Avionics II

AVIONICS II

JAMES FIORE

NSCC

Nova Scotia

CONTENTS

UNIT 1: JUNCTION FIELD EFFECT TRANSISTORS (JFETS)

Learning Objectives

After completing this chapter, you should be able to:

- Explain the basic operation of the JFET, detailing the different operating regions.
- Draw and explain a basic DC bias model for a JFET.
- Explain the terms transconductance and pinch-off voltage.
- Draw a JFET characteristic curve, detailing pertinent elements and points.
- Discuss the advantages and disadvantages of various JFET biasing topologies.
- Analyze basic DC biasing circuits for JFETs.
- Compare and contrast JFET operation to that of BJTs.

1.1 INTRODUCTION

The field effect transistor, or FET, is a semiconductor device that serves as an alternative to the bipolar junction transistor. FETs are available in two broad types: the junction FET, or JFET, and the metal oxide semiconductor FET, or MOSFET.

It is best not to think of FETs as either better or worse than the BJT. They have different characteristics and lend themselves to applications where BJT performance might be wanting. The inverse is also true and for some applications the judicious use of a combination of BJTs and FETs can produce superior performance when compared to either device used alone. Like the BJT's NPN and PNP variants, FETs comes in two "flavors": the Nchannel type and the P-channel type. We shall cover JFETs first and then discuss MOSFETs in subsequent chapters. In this chapter we will cover the internal structure of the JFET, its theory of operation and biasing techniques. In the next chapter we shall discuss small signal JFET amplifiers; both voltage amplifiers and voltage followers.

The JFET is fundamentally different from a bipolar junction transistor. While the JFET, like the BJT, relies on the PN junction for operation, the JFET is modeled as a voltage-controlled current source while the BJT is modeled as a current-controlled current source. Further, the BJT relies on a forward-biased base-emitter junction for proper operation while the JFET achieves current control via a reverse-biased junction. Consequently, JFET biasing circuits tend to be incompatible with BJT biasing schemes and one device cannot be swapped out for the other.

1.2 JFET INTERNALS

A simplified internal model of a JFET is shown in Figure 1.2.1. The main portion of the device is called the channel. The diagram illustrates an N-channel device. The channel is built upon a substrate (i.e., base layer) of oppositely doped material. Attached to the opposing ends of the channel are two terminals; the source and the drain. Embedded within the channel is a region using the opposite material type. A lead is attached to this as well and is called the gate. Although there is not perfect correspondence between them, the drain, source and gate are roughly analogous to the BJT's collector, emitter and base, respectively.

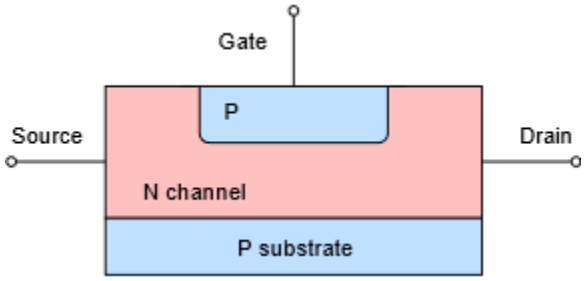


Figure 1.2.1 : Simplified internal structure of an N-channel JFET.

This diagram is drawn symmetrically. Some devices are designed in this fashion and their drain and source terminals can be swapped with no change in operation. This is not true for all devices, though. For small values of drain-source voltage, the channel exhibits a certain amount of resistance that is dependent on the doping level and physical layout of the device. Further, under normal operation, ◈◈ will equal ◈◈.

To understand how the device behaves, refer to Figure 1.2.2. Here we shall consider electron flow (shown as a dashed line). First, a positive voltage, ◈◈◈, is attached to the drain terminal along with a current limiting resistor, ◈◈. A negative supply, ◈◈◈, is applied to the gate terminal via resistor ◈◈. Let's start with the gate supply set to zero. If we start ◈◈◈ at zero, here is what happens as we increase its value. Initially, an increase in drain-source voltage will elicit a proportional increase in the current flowing through the channel. In other words, the channel acts like a resistor. As the voltage across the drain-source increases further, at some point the current will saturate, and no further increases in current will occur in spite of further increases in ◈◈◈ and ◈◈◈. At this point the device is behaving as a constant current source. The drain-source voltage where this transition occurs is called the pinch-off voltage, ◈◈. If the drain-source voltage increases too much, breakdown will occur and current will begin to increase rapidly.

What's particularly interesting is what happens when the gate supply is increased in the negative direction. This reverse-biases the gate-source PN junction and results in a larger depletion region being formed. The depletion region widens into the channel, thus restricting current flow sooner and at a lower level. The more negative we make ◈◈◈, the lower ◈◈ becomes. Eventually, when ◈◈◈ goes negative enough, the drain current will turn off. This voltage is called ◈◈◈(◈◈◈) and it has the

same magnitude as ◇◇ (i.e., ◇◇=|◇◇◇(◇◇◇)|). The action can be thought of as operating like a water valve: turning the gate source voltage more negative is like turning off the spigot and decreasing the flow.

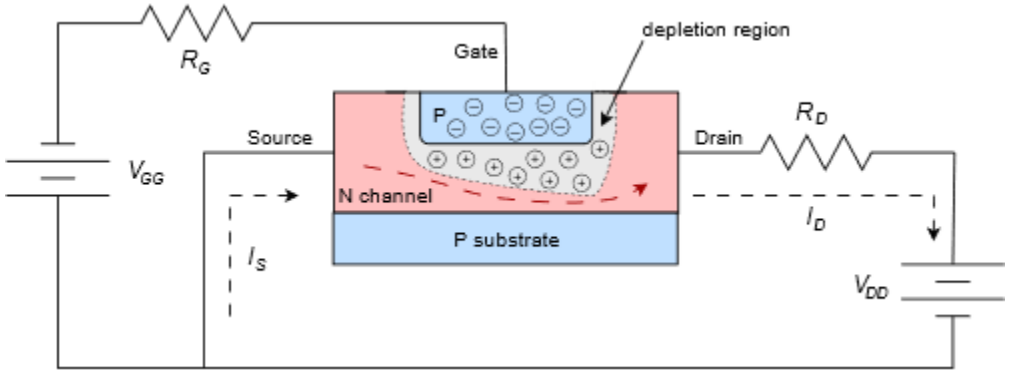


Figure 1.2.2 : Electron flow in an N-channel JFET.

The operation of the JFET can visualized nicely by plotting a set of drain curves, as shown in Figure 1.2.3.

