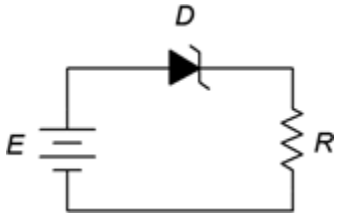


Figure 1.7.6

9. For the circuit of Figure 1.7.7, determine the resistor voltage. The source is 8 volts, the Zener potential is 3.3 volts and the resistor is 10 k.

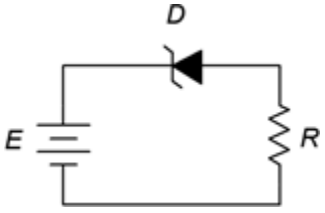


Figure 1.7.7

10. Determine the voltage across R_2 in Figure 1.7.8 if the source is 9 volts, the Zener is 6.8 volts, R_1 = 5.1 k and R_2 = 33 k.

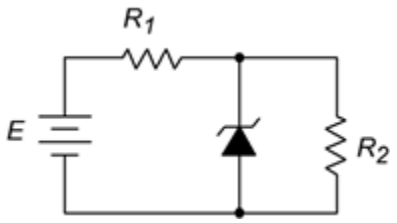


Figure 1.7.8

CHALLENGE PROBLEMS

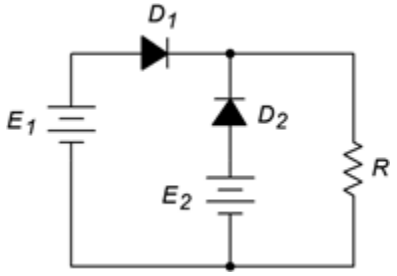


Figure 1.7.8

11. Determine the resistor voltage in Figure 1.7.9 if E_1 = 5 volts, E_2 = 9 volts and R= 1 k.

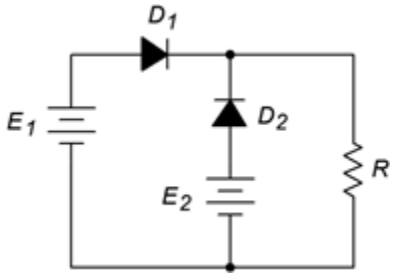


Figure 1.7.9

12. Determine the voltage across R_2 in Figure 1.7.8 if the source is 9 volts, the Zener is 5.6 volts, R_1 = 5.1 k and R_2 = 3.9 k.

DESIGN PROBLEMS

13. Determine a value for R in Figure 1.7.1 to set the current to 10 mA if the source is 5 volts.
14. Determine a value for R in Figure 1.7.4 that will set the LED current to approximately 20 mA if the source is 9 volts and the LED is a standard red type. Use a standard resistor value.

COMPUTER SIMULATION PROBLEMS

15. Simulate Problem 9.
16. Simulate the circuit designed in Problem 14 for verification.

1.8 AN ALTERNATE HYPOTHESIS REGARDING PN JUNCTIONS

And now for something completely different…
One of the great things about the Internet is that you can find almost anything on it. In contrast, one of the terrifying things about the Internet is that you can find almost anything on it. The following is presented in keeping with the dictum that "It must be true because I saw it on the Internet".

AN ALTERNATE HYPOTHESIS REGARDING PN JUNCTIONS

In order to learn how to design circuits and systems using transistors and other solid state devices, students of electronics are told in their courses how semiconductors function. The atomic structure of crystalline silicon is examined in its intrinsic and doped states. Discussion of energy levels, conduction band electrons and hole production quickly follow. Soon, the student encounters the PN junction, a basic building block of modern electronics, and learns about majority and minority carriers, depletion regions, barrier potentials, leakage current and other exotica. This information is intended to explain just how solid state devices really work and it can end up sounding quite obtuse. It may sound so complicated, in fact, that the student assumes that only a genius could design such devices and relegates him or herself to a lower paying engineering or technician job. This should not be the case!

Recent investigations by Farcebook authors have uncovered some startling facts:

- Semiconductors don't really work the way we've all been told. In fact, the fundamental theory is much simpler.
- This lie has been fabricated and perpetuated by the economic and political elite, a group of people with cushy, high paying jobs; jobs so easy in the light of the real theory that even CEOs, hedge fund managers, televangelists, TV psychics and other folk of nil capability could do it in their sleep (well, OK, the executives would still need an army of assistants and the televangelists would continue to be outraged by your private life, but you get the general idea). By making their jobs sound difficult these people get to sit around all day eating eclairs and reading Esquire for an eight figure annual income.

It is high time that the truth be told and this farcical sham be torn down! As an example, we shall see how a simple diode really works.

So, you think a diode is composed of semiconducting material? Think again! One of the chief researchers at Bell labs in the 1940s and 50s was a certain Doctor Schlocking. After several experiments involving solid state diodes, Dr. Schlocking wrote in his diary:

> "This stuff don't work at all. Better go back to tubes before they can me. Ooops, must let the dog out."

Schlocking was often pestered by his dog Melvin who reminded him of a small selfpropelled dust mop, and who was approximately as clean. The diary continued:

"If only there was a way in which Melvin could let himself out. Even better…if he couldn't get back in!"

This ominous tone led to Schlocking's invention of the now infamous one-way doggie door that can be seen mounted to the bottom of normal doors across the country. Schlocking knew how to milk an idea and took the form to its extremes by developing the cat door, the mouse door, the grasshopper door, the flea door (an early attempt at a flea and tick collar) and even the amoeba door. This last unit when properly designed could force microscopic parasites and bacteria out of the human body and not allow them to re- enter. This was instrumental in the development of the polio vaccine, in spite of the fact that the vaccine was produced some years earlier. Schlocking's greatest achievement, however, came when he shrunk the doggie door still further to produce the electron door. This is the fundamental unit of modern electronics.

In the figure below, we see a cross section of a diode and close-ups.

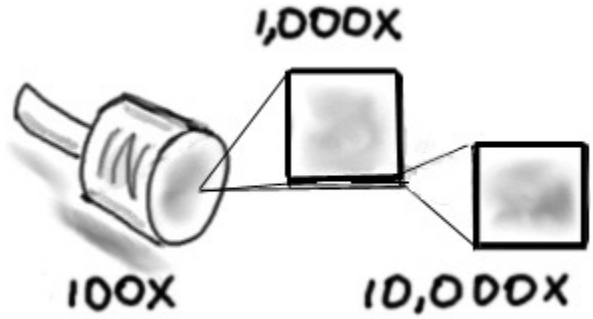


Figure 1.8.1

Even at 10,000X magnification we can still see nothing of the PN junction. If we go a bit further, something interesting comes into focus (see the second figure).

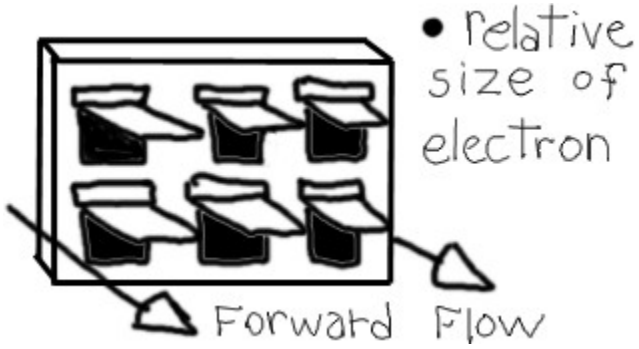


Figure 1.8.2

Yes! A PN junction is nothing more than a huge array of real tiny one-way doggie doors! Here's how it works: Electrons are a lot like marbles. When one hits a doggie door from behind, the door flips open allowing the marble through (i.e., allowing current to flow). If the electron hits the doggie door from the front, the flap closes and the electron can't get through (i.e., no current flow). Now obviously, if we hang the diode vertically, gravity should open all the doors and we'll get lots of electrons (i.e., current flow) in either direction. In truth, a real diode doesn't do this. Its operation will not matter on how the diode is oriented in space. This feature is accomplished by simply adding a small coil spring to the

doggie door's hinge, forcing it to stay shut in the face of gravity. This has the negative side effect of requiring somewhat higher energy levels from the electrons to force the door open.

This force happens to be the barrier potential of the diode! It has nothing to do with so-called depletion regions. If you were an electron, would you want to go through a place called a depletion region? Of course not! Neither would electrons. They're not stupid, you know. In any case, the stronger the spring, the greater the barrier potential. Presently diodes are made of either silicon or germanium with barrier potentials of approximately 0.7 volts or 0.3 volts, respectively. In fact, silicon and germanium are really code words meaning strong spring and weak spring! It took a while to develop small, strong springs and this is why germanium diodes were the first ones built.

Note that spring strength also plays a role in how tightly the flap can shut thus indicating the reverse leakage current. Here again we see strong spring "silicon" units having lower leakage. Theory also indicates that leakage should increase with temperature. This effect can be seen clearly in the doggie door model. At present it is impossible to create both the frame and the door out of precisely the same material and thus two different expansion coefficients exist. Because the flap is smaller than the frame it will tend to curl away at higher temperatures allowing more electrons to sneak through the gaps. At very low temperatures the flap tends to stick to the frame in much the same fashion that your tongue or lips will stick to a metal flag pole in freezing weather (also known as the Christmas Story phenomenon).

At very high forward energy levels the flaps may be literally torn off their hinges. This high volume of electrons at high energy will yield the maximum forward current. Also, note that if the energy level is high enough in the reverse direction, either the flaps will be bent and pushed through the frames or they will start to bounce violently at resonance, allowing electrons through. These two modes are referred to as avalanche and Zener conduction, respectively. The required energy level indicates the reverse breakdown voltage.

Other fine points can be explained equally by the doggie door model, as well as bipolar and field effect transistors, IGBTs and just about everything else in the field of solid state electronics with the exception of the original 7400 series TTL logic gates which utilized an array of small, edible fungi and miniature harvester ants.

More on that in a future exposé.

UNIT 2: DIODE APPLICATIONS

Learning Objectives

After completing this chapter, you should be able to:

- Solve basic AC rectifier circuits for resulting waveforms.
- Detail the differences between half-wave, full-wave and full-wave bridge diode rectifier configurations.
- Solve basic regulator circuits employing Zener diodes.
- Outline a complete AC-to-DC power supply with regulation, describing the function of each component, including power transformer.
- Solve AC clipper circuits for output waveforms.
- Solve AC clamper circuits for output waveforms.

2.1 INTRODUCTION

The preceding chapter was concerned primarily with introducing the practical considerations of diodes while presenting them in DC circuits. This chapter will extend the discussion by focusing on AC circuit applications. A prime example is AC to DC conversion, the concept behind most electronic power supplies. It also includes the basics behind regulation and limiting/level shifting circuits such as clippers and clampers. The inherent asymmetry in the conductance of diodes, that is, their sensitivity to the direction of current flow, is what makes these circuits possible. Non-ideal effects such as a diode's forward voltage drop might be ignored in some instances but may be quite important in others.

2.2 RECTIFICATION

Rectification is the process of turning an alternating current waveform into a direct current waveform, i.e., creating a new signal that has only a single polarity. In this respect it's reminiscent of the common definition of the word, for example where "to rectify the situation" means "to set something straight". Before continuing, remember that a DC voltage or current does not have to exhibit a constant value (like a battery). All it means is that the polarity of the signal never changes. To distinguish between a fixed DC value and one that varies in amplitude in a regular fashion, the latter is sometimes referred to as pulsating DC.

The concept of rectification is crucial to the operation of modern electronic circuits. Most electronic devices such as a TV or computer require a fixed, unchanging DC voltage to power their internal circuitry. In contrast, residential and commercial power distribution is normally AC. Consequently, some form of AC to DC conversion is required.[1] This is where the asymmetry of the diode comes in.

HALF-WAVE RECTIFICATION

To understand the operation of a single diode in an AC circuit, consider the diagram of Figure 2.2.1. This is a simple series loop consisting of a sine wave source, a diode and a resistor that serves as the load. That is, primarily we will be interested in the voltage developed across the resistor.

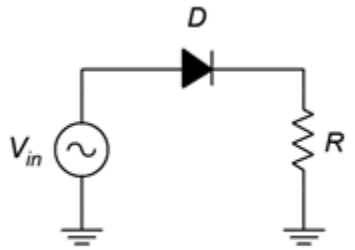


Figure 2.2.1: Basic AC diode-resistor circuit.

For positive portions of the input wave, the diode will be forward-biased. To a first approximation it will appear as a closed switch. Consequently, all of the input signal will drop across the resistor. In contrast, when the input signal switches to a negative polarity on the other half of the waveform, the diode will be reverse-biased. Therefore, the diode acts as an open switch. The circulating current drops to zero thereby producing no voltage across the resistor. All of the applied potential drops across the diode, as indicated by Kirchhoff's voltage law (KVL). The input and load resistor's voltage waveforms can be seen in Figure 2.2.2.

1. If you're wondering why we don't just use DC distribution instead in order to "cut out the middle man", the reasons are manifold. First, it is generally more efficient to distribute power via AC rather than DC. Second, even if DC is available, it may not be at the amplitude the circuitry requires. Therefore some form of DC-to-DC conversion would be needed. Depending on the application, this can turn out to be more expensive than AC-to-DC conversion.

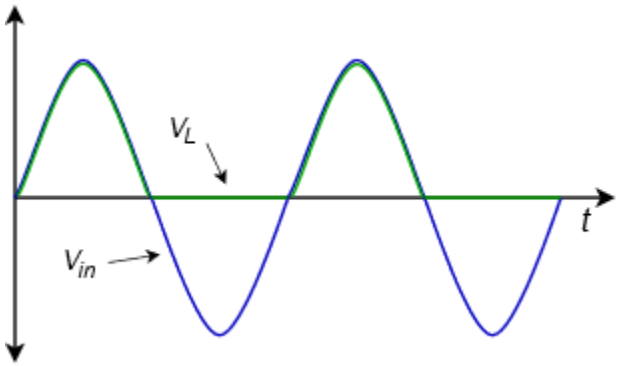


Figure 2.2.2- Half-wave rectification waveforms.

The resulting signal seen across the load resistor is a pulsating DC waveform. We have effectively removed the negative half of the waveform leaving just the positive portion. Because only half of the input waveform makes it to the load, this is referred to as half- wave rectification.

It is worth noting that if the AC peak input voltage is not particularly large, there can be an obvious discrepancy between the peak levels of the input and load signals. For example, if the peak input voltage is in the range of three or four volts and a silicon diode is used, the resulting waveforms would look more like Figure 2.2.3.

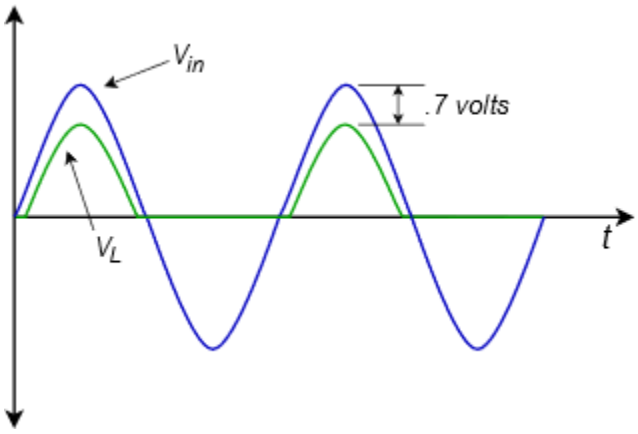


Figure 2.2.3: Half-wave rectification waveforms including forward diode drop.

In this case the 0.7 volt forward drop cannot be ignored as it represents a sizable percentage of the input peak. The positive pulses are also slightly narrowed as current will not begin to flow at reasonable levels until the input voltage reaches 0.6 to 0.7 volts.

If the diode was oriented in reverse, it would block the positive portion of the input and allow only the negative portion through. In this instance the load waveform would appear flipped top to bottom compared to Figures 2.2.2 and 2.2.3.

COMPUTER SIMULATION

A simulation schematic for a simple half-wave rectifier is shown in Figure 2.2.4. A sine wave source of 10 volts peak is used to feed a popular 1N4000 series rectifier diode connected to a 100 Ω load. The source frequency is 60 hertz, the North American standard for power distribution.

A transient analysis is run resulting in the waveforms shown in Figure 2.2.5. The source voltage waveform is shown in red while the load voltage waveform is depicted in blue. While the half-wave rectification is obvious, the loss due to the forward voltage drop of the diode is clearly evident. Based on the vertical scale, a value just under one volt would be a reasonable estimate. The simulation agrees nicely with the expected result as drawn in Figure 2.2.3, although not as extreme due to the increased source voltage.

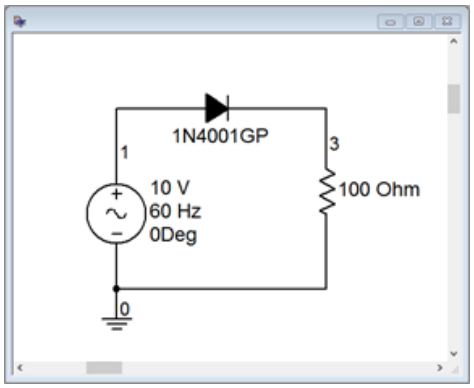


Figure 2.2.4: Simulation schematic for halfwave rectifier.

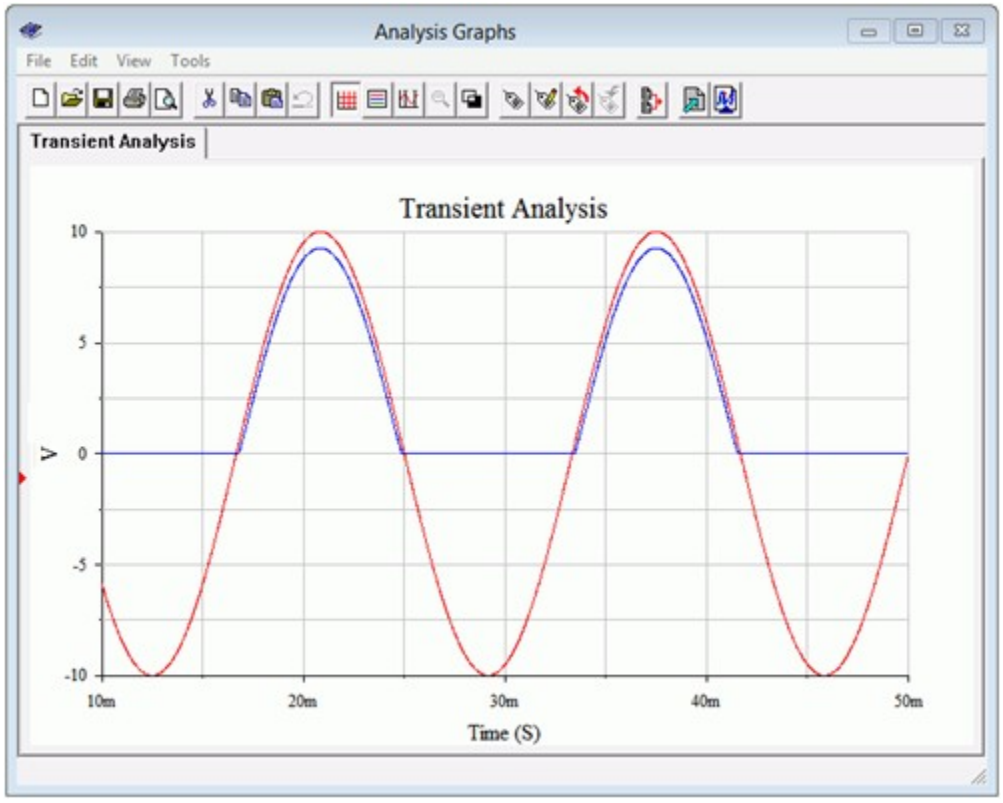


Figure 2.2.5: Transient analysis for halfwave rectifier.

On a practical note, there are still two items to consider when it comes to converting AC to DC. The first item is the issue of scaling the 120 VAC RMS outlet voltage to a more useful level. In many cases this means lowering the voltage although there are some applications such as high power amplifiers where the voltage will need to be increased. The second item involves smoothing the pulsating DC to produce a constant value, much like a battery.

A NOTE REGARDING TRANSFORMERS

The aforementioned voltage scaling issue can be addressed through the use of a transformer. While a complete exploration of transformers is beyond the scope of this chapter, we can present the basics. In simple terms, a transformer has an input side, or primary, and an output side, or secondary. Each side is made up of a coil of wire and these coils are wound around a common magnetic core. The current in the primary-side coil creates a magnetic flux in the core. This flux induces a current in the secondary coil. Ideally, the voltage is decreased and the current is increased by the ratio of the number of loops between these coils. For example, if the secondary-side coil has half as many turns as the primary-side coil then the secondary voltage will be half of the primary voltage and its current will be twice as large as the primary current. This implies that in the ideal case there is no power lost within the transformer. It simply transforms the power from high-voltage/low-current to low-voltage/high-current (or vice versa), hence the name. In reality, transformers do have voltage and current limits, and they are specified in terms of a volt-amp or VA rating which is simply the product of the nominal secondary voltage and maximum allowed secondary current. Transformers that decrease the voltage are referred to as step-down while those that increase the voltage are referred to as step-up. Finally, it is possible to create transformers with multiple primaries and secondaries (via either separate coils or multi-tapped coils). The resulting series and parallel coil configurations make them much more flexible.

SMOOTHING (FILTERING) THE OUTPUT

The second issue we have is smoothing and leveling the pulsating DC. The most straightforward method to achieve this is to add a capacitor in parallel with the load. The capacitor will charge up during the conduction phase, thus storing energy. When the diode turns off, the capacitor will begin to discharge, thus transferring its stored energy into the load. The larger the capacitor, the greater its storage capacity and the smoother the load voltage will be. It turns out that there is a down side to large capacitors, as we shall see. Consequently, the goal will not be to use as large of a capacitor as possible but rather to use an optimal size for a given application. A half-wave rectifier with transformer and capacitor is shown in Figure 2.2.6.

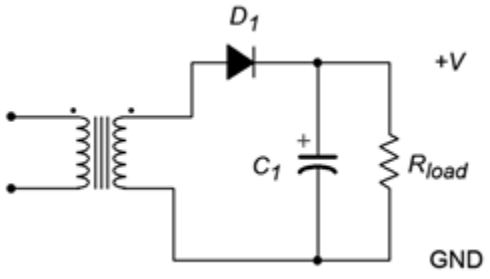


Figure 2.2.6: Half-wave rectifier with transformer and filter capacitor.

One way of looking at the inclusion of the smoothing capacitor is to consider that it, along with the load resistance, make up an *RC* discharge network. To achieve a smooth load voltage the discharge time constant should be much longer than the gap produced when the diode turns off. For 60 hertz operation, this gap is half of the period, or roughly 8.3 milliseconds. The time constant equation is

Recalling that in one time constant the capacitor voltage will fall to well below half of the starting value (roughly 37%), we will need a time constant several times larger than 8.3 milliseconds. For example, suppose our effective load resistance is 100 Ω. If we use a 1000 μF capacitor, the resulting time constant would be 100 milliseconds, or over ten times the gap duration. A much smaller capacitor, say around 50 μF, would not be nearly so effective at keeping the voltage constant.

The variation in output voltage due to capacitor discharge is referred to as ripple. It can be modeled as an AC voltage riding on a larger DC output. The magnitude of the ripple worsens as the load current increases. Under light load conditions, the output will tend to float to the peak voltage of the secondary with very little ripple. As load current demand goes up, the ripple magnitude increases and the nominal output voltage begins to drop.

COMPUTER SIMULATION

Two variations on a filtered half-wave rectifier are simulated below. Both versions use a 100 Ω load with a 10 volt source, similar to the prior simulation. The first version uses a 50 μF filter capacitor while the second ups this to 1000 μF. In both cases a 1 Ω resistor is added in series with the capacitor to serve as a current sensor. The first version is shown in Figure 2.2.7.

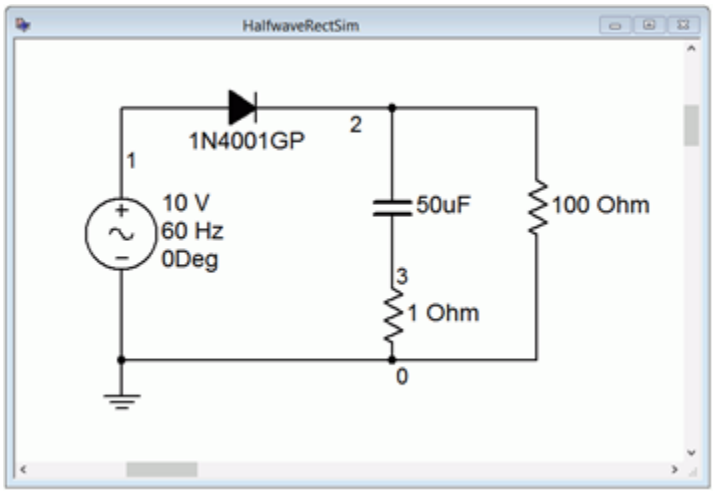


Figure 2.2.7: Simulation schematic for half-wave rectifier with 50 µF filter capacitor.

A transient analysis simulation graph is shown in Figure 2.2.8. The input waveform is colored blue while the load voltage is red. Comparing this waveform to that depicted in Figure 2.2.5 shows the effect of the capacitor stretching out the pulse and partially filling in the gap. It is obvious that this capacitor is too small given the load resistance and the resulting current demand. Indeed, by the time the next pulse arrives the capacitor is nearly depleted and the output voltage has dropped to around one volt.

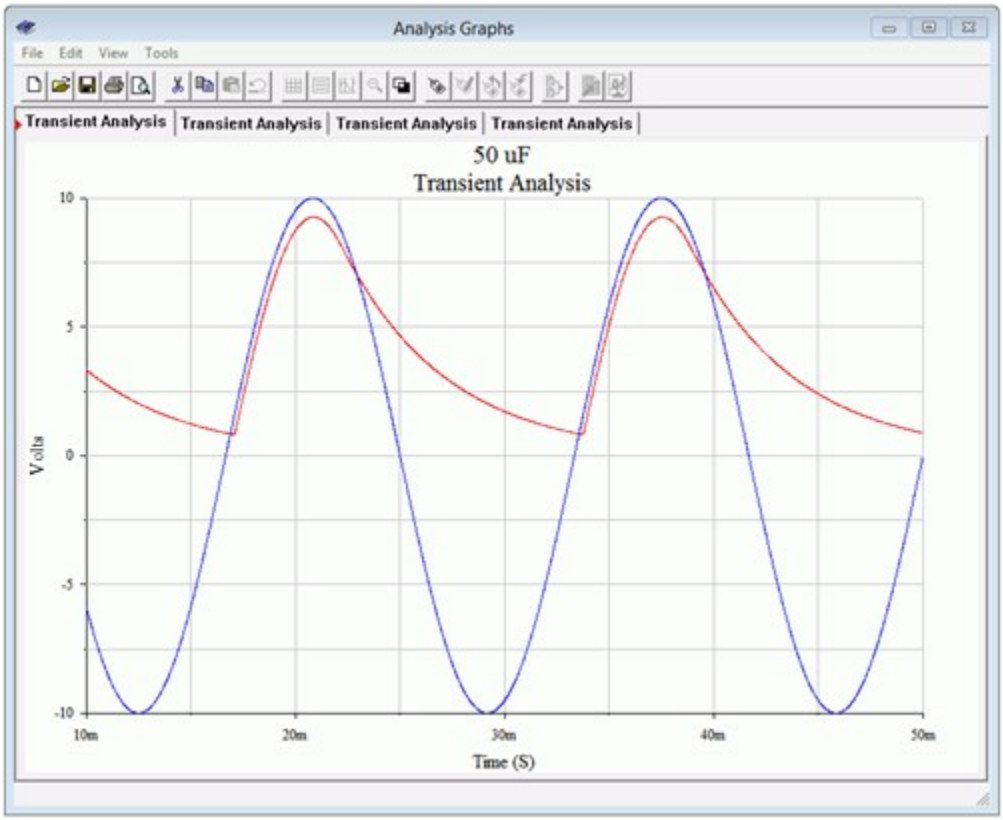


Figure 2.2.8: Transient analysis simulation for half-wave rectifier using a 50 μF filter capacitor.

In Figure 2.2.9 the simulation is rerun, but this time using a 1000 μF capacitor in place of the 50 μF. As expected, the increased *RC* time constant results in a much more stable load voltage. In this version the output has dropped from a little over nine volts to about eight volts yielding a peak-to-peak ripple of a volt and a half or so. The peak voltage of just over nine volts versus the applied ten volts is largely due to the voltage drop across the rectifying diode.

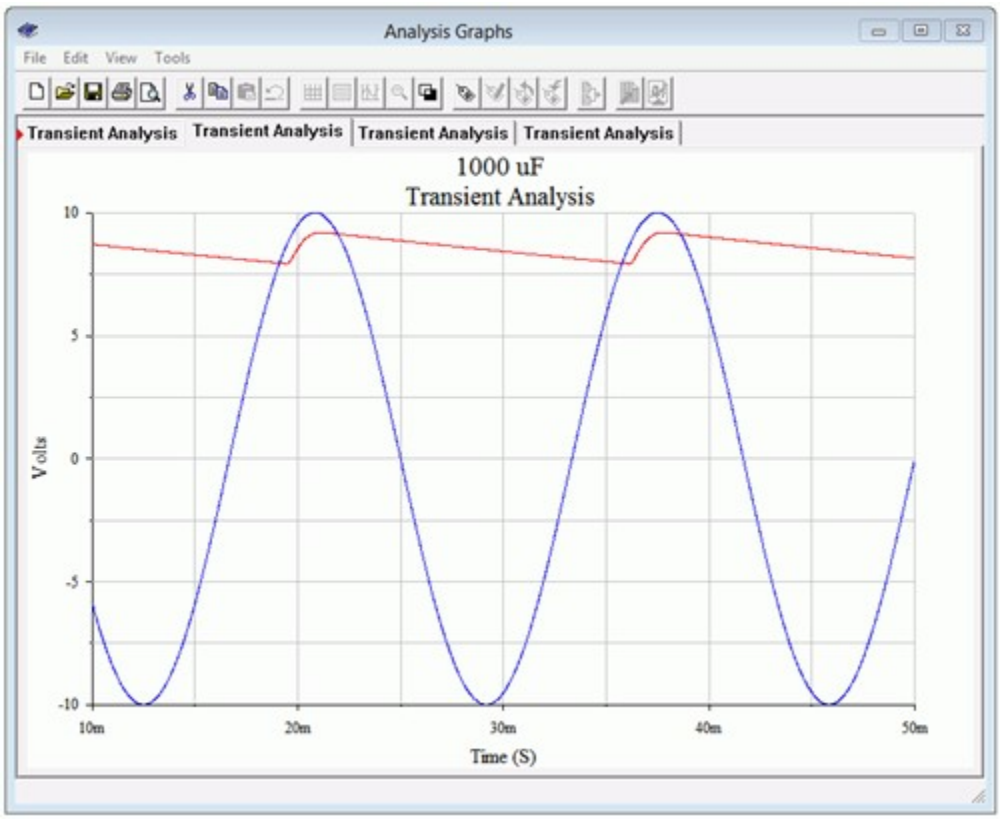


Figure 2.2.9: Transient analysis simulation for half-wave rectifier using a 1000 μF filter capacitor.

One thing that may not be apparent immediately is that the charge time for the larger capacitor is much shorter than for the smaller unit. This is perhaps counterintuitive. With a larger capacitor, the diode turns on for a shorter time because its cathode is held at a high voltage due to the capacitor. That is, it will only turn on when the input voltage exceeds the capacitor voltage by roughly 0.7 volts. It is only during this time that the capacitor will be replenished, and this can lead to very large current spikes.

To investigate this effect, the simulations are rerun, but this time adding the voltage across the 1 Ω sensing resistor. This relatively small value will have only a modest effect on the charging and discharging, and conveniently scales to the current value (i.e., 100 millivolts signifies 100 milliamps). First, examine the transient simulation of Figure 2.2.10using the 50 μF capacitor.

The red sweep is the output voltage while the blue sweep represents the capacitor current. The output voltage plot uses the left vertical axis while the current plot uses the right vertical axis. As the load voltage begins to rise, we see an abrupt spike in the capacitor current. This is current charging the capacitor and it peaks at about 180 milliamps. The total time for the charge phase is around 4 milliseconds. Once the output voltage peaks, the capacitor starts to discharge into the load. During the discharge phase note that the capacitor current's polarity has reversed. It is negative, peaking at roughly –80 milliamps, and delivering current to the load.

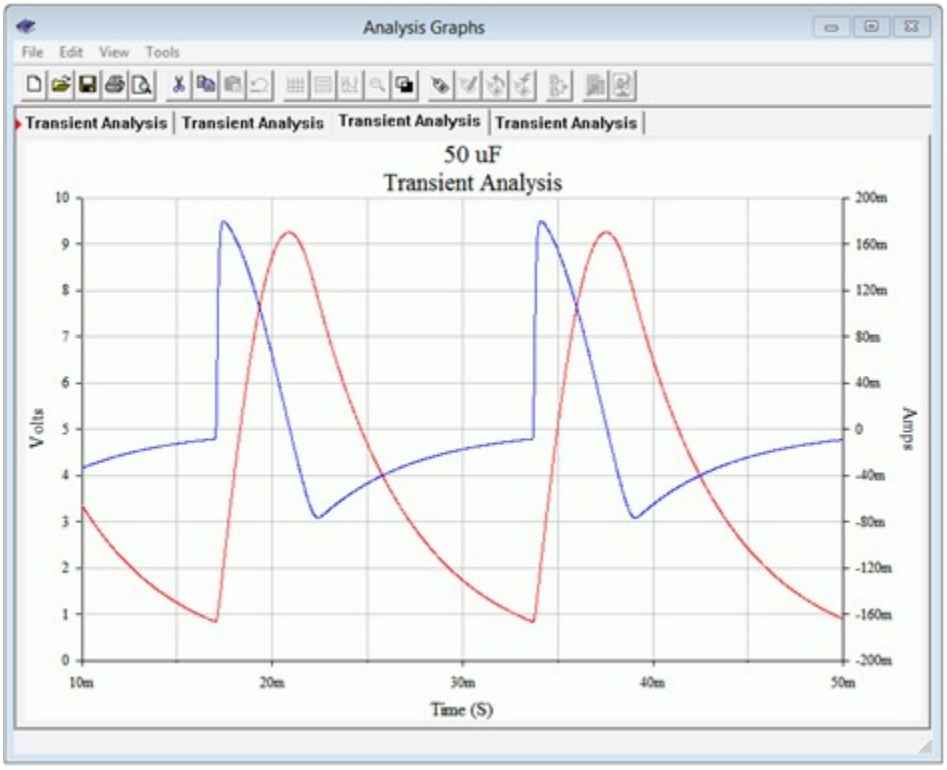


Figure 2.2.10: Transient analysis current waveform using a 50 μF filter capacitor.

This simulation is repeated using the 1000 μF capacitor. The results are shown in Figure 2.2.11.

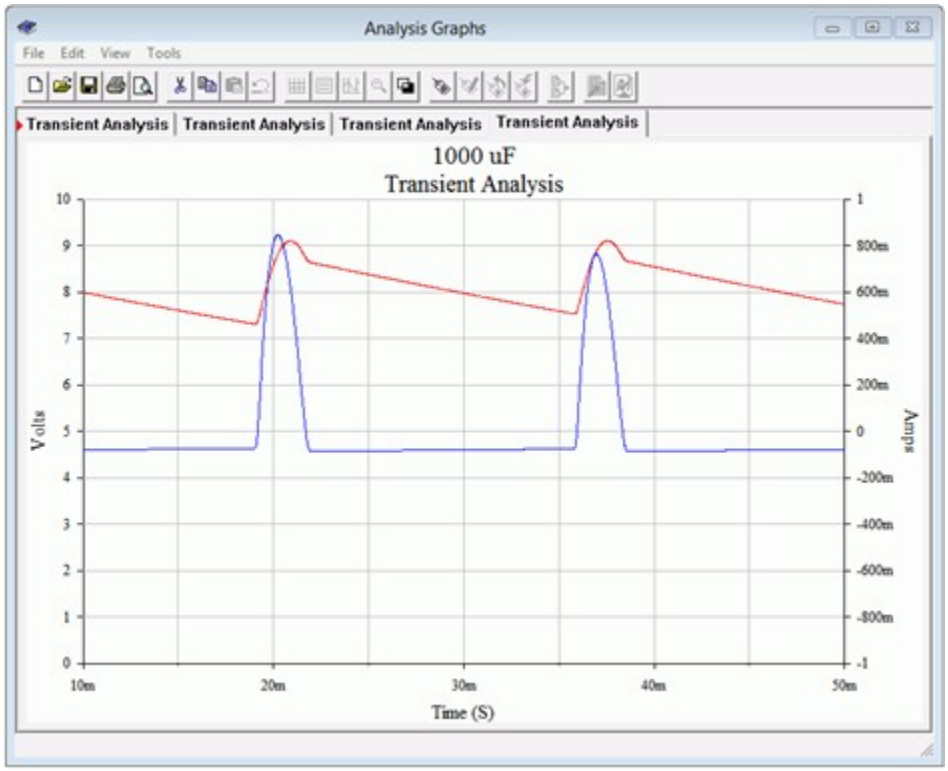


Figure 2.2.11: Transient analysis current waveform using a 1000 μF filter capacitor.

The blue current waveform peaks at approximately 800 milliamps, or over four times the value compared to using the smaller capacitor. Also, the width of the positive pulse has decreased to about

2.5 milliseconds. The discharge phase is nearly flat, implying that the output voltage must be more stable as this capacitor is the only source for load current during this phase.

FULL-WAVE RECTIFICATION

An improvement on half-wave rectification is full-wave rectification. Half-wave rectification is inefficient because it essentially throws away the negative portion of the input. In contrast, full-wave rectification makes use of the negative portion by inverting or flipping its polarity. The resulting circuit is modestly larger and more complicated but results in large performance improvements. For example, filter capacitor size is greatly reduced.

There are two popular methods to achieve full-wave rectification. The first method uses a pair of diodes with a center-tapped (i.e., split) secondary. The second method uses a four diode bridge network. The diode bridge form is also capable of producing a bipolar output (i.e., a positive output along with a negative output, typically of the same magnitude).

The two diode center-tapped secondary circuit is shown in Figure 2.2.12. This schematic also includes the filter capacitor.

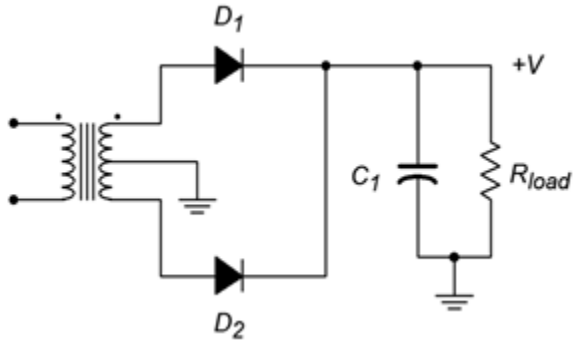


Figure 2.2.12: Full-wave center-tapped rectifier with capacitor.

The operation is as follows. During the positive half of the source voltage diode $D1$ is forward-biased while $D2$ is reverse-biased. Therefore the upper half of the secondary behaves like a simple half-wave rectifier allowing current to flow through $D1$ and into the load. Due to the reverse-bias on $D2$, the lower half presents an open circuit and is effectively removed. In mirror fashion, when the applied potential switches polarity $D1$ will be reverse-biased while $D2$ becomes forward-biased. Current is now free to flow through $D2$ into the load. Thus, both halves of the input waveform are used. The resulting waveforms are illustrated in Figure 2.2.13. For clarity, the filtering effect of the capacitor is not shown and V_{in} represents one half of the total secondary voltage.

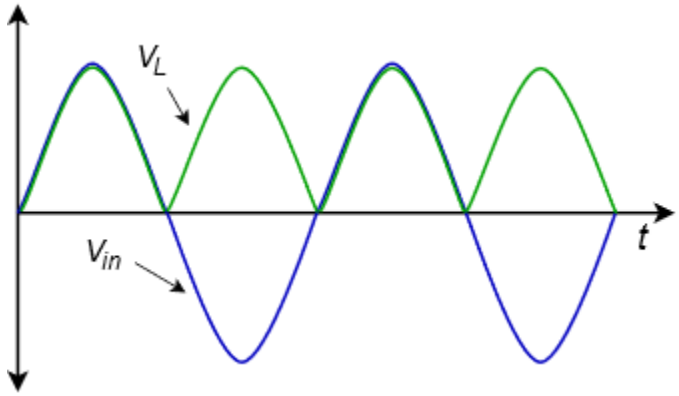


Figure 2.2.13: Full-wave rectifier waveforms.

An important point to remember about this configuration is that the load only "sees" half of the secondary at any given time. Therefore, the load voltage will only be half of the total secondary voltage (minus one forward diode drop). For example, if the transformer has a 10:1 turns ratio and is being fed from a standard 120 volt source, the secondary will produce 12 volts RMS. Ignoring the diode drop, the load would see half of this, or 6 volts RMS (about 8.5 volts peak). Typically, transformers are rated by their total secondary voltage so this transformer would be referred to as having a "12 volt center-tapped secondary".

A four diode bridge rectifier is shown in Figure 2.2.14. A filter capacitor is included. Also, note the usage of a standard, non center-tapped secondary. As this is a very common configuration, the four diode bridge is available as a single four-lead part in a variety of sizes and current capacities.

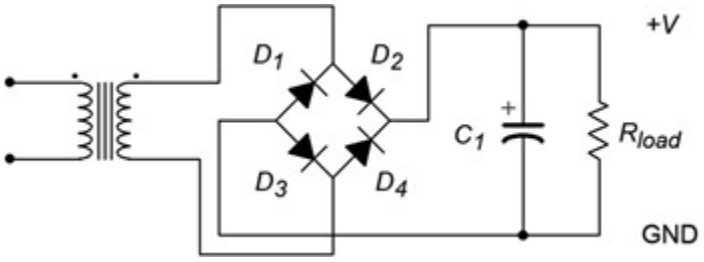


Figure 2.2.14: Full-wave bridge rectifier with capacitor.

The operation of this circuit is illustrated in Figure 2.2.15 for the positive portion of the input. First, current flows from the top of the secondary to the *D*1 /*D*2 junction. Only *D*2 offers a forward-bias path so current flows through *D*2 to the junction with *D*4 and the load. As *D*4 presents a reverse-bias path, current must flow down through the load. From ground, current continues to the *D*1 /*D*3 junction. Although at first glance it appears that current could flow through either diode, remember that the cathode of *D*1 is tied to the high side of the secondary. Therefore, its potential must be higher than the anode side, making it reverse-biased. Consequently, the current flows down through *D*3 . A similar situation occurs at *D*4 and current is directed back to the low side of the secondary. In short, *D*2 and *D*3 are forward-biased while *D*1 and *D*4 are reverse-biased. The load sees the entire secondary voltage minus two forward diode drops.

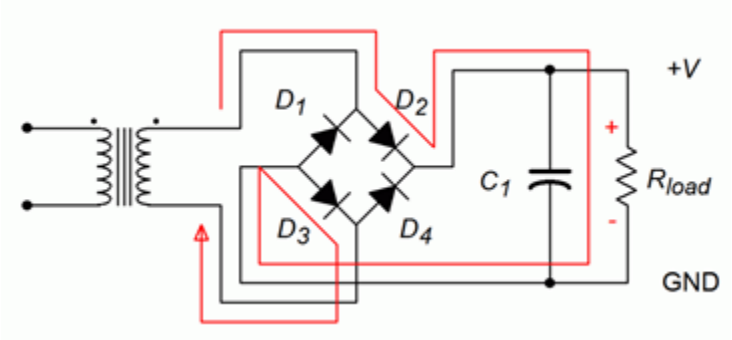


Figure 2.2.15: Full-wave bridge rectifier analysis, positive input.

During the negative polarity portion of the input the situation is reversed as illustrated in Figure 2.2.16. Current will flow from the bottom of the secondary through *D*4 , down through the load, and finally back to the top of the secondary via *D*1 . Thus, *D*1 and *D*4 are forward-biased while *D*2 and *D*3 are reverse-biased. The important thing is that in both cases, the current flows down through the load, top to bottom, resulting in a positive output voltage.

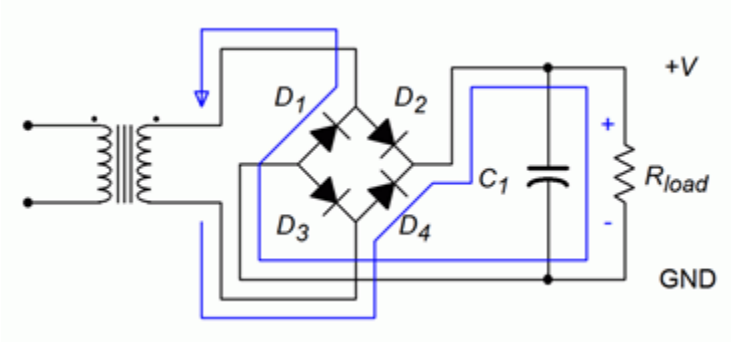


Figure 2.2.16: Full-wave bridge rectifier analysis, negative input.

Example 2.2.1

Design a rectifier/filter that will produce an output voltage of approximately 30 volts with a maximum current draw of 300 milliamps. It is to be fed from a 120 VAC RMS source. The ripple voltage should be less than 10% of the nominal output voltage at full load.

For this design we shall focus on using common off-the-shelf parts. As we have seen, the full-wave rectifiers are more efficient at converting AC to DC so we shall go that route, specifically, a four diode bridge arrangement. We will use the circuit of Figure 2.2.14 as a guide.

The first item to consider is the size of the transformer. A 30 volt output would require a peak secondary voltage of at least 32 volts as we must add in two forward diode drops. The equivalent RMS value is $32/\sqrt{2}$ or 22.6 volts. At full load the filtered output voltage will droop somewhat so a somewhat larger value is called for. A standard 24 volt secondary should suffice. Given the 300 milliamp load current rating, the transformer must be at least 0.3 amps 24 volts or 7.2 VA.

As far as the capacitor is concerned, it must be rated for the peak voltage. The peak equivalent is 24 VAC RMS $\cdot\sqrt{2}$ or 34 volts. Although a 35 volt rated capacitor might be tried, a standard 50 volt rating would

leave a generous safety margin and increase reliability. To find the capacitance value we must first find the effective worst case load impedance.

$$R = \frac{V_{out}}{I_{max}}$$

$$R = \frac{30V}{0.3A}$$

$$R = 100\Omega$$

It will be useful to compare this back to the simulation depicted in Figure 2.2.9. Our ripple specification is somewhat tighter than that achieved in the prior simulation. This is apparent by noting how far the output voltage has dropped by midway through the off portion of the cycle. Consequently, we will need a larger time constant, perhaps by a factor of two. That puts us at 200 milliseconds.

$$\tau = RC$$

$$C = \frac{\tau}{R}$$

$$C = \frac{0.2s}{100/Omega}$$

$$C = 2000$$

A 2200 μF standard value should be sufficient.

COMPUTER SIMULATION

To verify our results, the design from Example 2.2.1 is simulated. The schematic is shown in Figure 2.2.17. To simplify the simulation, a 24 volt RMS source is used in place of the transformer. The worst case load is simulated via a 100 Ω resistor. For the initial test the filter capacitor is omitted so that we can ensure the proper peak voltage and waveforms are created. The results of a transient analysis are shown in Figure 2.2.18. The secondary voltage is shown in red while the load voltage is shown in blue. The full-wave waveform is exactly as expected, including a slight reduction in the peak voltage value due to two forward diode drops. The output peak is just above 30 volts, as desired.

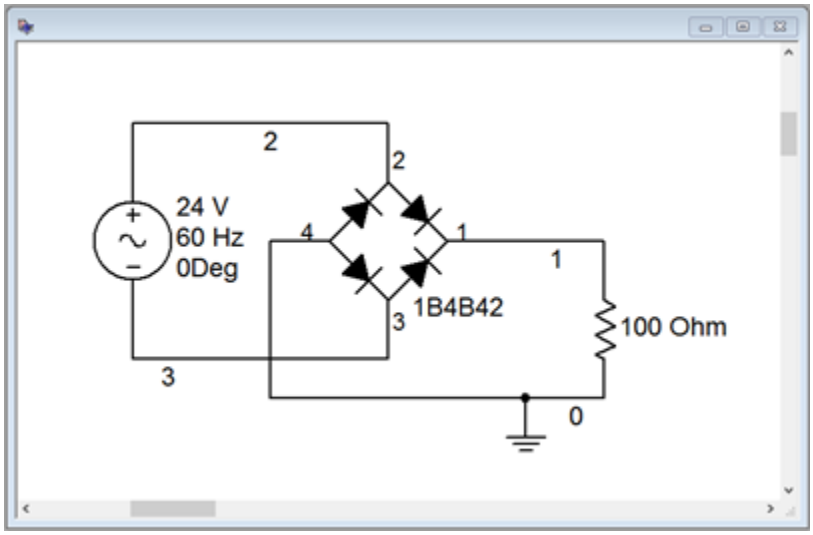


Figure 2.2.17: Simulation schematic for the design of Example 2.2.1 without capacitor.

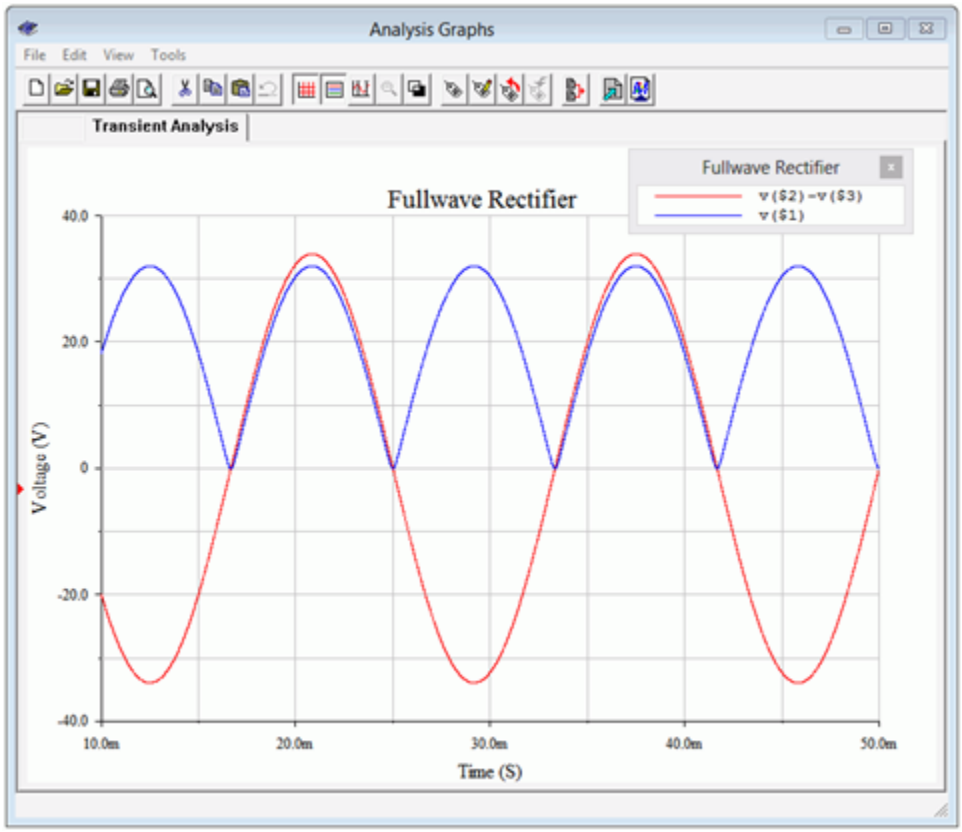


Figure 2.2.18: Transient analysis of the design of Example 2.2.1 without capacitor.

Now that we have confidence in the voltage level and waveform, the output filter capacitor is added as shown in Figure 2.2.19. A transient analysis is run again with the resulting input and load voltage waveforms depicted in Figure 2.2.20. The load voltage is shown in red. The average value is just over 30 volts and the peak-to-peak ripple is less than two volts, as desired. Note that the full-load peak voltage with the capacitor is slightly less than what was seen in the capacitor-less version. If the load current demand were to increase, both droop and ripple would get worse.

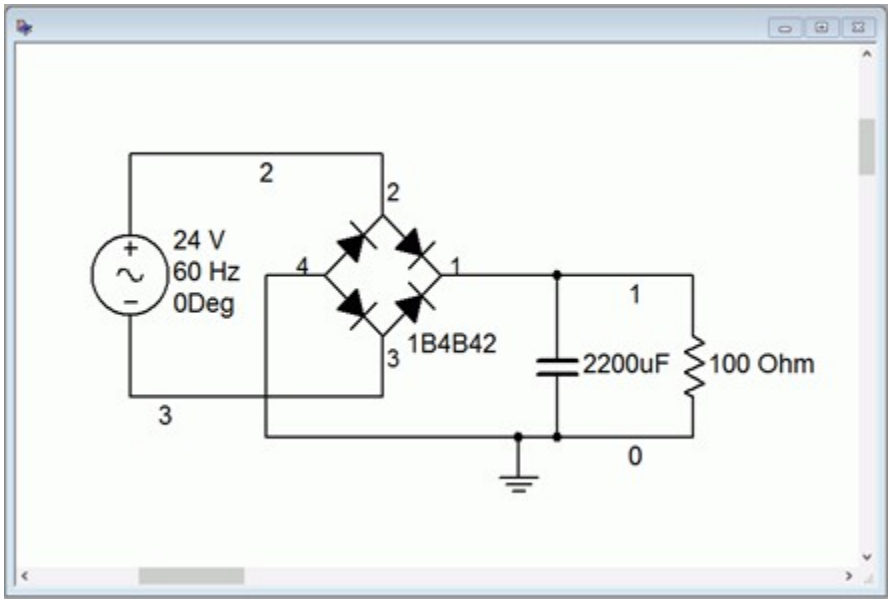


Figure 2.2.19: Simulation schematic for the design of Example 2.2.1 with capacitor.

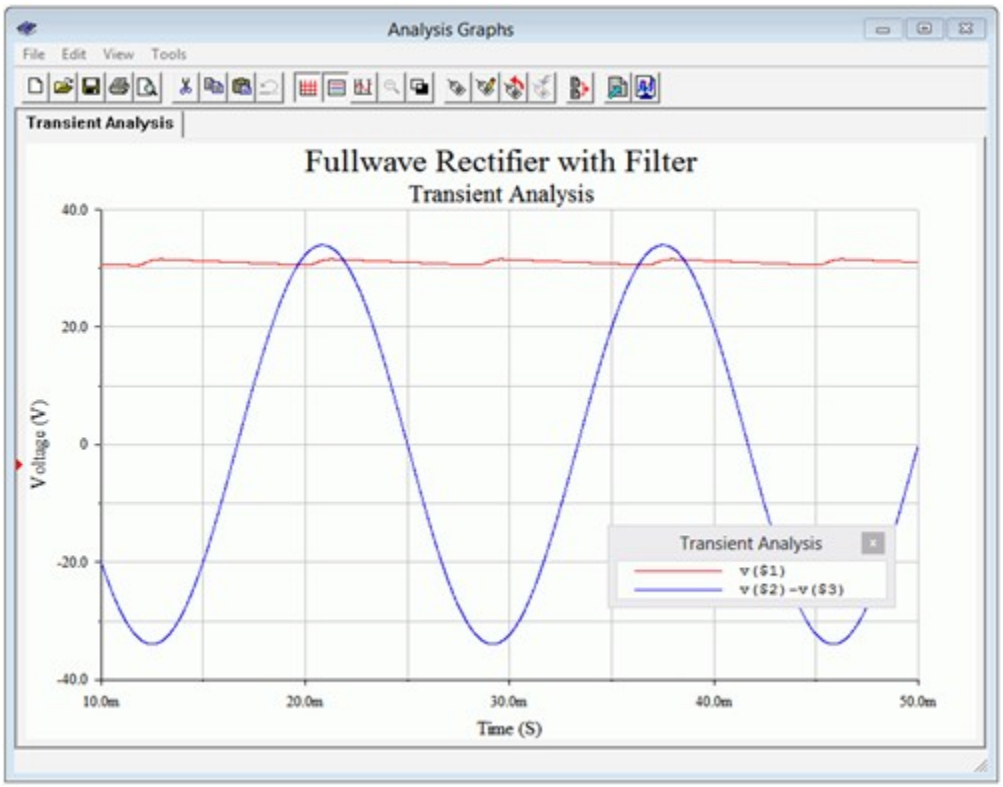


Figure 2.2.20: Transient analysis of the design of Example 2.2.1 with capacitor.

FULL-WAVE BRIDGE WITH DUAL OUTPUTS

As mentioned, the full-wave bridge can be configured to create a dual output bipolar supply. This is shown in Figure 2.2.21. Note the inclusion of the center tap on the secondary of the transformer and the location of the ground connection between the two loads and their associated capacitors.

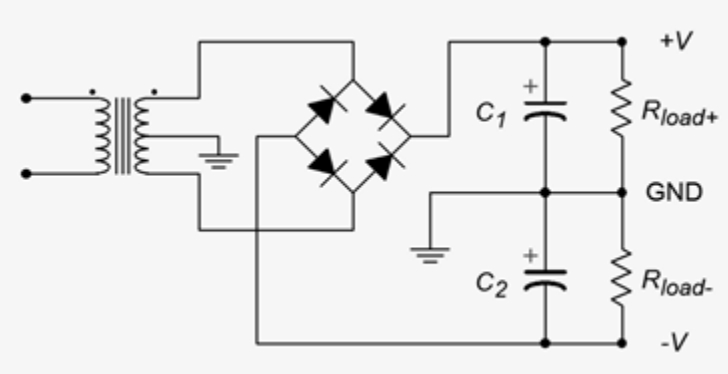


Figure 2.2.21: Dual output full-wave bridge rectifier.

One way of thinking of this is that we have simply created a new reference point, splitting in half the total output potential of the circuit presented in Figure 2.2.14. Alternately, it can be thought of as the upper half of the secondary driving R_{load}+ while the bottom half drives R_{load}–, as if the bridge and two-diode versions were somehow combined in a transporter accident, as in the 1958 movie The Fly, although it doesn't scream "Help me! Help me!" in a tiny little voice at the end.

ZENER REGULATION

Adding a large capacitor to a rectifier is necessary to store and transfer energy so that a smooth, ideally non-varying voltage results. As noted previously, under heavy load the ripple would increase in amplitude and the average voltage would drop. This issue can be greatly reduced by adding a Zener diode and current limiting resistor to the output, following the capacitor. This is called a Zener regulator and is shown in Figure 2.2.22.

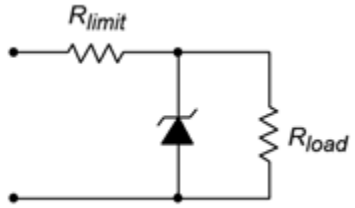


Figure 2.2.22: Simple Zener regulator.

The operation of the Zener regulator is fairly straightforward. Recall that when reverse-biased with a sufficiently large potential, the normal reverse diode behavior of an open switch abruptly changes to maintain a fixed voltage; the Zener potential. The current through the diode begins to increase dramatically once this potential is reached. If we place a Zener diode across the output of our filtered rectifier, the Zener will attempt to limit the output voltage to the Zener potential. To prevent excessive and possibly destructive current draw by the Zener diode, the voltage difference between the capacitor voltage and the Zener potential is dropped across a series current limiting resistor. This limiting resistor will set the maximum amount of output current. This current is then split between the Zener diode and the load. Under light load conditions, most of this current will flow through the Zener diode. Under heavy load conditions, most of the current will be drawn by the load with little flowing through the Zener diode. If the load current demand is too heavy, no current is available for the Zener diode and it stops conducting. Regulation is lost and the limiting resistor forms a voltage

divider with the load. A complete rectifier/filter/Zener regulator circuit is show in Figure 2.2.22. Let's examine how R_{limit} interacts with the load.

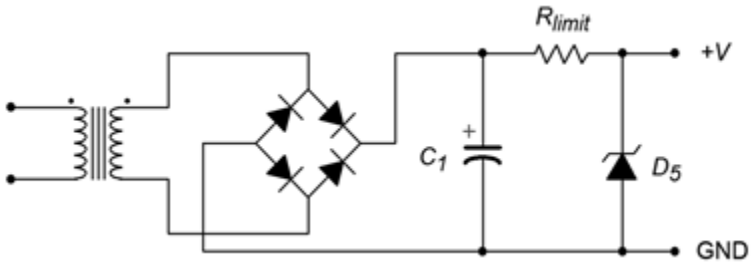


Figure 2.2.23: Full-wave bridge rectifier and filter with Zener regulator.

For proper operation, the Zener potential ($D5$) is the desired DC output voltage and the peak secondary voltage is set somewhat higher. We wish to guarantee that under full load conditions the lowest capacitor voltage due to ripple is still greater than the desired DC output voltage. The difference between the capacitor voltage and the Zener potential drops across R_{limit} . Therefore

$$I = \frac{V_{cap} - V_Z}{R_{limit}}$$

Under no-load conditions all of this current flows down through the Zener diode. The maximum load current is equal to this value (at which point no current flows through the Zener diode).

Example 2.2.2

Determine the maximum load current for a DC supply such as that found in Figure 2.2.22. The capacitor voltage is 15 volts average with ±1 volt of ripple (i.e., 16 volts dropping to 14 volts). The Zener potential is 12 volts and R_{limit} is 4.7 Ω.

The highest possible continuous load current is the current through R_{limit} (ignoring I_{ZT}). The limiting case for continuous draw will occur when the capacitor voltage is at its lowest value, or 14 volts.

$$I = \frac{V_{cap} - V_Z}{R_{limit}}$$

$$I = \frac{14V - 12V}{4.7\Omega}$$

$$I = 426mA$$

(actually a few mA less due to IZT)

The highest peak current through the Zener diode is found at the maximum capacitor voltage and assumes no current is drawn by the load.

$$I = \frac{V_{cap} - V_Z}{R_{limit}}$$

$$I = \frac{16V - 12V}{4.7\Omega}$$

$$I = 851mA$$

Note that this worst case current times the Zener potential results in a power dissipation of about 10 watts. Of course, during normal operation with a load drawing current, the diode dissipation is much reduced. It is interesting to note that the Zener dissipates maximal power when the load current is zero. Consequently, we can think of this circuit as shifting current from the Zener diode to the load as the load demands more current.[2]

2. As you might guess, this is not particularly efficient because even when the load demand is nil, the Zener diode is still drawing current from the transformer. An improved circuit may include a bipolar transistor, as examined in Chapter 4. For details on more sophisticated techniques to regulate voltage, see Fiore, J, Operational Amplifiers and Linear Integrated Circuits: Theory and Application, another free OER text.

2.3 CLIPPERS

Sometimes it is useful to limit the maximum amplitude of a signal. This might be done for protection, for example when too large of an input signal might damage the following circuit. It might also be employed as a means of wave shaping, that is, morphing a signal into another shape. A good example is the purely aesthetic desire to emulate the sound of "fuzz" guitar. In the early days of rock music it was discovered that over-driving a guitar amplifier in an attempt to make it louder created considerable distortion and this produced a new and interesting sound quality. Technically, this is largely caused by the power stage of the amplifier reaching its maximum output level. Any portion of the waveform above this level is simply cut off or clipped.[1] The practical problem here is that the only way to achieve this sound is to crank up the guitar amplifier's volume to ten[2] and live with the attendant high loudness level. Not too popular with the neighbors, that's for sure. In contrast, if the signal could be limited before the power amplifier stage in an attempt to mimic the clipping, the effect could be achieved without the resulting loudness. This proved to be so popular among guitarists that by the 1970s numerous companies were making "fuzz boxes" and "distortion pedals", each with their own twist on the concept.

The simplest form of clipper places a diode (or two parallel diodes of opposing polarity) in parallel with the load. The diode will limit the output voltage swing to its forward turn-on potential; 0.7 volts for a silicon device. This circuit is somewhat limiting (pun intended) as you are stuck with a 0.7 volt limit value. What if we need to limit at some other potential, say 12 volts? While it is possible to stack a bunch of diodes in series to increase the limit point, a more flexible and practical approach involves biasing the diode with a DC source. This is called a biased diode clipper.

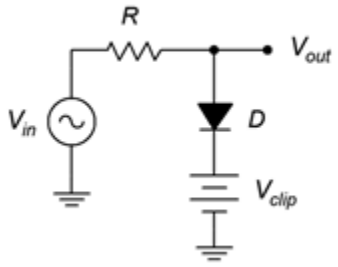


Figure 2.3.1: Biased positive clipper.

A biased diode clipper used to limit positive signals is shown in Figure 2.3.1. The operation is as follows. For any input signal that is less positive than the bias potential V_{clip} , the diode will be in reverse-bias. This means that the diode branch presents a high resistance and is effectively removed from the circuit. Therefore, V_{in} flows through R to the output unimpeded. If the input signal exceeds by V_{clip} by approximately 0.7 volts, the diode turns on resulting in a very low internal resistance.

1. We will take a closer look at amplifier clipping in the chapters that cover power amplifiers.
2. Or eleven, if you have a custom Spinal Tap amplifier.

As the internal resistance of the DC source is also very low, this creates a low impedance path to ground and results in a voltage divider with *R*. As *R* is a much greater resistance value, virtually all of the input signal above the turn-on voltage will be dropped across *R*, never reaching the output. Therefore, we can control or program the clip point by adjusting the bias voltage. Clipping will occur at approximately V_{clip} plus 0.7 volts, assuming a silicon diode is used.

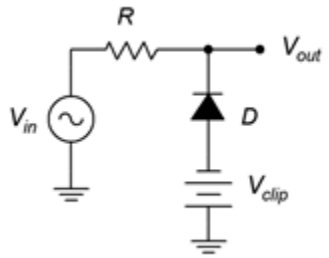


Figure 2.3.2: Biased negative clipper.

A negative biased clipper is shown in Figure 2.3.2. Both the diode and DC bias voltages have been flipped to the opposite polarity. The operation of this circuit is similar to the positive clipper. In this variant, the negative bias potential ensures that the diode is reverse-biased as long as the input level is more positive than V_{clip} minus 0.7 volts. Once the input signal goes below this voltage, the diode turns on creating the shorting path and limiting the output voltage.

A bipolar or dual-polarity clipper can be created by combining the positive and negative clippers. It is possible to limit the positive and negative swings independently, as illustrated in the following Example.

Example 2.3.1

Determine the output signal for the circuit of Figure 2.3.3. The input voltage is a 12 volt peak sine wave at 100 Hz. D_1 and D_2 are silicon switching diodes. R = 10 kΩ, V_1 = 4 volts and V_2 = 8 volts.

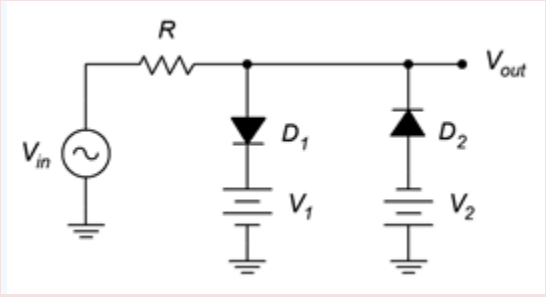


Figure 2.3.3: Dual clipper for Example 2.3.1.

First off, note that the precise value of R is unimportant here. It simply needs to be significantly larger than the on-resistance of the diodes. Based on our prior study of diode resistance in Chapter 2 (see the discussion around Figure 2.4.3) it is likely that the dynamic diode resistance will be under 100 Ω once full

turnon is reached. This resistance ratio is more than sufficient to create an effective voltage divider.

The positive clip level will be set by V1 . Adding in the forward potential of D1 we arrive at 4.7 volts. The negative clip level will be set by V2 . Including in the forward potential of D2 we arrive at –8.7 volts.

Thus, we expect to see a sine wave that is clipped at +4.7 volts and –8.7 volts. It should appear as a sort of lopsided cross between a sine wave and a square wave.

COMPUTER SIMULATION

To verify and visualize our computations, the circuit of Example 2.3.1 is simulated with a transient analysis. The circuit schematic is shown in Figure 2.3.4. For the diodes, common 1N914 switching diodes are used.