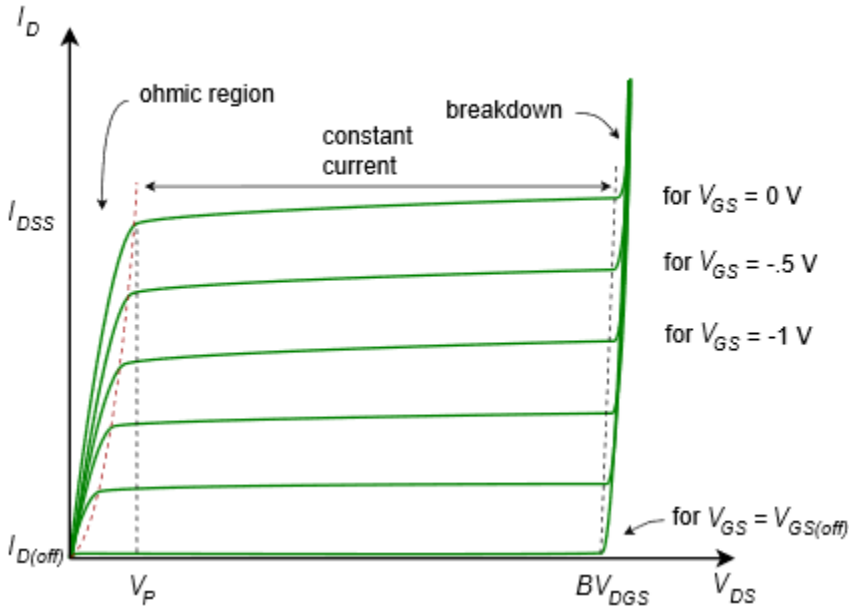


Figure 1.2.3: JFET drain curves.

The drain curve family plots drain current, ◇◇, versus drain-source voltage, ◇◇◇. We begin with the top-most curve. This is generated by setting the gate-source voltage, ◇◇◇, to zero. We then cycle ◇◇◇ from zero to some higher value. Initially, we see a proportional rise in ◇◇ as ◇◇◇ increases. This is called the ohmic or triode region. Eventually, the channel saturates and the current levels out. This is the constant current or saturation region and it occurs for ◇◇◇>◇◇. The breakdown voltage is called ◇◇◇◇◇, or alternately, ◇(◇◇)◇◇. Above this voltage the current increases rapidly. As usual, we do not wish to operate the device in this breakdown region.

If we now repeat the process but this time use a small negative value for ◇◇◇, we will trace out a curve of very similar shape. The transition to constant current mode will happen at a slightly lower voltage and the current value will be somewhat lower as well. This process continues in like fashion as we make ◇◇◇ more and more negative. Eventually, when ◇◇◇=◇◇◇(◇◇◇), the drain current

drops to virtually zero (in fact, a small leakage current flows called ◇◇(◇◇◇)). In contrast, if ◇◇◇ was allowed to go positive, operation would be lost because the PN junction would become forward-biased and we would lose control of the current via the depletion region. This means that the JFET's current control is entirely in the second quadrant and the largest drain current flows when ◇◇◇=0 V. This current is called ◇◇◇◇, which stands for the drain current with a shorted gate-source (i.e, if it's shorted, then ◇◇◇=0 V). The JFET cannot produce a continuous current larger than ◇◇◇◇ safely.

The characteristic equation relating drain current and gate-source voltage is shown below. This is valid for the constant current region (i.e., ◇◇◇>◇◇).

$$I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_{GS(off)}}\right)^2$$

(1.2.1)

Where

◇◇◇ is the gate-source voltage (◇◇◇(◇◇◇)≤◇◇◇≤0),

◇◇ is the drain current,

◇◇◇◇ is the maximum current,

◇◇◇(◇◇◇) is the turn-off voltage.

From this we see that the JFET is a square-law device rather than like the BJT which has a logarithmic characteristic.[1] In essence, this curve is a portion of a parabola. This means that the JFET's characteristic curve is much more gradual in slope than that of a BJT. This will have important implications when it comes to voltage gain potential and distortion, as we shall see in the following chapter.

It is useful to remember that ◇◇◇(◇◇◇) and ◇◇◇◇ are unique to a given device, rather like ◇ is for a BJT. There can also be a fairly large variation in these parameters. For example, a particular model of JFET might show an ◇◇◇◇ variation between 2 mA and 20 mA, and a ◇◇◇(◇◇◇) variation between –2 V and –8 V. Generally, the most negative ◇◇◇(◇◇◇) values will be associated with the largest ◇◇◇◇ values.

Equation 1.2.1 is plotted in Figure 1.2.4. Compare this curve to the curve generated by the Shockley equation for BJTs, Figure 7.2.1. The graph is shown in normalized form. Instead of plotting for specific values of ◇◇◇(◇◇◇) and ◇◇◇◇, the axes are presented as fractional portions of the maximums (i.e., the horizontal axis is –◇◇◇/◇◇◇(◇◇◇) and the vertical axis is ◇◇/◇◇◇◇).

1. As evidenced in the Shockley equation, Equation 2.1.1.

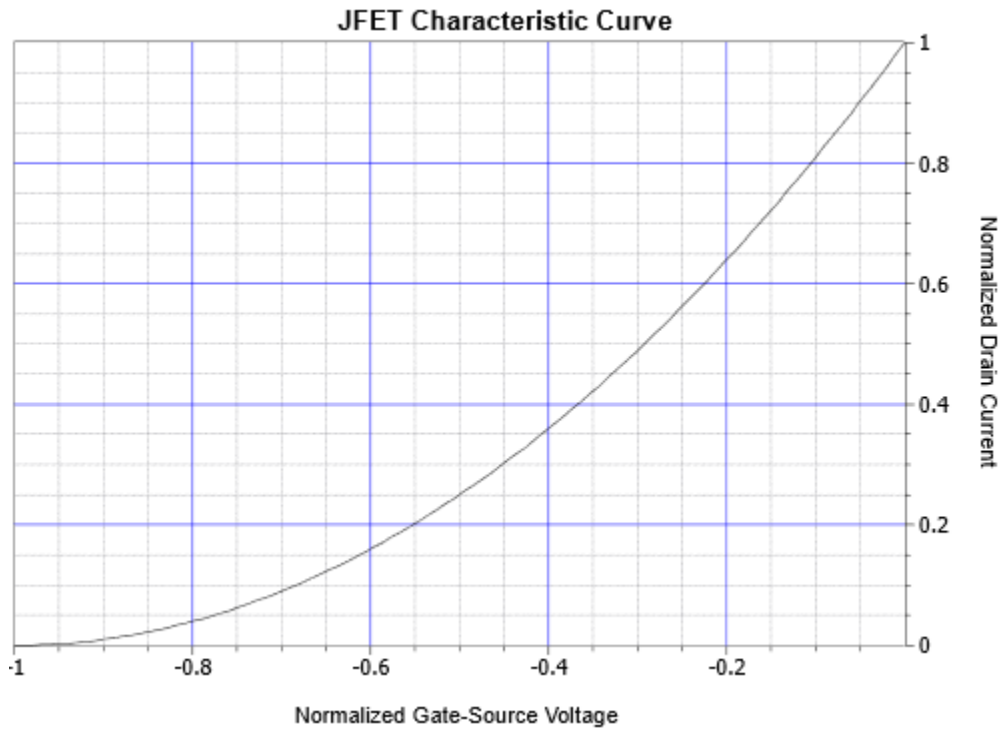


Figure 1.2.4: JFET normalized characteristic curve (note: this uses –◇◇/◇◇(◇◇◇) for the normalized voltage so that the curve does not appear reversed compared to a typical device curve).

Example 1.2.1

Using both Equation 1.2.1 and the graph of Figure 1.2.4, determine the drain current if the gate-source voltage is –1 V and the JFET specs are ◇◇◇ = 8 mA and ◇◇(◇◇◇) = –2 V.

First, using Equation 1.2.1

$$I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_{GS(off)}}\right)^2$$

$$I_D = 8mA\left(1 - \frac{-1V}{-2V}\right)^2$$

$$I_D = 2mA$$

Using the graph, ◇◇/◇◇(◇◇◇)is 1 V/–2 V, or –0.5. Find this value on the horizontal axis, follow up to the curve and then across to the right vertical axis. The normalized drain current is 0.25, thus ◇◇ is 0.25 ◇◇◇, or 2 mA.

As the characteristic curve plots output current versus input voltage, the slope of this represents the transconductance, an important characteristic for biasing and signal analysis. Device transconductance is denoted as ◇◇, or alternately as ◇◇◇, and given units of siemens. We can derive an equation for transconductance by taking the derivative of Equation 1.2.1.

$$I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_{GS(off)}}\right)^2$$

$$\frac{dI_D}{dV_{GS}} = -\frac{2I_{DSS}}{V_{GS(off)}}\left(1 - \frac{V_{GS}}{V_{GS(off)}}\right)$$

The coefficient –2I_{DSS}/$V_{GS(off)}$ is defined as g_{m0}, the transconductance when V_{GS}=0V. This is the maximum transconductance of the device. Substituting, we arrive at

$$g_{m0} = -\frac{2I_{DSS}}{V_{GS(off)}}$$

(1.2.2)

$$g_m = g_{m0}\left(1\frac{V_{GS}}{V_{GS(off)}}\right)$$

(1.2.3)

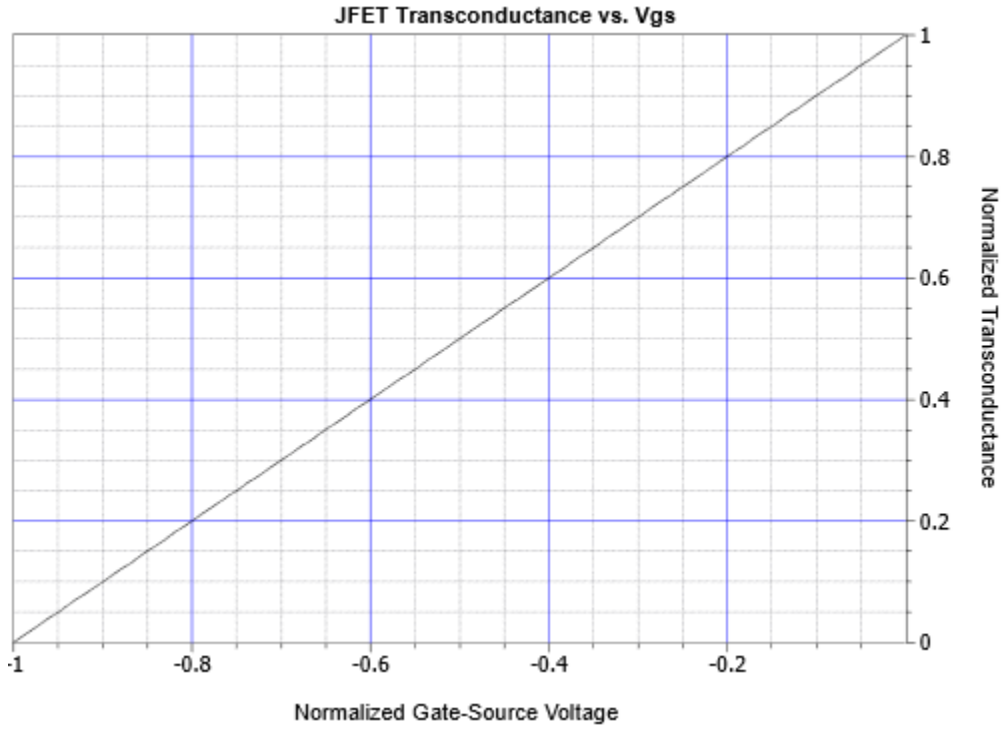


Figure 1.2.5: Curve of transconductance.

A normalized plot of transconductance versus V_{GS} is shown in Figure 1.2.5. The horizontal axis is –V_{GS}/$V_{GS(off)}$ and the vertical axis is g_m/g_{m0}.

From this graph we see that the transconductance is a linear function.

Another item of interest regarding these device equations: If we combine Equations 1.2.1 and 1.2.3, we generate two equations that will prove useful in upcoming work.

$$\frac{g_m}{g_{m0}} = \sqrt{\frac{I_D}{I_{DSS}}}$$

(1.2.4)

$$\frac{g_m}{g_{m0}} = \sqrt{\frac{I_D}{I_{DSS}}}$$

(1.2.5)

Before moving on, the schematic symbols for JFETs are shown in Figure 1.2.6. The middle vertical line represents the channel, and as is usually the case, the arrow points to N material. Sometimes the gate arrow is draw in the middle rather than toward the source. Also, as is the case the BJT, sometimes these symbols are drawn within a circle.