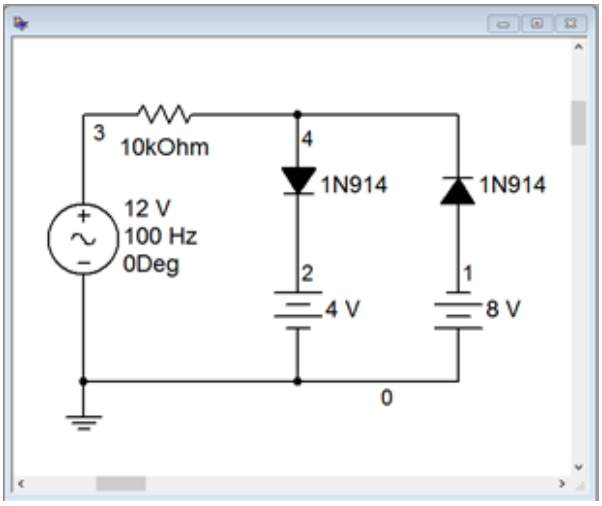


Figure 2.3.4: Simulation schematic for dual clipper of Example 2.3.1.

The results of the transient analysis are shown in Figure 2.3.5. The input waveform is shown in red while the output waveform is shown in blue. The input sine wave appears as expected with the specified 12 volt peak. The output follows the input perfectly for all values that are within the clip points. Beyond the clip points, the output voltage flattens. That is, it is limited to just below +5 volts (the programmed +4.7) and to just above –9 volts (the programmed –8.7 volts).

Careful inspection of the output waveform plot reveals that it is not perfectly flat at the voltage limits. In fact, there is a slight rounding that is most noticeable toward the transitions. This is due to the fact that the diodes do not turn on immediately. The dynamic resistance of the diodes change with the size of the signal. That is, the greater the input signal is above the clip point, the more current that will flow, and thus the dynamic resistance decreases, strengthening the effect of the voltage divider.

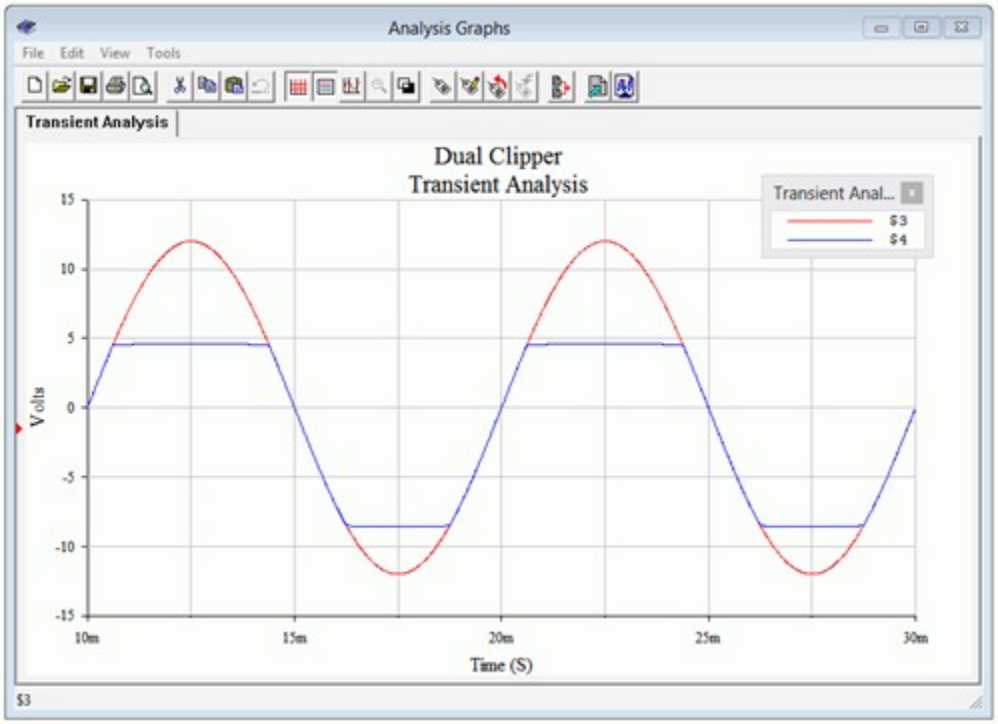


Figure 2.3.5: Transient analysis for dual clipper of Example 2.3.1.

2.4 CLAMPERS

A clamper is circuit that adds a DC offset to an AC signal in such a way that the resulting voltage is uni-polar. A positive clamper adds a positive offset such that the former negative peak now sits at zero volts. In like fashion, a negative clamper adds in a negative offset such that the former positive peak now sits at zero volts. Clampers are also referred to as DC restorers. Clampers can also be biased so that the new peak point is something other than zero volts.

The concept of a clamper is fairly simple; we just add a DC voltage to the existing AC signal. The trick is in getting the circuit to automatically determine what the DC shift needs to be. This way, if the amplitude of the input signal changes, the offset can track with it.

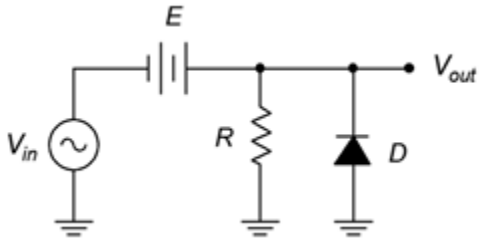


Figure 2.4.1: Prototype DC offset circuit.

First, let's consider the prototype circuit in Figure 2.4.1. This is a fixed DC offset circuit. The DC source *E* adds a positive offset to the input signal. If the offset is equal to the peak value of the input, the negative peak will rise up to zero volts and the diode will never turn on (meaning that it will not load the input and change the wave shape).

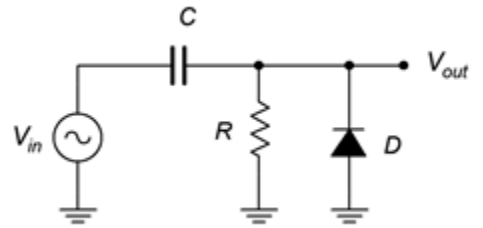


Figure 2.4.2: Positive clamper.

In Figure 2.4.2 the fixed DC source has been replaced with a capacitor. This capacitor is used to create the DC offset. Unlike the fixed source, the capacitor voltage will vary with the peak value of the input and therefore precisely compensate to produce an ideally clamped output signal. As long as the time constant for this capacitor and the surrounding resistance is much longer than the period of the input waveform, it will achieve proper clamping action.

Here is how the circuit operates. For the initial positive cycle, the capacitor is uncharged and the diode is reverse-biased. As the *RC* time constant is much longer than the input period, the output voltage merely follows the input voltage. Once the input signal swings negative, the diode turns on.

This bypasses the parallel resistor and drastically reduces the charge time constant. This means that the capacitor voltage will begin to track the negative portion of the input signal while the output stays near zero volts. Note that the capacitor voltage will have a polarity of minus-to-plus from left to right, in keeping with Kirchhoff's voltage law. The capacitor voltage will track the negative input voltage all the way down to the negative peak. Once the input begins to reverse slope and rise toward zero, the diode will be turned off due to the potential now held on the capacitor. At this point, the capacitor has a voltage across it that is equivalent to the negative peak value of the input signal and it will behave just like the fixed DC voltage source in the prototype. The input is just now starting to track in the positive direction from its negative peak while the capacitor holds this same magnitude of voltage. The result is that the output is at zero volts and as the input continues to swing positive, the output will track it, thus producing the desired level shift.

Of course, circuits are never perfect. First, the forward voltage drop of the diode will result in a negative peak that's not precisely at zero volts but is instead about –0.7 volts. Second, it may take more than one cycle of the input to "grab" the peak value, all depending on the period and the precise charge and discharge time constants. As you might guess, flipping the polarity of the diode will result in a negative clamper instead of a positive clamper. Also, if we add a DC source in series with the diode, like we did with the biased clipper, we can create a biased clamper. This is shown in Figure 2.4.3.

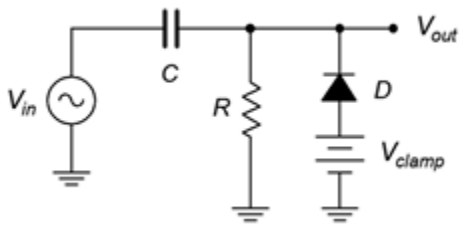


Figure 2.4.3: Biased positive clamper.

Example 2.4.1

Determine the output signal for the circuit of Figure 2.4.3. The input voltage is a 10 volt peak sine wave at 1 kHz. C = 10 μF, R = 10 kΩ, Vclamp = 5.7 volts and D is a silicon switching diode.

The configuration is a positive biased clamper. First, we need to ensure that the discharge time constant is much longer than the period. The period is 1/f, or 1 millisecond. The discharge time constant is

$$\tau = RC$$

$$\tau = 10k\Omega \times 10\mu F$$

$$\tau = 100milliseconds$$

The DC clamping source will produce a positive offset of 5 volts (5.7 volts minus the 0.7 volt forward diode drop). This means that we should see a 20 volt peak-to-peak sine wave that swings between +5 volts and +25 volts.

COMPUTER SIMULATION

To verify the analysis of Example 2.4.1, the circuit is captured as shown in Figure 2.4.4. A common 1N914 switching diode is used.

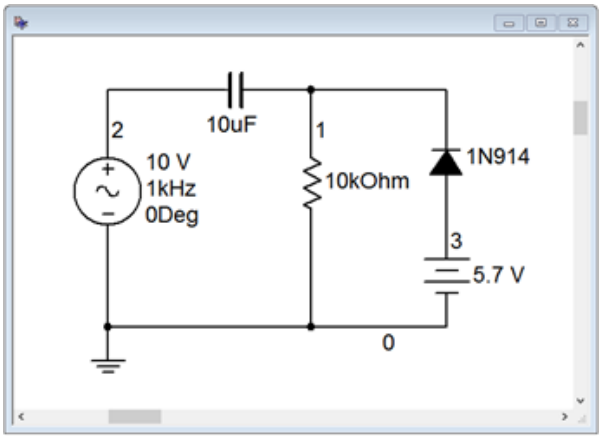


Figure 2.4.4: Simulation schematic for biased clamper of Example 2.4.1.

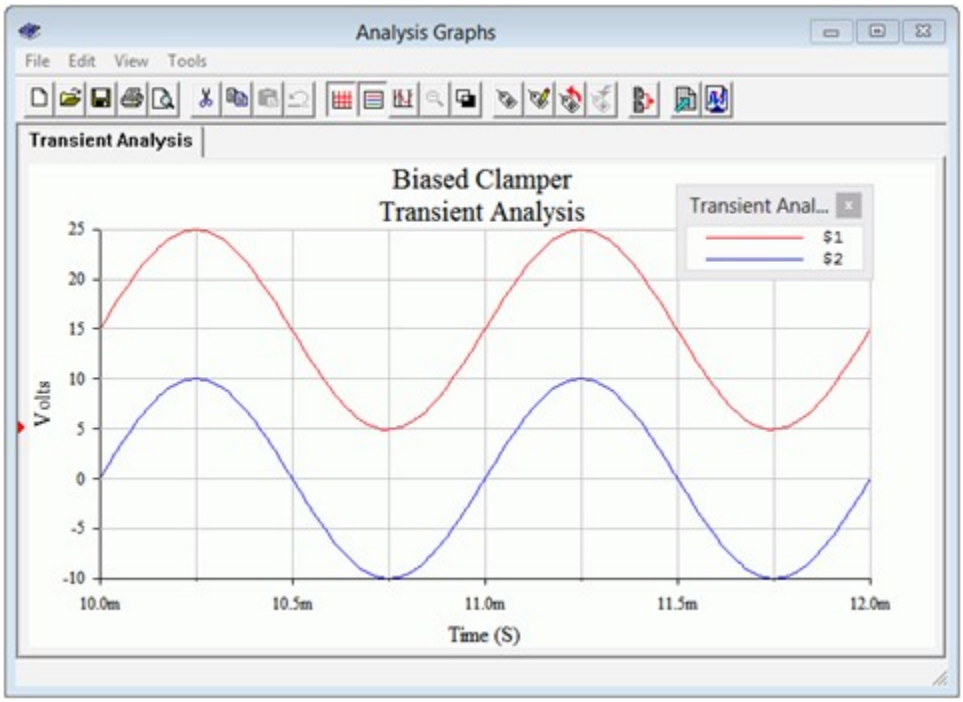


Figure 2.4.5: Transient analysis for biased clamper of Example 2.4.1.

The transient analysis is shown in Figure 2.4.5 and is precisely as predicted. The input waveform is blue and the clamped output is shown in red. The analysis was delayed ten milliseconds in order to get past the initial charge phase and observe the steady state operation.

2.5 SUMMARY

Numerous AC diode applications have been examined in this chapter. A single diode may be used to create half-wave rectification, producing pulsating DC from AC. This is achieved by simply blocking one of the two polarities with a diode. A more efficient form of rectification is full-wave rectification. In this scheme, one of the two polarities is effectively flipped. This may be achieved via a two diode circuit that employs a split secondary transformer or via a four diode bridge circuit using a non-tapped secondary. The addition of a split secondary to the bridge circuit enables a dual polarity output.

In order to smooth the pulsating DC into a relatively constant level, a filter capacitor is added in parallel with the load. The larger the capacitor, the greater the filtering and smoothing action, however, this will also increase peak charging current. A Zener diode may be employed to further stabilize the output voltage.

Clippers are used to limit the range of an input signal. They may be designed to clip the positive portion, the negative portion or both polarities of the input waveform. The positive and negative clip levels may be adjusted independently.

Clampers are used to create a DC level shift that is dependent on the peak level of the input waveform. The shift may be positive or negative, and may also include an optional bias. The operation of the clamper hinges upon the charge versus discharge time constants for a series capacitor and associated diode.

Review Questions

1. List the advantages and disadvantages of half-wave versus full-wave rectifiers.
2. Discuss the advantages and disadvantages of a full-wave bridge rectifier versus a two diode center-tapped rectifier.
3. What is the purpose of the capacitor in a rectifier/power supply circuit?
4. Under what load conditions will a Zener regulator fail to maintain regulation of the output voltage?
5. What is ripple? How might it be reduced?
6. What is the function of the DC source(s) in a biased clipper?
7. What is the function of the capacitor in a clamper circuit?

2.6 EXERCISES

Assume diodes are silicon unless stated otherwise

ANALYSIS PROBLEMS

1. For the circuit of Figure 2.6.1, determine the peak output voltage.
   V_{sec} = 12 volts RMS, R_{load} = 50 Ω, C_1 = 1500 μF.

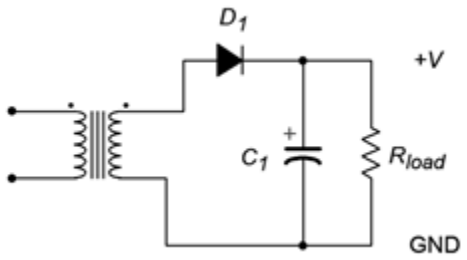


Figure 2.6.1

2. Sketch the output voltage waveform for the circuit of Problem 1, Figure 2.6.1, with and without the capacitor.
3. Determine the peak output voltage for the circuit of Figure 3.6.2.
   V_{sec} = 18 volts RMS, R_{load} = 75 Ω, C_1 = 470 μF.

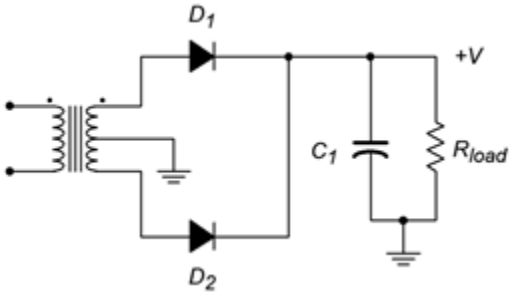


Figure 2.6.2

4. Sketch the output voltage waveform for the circuit of Problem 3, Figure 2.6.2, with and without the capacitor.
5. For the circuit of Figure 2.6.3, determine the peak output voltage.
   V_{sec} = 18 volts RMS, R_{load} = 40 Ω, C_1 = 1000 μF.

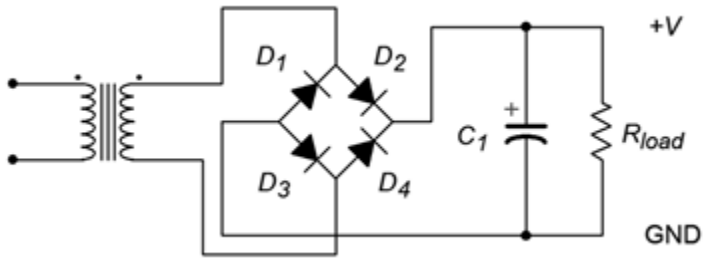


Figure 2.6.3

6. Sketch the output voltage waveform for the circuit of Problem 5, Figure 2.6.3, with and without the capacitor.
7. Determine the output voltage waveform and its amplitude for the circuit of Figure 3.6.4. V_{in} = 10 sin 2π100t, V_{clip}= 8 volts, R = 10 kΩ.

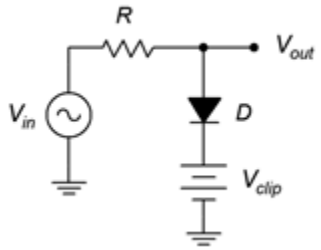


Figure 2.6.4

8. Draw the output waveform with its amplitudes for the circuit of Figure 2.6.5. V_{in} = 10 sin 2π100t, V_{clip} = 5 volts, R = 10 kΩ.

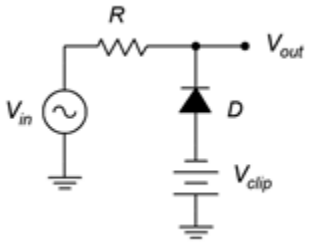


Figure 2.6.5

9. Draw the output waveform with its amplitudes for the circuit of Figure 2.6.6. V_{in} = 12 sin 2π200*t*, V_1 = 6 volts, V_2 = 4 volts, R = 10 kΩ.

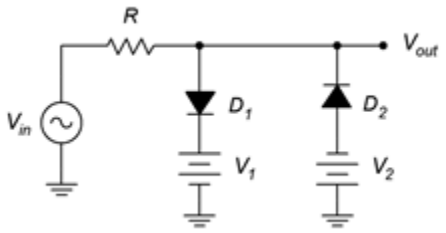


Figure 2.6.6

10. Draw the output waveform with its amplitudes for the circuit of Figure 2.6.7. V_{in} = 5 sin 2π2000*t*, C = 10μF, R = 4.7 kΩ

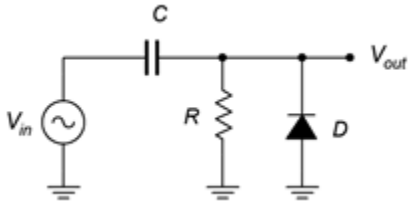


Figure 2.6.7

11. Draw the output waveform with its amplitudes for the circuit of Figure 2.6.8. V_{in} = 8 sin 2π500*t*, V clamp = 2 volts, C = 4.7 μF, R = 33 kΩ.

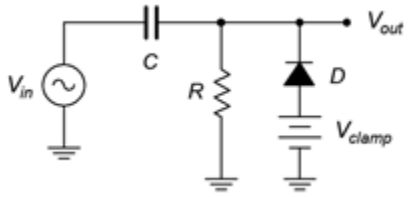


Figure 2.6.8

DESIGN PROBLEMS

12. Design a 15 volt AC to DC power supply capable of drawing 200 mA.
13. Design a circuit that will limit its output voltage to a range of –5 volts to +10 volts.
14. Design a circuit that will shift its output voltage so that it is always positive. The input frequency is 2 kHz.

CHALLENGE PROBLEMS

15. Design a circuit that will shift its output voltage so that its negative peak is at +3 volts. The input frequency range is from 100 Hz to 1 kHz.

COMPUTER SIMULATION PROBLEMS

16. Run a transient analysis of the circuit in Figure 2.6.1, Problem 1.
17. Run a transient analysis of the circuit in Figure 2.6.2, Problem 3.

18. Run a transient analysis of the circuit in Figure 2.6.3, Problem 5.
19. Run two transient analyses on the clamper circuit of Example 2.4.1, first using a capacitor 100 times larger than specified, and second using a capacitor 100 times smaller than specified. Discuss the resulting waveforms.

UNIT 3: BIPOLAR JUNCTION TRANSISTORS (BJTS)

Learning Objectives

After completing this chapter, you should be able to:

- Draw and explain the energy diagram for a biased bipolar junction transistor (BJT).
- Describe the differences between NPN and PNP BJTs.Explain forward-reverse bias operation.
- Define the transistor parameters alpha and beta, and determine them from device curves and/or circuit currents.
- Draw and explain the various regions of a BJT collector curve, along with Early voltage.
- Describe and utilize the Ebers-Moll BJT model. Describe the concept of the DC load line.
- Solve and design basic switching and driver circuits utilizing BJTs.

3.1 INTRODUCTION

The bipolar junction transistor, or BJT, is a foundational electronic component. It serves as the basis for a variety of applications ranging from simple amplifiers to device control to complex digital computing circuitry. Variations exist for applications spanning very low to very high frequency work; low, medium and high power; inexpensive general purpose through highly specialized niche items; and so forth.

No matter what a BJT has been optimized for, all BJTs can be considered to be current boosting devices. Of course, if you can boost current, then you can also boost voltage and power, depending on the associated impedances. Further, all BJTs share the same basic structure: three alternating layers of N-type and P-type material with one external lead attached to each layer. In this manner, the BJT can be thought of as an extension of the basic diode: just add another segment of oppositely doped material to one end of the diode creating a second PN junction. The configuration could be either PNP or NPN. There are uses for both types and circuits often work best when the two types are used together.

3.2 THE BIPOLAR JUNCTION TRANSISTOR

In prior work we discovered that the PN junction is the foundation of the basic diode. Under normal operating conditions the interface between the N-type and P-type materials is devoid of free charges and is referred to as a depletion region. The dissimilar Fermi levels of N-type and P-type materials lead to an "energy hill" between them, and without an external potential of the proper polarity, the junction will not allow current to flow. The required magnitude is a function of the material used but it is always the case that the P material (anode) must be positive with respect to the N material (anode). We extend this idea by adding a second portion of N material to the other side of the P material, creating an N-P-N "sandwich" of sorts. This is shown in Figure 3.2.1.

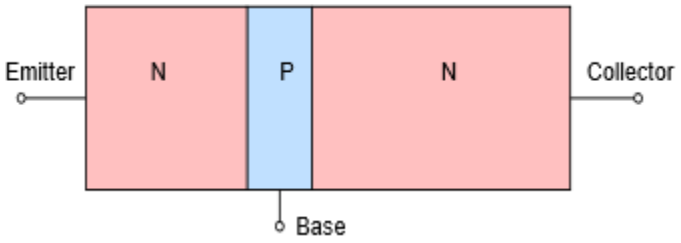


Figure 3.2.1: Basic configuration of NPN bipolar junction transistor.

This diagram is drawn to ease the understanding of the operation of the device, extending our earlier diode work. In contrast, real BJTs are built in more of a "layer cake" fashion, N-P-N bottom to top[1]. Of course, the spatial orientation of the device has no bearing on its operation so this is not a major issue for our purposes. The three terminals are named the emitter, base and collector. The collector is the largest of the three regions while the base is relatively thin and lightly doped.

Above absolute zero there will be recombination and two depletion regions will form as shown in Figure 2.2.2. Compare this figure to the basic PN junction drawing found at the beginning of Chapter 2, Figure 2.1.1.

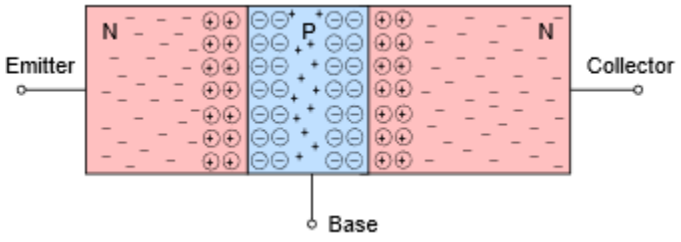


Figure 3.2.2: Charges in NPN BJT (base region widened to show detail).

A SIMPLE TWO-DIODE MODEL

Because this device contains two depletion regions, a much simplified model can be created using two diodes as shown in Figure 3.2.3. Please keep in mind that this is a very limited model (as we shall soon see).

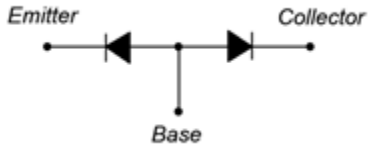


Figure 3.2.2: Charges in NPN BJT (base region widened to show detail).

If you were to test an NPN BJT with an ohmmeter, two leads at time, this model would successfully predict the results. If the red (positive) lead of the ohmmeter is connected to the base and the black (negative) lead is connected to either the emitter or collector, a low resistance will be indicated. This is because the ohmmeter will modestly forward-bias the base-emitter or base-collector junction. Similarly, if the leads are reversed, the meter will indicate high resistance because the junction under consideration will be reverse-biased. If the two leads are connected to the emitter and collector, a high reading will result regardless of the polarity. This is because one of the two junctions will be reverse-biased which results in no current flow through either of them due to the series connection.

BIASING THE BJT

Now let's consider adding external sources to bias the transistor. We begin by adding two DC sources with associated current limiting resistors as shown in Figure 3.2.4.

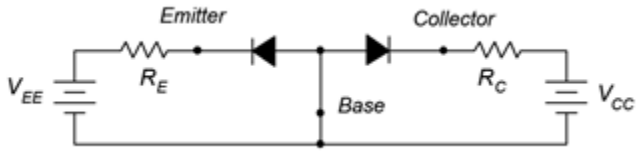


Figure 3.2.4: Double reverse-bias.

This circuit is comprised of two loops, one between the base-emitter and the second between the base-collector. In the B-E loop, the emitter supply V_{EE} reverse-biases the base-emitter diode. A similar situation occurs in B-C loop where the collector supply reverse-biases the base-collector diode. The result is that virtually no current flows anywhere in the circuit. If the two supplies are reversed in polarity then both diodes become forward-biased and we see currents flowing in both loops dependent on the precise values of the supplies and associated resistors. No surprises so far. Now consider if we forward-bias the base-emitter diode while simultaneously reversebiasing the base-collector diode, as shown in Figure 3.2.5.

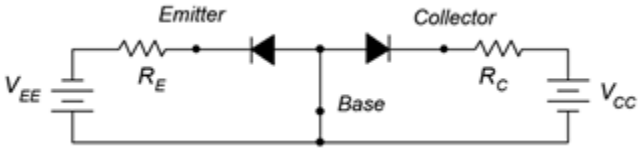


Figure 3.2.5: Forward-reverse bias.

With a simple pair of diodes we'd expect the B-E loop to show a high current and the B-C loop show negligible current. With a BJT, this is not what happens. Instead, what we see is a high current in both loops, and those currents are very nearly equal in magnitude. How does this come about?

The key to understanding this situation is that the base of the BJT is thin and lightly doped. In contrast, the dual diode model splits the base into two separate pieces of material and that makes all the difference. To get a better handle on what's happening here, let's take a closer look at this forward-reverse bias circuit but this time substituting the transistor diagram of Figure 3.2.2. Refer to Figure 3.2.6.

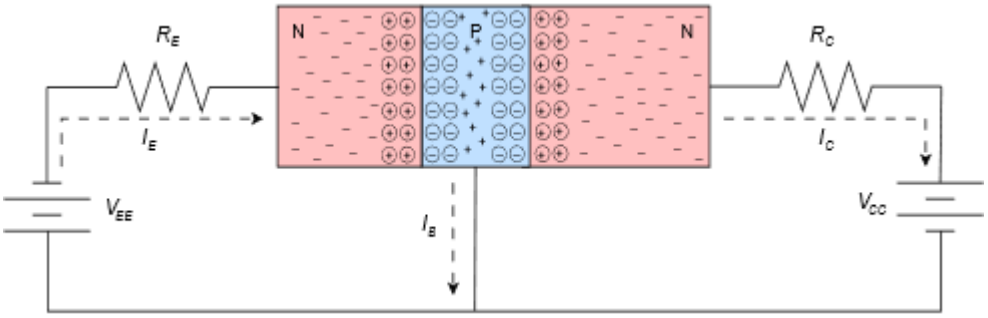


Figure 3.2.6: Forward-reverse bias, electron flow.

Electron flow will facilitate this explanation so we'll draw the current directions using dashed lines. From the left side of the diagram, electrons exit the emitter supply and enter the N emitter. Here they are the majority carrier. The base-emitter depletion creates an energy hill just as it did with a single PN junction. As long as there is sufficient potential from the emitter supply, the electrons will be pushed into the base. These electrons will attempt to recombine with the majority base holes, however, because the base is physically thin and lightly doped, only a small percentage of the injected electrons will recombine with base holes and exit the base terminal back to ground. This current is called the base current current or the recombination current. Meanwhile, the vast majority of the remaining electrons (95% to over 99%) will find their way to the base-collector depletion region and then to the collector. Once in the collector, the electrons are again the majority carrier and flow back to the positive terminal of the collector power supply. The energy diagram of the transistor is depicted in Figure 3.2.7. Compare this to the single PN junction energy diagram found at the beginning of Chapter 1, Figure 1.1.2.

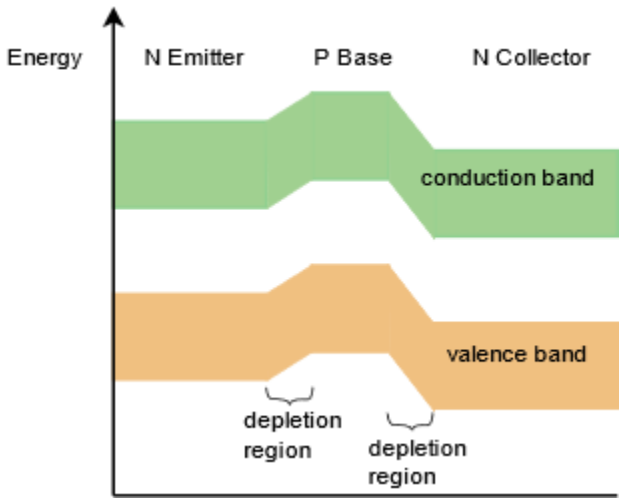


Figure 3.2.7: Energy diagram of BJT.

At first glance, it might appear as though the emitter and collector leads can be swapped with no change in operation. With real- world devices this is not possible generally because the emitter and collector regions are optimized and not physically identical.

Thus, placing transistors into a circuit backwards, with emitter and collector leads swapped, will usually result in unpredictable behavior.

Based on the foregoing discussion and what we already know about PN junctions, we can summarize transistor performance as follows:

- From KCL, $I_E = I_C + I_B$.
- $I_C \gg I_B$, therefore $I_E \approx I_C$.
- The base-emitter junction is forward-biased, therefore $V_{BE} \approx 0.7$ V (silicon).
- The base-collector junction is reverse-biased, therefore V_{CB} is large.
- Conventional current flows into the collector and base, and out of the emitter.

We can also define a couple of transistor performance parameters. The ratio of collector current to emitter current is called α (alpha). α typically is greater than 0.95. A somewhat more useful parameter is the ratio of collector current to base current. This is called β (beta) and can also be found on transistor spec sheets as h_{FE} (h_{FE} is one of four hybrid parameters). It is also referred to generically as current gain (if I_B is in the input signal and I_C is the output signal then β represents the amount of signal boost or gain). For small signal transistors β typically is in the range of 100 to 200, although it can be larger. For power transistors, β tends to be smaller, more like 25 to 50. Presented as formulas we have:

$$\alpha = I_C / I_E$$

(3.2.1)

$$\beta = I_C / I_B$$

3.2.2)

And with a little math,

$$\alpha = \beta / (\beta + 1)$$

$$\beta = \alpha / (1 - \alpha)$$

$$I_C = \beta I_B$$

Finally, we come to the schematic symbol of the NPN BJT, as shown in Figure 3.2.8. A common variation places the body of the device within a circle. Following the standard, the arrow points to N material and in the direction of easy conventional current flow.

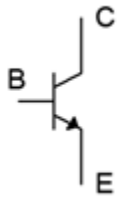


Figure 3.2.8: NPN Schematic symbol

THE PNP BIPOLAR JUNCTION TRANSISTOR

The PNP version of the BJT is created by swapping the material for each layer. The outcome is the logical inverse of the NPN regarding current directions and voltage polarities. That is, conventional current flows into the emitter, and out of the collector and base (echoing the electron flow of the NPN). Further, voltages across the device have reversed polarity, for example, $V_{BE} \approx -0.7$

V. All of the other characteristics remain unchanged so equations such 3.2.1 and 3.2.2 are still applicable. Just about any NPN-based circuit has its PNP counterpart. The schematic symbol of the PNP reverses the emitter arrow. As the base is now the N material, the arrow points toward the base. This is illustrated in Figure 3.2.9.

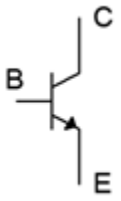


Figure 3.2.9: PNP Schematic symbol

3.3 BJT COLLECTOR CURVES

One of the more useful BJT device plots is the family of collector curves. This is a series of plots of collector current, I_C , versus collector-emitter voltage, V_{CE} , at varying levels of base current, I_B. To generate these curves we drive the base terminal with a fixed current source establishing I_B. A DC power supply is attached from the collector to emitter and then swept from zero volts to some upper value. This establishes V_{CE} . Simultaneously, we track the resulting collector current and plot the result. This results in one trace. The base current is then increased and the DC supply swept again for a second trace. This process is repeated to create a family of curves. An example is shown in Figure 3.3.1.

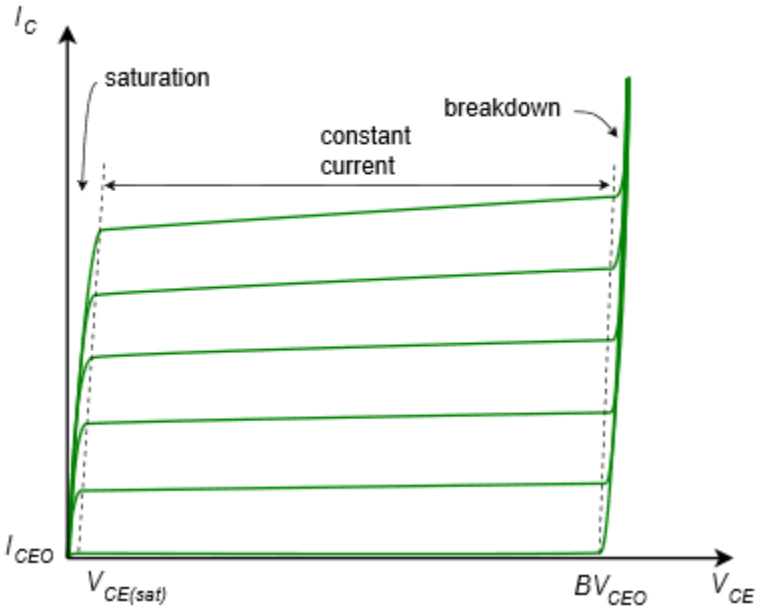


Figure 3.3.1: A family of collector curves.

The bottom curve results when I_B = 0. Ideally, the corresponding collector current would be 0 but a small leakage current occurs. This is usually referred to as I_{CEO} , meaning the *C*ollector-*E*mitter current with the base terminal *O*pen (i.e., no base current). The curves above this correspond to increasing levels of base current; each new base current stepping up a fixed amount for each subsequent trace (e.g., 0 μA, 10 μA, 20 μA, 30 μA, etc.).

The most striking thing about this set of curves is that there are three distinct regions or zones of operation. To the extreme left of the curve the current rises rapidly. This is known as the saturation region. The break-over point is fairly small at just a few tenths of a volt. This can be found on a data sheet as $V_{CE}(sat)$. The saturation region is used in transistor switching applications.

At the extreme right is another region where the collector current rises rapidly. This is called the breakdown region. This is the same effect we saw with individual diodes. We do not wish to operate devices in this region as damage may result. The breakdown voltage is denoted on most data sheets as

BVCEO (*C*ollector to *E*mitter voltage with an *O*pen base). For general purpose devices this will be in the range of 30 to 60 volts or so, but it can be much higher.

In between these two extremes is a region where the collector current is relatively constant, showing only a modest positive slope. This is the constant current region. This is where we want the transistor to operate for applications such as linear amplifiers.

A device called a curve tracer can be used to generate this family of curves in the lab. A very good approximation for β can be determined using these curves. First, we determine the approximate circuit values for I_C and V_{CE} , and locate this point on the graph. We then find the nearest plot line to that point. From the intersection of V_{CE} and and this plot line we track back to the vertical axis to find the precise value of I_C for that trace. We count the number of traces and multiply by the base current step size to determine the corresponding base current. We then divide the two values and arrive at β.

Example 3.3.1

Assume we have a BJT operating at V_{CE} = 30 V and IC = 4 mA. If the device is placed in a curve tracer and the resulting family of curves appears as in Figure 3.3.2, determine the value of β. Assume the base current is increased 10 μA per trace.

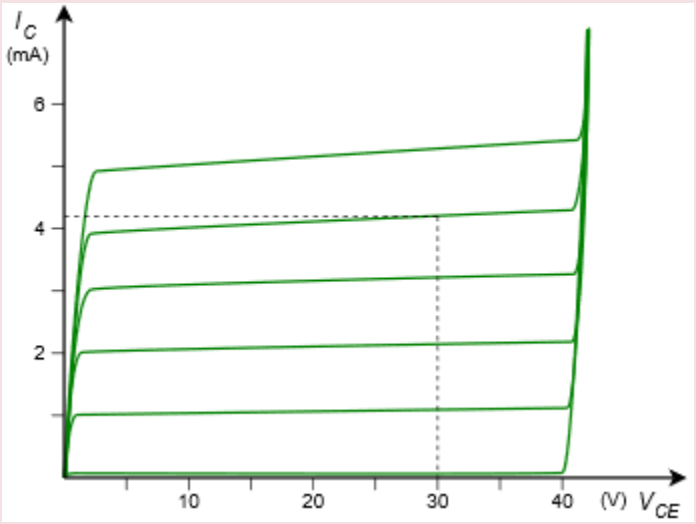


Figure 3.3.2: Curve tracer display for Example 3.3.1.

First, find a trace close to the operating point of 30 volts and 4 mA. Draw a vertical line at 30 volts and stop when that line intersects the trace nearest to 4 mA. In this example that's the second trace from the top. From that intersection point, track back to the vertical axis to determine the precise collector current. That's roughly 4.2 mA here. To determine the base current count up the number of traces to the selected trace. The selected trace is the fourth one up (do not include the bottom trace where I_B is 0). The base current is increased by 10 μA per trace so that leaves us with I_B = 40 μA.

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

$$\beta = \frac{4.2mA}{40\mu A}$$

$$\beta = 105$$

RISE IN *B* AND EARLY VOLTAGE

When looking at the collector curves, a good question we might ask is why the collector current rises as V_{CE} increases. This is due to the fact that the increased collector-emitter voltage is responsible for an increase in collector-base voltage (by definition, $V_{CE} = V_{CB} + V_{BE}$). V_{CB} is the reverse-bias potential on the collector-base PN junction. As this reverse potential increases, the collector-base depletion region widens. As it widens, it penetrates further into the base layer. Because the base is effectively narrowed, the chances for recombination are reduced, thus reducing base current and effectively increasing β.

If we extend the constant current region traces back into the second quadrant they intersect at a point called the Early Voltage, named after James Early, and denoted as V_A . This is illustrated in Figure 3.3.3.

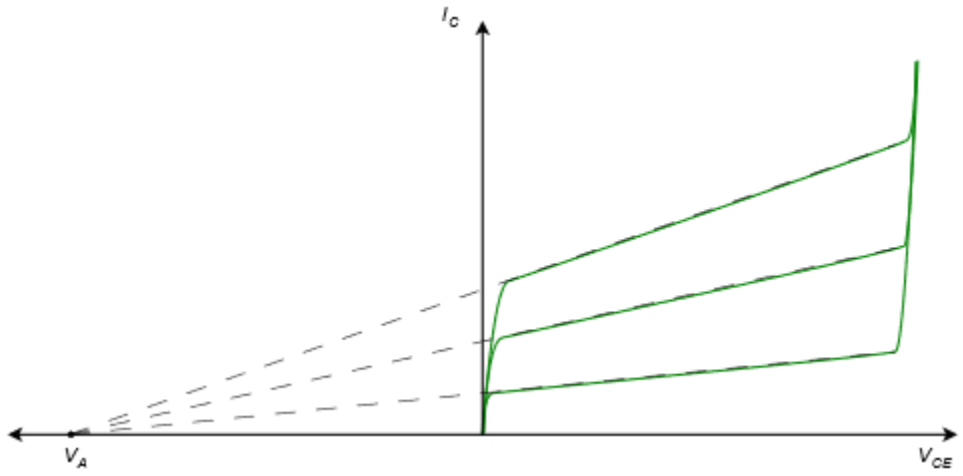
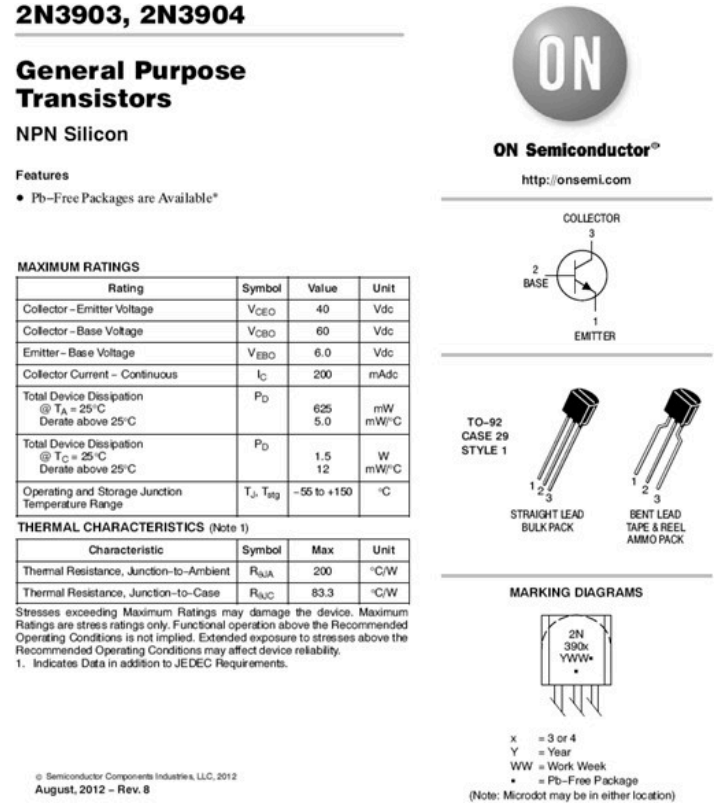


Figure 3.3.3: Early Voltage.

The data sheet for a common NPN transistor, the 2N3904, is shown in Figure 3.4.1. This model is available from several different manufacturers. First off, note the case style. This a TO-92 plastic case for through-hole mounting and is commonly used for small signal transistors. Under the maximums we find the device has a maximum power dissipation of 625 mW in free air (ambient temperature of 25° C), a maximum collector current of 200 mA and a maximum collector-emitter voltage of 40 V. Obviously, the device cannot withstand maximum current and voltage simultaneously.



2N3903, 2N3904

General Purpose Transistors

NPN Silicon

Features

- Pb–Free Packages are Available*

MAXIMUM RATINGS

Rating	Symbol	Value	Unit
Collector – Emitter Voltage	V_{CEO}	40	Vdc
Collector – Base Voltage	V_{CBO}	60	Vdc
Emitter – Base Voltage	V_{EBO}	6.0	Vdc
Collector Current – Continuous	I_C	200	mAdc
Total Device Dissipation @ T_A = 25°C Derate above 25°C	P_D	625 5.0	mW mW/°C
Total Device Dissipation @ T_C = 25°C Derate above 25°C	P_D	1.5 12	W mW/°C
Operating and Storage Junction Temperature Range	T_J, T_{stg}	−55 to +150	°C

THERMAL CHARACTERISTICS (Note 1)

Characteristic	Symbol	Max	Unit
Thermal Resistance, Junction–to–Ambient	$R_{\theta JA}$	200	°C/W
Thermal Resistance, Junction–to–Case	$R_{\theta JC}$	83.3	°C/W

Stresses exceeding Maximum Ratings may damage the device. Maximum Ratings are stress ratings only. Functional operation above the Recommended Operating Conditions is not implied. Extended exposure to stresses above the Recommended Operating Conditions may affect device reliability.
1. Indicates Data in addition to JEDEC Requirements.

© Semiconductor Components Industries, LLC, 2012
August, 2012 – Rev. 8

ON Semiconductor®
http://onsemi.com

COLLECTOR 3
2 BASE
1 EMITTER

TO–92
CASE 29
STYLE 1

STRAIGHT LEAD BULK PACK
BENT LEAD TAPE & REEL AMMO PACK

MARKING DIAGRAMS

2N
390x
YWW•
•

x = 3 or 4
Y = Year
WW = Work Week
• = Pb–Free Package
(Note: Microdot may be in either location)

Figure 3.4.1a: 2N3904 data sheet. Used with permission from SCILLC dba ON Semiconductor.

In Figure 3.4.1*b* we find a variety of characteristics including nominal values for β (listed here as h_{FE}) under various conditions. At particularly small or large collector currents β tends to drop off. Also, note the wide 3:1 variance at 10 mA. Perhaps more illustrative are the graphs from the third page, Figure 3.4.1*c*.

ELECTRICAL CHARACTERISTICS (T_A = 25°C unless otherwise noted)

Characteristic		Symbol	Min	Max	Unit
OFF CHARACTERISTICS					
Collector – Emitter Breakdown Voltage (Note 2) (I_C = 1.0 mAdc, I_B = 0)		$V_{(BR)CEO}$	40	–	Vdc
Collector – Base Breakdown Voltage (I_C = 10 μAdc, I_E = 0)		$V_{(BR)CBO}$	60	–	Vdc
Emitter – Base Breakdown Voltage (I_E = 10 μAdc, I_C = 0)		$V_{(BR)EBO}$	6.0	–	Vdc
Base Cutoff Current (V_{CE} = 30 Vdc, V_{EB} = 3.0 Vdc)		I_{BL}	–	50	nAdc
Collector Cutoff Current (V_{CE} = 30 Vdc, V_{EB} = 3.0 Vdc)		I_{CEX}	–	50	nAdc
ON CHARACTERISTICS					
DC Current Gain (Note 2)		h_{FE}			–
(I_C = 0.1 mAdc, V_{CE} = 1.0 Vdc)	2N3903		20	–	
	2N3904		40	–	
(I_C = 1.0 mAdc, V_{CE} = 1.0 Vdc)	2N3903		35	–	
	2N3904		70	–	
(I_C = 10 mAdc, V_{CE} = 1.0 Vdc)	2N3903		50	150	
	2N3904		100	300	
(I_C = 50 mAdc, V_{CE} = 1.0 Vdc)	2N3903		30	–	
	2N3904		60	–	
(I_C = 100 mAdc, V_{CE} = 1.0 Vdc)	2N3903		15	–	
	2N3904		30	–	
Collector – Emitter Saturation Voltage (Note 2)		$V_{CE(sat)}$			Vdc
(I_C = 10 mAdc, I_B = 1.0 mAdc)			–	0.2	
(I_C = 50 mAdc, I_B = 5.0 mAdc			–	0.3	
Base – Emitter Saturation Voltage (Note 2)		$V_{BE(sat)}$			Vdc
(I_C = 10 mAdc, I_B = 1.0 mAdc)			0.65	0.85	
(I_C = 50 mAdc, I_B = 5.0 mAdc)			–	0.95	
SMALL – SIGNAL CHARACTERISTICS					
Current – Gain – Bandwidth Product		f_T			MHz
(I_C = 10 mAdc, V_{CE} = 20 Vdc, f = 100 MHz)	2N3903		250	–	
	2N3904		300	–	
Output Capacitance (V_{CB} = 5.0 Vdc, I_E = 0, f = 1.0 MHz)		C_{obo}	–	4.0	pF
Input Capacitance (V_{EB} = 0.5 Vdc, I_C = 0, f = 1.0 MHz)		C_{ibo}	–	8.0	pF
Input Impedance		h_{ie}			k Ω
(I_C = 1.0 mAdc, V_{CE} = 10 Vdc, f = 1.0 kHz)	2N3903		1.0	8.0	
	2N3904		1.0	10	
Voltage Feedback Ratio		h_{re}			X 10^{-4}
(I_C = 1.0 mAdc, V_{CE} = 10 Vdc, f = 1.0 kHz)	2N3903		0.1	5.0	
	2N3904		0.5	8.0	
Small–Signal Current Gain		h_{fe}			–
(I_C = 1.0 mAdc, V_{CE} = 10 Vdc, f = 1.0 kHz)	2N3903		50	200	
	2N3904		100	400	
Output Admittance (I_C = 1.0 mAdc, V_{CE} = 10 Vdc, f = 1.0 kHz)		h_{oe}	1.0	40	μmhos
Noise Figure		NF			dB
(I_C = 100 μAdc, V_{CE} = 5.0 Vdc, R_S = 1.0 k Ω, f = 1.0 kHz)	2N3903		–	6.0	
	2N3904		–	5.0	
SWITCHING CHARACTERISTICS					
Delay Time	(V_{CC} = 3.0 Vdc, V_{BE} = 0.5 Vdc, I_C = 10 mAdc, I_{B1} = 1.0 mAdc)	t_d	–	35	ns
Rise Time		t_r	–	35	ns
Storage Time	(V_{CC} = 3.0 Vdc, I_C = 10 mAdc, I_{B1} = I_{B2} = 1.0 mAdc) 2N3903	t_s	–	175	ns
	2N3904		–	200	
Fall Time		t_f	–	50	ns

2. Pulse Test: Pulse Width ≤ 300 μs; Duty Cycle ≤ 2%.

Figure 3.4.1b: 2N3904 data sheet (cont).

The upper-most graph depicts the variation of β with both collector current and temperature. The normalized β is plotted on the vertical axis. That is, this is not the expected value but is a ratio used to compare β under varying conditions.

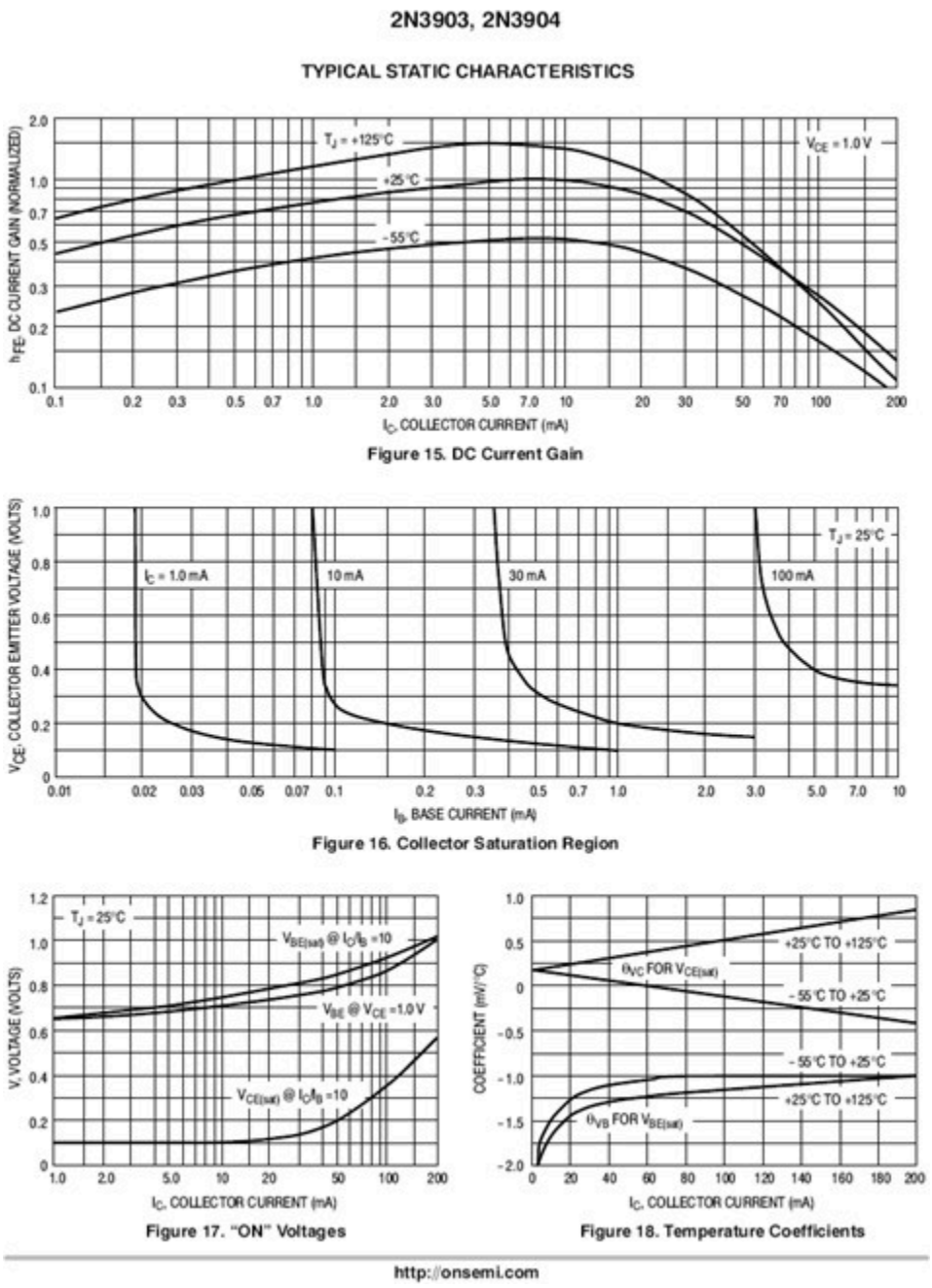


Figure 3.4.1c: 2N3904 data sheet (cont).

For example, at room temperature and 10 mA, the normalized value is 1.0. The second page indicated a range of 100 to 300 for the 2N3904's β under these conditions. Let's say we measure one particular transistor to have a β of 200. If we were to operate this transistor at a lower current, say 0.2 mA, the β would drop. From the graph, the normalized β value at 0.2 mA and 25° C is 0.7. Therefore, the β under these conditions would be 0.7/1.0 200, or 140. The graph also shows that, generally speaking, β tends to increase with increasing temperature.

The middle graph plots the collector-emitter saturation voltage, or *VCE*(*sat*) , for various current conditions. This is an important parameter when dealing with transistor switching circuits. We shall refer back to this graph a little later in this chapter.

3.5 EBERS-MOLL MODEL

A good, functional model of the BJT is the simplified Ebers-Moll model shown in Figure 3.5.1. This utilizes an ideal diode to model the base-emitter junction and a current-controlled current source located at the collector-base. This model is sufficient to achieve good analysis results with a variety of DC and low frequency circuits. It is important to remember, though, that β varies not only from device to device, but also varies with changes in temperature, collector current and collector-emitter voltage.

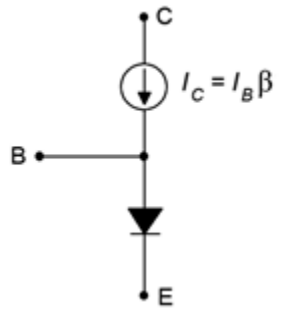


Figure 3.5.1: Ebers-Moll model of the NPN BJT.

We can put the Ebers-Moll model to use in basic DC biasing circuits. To properly bias the BJT we need to make the collector-base reverse-biased and the base-emitter forward-biased. In other words, $V_C > V_B > V_E$. There are many ways to achieve this. One method places the emitter at ground, a modest DC source in the base-emitter loop, and a somewhat higher DC source at the collector. An example is shown in Figure 3.5.2.

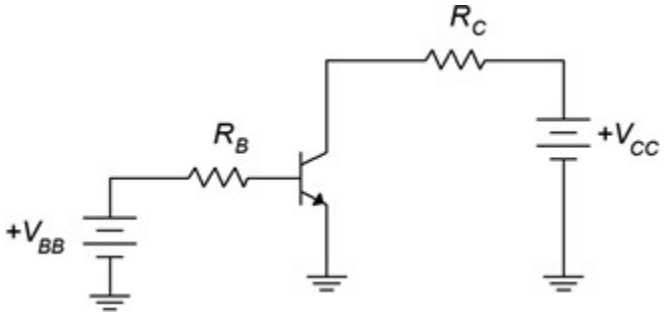


Figure 3.5.2: Simple base bias circuit.

The two resistors serve to limit the transistor's currents and voltages. Because the emitter is at ground, the common point, this circuit is classified as having a common emitter configuration. There are many possible common emitter circuits. We shall refer to this one specifically as base bias.

Now, let's replace the transistor with the Ebers-Moll model. The result, with added voltage polarities and current directions, is shown in Figure 3.5.3.

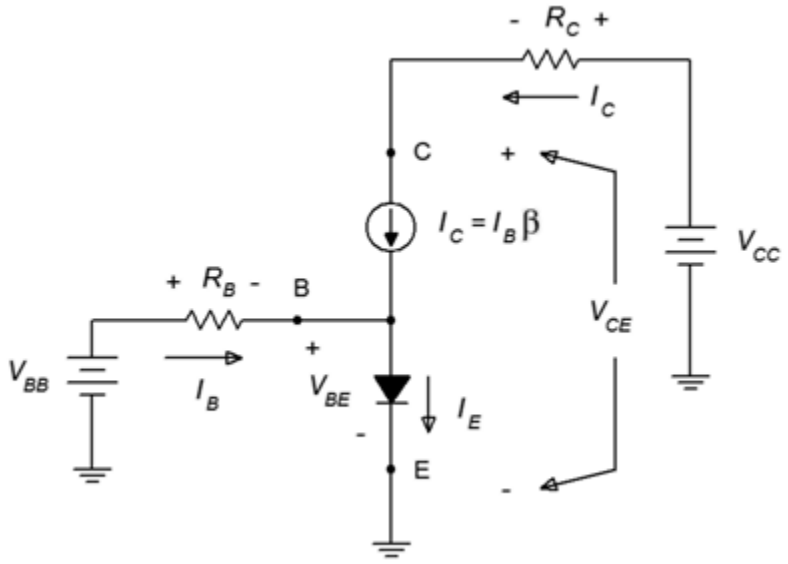


Figure 3.5.3: Base bias circuit with Ebers-Moll model.

Given values for the resistors, power supplies and β, all of the currents and voltages may be determined using fundamental circuit analysis techniques. The basic idea is to create KVL equations for the two loops and then expand using Ohm's law. We begin with the base-emitter loop.

$$V_{BB} = V_{RB} + V_BE$$

$$V_{BB} = I_B R_B + V_BE$$

$$I_B = \frac{V_{BB} - V_{BE}}{R_B}$$

(3.5.1)

And for the collector-emitter loop:

$$V_{CC} = V_{RC} + V_{CE}$$

$$V_{CC} = I_C R_C + V_{CE}$$

$$V_{CE} = V_{CC} - I_C R_C$$

(3.5.2)

To obtain I_C , recall that $I_C = \beta I_B$.

Example 3.5.1

Determine the circulating currents and device voltages for the circuit of Figure 3.5.2 if VBB = 10 V, VCC = 15 V, RB = 200kΩ, RC = 1 kΩ and β = 100. Assume that the transistor is silicon.

First, find the base current. KVL dictates that the voltage across RB is 9.3 volts.

$$I_B = \frac{V_{BB} - V_{BE}}{R_B}$$

$$I_B = \frac{10V - 0.7V}{200k\Omega}$$

$$I_B = 46.5\mu A$$

Now find the collector current and follow with Ohm's law and KVL.

$$I_C = \beta I_B$$

$$I_C = 100 \times 46.5\mu A$$

$$I_C = 4.65mA$$

$$V_{CE} = V_{CC} - I_C R_C$$

$$V_{CE} = 15V - 4.65mA \times 1k\Omega$$

$$V_{CE} = 10.35V$$

For the sake of completeness, the voltage across RC is 4.65 volts, V_{CB} is 9.65 volts and I_E is 4.6965 mA.

The preceding example illustrates that the place to start the analysis is in the base-emitter loop instead of the collector-emitter loop. This is because in the base-emitter loop we have the forward-biased base-emitter junction which has a known potential (approximately 0.7 volts). In contrast, the collector-emitter voltage is an unknown as it includes the reverse-biased collector-base junction. That voltage will depend on other circuit elements, most notably the collector resistor and associated supply.

An improvement on the circuit of Example 3.5.1 would be to redesign it for a single power supply rather than two supplies. This is easy to do. All that is needed is to keep the base current unchanged. If that remains at its original value then the collector current won't change and consequently nothing in the collector-emitter loop will change either. Using the 15 volt source for V_{BB} means that the voltage

across R_B will increase to 14.3 volts. Ohm's law then indicates that R_B must be 14.3 volts divided by 46.5 μA, or 307.5 kΩ.

B VARIATION ISSUES

There is a major problem with the circuit of Figure 3.5.2, namely, it lacks stability of collector current and collector-emitter voltage. As we shall see in upcoming chapters, it is important to keep these parameters stable in order to ensure consistent performance for many kinds of circuits. As we noted from our inspection of the 2N3904 data sheet, the variation of β can be quite large at a given operating point. If we also add in the variance due to temperature and other factors, we may be looking at a 10:1 range. If we repeat Example 3.5.1 with a doubled β of 200, the base-emitter loop is unchanged but the collector current doubles to 9.3 mA. This increases the voltage drop across R_C to 9.3 volts which then forces V_{CE} to drop to 5.7 volts.

Given a typical production run of transistors, this circuit might exhibit collector currents from less than 4 mA to more than 10 mA. In some applications this variation in current might be tolerable but not in all of them. For example, suppose an LED is placed in series with R_C . Because the brightness of an LED depends on its current level, the brightness will now depend on the β of the specific BJT used. If this is one LED in a larger display made up of similar circuits, then the illumination will be uneven between them causing the entire display to appear off kilter.

In fact, if this circuit was built in the lab, it is quite likely that after turning on the power, you could watch I_C slowly rise on your ammeter. This is because the BJT will begin to warm up as it dissipates power. As noted from the data sheet, β increases with increasing temperature. Because I_B is a fixed value, any rise in β means that I_C must also rise. This increased current will tend to cause a further rise in power dissipation and temperature which causes a further increase in β, and the process cycles. We have created an inadvertent thermal positive feedback loop. Left unchecked, devices could overheat and be destroyed. We will examine biasing circuits that achieve high stability in the next chapter.

There is another interesting aspect to this circuit. As noted, if we substituted the original BJT with another unit that had a higher β, the collector current would rise. What if we continued this to higher and higher β values? For example, if we increased β to 400 (admittedly, rather high) the new collector current would seem to jump up to 46.5 μA 400, or 18.6 mA. There is a "small" problem with this value. Ohm's law indicates that this current would develop a drop of 18.6 volts across the 1 kΩ R_C but that's impossible because V_{CC} is only 15 volts. The only way that "works" is if somehow the BJT is transformed magically into a 3.6 volt battery. No amount of prayer or letters to Santa will make that happen.[1]

1. Both being equally effective.

3.6 DC LOAD LINES

So how do we determine the range of possible values of collector current and collector-emitter voltage in any given DC BJT circuit? One answer is to employ the concept of the DC load line. In general, a load line is a plot of all possible coordinate pairs of I_C and V_{CE} for a transistor in a given circuit. Referring back to Figure 3.5.3, we pick up with Equation 3.5.2 and solve it for I_C :

$$V_{CE} = V_{CC} - I_C R_C$$

$$I_C = \frac{1}{R_C}(V_{CC} - V_{CE})$$

$$I_C = -\frac{1}{R_C}V_{CE} + \frac{V_{CC}}{R_C}$$

(3.6.1)

Equation 3.6.1 is a linear equation of the form $y = mx + b$. The y intercept (the value of I_C when $V_{CE} = 0$) is V_{CC}/R_C . This is the maximum collector current that can be achieved. At this point the transistor is saturated and this maximum is referred to as $I_{C(sat)}$. The x intercept (the value of V_{CE} when $I_C = 0$) is V_{CC}. This represents the largest possible voltage across the transistor's collector-emitter. At this point the current is cut off, and therefore this voltage is called $V_{CE(cutoff)}$. Lastly, the slope of the line is $-1/R_C$. A plot is shown in Figure 3.6.1.

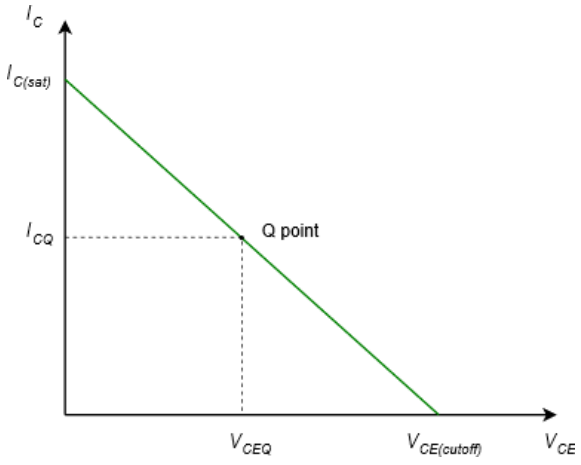


Figure 3.6.1: Generic DC load line.

To complete the graph, we also include the operating point for some specific transistor. This is called the quiescent point, or Q point, and the associated device current and voltage are called I_{CQ} and V_{CEQ}. All possible Q points lay on this line.

Referring back to Example 3.5.1 and using Equation 3.6.1, we can summarize the circuit as follows:

$$I_{C(sat)} = 15mA$$

$$V_{CE(cutoff)} = 15V$$

$$\text{Q Point for} \beta = 100 :$$

$$I_C = 4.65mA$$

$$V_{CE} = 10.35V$$

$$\text{Q Point for } \beta = 200 :$$

$$I_C = 9.3mA$$

$$V_{CE} = 5.7V$$

This is plotted in Figure 3.6.2.

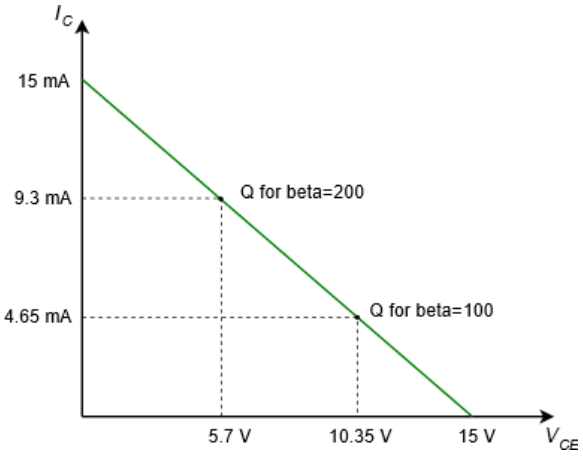


Figure 3.6.2: Load line for the variations on Example 3.5.1.

If we calculate a collector current that is greater than the saturation current, then we know that the actual current will be the saturation current maximum. For this circuit, any calculated value greater than 15 mA indicates that the transistor would produce only 15 mA (our earlier example using β = 400, for instance). In reality, the true value will be very slightly less. This is because the collector-emitter voltage does not go all the way down to zero volts when the device is saturated. Typically, *VCE*(*sat*) will be a tenth of a volt or so for small signal devices. Precise values can be determined from device graphs such as the middle graph of Figure 3.4.1c, labeled "Collector Saturation Region". As an example, if I_C = 10 mA and I_B = 0.3 mA, then $V_{CE(SAT)}$ is approximately 0.15 V. It turns out that we can use saturation to our advantage in switching circuits, as we are about see.

As mentioned, variation in β can cause changes in collector current. This can cause performance issues. For example, when driving an LED, this can lead to variance in brightness. But what if we purposely put the transistor into saturation? Saturation is a fixed value. It is inherently stable and β no longer matters. Effectively, when a BJT saturates, β is forced to drop to whatever value is needed to produce $I_C(sat)$. We just need to make sure that even the smallest β is large enough to cause saturation.

THE SATURATING SWITCH

A good example of this is the saturating LED driver circuit shown in Figure 3.7.1. To begin with, the whole point of the driver is to offload the current demand from the prior circuit. For example, we may wish to light an LED from the output of a logic gate or microcontroller chip. The problem is that those circuits might only be able to deliver, say, 5 mA when we might need well over 10 mA to achieve the desired brightness. The LED driver circuit is used to overcome this limitation.

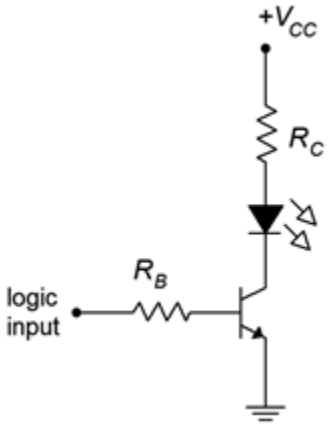


Figure 3.7.1: Saturating LED driver circuit (positive logic). Note: The negative terminal of VCC is connected to ground (not shown).

With the driver, the logic circuit will only need to supply base current, not LED current. Here is how it works: If the logic input voltage is zero, there will be no base current. This means that there will be no collector current and therefore the LED will be off. At this point the BJT is in cutoff. In contrast, when the logic level goes high, all of the logic voltage drops across R_B, with the exception of V_{BE} . This creates I_B. If properly designed, this current will be sufficient to put the BJT into saturation. The BJT acts as a switch, completing the circuit between the DC supply, the LED and the current limiting resistor, R_C . For this to work reliably, we have to make sure that the ratio of saturation current to base current is much less than β. A value of 10 or so would guarantee hard saturation.

If we would like to invert the logic, that is, have a logic low turn on the LED and a logic high turn it off, we can achieve that with a PNP version of the circuit as shown in Figure 3.7.2.

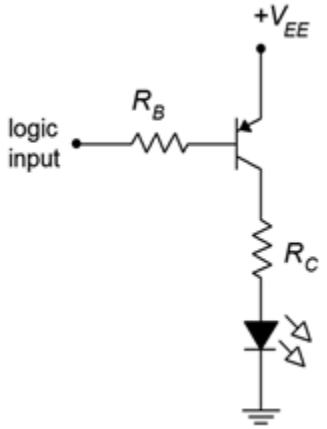


Figure 3.7.2 : Saturating LED driver circuit (negative logic).

Example 3.7.1

Determine the LED "on" current for the circuit of Figure 3.7.3. Assume the logic "on" voltage is 5 volts, V_{LED} = 1.8 volts and $V_{CE(sat)}$ = 0 .

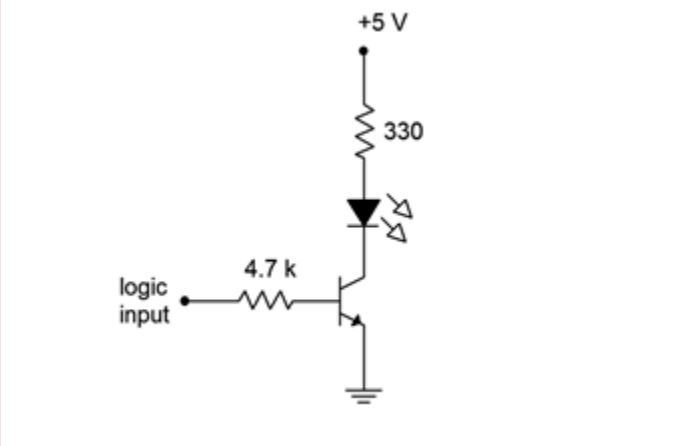


Figure 3.7.3 : Circuit for Example 3.7.1 .

First, find the base current.

$$I_B = \frac{V_{logic} - V_{BE}}{R_B}$$

$$I_B = \frac{5V - 0.7V}{4.7k\Omega}$$

$$I_B = 915\mu A$$

Now find IC(sat) , making sure the BJT is in saturation. This will be the LED current.

$$I_{C(sat)} = \frac{V_{CC} - V_{LED}}{R_C}$$

$$I_{C(sat)} = \frac{5V - 1.8V}{330\Omega}$$

$$I_{C(sat)} = 9.7mA$$

The ratio of these two currents is just over 10:1. This will guarantee hard saturation.

There are many different applications for saturating switches. Just about anywhere you can imagine a relay being used, you can consider a transistor switch. The transistor switch has the advantages of small size, no moving parts to wear out and very fast switching speeds. Relays have the advantage for very high currents. Figure 3.7.4 shows an example of direct motor drive using a saturating BJT switch.

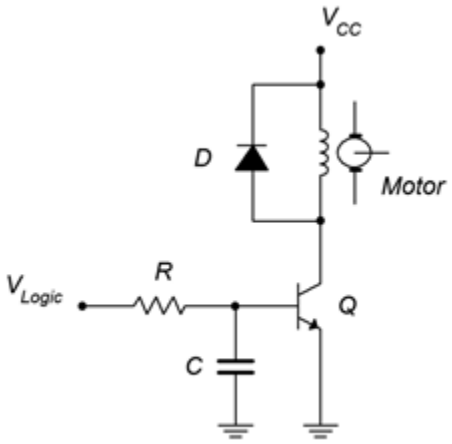


Figure 3.7.4: Direct DC motor drive.

This circuit is used to control the speed of a DC motor through a technique called pulse width modulation. The speed of the motor will depend on the average voltage applied to it. The trick here is that instead of applying a continuously variable voltage to the motor, we instead apply a series of pulses of varying width. These pulses are sufficient to saturate the BJT, causing it to behave as a switch. These pulses are so fast that the motor does not start and stop, but rather inertia keeps it going. Instead, the motor responds to the averaged value of these pulses. If the pulses are narrow and widely spaced, the average value will be low and the motor speed will be slow. If the pulses are wide and closely spaced, the average will high and the motor speed will be fast.

The resistor and capacitor at the base are used to shape the incoming pulse to improve performance. The diode across the motor winding is particularly important. It is referred to as a snubbing diode.[1] Without it, the switching transistor might experience large and damaging transient spikes. Here's

1. It is also known as a commutating diode, clamp diode, flyback diode and by a host of other names. But as Shakespeare said, "A snubbing diode by any other name would clamp a flyback voltage as well". Or something like that.

why: Let's assume the BJT is on and conducting fully. This current is the same current flowing through the motor's armature, which is little more than a huge coil of wire. That means it exhibits a lot of inductance.

When we turn off the transistor, we are attempting to turn off the armature current, but the current through an inductor cannot change instantaneously. The result is that the winding now generates a large flyback voltage (also called an "inductive kick") directly across the BJT. That is, the winding momentarily appears as a high voltage source of opposite polarity and, via KVL, this potential appears from collector to emitter. This could damage the BJT. The snubbing diode effectively short-circuits the winding when it reverses voltage polarity, preventing the large spike. The remainder of the time the diode is reverse-biased and effectively out of the circuit.

THE NON-SATURATING DRIVER

It is also possible to create a switch or driver that is non-saturating. An example of a non-saturating LED driver is shown in Figure 3.7.5

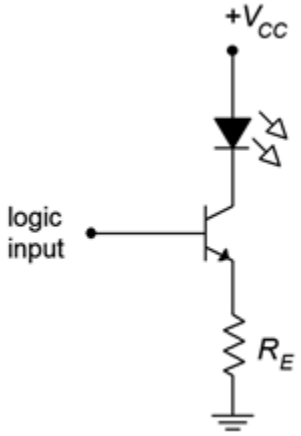


Figure 3.7.5: Non-saturating LED driver circuit (positive logic).

The advantage of this circuit is that it requires less current from the logic circuit. Unfortunately, it also exhibits higher transistor power dissipation and requires a DC source that is higher than the logic level. The operation is as follows: As in the saturating driver, if the logic level is zero, there is no rise in the base-emitter loop and the collector current will also be zero. With a high logic voltage, via KVL around the base-emitter loop, all of the logic input voltage drops across R_E, with the exception of V_{BE} . This creates I_E which is virtually the same as I_C (which is I_{LED}).

This circuit "programs" the emitter current via the resistor and logic voltage. Therefore it is fixed and stable. This process is sometimes referred to as bootstrapping. It might be said that the emitter voltage is "bootstrapped" to within 0.7 volts of the logic input level, keeping it stable.[2] In any case, if β varies, this will cause an inverse change in I_B with no change in I_C . A negative logic PNP version is also possible and left as an exercise.

2. This is in reference to the old phrase "pulling yourself up by your bootstraps". To be honest, that saying never made sense to this author and all that ever happened when I tried to do it was that my arms got tired.

Example 3.7.2

Determine the LED "on" current for the circuit of Figure 3.7.6. Assume the logic "on" voltage is 5 volts, VLED = 1.8 volts and β = 100.

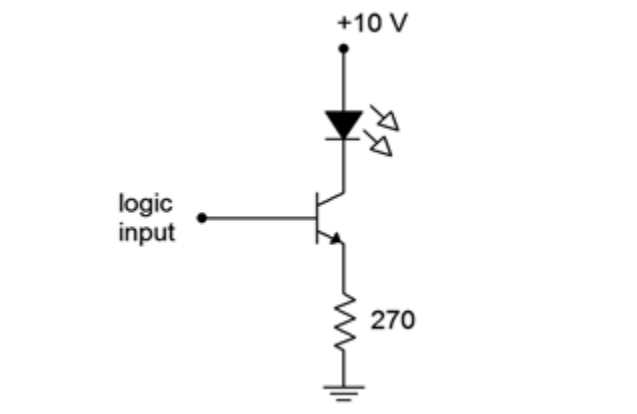


Figure 3.7.6 : Circuit for Example 3.7.2 .

We can find IC directly because IC ≈ IE . This will be the LED current.

$$I_C = \frac{V_{logic} - V_{BE}}{R_E}$$

$$I_C = \frac{5V 0.7V}{270\Omega}$$

$$I_C = 15.9mA$$

Note that β was not used. All it tells us is that IB = 15.9 mA/100, or 159 μA. A higher β would simply lead to a lower base current.

For the sake of completeness, we might also note that

$$V_{CE} = V_{CC} - V_{LED} - V_{RE}$$

$$V_{CE} = 10V - 1.8V - 4.3V$$

$$V_{CE} = 3.9V$$

Clearly, if VCE is 3.9 volts, the transistor is not in saturation.

3.7.3: THE ZENER FOLLOWER

In the prior chapter we examined a method of regulating the output voltage of a filtered full-wave rectifier through the use of a Zener diode. The downfall of that specific circuit is that it was not particularly efficient because it drew a fair amount of current even when the demand for load current

was light. Using the concept of locking one voltage to another, as in the non-saturating switch, we can create a nice improvement, the Zener Follower.

A Zener Follower is shown in Figure 3.7.7. The input signal is the positive rectified and filtered output of the AC-to-DC power supply.

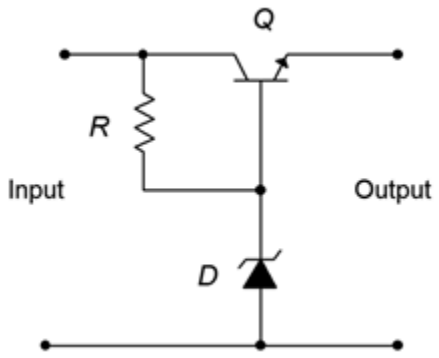


Figure 3.7.7: Zener Follower.

The primary thing to notice is that the Zener diode is reverse-biased via resistor *R*. That is, a current will flow down through *R* and into the Zener. The Zener presents a fixed potential, V_Z . Consequently, the difference between the input voltage and V_Z must drop across *R*, and by extension, V_{CB} . Further, the final output voltage is the voltage at the BJT's emitter which must be $V_Z - V_{BE}$. As these are both fixed, stable potentials, the output must likewise be a fixed, stable voltage. Lastly, because $V_{CE} = V_{CB} + V_{BE}$, it is apparent that any variation between the input voltage and the desired output (for example, due to ripple) must be dropped across the BJT.

The diode current is kept low in the Zener Follower and thus its power dissipation is also modest. Further, current draw from the input circuit is a direct reflection of load current demand. If the load current requirement is low, very little current will flow through the transistor, and ultimately, from the input circuit. This makes for a more efficient system.

3.8 SUMMARY

A bipolar junction transistor may be thought of as an extension of a simple diode or PN junction. Another layer of doped material is added, resulting in either an NPN or PNP configuration, both with two depletion regions. The two depletion regions create two hills in the energy diagram. The three terminals of the device are called the emitter, base (middle) and collector. BJTs are not normally constructed symmetrically and swapping the collector and emitter can result in unpredictable behavior.

For proper operation, the base-emitter junction is forward-biased while the collector-base junction is reverse-biased. This results in the emitter and collector currents being very nearly equal and much, much larger than the base current. The ratio of collector current to base current is called β (beta) while the ratio of collector current to emitter current is called α (alpha). β in particular is subject to wide variations and it can have a major impact on circuit parameters. A plot of collector current versus collector-emitter voltage reveals the three main regions of the BJT circuit: saturation, constant current and breakdown.

The Ebers-Moll model consists of a diode from the base to emitter and a controlled current source from the collector to base. This simple model of the BJT can be used to solve a variety of transistor circuits, particularly when used in conjunction with a DC load line. The DC load line is a plot of all possible operating points for a given transistor circuit.

Finally, it is possible to create switching and driver circuits using BJTs that produce stable output currents. These may utilize saturating or non-saturating configurations with NPN or PNP devices.

Review Questions

1. Describe the energy diagram for a forward-reverse biased BJT.
2. Define α
3. Define β
4. Define Early voltage. What is its significance?
5. What is a family of collector curves? What information can we derive from it?
6. Describe the Ebers-Moll BJT model.
7. Explain some of the issues involving variation of β.
8. What is a DC load line?
9. How is a saturating switch different from a non-saturating driver? What are the advantages and

disadvantages of each?

10. What is a Zener Follower?

3.9 EXERCISES

Assume diodes are silicon unless stated otherwise

ANALYSIS PROBLEMS

1. Determine β if $\alpha = 0.99$
2. Determine α if $\beta = 200$
3. Determine the currents for the circuit of Figure 3.9.1 if $V_{BB} = 5$ V, $V_{CC} = 20$ V, $R_B = 200$ kΩ, $R_C = 2$ kΩ, $\beta = 100$.
4. Determine the transistor voltages for the circuit of Figure 3.9.1 if $V_{BB} = 5$ V, $V_{CC} = 20$ V, $R_B = 200$ kΩ, $R_C = 2$ kΩ, $\beta = 200$.

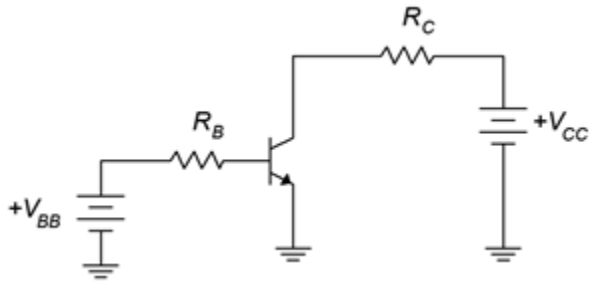


Figure 3.9.1

5. Determine the LED current in the circuit of Figure 3.9.2 if $V_{logic} = 5$ V, $V_{CC} = 5$ V, $V_{LED} = 2.1$ V, $R_B = 3.6$ kΩ, $R_C = 270$ Ω, $\beta = 100$.
6. Determine the LED current in the circuit of Figure 3.9.2 if $V_{logic} = 0$ V, $V_{CC} = 5$ V, $V_{LED} = 2.1$ V, $R_B = 3.6$ kΩ, $R_C = 270$ Ω, $\beta = 100$.
7. Determine the LED current in the circuit of Figure 3.9.3 if $V_{logic} = 5$ V, $V_{EE} = 5$ V, $V_{LED} = 2.2$ V, $R_B = 2.7$ kΩ, $R_C = 220$ Ω, $\beta = 100$.

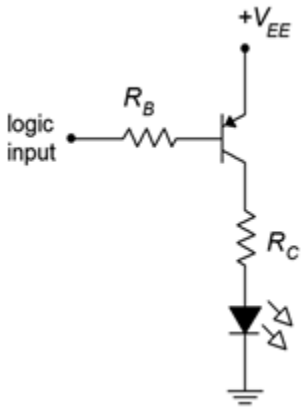


Figure 3.9.3

8. Determine the LED current in the circuit of Figure 3.9.3 if V_{logic} = 0 V, V_{EE} = 5 V, V_{LED} = 2.2 V, R_B = 2.7 kΩ, R_C = 220 Ω, β = 100.
9. Determine the LED current in the circuit of Figure 3.9.4 if V_{logic} = 3.6 V, V_{CC} = 10 V, V_{LED} = 2.3 V, R_E = 270 Ω, β = 200.

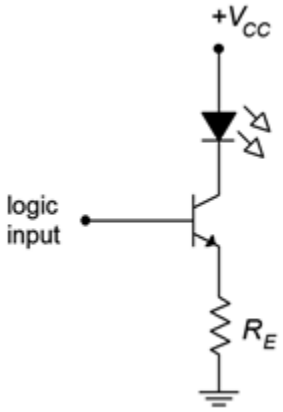


Figure 3.9.4

10. Determine the LED current in the circuit of Figure 3.9.4. V_{logic} = 0 V, V_{CC} = 10 V, V_{LED} = 2.3 V, R_E = 270 Ω, β = 200.
11. Using the 2N3904 data sheet, determine $V_{CE(sat)}$ if I_C = 30 mA and I_B = 1mA.
12. Using the 2N3904 data sheet, determine the percent change in β if I_C = 10 mA and the temperature rises from 25°C to 125°C.
13. Using the 2N3904 data sheet, determine the percent change in β if I_C = 40 mA and the temperature drops from 25°C to –55°C.

DESIGN PROBLEMS

14. Using Figure 3.9.1, determine a value for R_B to set I_C to 5 mA if V_{BB} = 5 V, V_{CC} = 25 V, R_C = 2 kΩ, β = 100.
15. Using Figure 3.9.1, determine a value for R_C to set V_{CE} to 6 V if V_{BB} = 10 V, V_{CC} = 25 V, R_B = 330 kΩ, β = 200.

16. For the circuit of Figure 3.9.2, determine a value for R_C to set the LED current to 15 mA. V_{logic} = 5 V, V_{CC} =5 V, V_{LED} = 1.6 V, R_B = 3.3 kΩ.
17. For the circuit of Figure 3.9.3, determine a value for R_C to set the LED current to 20 mA. V_{logic} = 0 V, V_{EE} = 5 V, V_{LED} = 2.0 V, R_B = 2.7 kΩ.
18. For the circuit of Figure 3.9.4, determine a value for R_E to set the LED current to 25 mA. V_{logic} = 5 V, V_{CC} = 9 V, V_{LED} = 2.8V.

CHALLENGE PROBLEMS

19. Determine the maximum and minimum values for I_C in the circuit of Figure 3.9.1 if all resistors have a 10% tolerance and β = ranges from 100 to 200. V_{BB} =5 V, V_{CC} = 20 V, R_B = 200 kΩ, R_C = 2 kΩ.
20. Derive and draw a PNP non-saturating LED driver circuit.

COMPUTER SIMULATION PROBLEMS

21. Simulate the circuit of Problem 3.
22. Simulate the circuit of Problem 5.
23. Simulate the circuit of Problem 7.
24. Verify the design of Problem 14 using a simulator.

UNIT 4: BJT BIASING

Learning Objectives

After completing this chapter, you should be able to:

- Explain the need for DC biasing of BJT amplifiers.
- Solve various BJT biasing circuits for device currents and voltages.
- Plot DC load lines for a variety of BJT biasing circuits.
- Discuss methods to increase circuit stability with regard to transistor parameter variation.

4.1 INTRODUCTION

As we saw in the preceding chapter, a bipolar junction transistor requires a forward-bias of the base-emitter junction and a reverse- bias of the collector-base junction in order to operate properly. One of the prime BJT parameters is the current gain, β. It can have a considerable impact on the operation of the circuit. Unfortunately, β also varies with changes in temperature, collector-emitter voltage, etc., and this can lead to circuit instability. In this chapter we shall investigate a variety of circuit topologies to bias the BJT, always with an eye toward stability.

4.2 THE NEED FOR BIASING

Why bias a transistor in the first place? After all, if the device exhibits current gain (i.e., β), why not just apply an AC signal at the base and obtain an amplified version of it at the collector? The first thing to remember is that current gain is an outgrowth of forward-reverse bias. Given that fact, and without an additional source of energy, amplification cannot be produced. Also, remember the magnitude of the energy hill required for forward-biasing the base-emitter. In order to achieve that, V_{BE} needs to be around 0.7 volts. If we simply applied an AC signal to the base, we could only hope to forward-bias the base-emitter when that signal exceeded 0.7 volts. The entire negative half of the AC signal would be ignored along with everything positive that's below 0.7 volts.

Seeing that the voltage generated from many input devices such as microphones and sensors may only be a few hundred millivolts, the entire signal could be ignored! The solution to these problems is to apply a DC bias to the transistor and then superimpose the AC signal on top of that. In other words, if the AC voltage is riding on a much larger DC voltage, then even the negative peak of the AC signal will be a net positive voltage, and we can maintain proper transistor function.

There are numerous ways to establish a proper polarity DC bias on a transistor. The trick is to find ways to make a stable bias, that is, to establish a Q point that doesn't move in spite of parameter changes such as changes in β. As we shall see in following chapters, an unstable Q point can have negative effects on the AC performance of an amplifier. For example, it could make the gain unstable, increase distortion or reduce output power. This lack of stability is a major problem with the base bias configuration examined in the prior chapter. What we would like is a circuit that will establish a collector current that does not shift even when β changes.

4.3 TWO-SUPPLY EMITTER BIAS

For proper functioning, the collector-base junction needs to be reverse-biased and the base-emitter junction needs to be forward- biased. For an NPN transistor that means that the collector must be at the highest potential, the base somewhat lower and the emitter at the lowest potential of the three. One way of doing this is to apply the usual positive supply to the collector, but instead of using a second potential at the base, the base is tied to ground through a resistor. The requisite forward-bias on the base-emitter is then achieved by connecting the emitter to a negative power supply. This circuit configuration is shown in Figure 4.3.1 using an NPN device. We shall refer to this as two-supply emitter bias.

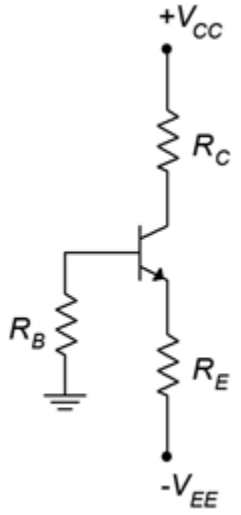


Figure 4.3.1: Two-supply emitter bias, NPN version.

We can derive an equation for the collector current by applying KVL to the baseemitter loop:

$$V_{EE} = V_{R_B} + V_{BE} + V_{R_E}$$

$$V_{EE} = I_B R_B + V_{BE} + I_E R_E$$

Recalling that $I_B = I_C/\beta$ and $I_E \approx I_C$,

$$V_{EE} = (I_C/\beta)R_B + V_B E + I_C R_E$$

Solving for I_C we arrive at

$$I_C = \frac{|V_{EE}|V_{BE}}{R_E + R_B/\beta}$$

(4.3.1)

The absolute value has been added to the emitter supply voltage so there is no confusion regarding the sign of this potential in the equation.

The thing to notice about Equation 4.3.1 is that $\beta R_E \gg R_B/\beta$, then the equation reduces to only partly determines the collector current. In fact, if we can make $R_E \gg R_B/\beta R_E \gg R_B/\beta$, then the equation reduces to

$$I_C \approx \frac{|V_{EE}| - V_{BE}}{R_E}$$

(4.3.2)

It is relatively easy to achieve the $R_E \gg R_B/\beta$ stipulation. Given typical values of β, this will be the case if R_E is approximately equal to or larger than R_B. What we find in this instance is that almost all of the emitter supply drops across R_E to establish a stable I_C with β playing virtually no role. If β changes, the result will be an inverse change in I_B with I_C remaining largely unchanged.

Now that we have the collector current, any other current or voltage in the circuit may be derived by applying Ohm's law, KVL and the like. For example, to find V_C , the voltage from the collector to ground,

$$V_C = V_{CC} - V_{R_C}$$

$$V_C = V_{CC} - I_C R_C$$

And to find the transistor's collector-emitter voltage, V_{CE} ,

$$V_{CE} = V_{CC} + |V_{EE}| - V_{RC} - V_{RE}$$

$$V_{CE} = V_{CC} + |V_{EE}| - I_C R_C - I_C R_E$$

$$V_{CE} = V_{CC} + |V_{EE}| - I_C(R_C + R_E)$$

Note that V_{CE} can also be found via $V_{CE} = V_C - V_E$. Dropping voltages along the base-emitter loop yields

$$V_E = -V_{R_B} - V_{BE}$$

$$V_E = I_B R_B - V_{BE}$$

(4.3.4)

Also, it is to our advantage to develop the DC load line for this configuration. The load line can serve as a "sanity check" for our computations. To find the endpoints, $I_C(sat)$ is the maximum current and will occur when $V_{CE} = 0$. If we imagine the current rising as *VCE* collapses, eventually all of the available supply voltage will have dropped across *RC* and R_E. Thus

$$I_{C(sat)} = \frac{V_{CC} + |V_{EE}|}{R_C + R_E}$$

(4.3.4)

Similarly, *VCE*(*cutoff*) occurs when I_C = 0. That means that there will be no potentials across R_C and R_E. Therefore, V_{CE} "absorbs" the entire available source voltage.

$$V_{CE(cutoff)} = V_{CC} + |V_{EE}|$$

(4.3.5)

Do not attempt to memorize all of the myriad equations presented. There are simply too many variations on the theme and it will only get worse when other biasing configurations are introduced. Instead, remember how to find the collector current and then get in the habit of applying Ohm's law and KVL to derive whatever else you may need.

At this point a comprehensive example is called for.

Example 4.3.1

Assuming β = 100, plot the Q point (I_C and V_{CE}) on the load line for the circuit of Figure 4.3.2.

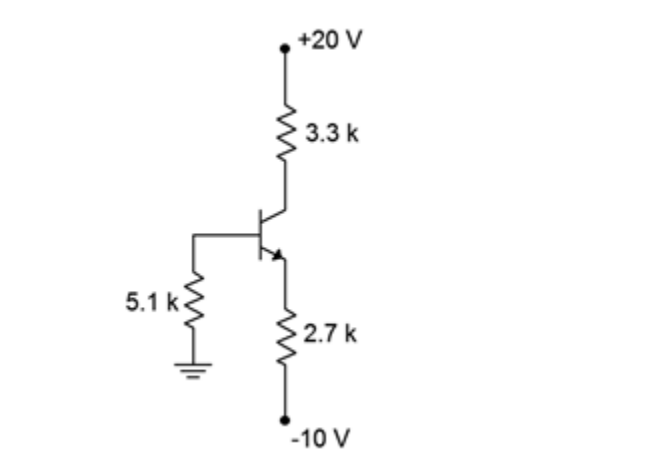


Figure 4.3.2 : Circuit for Example 4.3.1 .

Using Equation 4.3.1:

$$I_C = \frac{|V_{EE}| - V_{BE}}{R_E + R_B/\beta}$$

$$I_C = \frac{10 - 0.7}{2.7k\Omega + 4.1k\Omega/100}$$

$$I_C = 3.38mA$$

Noting the relative sizes of R_E and R_B, the approximation should be close.

$$I_C = \frac{|V_{EE}| - V_{BE}}{R_E}$$

$$I_C = \frac{10 - 0.7}{2.7k\Omega}$$

$$I_C = 3.44mA$$

To find VCE we can use the equation derived above (Equation 4.3.3).

$$V_{CE} = V_{CC} + |V_{EE}| - I_C(R_C + R_E)$$

$$V_{CE} = 20V + 10V - 3.38mA(3.3k\Omega + 2.7k\Omega)$$

$$V_{CE} = 9.72V$$

Now calculate the load line endpoints:

$$I_{C(sat)} = \frac{V_{CC} + |V_{EE}|}{R_C + R_E}$$

$$I_{C(sat)} = \frac{20V + 10V}{3.3K\Omega + 2.7K\Omega}$$

$$I_{C(sat)} = 5mA$$

$$V_{CE(cutoff)} = V_{CC} + |V_{EE}|$$

$$V_{CE(cutoff)} = 20V + -10V$$

$$V_{CE(cutoff)} = 30V$$

The load line for the circuit in Example 4.3.1 is shown in Figure 4.3.3.

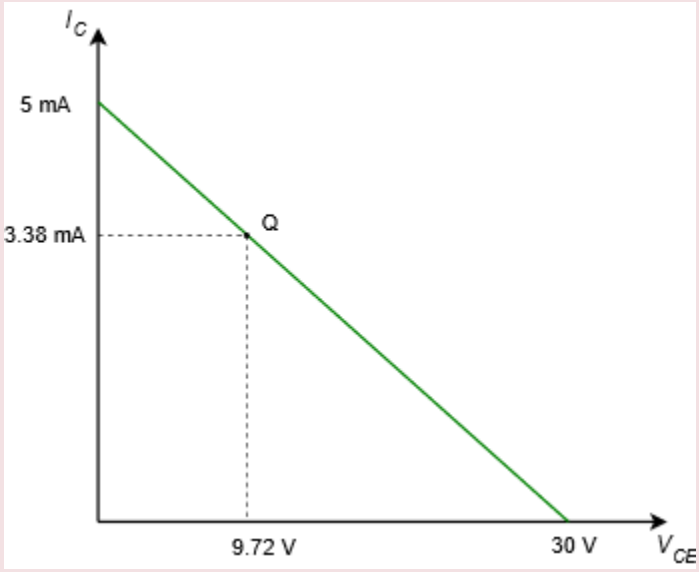


Figure 4.3.3: DC load line for the circuit of Figure 4.3.2.

Note the proportions between voltage and current for the Q point. The voltage is a little less than one-third of the maximum while the current is a little more than two-thirds of its maximum.

VERIFICATION OF STABILITY

The claim was made that two-supply emitter bias circuits like the one Figure 4.3.1 potentially have a stable Q point. If we were to plot a second Q point for a large change in β, it should hardly move, thus indicating very high stability. For example, doubling β to 200 results in I_C = 3.41 mA and V_{CE} = 9.53 V. The new Q point has edged just slightly closer to saturation, producing about a 1% change in current for a 100% change in β. Clearly, this configuration can produce very small changes in the Q point in spite of very large changes in β.

COMPUTER SIMULATION

The two-supply emitter bias circuit of Figure 4.3.4 is simulated using the DC Bias function. A quick estimation shows that we expect about 2 mA of collector current (9.3 V/4.7 kΩ) and a collector voltage of about 8 volts (15 V – 2 mA 3.6 kΩ). We also expect a small negative potential at the base ($-I_B R_B$). Given typical β values for the 2N3904 (200-ish at this current, refer back to the data sheet), we expect a base current of around 10 to 15 μA, leaving us with a V_B of approximately –0.1 volts. The emitter voltage would be about 0.7 volts less than that, perhaps –0.8 volts or so.

In short, for a properly designed circuit of this type we expect V_B to be pretty close to 0 V and V_E to be about –0.7 volts.

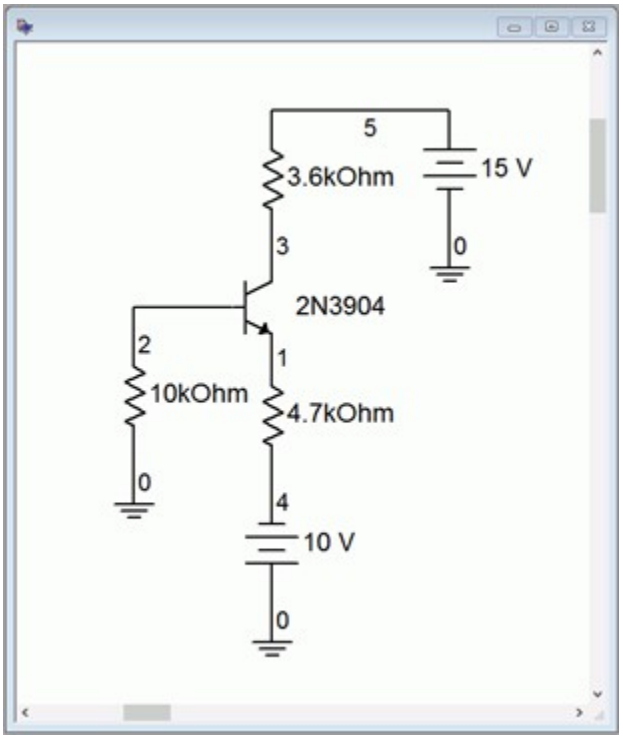


Figure 4.3.4: Schematic for two-supply emitter bias simulation.

The results are shown in Figure 4.3.5. The node voltages agree with our estimations. Node 3 is the collector voltage, very close to the estimation. The results for the base voltage (node 2) and the emitter voltage (node 1) are also in line with the estimates.

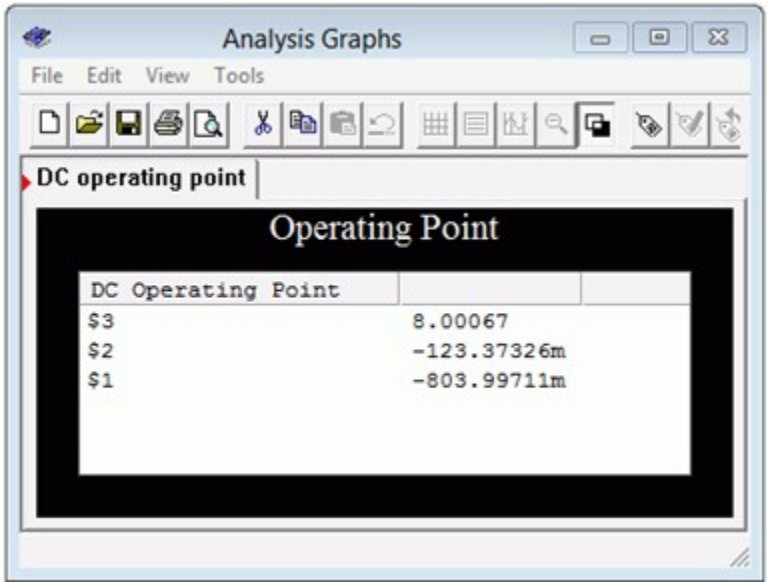


Figure 4.3.5: Simulation results for twosupply emitter bias circuit

PNP TWO-SUPPLY EMITTER BIAS

While it is possible to create a PNP version of bias circuits by simply swapping out the device and then changing the signs of the power supplies, it is common to "flip" the entire circuit from top to bottom so that the emitter winds up on top and the collector on the bottom. One advantage of this is

that, in a multi-transistor circuit schematic, all of the DC bias currents "run down the page", that is, the collector currents flow from the top of the page to the bottom of the page. Figure 4.3.6 shows a PNP two-supply emitter bias circuit.

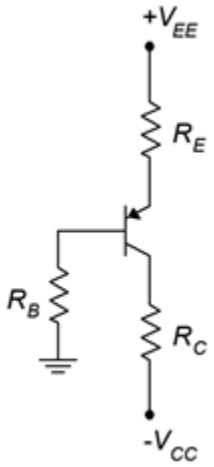


Figure 4.3.6: PNP two-supply emitter bias circuit.

All of the device current equations and component voltage equations derived for the NPN version will hold for the PNP version. The differences to remember are that the voltage polarities will be reversed (what was positive in the NPN is negative in the PNP) and the current directions will be reversed (e.g., conventional current flows into the NPN's collector but out of the PNP's collector. For example, in the NPN we expect the base current to flow into the base terminal. This creates a small negative voltage at the base and a somewhat more negative voltage (by 0.7 V) at the emitter. In the PNP, the base current flows out of the base. This creates a small positive voltage at the base and results in the emitter being slightly more positive (by 0.7 V). This is perhaps best illustrated with an example.

Example 4.3.2

Assuming β = 100, determine the Q point and load line endpoints of the circuit of Figure 4.3.7.

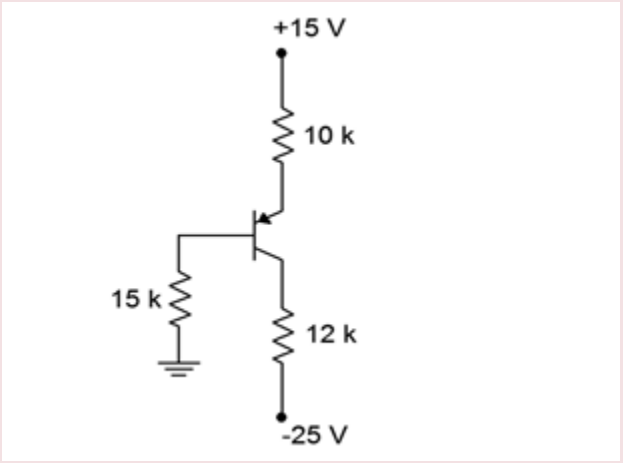


Figure 4.3.7: Circuit for Example 4.3.2.

First, note that this is a PNP drawn upside down so the emitter is at the top. Using Equation 4.3.1:

$$I_C = \frac{|V_{EE}| - V_{BE}}{R_E + R_B/\beta}$$

$$I_C = \frac{15 - 0.7}{10k\Omega + 15k\Omega/100}$$

$$I_C = 1.409mA$$

As a cross check, noting the relative sizes of R_E and R_B, the approximation should be close.

$$I_C = \frac{|V_{EE}| - V_{BE}}{R_E + R_B/\beta}$$

$$I_C = \frac{15 - 0.7}{10k\Omega}$$

$$I_C = 1.43mA$$

To find V_{CE} we can use Equation 4.3.3 with a slight modification.

$$V_{CE} = V_{EE} + |V_{CC}| - I_C(R_C + R_E)$$

$$V_{CE} = 15V + 25V - 1.409mA(12k\Omega + 10k\Omega)$$

$$V_{CE} = 9V$$

We complete the picture by determining the endpoints of the load line.

$$I_{C(sat)} = \frac{V_{EE} + |V_{CC}|}{R_C + R_E}$$

$$I_{C(sat)} = \frac{10V + 25V}{12K\Omega + 10K\Omega}$$

$$I_{C(sat)} = 1.818mA$$

$$V_{CE(cutoff)} = V_{EE} + |V_{CC}$$

$$V_{CE(cutoff)} = 15V + 25V$$

$$V_{CE(cutoff)} = 40V$$

The Q point is about 3/4ths of the maximum current and 1/4th of the maximum voltage.

4.4 VOLTAGE DIVIDER BIAS

Another configuration that can provide high bias stability is voltage divider bias. Instead of using a negative supply off of the emitter resistor, like two-supply emitter bias, this configuration returns the emitter resistor to ground and raises the base voltage. So as to avoid issues with a second power supply, this base voltage is derived from the collector power supply via a voltage divider. The bias template is shown in Figure 4.4.1.

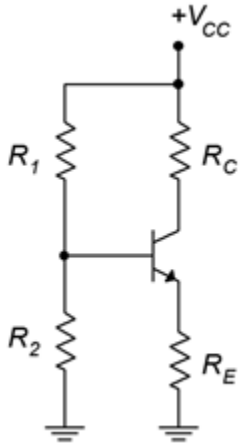


Figure 4.4.1 : Voltage divider bias.

Let's derive the equations for the load line. First, let's consider the saturation and cutoff endpoints. For saturation, assume V_{CE} goes to 0. What resistances are left to limit the current?

$$I_{C(sat)} = \frac{V_{CC}}{R_C + R_E}$$

(4.4.1)

V_{CE}(*cutoff*) occurs when I_C = 0 and that means that there will be no potentials across R_C and R_E. Therefore, V_{CE} takes on the entire available source voltage.

$$V_{CE(cutoff)} = V_{CC}$$

(4.4.2)

The key to finding the Q point (and pretty much any other current or voltage in the circuit) is to find I_C . To simplify the process, Thevenize the voltage divider as shown in Figure 4.4.2.

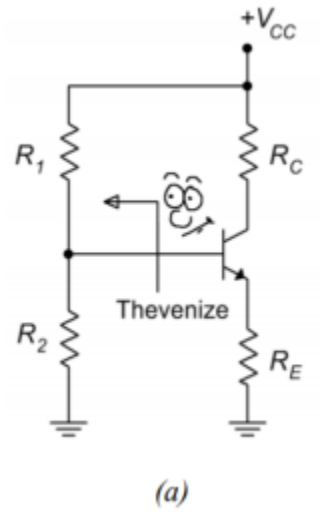


Figure 4.4.2a: Thevenizing the voltage divider.

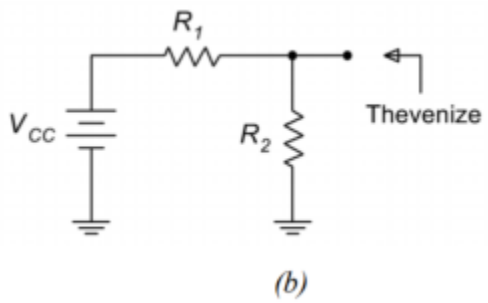


Figure 4.4.2b: Thevenizing the voltage divider.

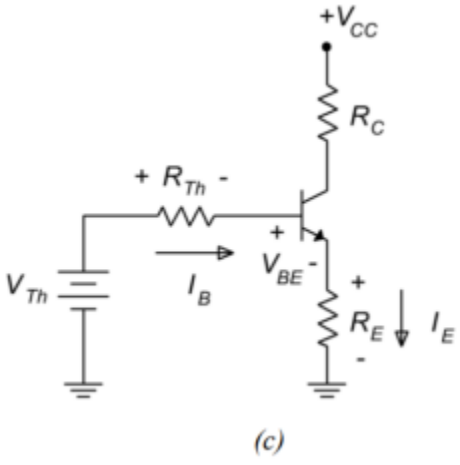


Figure 4.4.2c: Thevenizing the voltage divider.

By inspection of Figure 4.4.2*b*,

$$V_{TH} = V_{CC}\frac{R_2}{R_1 + R_2}$$

$$R_{TH} = R_1 || R_2 = \frac{R_1 R_2}{R_1 + R_2}$$

Now we can derive an equation for the collector current by applying KVL to the base-emitter loop of Figure 4.4.2*c*:

$$V_{TH} = V_{R_{TH}} + V_{BE} + V_{R_E}$$

$$V_{TH} = I_B R_{TH} + V_{BE} + I_E R_E$$

Recalling that $I_B = I_C/\beta$ and $I_E \approx I_C$,

$$V_{TH} = (I_C/\beta) R_{TH} + V_{BE} + I_C R_E$$

Solving for I_C we arrive at

$$I_C = \frac{V_{TH} - V_{BE}}{R_E + R_{TH}/\beta}$$

(4.4.3)

Can we find a quick approximation for I_C as well? If we assume that the voltage divider of R_1 and R_2 is lightly loaded, in other words, that the divider current is much, much less than the base current, finding I_C is easy. The divider voltage yields the base voltage. We then subtract the 0.7 volt drop on the base-emitter and what's left drops across R_E. From there it's one short application of Ohm's law to get I_E, which is approximately equal to I_C . But how do we know if the divider is lightly loaded in the first place without going through the Thevenin equivalent? Looking at Equation 4.4.3, as long as $R_E \gg R_{TH}/\beta$, we can ignore the second term in the denominator, leaving us with our quick approximation. Given typical values for β, as long as R_2 is not much larger than R_E, the approximation will be reasonably accurate.

Once I_C is obtained we can find the transistor's collector-emitter voltage, V_{CE} ,

$$V_{CE} = V_{CC} - V_{R_C} - V_{R_E}$$

$$V_{CE} = V_{CC} - I_C R_C - I_C R_E$$

$$V_{CE} = V_{CC} - I_C (R_C + R_E)$$

(4.4.4)

Time for yet another thrilling illustrative example.

Example 4.4.1

Assuming β = 200, plot the Q point (IC and VCE) on the load line for the circuit of Figure 4.4.3. Also determine the value of V_B.

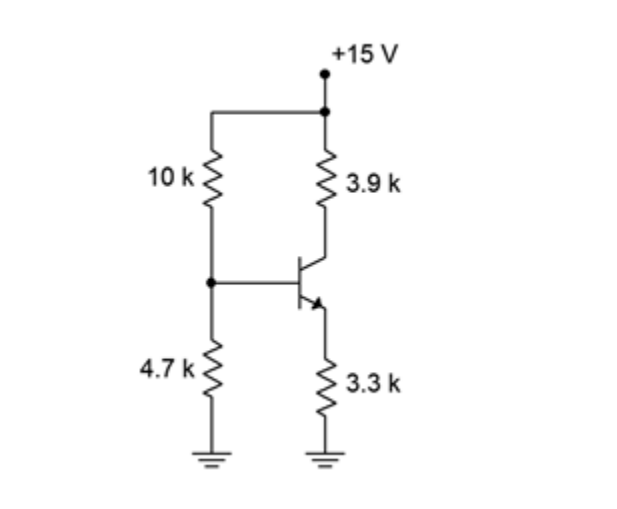


Figure 4.4.3: Circuit for Example 4.4.1.

Calculate the load line endpoints so we know the maximums.

$$I_{C(sat)} = \frac{V_{CC}}{R_C + R_E}$$

$$I_{C(sat)} = \frac{15V}{3.9K\Omega + 3.3K\Omega}$$

$$I_{C(sat)} = 2.08mA$$

$$V_{CE(cutoff)} = V_{CC}$$

$$V_{CE(cutoff)} = 15V$$

To obtain the Q point, first find the Thevenin values.

$$V_{TH} = V_{CC}\frac{R_2}{R_1 + R_2}$$

$$V_{TH} = 15V\frac{4.7k\Omega}{10k\Omega + 4.7k\Omega}$$

$$V_{TH} = 4.8V$$

$$R_{TH} = R_1 \| R_2 = \frac{R_1 R_2}{R_1 + R_2}$$

$$R_{TH} = \frac{10k\Omega \times 4.7k\Omega}{10k\Omega + 4.7k\Omega}$$

$$R_{TH} = 3.2k\Omega$$

Using Equation 4.4.3:

$$I_C = \frac{V_{TH} - V_{BE}}{R_E + R_{TH}/\beta}$$

$$I_C = \frac{4.8V - 0.7V}{3.3k\Omega + 3.2k\Omega/200}$$

$$I_C = 1.236mA$$

Noting the relative sizes of R_E and R_2, the approximation should be fairly accurate.

$$I_C = \frac{V_{TH} - V_{BE}}{R_E}$$

$$I_C = \frac{4.8V - 0.7V}{3.3k\Omega}$$

$$I_C = 1.242mA$$

To find V_{CE} we can use Equation 4.4.4.

$$V_{CE} = V_{CC} - I_C(R_C + R_E)$$

$$V_{CE} = 15V - 1.236mA(3.9k\Omega + 3.3k\Omega)$$

$$V_{CE} = 6.1V$$

As far as finding V_B is concerned, a decent approximation would be the value of V_{TH} because we have determined that the divider is lightly loaded. In a more general sense, we could also find the drop across R_E and then add V_{BE}. The approximation yields 4.8 volts and the more accurate method yields

$$V_B = V_{BE} + I_C R_E$$

$$V_B = 0.7V + 1.236mA \times 3.3k\Omega$$

$$V_B = 4.78V$$

The load line for the circuit in Example 5=4.4.1 is shown in Figure 4.4.4.

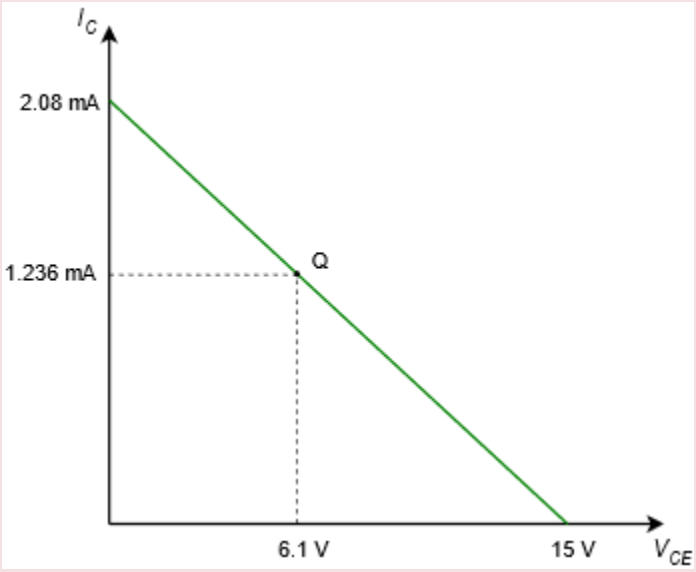


Figure 4.4.4 : DC load line for the circuit of Figure 4.4.3 .

Once again the proportions between voltage and current for the Q point appear to be proper when compared against the endpoints.

VERIFICATION OF STABILITY

How much does the Q point move if β were to get cut half? Recalculating with a β of 100 yields I_C = 1.23 mA and V_{CE} = 6.14V. This represents a shift in both current and voltage of less than 1%. This will, of course, cause a near doubling of I_B but this will be hardly noticed here as the divider current is so much larger; approximately 15V/ (10 k + 4.7 k) or 1 mA versus about 1.23 mA/100 or 12.3 μA.

PNP VOLTAGE DIVIDER BIAS

To create the PNP version of the voltage divider bias, we replace the NPN with a PNP and then change the sign of the power supply. As mentioned with the two-supply emitter bias, these circuits are usually flipped top to bottom resulting in the flow of DC current going down the page. All of the currents and component voltages are unchanged except that their directions and polarities are reversed. The current equations and so forth remain valid. Something a little odd-looking happens with the voltage divider bias, though: we end up with ground being the most positive potential and a negative supply at the bottom of the schematic. It works, but it's an issue if we're using a traditional positive supply elsewhere in the circuit. After all, why have two supplies where one will do? It turns out that we can make a positive supply version fairly easily. All we need to do is add the magnitude of the negative source voltage to the ground and power connections. This progression is shown in Figure 4.4.5.

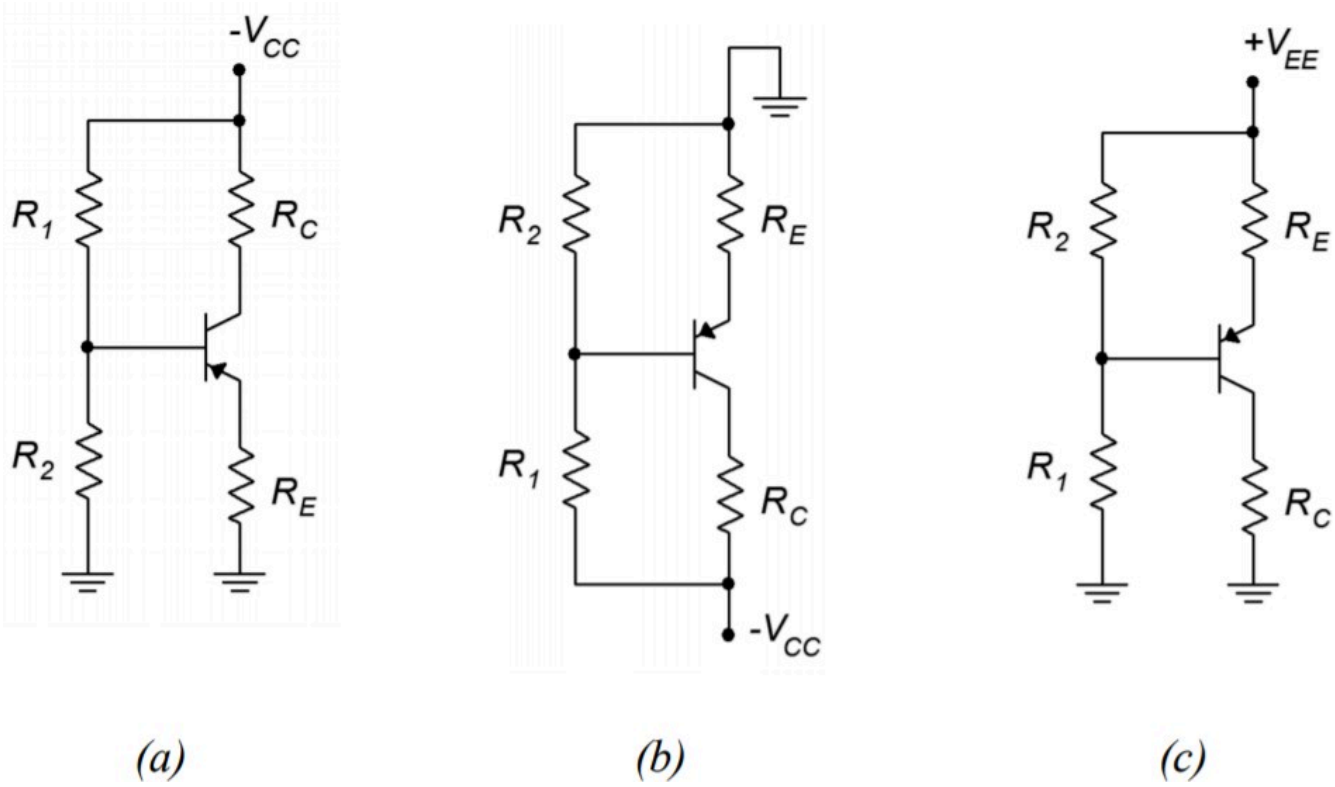


Figure 4.4.5: Progression of PNP voltage divider bias circuit. a. Direct conversion from NPN. b. Top-to-bottom flip. c. DC supply offset added to achieve a positive supply.

There is nothing magic about this procedure. In essence, all we've really done is renamed the reference point. All of the individual component voltages remain unchanged. For example, looking at Figure 4.4.5*c* versus 4.4.5*b*, it is still the case that the top connection to R_E is more positive than the bottom connection to R_C by the voltage V_{CC} (although we did rename the supply to V_{EE} to be consistent with where it's connected). What has happened is that all ground-referenced (i.e., single subscript) voltages have changed. For example, V_B in Figures 4.4.5*a* and 4.4.5*b* is the voltage across R_2 . In contrast, V_B in Figure 4.4.5*c* is the voltage across R_1 . That makes sense. If we move the reference then any voltage that is measured against the reference will change.

When analyzing the PNP voltage divider, we could simply parrot the collector current formula developed for the NPN, but there are other techniques. Two methods are illustrated in the following example.

Example 4.4.2

Assuming β=200, determine the Q point (I_C and V_{CE}) for the circuit of Figure 4.4.3. Also determine the values of V_C and V_B.

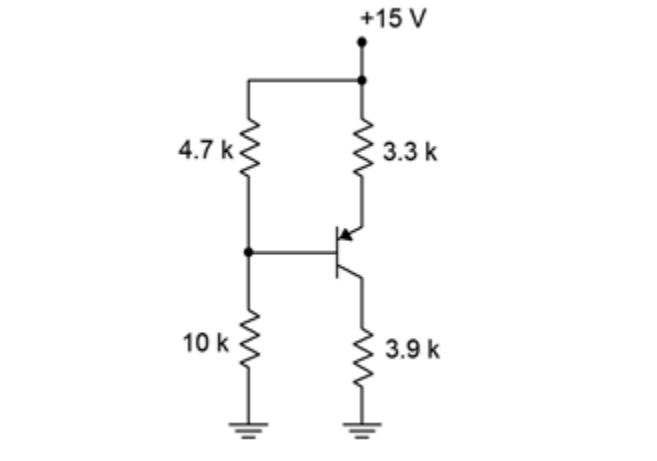


Figure 4.4.6 : Circuit for Example 4.4.2 .

First off, R_2 (now on top) is around the same size as R_E so the approximation method should be accurate and we can assume the divider is lightly loaded.

Method One

We will focus on the base-emitter loop as usual because V_{BE} is a known potential. Our immediate goal is to find the voltage across R_E so that we can use Ohm's law to find I_C. First we note that the voltage drop across R_2 is equal to the combined drops across R_E and V_{BE}. The drop across R_2 is found via the voltage divider rule.

$$V_{R_2} = V_{EE}\frac{R_2}{R_1 + R_2}$$

$$V_{R_2} = 15V\frac{4.7k\Omega}{10k\Omega + 4.7k\Omega}$$

$$V_{R_2} = 4.8V$$

And

$$V_{R_E} = V_{R_2} - V_{BE}$$

$$V_{R_E} = 4.8V - 0.7V$$

$$V_{R_E} = 4.1V$$

Therefore

$$I_C = \frac{V_{R_E}}{R_E}$$

$$I_C = \frac{4.1V}{3.3k\Omega}$$

$$I_C = 1.24mA$$

Method Two

Here we will determine all voltages with respect to ground.

$$V_B = V_{EE}\frac{R_1}{R_1 + R_2}$$

$$V_B = 15V\frac{10k\Omega}{10k\Omega + 4.7k\Omega}$$

$$V_B = 10.2V$$

The voltage from base to emitter has a – to + polarity, meaning it is a rise of 0.7 volts. Therefore

$$V_E = V_B + V_{BE}$$

$$V_E = 10.2V + 0.7V$$

$$V_E = 10.9V$$

The voltage across R_E is the difference between V_{EE} and V_E.

$$V_{R_E} = V_{EE} - V_E$$

$$V_{R_E} = 15V - 10.9V$$

$$V_{R_E} = 4.1V$$

This is the same value we arrived at using method one, so the collector current must be the same at 1.24 mA.

To find V_{CE} we also have options. One path is to use a slightly modified Equation 4.4.4

$$V_{CE} = -(V_{EE} - I_C(R_C + R_E))$$

$$V_{CE} = -15V + 1.24mA(3.9k\Omega + 3.3k\Omega)$$

$$V_{CE} = -6.07V$$

The collector is negative relative to the emitter, hence the negative sign. To avoid this, we could just swap the leads and refer to V_{EC} instead.

Alternately, we could find V_{CE} by determining V_C and then subtracting V_E from it.

$$V_C = I_C R_C$$

$$V_C = 1.24mA \times 3.9k\Omega$$

$$V_C = 4.84V$$

$$V_{CE} = V_C - V_E$$

$$V_{CE} = 4.84V - 10.9V$$

$$V_{CE} = -6.06V$$

We see a very slight difference here due to carried rounding errors.

It is instructive to compare the results of Example 4.4.2 back to Example 4.4.1. These circuits are otherwise identical except for the fact that one is NPN and the other is PNP. We find the same results for device currents (I_C) and component voltage magnitudes (V_{CE} or the voltage across R_E); only the signs and directions are reversed. On the other hand, we find that ground referenced potentials such as V_B, V_C and V_E are decidedly different between the two circuits.

4.5 FEEDBACK BIASING

While two-supply emitter bias and voltage divider bias can produce very high stability, there are other bias configurations available. Their stability tends not to be quite as good, but they are superior to simple base bias. They also tend to use fewer components than their high stability cousins. As a group, we refer to these as feedback biasing configurations. They use the concept of negative feedback. This is a technique where a change in the output can be reflected back to the input in such a way that it tends to partially offset the output change.

COLLECTOR FEEDBACK BIAS

With a simple move of R_B in the basic base bias configuration, we arrive at collector feedback bias. The NPN template is shown in Figure 4.5.1. Compared to base bias, all that has changed is that R_B is connected to the lower part of R_C rather than to the power supply. That small change can have a noticeable effect on stability.

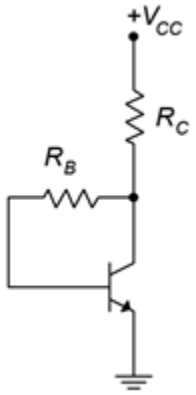


Figure 4.5.1: Collector feedback bias.

To understand how feedback works, assume that a current is flowing from the supply, through R_C , into the collector and finally, out of the emitter to ground. Via KVL, $V_{CE} = V_C = V_{CC} - I_C \cdot R_C$. Now suppose for some reason, a temperature change perhaps, β increases. This should cause an increase in I_C . An increase in I_C , though, would cause an increase in the drop across R_C due to Ohm's law. This, in turn, would force V_C to drop. Here is the key: V_C is also equal to the drop across R_B plus the voltage V_{BE} . The base-emitter potential is fixed at approximately 0.7 volts so any decrease in V_C is reflected as a decrease in voltage across R_B. By Ohm's law, that means that I_B must decrease by a similar proportion. This decrease tends to offset the initial tendency of the collector current to increase.

To derive an equation for the collector current, we can use KVL.

$$V_{CC} = V_{R_C} + V_{R_B} + V_{BE}$$

$$V_{CC} = I_E R_C + I_B R_B + V_{BE}$$

$$V_{CC} = I_C R_C + \frac{I_C}{\beta} R_B + V_{BE}$$

$$I_C = \frac{V_{CC} - V_{BE}}{R_C + R_B/\beta}$$

(4.5.1)

This equation is very similar to the current derivations for the two-supply emitter bias (Eq 4.3.1) and voltage divider bias (Eq 4.4.3). Again, if we can set $R_C \gg R_B/\beta$ then I_C will be relatively immune from Q point shifts due to β. The problem here is that it's not nearly so easy to meet that stipulation in this circuit. Consequently, collector feedback tends to have only modest stability.

Concerning the cutoff and saturation endpoints on the DC load line, once again, cutoff is determined by the DC power supply while saturation is determined by the amount of resistance in the collector-emitter to limit said power supply's current.

$$I_{C(sat)} = \frac{V_{CC}}{R_C}$$

(4.5.2)

$$V_{CE(cutoff)} = V_{CC}$$

(4.5.3)

Example 4.5.1

Assuming β=100β=100, determine the Q point (I_C and V_{CE}) for the circuit of Figure 4.5.2. How much does the Q point change if β is halved?

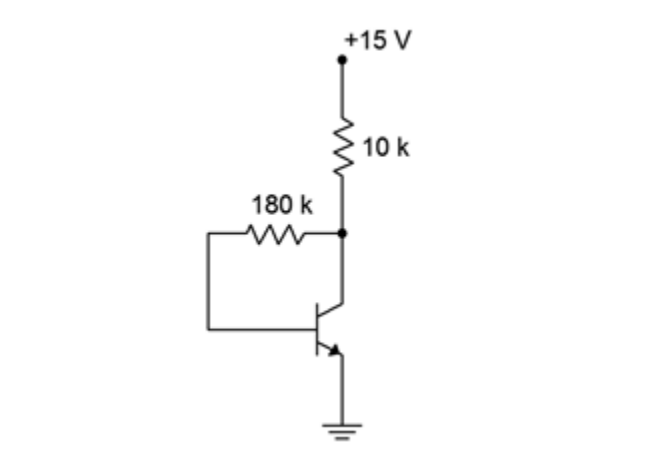


Figure 4.5.2 : Circuit for Example 4.5.1 .

Using Equation 4.5.1

$$I_C = \frac{V_{CC} - V_{BE}}{R_C + R_B/\beta}$$

$$I_C = \frac{15V - 0.7V}{10k\Omega + 180k\Omega/100}$$

$$I_C = 1.21mA$$

Using KVL

$$V_{CE} = V_{CC} - V_{RC}$$

$$V_{CE} = V_{CC} - I_C R_C$$

$$V_{CE} = 15V - 1.21mA \times 10k\Omega$$

$$V_{CE} = 2.9V$$

If β is halved to 50

$$I_C = \frac{V_{CC} - V_{BE}}{R_C + R_B/\beta}$$

$$I_C = 1.05mA$$

$$V_{CE} = V_{CC} - I_C R_C$$

$$V_{CE} = 15V - 1.05mA \times 10k\Omega$$

$$V_{CE} = 4.5V$$

For a 2:1 drop in β we see about a 13% reduction in I_C with a somewhat larger change in V_{CE}. This circuit is clearly not as stable as the two-supply emitter bias or the voltage divider bias but it is superior to base bias.

The PNP version of the collector feedback bias configuration should come as no surprise. The template is shown in Figure 4.5.3. Here, we use the same technique of power supply shifting that was used with the PNP voltage divider in order to wind up with a positive power supply. As with the PNP voltage divider, because we have changed the reference point, all ground referenced voltages will be different from their NPN counterparts. All currents and component voltages will have the same magnitudes but with opposite directions and polarities.

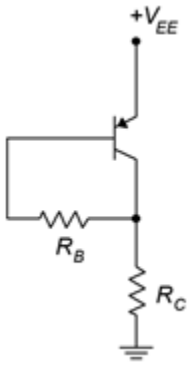


Figure 4.5.3: PNP Collector feedback bias.

EMITTER FEEDBACK BIAS

The emitter feedback bias uses the same overall idea as the collector feedback circuit, namely, that changes at the output will be reflected back to the input and thus help mitigate the initial change. While collector feedback focuses on collector current establishing V_C via R_C , emitter feedback uses the fact that emitter current establishes V_E via R_E. In both cases, these voltages are used to change the voltage across R_B, which results in a change in I_B that opposes the original collector current change.

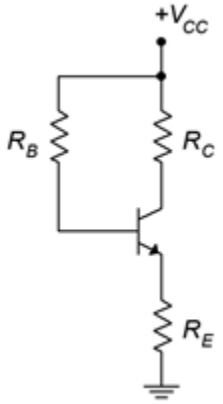


Figure 4.5.4: Emitter feedback bias.

A basic emitter feedback bias circuit is shown in Figure 4.5.4. We shall use KVL to develop an equation for collector current.

$$V_{CC} = V_{R_B} + V_{BE} + V_{R_E}$$

$$V_{CC} = I_B R_B + V_{BE} + I_E R_E$$

$$V_{CC} = \frac{I_C}{\beta} R_B + I_C R_E + V_{BE}$$

$$I_C = \frac{V_{CC} - V_{BE}}{R_E + R_B/\beta}$$

(4.5.4)

If we can set $R_E \gg R_B/\beta$ then the Q point will be stable in spite of changes in β. The problem here is the same as was the case in collector feedback, namely that this stipulation is not easy to achieve. Consequently, the emitter feedback configuration tends to have only modest stability. In any event, once the collector current is known, V_{CE} can be found using the techniques illustrated with the voltage divider configuration. The endpoints for the DC load line are found in the usual manner.

$$I_{C(sat)} = \frac{V_{CC}}{R_C + R_E}$$

(4.5.5)

$$V_{CE(cutoff)} = V_{CC}$$

(4.5.6)

Example 4.5.2

Assuming β = 100, determine the Q point (IC and VCE) for the circuit of Figure 4.5.5.

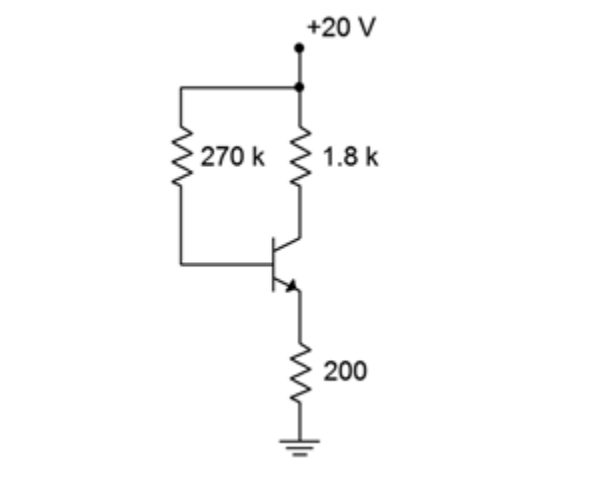


Figure 4.5.5 : Circuit for Example 4.5.2 .

Using Equation 4.5.4

$$I_C = \frac{V_{CC} - V_{BE}}{R_E + R_B/\beta}$$

$$I_C = \frac{20V - 0.7V}{200\Omega + 270k\Omega/100}$$

$$I_C = 6.66mA$$

Using KVL

$$V_{CE} = V_{CC} - V_{R_C} - V_{R_E}$$

$$V_{CE} = V_{CC} - I_C(R_C + R_E)$$

$$V_{CE} = 20V - 6.66mA(1.8k\Omega + 200\Omega)$$

$$V_{CE} = 6.68V$$

To complete the load line, we find

$$I_{C(sat)} = 10mA$$

$$V_{CE(cutoff)} = 20V$$

Dropping β to 50 will result in

$$I_C = 3.45mA$$

$$V_{CE} = 13.1V$$

We see less than a 2:1 change in I_C and V_{CE} but the stability is not dramatic.

COMBINATION FEEDBACK BIAS

The final feedback bias configuration combines both collector feedback and emitter feedback to arrive at the circuit depicted in Figure 4.5.6. We shall bestow upon it the highly original name of combination feedback bias.