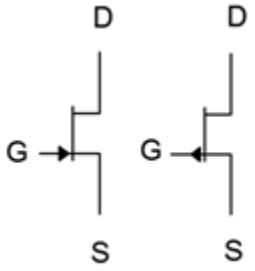
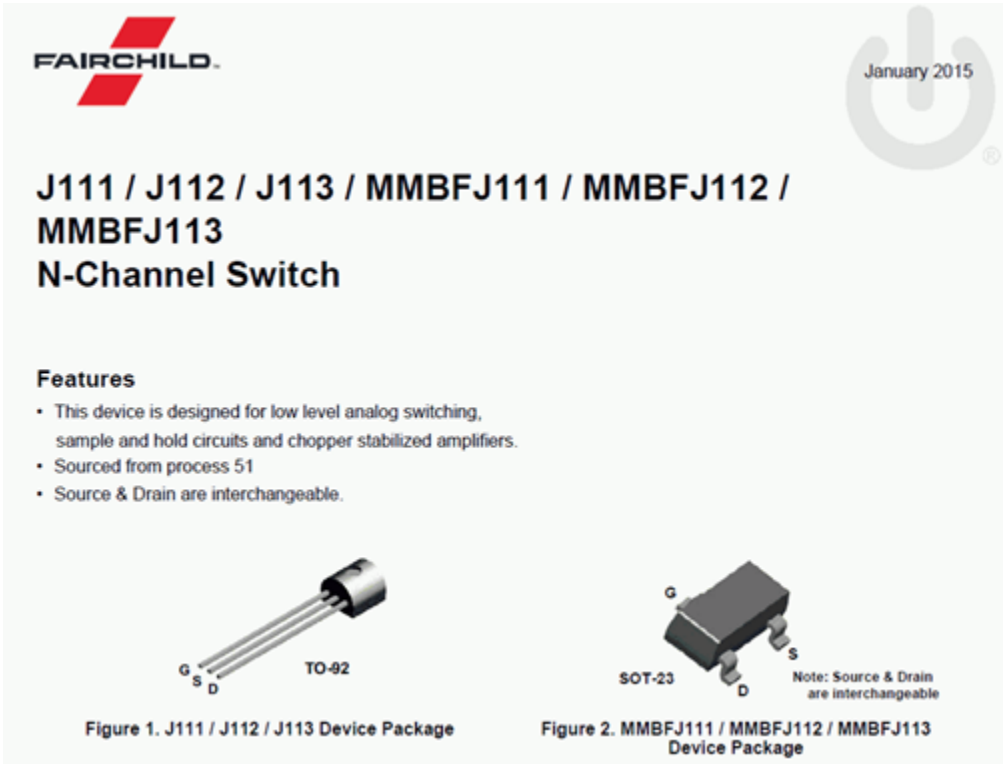


Figure 1.2.6 : JFET schematic symbols: N-Channel (left) P-Channel (right)

1.3 JFET DATA SHEET INTERPRETATION

A data sheet for the J111 series N-channel JFET is shown in Figure 1.3.1. This is a small signal device designed for audio frequency circuits. It is available in the common TO-92 through-hole package as well as in the surface mount SOT-23 package. Note that the source and drain are interchangeable for this device.



FAIRCHILD

January 2015

J111 / J112 / J113 / MMBFJ111 / MMBFJ112 / MMBFJ113
N-Channel Switch

Features

- This device is designed for low level analog switching, sample and hold circuits and chopper stabilized amplifiers.
- Sourced from process 51
- Source & Drain are interchangeable.

Figure 1. J111 / J112 / J113 Device Package

Figure 2. MMBFJ111 / MMBFJ112 / MMBFJ113 Device Package

Figure 1.3.1 : J111 series N-channel JFET data sheet. Used with permission from SCILLC dba ON Semiconductor.

Examining the absolute maximum ratings and thermal characteristics, we find values typical of small signal devices. Maximum drain-gate and gate-source voltages are 35 volts and the maximum power dissipation is 625 milliwatts.

Absolute Maximum Ratings[1], [2]

Stresses exceeding the absolute maximum ratings may damage the device. The device may not function or be operable above the recommended operating conditions and stressing the parts to these levels is not recommended. In addition, extended exposure to stresses above the recommended operating conditions may affect device reliability. The absolute maximum ratings are stress ratings only. Values are at T_A = 25°C unless otherwise noted.

Symbol	Parameter	Value	Unit
V_{DG}	Drain-Gate Voltage	35	V
V_{GS}	Gate-Source Voltage	-35	V
I_{GF}	Forward Gate Current	50	mA
T_J, T_{STG}	Operating and Storage Junction Temperature Range	-55 to 150	°C

Notes:

1. These ratings are based on a maximum junction temperature of 150°C.
2. These are steady-state limits. Fairchild Semiconductor should be consulted on applications involving pulsed or low-duty-cycle operations.

Thermal Characteristics

Values are at T_A = 25°C unless otherwise noted.

Symbol	Parameter	Max.		Unit
		J111 / J112 / J113[3]	MMBFJ111 / MMBFJ112 / MMBFJ113[4]	
P_D	Total Device Dissipation	625	350	mW
	Derate Above 25°C	5.0	2.8	mW/°C
$R_{\theta JC}$	Thermal Resistance, Junction-to-Case	125		°C/W
$R_{\theta JA}$	Thermal Resistance, Junction-to-Ambient	200	357	°C/W

Notes:

3. PCB size: FR-4, 76 mm x 114 mm x 1.57 mm (3.0 inch x 4.5 inch x 0.062 inch) with minimum land pattern size.
4. Device mounted on FR-4 PCB 36mm × 18mm × 1.5mm; mounting pad for the collector lead minimum 6cm^2.

Figure 1.3.1 : J111 series N-channel JFET data sheet (cont).

Electrical Characteristics

Values are at T_A = 25°C unless otherwise noted.

Symbol	Parameter	Conditions		Min.	Max.	Unit
Off Characteristics						
$V_{(BR)GSS}$	Gate-Source Breakdown Voltage	I_G = -1.0 µA, V_{DS} = 0		-35		V
I_{GSS}	Gate Reverse Current	V_{GS} = -15 V, V_{DS} = 0			-1.0	nA
V_{GS}(off)	Gate-Source Cut-Off Voltage	V_{DS} = 15 V, I_D = 1.0 µA	**111**	-3.0	-10.0	V
			112	-1.0	-5.0	
			113	-0.5	-3.0	
I_D(off)	Drain Cutoff Leakage Current	V_{DS} = 5.0 V, V_{GS} = -10 V			1.0	nA
On Characteristics						
I_{DSS}	Zero-Gate Voltage Drain Current[5]	V_{DS} = 15 V, V_{GS} = 0	**111**	20		mA
			112	5.0		
			113	2.0		
r_{DS}(on)	Drain-Source On Resistance	V_{DS} ≤ 0.1 V, V_{GS} = 0	**111**		30	Ω
			112		50	
			113		100	

Figure 1.3.1 : J111 series N-channel JFET data sheet (cont).

From the electrical characteristics, note the large variation in VGS(off)"> ( ). For the J111, this runs from a minimum of –3 V to a maximum of –10 V. The J112 and J113 exhibit even wider min/max ratios. Also, note how the larger VGS(off)"> ( ) ranges are associated with larger

maximums for IDSS">◇◇◇. Finally, let's take a look at a series of performance curves shown in Figure 10.3.1d">1.3.1◇.

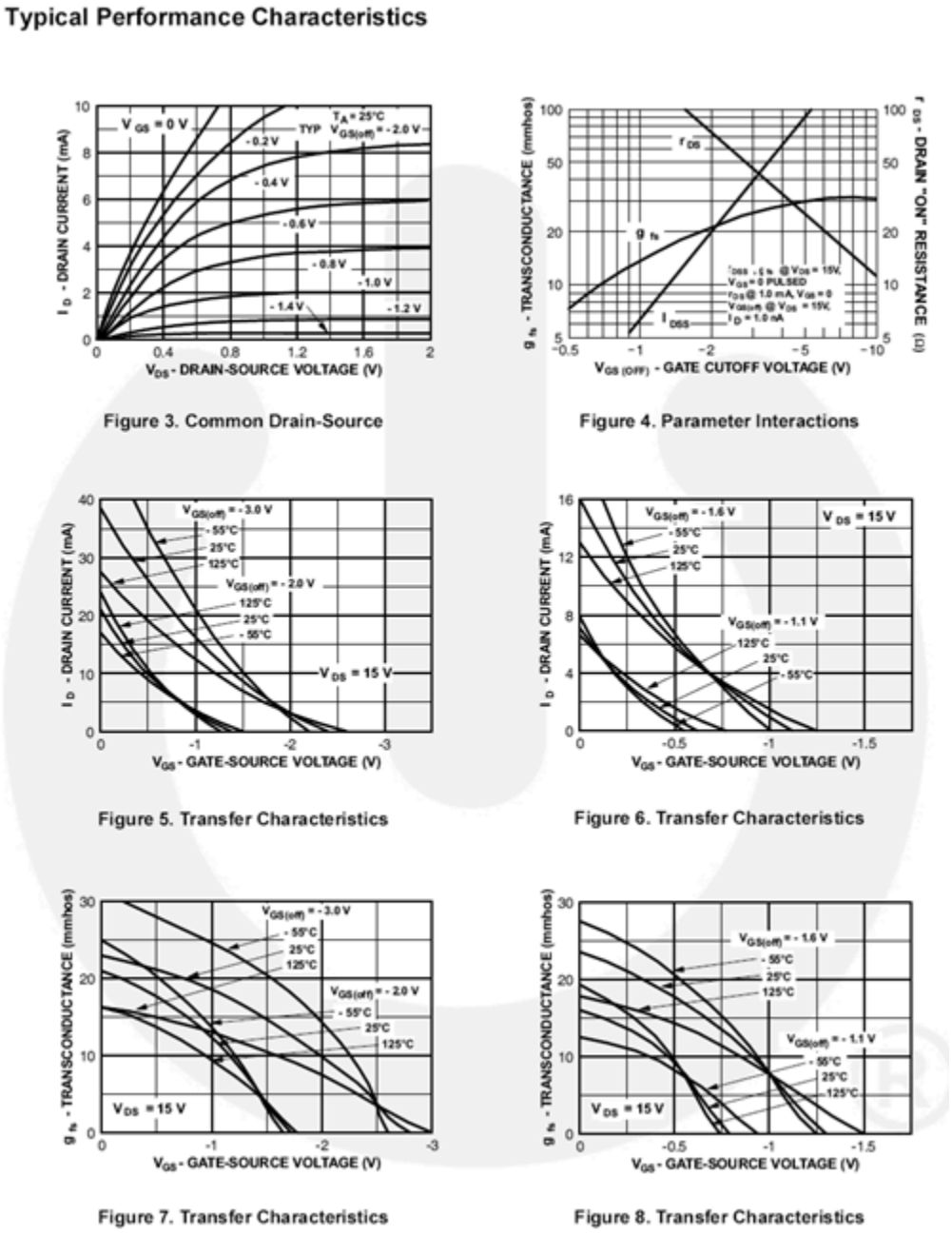


Figure 1.3.1◇: J111 series N-channel JFET data sheet (cont).

The upper-left graph is a family of drain curves and corresponds to Figure 1.2.3, presented earlier. Note that the horizontal scale for ◇◇◇ only shows the first 2 volts. The point here is to examine the ohmic region. The two middle graphs plot the characteristic curve of the device and correspond to Figure 1.2.4, although these graphs are drawn rotated around the vertical axis (note that ◇◇◇ is still shown as a negative value). Two important things may be noted here. First, as already mentioned, large values of ◇◇◇ tend to be associated with large values of ◇◇◇(◇◇◇). This graph shows that individual plots tend to scale both horizontally and vertically away from the origin. Second, thermal

variations are very much apparent: As the temperature increases, the characteristic curve tends to become less steep.

Finally, the two bottom-most graphs plot the variation of ◈◈ with ◈◈◈. These correspond to Figure 1.2.5, although again, the horizontal axis has been rotated around the vertical. Once again we see considerable variation due to temperature. Also, none of the plots exhibit perfect linearity. Further, at lower temperatures, the linearity of the plots decreases even more, warping a relatively straight line into a complex curve.

1.4 JFET BIASING

There are several different ways of biasing a JFET. For many configurations, I_{DSS} and $V_{GS(OFF)}$ will be needed. A simple way to measure these parameters in the lab is shown in Figure 1.4.1. To measure I_{DSS} we simply ground the gate and source terminals as this forces V_{GS} to be 0 V. We insert an ammeter between V_{DD} and the drain, and then set V_{DD} to a value higher than V_P (+15 V_{DC} generally being sufficient). The resulting ammeter reading is I_{DSS}. Obtaining $V_{GS(OFF)}$ is only slightly more work. Leaving the ammeter in the drain, unhook the gate from ground and instead connect it to an adjustable negative power supply. Turn the supply more negative until the ammeter reads zero (practically speaking, < 1% of I_{DSS}). At that point the voltage source will be equal to $V_{GS(OFF)}$.

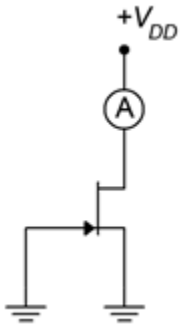


Figure 1.4.1: Measuring I_{DSS} and $V_{GS(OFF)}$.

DC MODEL

Before we begin examining the bias circuits themselves, we need a basic DC model of the JFET. A model sufficient for our analyses is shown in Figure 1.4.2.

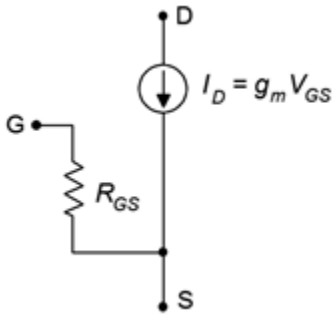


Figure 1.4.2: DC model of JFET.

The model consists of a voltage-controlled current source, I_D, that is equal to the product of the gate-source voltage, V_{GS}, and the transconductance, g_m. The resistance between the gate and source, R_{GS}, is that of the reverse-biased P_N junction, in other words, ideally infinity for DC. As

a consequence, in most practical circuits we can assume that gate current, I_G, is zero. Therefore, $I_R = I_D$.