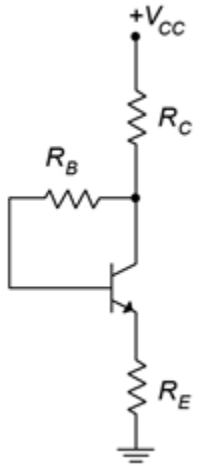


Figure 4.5.6 : Combination feedback bias.

This circuit applies feedback to R_B from both ends, so to speak, so it tends to have slightly better stability than either collector feedback or emitter feedback bias. Of course, it is now only one resistor shy from the voltage divider circuit which is considerably more stable.

The equations for the load line are listed below. The derivations are left as an exercise.

$$I_C = \frac{V_{CC} - V_{BE}}{R_C + R_E + R_B/\beta}$$

(4.5.7)

$$V_{CE} = V_{CC} - I_C(R_C + R_E)$$

(4.5.8)

$$I_{C(sat)} = \frac{V_{CC}}{R_C + R_E}$$

(4.5.9)

$$V_{CE(cutoff)} = V_{CC}$$

(4.5.10)

Example 4.5.3

Assuming β = 125, determine the Q point (IC and VCE) for the circuit of Figure 4.5.7.

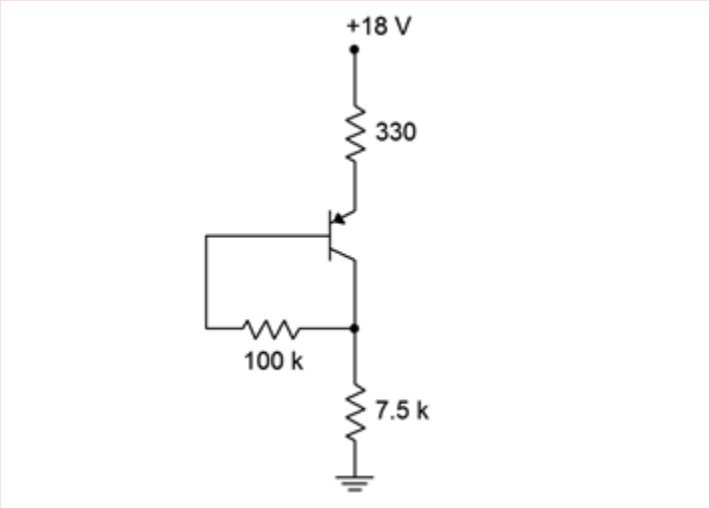


Figure 4.5.7 : Circuit for Example 4.5.3 .

Note that this is the upside down PNP version. Using Equation 4.5.7

$$I_C = \frac{V_{CC} - V_{BE}}{R_C + R_E + R_B/\beta}$$

$$I_C = \frac{18V - 0.7V}{7.5k\Omega + 330\Omega + 100k\Omega/125}$$

$$I_C = 2mA$$

Using KVL

$$V_{CE} = V_{CC} - V_{R_C} - V_{R_E}$$

$$V_{CE} = V_{CC} - I_C(R_C + R_E)$$

$$V_{CE} = 18V - 2mA(7.5k\Omega + 330\Omega)$$

$$V_{CE} = 2.34V$$

4.5 SUMMARY

DC biasing is required in order to maintain the proper junction potentials and operation of the BJT. Several different circuit configurations are available to establish a DC bias on both NPN and PNP transistors. These circuits vary in complexity and their ability to maintain a constant operating point, or Q point, in the face of variations of β.

The two-supply emitter bias topology offers very high Q point stability. It achieves this through the use of two powers supplies; one connected through a resistor to the emitter and a second unit connected through a resistor to the collector. It is unique in that the supplies are bipolar; one being positive and the other being negative.

The voltage divider bias circuit offers similar stability performance to the two-supply emitter bias circuit. It uses a single supply and a resistive voltage divider to establish a second, lower potential at the base terminal.

The three feedback bias configurations offer only modest enhancements in stability but use the least amount of parts. They all rely on a single DC power source.

A DC load line is a plot of all possible collector current and corresponding collector-emitter voltage operating points. No matter what the β for a circuit happens to be, the transistor's operating point must lie on this line. It is a valuable DC analysis tool.

Review Questions

1. Explain the need for DC biasing. Why can't we just apply an AC signal to the base of a BJT and expect proper amplification of the signal?
2. What is a Q point?
3. What are the four values found on a DC load line?
4. Rank the bias configurations presented in this chapter in terms of their Q point stability relative to β
5. Rank the bias configurations presented in this chapter in terms of their circuit complexity.
6. Describe the process of making a PNP version of an NPN bias circuit.

4.6 EXERCISES

Assume diodes are silicon unless stated otherwise

ANALYSIS PROBLEMS

1. Plot the load line for the circuit of Figure 4.7.1.
   V_{CC} = 20 V, V_{EE}= –8 V, R_B = 7.5 kΩ, R_E = 10 kΩ, R_C= 12 kΩ.

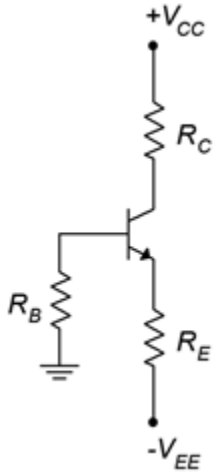


Figure 4.7.1

2. Determine the new Q point for Problem 1 if β = 250.
3. Plot the load line for the circuit of Figure 4.7.2.
   V_{EE} = 5 V, V_{CC}= –18 V, R_B = 22 kΩ, R_E = 1.2 kΩ, R_C = 1.5 kΩ.

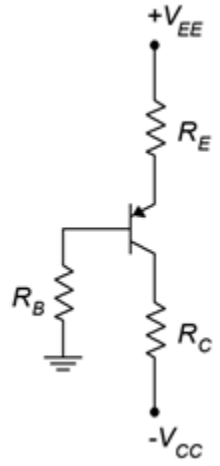


Figure 4.7.2

4. Determine the new Q point for Problem 3 if β = 50.

5. Plot the load line for the circuit of Figure 4.7.3.
   V_{CC} = 20 V, R_1 = 15 kΩ, R_2 = 5 kΩ, R_E = 4.3 kΩ, R_C = 9.1 kΩ.
6. Determine the new Q point for Problem 5 if β = 150.
7. Plot the load line for the circuit of Figure 4.7.4.
   V_{EE} = 16 V, R_1 = 12 kΩ, R_2 = 4.7 kΩ, R_E = 6.2 kΩ, R_C = 10 kΩ.

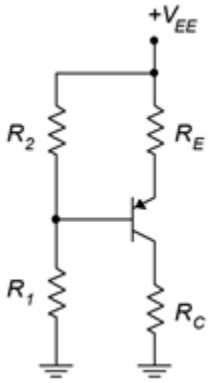


Figure 4.7.4.

8. Determine the new Q point for Problem 7 if β = 200.
9. Plot the load line for the circuit of Figure 4.7.5.
   V_{CC} = 12 V, R_B = 560 kΩ, R_C = 3.3 kΩ.

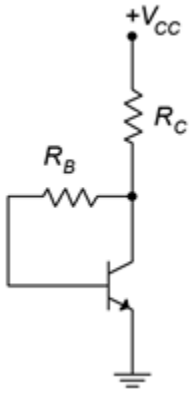


Figure 4.7.5.

10. Determine the new Q point for Problem 9 if β = 75.
11. Plot the load line for the circuit of Figure 4.7.6.

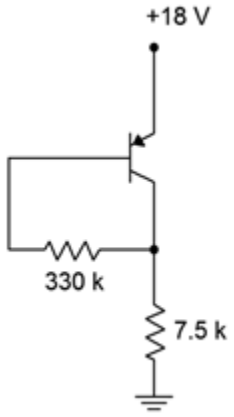


Figure 4.7.6

12. Determine the new Q point for Problem 11 if β = 200.
13. Plot the load line for the circuit of Figure 4.7.7.
    V_{CC} = 15 V, R_B = 470 kΩ, R_E = 560 Ω, R_C = 3.3 kΩ.

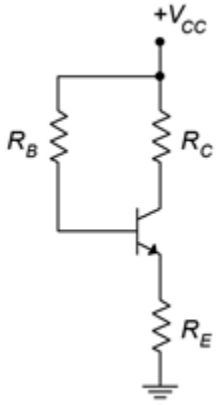


Figure 4.7.7.

14. Determine the new Q point for Problem 13 if β = 170.
15. Plot the load line for the circuit of Figure 4.7.8.
16. Determine the new Q point for Problem 15 if β = 75.
17. Plot the load line for the circuit of Figure 4.7.9.
    V_{EE} = 18 V, R_B = 680 kΩ, R_E = 270 Ω, R_C = 3.9 kΩ.
18. Determine the new Q point for Problem 17 if β = 200.

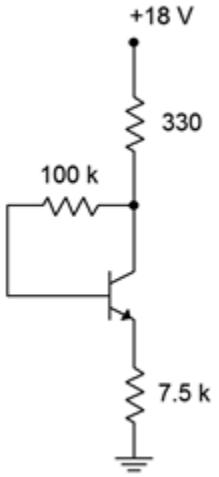


Figure 4.7.8

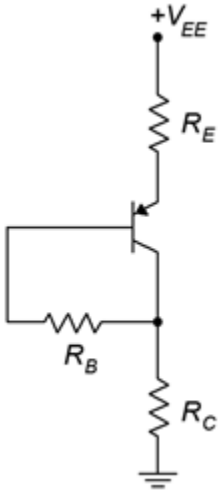


Figure 4.7.9

DESIGN PROBLEMS

19. Determine a value for R_E in the circuit of Figure 4.7.1 to set I_C = 2 mA.
Use V_{CC} = 20 V, V_{EE} = –8 V, R_B = 10 kΩ, R_C = 4.6 kΩ.
20. Determine a value for R_C in the circuit of Figure 4.7.2 to set V_{CE} = 10 V.
Use V_{CC} = –25 V, V_{EE} = 6 V, R_B = 15 kΩ, R_E = 6.8 kΩ.
21. 21. Determine a value for R_C in the circuit of Figure 4.7.3 to set V_{CE} = 8 V.
Use V_{CC} = 24 V, R_1 = 22 kΩ, R_2 = 10 kΩ, R_E = 4.6 kΩ.
22. Determine new values for R_1 and R_2 in the circuit of Figure 4.7.4 in order to set I_C = 500 μA.
V_{EE} = 22 V, R_E = 15 kΩ, R_C = 6.8 kΩ.

CHALLENGE PROBLEMS

23. Determine the maximum and minimum values for I_C in Problem 1 if every resistor has a 10% tolerance.

24. Determine the maximum and minimum values for V_{CE} in Problem 3 if every resistor has a 5% tolerance.
25. Determine a value for R_E in the circuit of Figure 4.7.3 to set V_{CE} = 10 V. V_{CC} = 30 V, R_1 = 12 kΩ, R_2 = 3 kΩ, R_C = 8.2 kΩ.
26. Derive Equation 4.16.
27. Determine the the power drawn from the supply for the circuit of Problem 5. 28. Using a 15 volt power supply, design a bias circuit to create a very stable Q point of 2 mA and 5 volts.

COMPUTER SIMULATION PROBLEMS

1. Perform a series of DC simulations to test the Q point stability versus β of the circuit of Problem 1.
2. Perform a Monte Carlo simulation to investigate the Q point stability of the circuit of Problem 5 if the emitter resistor has a 10% tolerance.

UNIT 5: AMPLIFIER CONCEPTS

Learning Objectives

After completing this chapter, you should be able to:

- Explain the differences between voltage gain, current gain and power gain.
- Describe a basic voltage amplifier model using voltage gain, input impedance and output impedance.
- Determine the effects of source and load impedance on system gain and explain how they interact with an amplifier's input and output impedance.
- Describe and distinguish the concepts of noise and waveform distortion.
- Define the concept of output compliance.
- Discuss the frequency limits of an amplifier in general terms. Define Miller's Theorem.

5.1 INTRODUCTION

The concept of signal amplification finds numerous uses in the field of electronics. This includes applications such as boosting the signal level from a sensor or driving loads like loudspeakers or antennas. Reduced to its most simple terms, amplification is just multiplication. The ideal amplifier multiplies the amplitude of the input signal by a constant. It should not change the frequency of the signal, alter its shape, add noise or in any other way warp or distort the signal.

Amplifiers can be designed to be voltage sensing or current sensing and can be modeled as either controlled voltage sources or controlled current sources. As a functional block, we are primarily interested in describing an amplifier in terms of its amplification factor, input impedance and output impedance. The amplification factor is also referred to as the gain and may be expressed in terms of voltage gain, current gain or power gain, depending on the application. Other items of interest include the maximum output level or compliance, useful frequency range, noise and distortion characteristics.

5.3 COMPLIANCE AND DISTORTION

At some point the idealization that the output signal is merely the input signal times the gain fails. All amplifiers have a limit on just how large the output signal can be. This is set by the DC power supply and the amplifier design. The maximum output signal (typically, the maximum output voltage) is referred to as the compliance. Any attempt to produce an output signal that swings beyond the compliance will result in waveform distortion. In the simplest case, the output signal is strictly and abruptly limited to the compliance level and any portion of the output waveform that would otherwise lay above that will be removed. It is as if some form of electronic scissors clipped off the top of the waveform. Hence, this is often referred to as clipping. An example of clipping is illustrated in Figure 5.3.1. The ideal output waveform is shown in brown and the clipped waveform is shown in blue. The clipping is so severe here that the clipped waveform now looks less like a sine wave and more like a square wave. This is extreme waveform distortion and has important consequences.

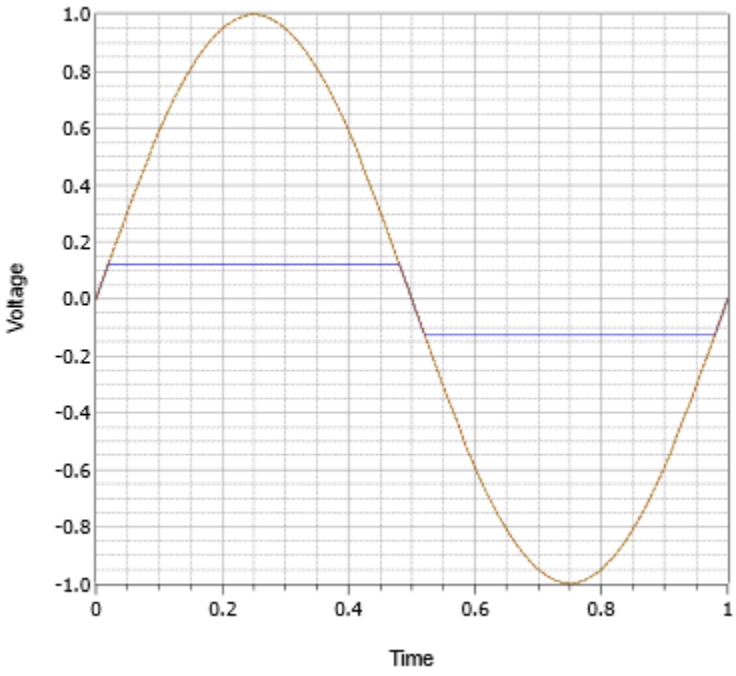


Figure 5.3.1: Clipped waveform.

Whenever a signal is altered in the time domain, there will be an alteration of its frequency content. Depending on what we started with and the manner in which the wave shape is altered, new frequency components may be added to the signal and the levels of existing components may be changed or even deleted. The extreme clipping that created a near-square wave added a large number of new frequency components to the signal.

To see how the wave shape and frequency content are connected, consider the waveform depicted in Figure 5.3.2. We start with a simple sine wave shown in green. We refer to this as the fundamental or

base frequency. We then add a harmonic. A harmonic is another sine wave that is an integer multiple of the fundamental frequency. It may be larger or smaller in amplitude and the phase may be shifted. In this example we have a single sine at three times the fundamental frequency (blue). When we add this harmonic to the fundamental we arrive at a new waveform shown in red. This new waveform looks something like a square wave but with a "lumpy" top and bottom. If we add more harmonics, these variations will begin to smooth out, as shown in Figure 5.3.3. This waveform appears to be fairly close to a square wave and not too distant from our earlier clipped waveform. Based on this, we can conclude that the clipped sine wave has new frequency components added to it. We can also conclude that if our amplifier clips a more complicated waveform such as a snippet of music, that process will add new harmonics as well. Further, it is likely that these harmonics will be audible and can change our perception of the music, perhaps subtly but maybe drastically.

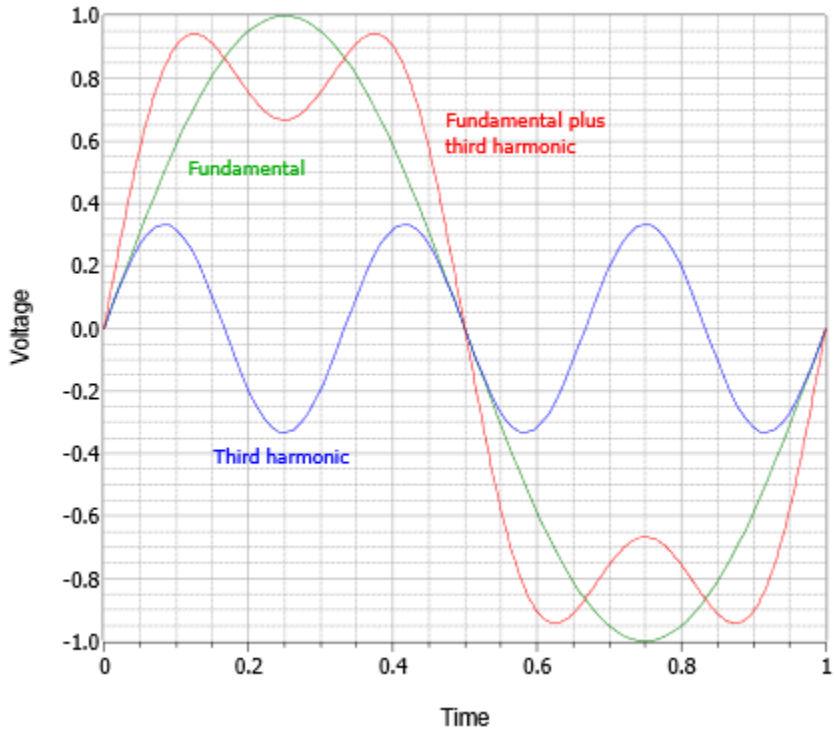


Figure 5.3.2: Sine wave with third harmonic.

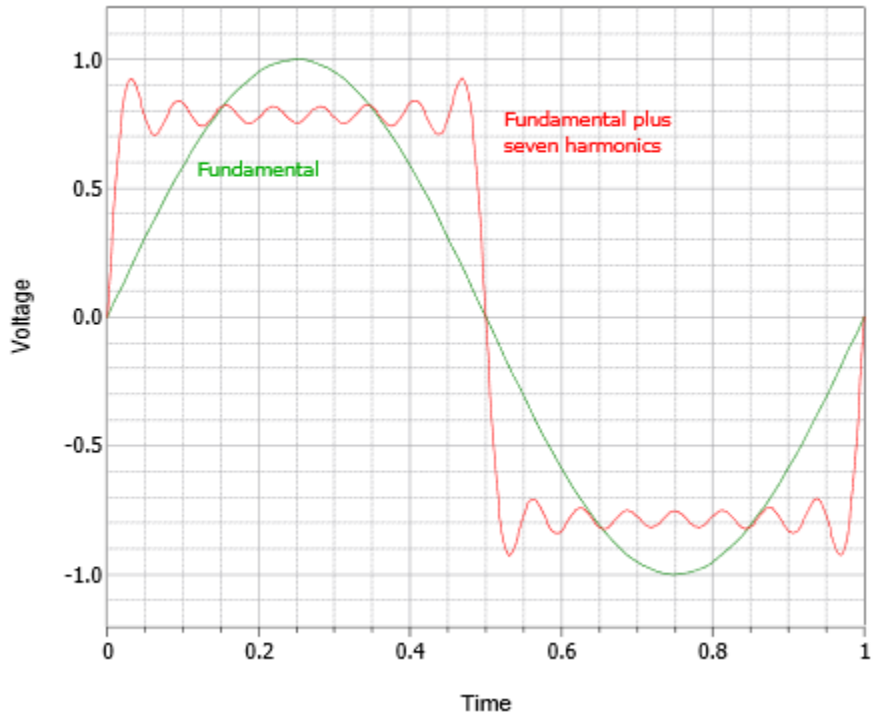


Figure 5.3.3 : Sine wave with seven harmonics, approaching a square wave.

Along with clipping, amplifiers can exhibit more subtle forms of distortion due to internal nonlinearity. For example, it is possible for the gain to vary slightly as the signal swings from low to high or from negative to positive. An example is shown in Figure 5.3.4 with the distorted wave shown in red.

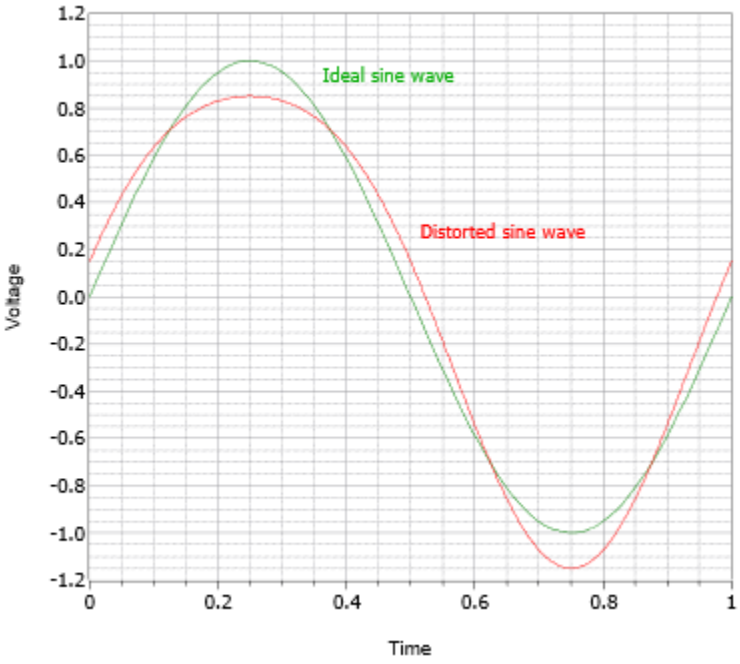


Figure 5.3.4: Sine wave with distortion.

At first glance it may appear as though the wave is merely offset negatively. This is not the case. If we shift the wave vertically, as in Figure 5.3.5, it becomes apparent that the wave is truly distorted and is no longer a pure sine wave.

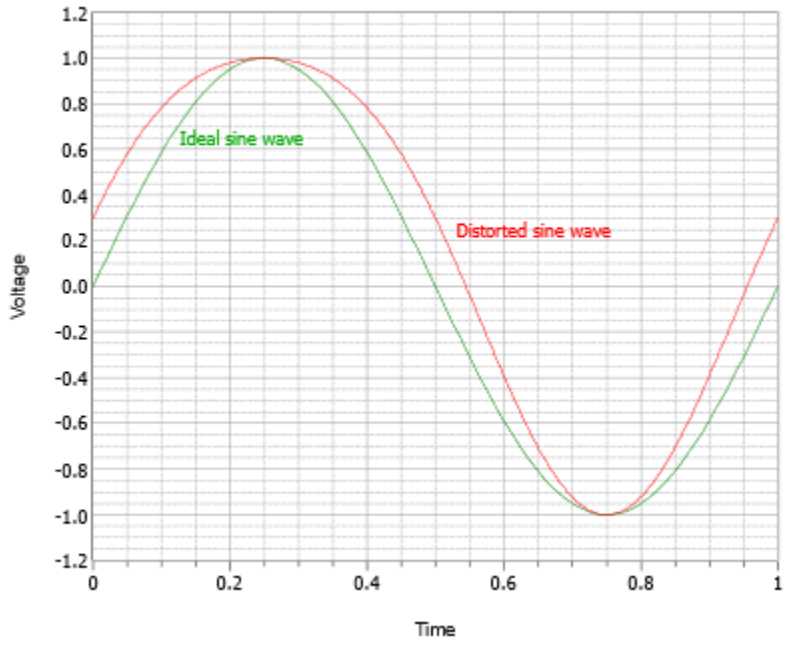


Figure 5.3.5: Sine wave with distortion, level shifted.

Unlike the clipped wave, the distorted wave in Figure 5.3.5 exhibits an asymmetry; the negative portion does not appear to be a mirror image of the positive portion. In other words, this wave lacks half-wave symmetry. Waves that exhibit half-wave symmetry contain only odd harmonic distortion (harmonics that are odd integer multiples of the fundamental). In contrast, waves that lack half-symmetry have at least one even harmonic. Here is how to test for half-wave symmetry. First, consider the sawtooth wave shown in Figure 5.3.6.

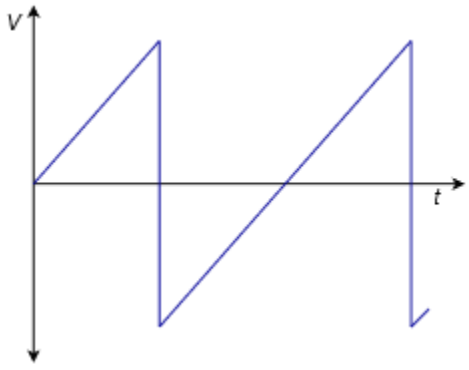


Figure 5.3.7: Half-wave symmetry test: Sawtooth wave, negative portion rotated. Finally, slide the negative portion over the positive portion and see if they're identical, as in Figure 5.3.8.

Rotate the negative portion of the wave around the time axis as shown in Figure 5.3.7.

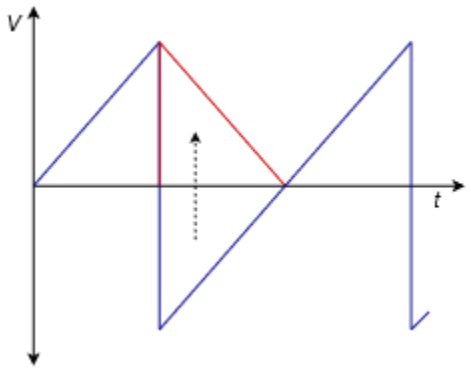


Figure 5.3.7: Half-wave symmetry test: Sawtooth wave, negative portion rotated.

Finally, slide the negative portion over the positive portion and see if they're identical, as in Figure 5.3.8.

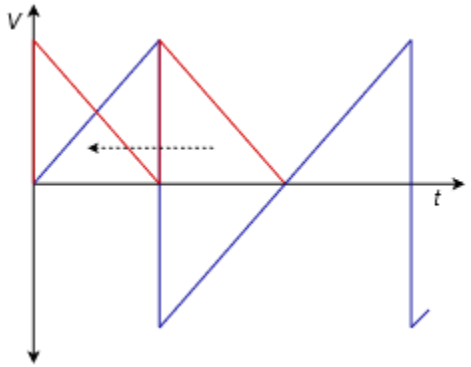


Figure 5.3.8 : Half-wave symmetry test: Sawtooth wave, negative portion rotated and slid.

If the two halves are identical then the wave has half-wave symmetry. The sawtooth wave does not exhibit half-wave symmetry, therefore it must contain at least one even harmonic.

An amplifier's linearity is often quantified through a Total Harmonic Distortion, or THD, measurement. The measurement is carried out by applying a very pure, low distortion sine wave to the amplifier. This is the fundamental. At the output of the amplifier, a very selective filter is used to remove the fundamental. This leaves behind just the added distortion harmonics.[1] These harmonics are then treated as a lumped value and presented as a percentage of the total signal. On an oscilloscope, it is relatively easy for a person to discern THD levels in the double digits. On the other hand, it is very difficult, if not impossible, for an individual to discern THD levels much below 1% by eye. Of course, what matters is what we can hear, not how the waveform looks. To put this in perspective, many high fidelity audio amplifiers exhibit THD levels below 0.1% while an over-driven guitar amplifier might be running over 20%. THD is not the final word on distortion though. It has its limits. For example, all of the distortion products are lumped together. It says nothing about which

1. To be strictly accurate, the residual consists of the harmonics plus any noise produced by the amplifier. Therefore it is more accurate to refer to this as a THD+noise spec.

harmonics are particularly strong or their distribution. It also doesn't say much about what happens when multiple frequencies interact. One method of trying to quantify that is to apply two sine waves at different frequencies to the amplifier simultaneously. The result is called an Intermodulation Distortion rating, or IMD. This is also expressed as a percentage.

5.2 AMPLIFIER MODEL

The ideal amplifier does nothing except increase the amplitude of the input signal. The factor of increase is defined as the ratio of the output signal to the input signal. It is a unit-less quantity. For example, if the input signal has a power of 10 milliwatts and the circuit boosts the signal up to 50 milliwatts, we say it has a power gain of 50 milliwatts /10 milliwatts, or 5. Similarly, if the input signal is 2 volts and the output signal is 16 volts, we say it has a voltage gain of 16 volts/2 volts, or 8. Historically, power gain is denoted as *G*. For voltage gain and current gain we use A_v and A_i , where *A* stands for *A*mplification factor. Some amplifiers invert the signal from input to output. Basically, they flip the wave shape upside down. For a simple sine wave this is equivalent to shifting the phase of the signal by 180° , and for a sine wave input the amplifier produces a –sine output. To reflect this effect, the amplification factor is denoted as negative. For example, an A_v of –10 indicates an amplification factor of 10 with a signal inversion.

The size and complexity of an amplifier circuit can vary considerably, ranging from a single transistor to dozens of transistors. To ease system design it is helpful to use simplified functional models. Typically, these models use a resistor to represent the impedance seen looking into the amplifier along with a controlled source and its associated internal resistance. An example is shown in Figure .2.1.

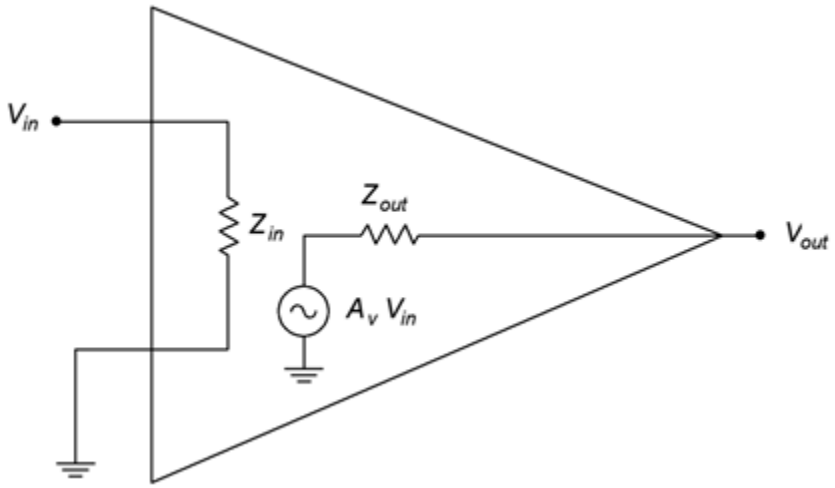


Figure 5.2.1: A simple voltage amplifier model.

This is a model of a voltage amplifier (note that V_{in} and V_{out} are specified along with a controlled voltage source within the model). The controlled voltage source and its series Z_{out} is the Thevenin equivalent of the output when viewed from the load (i.e., the V_{out} pin). Likewise, Z_{in} is the equivalent impedance seen by the driving source. Because a voltage amplifier is designed to maximize voltage transfer, the input impedance tends to be high to minimize loading (think of a voltmeter). Similarly, the output impedance would tend to be low (think of an ideal voltage source). In contrast, a circuit

designed for maximum current transfer would tend to have a low Z_{in} and a high Z_{out}. Precisely how the circuit creates the signal boost is not a concern of this model, we only care that it does.

LOADING EFFECTS

Once the model is established it is relatively easy to recognize and compute loading effects. Loading effects are signal losses caused by interactions between the amplifier's impedances and those of the circuits and loads connected to it. A generic model including loading effects is shown in Figure 5.2.2.

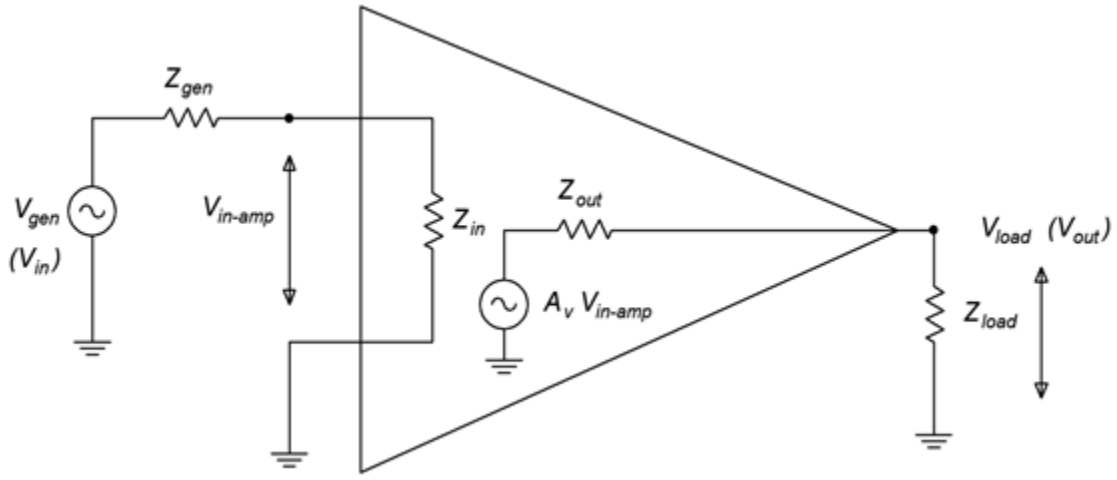


Figure 5.2.2: Voltage amplifier model with loading effects.

A cursory examination of Figure 5.2.2 shows that there is a voltage divider between Z_{gen} and Z_{in} along with a second divider between Z_{out} and Z_{load}. Each of these dividers causes signal loss, that is, they reduce the final output voltage. The input voltage to the amplifier is reduced as follows:

$$V_{in-amp} = \frac{Z_{in}}{Z_{in} + Z_{gen}} \times V_{gen}$$

The load voltage is reduced as follows:

$$V_{in-amp} = \frac{Z_{load}}{Z_{load} + Z_{out}} x A_v \times V_{in} - amp$$

Therefore the combined gain is:

$$A_{system} = \frac{V_{load}}{V_{gen}} = \frac{Z_{in}}{Z_{in} + Z_{gen}} \times A_v \times \frac{Z_{load}}{Z_{load} + Z_{out}}$$

To minimize these losses we'd like $Z_{in} \gg Z_{gen}$ and $Z_{load} \gg Z_{out}$.

A voltage amplifier has the following specifications: A_v = 20 , Z_{in} = 10 kΩ, Z_{out} = 200Ω. It is driven by a 30 millivolt source with a 600 Ω internal impedance and drives a 1 kΩ load. Determine the load voltage.

The voltage that appears at the amplifier's input is

$$V_{in-amp} = \frac{Z_{in}}{Z_{in} + Z_{gen}} \times V_{gen}$$

$$V_{in-amp} = \frac{10k\Omega}{10k\Omega + 600\Omega} \times 30mV$$

$$V_{in-amp} = 28.3mV$$

This is multiplied by the voltage gain of 20 and then reduced by the output divider.

$$V_{in-amp} = \frac{Z_{load}}{Z_{load} = Z_{out}} \times A_v \times V_{in-amp}$$

$$V_{in-amp} = \frac{1k\Omega}{1k\Omega + 200\Omega} \times 20 \times 28.3mV$$

$$V_{in-amp} = 471.7mV$$

Without the loading effects the output signal would be simply 30 millivolts times the voltage gain of 20, or 600 millivolts. Further, note that if the source is replaced with a typical laboratory grade function generator exhibiting an internal impedance of 50 Ω and the load is removed, being replaced by an oscilloscope exhibiting a typical 1 MΩ input impedance, the loading effects would be minimal and we would measure just a few millivolts shy of the ideal 600 millivolts.

5.7 EXERCISES

ANALYSIS PROBLEMS

1. Determine the load voltage for the model of Figure 5.7.1 if V_{gen} = 10 mV, Z_{gen} = 50 Ω, Z_{in} = 1 MΩ, Z_{out} = 75 Ω, Z_{load}= 1 k Ω and A_v= 50.

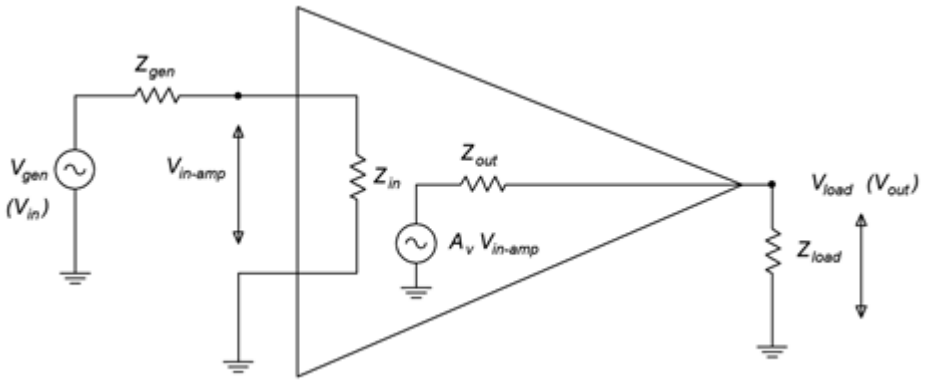


Figure 5.7.1

2. Determine the load voltage for the model of Figure 5.7.1 given V_{gen} = 8 mV, Z_{gen} = 1 k Ω, Z_{in} = 6 kΩ, Z_{out} = 500 Ω, Z_{load} = 2 k Ω and A_v = 100.
3. If the circuit of Problem 1 has a compliance of 2 volts, will the output clip? What if the input is increased to 100 mV?
4. If the circuit of Problem 2 has a compliance of 5 volts, will the output clip? What if the input is increased to 200 mV?
5. If an amplifier has A_v = 25, V_{in} = 20 mV and there is no appreciable loading, determine the output signal-to-noise ratio if the amplifier generates an output noise voltage of 10 μV.
6. Determine which waveforms from Figures 5.7.26 through 5.7.6 exhibit halfwave symmetry.

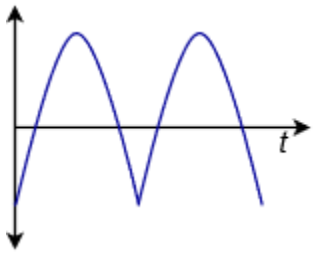


Figure 5.7.2

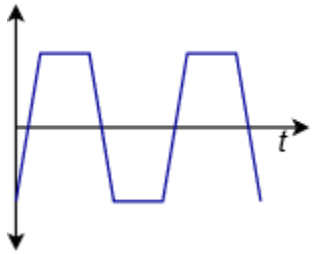


Figure 5.7.3

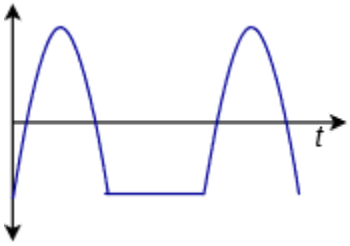


Figure 5.7.4

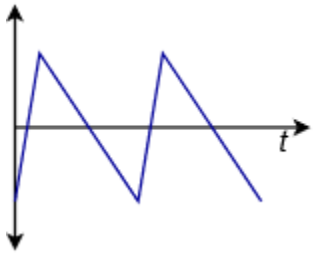


Figure 5.7.5

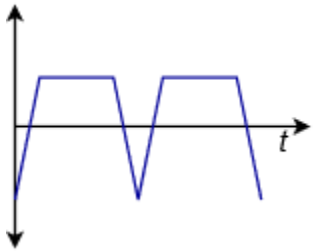


Figure 5.7.6

7. Determine the Miller equivalent resistances for the circuit of Figure 5.7.7 if $A_v = -20$ and $R = 60\ k\Omega$.

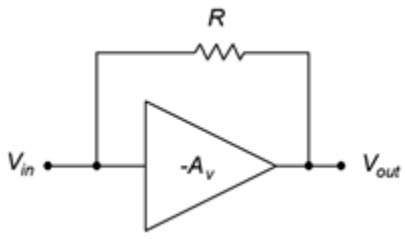


Figure 5.7.7

8. Determine the Miller equivalent capacitances for the circuit of Figure 5.7.8 assuming $A_v = -30$ and $C = 200$ pF.

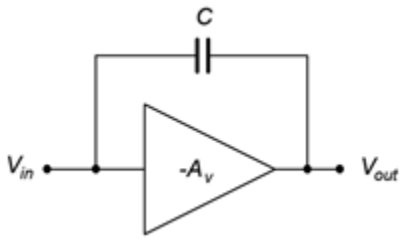


Figure 5.7.8

CHALLENGE PROBLEMS

9. If the circuit of Problem 1 has a compliance of 20 volts, how large can the input signal be before the load voltage is clipped?
10. If the circuit of Problem 2 has a compliance of 10 volts, how large can the input signal be before the load voltage is clipped?
11. Using Figure 5.7.7 as a guide and assuming that $R = 100$ kΩ, how large would the gain have to be such that the input equivalent resistance is 4 kΩ?
12. Using Figure 5.7.8 as a guide and assuming that $A_v = -35$, determine a value for C such that the input equivalent capacitance is 1.2 nF.

COMPUTER SIMULATION PROBLEMS

13. Simulate the circuit of Problem 1 and verify the load voltage.
14. Simulate the circuit of Problem 2 and verify the load voltage.

5.5 MILLER'S THEOREM

Some inverting voltage amplifier designs employ an impedance bridged between the input and output, as shown in Figure 5.5.1.

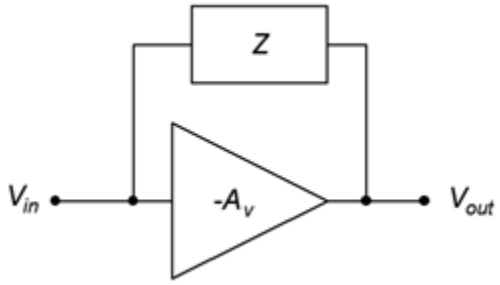


Figure 5.5.1: An input/output bridged impedance.

There are different reasons for doing this, a prime example involving shaping the frequency response of the amplifier. To simplify the analysis of such a configuration, we may employ Miller's Theorem, named after American engineer John Milton Miller.

The goal here is to determine equivalent impedances that lie in parallel with the input and output of the amplifier. These equivalents simply become part of the input and output networks around the amplifier.

First, let's consider the equivalent impedance at the input, $Z_{in–miller}$. By definition, this is the impedance in parallel with the input of the amplifier that would draw the same amount of current as the original bridging Miller impedance. The current through the Miller impedance is simply the the voltage across it divided by the Miller impedance, Z . The voltage across it is the difference between the input and output voltages.

$$i_{in-miller} = \frac{V_{in} - V_{out}}{Z}$$

$$i_{in-miller} = \frac{V_{in} - A_v V_{in}}{Z}, \text{ the gain is negative so}$$

$$i_{in-miller} = \frac{V_{in}(|A_v| + 1)}{Z}$$

Dividing this current into the input voltage yields the equivalent impedance.

$$Z_{in-miller} = \frac{Z}{|A_v| + 1}$$

A similar derivation yields the output equivalent.

$$Z_{out-miller} = \frac{Z|A_v|}{|A_v| + 1}$$

The Miller equivalent circuit for a general impedance Z is shown in Figure 5.5.2.

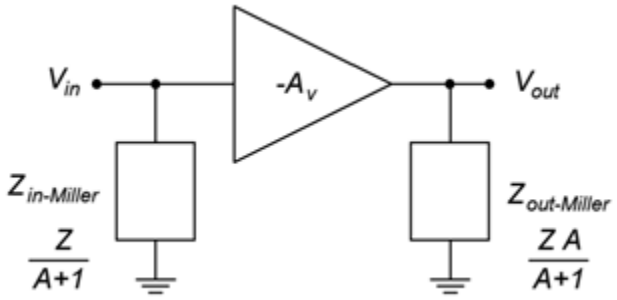


Figure 5.5.2: Miller general equivalent.

The general rule to remember is that the Miller equivalent presents equivalent impedances that are less than the original bridging impedance. In the case of the input section, the reduction effect is very large at higher gains.

Two typical cases for the impedance Z are a resistance, R, and a capacitance, C. For a pure resistance, we can perform a direct substitution for Z . The original and Miller equivalents are shown in Figure 5.5.3

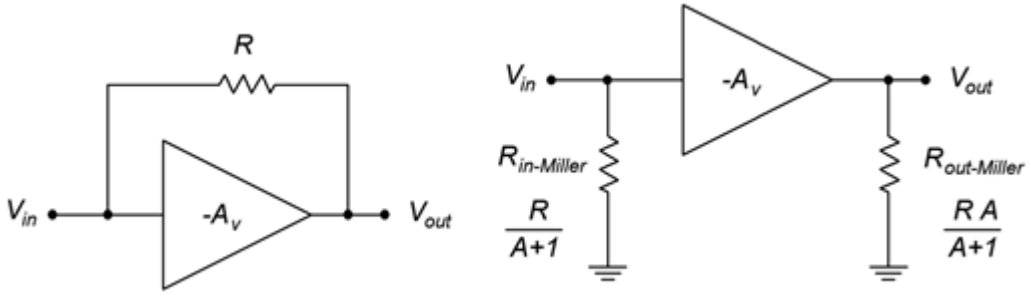


Figure 5.5.3 : Original and Miller equivalent for a resistor.

For a capacitor, the situation is similar, however, we will substitute X_c for Z and recall that $C = 1/(2\pi f X_c)$.

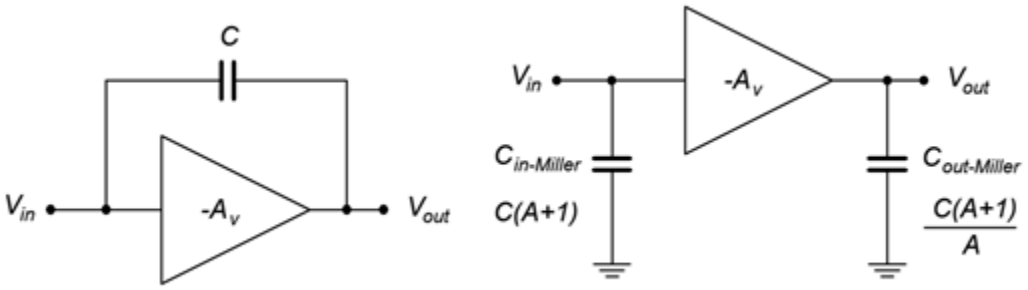


Figure 5.5.4: Original and Miller equivalent for a capacitor.

Note that for the capacitor, there is a multiplicative effect. That is, the effect of the original bridging capacitor on a high gain amplifier is equivalent to a much larger input shunt capacitor..

5.6 SUMMARY

Complex amplifier circuits can be modeled with a functional block. The ideal model includes the input and output impedances along with a controlled source. This source would exhibit a signal gain or amplification factor. Usually, *G* stands for power gain while *Av* and *Ai* represent voltage and current amplification, respectively. This amplification factor may be negative which indicates that the amplifier inverts the phase of the input, that is, the waveform is flipped upside down. The impedances allow calculation of loading effects while the gain determines the the size of the output signal.

If the input signal is too large, the output signal may be limited in amplitude or clipped. The maximum output amplitude is referred to as the compliance. Clipping is a gross form of distortion but more subtle forms exist as well. In general, distortion creates new frequency components. If these new components are integer multiples of the original input frequency, which they are typically, they are referred to as harmonics. One method of quantifying distortion performance is to sum all of the harmonics and compare that to the original signal. This is called THD or total harmonic distortion. Along with distortion, the amplifier might also add undesirable noise to the output signal. Noise is a random signal that contains many different frequencies. Typically, this is measured via a signal-to-noise ratio, or *S/N*, at the output.

An amplifier also operates over a given range of frequencies, from a lower limit, *f*1, to a high limit, *f*2. Some amplifiers are able to amplify down to 0 Hz (DC) and effectively do not have an *f*1 but all amplifiers do have an upper limit.

Finally, Miller's Theorem is an analysis technique that allows an impedance that bridges from the input of an inverting voltage amplifier to its output to be split into equivalent input and output parallel impedances. These impedances will be smaller than the original bridging impedance and are a function of the gain of the amplifier.

Review Questions

1. Explain how an amplifier's input impedance might react with a source to produce a signal loss.
2. Explain how an amplifier's output impedance might react with a load to produce a signal loss.
3. What is compliance?
4. Describe clipping.
5. Describe half-wave symmetry. What does it have to do with amplifier distortion?
6. What is noise? How does it differ from distortion?
7. Draw a generic frequency response plot for an amplifier.

8. Detail the purpose and use of Miller's Theorem.

5.4 FREQUENCY RESPONSE AND NOISE

Like compliance and distortion, two other practical limits on amplifier performance are its frequency response and output noise. First, let's discuss frequency response.

Although we describe an amplifier as having a specific gain or amplification factor, this is true only for a certain range of frequencies. All amplifiers are limited in terms of the range of frequencies over which they can operate. If we examine an amplifier's performance at extreme frequencies, the gain may be much less than the nominal value. In fact, if we go far enough, the gain may even be fractional, meaning that the "amplifier" is actually reducing the signal level.

The region where the nominal gain is accurate is referred to as the mid-band. This range is defined by one or two corner or break frequencies. The lower limit is referred to as f_1 while the upper limit is referred to as f_2 . At these frequencies, the output level has dropped to half the power exhibited by a mid-band frequency of the same input level. A gain versus frequency response plot that encapsulates this concept is shown in Figure 5.4.1.

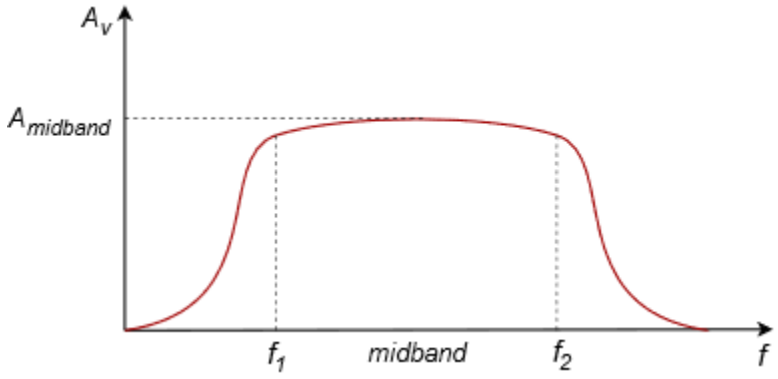


Figure 5.4.1: A generic gain versus frequency plot.

In this representative plot it is apparent that only input signals whose frequencies lay between f_1 and $f2$ will receive full amplification. As the input frequency moves to either side of the middle band, the gain begins to drop off. The drop-off increases as the signal frequency moves farther and farther away. Eventually the gain will fall to practically zero and virtually no trace of the input signal will appear at the output.

Precise values of the corner frequencies will depend on the application. For example, a high fidelity audio amplifier will most likely have an f_1 below 20 Hz and an f_2 above 20 kHz21 while an amplifier used for telephone systems might range from 300 Hz to 4 kHz22. In contrast, a radio frequency amplifier may be operating at frequencies orders of magnitude higher than these.

Without exception, all amplifiers have an upper limit frequency, f_2 , but not all of them have a lower frequency limit, f_1 . Amplifiers without a lower limit can amplify signals with frequencies all the way down to DC. They are referred to as direct coupled or DC amplifiers. The lower frequency limit is

usually caused by in-line coupling capacitors, and in some cases, transformers. Among other uses, these components are added to purposely block DC. There are good reasons to do this, as we shall see in upcoming work, however, it is possible to design amplifiers without them. The resulting amplifier will then have no limit on how low of a frequency it can amplify.

The upper limit frequency is another story. While components are often added to tailor the upper frequency response of an amplifier, even if no tailoring was desired the amplifier would still have an upper limit frequency. This would be due to small and unavoidable capacitances and inductances that exist in the circuit, for example stray wiring capacitance. Ultimately, the corresponding reactances will cause a signal level reduction that worsens as frequency increases. As you might guess, these reactances also cause varying phase shifts between the input signal and the output signal.

Amplifier performance is also limited by its internal noise. Noise is an undesired signal that appears at the output of an amplifier. Unlike distortion, noise is usually not correlated with the input signal level. Generally, noise is broad-band, meaning that it contains a very wide range of frequencies. As such, it does not have a discernible pitch. Examples in nature include the sound of leaves rustling in the wind or the sound of a waterfall. Noise is best thought of as a truly random signal. As such, it cannot be accurately predicted and therefore there is no easy way to remove it once it has been added to a desired signal. There are many potential sources of noise in an amplifier. They range from process issues in semiconductors to thermal effects in resistive elements. In general, noise gets worse as temperature, resistance and frequency range increase. Noise is unavoidable in absolute terms but ultimately what we care about is whether or not it is low enough for a given application. In other words, is the noise level significantly lower than the signal level, to the point where it is no longer a problem? This is quantified by simply creating a ratio between the nominal output signal level and the output noise level. This ratio is given the very creative name signal-to-noise ratio, or S/N for short. All other factors being equal, the higher the S/N, the better.

REFERENCES

[1]20 Hz – 20 kHz is the range of frequencies heard by a typical healthy young human.

[2]Decidedly not hi-fi, but do we really need high fidelity to call-in a take-out order?

UNIT 6: BJT SMALL SIGNAL AMPLIFIERS

Learning Objectives

After completing this chapter, you should be able to:

- Determine the voltage gain, input impedance and output impedance of simple BJT amplifiers.
- Detail the functional differences between voltage amplifiers and voltage followers.
- Explain the advantages and disadvantages of using localized feedback (swamping). D
- etermine the combined characteristics of multistage BJT amplifiers.
- Detail the advantages and disadvantages of using direct coupling versus capacitor coupling in multistage amplifiers.
- Explain the operation of the Darlington pair.

6.1 INTRODUCTION

In the prior chapter we discussed the general operational characteristics of amplifiers including voltage gain, input and output impedance, compliance, distortion and so forth. In this chapter we shall focus on the analysis of small signal amplifiers, specifically, their voltage gain and input/output impedances. As we will be performing a small signal analysis, we will not be concerned with compliance, maximum load power, device dissipation or the like. There is no specific definition of small signal versus large signal but for our purposes we shall define small signal as output signals that are well below the clipping limit and with power dissipation of no more than a few hundred milliwatts for either the load or transistor.

There are two popular techniques used to analyze BJT amplifier circuits. One is through the use of hybrid parameters. There are four different hybrid parameters. We have already seen one of them, the forward current gain, h_{fe}. We simply call it β. The other three are h_{ie} , the input impedance; h_{oe}, the output admittance; and h_{re} , the reverse voltage gain. The second letter of the subscript (the "*e*" in h_{fe}) indicates it is for the common emitter configuration (that is, input applied to the base, output taken at the collector and the emitter at the common ground).

The second approach uses r' parameters (pronounced "r prime"). The r' approach is sufficient for all of our analyses and, given an understanding of Ohm's law, KVL and KCL, produces straightforward equations for circuit gain, input impedance and the like. As a consequence, we shall focus on the r' system.

6.2 SIMPLIFIED AC MODEL OF THE BJT

Just as we created a DC model to ease the analysis of DC bias circuits, we shall make use of an AC BJT model for our AC analyses. In fact, our AC model is based on the DC model. The collector-base region is still represented with a current-controlled current source although it's AC instead of DC: $i_C = \beta i_B$. The base-emitter junction is a bit trickier. Although a simple 0.7 volt junction worked fine for DC, we now have to consider the AC resistance of the diode.

To find the dynamic resistance of the junction, first recall that the AC signal is riding on the DC bias current, as plotted in Figure 6.2.1.[1] We can imagine that the AC signal is causing this point to trace back and forth along the curve. Of course, as this is a small signal analysis, this sweep will be very small, perhaps only a few percent of the quiescent current and can be approximated as a straight line segment. The slope of this line segment represents its conductance.[2] The reciprocal of conductance is resistance; therefore, the reciprocal of the slope represents the resistance of the device. Consequently, we can approximate the dynamic resistance of the device as the reciprocal of the slope of the line tangent to the operating point (that is, the reciprocal of the slope of the line tangent to the quiescent bias current I_C). In reality, this slope is changing slightly as the signal swings back and forth along the base-emitter I-V curve. As the signal swings positive and goes above the quiescent point the slope is a little steeper producing a slight reduction in dynamic resistance. In contrast, as it swings negative, going below the quiescent point, the slope becomes a bit more shallow and produces a slightly higher resistance. As a result, we are effectively computing an average value for the dynamic resistance by assuming this is a straight line segment. The variance in this resistance will be a source of asymmetrical distortion in the amplifier of the type shown in Chapter 6, Figure 6.3.4. We shall see more on this later.

1. The values plotted along the current axis are typical of a generic device and do not represent the current values for all BJTs.
2. Technically, this value is called the device's transconductance and is denoted as g_m. We shall be seeing this again in upcoming work.

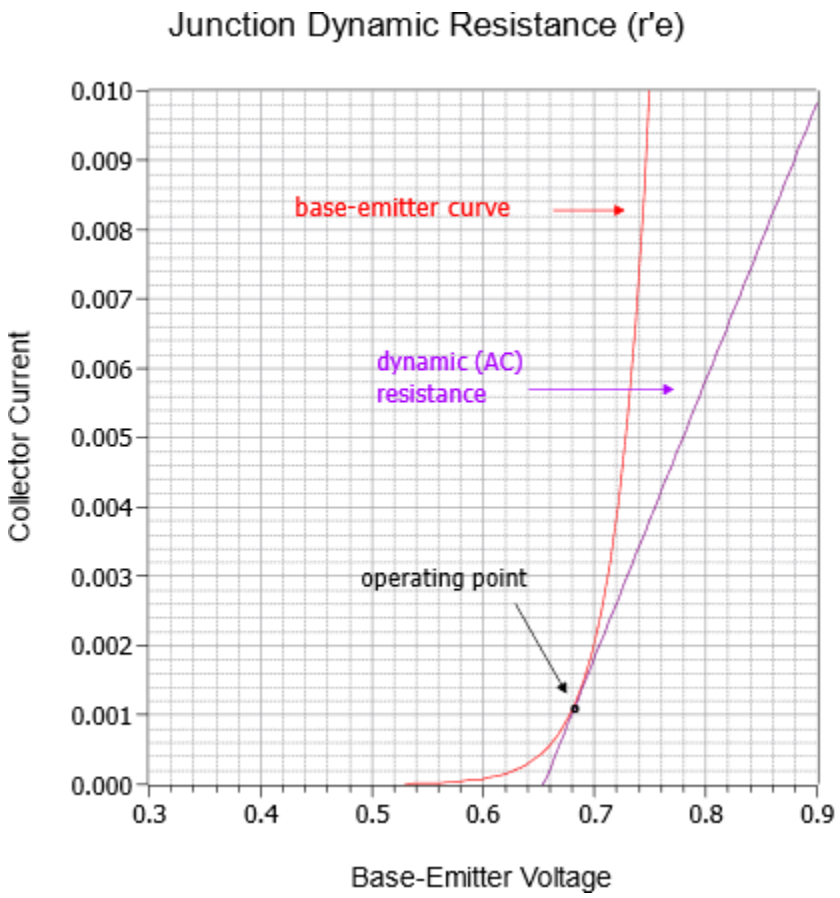


Figure 6.2.1: Base-emitter junction plot and dynamic resistance.

In order to derive an equation for the dynamic resistance, we begin with the Shockley equation from Chapter 2, Equation 2.2.1, slightly modified to reflect the terminal names of a BJT.

$$I_C = I_S \left(e^{\frac{V_{BE} q}{nkT}} 1 \right)$$

I_C is the junction (collector) current,

I_S is the reverse saturation current,

V_{BE} is the voltage across the junction (base-emitter),

q is the charge on an electron, 1.6E–19 coulombs,

n is the quality factor (typically between 1 and 2),

k is the Boltzmann constant, 1.38E–23 joules/kelvin,

T is the temperature in kelvin.

At 300 kelvin (about 80°F), q/kT is approximately 38.6, thus for any reasonable value of V_{BE} the "–1" term is small enough to ignore. Also, we shall take n as 1.

The equation then reduces to

$$I_C = I_S e^{38.6 V_{BE}}$$

(6.2.1)

To find the slope we take the first derivative of Equation 6.2.1 with respect to V_{BE} .

$$\frac{dI_C}{dV_{BE}} = 38.6 I_S e^{38.6 V_{BE}}$$

(6.2.2)

Substituting Equation 6.2.1 into Equation 6.2.2 yields

$$\frac{dI_C}{dV_{BE}} = 38.6 I_C$$

By definition, the dynamic junction resistance is the reciprocal of the slope.

$$\frac{dV_{BE}}{dI_C} = 25.9mVI_C\}Wecallthis < em > r < /em > .Thisvalueisslightlylowasitdoesn'tincludebulkresistancesoagoodapproximationisr'_e = \frac{26mV}{I_C}$$

(6.2.3)

It is important to note that I_S in Equation 6.2.2 varies with temperature. Therefore r' varies with temperature as well, decreasing with increasing temperature. This carries important ramifications with the thermal stability of higher power amplifiers as we shall see in subsequent work.

One of the most important things to remember here is that the DC collector current sets up the resistance of the AC model. In other words, the stability of the AC circuit will depend in part on the stability of the DC bias (hence our emphasis on stable bias circuits in Chapter 5).

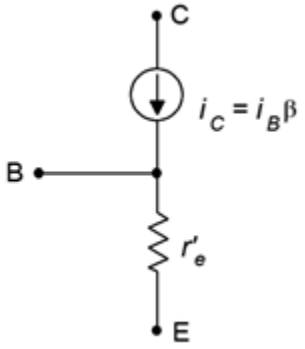


Figure 6.2.2: Simplified AC model of BJT.

We now have our AC model, as shown in Figure 6.2.2. This is a simplified model in that it does not include junction capacitance effects, lead inductance and the like. It is appropriate, therefore, as a

low to mid-band frequency model. To summarize, the AC collector current, I_C , is determined by the AC input current, I_B ; which is in turn a function of the the size of the applied input e signal. In contrast, r'_e is set by the DC bias current, I_C. The AC input can produce small variations in r'_e which are manifested as waveform distortion.

Given the model, there are three ways to configure the transistor as an amplifier:

- Common Emitter. The input is applied to the base and the output is taken at the collector. The emitter terminal is at the common or ground point. This configuration exhibits both voltage gain and current gain. It also inverts the phase of the signal.
- Common Collector. The input is applied to the base and the output is taken at the emitter. The collector terminal is at the common or ground point. This configuration offers a voltage gain of about unity but does exhibit current gain. It maintains the phase of the input signal. It is also referred to as an emitter follower or voltage follower.
- Common Base. The input is applied to the emitter and the output is taken at the collector. The base terminal is at the common or ground point. This configuration exhibits voltage gain but the current gain is unity at best. It also maintains the phase of the input signal.

We shall examine each of these topologies in turn. Each of these can be made using a variety of DC bias techniques. For example, a two-supply emitter bias or voltage divider bias could be used for any of the three AC topologies, and further, they could utilize either an NPN or PNP transistor.

6.3 COMMON EMITTER AMPLIFIER

The common emitter configuration finds wide use as a general purpose voltage amplifier. We begin with a basic DC biasing circuit and then add a few other components. For example, refer to Figure 6.3.1.

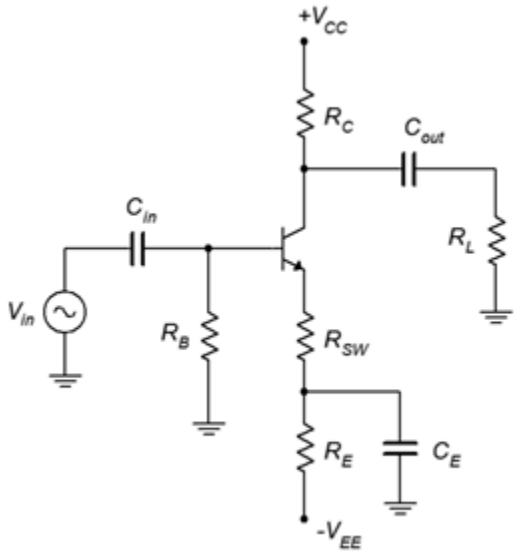


Figure 6.3.1: Common emitter amplifier using two-supply emitter bias.

This amplifier is based on a two-supply emitter bias circuit. The notable changes are the inclusion of an input signal voltage, V_{in}, and a load, R_L. So that these components do not alter the bias, we isolate the input and load through the use of coupling capacitors C_{in} and C_{out} . These capacitors will act as opens to DC creating the desired isolation. As for the AC signal, the capacitances will be chosen such that their reactances will be much smaller than the surrounding resistors at the frequency of the input. Consequently, the capacitors will appear as shorts and allow the AC signal to pass through the amplifier.

The final alteration involves the emitter resistor. The single resistor of the bias network is replaced by a pair of resistors, R_E and R_{SW} , along with a bypass capacitor, C_E. For DC, the capacitor is open and the effective emitter bias resistance is $R_E + R_{SW}$. For AC, the capacitor will behave ideally as a short so the AC emitter resistance will fall to just R_{SW} . This resistor is called a swamping or emitter degeneration resistor. It is used primarily to help control the voltage gain of the amplifier.

We can use our AC transistor model along with the Superposition Theorem to arrive at an equivalent AC circuit of the amplifier, as shown in Figure 6.3.2.

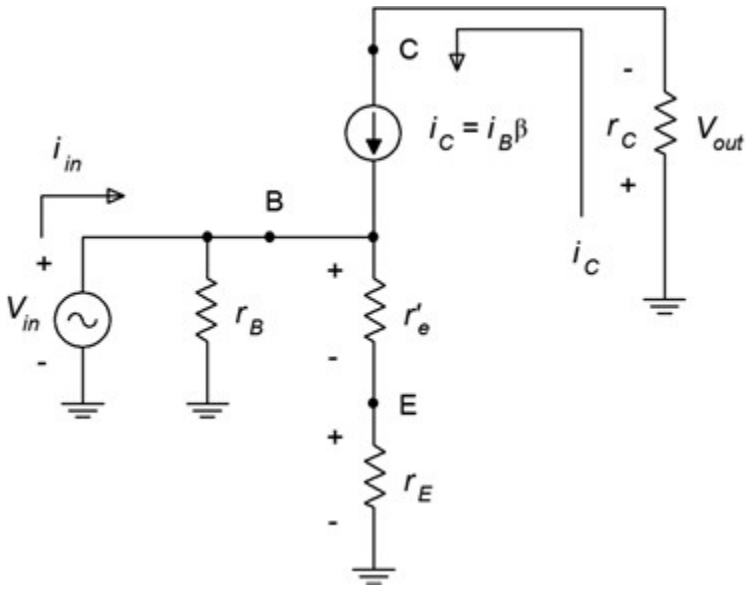


Figure 6.3.2: AC equivalent of common emitter amplifier.

First, we have shorted all of the capacitors. Second, we have replaced the DC sources with their ideal internal resistance (a short) which places those points at AC ground. Third, we swapped out the transistor for the model. Lastly, we have combined and/or renamed resistances where needed. Because this is an AC circuit, we use the convention of lower case *r* for resistance to avoid confusion with the DC resistance (which are upper case). Thus, r_E is the AC resistance from the emitter to AC ground. This corresponds to R_{SW} in the original schematic. Similarly, r_C represents the total resistance seen from the collector to AC ground. In the original schematic this corresponds to R_C in parallel with R_L. If this circuit was unloaded, then r_C would just be equal to R_C . Finally, r_B corresponds to R_B but in a voltage divider bias it would be equal to R_1 in parallel with R_2 .