CONSTANT VOLTAGE BIAS

The simplest form of bias is the constant voltage bias. The prototype is shown in Figure 1.4.3 with current directions and voltage polarities shown.

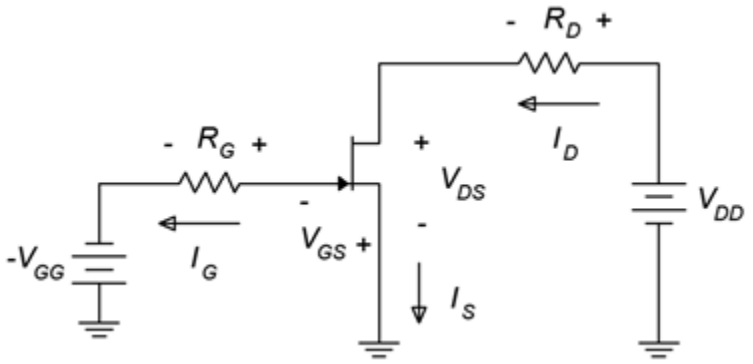


Figure 1.4.3: Constant voltage bias prototype.

This is a fairly straightforward design using only a couple of resistors and power sources. Figure 1.4.4 shows the same circuit but with the JFET model inserted, ready for analysis.

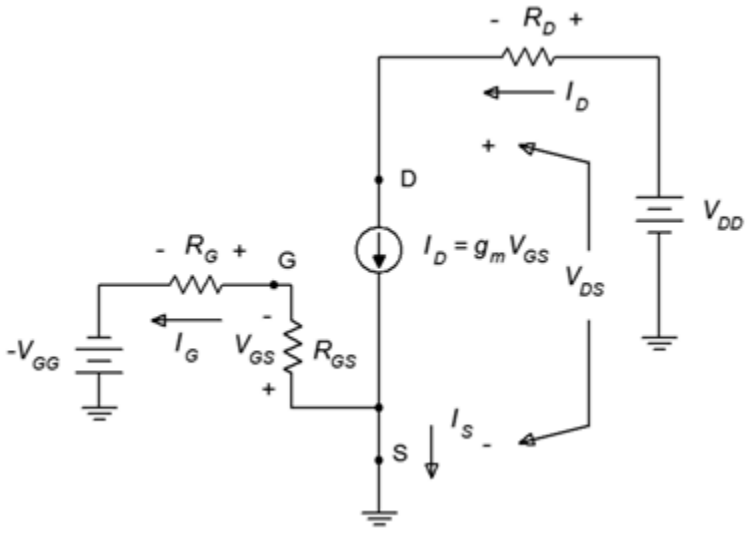


Figure 1.4.4 : Constant voltage bias with model.

Ultimately, the goal here is to determine a means for finding the transistor's drain current and drain-source voltage, along with the potentials across any other components.

To begin, consider the gate-source loop. By KVL, the V_{GG} source must drop across R_G and the gate-source junction, V_{GS}.

$$V_{GG} = V_{R_G} + V_{GS}$$

$$V_{GG} = I_G R_G + V_{GS}$$

I_G is approximately zero so this simplifies to

$$V_{GS} = V_{GG}$$

Given the transconductance, ◇◇gm, we can find ◇◇. Alternately, ◇◇ may be found using Equation 1.2.1 along with the device parameters ◇◇◇◇ and ◇◇◇(◇◇◇). For this circuit, the latter technique tends to be more practical. Once ◇◇ is found, the voltage drop across ◇◇ may be found, and then ◇◇◇ is determined from KVL.

Example 1.4.1

For the circuit of Figure 1.4.5, determine ◇◇ and ◇◇◇. Assume ◇◇◇◇ = 10 mA and ◇◇◇(◇◇◇) = –5 V.

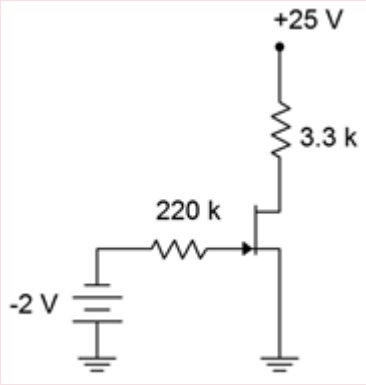


Figure 1.4.5: Schematic for Example 1.4.1

.

First, because ◇◇≈0, the drop across ◇◇ is ≈0 and ◇◇◇=◇◇◇. Using Equation 1.2.1

$$I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_{GS(off)}}\right)^2$$

$$I_D = 10mA\left(1 - \frac{-2V}{-5V}\right)^2$$

$$I_D = 3.6mA$$

Looking at the drain-source loop, KVL shows

$$V_{DD} = I_D R_D + V_{DS}$$
$$V_{DS} = V_{DD} - I_D R_D$$
$$V_{DS} = 25V - 3.6mA \times 3.3k\Omega$$
$$V_{DS} = 13.1V$$

While the computation for the constant voltage bias is relatively simple, it does not exhibit a stable Q point. For example, if Example 1.4.1 is repeated with another JFET, this one with ◇◇◇◇ = 12 mA and ◇◇◇(◇◇◇) = –6 V, the results are starkly different: ◇◇ grows to 5.33 mA and ◇◇◇ shrinks to 7.4 V. These are considerable changes given the relatively modest shifts in the device parameters. In this regard, the constant voltage bias is reminiscent of the simple base bias configuration used with BJTs.

To get a better understanding of the Q point stability issue, refer to Figure 1.4.6.

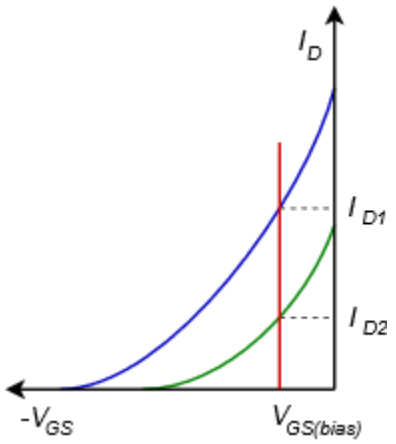


Figure 1.4.6: Variation for constant voltage bias.

Characteristic curves are plotted here for two different devices, one in green and one in blue. These represent the sort of device parameter variations we might expect to see across a product model. The fixed value of gate bias voltage is shown in red. From this graph it should be obvious that this form of bias will produce a wide variation in drain current, and thus, is not a good choice for applications that require a stable Q point. If the application does not have this requirement, constant voltage bias offers the advantage of requiring a minimum of components.

SELF BIAS

Self bias uses a small number of components and only a single power supply, yet it offers better stability than constant voltage bias. The name comes from the fact that the drain current will be used to create a voltage drop that sets up the gate-source, hence the circuit "biases itself". It is also referred to as automatic bias. The self bias prototype is shown in Figure 1.4.7.