VOLTAGE GAIN

Voltage gain, A_v, is defined as the ratio of v_{out} to v_{in}. Using Ohm's law we find

$$A_v = \frac{v_{out}}{v_{in}} = \frac{v_C}{v_B}$$

$$A_v = \frac{-i_C r_C}{i_C(r'_e + r_E)}$$

$$A_v = -\frac{r_C}{r'_e + r_E}$$

(6.3.1)

First, the negative sign indicates that this amplifier inverts the waveform, top to bottom. For a sine wave, this is equivalent to shifting the phase 180° . In some applications this can be a major issue, in others, not so much. If it is an issue, it can be resolved by using a second inverting gain amplifier in sequence with the first (inverting the inversion).

The second thing we see is that the gain is little more than a ratio of collector to emitter resistances. This is where splitting the emitter resistor into two parts comes in. In the equation, r_E is the swamping resistor R_{SW} . The larger the swamping resistor, the lower the gain. The maximum gain will be achieved when $R_{SW} = 0$. That is, when the emitter is completely bypassed. The down side of this is that the gain will now depend entirely on r' . This will increase the distortion. The reason is because R_{SW}, being so
much larger, effectively "swamps out" the variation in r' and reduces distortion. The larger R_{SW} is relative to r' , the greater the reduction in distortion, but with the cost of reduced gain. This is why a swamping resistor is also called an emitter degeneration resistor: it degrades the voltage gain.

INPUT IMPEDANCE

Input impedance, Z_{in}, is defined as the ratio of v_{in} to i_{in} . In Figure 6.2.2 this is equal to r_B in parallel with the impedance looking into the base terminal, $Z_{in(base)}$. Using Ohm's law we find

$$Z_{in(base)} = \frac{v_B}{i_B}$$

$$Z_{in(base)} = \frac{i_C(r'_e + r_E)}{i_B}$$

$$Z_{in(base)} = \frac{i_C(r'_e + r_E)}{i_C/\beta}$$

$$Z_{in(base)} = \beta(r'_e + r_E)$$

(6.3.2)

Therefore

$$Z_{in} = r_B||Z_{in(base)}$$

(6.3.3)

We see that both the swamping resistor and β play a role in setting the input impedance. Larger values of R_{SW} and β produce larger input impedances. In sum, we find that while swamping decreases voltage gain, it reduces distortion and increases input impedance, the latter two generally desirable for a voltage amplifier. A non-swamped amplifier will have the largest gain but will suffer from the worst distortion and a low input impedance. This is a classic "quality versus quantity" trade-off: a large low quality gain versus a modest high quality gain.[1]

1. The obvious question is, "How do we get both high gain and low distortion?" One answer is to use multiple low gain stages in cascade.

OUTPUT IMPEDANCE

Output impedance, Z_{out} , is defined as the internal impedance of the equivalent source that drives the load. If we position ourselves at the load and look back into the amplifier shown in Figure 6.2.1, C_{out} is shorted ideally and V_{CC} is at AC ground. This leaves us with R_C in parallel with the transistor. The transistor is modeled as a current source and its ideal internal resistance would approach infinity. In reality, the effective value, r' , is likely in the region of 100 kΩ or so, depending on bias current. This parallel combination comprises the output impedance of the current source. We model this circuit as a voltage amplifier so to be proper, we'd convert the current source with parallel internal resistance to a voltage source with series internal resistance. Those resistance values are identical, though, and we arrive at

$$Z_{out} = r'_C || R_C$$

In many circuits, RC is considerably smaller than r' , therefore

$$Z_{out} \approx R_C$$

(6.3.4)

Example 6.3.1

Determine the input and output impedances of the amplifier shown in Figure 6.3.3. Also compute the voltage gain. Assume β = 150.

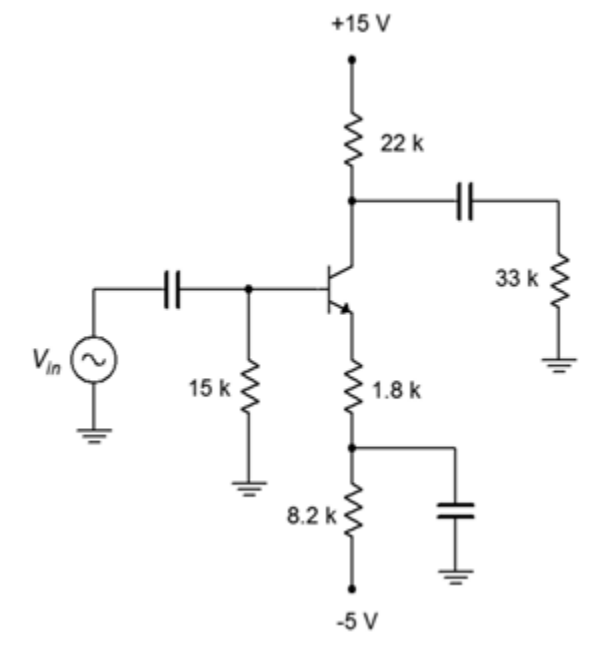


Figure 6.3.3: Schematic for Example 6.3.1.

First, the easy bit. We can determine the output impedance by inspection. It is approximately equal to R_C , or 22 kΩ.

In order to find Z_{in} and A_v, we will need to determine r' . To obtain r' we need to find I_C . Using KVL around the base-

emitter loop, if we approximate the DC base voltage to be near zero, then all of the emitter supply drops across the DC emitter resistance, with the exception of V_{BE} .

$$I_C = \frac{|V_{EE}| - V_{BE}}{R_E + R_{SW}}$$

$$I_C = \frac{5V - 0.7V}{8.2k\Omega + 1.8k\Omega}$$

$$I_C = 0.43mA$$

$$r'_e = \frac{26mV}{I_C}$$

$$r'_e = \frac{26mV}{0.43mA}$$

$$r'_e = 60.5\Omega$$

$$Z_{in-base} = \beta(r'_e + r_E)$$

$$Z_{in-base} = 150(60.5\Omega + 1.8k\Omega)$$

$$Z_{in-base} = 279k\Omega$$

This value in parallel with the base biasing resistor creates the input impedance.

$$Z_{in} = R_B||Z_{in(base)}$$

$$Z_{in} = 15k\Omega||279k\Omega$$

$$Z_{in} = 14.2k\Omega$$

First, method one.

$$r_C = R_C||R_L$$

$$r_C = 22k\Omega||33k\Omega$$

$$r_C = 13.2k\Omega$$

$$A_v = -\frac{r_C}{r'_e + r_E}$$

$$A_v = -\frac{13.2k\Omega}{60.5\Omega + 1.8k\Omega}$$

$$A_v = -6.1$$

And now method two; first the unloaded gain, then the divider effect and finally, the composite gain.

$$A_{v(unloaded)} = -\frac{r_C}{r'_e + r_E}$$

$$A_{v(unloaded)} = -\frac{22k\Omega}{60.5\Omega + 1.8k\Omega}$$

$$A_{v(unloaded)} = -11.82$$

$$A_{divider} = \frac{R_L}{R_L + R_C}$$

$$A_{divider} = \frac{33k\Omega}{33k\Omega + 22k\Omega}$$

$$A_{divider} = 0.6$$

$$A_v = A_{v(unloaded)} \times A_{divider}$$

$$A_v = -11.82 \times 0.6$$

$$A_v = 6.1$$

We shall repeat the prior example using the same circuit but with one change: the emitter resistor will be completely bypassed. This will show the effect that swamping has on voltage gain and input impedance.

Example 6.3.2

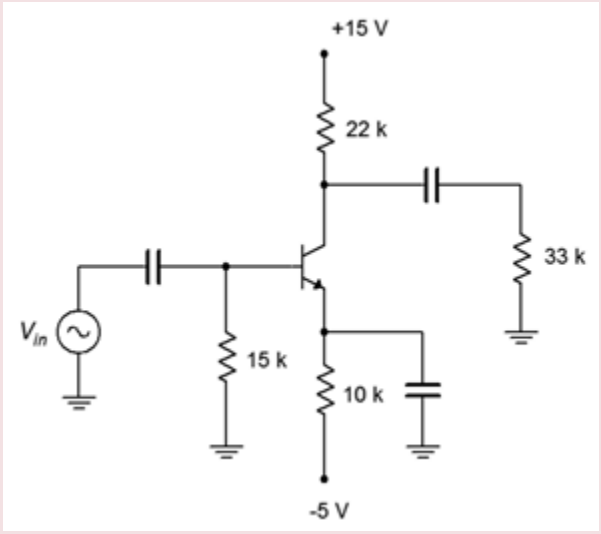


Figure 6.3.4: Schematic for Example 6.3.2 .

Determine the voltage gain and input impedance of the amplifier shown in Figure 6.3.4. Assume β=150.

The DC equivalent of this circuit is identical to that of the circuit shown in Figure 6.3.3. In both cases, the DC emitter resistance is 10 kΩ. Therefore, I_C and r'_e are unchanged. The bypass capacitor shorts this entire value for the AC equivalent because there is no swamping resistor. Consequently, $r_E = 0$. We can simply use 0 for r_E in the equations previously derived.We begin with the input impedance.

$$Z_{in-base} = \beta(r'_e + r_E)$$

$$Z_{in-base} = 150(60.5\Omega + 0)$$

$$Z_{in-base} = 9075\Omega$$

This value is considerably smaller than the value obtained from the swamped circuit. Continuing,

$$Z_{in} = R_B||Z_{in(base)}$$

$$Z_{in} = 15k\Omega||9075\Omega$$

$$Z_{in} = 5654\Omega$$

$$A_v = -\frac{r_C}{r'_e + r_E}$$

$$A_v = -\frac{13.2k\Omega}{60.5\Omega + 0}$$

$$A_v = -218.2$$

The end result is an input impedance less than half of the swamped case and a voltage gain over 30 times greater. What these calculations do not show is the increase in distortion that will be created by this change. More on that in a moment.

Let's consider something slightly different: a voltage divider bias PNP amplifier.

Example 6.3.3

Determine the input impedance and voltage gain for the circuit shown in Figure 6.3.5.
Also determine V_{load} if V_{in} = 20 mV peak. Assume ß=100.

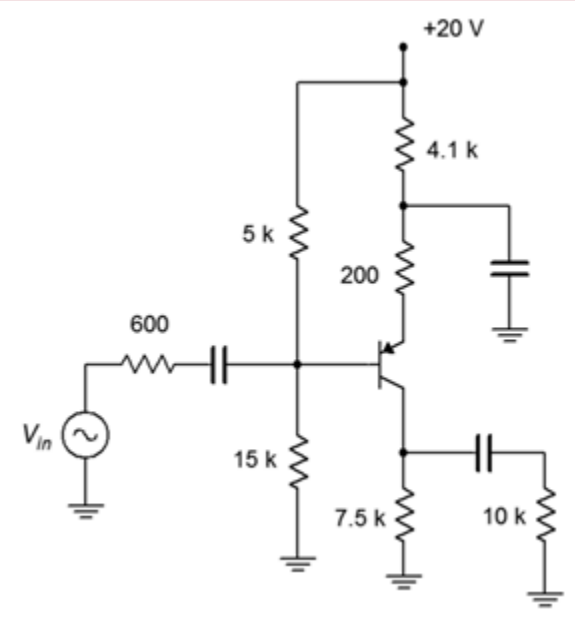


Figure 6.3.5: Schematic for Example 6.3.3.

We need to first determine r'_e which means that we need to find the collector current. If we assume a lightly loaded divider, the base voltage will be approximately 15 volts and the emitter will be 0.7 volts higher, or 15.7 volts. This leaves 20 volts – 15.7 volts, or 4.3 volts, across the DC equivalent emitter resistance. That's 4.1 k Ω + 200 Ω, or 4.3 kΩ, yielding 1 mA for I_C. This will produce r'_e = 26 Ω.

$$Z_{in(base)} = \beta(r'_e + r_E)$$

$$Z_{in(base)} = 100(26\Omega + 200\Omega)$$

$$Z_{in(base)} = 22.6k\Omega$$

This value is in parallel with the voltage divider biasing resistors, creating the input impedance.

$$Z_{in} = R_1||R_2||Z_{in(base)}$$

$$Z_{in} = 15k\Omega||5k\Omega||22.6k\Omega$$

$$Z_{in} = 3.22k\Omega$$

$$A_v = -\frac{r_C}{r'_e + r_E}$$

$$A_v = -\frac{7.5k\Omega||10k\Omega}{26\Omega + 200\Omega}$$

$$A_v = 19$$

We also need to include the effect of the 600 Ω source impedance. This will create a voltage divider with the input impedance.

$$_{divider} = \frac{Z_{in}}{Z_{in} + Z_{source}}$$

$$A_{divider} = \frac{3.22k\Omega}{3.22k\Omega + 600\Omega}$$

$$A_{divider} = 0.843$$

$$A_{v(system)} = A_v \times A_{divider}$$

$$A_{v(system)} = -19 \times 0.843$$

$$A_{v(system)} = -16$$

Finally, we get to the load voltage.

$$V_{load} = A_{v(system)} \times V_{in}$$

$$V_{load} = -16 \times 20mV$$

$$V_{load} = 320mV \text{ peak, inverted}$$

If we were to inspect the circuit of Figure 6.3.5 using a direct coupled oscilloscope, we would see the superposition of the AC and DC components. In other words, we'd see the AC signal riding on a DC offset. In some cases, the AC signal would be too small to notice compared to the DC portion. In proper scale it might be no thicker than the trace itself. In order to measure it accurately, we'd have to AC couple the oscilloscope.

The voltages at the source and load would be just AC as the coupling capacitors serve to block DC. At the base we'd have 15 volts DC with an AC signal riding on top of it. The AC would be the 20 mV

input times the input impedance/source impedance divider of 0.843, or 16.86 mV. Recalling that I_C is 1 mA, the DC drop across R_C must be 7.5 volts. This is, of course, V_C . Therefore, at the the collector we'd see an inverted 320 mV signal riding on 7.5 volts DC.

COMPUTER SIMULATION

In order to get some insight into the swamping-versus-distortion issue, we shall take a look at a more involved circuit simulation. This will echo Examples 6.3.1 and 6.3.2 in that we will simulate two circuits with the same DC equivalents. The only circuit change will be that one version will have a fully bypassed emitter while the other version will utilize a swamping resistor. In order to keep the comparison fair, we will increase the input signal voltage of the lower gain swamped amplifier so that both versions have a similar load voltage. In this way we guarantee that they are both using a similar percentage of the junction curve.

The unswamped circuit is shown in Figure 6.3.6. This utilizes a straightforward two-supply emitter bias.

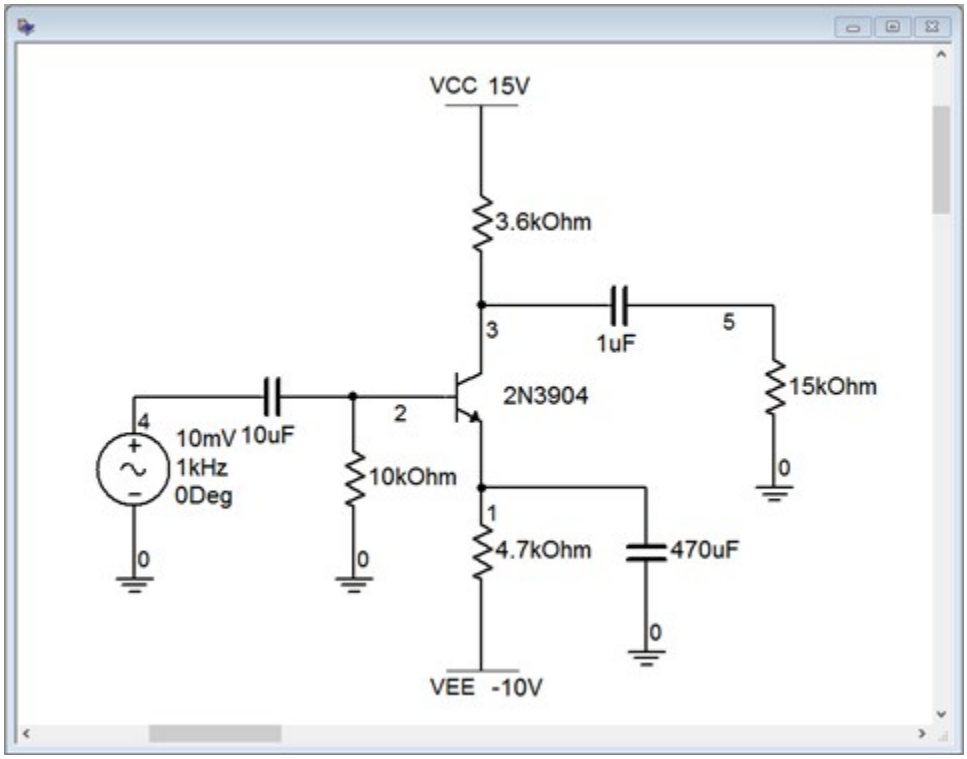


Figure 6.3.6: Unswamped CE amplifier in simulator.

A quick "back-of-an-envelope" estimate gives $I_C \approx 2$ mA, yielding $r' \approx 13\Omega$. The load will be around 3 kΩ which gives a gain in the low 200s. Thus, we expect the load voltage to be around 2 volts. The transient analysis graph is depicted in Figure 6.3.7. Several traces are shown.

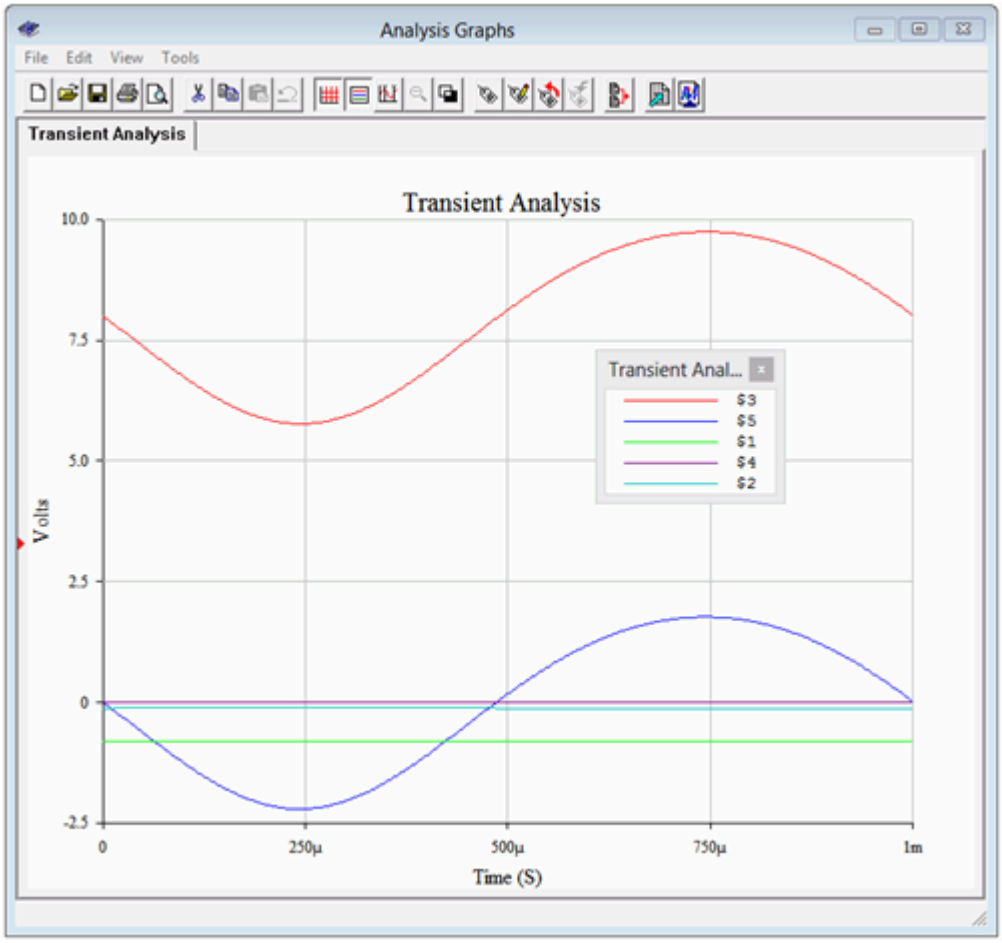


Figure 6.3.7: Unswamped CE amplifier, Transient Analysis

At this scale, the AC signal at the input (node 4, purple) and the base (node 2, aqua) cannot be seen. As expected, we see a small negative DC value at the base and at the emitter, around –0.7 VDC. The DC offset at the collector is around 8 volts, as expected. Finally, the load voltage (node 5, blue) is sitting right around 2 volts.

What might not be visible immediately in the load voltage plot is some waveform asymmetry distortion. This can be quantified through a THD simulation, the output of which is shown in Figure 6.3.8. THD is nearly 8%. Not so good.

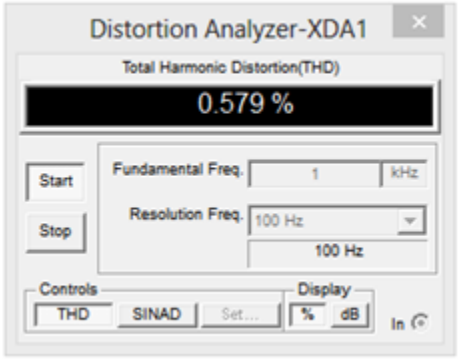


Figure 6.3.8: Non-swamped CE amplifier, THD Analysis.

For the second pass, the circuit is modified to include a swamping resistor, as illustrated in Figure 6.3.9. The original 4.7 kΩ emitter resistor has been split into a 4.5 kΩ and a 200 Ω swamping resistor. The bias in this circuit is identical to the first, therefore re′ is unchanged. This will lower our expected gain to around 13, decreasing by a factor of 15. The input signal is raised by a factor of 15 to compensate so that our load voltage will still be around 2 volts.

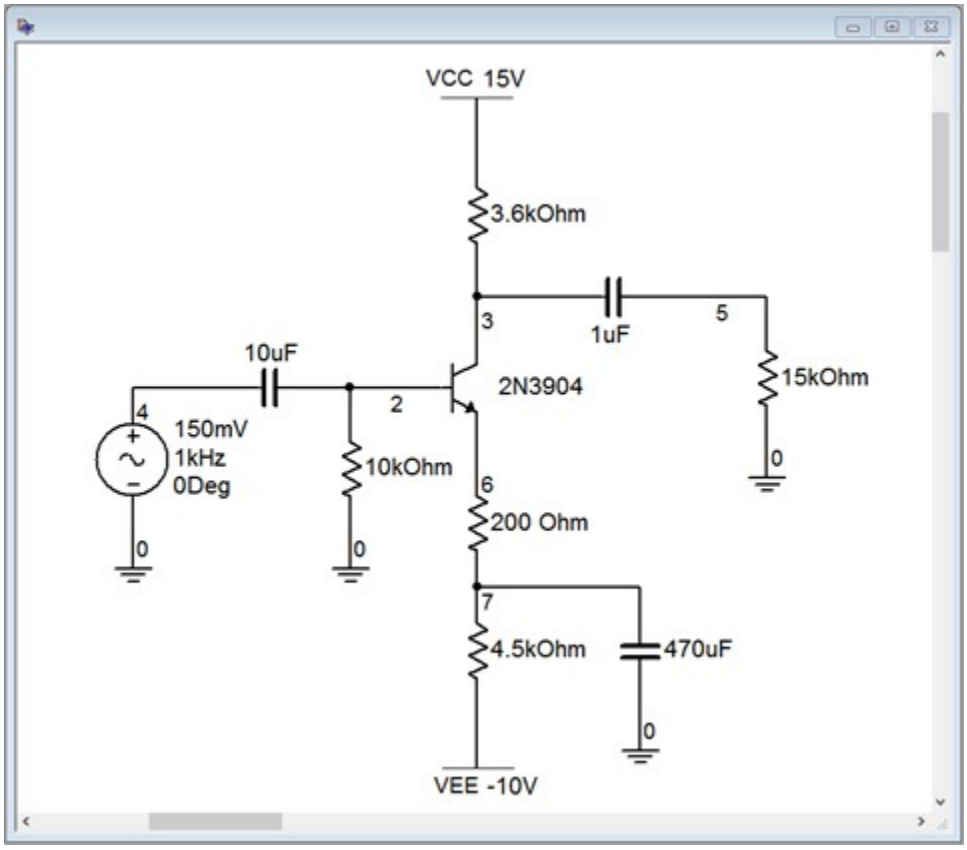


Figure 6.3.9: Swamped CE amplifier in simulator

Once again, we run a transient analysis. The results are shown in Figure 6.3.10. In this case we have done something a little different. By zooming in, we can now confirm the signal inversion. The input signal is the purple trace at node 4. We can also see this signal at the base, riding on the small negative DC bias voltage (aqua trace, node 2). The DC offset is about –0.1 volts. Looking at the emitter we see the expected 0.7 volt DC base-emitter drop below this, or about –0.8 volts DC. Notice that there is no AC signal at the emitter whatsoever. This is expected as the emitter bypass capacitor forces this point to an AC ground.

The load voltage is the blue trace, node 5. While much of it is not visible at this zoom level, clearly it is an inverted waveform when compared to the input signal.

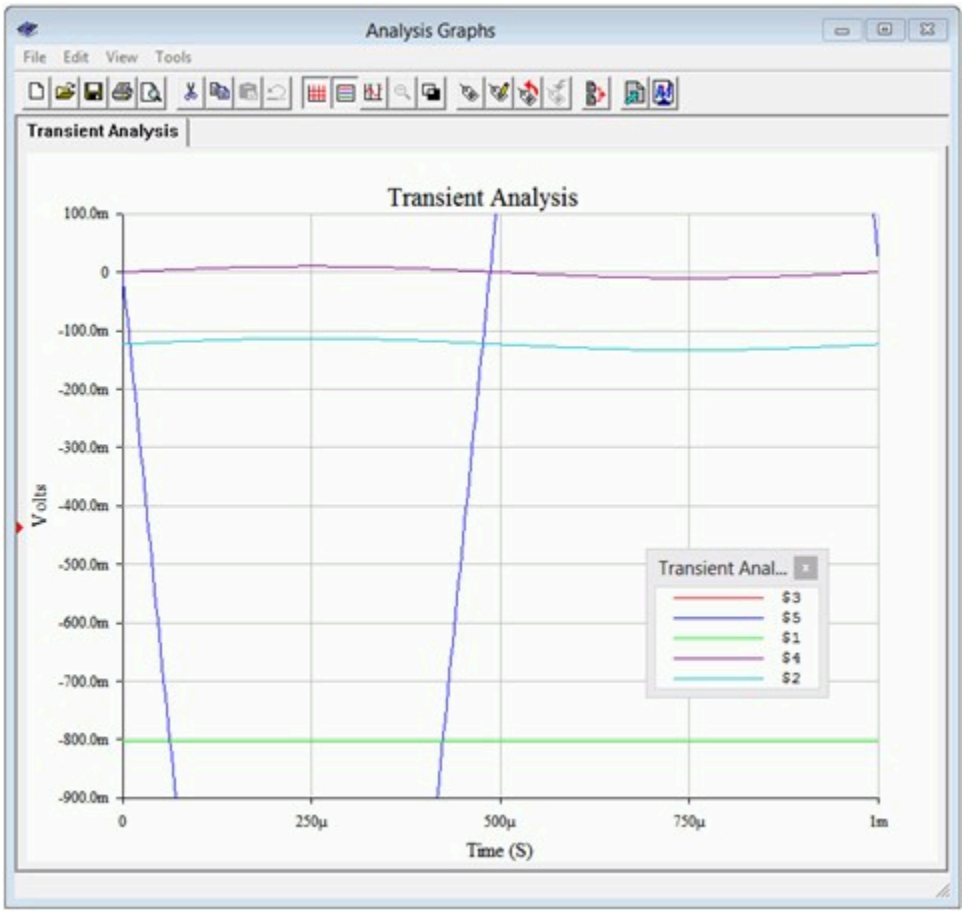


Figure 6.3.10: Swamped CE amplifier, Transient Analysis.

But what about the load voltage distortion? A THD simulation is performed on the swamped amplifier with the results shown in Figure 6.3.11. The THD is now under .6 %, a considerable improvement, even if not audiophile quality. Interestingly, as a ratio, the reduction in distortion is roughly equal to the reduction in gain. The more you give up, the more you get.

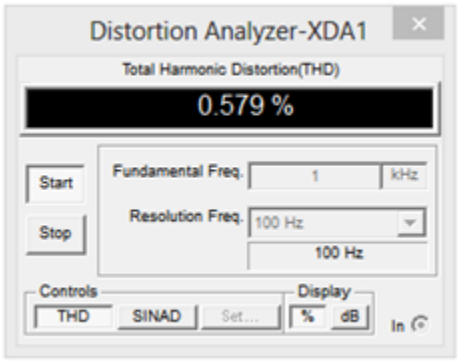


Figure 6.3.11: Swamped CE amplifier, THD Analysis.

Finally, the change in signal quality can be seen readily by plotting both load voltages concurrently, as shown in Figure 6.3.12. The non-swamped output (in blue) exhibits telltale asymmetry. Notice that the positive peak does not quite reach 2 volts but the negative peak exceeds –2 volts. The positive peak is also broadened and flattened, while the negative peak is sharper. In contrast, the swamped output

(in red) has virtually identical positive and negative peak values with no apparent shape changes on them. Compare this simulation to the waveform distortion discussion from Chapter 6. In particular, compare Figure 6.3.12 to Figure 6.3.4.

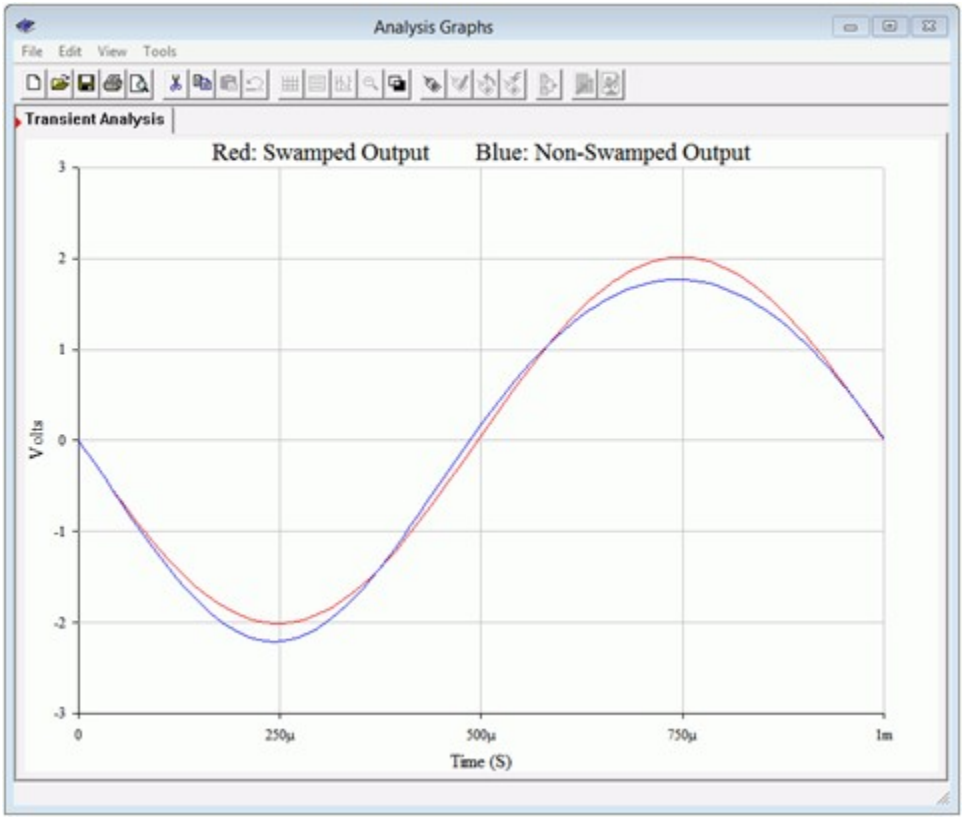


Figure 6.3.12: Swamped versus non-swamped CE amplifiers, Transient Analysis.

POWER SUPPLY BYPASS AND DECOUPLING

In the prior analyses we have assumed ideal behavior from the DC power sources. First, we assumed that they present a perfect AC ground and second, that they exhibit no ripple or noise. In reality, this may not be the case and non-ideal behavior may lead to a number of problems that diminish the quality of the amplified output signal, including hum and oscillations.

To battle the first issue, power supply bypass capacitors may be used. These capacitors are usually modest in size, perhaps 1 μF or so, although they can be much larger, particularly with high output power amplifiers. Power supply bypass capacitors are located physically close to the active devices. This location minimizes the resistive and inductive effects of power supply circuit board traces and wiring that could result in the power supply not being a good AC ground.

The second issue involves the noise and ripple from the power supply finding its way into the input signal and becoming part of the output signal. A classic example of this is amplifiers that use voltage divider biasing such as the one shown in Figure 6.3.5. Not only does the divider create the needed DC potential at the base terminal, but it also couples in any noise or ripple that might be riding on the DC voltage. This is particularly nasty because this undesirable signal is being applied to the base where it will get amplified.

The obvious solution to this problem is to create a very high quality, regulated DC supply, but this is not always practical given cost constraints. A relatively simple solution is to decouple the undesirable AC components through an *RC* network as shown in Figure 6.3.13.

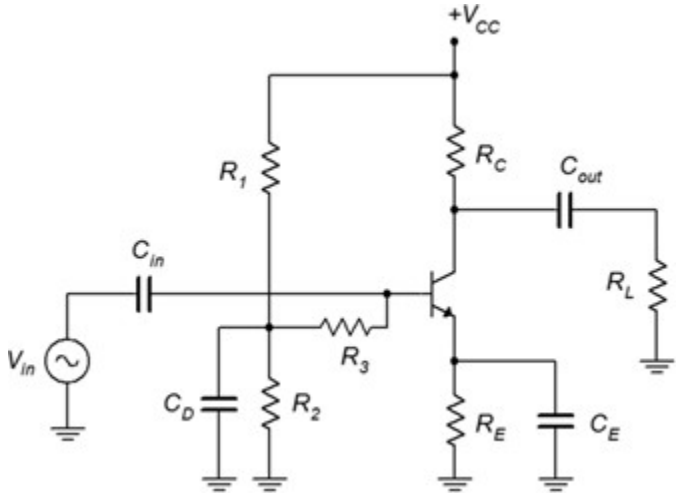


Figure 6.3.13: Decoupled voltage divider.

The capacitor C_D is used to create an AC ground at the divider junction, thus shunting any noise or ripple to ground. Unfortunately, this would also short out the input signal so R_3 impedance.

6.4 COMMON COLLECTOR AMPLIFIER

The common collector amplifier is often referred to as an emitter follower, or more generically, as a voltage follower. The key characteristics of a voltage follower are a high input impedance, a low output impedance and a non-inverting voltage gain of approximately one. The name comes from the fact that output voltage follows the input, that is, it's at the same voltage level and is in phase with the input. While this configuration does not produce voltage gain, it does produce current gain, and therefore, power gain. It's primary purpose is to reduce impedance loading effects, for example, to match a high impedance source to a low impedance load. Consequently, they are used as high-Z input buffer stages or as drivers for low impedance loads such as loudspeakers.

A common collector amplifier using two-supply emitter bias is shown in Figure 6.4.1. The input is coupled into the base like the common emitter amplifier, however, the output signal is taken at the emitter instead of at the collector. Because the collector is at the AC common, there is no need for a collector resistor.

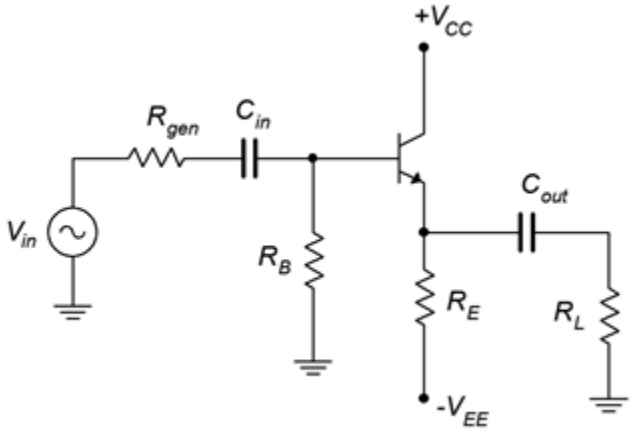


Figure 6.4.1: Common collector amplifier.

Perhaps the best way to think about the follower is not that it gives a voltage gain of one, but that it will prevent signal loss. The analysis follows, using Figure 6.4.2.

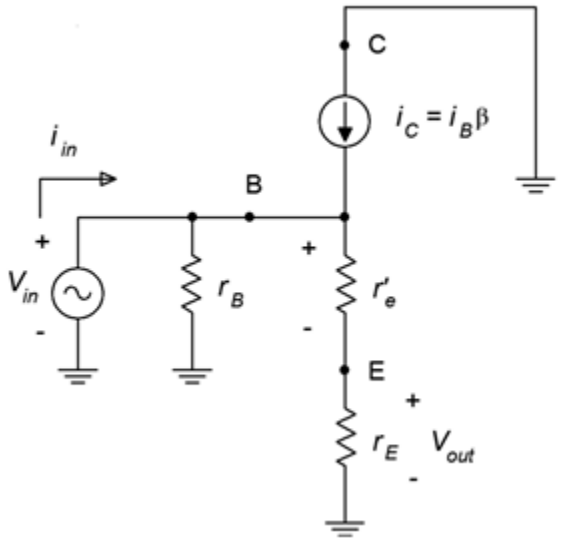


Figure 6.4.2: AC equivalent of common collector amplifier.

First, the AC emitter resistance, r_E , is either the emitter bias resistor, R_E, or the parallel combination of R_E

and the load resistance, R_L. We'll use the former in order to determine the unloaded gain and the latter to determine the loaded gain, similar to what we did with the common emitter amplifier concerning R_C and R_L. The AC base resistance, r_B, typically boils down to the base biasing resistor just as we saw with the common emitter amplifier (R_B in a two-supply emitter bias or $R_1||R_2$ for a voltage divider bias).

VOLTAGE GAIN

The derivation for the emitter follower's voltage gain equation is similar to that shown for the common emitter amplifier. We begin with the basic definition of voltage gain and then expand using Ohm's law.

$$A_v = \frac{v_{out}}{v_{in}} = \frac{v_E}{v_B}$$

$$A_v = \frac{i_C r_E}{i_C(r'_e + r_E)}$$

$$A_v = \frac{r_E}{r'_e + r_E}$$

(6.4.1)

This equation is very similar to that of Equation 7.3.1. Here we see that the output signal is in phase with the input and that if $r_E \gg r'_e$, the gain approaches unity. Signal distortion tends to be low in followers because a gain of one is a desired goal.

INPUT IMPEDANCE

The derivation for Zin and Zin(base) are unchanged compared to the common emitter configuration. The formulas are repeated below for convenience.

$$Z_{in(base)} = \beta(r'_e + r_E)$$

$$Zin = r_B || Z_{in(base)}$$

OUTPUT IMPEDANCE

The derivation for common collector output impedance varies considerably from that of the common emitter. We shall use Figure xx for the analysis.

First, note that this diagram splits the AC emitter resistance into its two components, R_L and the biasing resistor R_E. This is because we want to find the effective resistance of the source that drives the load, so logically we can't include the load in that value. We begin by looking back into the emitter from the perspective of the load. We see the emitter bias resistor in parallel with whatever the impedance is looking back into the emitter terminal.

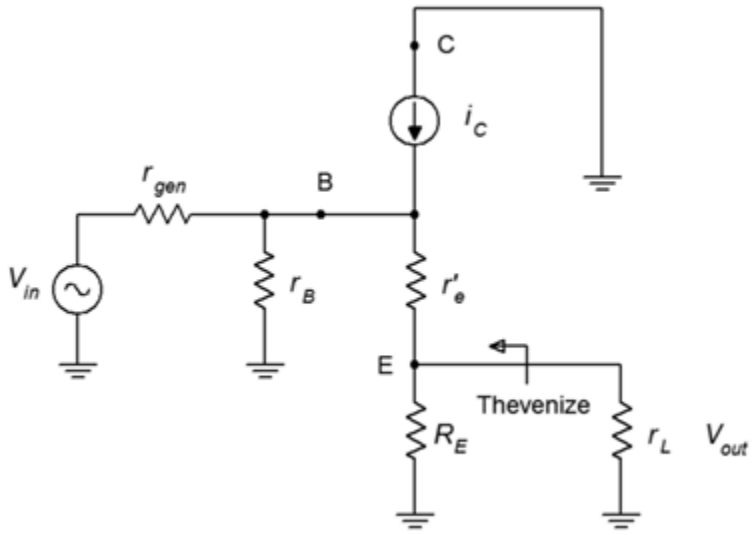


Figure 6.4.3: Common collector output impedance analysis.

$$Z_{out} = R_E || Z_{out(emitter)}$$

(6.4.2)

$Z_{out(emitter)}$ is equal to r'_e in series with the equivalent resistance of the network above it and to the left. The internal resistance of e the current source is high enough to ignore so we're left with the equivalent resistance looking back off the base. We'll call this $Z_B(equivalent)$. At first glance this might appear to be the parallel combination of r_{gen} and r_B, but this ignores the effect of the e collector current source. What we really want is the effective resistance as seen from the perspective of r' , not as seen from the base terminal.

$$Z_{out(emitter)} = r'_e + Z_{B(equivalent)}$$

(6.4.3)

$$Z_{B(equivalent)} = \frac{v_B}{i_C}$$

$$Z_{B(equivalent)} = \frac{i_B(r_B||r_gen)}{\beta i_B}$$

$$Z_{B(equivalent)} = \frac{r_B||r_{gen}}{\beta}$$

6.4.4)

Combining Equations 6.4.2, 6.4.3 and 6.4.4 yields

$$Z_{out} = R_E|| \left(r'_e + \frac{r_B||r_{gen}}{\beta} \right)$$

(6.4.5)

In many instances the emitter bias resistor is large enough to ignore.

Example 6.4.1

For the follower shown in Figure 6.4.4, determine the input impedance, output impedance and load voltage. Assume β = 100 and V_{in} = 100 mV.

First, find I_C in order to find r' . Assuming an unloaded divider, V will equal half of the DC supply, or 10 volts. We lose 0.7 volts across the base-emitter junction leaving 9.3 volts across the 10 kΩ. This results in a collector current of 930 μA and an r'_e of 28 Ω.

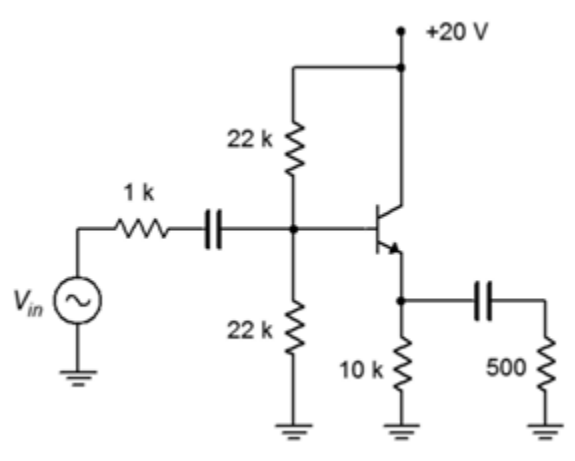


Figure 6.4.4: Schematic for Example 6.4.1.

To find Z_{in}

$$Z_{in(base)} = \beta(r'_e + r_E)$$

$$Z_{in(base)} = 100(28\Omega + 10k\Omega||500\Omega)$$

$$Z_{in(base)} = 50.4k\Omega$$

$$Z_{in} = R_1||R_2||Z_{in(base)}$$

$$Z_{in} = 22k\Omega||22k\Omega||50.4k\Omega$$

$$Z_{in} = 9.03k\Omega$$

This value is not particularly high when compared with the rather large source resistance of 1 kΩ. There will be some signal loss here due to the voltage divider effect between the two impedances. And now for Z_{out}

$$Z_{out} = R_E|| \left(r'_e + \frac{r_B||r_{gen}}{\beta} \right)$$

$$Z_{out} = 10k\Omega|| \left(28\Omega + \frac{22k\Omega||22k\Omega||1k\Omega}{100} \right)$$

$$Z_{out} = 37\Omega$$

This value is much, much lower than anything we saw with the common emitter amplifiers. Therefore this circuit can drive much lower impedance loads with minimal signal loss. The loaded gain from base to emitter is

$$A_v = \frac{r_E}{r'_e + r_E}$$

$$A_v = \frac{500\Omega||10k\Omega}{28\Omega + 500\Omega||10k\Omega}$$

$$A_v = 0.9444$$

As mentioned, we need to include the effect of the 1 kΩ source impedance. This will create a voltage divider with the input impedance.

$$A_{divider} = \frac{Z_{in}}{Z_{in} + Z_{source}}$$

$$A_{divider} = \frac{9.03k\Omega}{9.03k\Omega + 1k\Omega}$$

$$A_{divider} = 0.9$$

$$A_{v(system)} = A_v \times A_{divider}$$

$$A_{v(system)} = 0.9444 \times 0.9$$

$$A_{v(system)} = 0.85$$

Finally, we get to the load voltage.

$$V_{load} = A_{v(system)} \times V_{in}$$

$$V_{load} = 0.85 \times 100mV$$

$$V_{load} = 85mV$$

At this point the question might be, "Why did we go to the trouble of building this circuit when we lost 15% of the input signal?" Well, consider what would have happened without the circuit. If we had connected the source directly to the load, the resulting 1 kΩ/500 Ω voltage divider would have dropped the load voltage to 33 mV. This circuit prevented that loss.

A HIGH IMPEDANCE SOURCE: THE GUITAR PICKUP

In Example 6.4.1, the source had an internal impedance of 1 kΩ, much higher than we would see with, say, a laboratory function generator (probably 50 Ω). Things could be much worse. Consider the electric guitar pickup. The job of a pickup is to transform the vibrations of the guitar strings into an electrical signal so that it can be amplified. It is commonly thought that a pickup is some form of microphone but this is not true.

A guitar pickup is little more than a magnet surrounded by numerous turns of fine wire, as shown in Figure 6.4.5. This particular pickup is for a bass guitar but the construction is similar for all types of guitars and basses.

Figure 6.4.5: Electric bass pickup (cover removed).

Here is it how works: The magnet creates a field around the guitar strings. Because the strings are steel, their reluctance is much less than the surrounding air, therefore, they distort or bend the magnetic field. When a string is plucked, the field moves back and forth along with it. As the field moves, the flux lines cut across the coil's wire and this action induces a current in the conductor in accordance with Faraday's Law of Induction. This current is then fed to the amplifier.

A typical guitar pickup consists of perhaps 5000 turns of very fine wire, 42 AWG being typical. 42 gauge copper wire has a resistance of around 1.6 Ω per foot so the DC resistance of the coil can be over 5 kΩ. Further, that many turns of wire around a magnet can produce a very large inductance, perhaps several henries, that is in series with this resistance. There is also distributed capacitance and cable capacitance in parallel that could be upwards of 1 nF. The result is a complex impedance with resonance effects, regions of which can be tens of kΩ in magnitude. What makes this more challenging is that because the impedance is a function of frequency, the voltage divider effect with the amplifier's input impedance also becomes a function of frequency. For example, the increasing impedance due to X_L will result in increasing attenuation with frequency. This is akin to turning down the treble on the amplifier. Generally, not a good result. How do we limit this effect? Simple. We make a circuit with a very, very high input impedance. How do we do that? Well, there are several ways, including the use of field effect transistors and operational amplifiers, but we can also obtain high input impedances through the use of a dual BJT configuration called the Darlington pair.

THE DARLINGTON PAIR

The Darlington pair was invented by Sidney Darlington, an American engineer. The configuration leads to a compound device with a very high β. Used properly, this can lead to amplifier circuits with very high input impedance. A Darlington pair is shown in Figure 6.4.6.

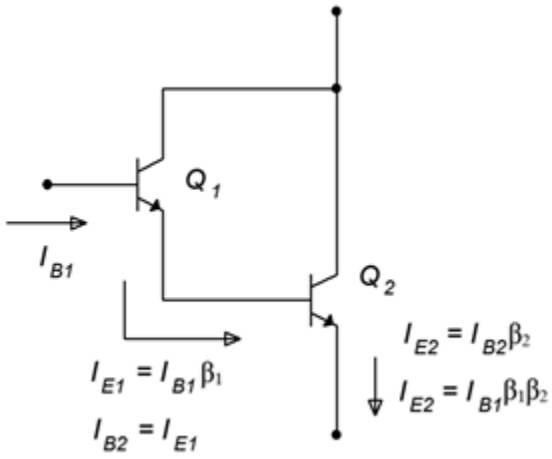


Figure 6.4.6 : Darlington pair.

The operation is as follows. The base current of the first transistor, Q_1, is multiplied by the β of Q_1 resulting in Q_1's emitter current. This current is fed into the base of the second transistor, Q_2, where it is multiplied by the β of Q_2 resulting in Q_2's emitter current. If we treat the pair as a single device, then the effective β of the pair is $\beta_1\beta_2$. Given typical values for β, the compound value can be in the vicinity of 5000 to 10,000. The functional downside to this arrangement is that V_{BE} is now doubled to 1.4 volts (for silicon) and the effective r'_e of the pair is doubled as well. These issues are minor when compared to the advantage of the huge current gain that can be obtained. The bottom line when using a Darlington pair is to treat it like an ordinary transistor except that it has a very large β and both V_{BE} and r'_e are doubled compared to the ordinary values.

Example 6.4.2

Determine the output voltage for the follower shown in Figure 6.4.7. Assume the input is 100 mV peak and the β for the Darlington pair is 10,000.

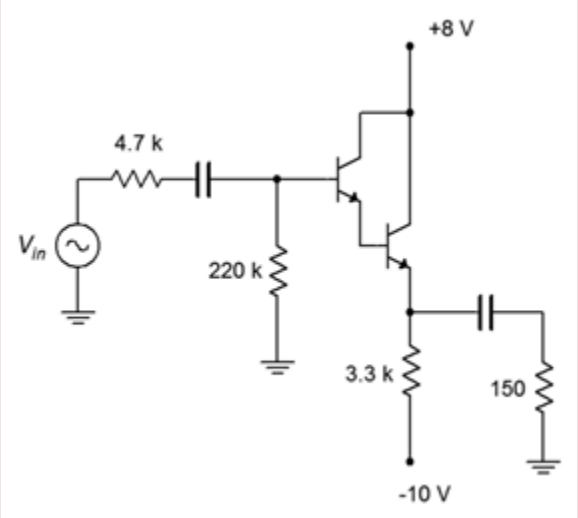


Figure 6.4.7: Schematic for Example 6.4.2.

The first thing that might look a little odd, at least compared to previous bias circuits, is that the base biasing resistor is so much larger than the emitter biasing resistor. Normally, this would lead to an unstable Q point but that's not a problem here. Because β is so large, RB can be much larger than normal and we'll still achieve good stability. In fact, we can still use the approximation that the base is at DC ground. This being true, the analysis proceeds as follows

$$I_C = \frac{|V_{EE}|V_{BE}}{R_E}$$

$$I_C = \frac{10V - 1.4V}{3.3k\Omega}$$

$$I_C = 2.61mA$$

$$r'_e = 2 \times \frac{26mV}{I_C}$$

$$r'_e = \frac{52mV}{2.61mA}$$

$$r'_e = 20\Omega$$

$$Z_{in(base)} = \beta(r'_e + r_E)$$

$$Z_{in(base)} = 10,000(20\Omega + 3.3k\Omega||150\Omega)$$

$$Z_{in(base)} = 1.63M\Omega$$

This value is in parallel with the base biasing resistor, creating the input impedance.

$$Z_{in} = R_B||Z_{in(base)}$$

$$Z_{in} = 220k\Omega||1.63M\Omega$$

$$Z_{in} = 194k\Omega$$

This is much higher than we have seen in previous circuits. The loaded gain from base to emitter is

$$A_v = \frac{r_E}{r'_e + r_E}$$

$$A_v = \frac{150\Omega||3.3k\Omega}{20\Omega + 150\Omega||3.3k\Omega}$$

$$A_v = 0.88$$

Now to include the effect of the 4.7 kΩ source impedance. This will create a voltage divider with the input impedance, minimal as it turns out.

$$A_{divider} = \frac{Z_{in}}{Z_{in} + Z_{source}}$$

$$A_{divider} = \frac{194k\Omega}{194k\Omega + 4.7k\Omega}$$

$$A_{divider} = 0.976$$

$$A_{v(system)} = A_v \times A_{divider}$$

$$A_{v(system)} = 0.88 \times 0.976$$

$$A_{v(system)} = 0.86$$

The load voltage is

$$V_{load} = A_{v(system)} \times V_{in}$$

$$V_{load} = 0.86 \times 100mV$$

$$V_{load} = 86mV$$

If we had connected the source directly to the load, the 4.7 kΩ/150 Ω divider would have squashed the applied signal into a shadow of its former size, leaving us with just 3 mV.

COMPUTER SIMULATION

To verify the results of Example 6.4.2, we'll run a transient analysis. The input schematic is shown in Figure 6.4.8.

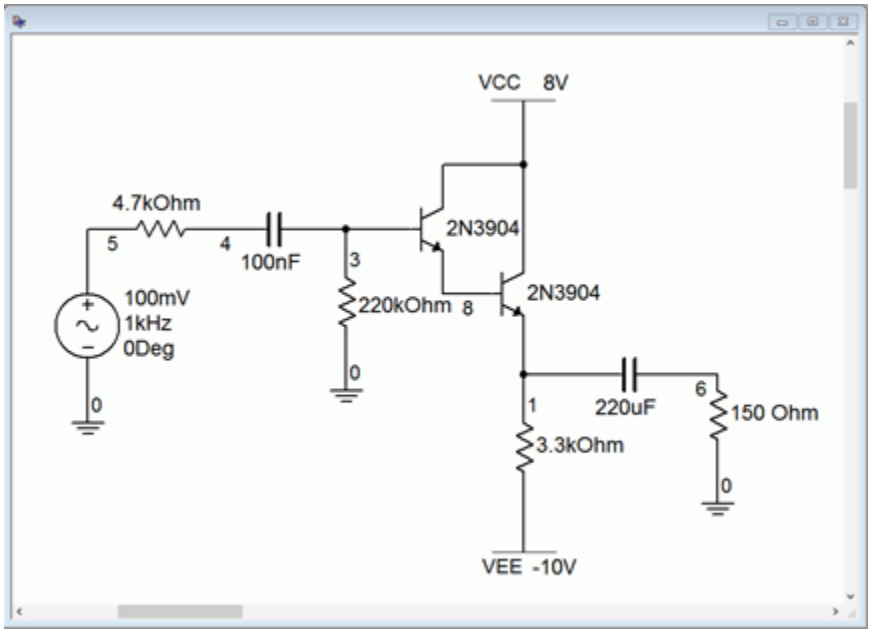


Figure 6.4.8: Simulation schematic for Darlington pair follower.

Of interest here will be the voltages at the source, base and load. As the input impedance/source impedance divider was 0.976, we expect 97.6 mV at node 4. At the output, node 6, we expect to see our final computed value of 86 mV. The output plot of the simulation is shown in Figure 6.4.9. The simulation concurs.

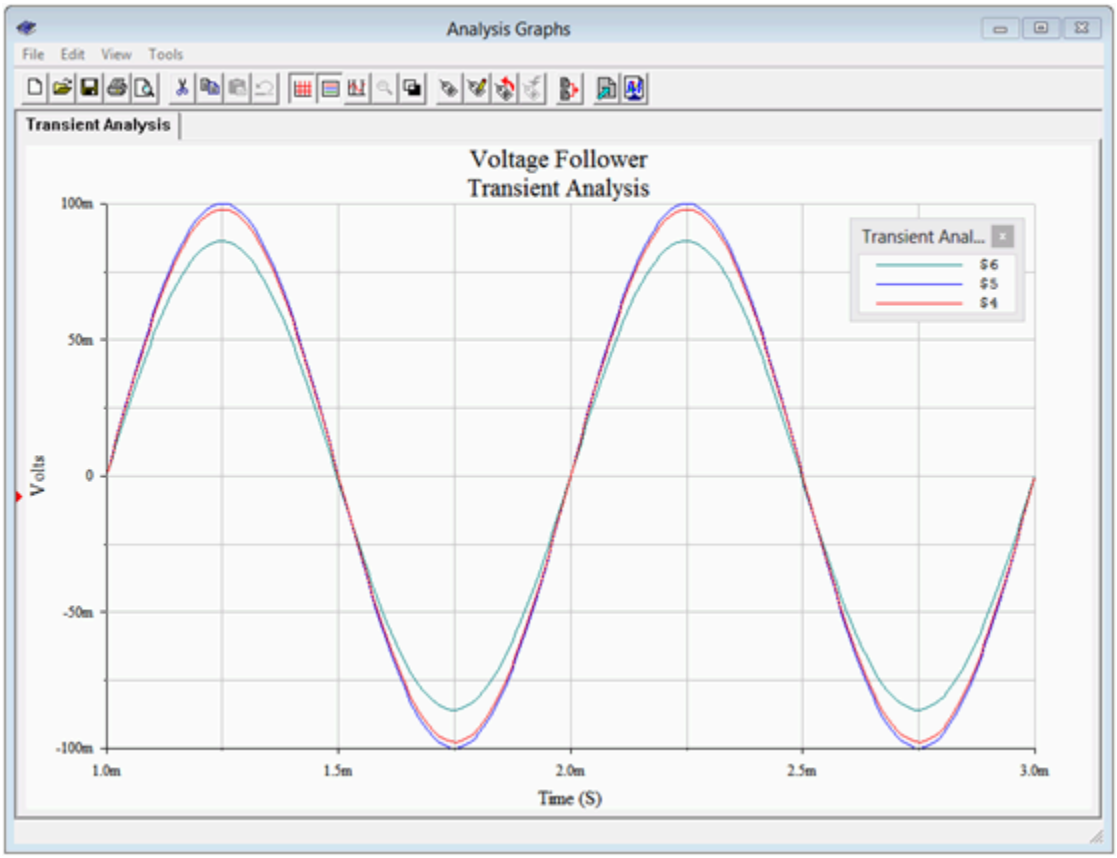


Figure 6.4.9 : Transient analysis for Darlington pair follower.

THE PHASE SPLITTER

Figure 6.4.10: A simple phase splitter.A phase splitter is a combination of a common emitter amplifier and a common collector follower using a single transistor. The purpose of the circuit is to produce two versions of the input signal: a buffered version identical to the input and an inverted version, both waves having the same amplitude. The circuit is used for differential line driver systems. This scheme helps to minimize outside noise and interference picked up by communications cables. There are other ways to create phase splitters, including using differential amplifiers or op amps, but this BJT-based version is a minimalist solution.[1] The basic circuit is shown in Figure 6.4.10.

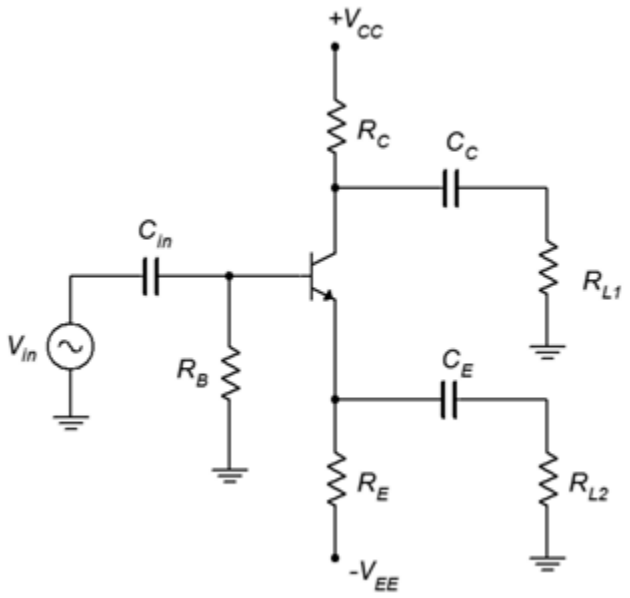


Figure 6.4.10: A simple phase splitter.

For proper operation, the circuit is largely symmetrical. That is, $R_L1 = R_L2$, $R_E = R_C$, and $C_C = C_E$. That being the case, the AC collector and emitter resistances will be equal ($r_C = r_E$). If we then look at the basic gain equations, we find that both loads will receive the same gain magnitude (just under unity), although R_{L1} will see the signal inverted.

$$A_v = \frac{r_C}{r'_e + r_E} \text{ Common emitter amplifier}$$

$$A_v = \frac{r_E}{r'_e + r_E} \text{ Common collector follower}$$

1. For details on alternate methods, see Fiore, J, Operational Amplifiers and Linear Integrated Circuits: Theory and Application, another free OER text.

6.5 COMMON BASE AMPLIFIER

The third and final prototype is the common base amplifier. In this configuration the input signal is applied to the emitter and the output is taken from the collector. The base terminal is at the common ground point. An example, using two-supply emitter bias, is shown in Figure 6.5.1. Note that because neither the input nor output is connected to the base, there is no need for a base resistor. Consequently, the base terminal is connected directly to ground.

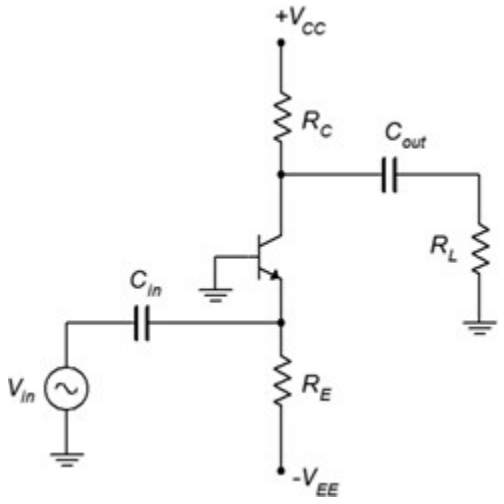


Figure 6.5.1: Common base amplifier.

Some people find that redrawing the schematic horizontally helps to visualize the signal flow. This version is shown in Figure 6.5.2.

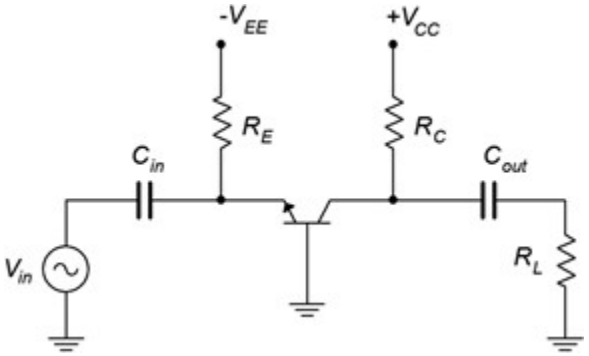


Figure 6.5.2: Common base amplifier redrawn.

One nice thing about the horizontal version is that when we make the AC equivalent, it becomes obvious that *RE* is in parallel with the input and *RC* is in parallel with the load. For the AC analysis we shall modify Figure 6.5.2 by substituting the BJT model for transistor, shorting the capacitors and taking the DC sources to AC ground. The result is shown in Figure 6.5.3.

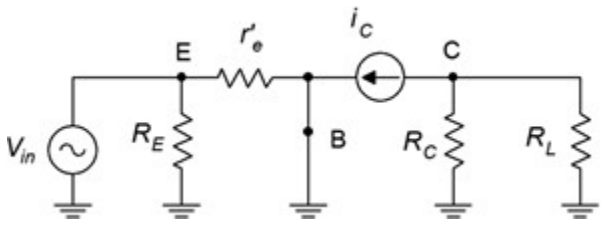


Figure 6.5.3: Common base amplifier with BJT model.

VOLTAGE GAIN

We begin with the basic definition of voltage gain and then expand using Ohm's law.

$$A_v = \frac{v_{out}}{v_{in}} = \frac{v_C}{v_E}$$

$$A_v = \frac{i_C r_C}{i_E r'_e}$$

$$A_v = \frac{r_C}{r'_e}$$

(6.5.1)

This equation is very similar to that of a non-swamped common emitter amplifier except that it does not invert the input signal. Therefore, gain potential is fairly high.

INPUT IMPEDANCE

The derivation for Z_{in} is obtained via direct inspection of the schematic.

$$Z_{in} = R_E || r'_e$$

(6.5.2)

r'_e normally dominates and thus we see that the common base configuration tends to have a low input impedance. For audio frequencies this can be an issue but it is less of a problem at higher frequencies as, generally speaking, system impedances need to be lower to avoid complications with capacitive effects.

OUTPUT IMPEDANCE

The derivation for Z_{out} unchanged compared to the common emitter configuration. The formula is repeated below for convenience.

$$Z_{out} \approx R_C$$

Example 6.5.1

For the amplifier shown in Figure 6.5.4, determine the voltage gain and input impedance.

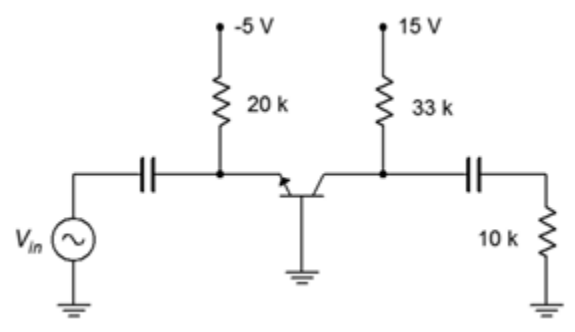


Figure 6.5.4: Schematic for Example 6.5.1.

$$I_C = \frac{V_{EE} V_{BE}}{R_E}$$

$$I_C = \frac{5V - 0.7V}{20k\Omega}$$

$$I_C = 0.215mA$$

$$r'_e = \frac{26mV}{I_C}$$

$$r'_e = \frac{26mV}{0.215mA}$$

$$r'_e = 121\Omega$$

$$Z_{in} = 120\Omega$$

$$A_v = \frac{r_C}{r'_e}$$

$$A_v = \frac{33k\Omega || 10k\Omega}{120\Omega}$$

$$A_v = 64$$

6.6 MULTI-STAGE AMPLIFIERS

In order to achieve a higher gain than we can obtain from a single stage, it is possible to cascade two or more stages. Different biasing types might be used along with a mix of AC configurations such as a common collector follower for the first stage that drives a common emitter voltage amplifier. A mix of NPN and PNP devices may also be present.

In general terms, each stage serves as the load for the preceding stage. That is, the Z_{in} of one stage is the R_L of the previous stage. The gains of the individual stages are then multiplied together to arrive at the system gain. The system input impedance is the input impedance of the first stage only. The source drives the first stage alone. The first stage, in turn, drives the second stage, and so on. Therefore the source only "sees" the first stage because it is the only stage to which it delivers current. In a similar fashion, the output impedance of the system is the *Zout* of the last stage. An example is shown in Figure 6.6.1.

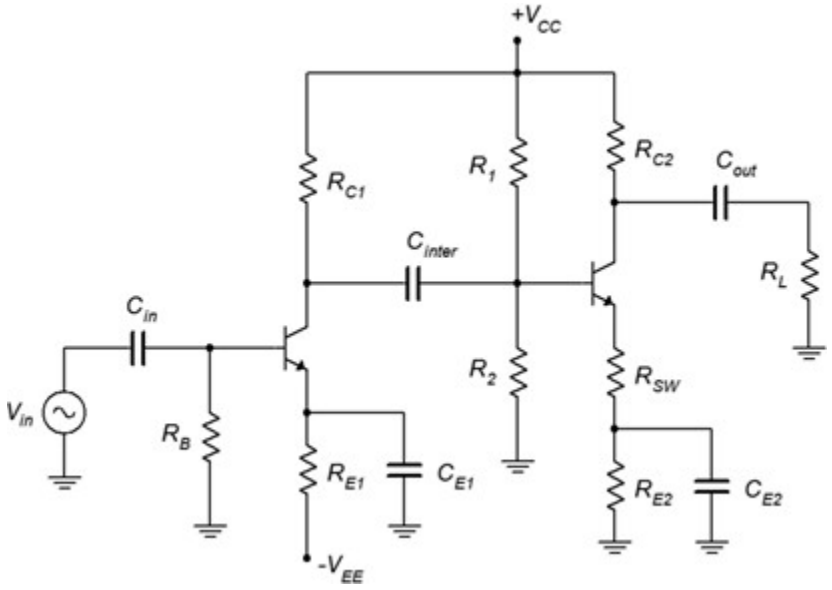


Figure 6.6.1: Two stage amplifier.

In this circuit, stage one is a non-swamped common emitter amplifier utilizing two supply emitter bias. Stage two is a swamped common emitter amplifier using voltage divider bias. As far as the DC analysis is concerned, these are two separate circuits. The inter-stage coupling capacitor, C_{inter} , prevents the DC potential at the collector of the first transistor from interfering with the bias established by R_1 and R_2 for transistor number two. For the AC computation, the first stage is analyzed in normal fashion except that its load resistance is comprised of $R_1 \| R_2 \| Z_{in-base2}$ (i.e., *Zin* of stage 2). The second stage is analyzed without changes and its gain is multiplied by the first stage's gain to arrive at the final gain for the pair. The input impedance of the system is $RB \|_{Zin-base1}$ (i.e., Z_{in} of stage 1).

It should be obvious that by cascading several stages it is possible to achieve very high system gains, even if each stage is heavily swamped in order to reduce distortion. For example, three swamped common emitter stages with voltage gains of just 10 each would produce a system voltage gain of 1000.

DIRECT COUPLING

With a little creativity, it is possible to create multi-stage designs that use fewer components but which achieve higher performance. One technique is to employ direct coupling of the stages. Direct coupling allows DC to flow from stage to stage. As such, it is possible to design an amplifier that has no lower frequency limit. An example is shown in Figure 6.6.2.

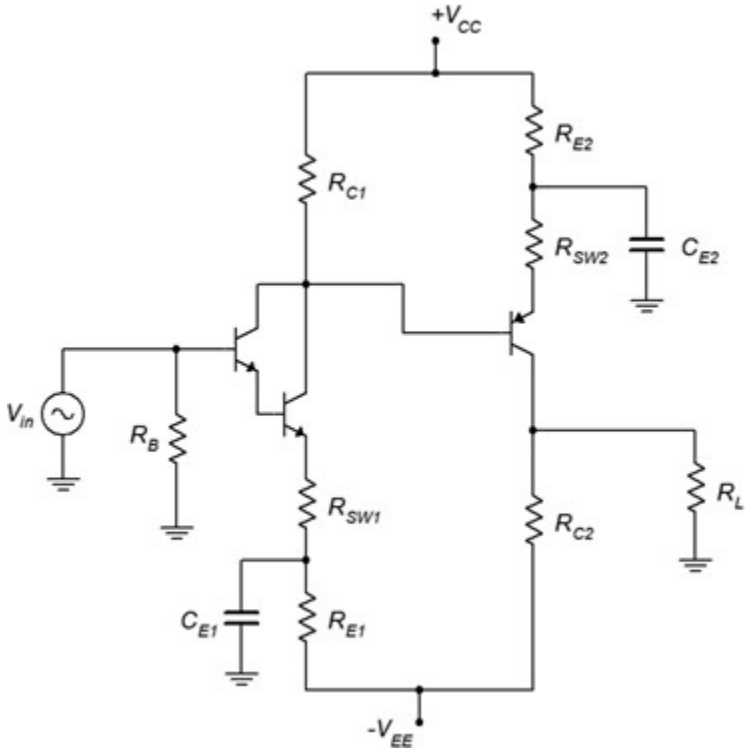


Figure 6.6.2: Direct coupled amplifier.

This two-stage amplifier uses no coupling capacitors nor does it rely on voltage divider resistors for the second stage.[1] Here is how it works: The first stage is a fairly ordinary swamped common emitter amplifier using two-supply emitter bias. It also uses a Darlington pair to maximize the input impedance. Because the base current is so low, the DC drop on R_B could be small enough to ignore so we may dispense with the input coupling capacitor. The DC potential at the collector of the Darlington is applied directly to the base of the second stage. This is used to set up the bias of the second stage via the stage two emitter resistors. This is precisely what we did with the circuit of Figure 7.3.5. The only difference is that here the base voltage is derived from the preceding stage instead of from a voltage divider. The computations for IC , r'_e and the like would proceed unchanged. In any event, this eliminates two biasing resistors and another coupling capacitor.

1. This circuit does use emitter bypass capacitors so the DC gain will be less than the AC gain. In that sense we might say that this amplifier is not fully DC coupled.

Note the use of the PNP device for the second stage. By using a PNP, its collector voltage must be less than its emitter voltage. As we're also using a bipolar power supply, we can eliminate the need for the final output coupling capacitor. All we need to do is set up the resistor values such that the drop across *RC*2 is the same as *VEE*. This will place the stage two DC collector voltage at 0 volts. If there's no DC voltage then there's nothing to block, and therefore no need for the coupling capacitor.

6.7 SUMMARY

A simplified AC model of the bipolar junction transistor consists of a controlled current source in the collector and a dynamic resistance in the emitter called r'_e. This resistance is a function of the DC bias current, I_C; the higher the biasing current, the lower the resistance. Fluctuations in this resistance can lead to waveform distortion. Swamping, also known as emitter degeneration, is a technique used to reduce distortion and stabilize gain. The basic idea is to add a fixed resistor in series with the emitter so as to buffer or "swamp out" the changes of r'_e.

There are three basic AC amplifier configurations: common emitter, common collector and common base. The common emitter configuration produces a voltage amplifier with high gain and intermediate input impedance. It also inverts the signal. Because it exhibits both voltage gain and current gain, it has a potential for high power gain. The common collector configuration is known as a follower because its output follows the input. It produces a non-inverting voltage gain of one and exhibits high input impedance and low output impedance. Therefore, it is useful as either an input buffer or as a final drive stage to a low impedance load. The common base configuration exhibits high non-inverting voltage gain. It has a low input impedance and a high output impedance.

The Darlington pair is a two-transistor configuration that may be treated as a single device. As such, it exhibits a doubling of both V_{BE} and r'_e, and a very large ß.

In order to achieve higher gains, multiple stages may be cascaded. Their gains multiply together to produce the combined system gain. The stages may be coupled through capacitors or via a capacitor-less direct coupling technique that can improve performance while reducing component count.

Review Questions

1. How does the AC BJT model compare with the DC model? What are the differences and similarities?
2. Explain how swamping reduces waveform distortion.
3. Compare and contrast common emitter, common collector and common base amplifiers in terms of voltage gain, power gain, input impedance and output impedance.
4. Why might biasing circuits that produce stable Q points be preferred for non-swamped amplifiers?
5. What is a phase splitter?
6. What are the advantages of direct coupling?
7. Give at least one example of a high internal impedance source.

6.8 EXERCISES

Unless otherwise specified, use ß=100.

ANALYSIS PROBLEMS

1. Determine the input and output impedances of the circuit of Figure 6.8.1.

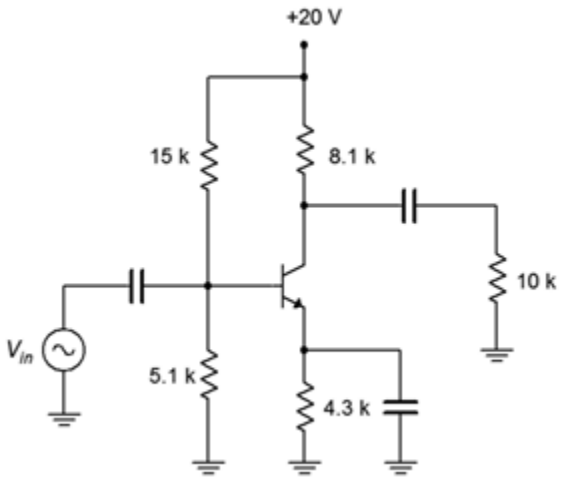


Figure 6.8.1

2. Determine the load voltage for the circuit of Figure 6.8.1 if V_{in} is 10 mV.
3. Determine Z_{in}, Z_{out}, and the load voltage for the circuit of Figure 6.8.2 if V_{in} is 70 mV.

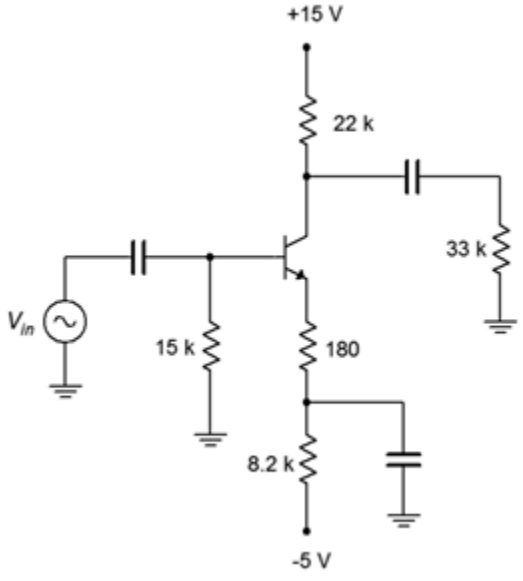


Figure 6.8.2

4. Determine Z_{in}, Z_{out}, and the load voltage for the circuit of Figure 6.8.3 if V_{in} is 50 mV.

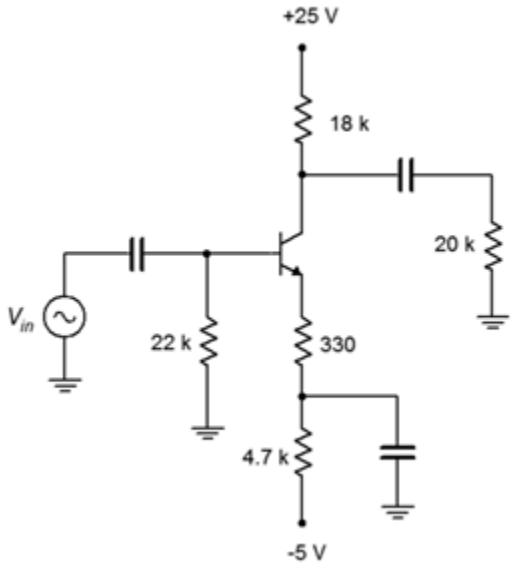


Figure 6.8.3

5. Determine Z_{in}, Z_{out}, and the load voltage for the circuit of Figure 6.8.4 if V_{in} is 25 mV.

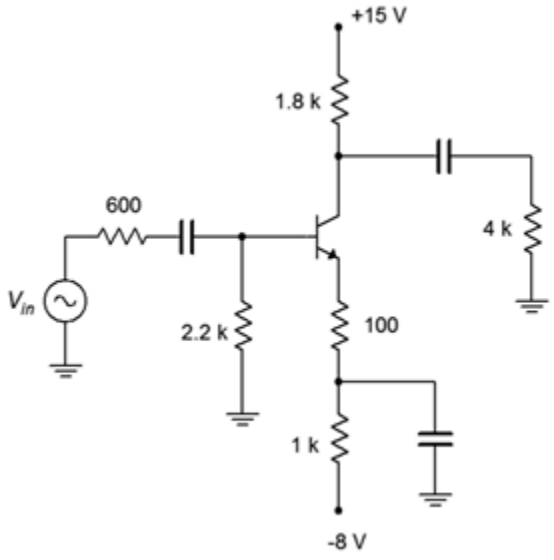


Figure 6.8.4

6. Determine Z_{in}, Z_{out}, and the load voltage for the circuit of Figure 6.8.5 if V_{in} is 30 mV.

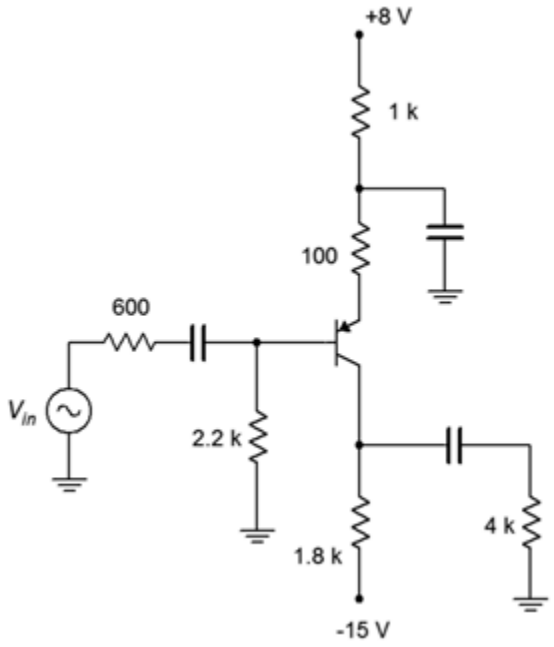


Figure 6.8.5

7. Determine Z_{in}, Z_{out}, and the load voltage for the circuit of Figure 6.8.6 if V_{in} is 60 mV.

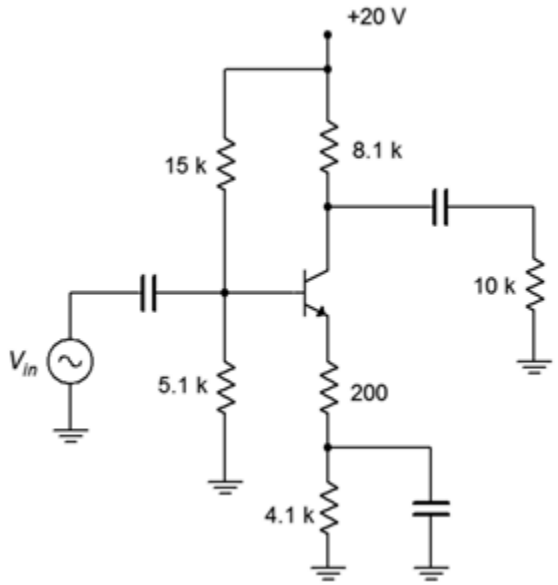


Figure 6.8.6

8. Determine Z_{in}, Z_{out}, and the load voltage for the circuit of Figure 6.8.7 if V_{in} is 150 mV.

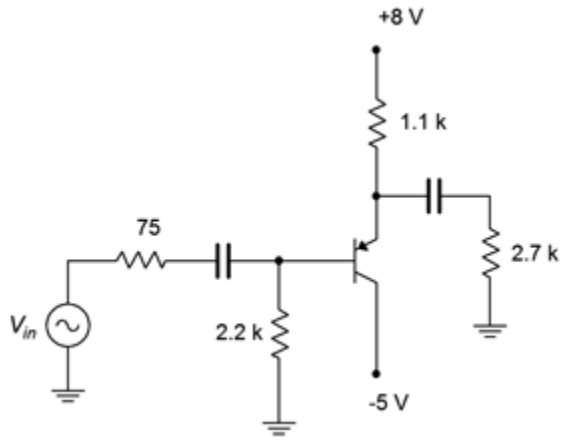


Figure 6.8.7

9. Determine Z_{in}, Z_{out}, and the load voltage for the circuit of Figure 6.8.8 if V_{in} is 200 mV.

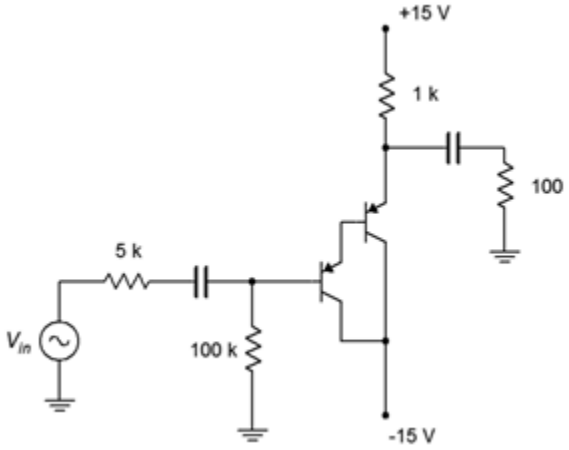


Figure 6.8.8

10. Determine Z_{in}, Z_{out}, and the load voltage for the circuit of Figure 6.8.9 if V_{in} is 250 mV.

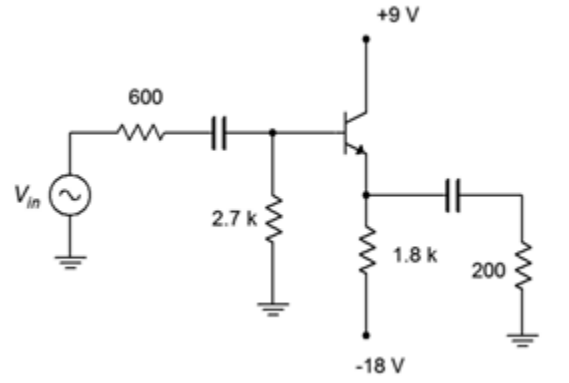


Figure 6.8.9

11. Determine Z_{in}, Z_{out}, and the load voltage for the circuit of Figure 6.8.10 if V_{in} is 300 mV.

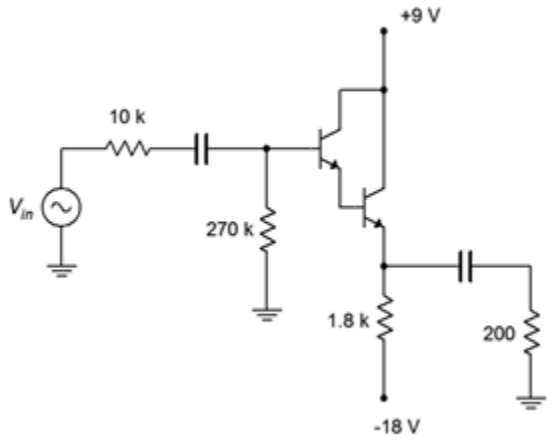


Figure 6.8.10

12. Determine Z_{in}, Z_{out}, and the load voltage for the circuit of Figure 6.8.11 if V_{in} is 50 mV.

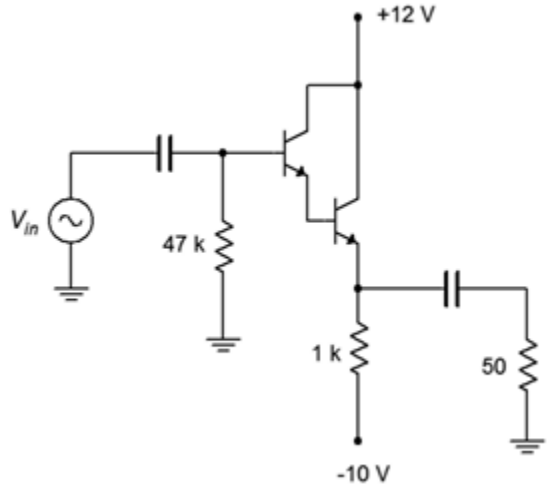


Figure 6.8.11

13. Determine Z_{in}, Z_{out}, and the load voltage for the circuit of Figure 6.8.12 if V_{in} is 2 mV.

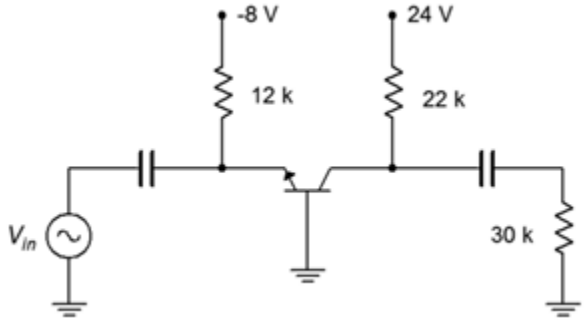


Figure 6.8.12

DESIGN PROBLEMS

14. Redesign the circuit of Figure 6.8.2 to halve the existing gain while keeping the Q point where

it is currently.

15. By using a Darlington pair, redesign the circuit of Figure 6.8.3 to double Z_{in}.
16. Redesign the circuit of Figure 6.8.3 so that it exhibits the same performance parameters but uses a PNP device.
17. Redesign the circuit of Figure 6.8.5 to double the existing gain while keeping the Q point where it is currently.
18. Redesign the circuit of Figure 6.8.7 so that it exhibits the same performance parameters but uses an NPN device.

CHALLENGE PROBLEMS

19. Determine the gain and input impedance for the circuit of Figure 6.8.13. V_{CC} = 20 V, *VEE* = –10 V, R_B = 18 kΩ, R_{E1}= 10 kΩ, R_{C1} = 12 kΩ, R_1 = 33 kΩ, R_2 = 15 kΩ, R_{E2} = 5.6 kΩ, R_{SW} = 400 kΩ, R_{C2} = 6.8 kΩ, R_L = 24 kΩ.

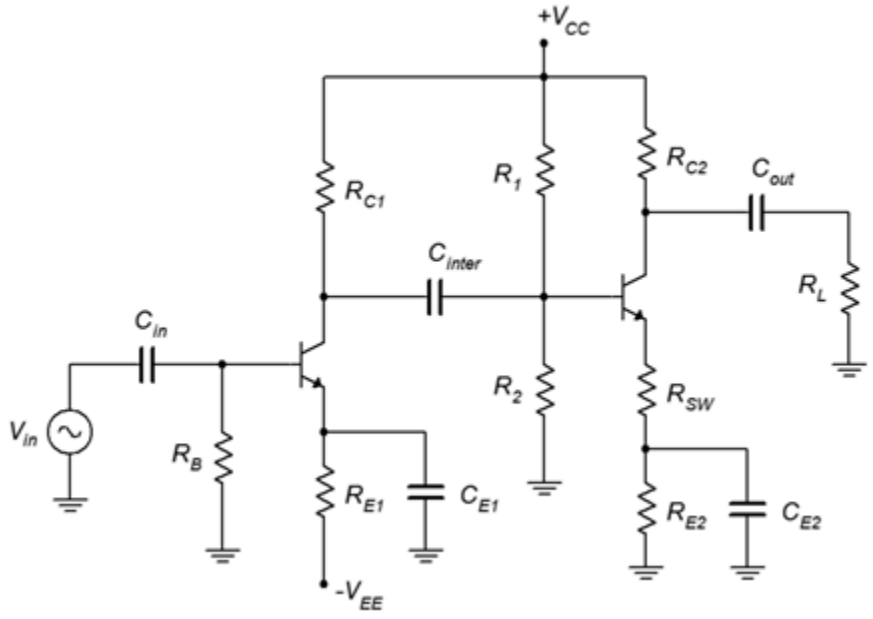


Figure 6.8.13

20. For the circuit of Figure 6.8.10, replace its load resistor with the circuit of Figure 6.8.6 and determine the combined gain and input impedance of the system.

COMPUTER SIMULATION PROBLEMS

21. Use a transient analysis to verify the load voltage of problem 3.
22. Use a transient analysis to verify the load voltage of problem 4.
23. Use a transient analysis to verify the load voltage of problem 8.
24. Consider the amplifier of Figure 6.8.1. Replace the 4.3 kΩ emitter resistor with a potentiometer of the same value. Connect the wiper arm to the emitter bypass capacitor. Run several transient analyses at different pot settings (0%, 25%, 50%, etc.). What can you conclude

from the results?

UNIT 7: BJT CLASS A POWER AMPLIFIERS

Learning Objectives

After completing this chapter, you should be able to:

- Define class A operation.
- Determine AC load lines for class A amplifier stages.
- Determine the compliance and maximum load power for class A amplifier circuits.
- Determine the efficiency and required device ratings for class A amplifier circuits.
- Describe the operation of a dynamic loudspeaker.
- Understand the need for heat sinks and other thermal management techniques.

7.1 INTRODUCTION

Now that we have examined BJT voltage amplifiers in terms of their gain, input impedance and output impedance, it is time that we extend the small signal analysis to larger signals. Of primary importance will be determination of the maximum output voltage swing, or compliance, along with maximum load power, device dissipation requirements and amplifier efficiency. To assist with this, we introduce the concept of the AC load line. In general, we will not concern ourselves with input impedance, voltage gain or even β_{ac}. Indeed, we shall simply consider β_{ac} as a source of distortion. Most power amplifiers are configured as voltage followers so we will focus largely on those.

7.2 AMPLIFIER CLASSES

There are several classes of amplifier operation. The class of an amplifier has nothing to do with the fidelity or quality of the amplifier. Rather, the class indicates the fundamental operational principle of the circuit. In general, as the class letter increases, the designs become more complicated but also more efficient. For audio and other linear applications, classes A, B and D are relatively common these days. Class C is largely relegated to high power telecommunications while classes G and H are essentially variations of class B.

The definition of class A is that signal current in the collector flows 360° out of the cycle. In other words, it flows for the entire cycle without interruption. All of the amplifiers that were presented in the prior chapter are class A designs. In class B, I_C flows for just 180°, and for class D, I_C is discontinuous; the transistor is used as a switch. Class B and D designs are examined in later chapters.

7.3 CLASS A OPERATION AND LOAD LINES

The signal current in the class A amplifier flows continuously throughout the entire cycle of the waveform. Ultimately, we would like to known just how large this signal can be before it is limited and grossly distorted. To do so, we need to examine the AC equivalent of the amplifier. A generic AC equivalent is shown in Figure 7.3.1. This includes both AC collector and emitter resistances so it can be used for either swamped or unswamped common emitter amplifiers or for emitter followers. If one of the resistances is not used (for example, r_C in a follower), we can just substitute a value of zero for it.

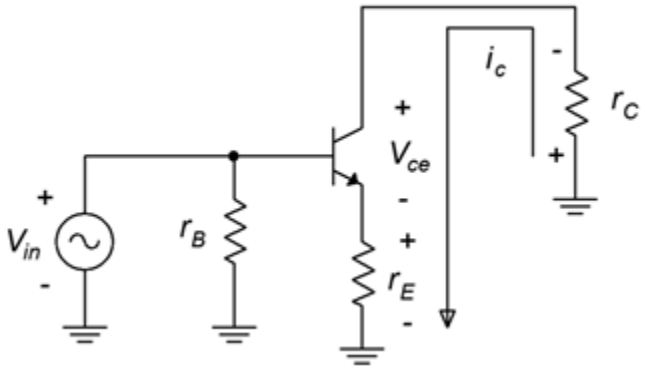


Figure 7.3.1: AC equivalent circuit.

The voltage polarities and current direction are shown for a positive input voltage. To determine the maximum load voltage swing (compliance), we will need to construct an AC load line as shown in Figure 7.3.2.

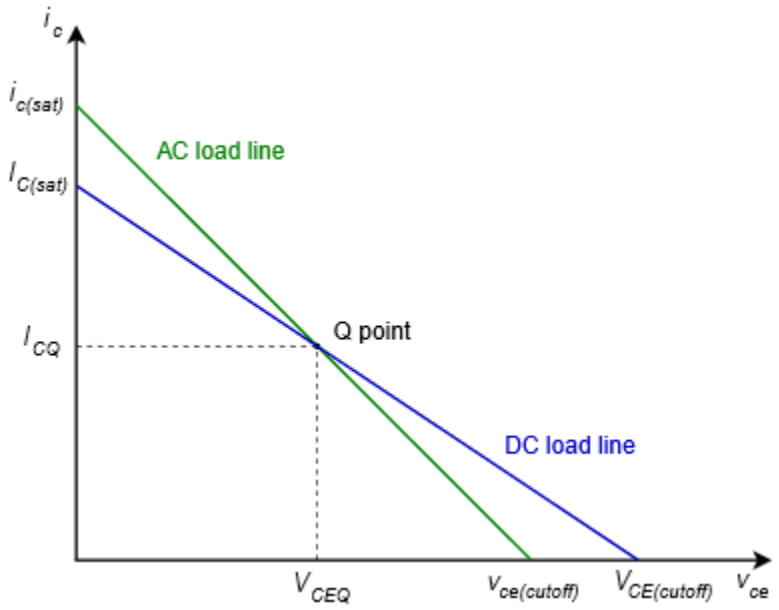


Figure 7.3.2: AC and DC load lines.

The AC load line is similar to the DC load line that was used for analyzing biasing circuits. As in the DC version, there will be a cutoff voltage, $V_{CE(cutoff)}$, and a saturation current, $I_{C(sat)}$. The AC and DC load lines normally are not the same, however, they must share one point in common, and that's the Q point. Usually, the slope of the AC load line is steeper than that of the DC load line. This is because the AC resistance tends to be less than the DC resistance due to loading and capacitor bypassing. Consequently, $V_{CE(cutoff)}$ tends to be smaller than $V_{CE(cutoff)}$ and $I_{C(sat)}$ tends to be larger than $I_{C(sat)}$.

I_{CQ} and a no-signal transistor voltage of V_{CEQ}. As the input signal grows, i_C increases. The effect of this is to increase the voltage drops across r_E and r_C due to Ohm's law. This, in turn, forces V_{CE} to decrease due to KVL. The collector current can only increase to the point where V_{CE} drops to 0 V. This is a maximum increase of $V_{CEQ}/(r_C + r_E)$. Therefore

$$I_{C(sat)} = I_{CQ} + \frac{V_{CEQ}}{r_E + r_C}$$

(7.3.1)

In terms of cutoff voltage, the transistor starts with V_{CEQ} and I_{CQ}. The largest v_{CE} increase that can occur is if the current falls to zero. Then, all of the potential originally developed across r_E and r_C by I_{CQ} must be absorbed by the transistor. Therefore

$$V_{CE(cutoff)} = V_{CEQ} + I_{CQ}(r_E + r_C)$$

(7.3.2)

There are three possible ways this can be configured: Q point closer to saturation, Q point closer to cutoff, or Q point centered on the AC load line. Let's first consider the Q point closer to saturation. This is shown in Figure 7.3.3.

Here we have plotted the input voltage in red and drawn the corresponding collector current and collector-emitter voltage in blue. It is apparent that as the input signal increases, eventually, the output signal is limited at zero for V_{CE} and at $I_C(sat)$ for I_C . The two blue waveforms are severely clipped and distorted. The largest unclipped peak voltage swing is V_{CEQ} and the largest peak current swing is $I_C(sat) - I_{CQ}$, or more conveniently, $V_{CEQ}/(r_E + r_C)$.

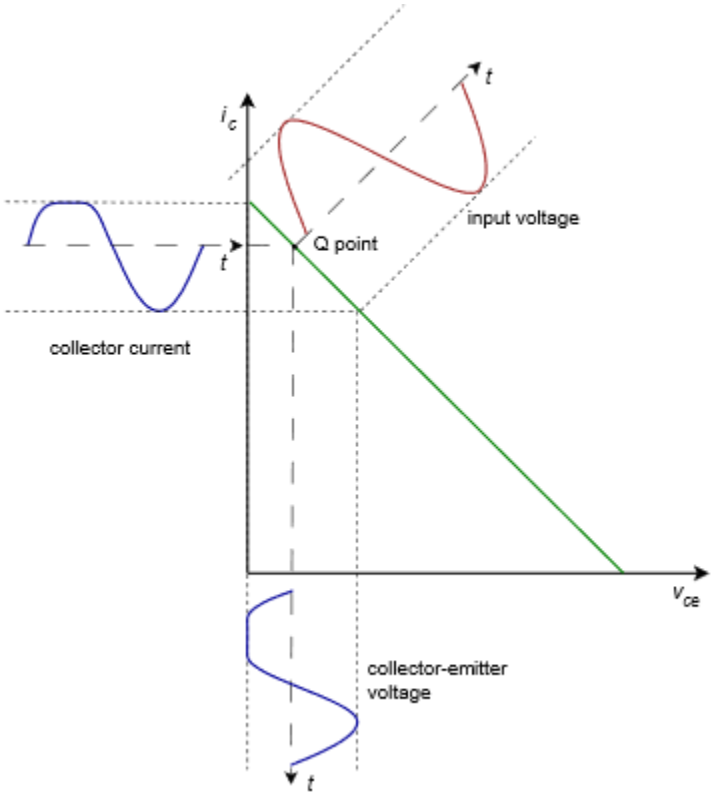


Figure 7.3.3: AC load line, Q point closer to saturation.

If we shift the Q point toward cutoff, we solve the saturation clipping problem but now we have a new problem, as illustrated in Figure 7.3.4. It should come as no surprise that we now have cutoff clipping.

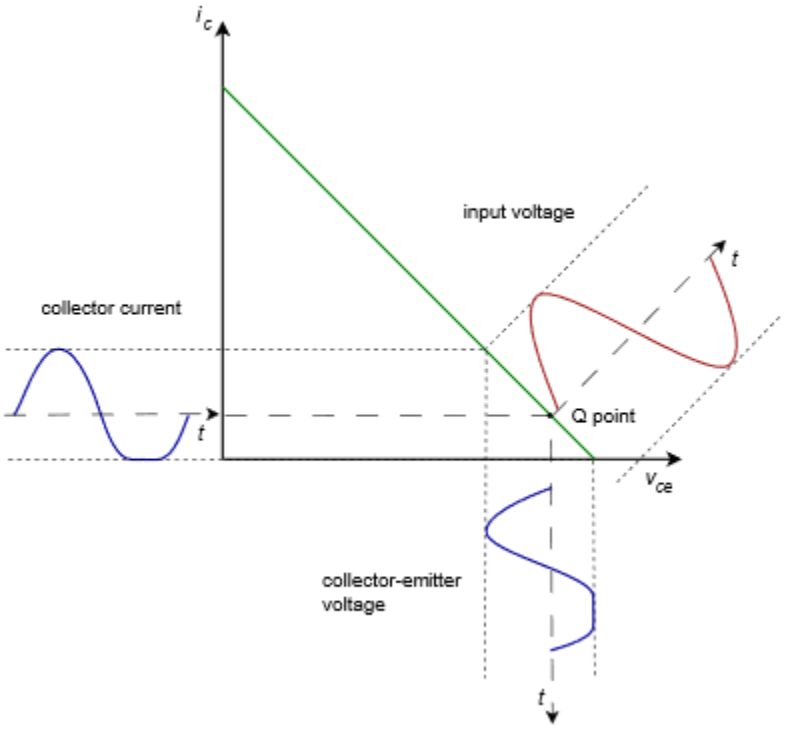


Figure 7.3.4: AC load line, Q point closer to cutoff.

In this version the largest unclipped peak voltage swing is $V_{CE}(cutoff) - V_{CEQ}$ (or alternately, $I_{CQ}(r_E + r_C)$) and the largest peak current swing is I_{CQ}. What's important here is that the waveform has been clipped. It doesn't really matter which side has been clipped, either way it's gross distortion. Eventually, every amplifier will have a limit but we will be able to produce the largest unclipped voltage swing if the Q point is centered on the AC load line. This is shown in Figure 7.3.5.

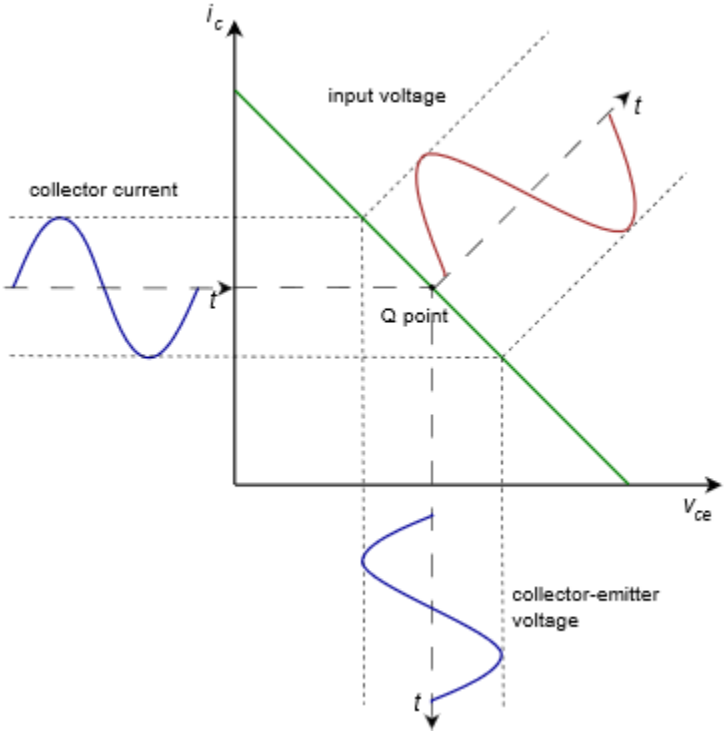


Figure 7.3.5: AC load line, centered Q point.

V_{CEQ} and the largest unclipped peak current swing is I_{CQ}. By examining equations 7.3.1 and 7.3.2 it is apparent that in order to achieve a centered Q point on the AC load line, the following must be true:

$$\frac{V_{CEQ}}{I_{CQ}} = r_E + r_C$$

(7.3.3)

Of course, while it is useful to determine the maximum voltage across the transistor, it is more important to determine the maximum voltage across the load. Looking back at the circuit of Figure 7.3.1, most times the maximum load voltage (i.e., the compliance) will equal the maximum transistor voltage. This will be the case in voltage followers and unswamped amplifiers. The only time there will be a noticeable reduction is with very heavily swamped amplifiers. In this case the compliance will be reduced by the voltage divider between the load and swamping resistors. For example, a swamped amplifier with a voltage gain of 4 would lose about 20% of the maximum swing. Swamping has to be very heavy resulting in very low gains before appreciable signal is lost.

Thus we arrive at the following general rule:

$$\text{Peak compliance is the smaller of } V_{CEQ} \text{ or } I_{CQ}(r_E + r_C)$$

Knowing the compliance, the maximum load power may be determined using power law. Power is determined using RMS values, so the peak compliance will need to be divided by √2 (or multiplied by 0.707) before continuing.

$$P_{load(max)} = \frac{Compliance_{RMS}{}^2}{R_L}$$

(7.3.5)

There is something important to note about this equation. It uses the load resistance value, not the total AC effective value (i.e., not r_L which is R_L in parallel with a biasing resistor). If r_L was used, we'd be calculating the power in the load plus the power in the biasing resistor.

We would also like to determine the maximum power dissipated by the transistor. Because the transistor's current and voltage are fluctuating with the input signal, we need to determine the magnitude of the load voltage that produces maximum power in the transistor. Intuitively, we might guess that this occurs at maximum load power but it turns out that this guess is incorrect. Under no- signal conditions the transistor is operating statically at the Q point. Therefore, quiescent power dissipation is

$$P_{DQ} = V_{CEQ} I_{CQ}$$

(7.3.6)

In contrast, at full load for a centered Q point, we have

$$V_{CE} = V_{CEQ}(1 - \sin 2\pi ft)$$

$$I_C = I_{CQ}(1 + \sin 2\pi ft)$$

$$P_D = v_{CE} i_C$$

$$P_D = V_{CEQ}(1 \sin 2\pi ft) \times I_{CQ}(1 + \sin 2\pi ft)$$

$$P_D = V_{CEQ} I_{CQ}(1 - \sin^2 2\pi ft)$$

$$P_D = V_{CEQ} I_{CQ}(.5 + .5 \cos 4\pi ft)$$

$$P_D = \frac{P_{DQ}}{2} + \frac{P_{DQ}}{2} \cos 4\pi ft$$

(7.3.7)

The first term of Equation 7.3.7 is a fixed offset while the second term is a sinusoid at twice the signal frequency. Because the peak amplitude of this sinusoid is the same as the fixed offset, the average over time is simply the offset value. These waveforms are illustrated in Figure 7.3.6.

Figure 7.3.6: Transistor power dissipation at full load power.

I_{CQ}. That current times the supply voltage yields the supplied power. What's happening is that as the signal increases in amplitude, more and more of the power dissipated by the transistor is shifted to the

load. At maximum load swing, both the transistor and the load will be dissipating $P_{DQ}/2$. As strange as it might seem, if you want to keep the output transistor of a class A amplifier cool, don't turn the volume down, turn it up.

The foregoing implies that class A designs are not power efficient. This is indeed the situation. As we have just seen, the best case maximum load power will be one half of P_{DQ}, assuming a centered Q point (non-centered will be worse). To achieve this swing, the power supply will have to be at least twice as large as V_{CEQ} because it has to cover the peak-to-peak swing, while V_{CEQ} represents the peak swing for a centered Q point.[1] In any event, the best case efficiency turns out to be dismal, as follows.

$$\eta = \frac{P_{out}}{P_{in}} = \frac{P_{load}}{P_{DC}}$$

$$\eta = \frac{P_{DQ}/2}{2V_{CEQ}I_{CQ}}$$

$$\eta = \frac{P_{DQ}/2}{2P_{DQ}}$$

$$\eta = 25\%$$

This represents the maximum or best case efficiency for an *RC* coupled class A amplifier. It may be considerably less depending on precisely how it is biased. This, truly, is the Achilles heel of the class A topology: it is wasteful. It draws full power from the supply regardless if signal is present and, at best, will translate only one quarter of that power into useful load power. At the same time, the power dissipation of the transistor will need to be at least twice that of the delivered load power, and might need to be much greater. Why use it then? To its advantage, it is a relatively simple design so if large output powers are not needed, it can

Example 7.3.1

For the amplifier shown in Figure 7.3.7, determine the compliance, maximum load power, worst case transistor dissipation and efficiency.

1. This is the case if the AC and DC load lines are identical. This is atypical. Consequently, the power supply will tend to be larger than twice V_{CEQ} which makes the situation even worse.

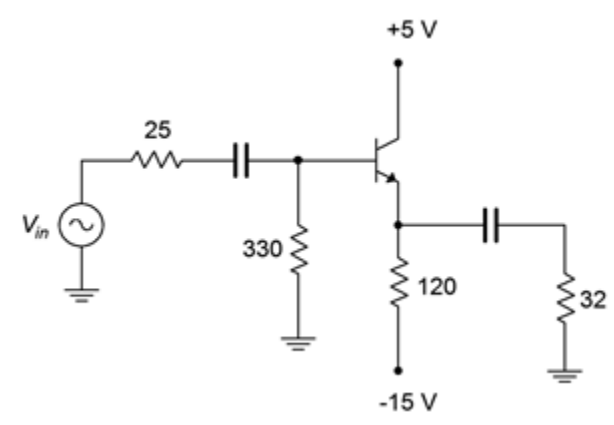


Figure 7.3.7: Schematic for Example 7.3.1.

$$I_{CQ} = \frac{|V_{EE}| - V_{BE}}{R_E}$$

$$I_{CQ} = \frac{15V - 0.7V}{120\Omega}$$

$$I_{CQ} = 119mA$$

By inspection, V_{CEQ} = 5.7 V. The AC cutoff voltage is

$$V_{CE(cutoff)} = V_{CEQ} + I_{CQ}(r_C + r_E)$$

$$V_{CE(cutoff)} = 5.7V + 119mA(0 + 120\Omega||32\Omega)$$

$$V_{CE(cutoff)} = 5.7V + 119mA(25.3\Omega)$$

$$V_{CE(cutoff)} = 5.7V + 3V$$

$$V_{CE(cutoff)} = 8.7V$$

The smaller of V_{CEQ} and $I_{CQ}(r_C + r_E)$ is the peak compliance, so

$$compliance = 3Vpeak$$

Given the compliance, we can use power law to find the load power

$$P_{load(max)} = \frac{Compliance_{RMS}{}^2}{R_L}$$

$$P_{load(max)} = \frac{(.707 \times 3V)^2}{32\Omega}$$

$$P_{load(max)} = 141mW$$

This is not a lot of power for something like a loudspeaker but is a fair amount to drive something like a pair of headphones.

The transistor's worst case power dissipation is

$$P_{D(max)} = P_{DQ} = I_{CQ}V_{CEQ}$$

$$P_{D(max)} = 119mA \times 5.7V$$

$$P_{D(max)} = 678mW$$

The supplied circuit power is the average current draw times the total supplied voltage differential

$$P_{DC} = I_{CQ}(V_{CC}V_{EE})$$

$$P_{DC} = 119mA \times 20V$$

$$P_{DC} = 2.38W$$

The efficiency is the ratio of maximum load power to supplied DC power

$$\eta = \frac{P_{load(max)}}{P_{DC}}$$

$$\eta = \frac{141mW}{2.38W}$$

$$\eta = 5.9\%$$

This is much worse than the theoretical best case. This is due, at least in part, to the fact that the Q point is not centered on the AC load line.

To complete the analysis, note that the transistor's breakdown rating (BV_{CEO}) should be at least as large as $V_{CE(cutoff)}$ (8.7 volts), and the maximum current rating should be at least as large as $I_{C(sat)}$ (119 mA+5.7 V/ 25.3 Ω = 344 mA).

COMPUTER SIMULATION

A computer simulation of a class A emitter follower using a Darlington pair is examined next. Of primary interest here is the verification of the output compliance so a transient analysis will be used. The simulator schematic is shown in Figure 7.3.8.

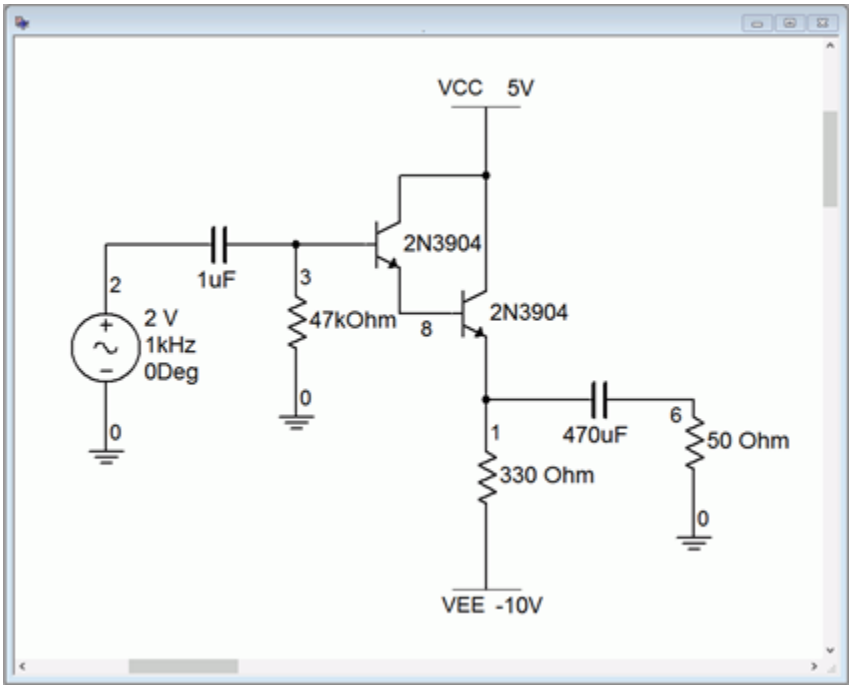


Figure 7.3.8: Class A follower in simulator.

We can make a few quick computations to determine the compliance. First, we find the collector Q point current.

$$I_{CQ} = \frac{|V_{EE}| - V_{BE}}{R_E}$$

$$I_{CQ} = \frac{10V - 1.4V}{330\Omega}$$

$$I_{CQ} = 26mA$$

By inspection, the emitter is two base-emitter junction potentials below ground, or –1.4 V. As the collectors are tied to V_{CC} , this means that V_{CEQ} = 6.4 V. The other half of the swing, from V_{CEQ} to *vCE*(*cutoff*) is

$$V_{CE(cutoff)} - V_{CEQ} = I_{CQ}(r_C + r_E)$$

$$V_{CE(cutoff)} - V_{CEQ} = 26mA(0 + 330\Omega||50\Omega)$$

$$V_{CE(cutoff)} - V_{CEQ} = 26mA(43.4\Omega)$$

$$V_{CE(cutoff)} - V_{CEQ} = 1.13V$$

The Q point is not centered and is closer to cutoff. This means that the amplifier will produce cutoff clipping around 1.1 volts and saturation clipping around 6 volts. In other words, there is more room for the current to swing up to saturation than to swing down to zero. As this is the current flowing

through the load and we have a non-inverting follower, we expect to see the load voltage echo this. That is, the negative portion of the load voltage should clip before the positive portion.

The transient analysis results are shown in Figure 7.3.9. A two volt peak input signal is applied (blue trace). The negative portion of the load voltage clips at approximately 1.1 volts as expected (red trace). The input signal is not large enough to cause saturation clipping. This was done on purpose to verify the voltage gain of the follower. It should be very close to unity. In fact, the trace shows that the gain is around 0.95 or so.

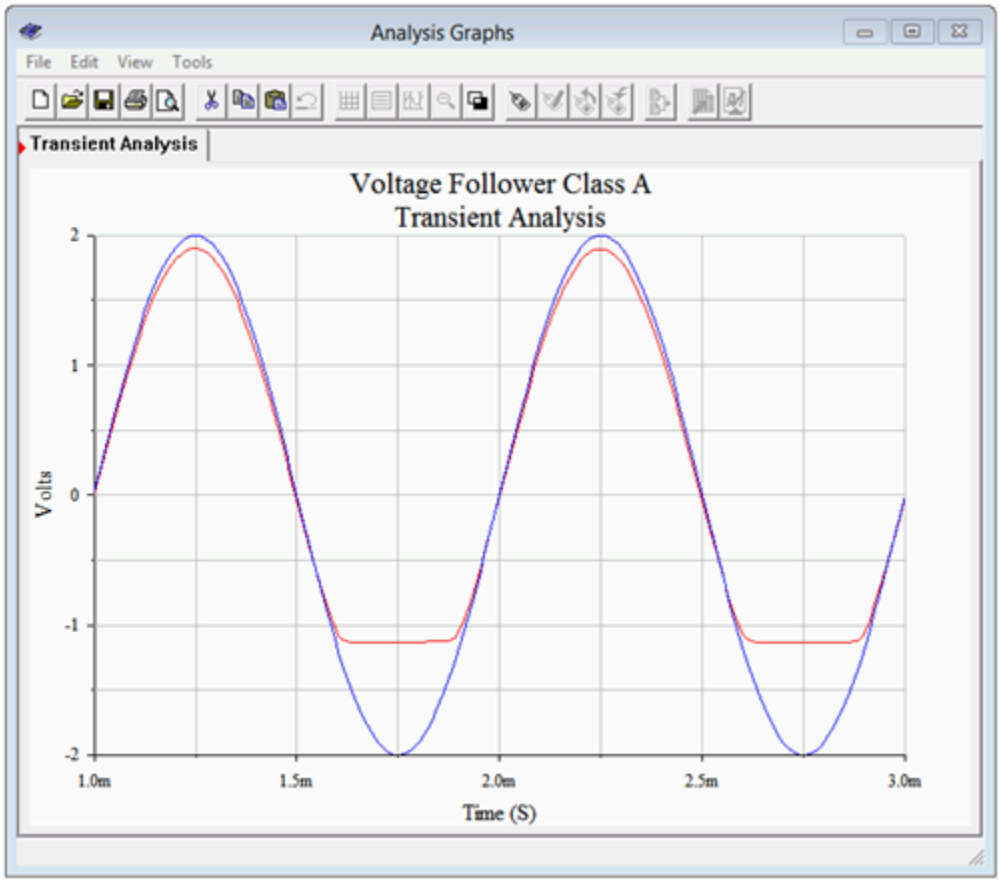


Figure 7.3.9: Class A follower transient analysis.

If this had been a voltage amplifier instead of a follower, these waveforms would appear flipped vertically. To verify this, the circuit is modified to produce a voltage amplifier with a gain of approximately one. This is achieved by moving the load to the collector and adding a 330 Ω biasing resistor. This will result in the same AC load impedance. To maintain a similar V_{CEQ}, V_{CC} is raised by 10 volts. Finally, the original 330 Ω emitter biasing resistor is split in two: 287 Ω and 43 Ω. This will yield the same I_{CQ} and achieve a voltage gain of unity. As a result, we expect to see clipping at approximately 1.1 volts on the positive portion. The modified circuit is shown in Figure 7.3.10and the resulting transient simulation in Figure 7.3.11.

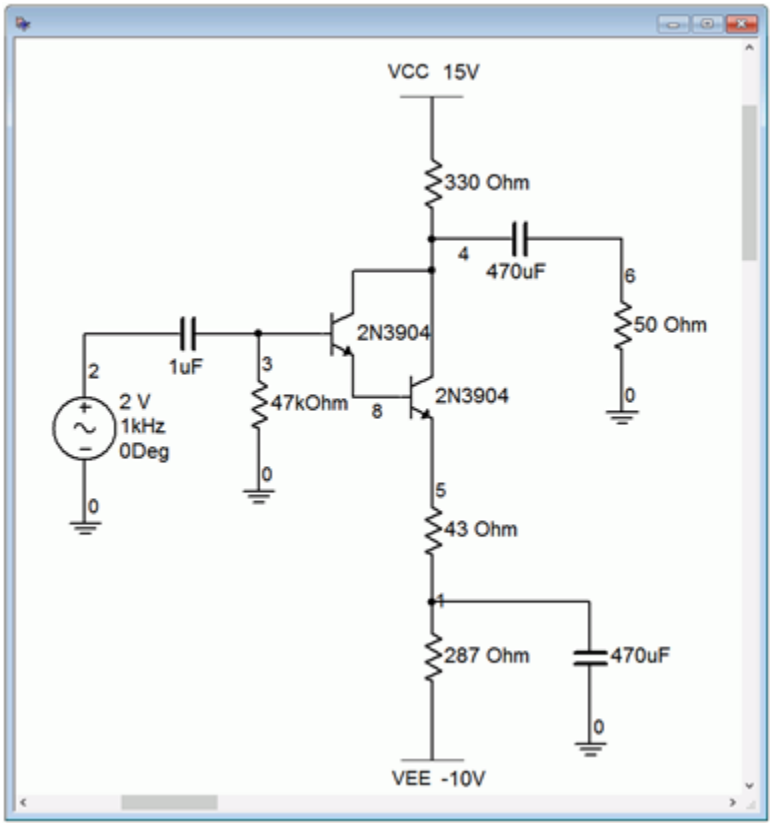


Figure 7.3.10: Class A amplifier in simulator.

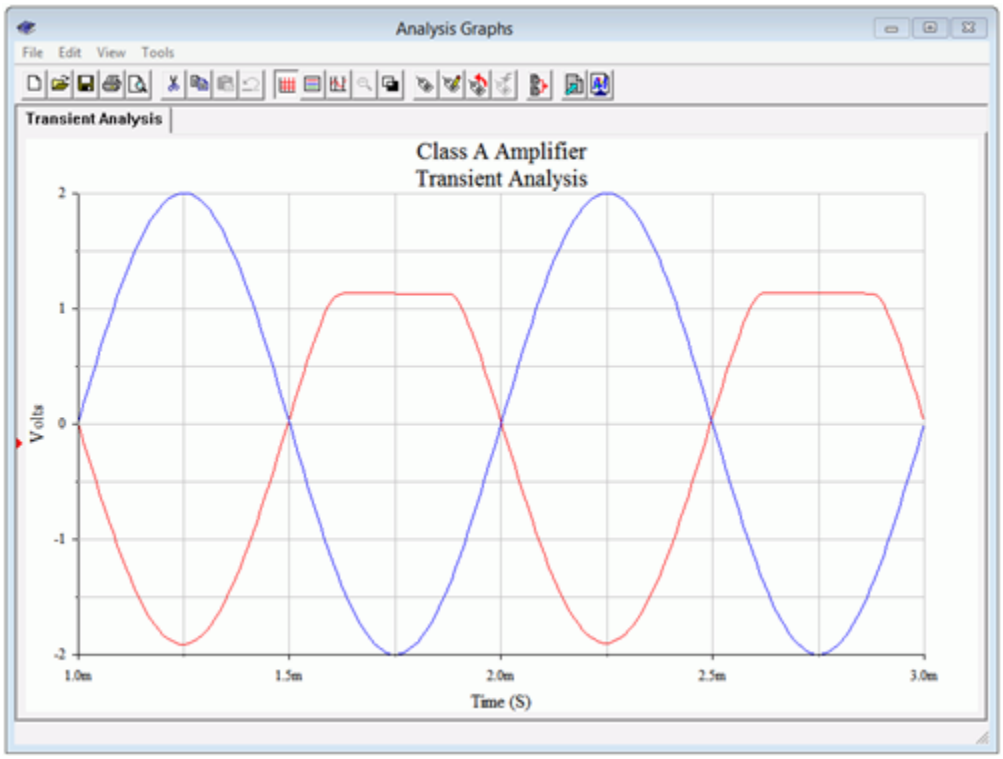


Figure 7.3.11: Class A amplifier transient analysis.

One final item of interest regarding the simulations: If the input level is increased in an attempt to see clipping on the other half of the waveform, something strange happens. At first it will appear as

though it never clips. A careful examination reveals something different, though. Given the values in these circuits, they will exhibit a certain amount of clamping action (clamping was presented in Chapter 3). This will cause the waveform to shift. If you inspect the peak-to-peak value, it will be close to the value of $V_{CE(cutoff)}$. It will be a little less due to the fact that, particularly for a Darlington pair, $V_{CE(sat)}$ is not 0 V.

7.4 LOUDSPEAKERS

One of the more common loads for amplifiers is a loudspeaker. It makes sense then to look at how they are constructed and note anything interesting or peculiar as far as their electrical characteristics are concerned. The most common form of loudspeaker is the dynamic loudspeaker.[1] All dynamic loudspeakers share certain common elements regardless of size or acoustic output capability. A cutaway view of a low frequency driver is shown in Figure 7.4.1.

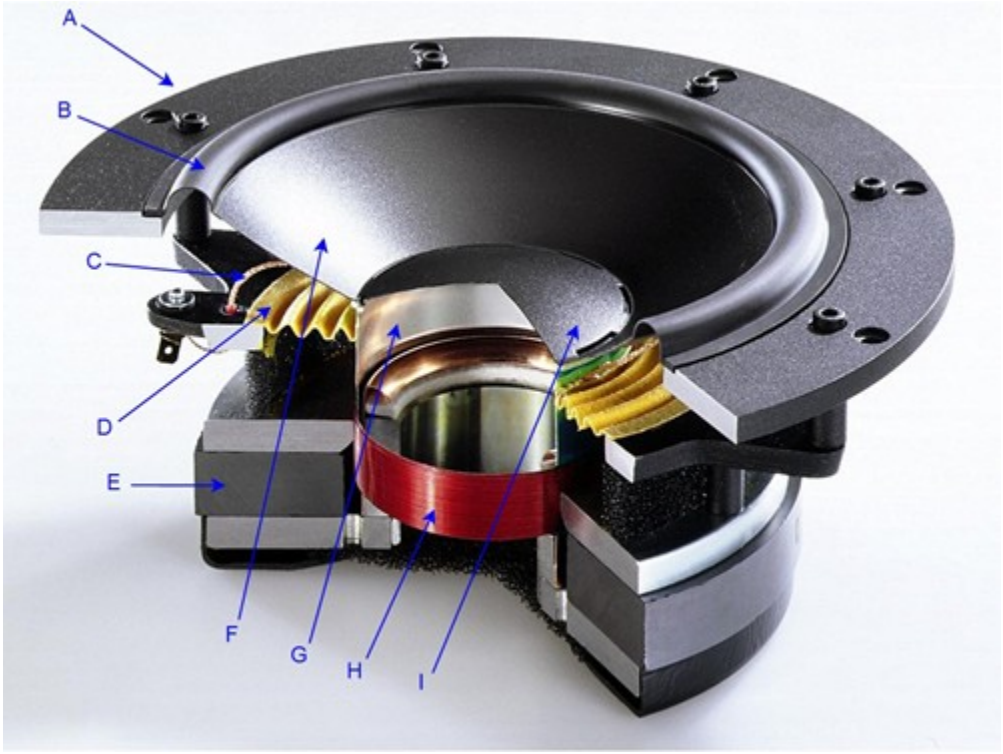


Figure 7.4.1: Dynamic loudspeaker. A. Frame B. Suspension C. Lead wire D. Spider E. Magnet F. Diaphragm G. Voice coil former H. Voice coil I. Dust cap Image courtesy of Audio Technology

The idea behind its operation is magnetic repulsion and attraction. The heart of the unit is the voice coil (H). This is a coil of magnet wire wound around a former (G) that typically is made of aluminum or some other high temperature material. The voice coil might be a single layer of edge-wound ribbon wire or perhaps several layers of ordinary round wire. Depending on the design, the voice coil might be anywhere from a fraction of an inch to several inches in diameter. The coil ends are connected to flexible lead wires (C) that terminate on the loudspeaker frame (A). Ultimately, that's what the amplifier will connect to.

The voice coil is fixed to a diaphragm (F) and is freely suspended by an outer edge suspension (B) and an inner element known as a spider (D). The voice coil sits in a strong magnetic field that is created

1. It's a rather odd name given that all loudspeakers are dynamic in some respect. If they weren't they wouldn't produce sound.

by a powerful permanent magnet (E) that commonly uses ceramic, alnico or rare earth construction. When current from the amplifier flows through the coil, it will create it's own magnetic field that will either aid or oppose the fixed field created by the permanent magnet, depending on the direction of the current. This results in a force that causes the coil to move within the fixed field. As the coil moves, the diaphragm moves with it, pushing on the surrounding air and creating sound. The larger the current, the stronger the newly created field and the greater the resulting aid or opposition, which results in greater movement of the diaphragm and a larger sound pressure. This fundamental design has changed little since its invention in the 1920s. Modern magnets, suspension and diaphragm materials have improved considerably in the intervening years but the operational principle is pretty much the same.

It is very difficult to create a driver that can cover the full audio spectrum of 20 Hz to 20 kHz while achieving sufficient listening volume at low distortion. Consequently, drivers are often designed to cover a limited portion of the audio spectrum. Low frequency drivers are commonly referred to as woofers while high frequency drivers are called tweeters. Drivers that cover the middle range of frequencies are given the highly inventive name midranges (although once upon a time they were called squawkers). A combination of these devices will be wired together with other components to create a complete home or auto loudspeaker system. Although very high quality systems can be produced, virtually all direct radiating dynamic loudspeaker systems suffer from low conversion efficiency. For a typical consumer system, only about 1% to 2% of the applied electrical power is turned into useful acoustic output power. The vast majority of the applied power simply makes the voice coil hot.

LOUDSPEAKER IMPEDANCE

Loudspeakers are given a nominal impedance value. The most common impedance for home use is 8 Ω while 4 Ω is common in automotive systems. It is important to remember that this is a nominal value and the true value varies with frequency. While it is common to test power amplifiers with large power resistors, they are only a coarse approximation of a real loudspeaker.

The electrical and mechanical characteristics of a loudspeaker combine to create an equivalent circuit with resistive, inductive and capacitive elements. A typical electrical circuit model[2] of a single loudspeaker driver is shown in Figure 7.4.2. R_{VC} and L_{VC} are the resistance and inductance of the voice coil, respectively. The other components are electrical equivalents of mechanical properties such as suspension losses.

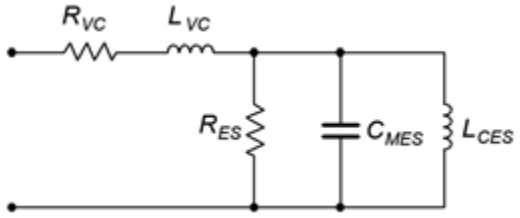


Figure 7.4.2: Dynamic loudspeaker electrical model.

2. 2Adapted from R. H. Small, "Direct Radiator Loudspeaker System Analysis", Journal of the Audio Engineering Society, June, 1972.

Clearly, this is not a simple 8 Ω resistor. In fact, we see a very complex impedance: The three parallel elements will create a resonant peak and the series inductance will cause the impedance to rise with frequency. Typically, the resonant peak occurs at the lower end of the spectrum and the associated resonant frequency is denoted on a spec sheet as f_S , the free-air resonance. For a nominal 8 Ω woofer, the peak impedance can be over 30 Ω. An example of a loudspeaker impedance plot is shown in Figure 7.4.3.

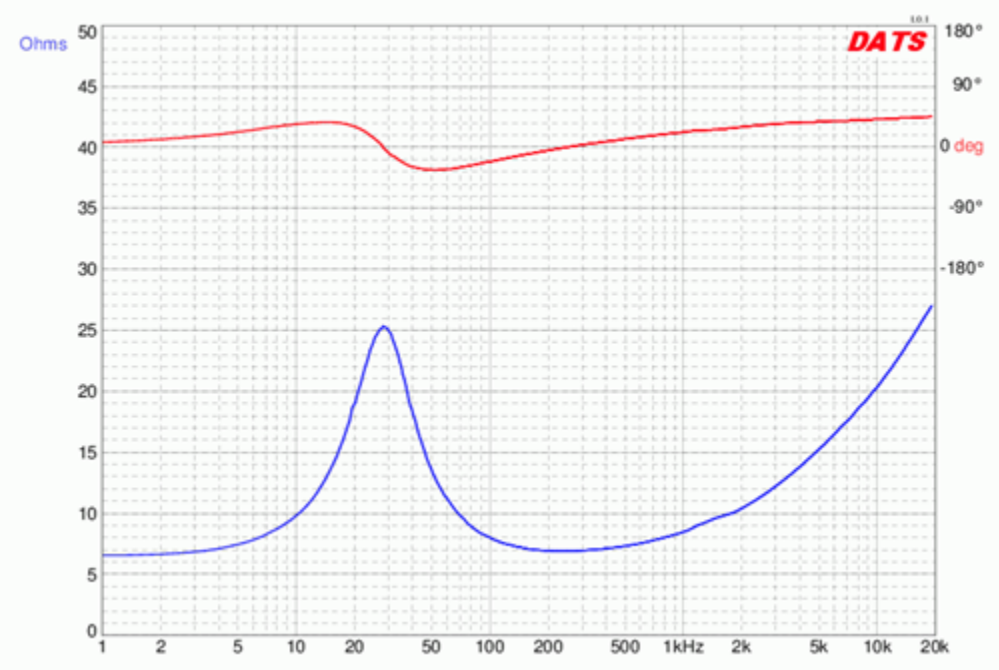


Figure 7.4.3: Dynamic loudspeaker impedance plot. Courtesy of Dayton Audio.

This loudspeaker is a nominal 8 Ω unit yet the impedance can be many times this value, and at some frequencies, less than 7 Ω. This plot also includes the phase angle of the loudspeaker and we can see that it can be upwards of 40° capacitive or inductive, depending on the frequency. What makes this more interesting is that a consumer loudspeaker system is a combination of multiple drivers plus other electrical components, and this can result in an even more complex impedance plot.

The obvious question is, "Does this have any effect on the analysis of the power amplifier?" The simple answer is, "Yes". Areas of the spectrum where the impedance magnitude drops below the nominal value will require more current for any given load voltage. Further, the phase shift caused by a partly reactive load will impact the power dissipation of the transistor.

Consider the power graph shown in Figure 8.3.6 for a purely resistive load. If we repeat the plot but add a noticeable phase shift to simulate a partly reactive load, something interesting happens, as shown in Figure 7.4.4.

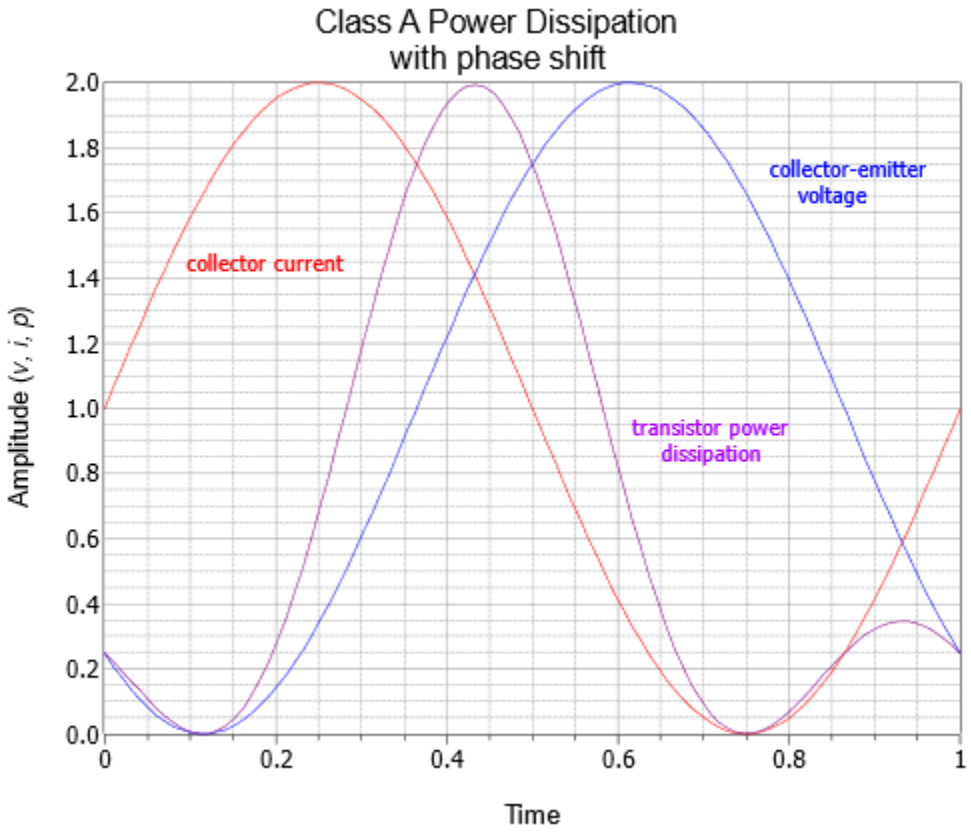


Figure 7.4.4: Power dissipation with reactive load.