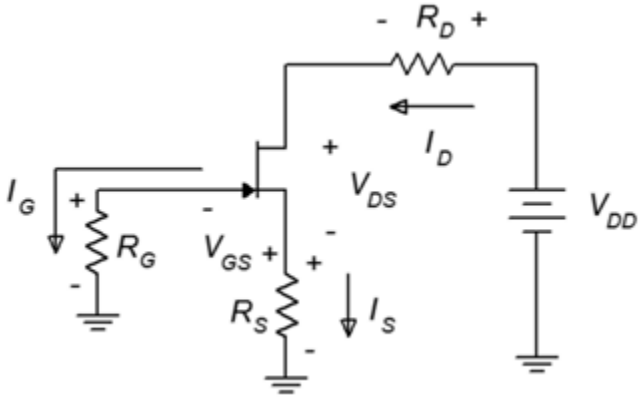


Figure 1.4.7: Self bias prototype.

Once again, we may assume that ⍰⍰ is 0. As ⍰⍰ is connected directly to ground, this means that ⍰⍰≈0. This being true, inspection of the schematic reveals that the magnitude of ⍰⍰⍰ must be the same as the voltage across ⍰⍰. Because ⍰⍰=⍰⍰ then

$$V_{GS} = -I_D R_S$$

(1.4.1)

This value of ⍰⍰⍰ is what generates the drain current. The definition is self-referential. This being

the case, how do we analyze the circuit? A proper derivation of the equation for drain current is not trivial. We start with the characteristic equation (Equation 11.2.1) and expand it.

$$I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_{GS(off)}}\right)^2$$

$$I_D = I_{DSS}\left(1 - \frac{2V_{GS}}{V_{GS(off)}} + \frac{{V_{GS}}^2}{{V_{GS(off)}}^2}\right)$$

$$I_D = I_{DSS} - \frac{2I_{DSS}V_{GS}}{V_{GS(off)}} + \frac{I_{DSS}{V_{GS}}^2}{{V_{GS(off)}}^2}$$

Substitute using Equation 1.2.2

$$g_{m0} = -\frac{2I_{DSS}}{V_{GS(off)}}$$

$$I_D = I_{DSS} + g_{m0}V_{GS} + \frac{I_{DSS}{V_{GS}}^2}{{V_{GS(off)}}^2}$$

Using Equation 1.4.1 this can be expanded to

$$I_D = I_{DSS} - g_{m0}I_D R_S + \frac{I_{DSS}{I_D}^2{R_S}^2}{{V_{GS(off)}}^2}$$

Rearranging yields

$$0 = \frac{I_{DSS}{R_S}^2}{{V_{GS(off)}}^2}{I_D}^2 - (1 + g_{m0}R_S)I_D + I_{DSS}$$

This is a quadratic equation in the form ◈◈^{2}+◈◈+◈ax^2+bx+c and can be solved using the quadratic formula:

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

The positive option in the numerator may be ignored as this occurs or ◈◈◈ beyond ◈◈◈(◈◈◈). The result is

$$I_D = 2I_{DSS}\left(\frac{1 + g_{m0}R_S - \sqrt{1 + 2g_{m0}R_S}}{(g_{m0}R_S)^2}\right)$$

(1.4.2)

Although this is an accurate analytical solution, it's certainly not the sort of equation most people want to memorize or derive as needed. As the ◈◈0◈◈ term is repeated in this equation multiple times, it is useful to plot this equation in terms of normalized ◈◈ versus ◈◈0◈◈. This curve is plotted in Figure 1.4.8.

$$g_{m0} = -\frac{2I_{DSS}}{V_{GS(off)}}$$

The value of ◈0◈◈ is found on the horizontal axis, traced up to the curve and then over to the normalized ◈◈ ratio. This number is multiplied by ◈◈◈◈ to determine the value of ◈◈.

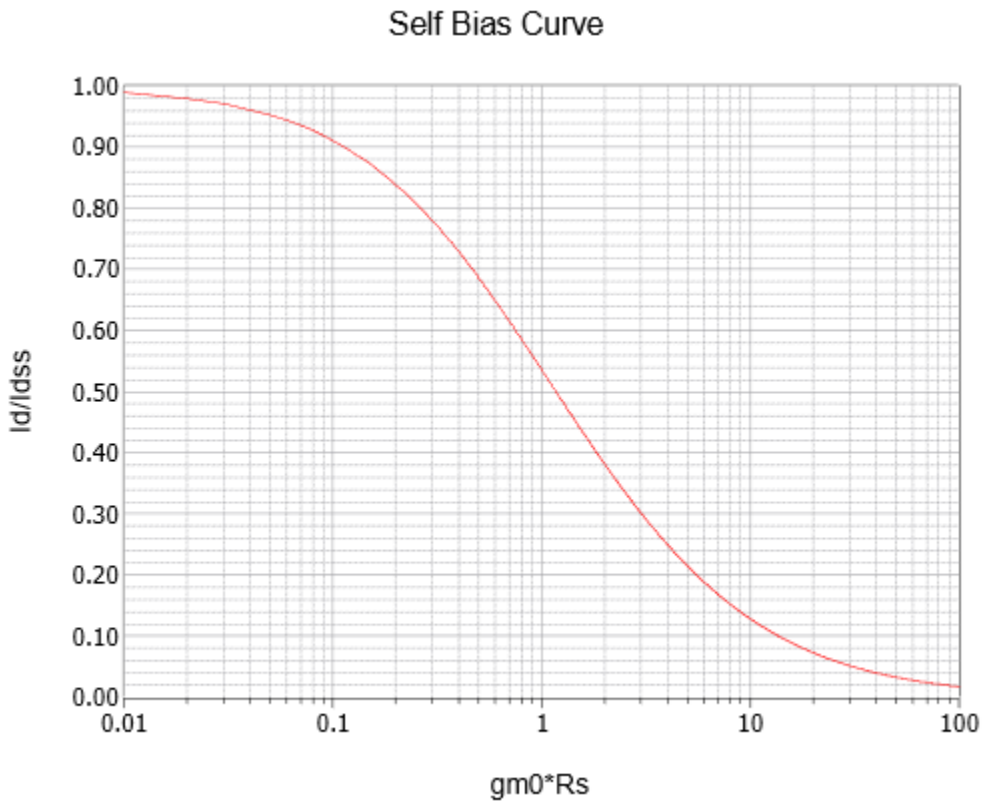


Figure 1.4.8: Self bias curve.

Example 1.4.2

Determine ◈◈ and ◈◈◈ for the circuit shown in Figure 11.4.9. Assume ◈◈◈◈ = 10 mA and ◈◈◈(◈◈◈) = –4 V.