The purple trace represents the power dissipation of the transistor. The peak current-voltage product is twice the value seen with a purely resistive load. This combination might lay outside the safe operating area of the transistor. Ultimately, reactive loads are somewhat more "challenging" than simple resistive loads. Therefore, the transistors may need to be rated higher than the values computed for an idealized resistive load.

Another way of looking at the issue of phase shift induced by loudspeakers and other complex loads is to examine the AC load line. Our previous work with load lines always assumed that the load was purely resistive. What happens in the complex impedance case?

If we examine a generic complex load at a single frequency, our former straight line load line turns into an ellipse, as shown in Figure 7.4.5.

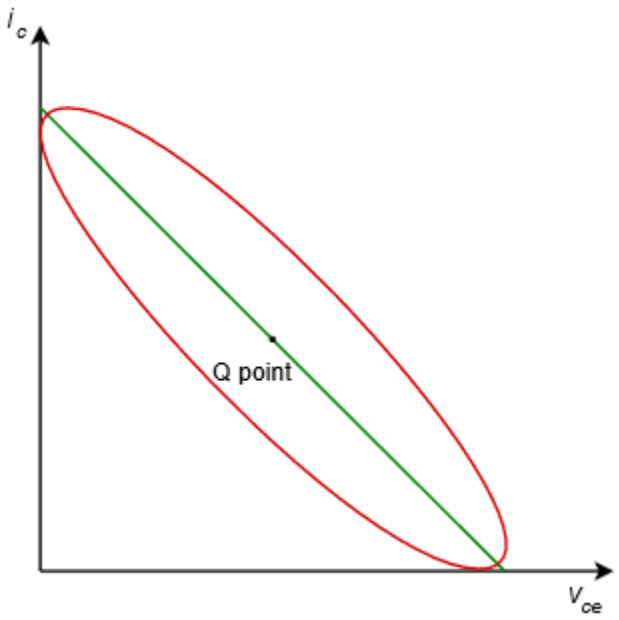


Figure 7.4.5: AC load line with complex load.

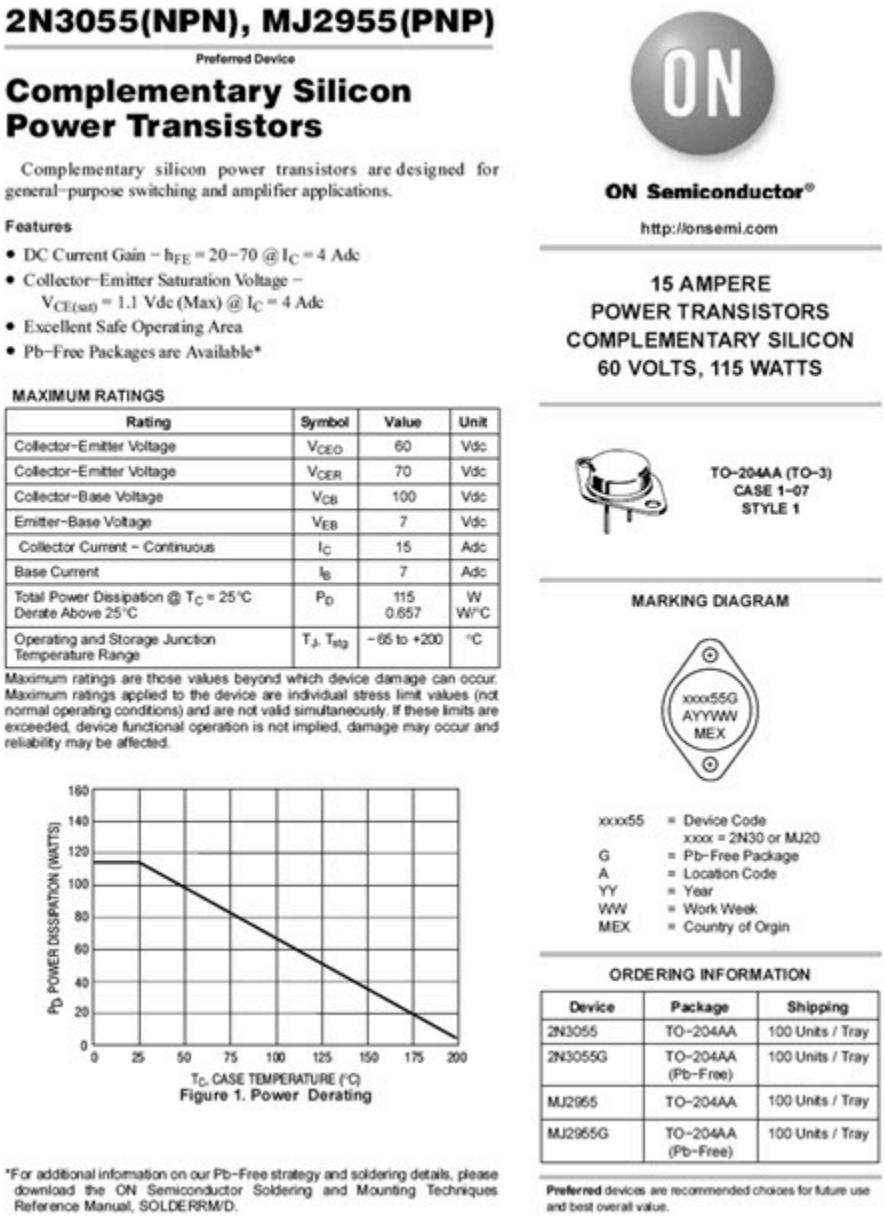
This plot assumes that the circuit has a centered Q point. The normal resistive load line is shown in green. We can visualize the signal starting at zero amplitude, meaning all we see is the Q point. As the signal gets larger and larger, it swings along the green line until, eventually, it maxes out at the two axes. In the complex impedance case, we also start at the Q point. As the signal increases, it traces out an ellipse around the Q point. Further increases would create a larger ellipse, and then a larger ellipse, and so on. Eventually, we would see the maximum swing just touching the axes. This is what is plotted above in red, the maximum case (i.e., at full compliance).

If the impedance angle changes, the aspect ratio of the ellipse changes in reaction to it. The larger the angle, the more open the ellipse becomes. The extremes are 0°, the purely resistive case that yields a collapsed ellipse or straight line; and 90°, the purely reactive case that yields a fully open ellipse, or circle. We have already seen that the phase angle of a loudspeaker changes with frequency, therefore, the load line also changes with frequency. As we sweep the input frequency from low to high, we can imagine the load line wavering back and forth between straight lines and various elliptical shapes. The important thing, though, is that some of these new operating regions (the areas where the red curve is above and to the right of the green line) may go outside the safe operating area of the transistor.

7.5 POWER TRANSISTOR DATA SHEET INTERPRETATION

The data sheet for a popular NPN power transistor, the 2N3055, is shown in Figure 7.5.1. This model is available from several different manufacturers. Due to the high power dissipation, the TO-92 plastic case that is used for small signal devices is not appropriate. Instead, this device uses the all-metal TO-3 case. Under the maximum ratings we find the device has a maximum power dissipation of 115 W at a case temperature of 25° C, a maximum collector current of 15 A and a maximum collector-emitter voltage of 60 V. Obviously, the device cannot withstand maximum current and voltage simultaneously.

2N3055 data sheet



2N3055(NPN), MJ2955(PNP)

Preferred Device

Complementary Silicon Power Transistors

Complementary silicon power transistors are designed for general−purpose switching and amplifier applications.

Features

- DC Current Gain − h_{FE} = 20−70 @ I_C = 4 Adc
- Collector−Emitter Saturation Voltage − $V_{CE(sat)}$ = 1.1 Vdc (Max) @ I_C = 4 Adc
- Excellent Safe Operating Area
- Pb−Free Packages are Available*

MAXIMUM RATINGS

Rating	Symbol	Value	Unit
Collector−Emitter Voltage	V_{CEO}	60	Vdc
Collector−Emitter Voltage	V_{CER}	70	Vdc
Collector−Base Voltage	V_{CB}	100	Vdc
Emitter−Base Voltage	V_{EB}	7	Vdc
Collector Current − Continuous	I_C	15	Adc
Base Current	I_B	7	Adc
Total Power Dissipation @ T_C = 25°C Derate Above 25°C	P_D	115 0.657	W W/°C
Operating and Storage Junction Temperature Range	T_J, T_{stg}	−65 to +200	°C

Maximum ratings are those values beyond which device damage can occur. Maximum ratings applied to the device are individual stress limit values (not normal operating conditions) and are not valid simultaneously. If these limits are exceeded, device functional operation is not implied, damage may occur and reliability may be affected.

Figure 1. Power Derating

*For additional information on our Pb−Free strategy and soldering details, please download the ON Semiconductor Soldering and Mounting Techniques Reference Manual, SOLDERRM/D.

ON Semiconductor®

http://onsemi.com

15 AMPERE POWER TRANSISTORS COMPLEMENTARY SILICON 60 VOLTS, 115 WATTS

TO−204AA (TO−3)
CASE 1−07
STYLE 1

MARKING DIAGRAM

xxxx55 = Device Code
xxxx = 2N30 or MJ20
G = Pb−Free Package
A = Location Code
YY = Year
WW = Work Week
MEX = Country of Orgin

ORDERING INFORMATION

Device	Package	Shipping
2N3055	TO−204AA	100 Units / Tray
2N3055G	TO−204AA (Pb−Free)	100 Units / Tray
MJ2955	TO−204AA	100 Units / Tray
MJ2955G	TO−204AA (Pb−Free)	100 Units / Tray

Preferred devices are recommended choices for future use and best overall value.

Figure 7.5.1*a*: 2N3055 data sheet. Used with permission from SCILLC dba ON Semiconductor.

In the drawing of the TO-3 case, only two leads are shown. These are for the emitter and base. The entire body of the device is the collector. This is because the device will most likely be attached to a metal heat sink (see next section) to help dissipate the heat generated. The greater the contact area, the more effective the heat flow will be. The curves presented in Figure 7.5.1*b* indicate that β is considerably lower than what we saw for small signal devices. Further, $I_C(sat)$ tends to be larger for higher power transistors. For very high currents, β might fall to less than 20 while $IC(sat)$ can be upwards of half of a volt.

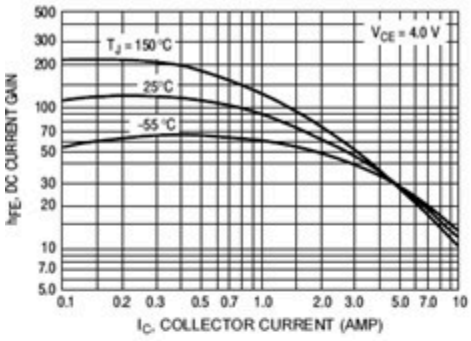


Figure 3. DC Current Gain, 2N3055 (NPN)

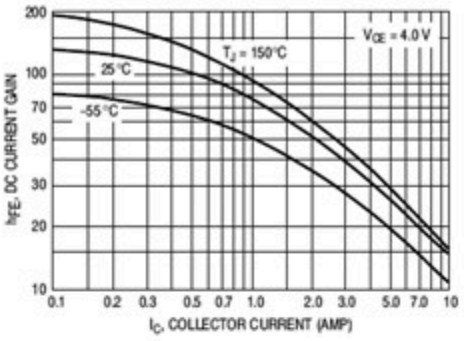


Figure 4. DC Current Gain, MJ2955 (PNP)

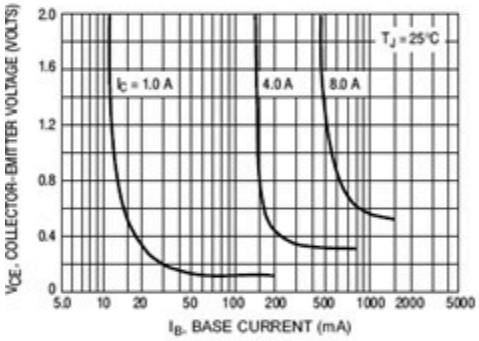


Figure 5. Collector Saturation Region, 2N3055 (NPN)

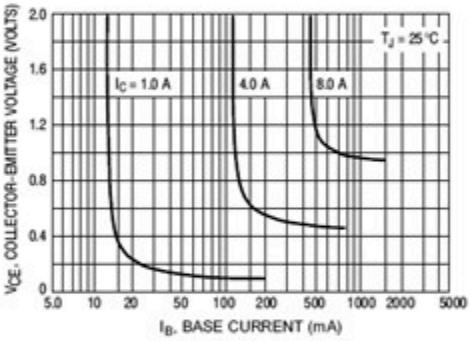


Figure 6. Collector Saturation Region, MJ2955 (PNP)

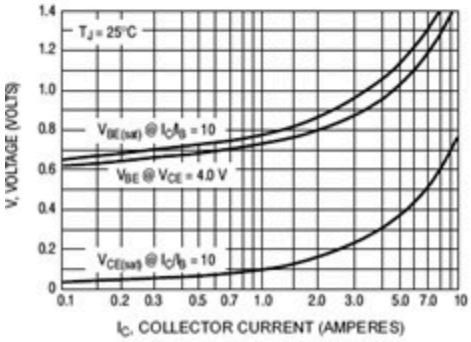


Figure 7. "On" Voltages, 2N3055 (NPN)

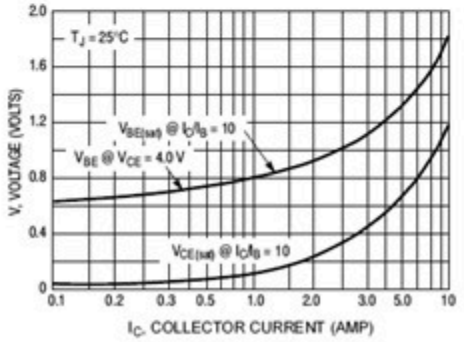


Figure 8. "On" Voltages, MJ2955 (PNP)

Figure 7.5.1*b*: 2N3055 data sheet (cont).

THERMAL CHARACTERISTICS

Characteristic	Symbol	Max	Unit
Thermal Resistance, Junction-to-Case	$R_{\theta JC}$	1.52	°C/W

ELECTRICAL CHARACTERISTICS (T_C = 25°C unless otherwise noted)

Characteristic	Symbol	Min	Max	Unit
OFF CHARACTERISTICS*				
Collector-Emitter Sustaining Voltage (Note 1) (I_C = 200 mAdc, I_B = 0)	$V_{CEO(sus)}$	60	–	Vdc
Collector-Emitter Sustaining Voltage (Note 1) (I_C = 200 mAdc, R_{BE} = 100 Ω)	$V_{CER(sus)}$	70	–	Vdc
Collector Cutoff Current (V_{CE} = 30 Vdc, I_B = 0)	I_{CEO}	–	0.7	mAdc
Collector Cutoff Current (V_{CE} = 100 Vdc, $V_{BE(off)}$ = 1.5 Vdc) (V_{CE} = 100 Vdc, $V_{BE(off)}$ = 1.5 Vdc, T_C = 150°C)	I_{CEX}	– –	1.0 5.0	mAdc
Emitter Cutoff Current (V_{BE} = 7.0 Vdc, I_C = 0)	I_{EBO}	–	5.0	mAdc
ON CHARACTERISTICS* (Note 1)				
DC Current Gain (I_C = 4.0 Adc, V_{CE} = 4.0 Vdc) (I_C = 10 Adc, V_{CE} = 4.0 Vdc)	h_{FE}	 20 5.0	 70 –	–
Collector-Emitter Saturation Voltage (I_C = 4.0 Adc, I_B = 400 mAdc) (I_C = 10 Adc, I_B = 3.3 Adc)	$V_{CE(sat)}$	–	 1.1 3.0	Vdc
Base-Emitter On Voltage (I_C = 4.0 Adc, V_{CE} = 4.0 Vdc)	$V_{BE(on)}$	–	1.5	Vdc
SECOND BREAKDOWN				
Second Breakdown Collector Current with Base Forward Biased (V_{CE} = 40 Vdc, t = 1.0 s, Nonrepetitive)	$I_{s/b}$	2.87	–	Adc
DYNAMIC CHARACTERISTICS				
Current Gain – Bandwidth Product (I_C = 0.5 Adc, V_{CE} = 10 Vdc, f = 1.0 MHz)	f_T	2.5	–	MHz
*Small-Signal Current Gain (I_C = 1.0 Adc, V_{CE} = 4.0 Vdc, f = 1.0 kHz)	h_{fe}	15	120	–
*Small-Signal Current Gain Cutoff Frequency (V_{CE} = 4.0 Vdc, I_C = 1.0 Adc, f = 1.0 kHz)	f_{hfe}	10	–	kHz

*Indicates Within JEDEC Registration. (2N3055)
1. Pulse Test: Pulse Width ≤ 300 μs, Duty Cycle ≤ 2.0%.

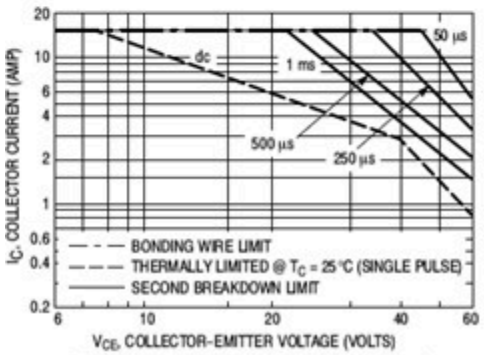


Figure 2. Active Region Safe Operating Area

There are two limitations on the power handling ability of a transistor: average junction temperature and second breakdown. Safe operating area curves indicate $I_C - V_{CE}$ limits of the transistor that must be observed for reliable operation; i.e., the transistor must not be subjected to greater dissipation than the curves indicate.

The data of Figure 2 is based on T_C = 25°C; $T_{J(pk)}$ is variable depending on power level. Second breakdown pulse limits are valid for duty cycles to 10% but must be derated for temperature according to Figure 1.

Figure 7.5.1*c*: 2N3055 data sheet (cont).

One item of note in Figure 7.5.1*c* is the small graphic at the bottom of the sheet. This is a plot of safe operating area. Basically, the combination of V_{CE} and I_C must fall within the lower-left zone. What is of particular interest is that the safe zone extends out further if the current/voltage combination is the result of a short pulse rather than a continuous condition.

POWER DERATING

One final item of concern is the graph found in Figure 7.5.1*a*, and magnified in Figure 7.5.2. This is a power derating curve.

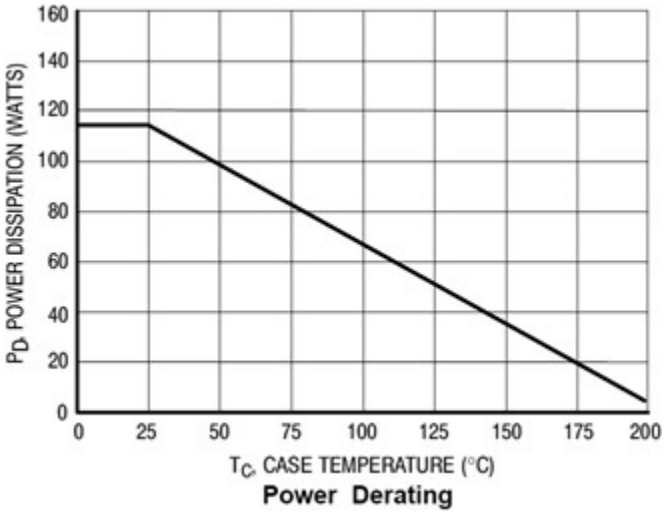


Figure 7.5.2: Power derating curve for 2N3055. Used with permission from SCILLC dba ON Semiconductor.

Although the device is rated for 115 watts, that is only true at case temperatures of 25° C or lower. At higher temperatures, the power dissipation capability decreases. For example, at 100° C this device can only dissipate about 65 watts. A precise value can be computed via the following formula:

Where

$$P_D = P_{25} D(T_{case} 25°C)$$

(7.5.1)

P_D is the power dissipation at the new case temperature,

P_{25} is the power dissipation at 25°C,

D is the derating factor (units of W/C°),

T_{case} is the new case temperature.

Example 7.5.1

Determine the power dissipation of a 2N3055 at 75°C.

Using the graph, this is a little over 80 watts.

For a more accurate result, we'll use Equation 7.5.1. From the data sheet the dissipation at 25°C, P_{25}, is 115 watts. The derating factor, D, is 0.657 W/C° (the derating factor is found directly above the graph of power derating in Figure 7.5.1a).

$$P_D = P_{25} - D(T_{ambient} - 25°C)$$

$$P_D = 115W - 0.657W/C°(75°C - 25°C)$$

$$P_D = 82.1W$$

7.6 HEAT SINKS

The issue with power transistors is always heat. As noted in Example 7.5.1, as the transistor heats up due to internal power dissipation, its ability to dissipate heat is compromised. The trick, then, is to efficiently move the heat from the transistor to someplace else. This is normally achieved through the use of a heat sink.

A heat sink is a metal device that is attached to the power transistor. Typically, they are made of aluminum and feature an array of fins. By increasing the surface area, heat can be moved away from the transistor more efficiently than by the transistor alone.

Heat sinks are designed to mount specific device case styles. The most common case styles include the TO-3 "can" along with the various "power tab" styles such as the TO-220 and TO-202. Special mounting hardware and insulation spacers are also required in order to maintain electrical isolation between the transistor and the heat sink as we do not want the heat sink to be electrically live. This usually takes the form of a mica sheet, and plastic washers and bushings for the mounting machine screws (for small heat sinks, sometimes nylon machine screws are used).

There are a few general rules that should be followed when using heat sinks:

- Always use some form of heat sink grease or thermally conductive pad between the heat sink and the device. This will increase the thermal transfer between the two parts, however, excessive quantities of heat sink grease will decrease performance.
- Mount fins in the vertical plane for optimum natural convective cooling.
- Do not overcrowd or obstruct devices that use heat sinks.
- Do not block air flow around heat sinks – particularly directly above and below items that rely on natural convection.
- If thermal demands are particularly high, consider using forced convection (i.e., a small fan directed at the heat sink).

Some typical heat sinks are shown below. Figure 7.6.1 shows a heat sink and thermal data plot for use with a single TO-3 case device.

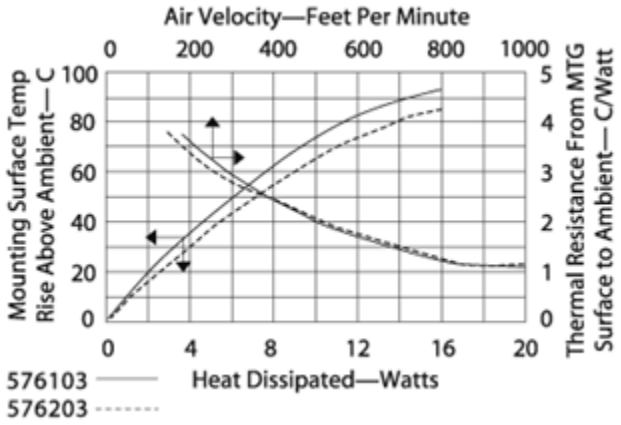


Figure 7.6.1: Heat sink for TO-3. Reprinted courtesy of Aavid Thermalloy, Inc.

Figure 7.6.2 shows a heat sink designed for a pair of transistors using TO-220 cases. In this photo, the white insulating pads can be seen between the transistors and heat sink.

Figure 7.6.2: Heat sink for dual TO-220. Reprinted courtesy of Aavid Thermalloy, Inc.

THERMAL RESISTANCE

In order to specify a particular heat sink for a given application, a more technical explanation is in order. What we are going to do is create a thermal circuit equivalent. In this model, the concept of thermal resistance is used. Thermal resistance denotes how easy it is to transfer heat energy from one mechanical part to another. The symbol for thermal resistance is θ, and the units are Celsius degrees per watt. In this model, temperature is analogous to voltage, and thermal power dissipation is analogous to current. A useful equation is,

$$P_D = \frac{\Delta T}{\theta_{total}} \tag{7.6.1}$$

Where P_D is the power dissipated by the semiconductor device in watts, ΔT is the temperature differential, and θ_{total} is the sum of the thermal resistances. Basically, this is a thermal version of Ohm's law.

In order to construct our model, let's take a closer look at the power-device/heat-sink combination. This is shown in Figure 7.6.3. The subscript *j* stands for junction, *c* is for case (of the transistor), *s* is for heat sink and a is for the ambient air. T_j is the semiconductor junction temperature and is created by the product of the transistor's current and voltage. This thermal source heats the device case to T_c . The thermal resistance between the two entities is θ_{jc} . The case, in turn, heats the heat sink via the interconnection. This thermal resistance is θ_{cs} , and the resulting temperature is T_s . Finally, the heat sink passes the thermal energy to the surrounding air which is sitting at T_a . The thermal resistance of the heat sink to the air is θ_{sa} . The equivalent thermal model is shown in Figure 7.6.4. Although this "thermal circuit" does not have perfect correspondence with normal circuit analysis, it does illustrate the main points.

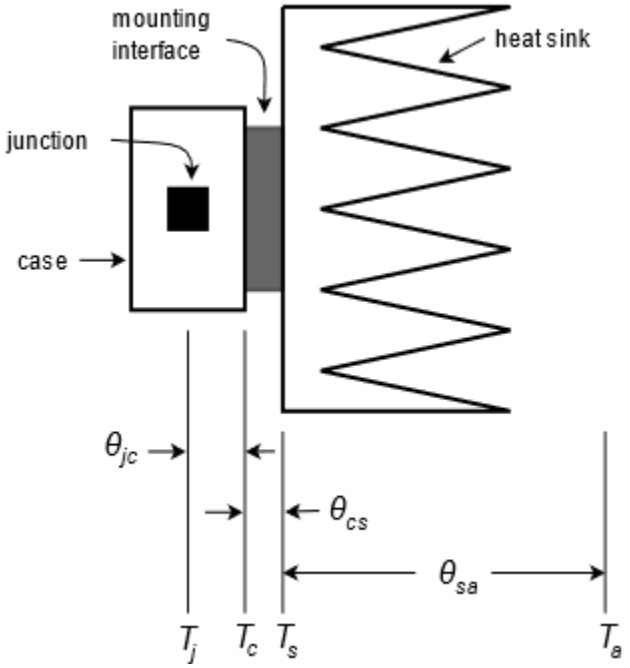


Figure 7.6.3: Device and heat sink.

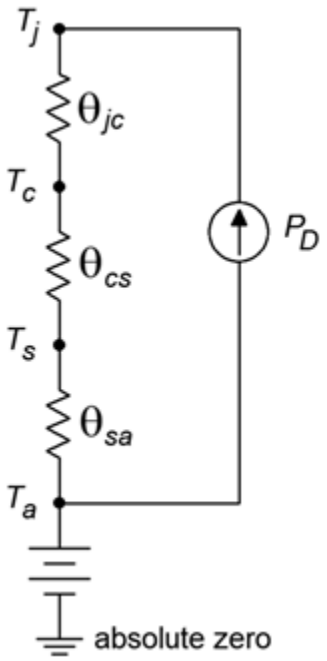


Figure 7.6.4: Equivalent thermal model of Figure 7.6.3.

In this model, ground represents a temperature of absolute zero. The circuit is sitting at an ambient temperature T_a , thus a voltage source of T_a is connected to ground and the heat sink. The three thermal resistances are in series and are driven by a current source that is set by the present power dissipation of the device. Note that if the power dissipation is high, the resulting "voltage drops" across the thermal resistances are high. Voltage is analogous to temperature in this model, so this indicates that a high temperature is created. Because there is a maximum limit to T_j, a higher power dissipation requires lower thermal resistances. As θ_{jc} is set by the device manufacturer, we have no control over that element. However, θ_{cs} is a function of the case style and the insulation material used, so we do have some control (but not a lot) over that. On the other hand, as the person who specifies the heat sink, we have a great deal of control over θ_{sa} . Values for θ_{sa} are given by heat sink manufacturers. A useful variation of Equation 7.6.1 is

$$P_D = \frac{T_j - T_a}{\theta_{jc} + \theta_{cs} + \theta_{sa}}$$

(7.6.2)

Normally, power dissipation, junction and ambient temperatures, θ_{jc} and θ_{cs} are known. The idea is to determine an appropriate heat sink. Both T_j and θ_{jc} are given by the semiconductor device manufacturer. The ambient temperature, T_a , may be determined experimentally. Due to localized warming, it tends to be higher than the actual "room temperature". Standard graphs, such as those found in Figure 7.6.5, may be used to determine θ_{cs} . Note the generally lower values of θ_{cs} for the TO-3 case relative to the TO-220. This is one reason why TO-3 cases are used for higher power devices. This case also makes it easier for the manufacturer to reduce θ_{jc} .

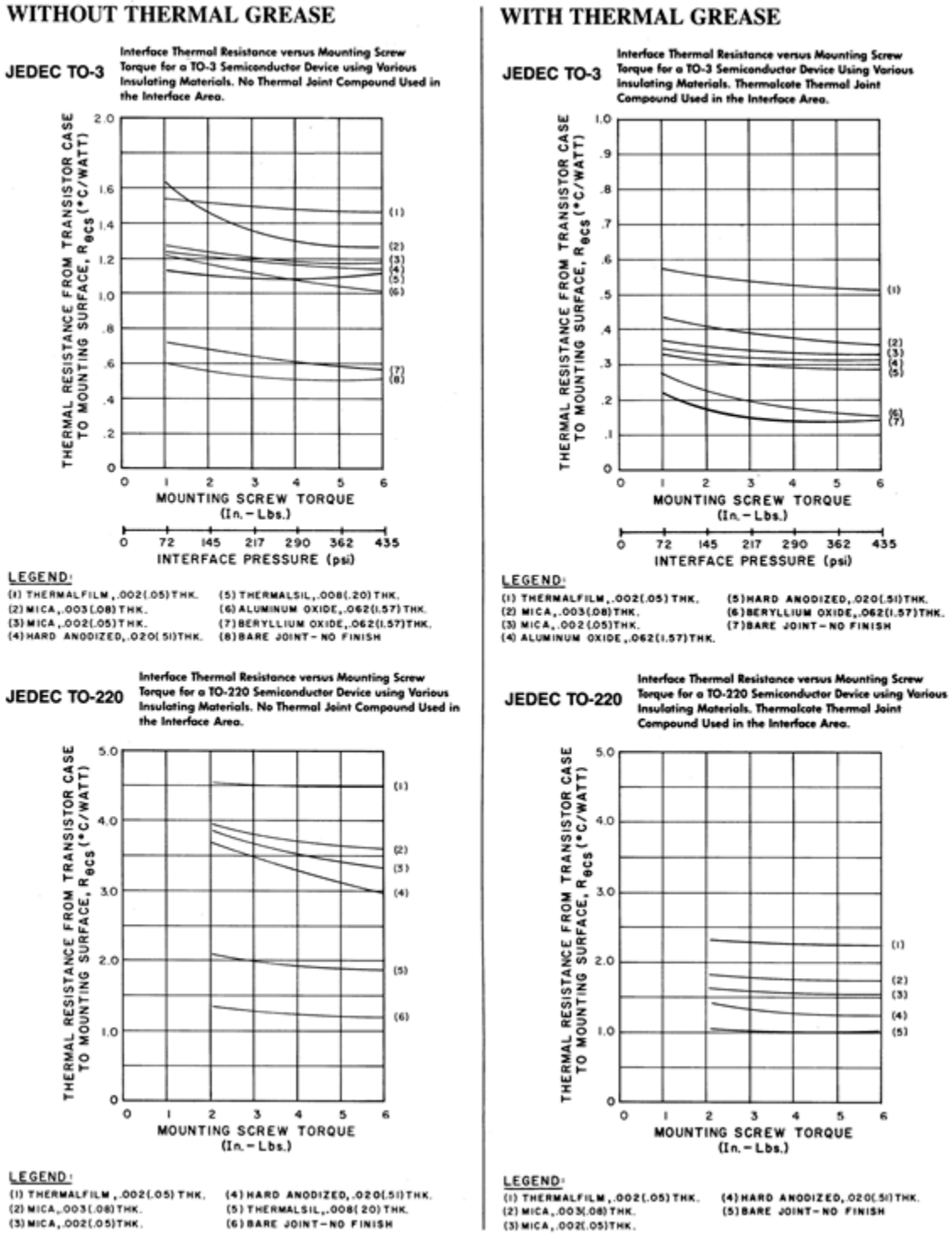


Figure 7.6.5: θCS for TO-3 and TO-220 Reprinted courtesy of Thermalloy, Inc.

Example 7.6.1

Determine the appropriate heat sink rating for a power device rated as follows: $T_{j(max)}$ = 175°C, TO-3 case style, θ_{jc} = 1.5C°/W. The device will be dissipating a maximum of 15 W in an ambient temperature of 40C°.

Assume that the heat sink will be mounted with heat sink grease and a 0.002 mica insulator.

First, find θ_{cs} from the TO-3 "With Thermal Grease" graph in Figure 7.6.5. Curve 3 is used. The approximate value is 0.35C° /W.

$$P_D = \frac{T_j - T_a}{\theta_{jc} + \theta_{cs} + \theta_{sa}}$$

$$\theta_{sa} = Tj - TaPD\theta jc - \theta cs$$

$$\theta_{sa} = \frac{175°C - 40°C}{15W} - 1.5C°/W - 0.35C°/W$$

$$\theta_{sa} = 7.15C°/W$$

This is the maximum acceptable value for the heat sink's thermal resistance. Note that the use of heat sink grease gives us an extra 0.8 C°/W or so. For this application, the heat sink pictured in Figure 7.6.1 will most likely be sufficient without added forced air cooling (the graph stops at less than 4 C°/W with an air flow of under 200 feet/minute).

If we repeat this problem with a much higher PD, things work out a little differently. Let's use 40 W this time.

$$P_D = \frac{T_j - T_a}{\theta_{jc} + \theta_{cs} + \theta_{sa}}$$

$$\theta_{sa} = \frac{T_j - T_a}{P_D} - \theta_{jc}\theta_{cs}$$

$$\theta_{sa} = \frac{175°C - 40°C}{40W} - 1.5C°/W - 0.35C°/W$$

$$\theta_{sa} = 1.53C°/W$$

If we hope to use that same heat sink, we will have to add forced air cooling of at least 700 feet/minute. The other option would be to find a more thermally efficient (and probably much larger) heat sink if we hope to use natural convection alone.

7.7 SUMMARY

Class A operation is defined as having collector current flow for 360° of the cycle. This means that a single output device can be used to amplify the entire input waveform. To determine the maximum signal swing, or compliance, an AC load line is used. This is similar to a DC load line and plots all possible transistor voltage and current coordinate pairs. The efficiency of the class A amplifier tends to be low. The maximum theoretical efficiency is only 25%. Further, the amplifier draws full current from the power supply regardless of whether or not a signal is present. As a consequence, the transistor runs hottest when there is no signal. With an applied signal, some of the power formerly dissipated within the transistor is shifted to the load.

Loudspeakers offer a complex impedance as a load. As such, they are more challenging than simple resistive loads and may require the output transistor to be rated for a higher-than-normal power dissipation.

Heat sinks are used to efficiently move heat from the transistor's internal structure to the surrounding air. The thermal effectiveness of a heat sink is measured by thermal resistance, θ. The lower the value of θ, the more effective the heat sink is at transferring heat from the transistor to the surrounding air. For high power applications where a good deal of heat is generated, heat sinks are augmented with forced air cooling.

Review Questions

1. Define class A operation.
2. Why are voltage followers generally preferred over voltage amplifiers for power output applications?
3. How does an AC load line differ from a DC load line?
4. What is the advantage of having a centered Q point on the AC load line?
5. What effect does a reactive load have on an AC load line?
6. Describe the operation of a dynamic loudspeaker.
7. What are heat sinks? What are they used?
8. What is thermal resistance?

7.8 EXERCISES

Assume diodes are silicon unless stated otherwise

ANALYSIS PROBLEMS

1. Draw the AC load line for the circuit of Figure 7.8.1. Also determine the compliance, maximum load power, maximum transistor dissipation and efficiency.
   V_{CC} = 6 V, V_{EE} = –12 V, R_{gen} = 50 Ω, R_B = 2.2 kΩ, R_E = 470 Ω, R_L = 75 Ω.

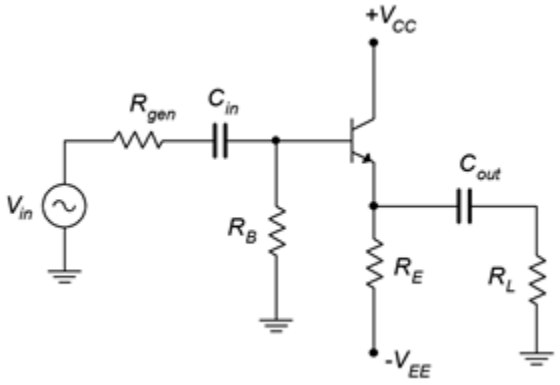


Figure 7.8.1

2. Recalculate Problem 1 if the load is halved.
3. Determine if the circuit of Figure 7.8.2 has a centered Q point on its AC load line.
   V_{CC} = –10 V, V_{EE} = 15 V, R_B = 1 kΩ, R_E = 330 Ω, R_L = 50 Ω.

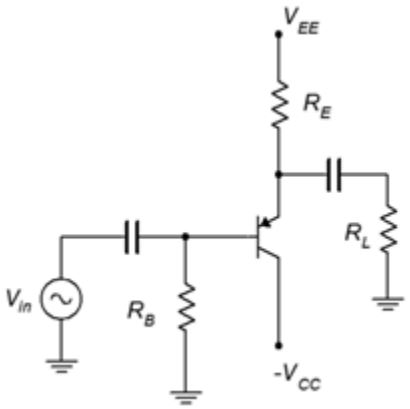


Figure 7.8.2

4. Draw the AC load line for the circuit of Figure 7.8.2. Also determine the compliance, maximum load power, maximum transistor dissipation and efficiency.
   V_{CC} = –8 V, V_{EE} = 12 V, R_B = 1 kΩ, R_E = 330 Ω, R_L = 32 Ω.

5. Draw the AC load line for the circuit of Figure 7.8.3. Also determine the compliance, maximum load power, maximum transistor dissipation and efficiency.
   V_{CC} = 15 V, V_{EE} = –20 V, R_B = 10 kΩ, R_E = 100 Ω, R_L = 16 Ω.

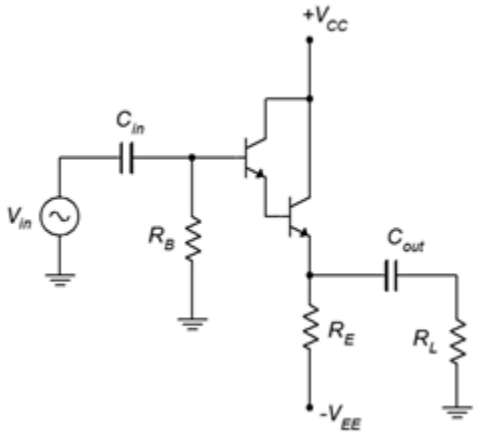


Figure 7.8.3

6. Determine if the circuit of Figure 7.8.4 has a centered Q point on its AC load line.
   V_{CC} = 30 V, R_1 = 3.9 kΩ, R_2 = 3.3 kΩ, R_E = 560 Ω, R_L = 50 Ω.

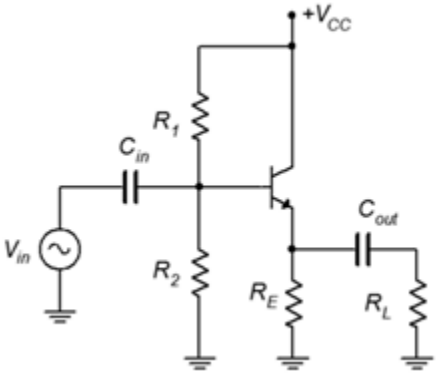


Figure 7.8.4

7. Draw the AC load line for the circuit of Figure 7.8.4. Also determine the compliance, maximum load power, maximum transistor dissipation and efficiency.
   V_{CC} = 30 V, R_1 = 2.2 kΩ, R_2 = 2.2 kΩ, R_E = 470 Ω, R_L = 32 Ω.
8. Determine if the circuit of Figure 7.8.5 has a centered Q point on its AC load line.
   V_{CC} = 15 V, V_{EE} = –15 V, R_B = 1 kΩ, R_E = 510 Ω, R_{SW} = 10 Ω, R_C = 270 Ω, R_L = 50 Ω.

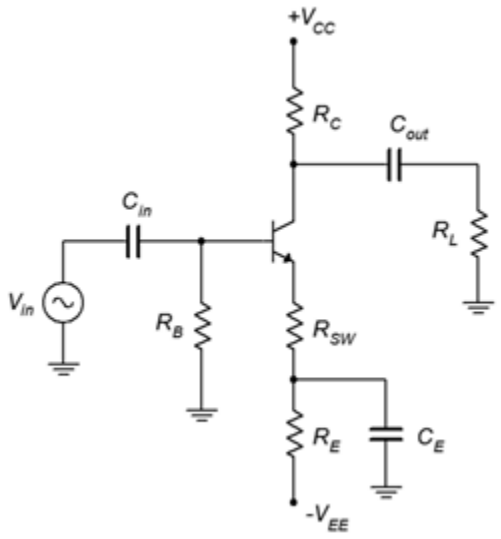


Figure 7.8.5

9. Draw the AC load line for the circuit of Figure 7.8.5. Also determine the compliance, maximum load power, maximum transistor dissipation and efficiency.
   V_{CC} = 25 V, V_{EE} = −15 V, R_B = 1 kΩ, R_E = 270 Ω, R_{SW} = 6.8 Ω, R_C = 330 Ω, R_L = 16 Ω.
10. A power transistor has a $P_{D(max)}$ of 50 watts at 25°C. It has a derating factor of 0.4 W/C°. Will this transistor be sufficient for a circuit that needs to dissipate 40 watts at 85°C?
11. A power transistor has a $P_{D(max)}$ of 100 watts at 25°C. It has a derating factor of 0.6 W/C°. Will this transistor be sufficient for a circuit that needs to dissipate 65 watts at 75°C?
12. Determine the appropriate heat sink rating for a power device rated as follows: $T_{j(max)}$ = 175°C, TO-3 case style, θ_{jc} = 1.5 C°/W. The device will be dissipating a maximum of 25 W in an ambient temperature of 35°C. Assume that the heat sink will be mounted with heat sink grease and a 0.003 mica insulator.
13. Determine the appropriate heat sink rating for a power device rated as follows: $T_{j(max)}$ = 165°C, TO-220 case style, θ_{jc} = 3 C°/W. The device will be dissipating a maximum of 15 W in an ambient temperature of 35°C. Assume that the heat sink will be mounted with heat sink grease and a 0.002 mica insulator.

DESIGN PROBLEMS

14. Alter the emitter power supply in the circuit described in Problem 1 to achieve a centered Q point.
15. Alter the emitter power supply in the circuit described in Problem 4 to achieve a centered Q point.

CHALLENGE PROBLEMS

16. Find a heat sink (make and model number) that will meet the thermal resistance requirement for Problem 12 with no more than 400 feet/minute of forced air.
17. Alter the voltage divider in the circuit described in Problem 6 to achieve a centered Q point.

COMPUTER SIMULATION PROBLEMS

18. Perform a transient analysis for the circuit described in Problem 1 to verify the compliance.
19. Perform a transient analysis for the circuit described in Problem 4 to verify the compliance.
20. Perform a transient analysis for the circuit described in Problem 9 to verify the compliance.

UNIT 8: BJT CLASS B POWER AMPLIFIERS

Learning Objectives

After completing this chapter, you should be able to:

- Define class B operation.Determine AC load lines for class B amplifier stages.
- Determine the compliance, maximum load power, efficiency and required device ratings for class B circuits.
- Discuss the advantages and disadvantages of class B operation versus class A operation.
- Discuss the origin of notch distortion and methods used to mitigate it.
- Explain the operation of a current mirror.
- Explain the operation of a Sziklai pair.
- Explain the operation and use of a VBE multiplier.
- Outline the operation of fully complimentary and quasi complimentary output stages utilizing direct coupled driver stages.
- Discuss methods to protect the output devices from overload.

8.1 INTRODUCTION

B amplifiers have long been a mainstay of linear amplifier design. Compared to class A operation, class B amplifiers offer much greater power efficiency. That is, a much larger percentage of the applied DC power can be turned into useful AC output to the load. This also means that the power dissipation requirements for the transistors are lowered. Further, unlike class A operation, class B designs do not continuously draw full power from their DC supplies. Instead, they draw current as it is needed, and therefore run relatively cool at idle and at low output power.

The downside for this improved efficiency is added complexity. For starters, two transistors are required for linear class B operation. Also, the biasing can be a little trickier which requires modifications to the relatively straightforward class A biasing circuits we have already examined. Also, class B amplifiers suffer from a unique form of distortion that class A amplifiers do not.

In this chapter we shall also examine the use auxiliary device configurations and sub-circuits to improve performance. These include the current mirror, Sziklai pair, V_{BE} multiplier, and overload protection and prevention circuitry. As with the class A amplifier, the common collector or voltage follower configuration tends to be the most widely used configuration for class B circuits, hence, we shall focus on followers and not on voltage amplifiers.

8.2 THE CLASS B CONFIGURATION

Class B operation is defined as having AC collector current flow 180° out of the cycle. Consequently, in order to amplify the entire signal, two devices will be needed. Further, we will need to pay attention to how the two waveform halves are "stitched together" as this could be a problem area. The obvious question at this point is, why do we bother separating the positive and negative half- waves if it leads to circuit complexity and possible waveform issues? The answer is improved efficiency.

In the previous chapter we discovered that class A amplifiers are not efficient. In fact, at best they only transform 25% of the DC input power into useful load power. Why does this occur and how does the class B topology address this situation?

The basic idea of class B is to push the Q point down so that it is sitting right at cutoff on the AC load line. This means that I_{CQ} is 0 A and virtually no power is drawn from the supply at idle. Locating the Q point at cutoff also means that the transistor will immediately clip the negative portion of the wave. Consequently, we will need a mirror image circuit to produce that portion (and which will clip the positive portion).

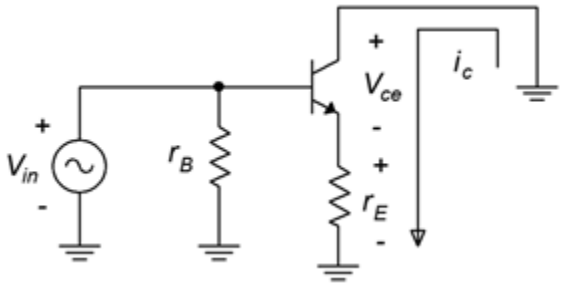


Figure 8.2.1: Voltage follower simplified AC circuit.

To gain a better understanding of how class B operation works, consider the simplified AC circuit of a voltage follower shown in Figure 8.2.1. If we situate the Q point directly at $V_{CE(cutoff)}$ then the associated I_{CQ} is 0 A. As the input signal swings positive, the collector current increases. As it does so, the voltage across the load (r_E) begins to increase and the voltage across the transistor's collector-emitter begins to decrease (due to KVL). When the input signal swings negative, the transistor is turned off. As a result, no collector current is created, no voltage is developed across the load and V_{CE} stays at cutoff. It is as if the input waveform has been half-wave rectified. This action is shown in Figure 8.2.2. As the input signal swings positive, the operating point slides up the load line, moving toward saturation, and this increased current creates a load voltage that follows the input signal. In contrast, when the input tries to swing negative, there is no place else to go on the load line and the negative portion of the wave is simply clipped.

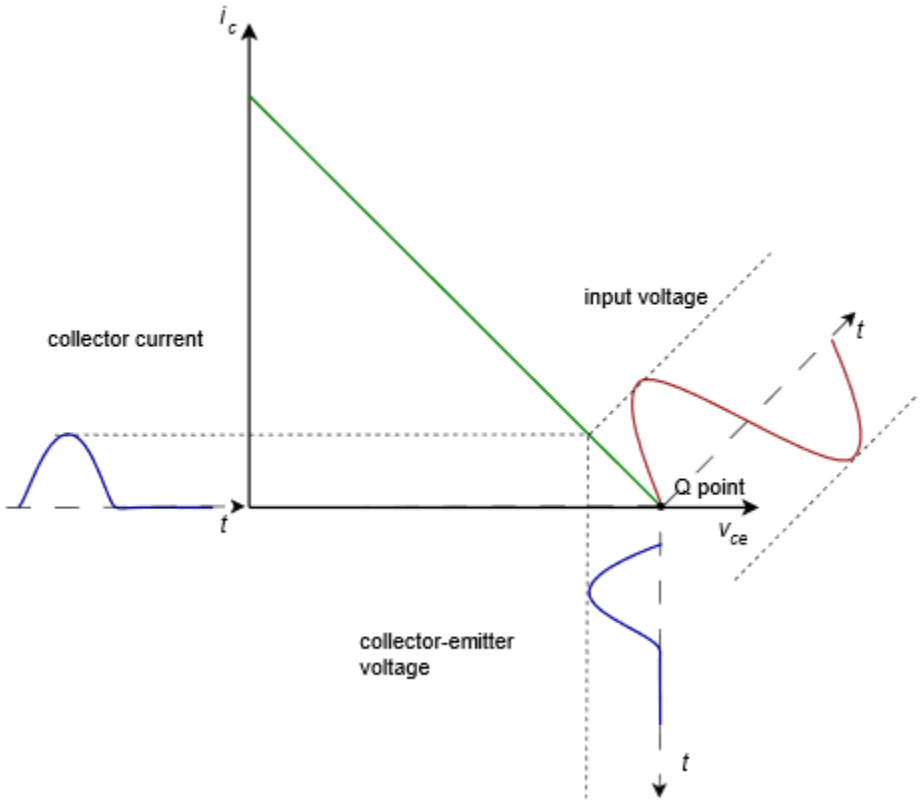


Figure 8.2.2: AC load line for class B operation.

If we made a PNP version of the circuit depicted in Figure 8.2.1, the exact opposite would happen: the amplifier would reproduce the negative portion of the wave and clip the positive portion. The next question is, how do we bias the transistor at cutoff and splice together the NPN and PNP versions into a workable whole?

Let's begin by making two bare-bones emitter followers, one NPN and the other PNP. We'll connect their emitters together and tie that to the load. We'll connect the NPN's collector directly to a DC supply and the PNP's collector to ground. Remember, collector resistors will not be needed because these are followers. We will include no biasing components on the base because we want to set I_{CQ} to 0 A. We will also have to add input and output capacitors to prevent the source and load from inadvertently shorting out or shunting portions of the DC circuit. The result is seen in Figure 8.2.3.

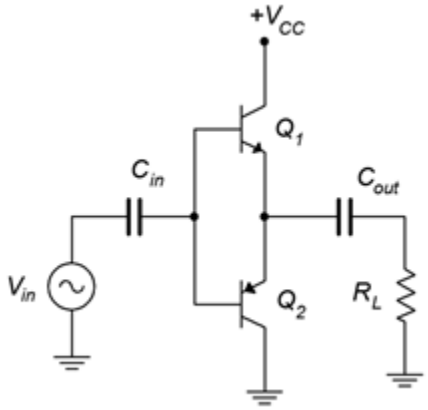


Figure 8.2.3: Prototype Class B circuit.

With no signal supplied, both transistors must be off. This is because their bases are tied together, and without some other applied potential, both base-emitter voltages must be zero. Assuming Q_1 and Q_2 are matched, the power supply voltage should split evenly between them, leaving half of V_{CC} reaching R_L. at the emitters. This potential also appears across C_{out} , preventing the DC voltage from When the input signal goes positive, it raises V_{B1} and V_{B2} above 0.5 V_{CC} . This keeps Q_2 off but turns on Q_1 . Current is now free to flow down through Q_1 and into the load. When the input signal swings negative, the inverse happens: Q_1 is turned off and Q_2 is turned on. This allows current to flow up from the load and down through Q_2 (if this is confusing, remember that a DC voltage had already been established across C_{out} that is equal to 0.5 V_{CC} and this is what allows the current to flow from ground up through R_L and then down through Q_2 as Q_2 begins to conduct). You can think of Q_1 pushing current into the load (sourcing) and Q_2 pulling current from the load (sinking). Consequently, class B amplifiers are sometimes called a push-pull amplifiers.

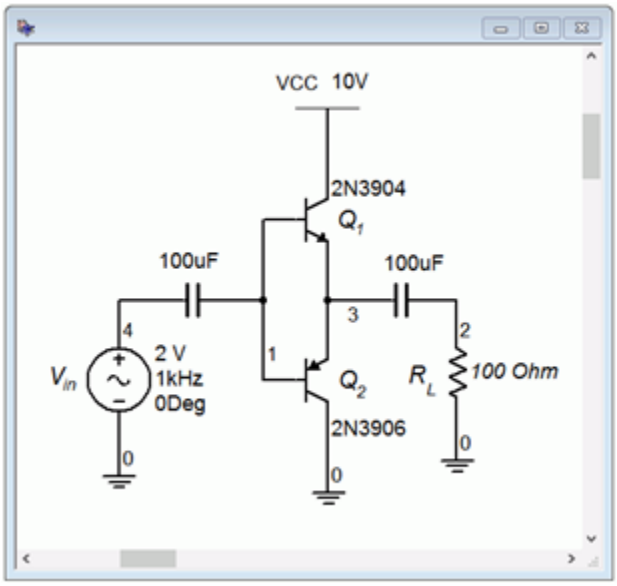


Figure 8.2.4: Prototype Class B circuit in simulator.

To see how well this prototype works, we'll enter a version into a simulator, as shown in Figure 8.2.4. A transient analysis is performed with the results shown in Figure 8.2.5.

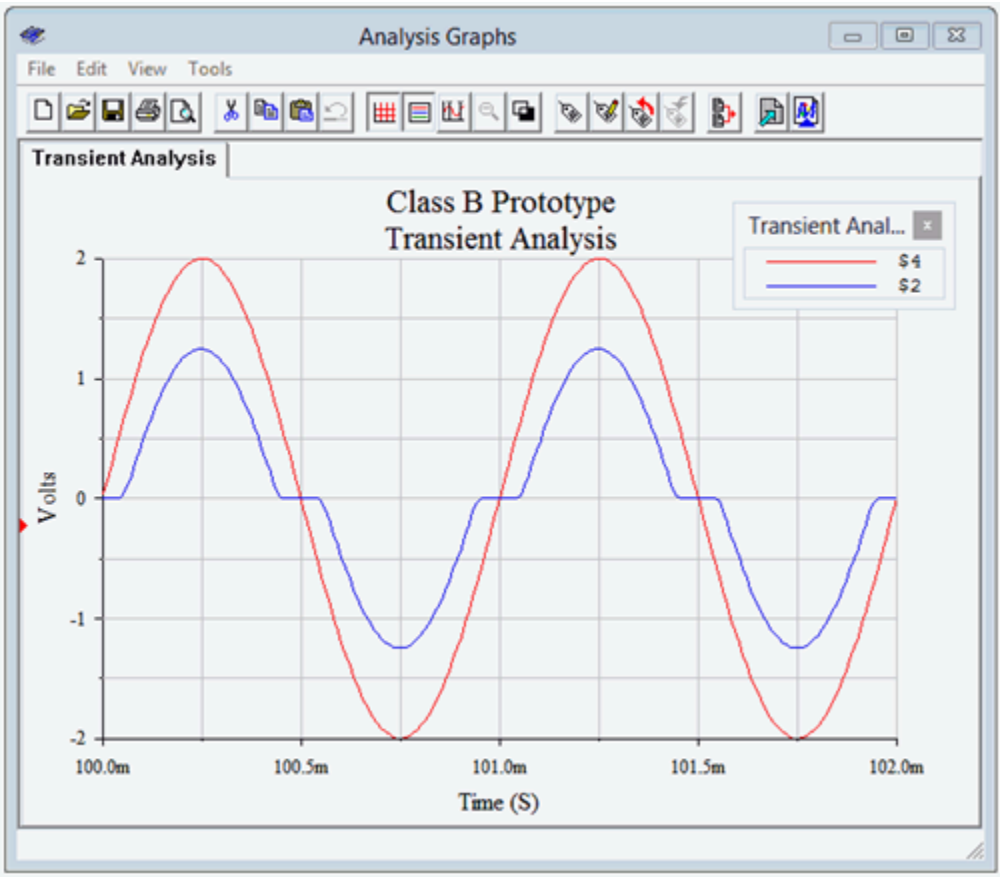


Figure 8.2.5: Transient analysis of prototype Class B circuit.

Clearly, there are issues with the output waveform (node 2, in blue). First off, the signal amplitude is noticeably smaller than the 2 volt peak input signal. The second issue is the bizarre "flat spotting" of the output waveform near the zero-crossing points. It turns out that these two problems are manifestations of the same root cause. If we look carefully at the peaks, we can get a clue as to what is going on. The peak output voltage is about 0.75 volts below the input, or just about one PN junction forward potential. The problem is that the input signal will not truly turn on the NPN transistor until the signal exceeds approximately 0.7 volts or drops below –0.7 volts for the PNP side. That region between –0.7 volts and +0.7 volts is a dead zone that the amplifier will not respond to. Essentially, the amplifier "rips out" anything between ±0.7 volts. This is a gross form of distortion and goes by many names including notch distortion and cross-over distortion. The particularly nasty part about this form of distortion is that it hits small signals worse than large signals. Most other forms of nonlinearities tend to get worse as the signal level increases.

CLASS AB OPERATION

The basic solution to this problem is to provide a small idle current so that the transistors are almost on. In this way, only a very small input signal will be needed to turn on the devices. As this would slightly increase the conduction angle, this form of operation is referred to as class AB operation. One potential solution is to add a voltage divider as depicted in Figure 8.2.6.

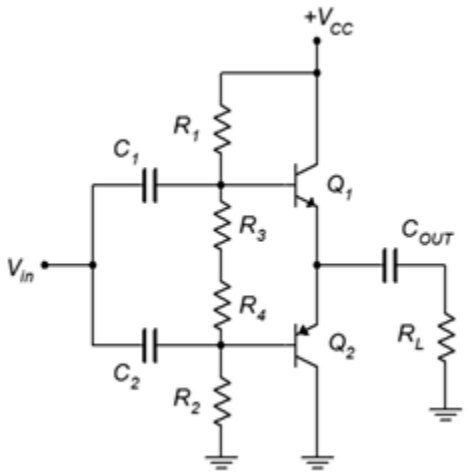


Figure 8.2.6: Prototype Class AB circuit.

This circuit uses a symmetrical layout: everything above a horizontal line drawn through the middle is echoed by a component below the line. In other words, $R_1 = R_2$, $R_3 = R_4$, $C_1 = C_2$, and Q_1 and Q_2 are a complimentary pair. The divider is configured so that the voltage drops across R_3 and R_4 are about 0.7 volts each. Properly designed, the circuit of Figure 8.2.6 will reduce notch distortion. Unfortunately, it has other problems. The first issue involves the three capacitors. C_{OUT} in particular might be quite large. These can be removed if we went to a symmetrical bipolar power supply. Instead of running the PNP's collector to ground, we'll tie it to a negative DC supply. To keep the same total voltage, we'll set Q_1's collector to half of the original V_{CC} and Q_2's collector to half of the original V_{CC} but negative. On the input side, we could run the input signal to the junction of R_3 and R_4 .

The second issue plaguing the circuit of Figure 8.2.6 is the stability of the bias. The voltage divider resistors have to be very accurate in order to set the transistors right where we want them and stability is a concern. The problem is that we are trying to use a device with a linear current-voltage characteristic (a resistor) to match the exponential current-voltage characteristic of a PN junction. This problem is exacerbated by the fact that these devices will drift with temperature, and drift in different ways. The solution to this problem is to use a device with better matching characteristics. What better device to match a PN junction than another PN junction?

THE CURRENT MIRROR

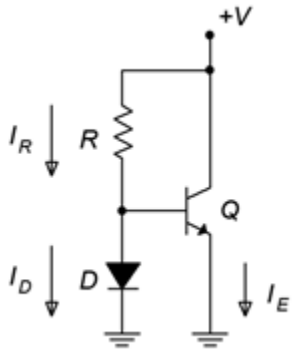


Figure 8.2.7: A simple current mirror.

Consider the circuit shown in Figure 8.2.7. This is called a current mirror. Here is how it works: First, look at the divider between R and D. The voltage across R must equal the supply voltage minus the diode drop, or approximately $V - 0.7$. This sets up a current, I_R. If the base current is small enough to ignore, this same current flows down through the diode as I_D. This diode current sets up a specific voltage across the diode (somewhere in the vicinity of 0.7 volts although the exact voltage is not important). Because the diode is in parallel with the base-emitter junction, then $V_{BE} = V_D$. If the transconductance curve (I-V curve) of the transistor is identical to that of the diode, then the emitter current must be the same as the diode current. Any change in the diode current would cause a slight change in diode voltage, and since diode voltage and base-emitter voltage are the same, then the emitter current must change in response. In other words, the emitter current mirrors the diode current. We can program the diode current (and hence, the emitter current) by setting an appropriate value for R. Whatever the current through R is, that's also the collector current.

The idea of the current mirror is used with great effect in integrated circuits where it is easy to match device characteristics. When it comes to discrete components, it is not nearly so easy to match a diode to a transistor. Fortunately, we don't have to have a perfect match. Simply using any signal diode will provide a much better match for V_{BE} than using a resistor. Although the diode current won't match the collector current precisely, the bias will be more stable and the components will track much better than when using a resistor.

Combining these ideas leads us to the circuit of Figure 8.2.8. This is our first practical class B amplifier with the potential for decent values of distortion and stability.

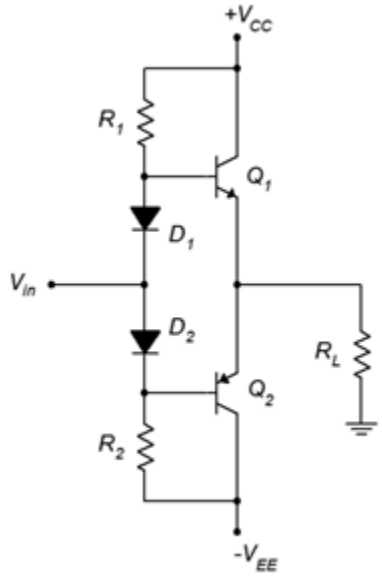


Figure 8.2.8: Class B amplifier with diode bias and bipolar supplies.

One thing that sometimes bothers people when they first see a diode biased amplifier is how the AC signal will pass through the diodes to the bases. At first glance, it appears that the positive portion of the input signal would be "going the wrong way" against diode D_1. What we need to remember is that

D_1 is already forward-biased due to the DC supply and surrounding resistors. The signal won't see an open, it will see the dynamic resistance of the diode.[1]

CLASS B CIRCUIT MAXIMUMS

Now that we have a workable circuit, we need to derive formulas for the endpoints of the AC load line, meaning $V_{CE(cutoff)}$ and $I_{C(sat)}$, and determine the compliance. The first item of note is cutoff. Because the two transistors will split the available supply, V_{CEQ} will always equal half of the total supply. Further, we have biased these devices at cutoff, and therefore

$$V_{CE(cutoff)} = V_{CEQ} = 0.5 \cdot \text{ Total DC Supply}$$

(8.2.1)

In the case of a bipolar supply, that's the same as one of the two sides. Because the class B uses two transistors, peak compliance will be the same as cutoff.

$$Compliance_{peak} = V_{CEQ} = 0.5 \cdot \text{ Total DC Supply}$$

(8.2.2)

The maximum voltage rating of the transistors will occur when they are off. In that instance, if the opposite transistor is fully conducting it will have a negligible voltage across its collector-emitter. Consequently, the off-state transistor can see the entire power supply.

$$BV_{CEO} = \text{ Total DC Supply}$$

(8.2.3)

The saturation current is dictated by the compliance and the load. The only thing that limits AC current is the load. Therefore

$$I_{C(sat)} = \frac{Compliance_{peak}}{r_L}$$

(8.2.4)

Based on that, we can say

$$P_{load(max)} = \frac{Compliance_{RMS}{}^2}{r_L}$$

(8.2.5)

1. Of course, the dynamic resistance is a function of the current flowing through the diode so it will fluctuate as the signal changes. This is best thought of as a distortion generating mechanism because it will slightly alter the signal that reaches the base. This distortion is most likely orders of magnitude smaller than the notch distortion the diode mitigates, so it's a good trade.

Before we go any further, there is an important item to note about the circuit of Figure 8.2.8 (and the variants we shall discuss). If you take another look at the circuit you will see that there is nothing in the collector-emitter line to limit DC current. In fact, if we were to plot the DC load alongside the AC load, we'd get something like Figure 8.2.9.

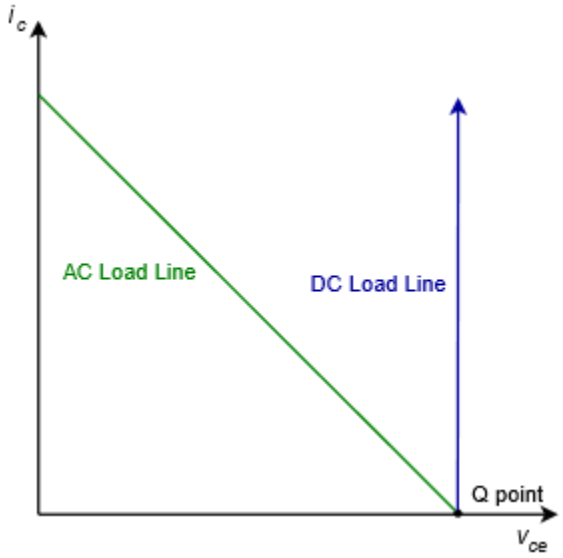


Figure 8.2.9: Comparison of AC and DC load lines for class B operation

The DC load line goes straight up. There is no saturation current value short of infinity. What this really means is that if we're not careful with the bias, it is possible to destroy the transistors. The same is true if we accidentally short the load. A quick examination of Equations 8.2.4 and 8.2.5 shows that a shorted load condition would lead to huge currents and powers, and destroy the transistors in the process. With an audio amplifier, this could happen if one of the strands of loudspeaker wire unraveled and touched the adjacent lead. Obviously, not a happy situation unless you prefer the smell of burnt silicon over the sound of music. We will examine means of protecting the transistors from accidental overloads later in the chapter.

CLASS B POWER DISSIPATION

Transistor power dissipation for the class B configuration is a bit trickier than it was for the class A. The idle (Q point) power dissipation is very low. It is found by multiplying I_{CQ} by V_{CEQ}. I_{CQ} is generally set at a few percent of $iC(sat)$ so it's obviously way below the maximum load power. Unlike the class A design, P_{DQ} does not represent the worst case for class B.

To determine the worst case transistor power dissipation, we begin by describing the current and voltage waveforms during the conduction phase of the NPN. The collector current will appear as the positive half of a sine wave. The maximum case has the current starting at zero and peaking at $I_C(sat)$ (or alternately, V_{CEQ}/r_L). For v_{CE}, it starts at V_{CEQ} and then swings down to zero as a negative half sine.

Nothing says that the maximum load power case must cause the maximum transistor dissipation. In fact, we saw this wasn't the case for class A. Consequently, we will introduce a coefficient, k,

that represents the percentage of the maximum current. We now arrive at our general equations for transistor current and voltage for the first half cycle.

$$I_C = k\frac{V_{CEQ}}{r_L}\sin 2\pi ft$$

(8.2.6)

$$V_{CE} = V_{CEQ}(1 - k\sin 2\pi ft)$$

(8.2.7)

Where 0≤◈≤1

For convenience, we'll set $2\pi f$ to 1. To get the power dissipation, we find the product of the transistor's current and voltage.

Remove ugly $\sin^2$ term…

$$P_D = \frac{{V_{CEQ}}^2}{r_L}\left(k\sin t + \frac{k^2}{2}\cos 2t - \frac{k^2}{2}\right)$$

..and integrate to get:

$$P_D = \frac{{V_{CEQ}}^2}{r_L}\left(k\cos t - \frac{k^2}{4}\sin 2t - \frac{k^2 t}{2}\right)|_0^\pi$$

Note that $\frac{V_{CEQ}^2}{r_L} = 2P_{load(max)}$ This is a constant, so replace it to simplify and then evaluate the expression. Finally, divide by 2π to find the average over one full cycle.

$$P_D = \frac{2P_{load(max)}\left(2k\frac{k^2\pi}{2}\frac{k^2 t}{2}\right)}{2\pi}$$

$$P_D = 2P_{load(max)}\left(\frac{k}{\pi}\frac{k^2}{4}\right)$$

(8.2.8)

Equation 8.2.8 is the general case. For the worst case, we need the min/max *k* value. We will take the derivative of Equation 8.2.8 and then set it to zero to find the worst case value of *k*.

$$P_D = 2P_{load(max)}\left(\frac{k}{\pi} - \frac{k^2}{4}\right)$$

$$\frac{dP_D}{dk} = 2P_{load(max)}\left(\frac{1}{\pi} - \frac{k}{2}\right)$$

Worst case occurs at $k = 2/\pi$. This means that only $2/\pi$, or 63.7%, of the load line is used at maximal heating of the transistor. 63.7% of the load line corresponds to a load power of about 40% of $P_{load}(max)$ (i.e., 0.637^2). We now substitute this value back into Equation 8.2.8 to find the worst case P_D.

$$P_D = 2P_{load(max)}\left(\frac{2/\pi}{\pi}\frac{(2/\pi)^2}{4}\right)$$

$$P_D = \frac{2}{\pi^2}P_{load(max)} \approx \frac{P_{load(max)}}{5}$$

(8.2.9)

The end result is that when the load is receiving about 40% of its maximum power, the transistors will be at their hottest and will be dissipating approximately 20% of the maximum load power (or about half of the power delivered to the load at that point). Thus, if a class B amplifier is rated to produce a maximum load power of 100 watts, the transistors will get their hottest when the load is receiving 40 watts, and each transistor will be dissipating 20 watts. The transistors will be dissipating less power when the load is at maximum. This is easily verified by substituting $k = 1$ into Equation 8.2.8. The result is a power dissipation of 13.7% of maximum load power, or 13.7 watts for the preceding example.

To help gain a deeper understanding of precisely what's happening here, the transistor waveforms are plotted in Figures 8.2.10 and 8.2.11. The maximum load power case is presented in Figure 8.2.10. Here, we see the full I_C and V_{CE} swings. Note that when I_C is maximum, V_{CE} is 0, hence the power is 0. In contrast, Figure 8.2.11 shows the worst case. Collector current peaks at just under 64% of maximum but V_{CE} drops to only about 36% rather than 0% of its maximum. This results in a power dissipation curve with considerably greater area underneath it, indicating a higher average power.

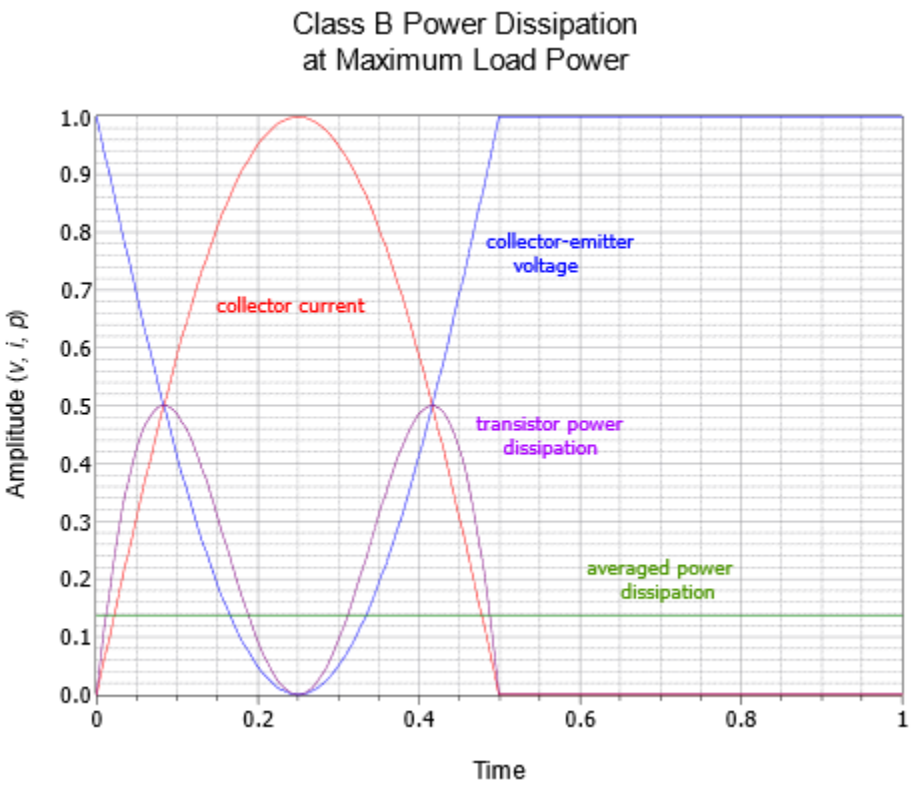


Figure 8.2.10: Class B transistor power dissipation at $P_{Load(max)}$.

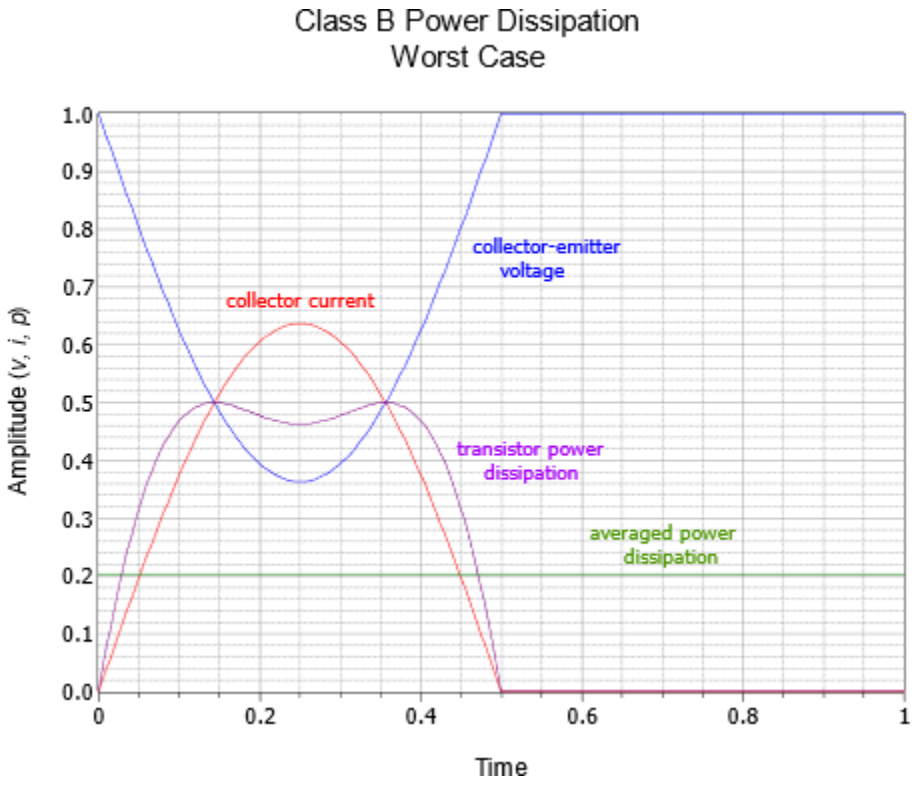


Figure 8.2.11: Class B transistor power dissipation, worst case.

Finally, it is worth remembering that the reactive load issues discussed for class A amplifiers still apply to class B amplifiers. Loads with a complex impedance may be harder to drive than the ideal purely resistive loads examined here. Therefore, we may have to over-rate the transistors above $P_{load(max)}/5$.

As a side note, the AC load line for a class B amplifier with a reactive load will appear as an ellipse that has been cut in half (refer back to Figure 8.4.5 and imagine a horizontal cut line running through the Q point).

CLASS B EFFICIENCY

Efficiency is defined as useful output or load power versus supplied DC power.

$$\eta = \frac{P_{out}}{P_{in}} = \frac{P_{load(max)}}{P_{DC}}$$

This is dynamic for class B amplifiers. At *Pload*(*max*) the DC supply is delivering the full voltage of $2V_{CEQ}$ and the corresponding peak value of the current is i_C(*sat*).

$$I_{C(sat)} = \frac{V_{CEQ}}{r_L}$$

The average of this over a half-cycle is

$$I_{C(avg)} = \frac{1}{\pi} \times \frac{V_{CEQ}}{r_L}$$

Therefore, the supplied power must be

$$P_{DC} = 2V_{CEQ} i_{C(sat)}$$

$$P_{DC} = 2V_{CEQ} \times \frac{1}{\pi} \frac{V_{CEQ}}{r_L}$$

$$P_{DC} = \frac{2}{\pi} \times \frac{{V_{CEQ}}^2}{r_L}$$

As noted previously V CEQ 2 r L = 2 P_{load} (max) therefore

$$P_{DC} = \frac{4}{\pi} \times P_{load(max)}$$

Finally, substitute this expression back into the original definition for efficiency.

```
*** QuickLaTeX cannot compile formula:
\[\eta_{max} = \frac{P_{load (max)}}{{P_{DC}} \]

*** Error message:
File ended while scanning use of \frac .
Emergency stop.
```

$$\eta_{max} = \frac{P_{load(max)}}{\frac{4}{\pi}P_{load(max)}}$$

$$\eta_{max} = \frac{\pi}{4}$$

$$\eta_{max} \approx 78.5\%$$

We find that the maximum theoretical efficiency of a class B amplifier is over three times that of a class A amplifier.

Time for an example.

Example 8.2.1

The amplifier shown in Figure 8.2.12 is driving a nominal 8 Ω loudspeaker. Determine the compliance, maximum load power and worst case transistor dissipation. Also estimate Zin assuming β = 50 and determine the transistor ratings for maximum current and BV_{CEO} .

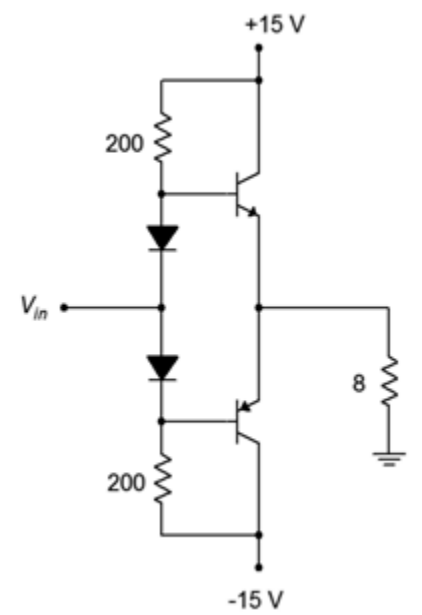


Figure 8.2.12: Schematic for Example 8.2.1.

By inspection, V_{CEQ} = 15 V. This is the peak compliance.

$$compliance = 15Vpeak = 10.6VRMS$$

Given the compliance, we can use power law to find the load power

$$P_{load(max)} = \frac{Compliance_{RMS}{}^2}{R_L}$$

$$P_{load(max)} = \frac{(0.707 \times 15V)^2}{8\Omega}$$

$$P_{load(max)} = 14W$$

That's not huge but it might be enough to irritate the neighbors.

The transistors' worst case power dissipation is

$$P_D = \frac{P_{load(max)}}{5}$$

$$P_D = \frac{14W}{5}$$

$$P_D = 2.8W$$

The breakdown voltage is the entire supply so BV_{CEO} > 30 V. The maximum current through the transistors is the same as the maximum load current or $i_{C(sat)}$.

$$I_{C(sat)} = \frac{V_{CEQ}}{r_L}$$

$$I_{C(sat)} = \frac{15V}{8\Omega}$$

$$I_{C(sat)} = 1.88A$$

The input impedance is found in the usual manner but with a minor twist. $Z_{in(base)}$ is approximately equal to βr_L, or about 400 Ω. Only one transistor is on at any given time, though, so this is in parallel with the two 200 Ω biasing resistors but not the other $Z_{in(base)}$ (the "off" transistor has a very high input impedance because it is not conducting). This leaves a Z^2 of about 80 Ω, or ten times the load impedance.

COMPUTER SIMULATION

To verify the basic operation of a class B amplifier, the circuit of Figure 8.2.13 is entered into a simulator. The amplifier should clip just below the ±10 volt power rails so a 10 volt peak source is used to verify this. Also, A_v should be about 1.

2. *in*(*base*)

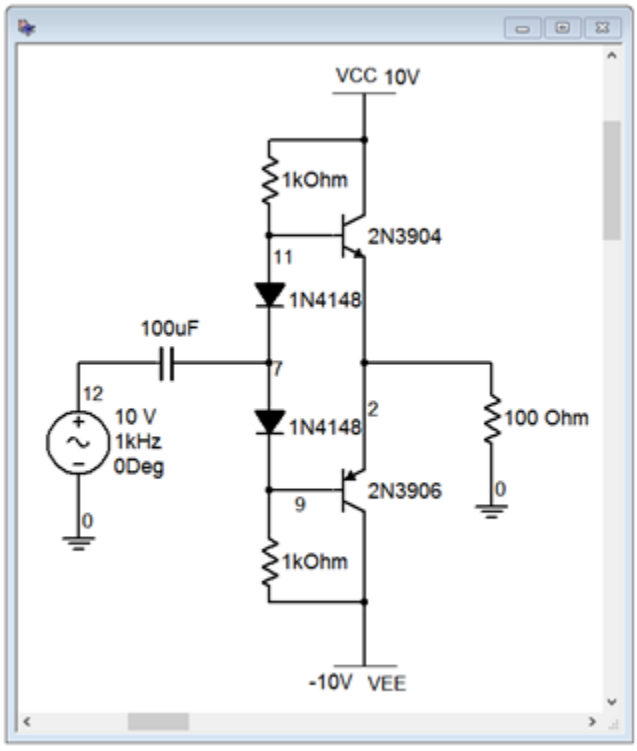


Figure 8.2.13: Class B amplifier in simulator.

The results of a transient analysis are shown in Figure 8.2.14. First, it is apparent that the voltage gain is approximately unity as the input and output waves are nearly coincident (with the exception of the clipped portion). Clipping occurs at about 8.5 volts. This is largely due to limiting from the biasing diodes. Once the input gets close in value to the power supply, the diodes become reverse- biased and the signal does not make it to the base. Thus, the output clips prematurely.

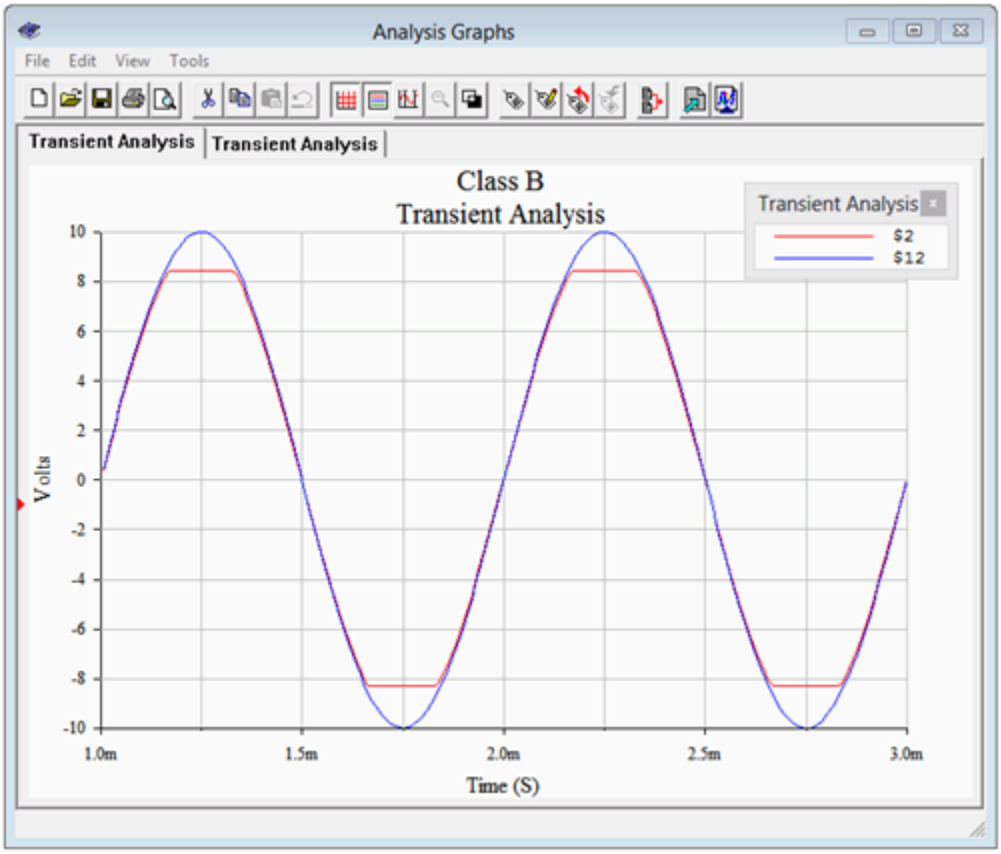


Figure 8.2.14: Class B amplifier transient analysis.

To verify earlier commentary regarding the biasing diodes, the circuit is modified so that the diodes are shorted out and the transient analysis is run again. The results are shown in Figure 8.2.15.

Two things should be apparent in the new simulation. First, the output waveform is suffering from obvious notch distortion. Therefore, we can be pleased that the biasing diodes have done their job to reduce this effect. The second item involves the compliance. The new output is not clipped. Granted, the lack of biasing diodes has reduced the signal by about 0.7 volts, but a careful examination of the output waveform shows that it has reached over 9 volts peak without clipping. In fact, if we increase the input to 12 volts peak, as in Figure 8.2.16, we can see that clipping occurs just under the power rails. Thus, the premature clipping is due to the diodes and their interaction with the power supplies and surrounding resistors. In short, we lose a little bit of compliance due to the diodes but that's much better than getting notch distortion.

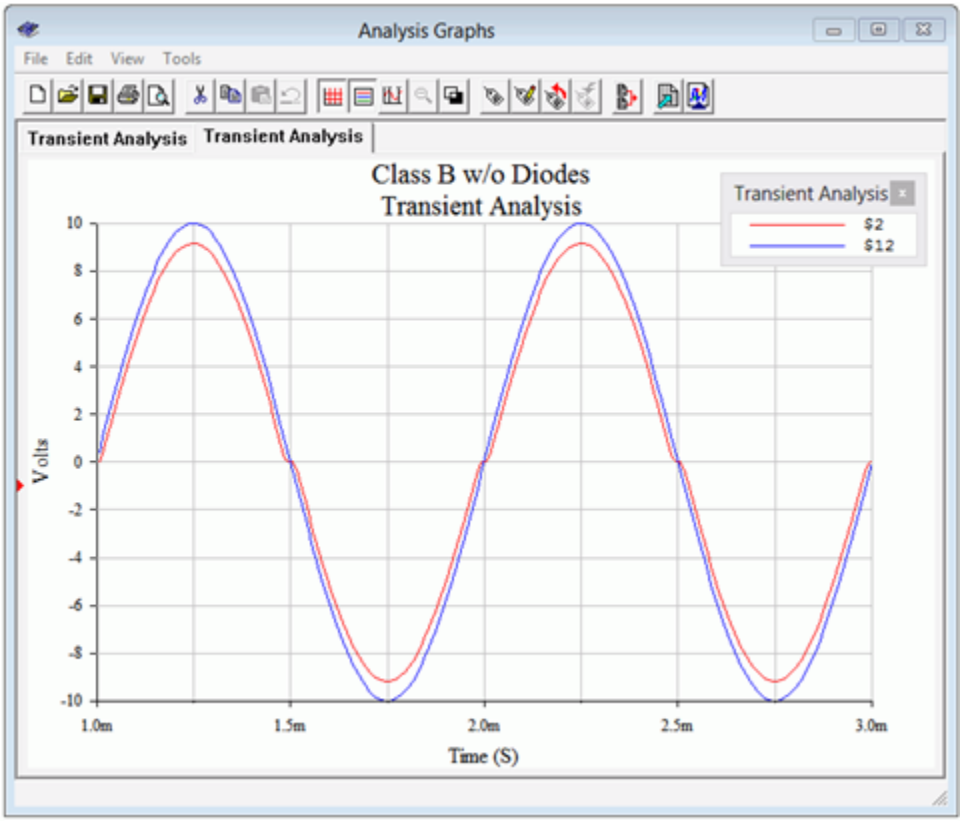


Figure 8.2.15: Class B amplifier transient analysis, without biasing diodes.

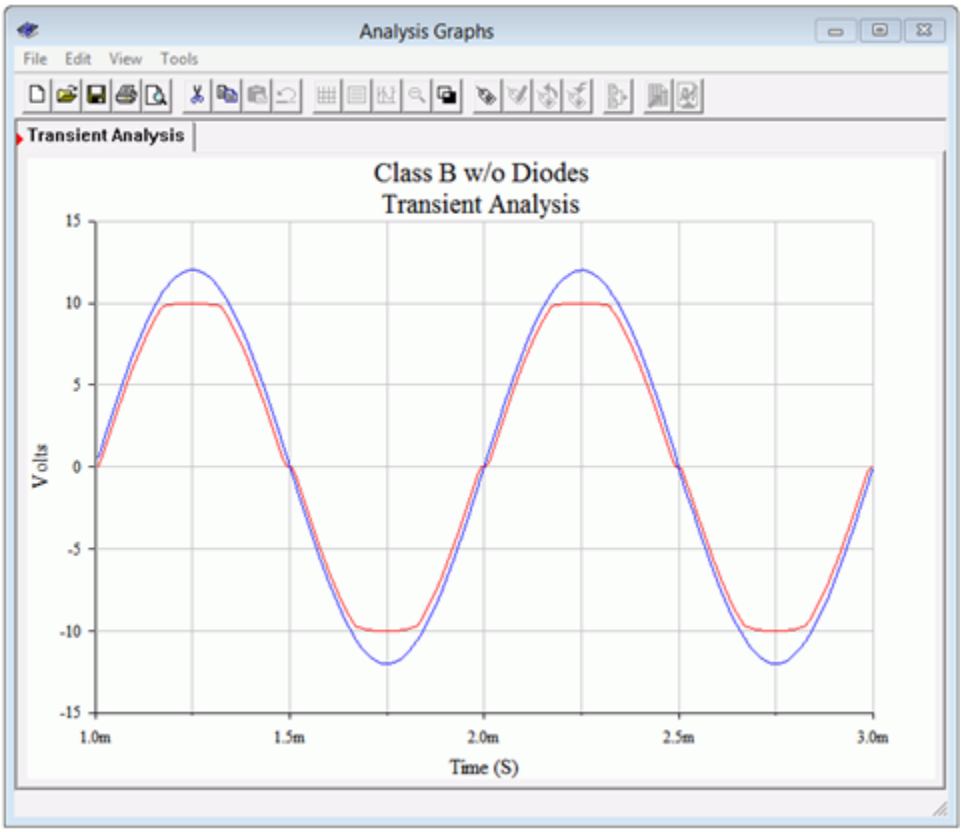


Figure 8.2.16: Class B amplifier transient analysis, without biasing diodes and showing clipping.

DIRECT COUPLED DRIVER

The circuits we have examined offer both current gain and power gain but not voltage gain. In order to increase the signal voltage, some preceding voltage gain stages will most likely be needed. These stages can be connected to the previous class B follower circuits with coupling capacitors but this is not the most effective method. A more common technique is the use of a direct coupled driver.

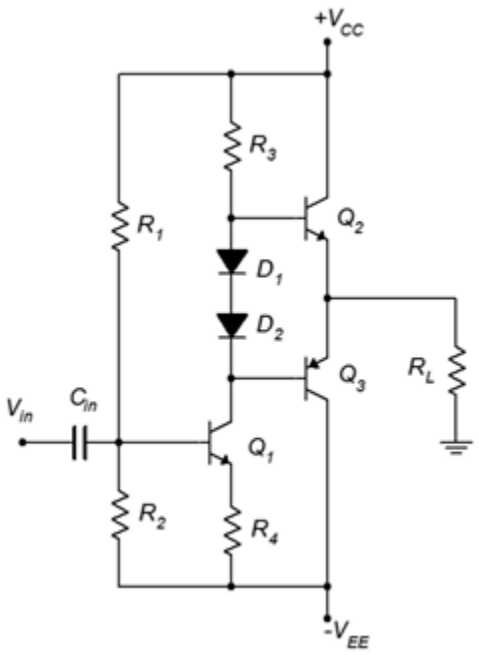


Figure 8.2.17: Class B amplifier with direct coupled driver.

A class B follower with a direct coupled driver stage is shown in Figure 8.2.17. What we've done here is combined an ordinary class A common emitter amplifier (in this case, using voltage divider bias) with a class B follower. The follower is positioned where the common emitter stage's collector resistor would be normally. This eliminates three components: the collector resistor, the interstage coupling capacitor, and the lower base biasing resistor of the class B output stage. The removal of the resistors raises the effective load resistance for the first stage, thus producing a higher voltage gain from the driver stage.

Biasing the direct coupled driver is not difficult if we remember one thing: the DC voltage across R_3 must be equal to approximately V_{CC} – 0.7 V. If this is not the case, the class B stage will not be symmetrical, or in other words, $V_{CEQ2} \neq V_{CEQ3}$, and the final output will not be sitting at 0 VDC as it needs to. Knowing the value of R_3 and its voltage, we can determine its

current. This current flows down into $Q1$ as I_{CQ1} . Knowing I_{CQ1} , we can determine the voltage across R_4 determine the appropriate divider ratio for R_1 and R_2 to achieve this value.

Example 8.2.2

Using the two-stage amplifier of Figure 8.2.17, first determine values for R_1 and R_2 to obtain proper system bias. Determine the output compliance, maximum load power and worst case transistor dissipation. Also estimate A_v. Assume β = 50 for the output devices and 100 for the first stage. V_{CC} = 20 V, V_{EE} = –20 V, R_L = 16 Ω, R_3 = 560 Ω, R_4 = 75 Ω.

For the output section (assuming it will be biased properly), by inspection, V_{CEQ} = 20 V. This is the peak compliance.

$$compliance = 20Vpeak = 14.1VRMS$$

Given the compliance, we can use power law to find the load power

$$P_{load(max)} = \frac{Compliance_{RMS}{}^2}{R_L}$$

$$P_{load(max)} = \frac{(14.1V)^2}{16\Omega}$$

$$P_{load(max)} = 12.5W$$

The transistors' worst case power dissipation is

```
*** QuickLaTeX cannot compile formula:
\[P_D = \frac{P_{load (max)}{5} \]

*** Error message:
File ended while scanning use of \frac .
Emergency stop.
```

$$P_D = \frac{12.5W}{5}$$

$$P_D = 2.5W$$

To determine the biasing resistors, we start with R_3 . To achieve bias symmetry, all of V_{CC} drops across R_3 with the exception of 0.7 volts for D_1 . This is the same as I_{CQ1} .

$$I_{CQ_1} = \frac{V_{CC} - 0.7V}{R_3}$$

$$I_{CQ_1} = \frac{19.3V}{560\Omega}$$

$$I_{CQ_1} = 34.5mA$$

The drop across R4 is found via Ohm's law

$$V_{R_4} = I_{CQ_1} R_4$$

$$V_{R_4} = 34.5mA \times 75\Omega$$

$$V_{R_4} = 2.6V$$

This implies that the voltage across R_2 must be 0.7 volts more, or 3.3 volts. If we ignore the base current of Q_1, then the ratio of R_1 to R_2 must be the same as the ratio of their voltages, 36.7 to 3.3, or 11.1 to 1. In other words, R_1 must be 11.1 times larger than R_2 . For good bias stability we don't wish to set R_2 too much larger than R_4 . If we set it to 200 Ω, for example, then R_1 would need to be about 2.2 k Ω. Due to component tolerances, one of these resistors would need to be a potentiometer (connected as a rheostat) in order to "tweak" the final output to 0 VDC. A resistor/pot combo might be even better as it won't be so "touchy". For example, the 2.2 k Ω could be replaced with a series combination of a 1.8 k Ω and a 1 k Ω pot. This would be much easier to adjust than if the resistor was replaced with a standard 5 k Ω pot.

Now for the system voltage gain. The gain of the follower is approximately one so we need only concern ourselves with the first stage common emitter amplifier. This amplifier is swamped by R4 and given that the collector current is over 34 mA, r′e will be less than an ohm and can be ignored. All we need to do is find the effective load at the collector of Q1. This is R_3 in parallel with a single $Z_{in(base)}$ (remember that only one transistor is on at any given time and the off transistor will appear as a high impedance).

$$Z_{in(base)} = \beta r_E$$

$$Z_{in(base)} = 100 \times 75\Omega$$

$$Z_{in(base)} = 7500\Omega$$

$$A_v = -\frac{r_L}{r_E}$$

$$A_v = -\frac{7500\Omega||560\Omega}{75\Omega}$$

$$A_v = -6.95$$

Before leaving this section, there are a few items to note. First, the load power calculations have assumed that the entire power supply can be used by the output devices. As we saw with the diode bias version, this is not always the case. There is another situation that can limit the output swing. The class B stage is a follower and thus has a voltage gain of one. It the driver stage can't produce the full swing then the output stage can't either. Consequently, a class A analysis (i.e., AC load line) needs to be performed on the driver in order to determine just how large the signal can be before clipping. Indeed, it's quite likely that the driver stage will clip before the output stage.

Also, the direct coupled driver does not have to be an NPN as depicted in Figure 8.2.17. A PNP can be used instead, it simply needs to be shifted to the upper section rather than the lower section, as shown in Figure 8.2.18.

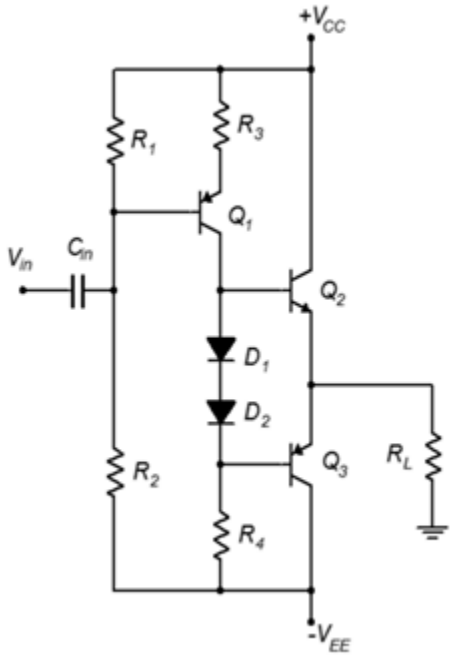


Figure 8.2.18: Class B amplifier with direct coupled driver, PNP version.

8.3 EXTENSIONS AND REFINEMENTS

The foregoing discussion has covered the basics of class B amplifier design and operation but there are a variety of things we might add.

CURRENT LIMITING

One of the more useful additions to an amplifier is some form of protection circuit. As noted earlier, there is nothing in the basic class B amplifier to limit current, so if the load is accidentally shorted, the transistor current will spike to very high levels and possibly damage the output transistors. How might we prevent this from happening?

Perhaps the most basic protection technique is to place a fuse in-line with the load. This is a somewhat tricky proposition because voice and music waveforms are very dynamic. Fuses are not fast response devices and selection of the proper current rating is a trade-off. If we want to catch large, fast transients, the fuse will have to be rated on the low side. Unfortunately, the fuse might then blow on moderately loud sustained low frequency content. Relays suffer from similar problems. Also, there is the issue that if the fuses are replaceable by the consumer, they may use the wrong value or some other item[1] that will allow too much current, resulting in blown output transistors. This can be avoided by having a second set of fuses mounted on the PC board, connected to the power rails, but this is also an imperfect solution.

One means of dealing with this situation is to employ an active current limiter, such as the circuit shown in Figure 8.3.1. In this circuit, Q_1 is the main NPN output transistor, the PNP side not shown. At the top is the power supply connection and the load is connected off to the right. A small resistor, R_E, is inserted into the emitter current path. The resistance is small enough that the voltage across it is normally less than 0.5 volts or so. Across the resistor is another transistor, Q_2. Under normal operation Q_2 is not on and does not affect the circuit. If the load current gets large enough (beyond the safe limit), the voltage drop across R_E will reach 0.7 V. At this point Q_2 turns on and begins to conduct current away from the base of Q_1, limiting the amount of load current to approximately $0.7V/R_E$. Unlike a fuse, once the fault is removed and the load current falls to safe levels, Q_2 disengages and normal amplifier function resumes.

1. Such as bits of wire, aluminum foil, small bolts or screws, etc. Just ask any seasoned repair technician.

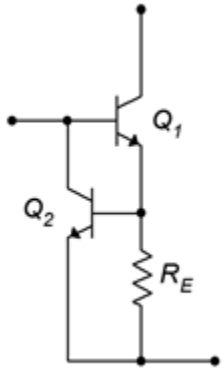


Figure 8.3.1: Active current limiter.

Two of these circuits are required for the amplifier; one for the positive half and one for the negative half. Figure 8.3.2 shows a class B amplifier with the added current protection circuits (within the dashed red box).

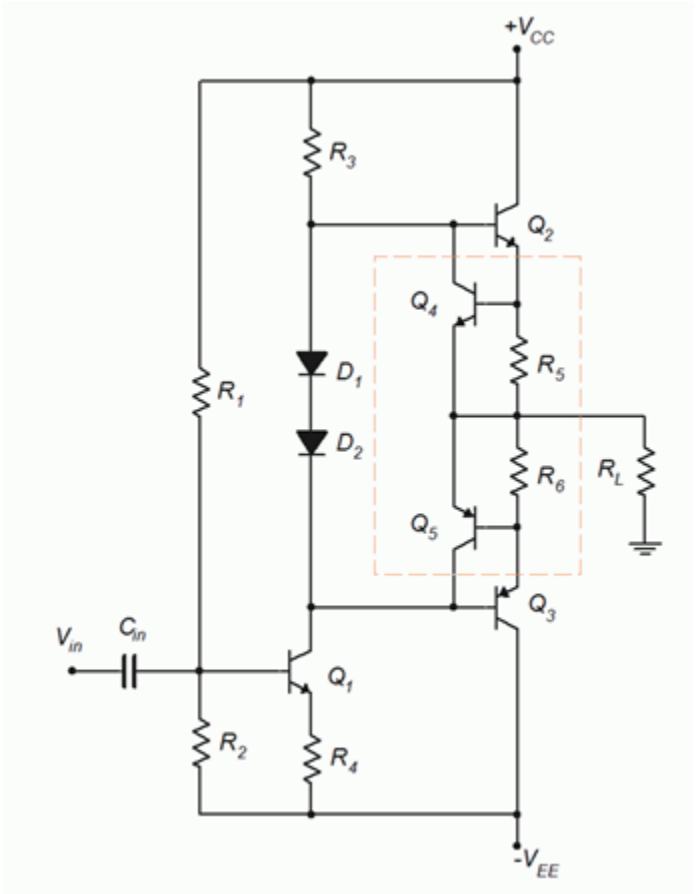


Figure 8.3.2: Class B amplifier with current limiters

HIGH CURRENT GAIN CONFIGURATIONS

For some applications the β of typical power transistors may prove to be insufficient. In such cases we can use high current gain configurations. There are several ways to configure the output devices. A Darlington scheme is shown in Figure 8.3.3. Q_3 and Q_5 are the main output devices. Q_2 and Q_4

can be thought of as drive transistors. Although their BV_{CEO} rating will need to be as high as that of Q_3/Q_5, they will be handling less current and therefore will dissipate less power. Sometimes Q_2/Q_4 are configured as seen here and sometimes emitter resistors may be added so that the output appears to be a cascade of emitter followers. Either way, there are now four base-emitter junctions to be compensated for, thus the addition of D_3 and D_4.

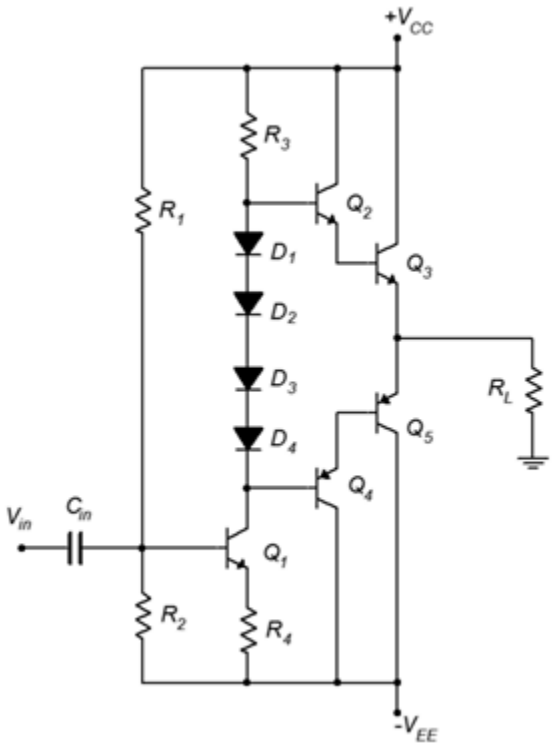


Figure 8.3.3: Darlington output devices for high current gain.

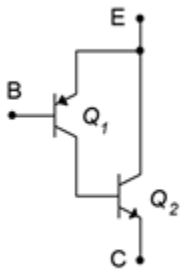


Figure 8.3.4: Sziklai pair, AKA composite PNP.

Some years ago, high power PNP transistors exhibiting decent audio quality were not available. Instead of using a PNP Darlington configuration, a Sziklai pair was used. This dual transistor configuration is named after Hungarian, and later, American, engineer George Sziklai. The Sziklai pair is also known as a composite pair. A composite PNP is shown in Figure 8.3.4. The operation is similar to that of a Darlington pair. The input transistor, Q_1, drives its collector current into the base of Q_2, the output transistor.

Q_2's base current is multiplied by its β, thus its collector current is the product of I_{B1} , β_1 and β_2, just like a Darlington. The differences are that the main power device for the composite PNP is

an NPN, and that there is only a single V_{BE} to compensate for. An example amplifier based on the earlier direct coupled driver circuit is shown in Figure 8.3.5. This configuration is known as a quasi-complementary output.[2] Note the use of three compensating diodes; two for the Darlington NPN and one for the composite PNP/Sziklai pair. One advantage here is that the power devices, Q_3 and Q_5, can be identical models.

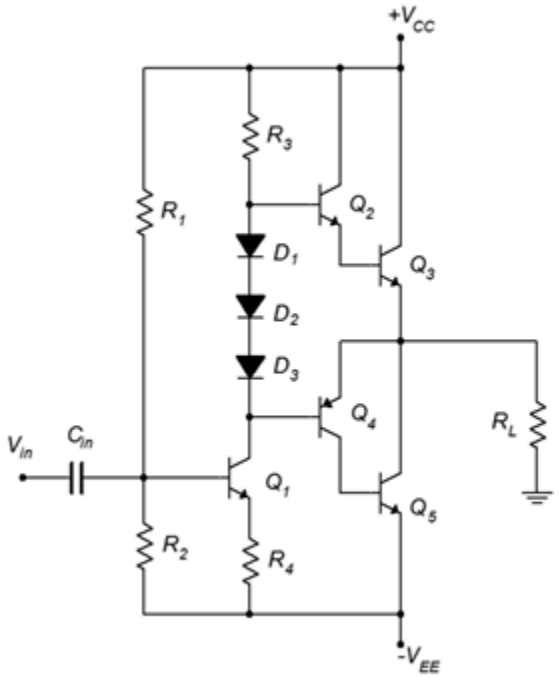


Figure 8.3.5: Quasi-complementary output.

The current limiting circuit of Figure 8.3.1 can be added to both the Darlington and quasi-complementary output stages. Current limiting can also be added to the circuits presented in the next section, but have been left off for clarity.

CURRENT SHARING

For higher output powers, and especially for amplifiers driving very low load impedances such 1 or 2 ohms, it may not be possible to use a single NPN and PNP for the output stage. Instead, the duties of each device will be shared among two or three transistors operating in parallel. Generally, current-sharing schemes will also require a Darlington-based approach due to the very high total load current.

Initially, it may be tempting to just place two or three transistors in parallel, each base terminal being fed by a common drive transistor. There is a fatal flaw with this approach. The issue is referred to as either thermal runaway or current hogging. The basic e problem is that the temperature coefficient of transconductance for a bipolar transistor is positive. In other words, r' gets smaller as temperature rises. This means that, as the device heats up, it tends to more easily conduct current. Of course, if the device draws more current, it will dissipate more power, which means that it will get hotter, which means that it will conduct more current, which means that it will dissipate more power, and so on. This process spirals out of control and is particularly evil when we have multiple devices in parallel. Devices are never perfectly matched so the load current will never be perfectly shared

2. Not to be confused with the quasi-complimentary output of which we have only partially nice things to say.

among the devices. This means that one device will tend to get hotter than the other(s) which leads to it getting even hotter and grabbing a larger and larger share of the total current. Eventually, this device "hogs" all of the available current and winds up destroying itself. What we have is a positive thermal feedback loop. Transistor harakiri is a decidedly sub-optimal performance characteristic. To circumvent this problem we can add small resistors to the emitters of the devices. As all of the paralleled devices are being driven from a common drive transistor, they all see the same base voltage. If one output transistor starts to grab a larger share of the load current, the voltage drop across the emitter resistor will increase, thus forcing a reduction in that transistor's base-emitter voltage. This reduction compensates for the initial tendency of the current to increase. This technique is referred to as local negative feedback. We have seen this concept before, for example, a swamping resistor falls into this category as does the collector feedback bias scheme.

An example of a current-sharing output section is shown in Figure 8.3.6. This version uses a Darlington scheme. The newly added transistors are shown in red while the thermal/current feedback resistors are shown in blue.

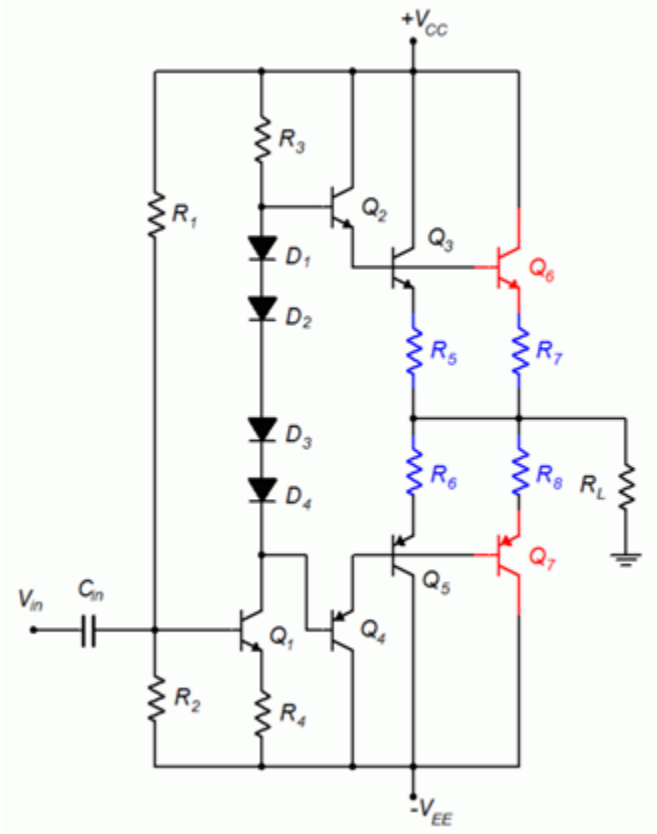


Figure 8.3.6: Current-sharing output section.

V_{BE} MULTIPLIER AND MILLER CAPACITOR

With the ever increasing complexity of the output section, the simple diode biasing scheme lacks the ability to set the optimal idle bias to achieve minimum crossover distortion. A more flexible approach is to use a V_{BE} multiplier, as illustrated in Figure 8.3.7.

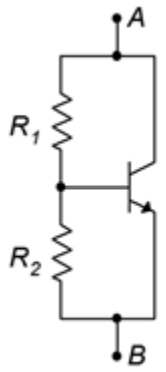


Figure 8.3.7: V_{BE} multiplier.

If we ignore base current, the currents through the two resistors are identical. Therefore, their voltages have the same ratio as their resistances. $R2$ is in parallel with V_{BE} so its voltage must be equal to V_{BE} . Consequently, the voltage across $R2$ must be a multiple of V_{BE} . For example, if we want to generate the equivalent of four base-emitter drops from point A to point B, we make R_1 three times as large as R_2 . What makes this particularly useful as that we can set any ratio we want, and by replacing either resistor with a potentiometer, we can make that voltage adjustable. And example is shown in Figure 8.3.8, using the current-sharing amplifier of Figure 8.3.6. The V_{BE} multiplier is shown in the dashed red box and replaces the four biasing diodes. A capacitor, C_B, shunts the multiplier, making sure it behaves as a short for AC signals.

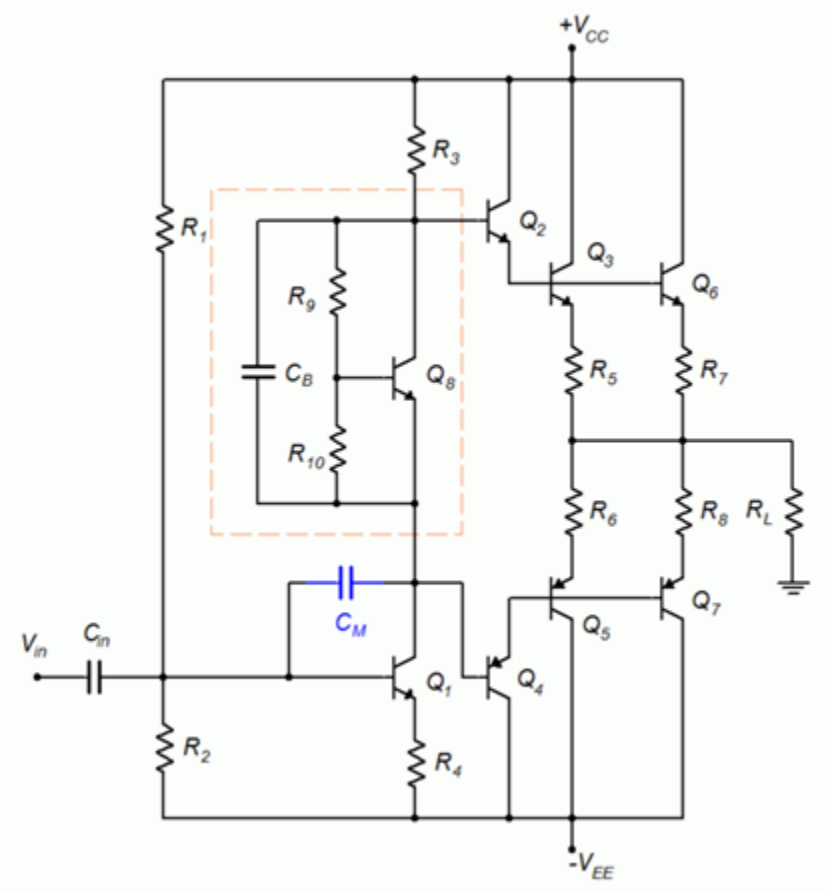


Figure 8.3.8: Amplifier with VBE multiplier and Miller compensation capacitor.

The circuit of Figure 8.3.8 also includes another capacitor, C_M . This is a Miller compensation capacitor. It uses the Miller Effect (see Chapter 6, Miller's Theorem) to create a much larger equivalent input capacitance. This capacitance appears in parallel with the input of Q_1 and creates a lag network. The inclusion of C_M allows the designer to tailor the high frequency response of the amplifier.

BRIDGING

Some amplifiers employ a bridged output scheme. This is particularly true where power supply voltages are limited, for example, in automobiles (nominally 12 V_{DC} but closer to 13.8 V_{DC} from the alternator). Without resorting to an expensive DC-to-DC converter, an amplifier designed for automotive use is in a bit of a bind. If we assume a +12 volt DC source, then a basic class B amplifier would use a capacitor coupled configuration like Figure 8.2.6. This would yield a V_{CEQ} of 6 volts and a compliance of about 4.2 volts RMS. If we were to use a standard home loudspeaker of 8 Ω, the maximum load power would be a mere 2.25 watts (this is one reason why car audio systems typically use 4 Ω loudspeakers, because it doubles the output power, in this case to 4.5 watts). A bridged output can double this again.

A block diagram of a bridged drive is shown in Figure 8.3.9. The left half is what we would see normally. On the right half we have a second identical amplifier but we drive it with an inverted copy of the original signal. What ends up happening is that, as the left output goes positive, the right output goes negative by the same amount. The end result is that the load sees twice the voltage it would have from the left amplifier alone. Power varies as the square of voltage so doubling the voltage quadruples the power. The 2.25 watt output of that car amplifier jumps up to 9 watts.[3] If we want to go higher than this our only options are to further decrease the loudspeaker impedance (there are practical limits that will not let us go much further) or increase the DC power supply. In the home, increasing the supply is relatively easy as we have an AC power source. In an automotive system this is a much more expensive proposition because only DC is generally available.

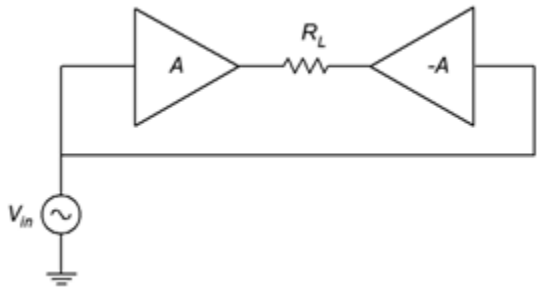


Figure 8.3.9: Block diagram of bridged amplifier.

An illustration of a bridged system at the transistor level is shown in Figure 8.3.10. For simplicity, this schematic uses basic class B stages. Input and right-side inversion circuitry are not shown.

3. OK, not huge, but probably still loud enough to obscure the sirens of emergency vehicles until they're right behind you.

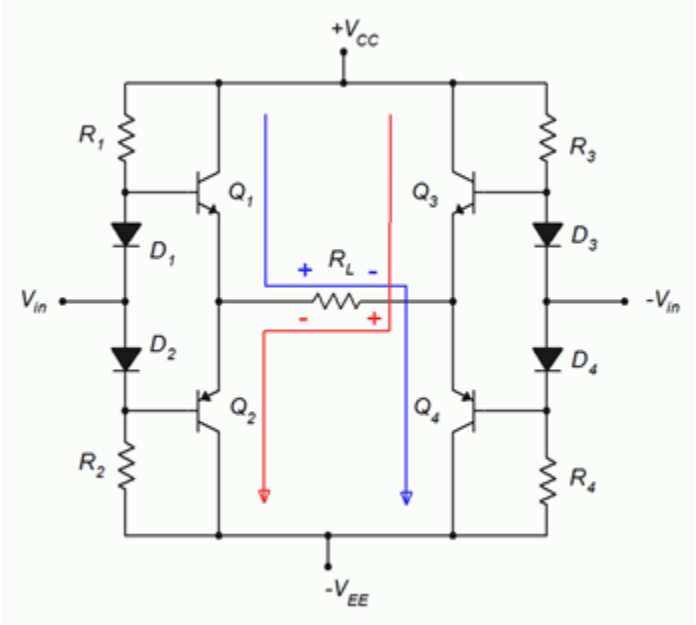


Figure 8.3.10: Bridged amplifier, transistor level.

The four transistors create a classic H bridge with the load in the center. For a positive V_{in}, Q_1 turns on. At the same time, because it is being fed by an inverted version of V_{in}, $Q4$ turns on. Thus, current flows down from V_{CC} , through $Q1$, across R_L from left-to-right, down $Q4$, and finally to V_{EE} (blue trace). At full swing, the voltage across the load will be $V_{CC} – V_{EE}$ from left to right.

For a negative V_{in} the opposite happens: Q_3 and Q_2 turn on, allowing current to flow through R_L from right left. The maximum voltage will be the same but with inverted polarity, hence an effective doubling of load voltage and a quadrupling of load power.

This increased output does not come without downsides. The most obvious problem is a doubling of the output circuitry. The second issue may not be immediately apparent: the load is not grounded. In simple automobile audio systems, the chassis of the vehicle is used as the system common or "ground". Therefore, it is possible to just run a single wire to a loudspeaker from an amplifier's output. The other loudspeaker terminal is then connected to the chassis at a convenient location. With a bridged output, two wires must be run to the loudspeaker. If a single-lead loudspeaker system is upgraded with a bridged output amplifier, new loudspeaker wires must be installed. Otherwise, the chassis return lead winds up shorting out one side of the bridge.

POWER SUPPLY BYPASSING

A final comment regarding high power amplifiers, particularly audio amplifiers, is in order. All through this chapter we have assumed that the DC power supplies will act ideally, mainly that they will present themselves as good AC grounds. This is not always easy to achieve given the practical limitations of PC board layouts, wiring constraints and so forth. Consequently, power supply bypass capacitors are often used to ensure a good AC ground. While this was mentioned in Chapter 7 regarding small signal amplifiers, it is perhaps more important for power amplifiers. The power supply bypass capacitors used for power amplifiers tend to be larger and the quality more stringent. A simple 1 μF bypass capacitor would hardly be the norm for a high power audio amplifier. The practical

problem is that large, high quality capacitors are not inexpensive. For example, a 10 μF polypropylene capacitor will be at least an order of magnitude more expensive than a similar size aluminum electrolytic (dollars versus cents). Unfortunately, the electrolytic will have higher leakage and ESR, and will not behave nearly as well at high frequencies (indeed, the impedance will actually start to increase due to inductive effects once the frequency gets high enough). To both improve performance and save money, bypass capacitors are sometimes doubled or tripled. For example, a large aluminum electrolytic might be placed in parallel with a much smaller polyester or polypropylene capacitor. The aluminum electrolytic will give the small X_C needed at lower frequencies, and when it starts to behave less ideally at higher frequencies, the higher quality poly capacitor effectively shunts it and extends the operating range. The result is nearly as effective as the single large, high quality capacitor but much less expensive.

8.4 SUMMARY

Class B operation is defined as having transistor collector current active for 180° out of the waveform cycle. In order to amplify the entire 360° in linear fashion, two devices are required. Each transistor of a complementary pair is biased at cutoff to achieve a 180° conduction angle. This means that no-signal collector current is zero, leading to very low power consumption at idle and, unlike class A amplifiers, dynamic power consumption. Unfortunately, pure class B operation also results in notch or crossover distortion as the switch over from one transistor to the other is not seamless. This can be mitigated by biasing the devices slightly "on", which is known as class AB operation. While simple resistor voltage dividers may be used for this purpose, a generally superior method uses diodes as they mimic the base-emitter current-voltage characteristic.

The compliance of a class B amplifier is based on its power supplies. Ideally, the peak-to-peak compliance of the amplifier will equal the total power supply differential. Worst case power dissipation and efficiency are far superior when compared to class A topology: Device power dissipation is only one-fifth of maximum load power and the theoretical efficiency at maximum load power is 78.5%. Maximum heating of the transistors occurs at approximately 40% of maximum load power.

A direct coupled drive stage is often used as it reduces parts count and improves performance. Amplifiers can also be extended through the use of Darlington pairs and current sharing schemes. Other refinements include the use of active current limiting for device protection, and V_{BE} multipliers in place of simple biasing diodes for greater flexibility.

Bridging is a technique used to increase output power. It relies on driving a floating load from both sides. Two amplifiers are needed for this configuration but it can offer a quadrupling of load power for the same power supplies.

Review Questions

1. Define class B operation and compare it to class A.
2. How do the AC and DC load lines differ between class A and class B operation?
3. What is the purpose of the biasing diodes in a class B amplifier?
4. Explain the source of notch distortion and discuss how it can be reduced.
5. What is thermal runaway? How might it be controlled?
6. What is a Sziklai pair?
7. What is a V_{BE} multiplier?

8. Explain the operation of an active current limiter.
9. What is a bridged output configuration? What are its benefits?
10. What's the difference between a complement and a compliment?

8.5 EXERCISES

ANALYSIS PROBLEMS

1. For the circuit of Figure 8.5.1, determine compliance, $P_{load(max)}$, $P_{D(max)}$, BV_{CEO} and $I_{C(max)}$.
   V_{CC} = 15 V, V_{EE} = –15 V, ß = 75, R_L = 16 Ω, R_1 = 680 Ω, R_2 = 680 Ω.
2. For the circuit of Figure 8.5.1, determine Z_{in}.
   V_{CC} = 15 V, V_{EE} = –15 V, ß = 75, R_L = 16 Ω, R_1 = 680 Ω, R_2 = 680 Ω.
3. For the circuit of Figure 8.5.1, determine Z_{in}.
   V_{CC} = 25 V, V_{EE} = –25 V, ß = 70, R_L = 8 Ω, R_1 = 560 Ω, R_2 = 560 Ω.

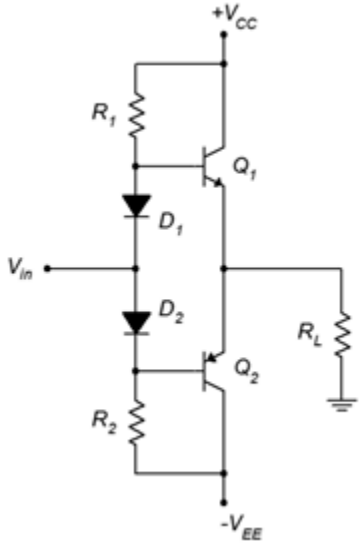


Figure 8.5.1

4. For the circuit of Figure 8.5.1, determine compliance $P_{load(max)}$, and $P_{D(max)}$, BV_{CEO} and $I_{C(max)}$.
   V_{CC} = 25 V, V_{EE} = –25 V, ß = 70, R_L = 8 Ω, R_1 = 560 Ω, R_2 = 560 Ω.
5. For the circuit of Figure 8.5.2, determine $P_{load(max)}$, and $P_{D(max)}$, BV_{CEO} and $I_{C(max)}$.
   V_{CC} = 15 V, ß = 75, R_L = 16 Ω, R_1 = 630 Ω, R_2 = 630 Ω.

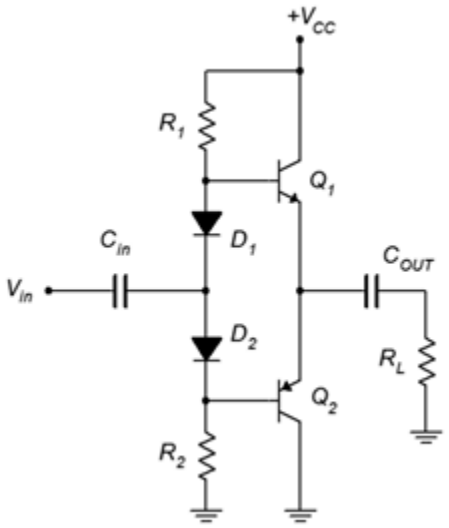


Figure 8.5.2

6. For the circuit of Figure 8.5.2, determine Z_{in}.
   V_{CC} = 15 V, ß = 75, R_L = 16 Ω, R_1 = 630 Ω, R_2 = 630 Ω.
7. For the circuit of Figure 8.5.2, determine Z_{in}.
   V_{CC} = 25 V, ß = 70, R_L = 8 Ω, R_1 = 560 Ω, R_2 = 560 Ω.
8. For the circuit of Figure 8.5.2, determine $P_{load(max)}$, and $P_{D(max)}$, BV_{CEO} and $I_{C(max)}$.
   V_{CC} = 25 V, ß = 70, R_L = 8 Ω, R_1 = 510 Ω, R_2 = 510 Ω.
9. For the circuit of Figure 8.5.3, determine $P_{load(max)}$, and $P_{D(max)}$, BV_{CEO} and $I_{C(max)}$ for the output transistors.
   V_{CC} = 24 V, V_{EE} = –24 V, ß = 75, R_L = 8 Ω, R_1 = 2.5 kΩ, R_2 = 300 Ω, R_3 = 330 Ω, R_4 = 63 Ω.

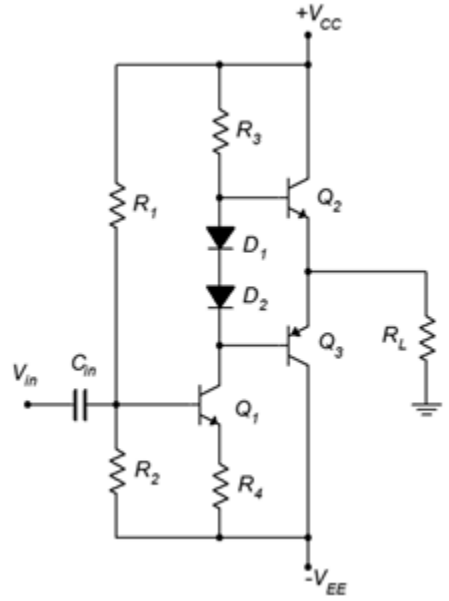


Figure 8.5.3

10. For the circuit of Figure 8.5.3, determine A_v and Z_{in}.
    V_{CC} = 24 V, V_{EE} = –24 V, ß = 75, R_L = 8 Ω, R_1 = 2.5 kΩ, R_2 = 300 Ω, R_3 = 330 Ω, R_4 = 63 Ω.

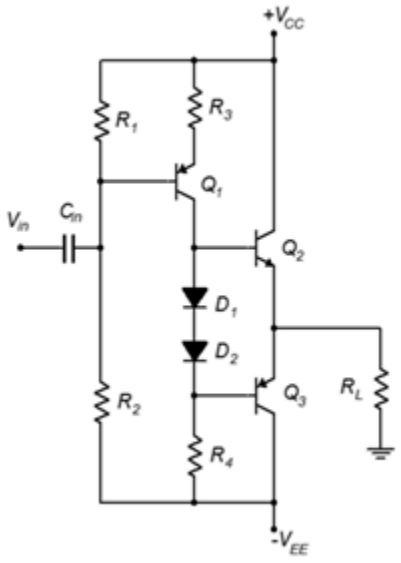


Figure 8.5.4

11. For the circuit of Figure 8.5.4, determine $P_{load(max)}$, and $P_{D(max)}$, BV_{CEO} and $I_{C(max)}$ for the output transistors.
V_{CC} = 24 V, V_{EE} = –24 V, ß = 75, R_L = 16 Ω, R_1 = 600 Ω, R_2 = 5 kΩ, R_3 = 63 Ω, R_4 = 330 Ω.
12. For the circuit of Figure 8.5.4, determine A_v and Z_{in}.
V_{CC} = 24 V, V_{EE} = –24 V, ß = 75, R_L = 16 Ω, R_1 = 600 Ω, R_2 = 5 kΩ, R_3 = 63 Ω, R_4 = 330 Ω.
13. Determine the limit current for the circuit of Figure 8.5.5 if R_E = 0.2 Ω.

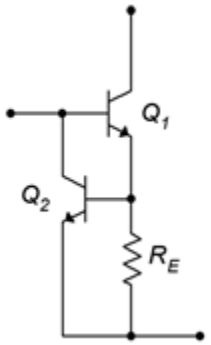


Figure 8.5.5

14. Determine $P_{load(max)}$, and $P_{D(max)}$, BV_{CEO} and $I_{C(max)}$ for the output and driver transistors of Figure 8.5.6.
V_{CC} = 50 V, V_{EE} = –50 V, ß = 75, R_L = 8 Ω, R_5 through R_8 = 0.05 Ω. Assume all other components produce proper bias.

DESIGN PROBLEMS

15. For the circuit of Figure 8.5.3, determine values for R_1 and R_2 for proper bias.
V_{CC} = 32 V, V_{EE} = –32 V, ß = 75, R_L = 8 Ω, R_3 = 330 Ω, R_4 = 63 Ω.
16. Determine a value for *RE* to set the limit current for the circuit of Figure 8.5.5 to 2 A.

CHALLENGE PROBLEMS

17. For the circuit of Figure 8.5.6, determine values for R_1 and R_2 for proper bias.
V_{CC} = 50 V, V_{EE} = –50 V, ß = 85, R_L = 8 Ω, R_5 through R_8 = 0.05 Ω, R_3 = 2.2 kΩ, R_4 = 330 Ω.

18. For the circuit of Figure 8.5.7, determine a value for R_5 for proper bias.
V_{CC} = 30 V, V_{EE} = –30 V, ß = 100, R_L = 16 Ω, R_1 = 2.2 kΩ, R_2 = 8.2 kΩ, R_3 = 1.2 kΩ, R_4 = 47 Ω, R_5 = 330 Ω, R_6 = 470 Ω, R_7 = 68 Ω.

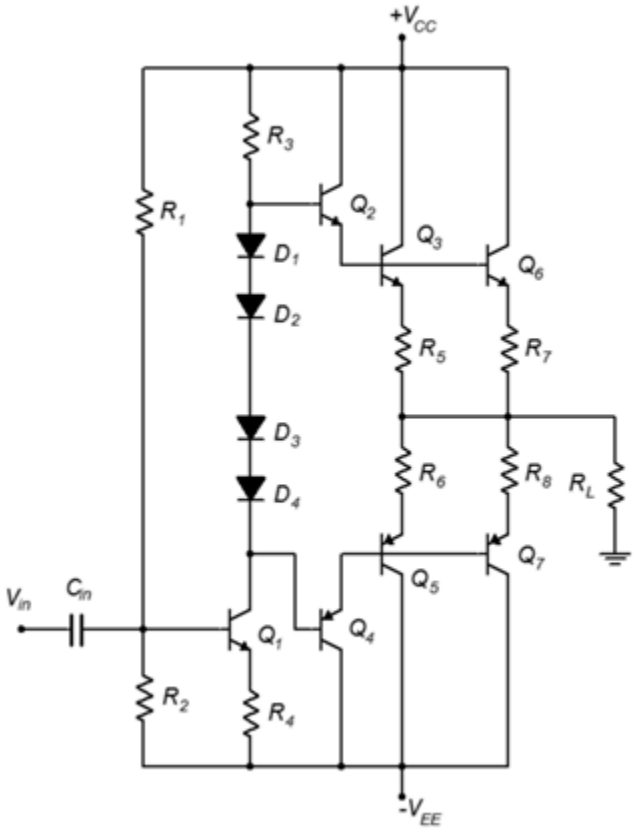


Figure 8.5.6

COMPUTER SIMULATION PROBLEMS

19. Perform a transient analysis on the circuit of Problem 1 to verify the compliance.
20. Perform a transient analysis on the circuit of Problem 4 to verify the compliance.
21. Perform a DC analysis on the design from Problem 15 to verify the results.
22. Perform a DC analysis on the design from Problem 17 to verify the results.

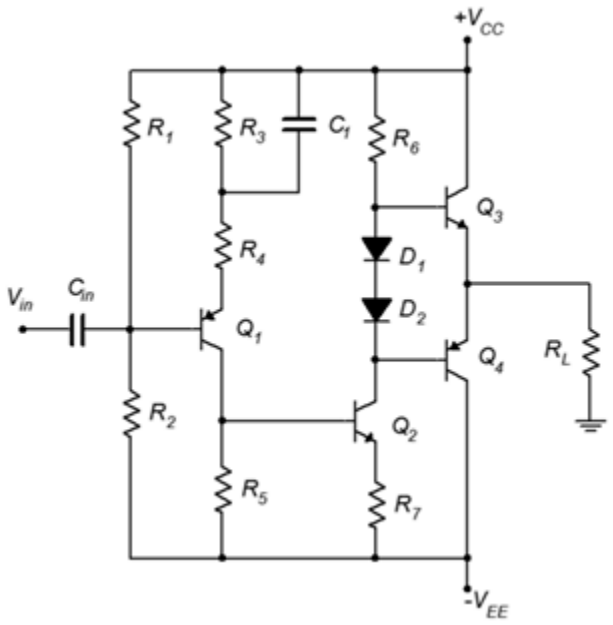


Figure 8.5.7

VERSION HISTORY

The NSCC edition of Semiconductor Devices: Theory and Application is a condensed adaptation of Semiconductor Devices: Theory and Application by James M. Fiore, Mohawk Valley Community College used under a CC BY-NC-SA license.

NSCC Edition	Semiconductor Devices: Theory and Application
Chapter 1	Chapter 2
Chapter 2	Chapter 3
Chapter 3	Chapter 4
Chapter 4	Chapter 5
Chapter 5	Chapter 6
Chapter 6	Chapter 7
Chapter 7	Chapter 8
Chapter 8	Chapter 9