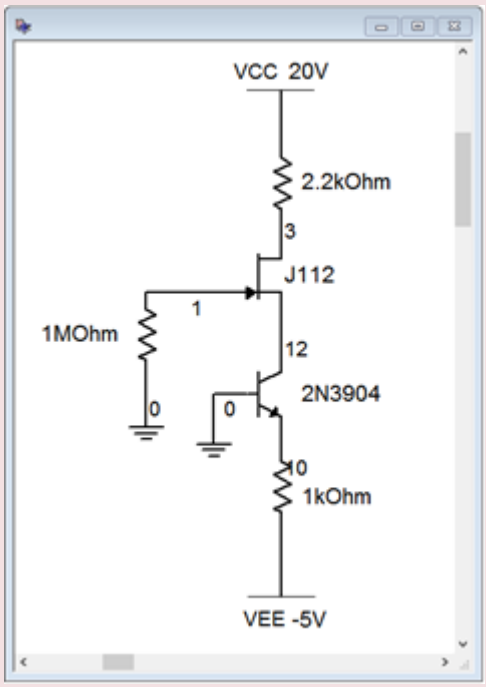


Figure 1.4.18: Constant current bias circuit in simulator.

Using the graphical method, first determine ◈◈0◈◈.

$$g_{m0} = -\frac{2I_{DSS}}{V_{GS(off)}}$$

$$g_{m0} = -\frac{2 \times 10mA}{4V}$$

$$g_{m0} = 5mS$$

Therefore ◈◈0◈◈ = 5 mS · 2.2 k Ω=11. The self bias graph yields approximately 0.12 for the normalized current ratio. Therefore

$$I_D = 0.12I_{DSS}$$
$$I_D = 0.12 \times 10mA$$
$$I_D = 1.2mA$$

Using Ohm's law and KVL

$$V_D = V_{DD} - I_D R_D$$
$$V_D = 20V - 1.2mA \times 3.9k\Omega$$
$$V_D = 15.32V$$
$$V_S = I_D R_S$$
$$V_S = 1.2mA \times 2.2k\Omega$$
$$V_S = 2.64V$$
$$V_{DS} = V_D - V_S$$

$$V_{DS} = 15.32V - 2.64V$$
$$V_{DS} = 12.68V$$

An alternate technique is to make an initial guess for ◇◇◇, typically one half of ◇◇◇(◇◇◇). The value of ◇◇ is then computed from the characteristic equation (Equation 1.2.1) and compared with the Ohm's law relation, Equation 1.4.1, rewritten as ◇◇=–◇◇◇/◇◇. Chances are, the two results will not agree so adjust the ◇◇◇ estimate and repeat the process. If done properly, the currents should be closer. Iterate this process until you converge on the answer.

To use this technique for the preceding problem we'd start by assuming ◇◇◇ = –2 V (half of ◇◇◇(◇◇◇◇)). Using this in Equation 1.2.1 yields ◇◇ = 2.5 mA, while using Equation 1.4.1 produces ◇◇ = 910 ◇A. Obviously the initial estimate was not correct. The second estimate for ◇◇◇VGS needs to increase negatively as this will decrease the result from Equation 1.2.1 and increase the result from Equation 1.4.1, hopefully meeting in the middle. We might try –2.5 volts. This will yield 1.4 mA from Equation 1.2.1 and 1.14 mA from Equation 1.4.1. As the gap has narrowed, the adjustment for the third estimate will be smaller, so we could try –2.6 volts. This would be relatively close to the value as computed in Example 1.4.2 (◇◇◇=–◇◇).

This approximation technique also offers a clue as to how self bias gains stability over constant voltage bias. If for some reason ◇◇ was to increase, this would create a larger voltage drop across ◇◇. Because this voltage is the same magnitude as ◇◇◇, this means that ◇◇◇ grows negatively. A more negative ◇◇◇ reduces ◇◇, thus opposing the initial change in drain current. This feedback mechanism is similar in function to the BJT collector feedback bias. The stability issue is visualized in Figure 1.4.1.

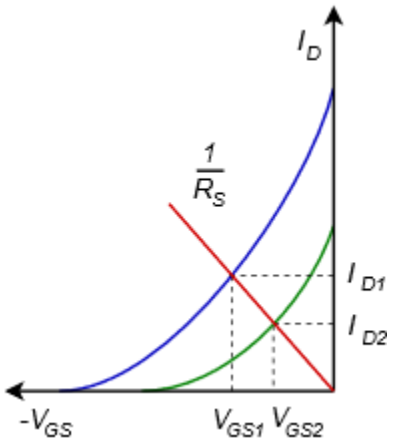


Figure 1.4.10 : Variation for self bias.

Two device curves are plotted to represent parameter variation (green and blue). Equation 1.4.1 shows the relationship between ◇◇ and ◇◇◇. If we put this in the form ◇=◇◇+◇y=mx+b, we find that the line goes through the origin and has a slope of 1/◇◇. This line is plotted in red. Where the line intersects the device curve yields the drain current and gate-source voltage for that particular device. Unlike constant voltage bias, self bias shifts some variation over to ◇◇◇, making ◇◇ more stable. In fact, if there is a particular design target for ◇◇ or ◇◇◇, a rearrangement of Equation 1.4.1 can be used to find the needed value of ◇◇ along with the characteristic curve or equation.

$$R_S = -\frac{V_{GS}}{I_D}$$

For example, if a certain ◈◈ is desired, this value could be used with Equation 1.2.1 to determine the corresponding ◈◈◈. These values are then used to find the required ◈◈. Alternately, the normalized values could be obtained via Figure 11.2.4.

Example 1.4.3

Determine a value for ◈◈ to set ◈◈◈ = –2 V for the circuit shown in Figure 1.4.11. Assume ◈◈◈◈= 20 mA and ◈◈◈(◈◈◈) = –4 V.

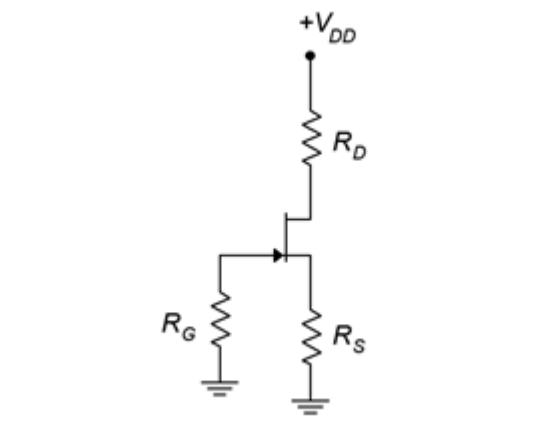


Figure 1.4.11: Schematic for Example 1.4.3 .

We can determine the drain current using Equation 1.2.1.

$$I_D = I_DSS\left(1 - \frac{V_{GS}}{V_{GS(off)}}\right)^2$$

$$I_D = 20mA\left(1 - \frac{-2V}{-4V}\right)^2$$

$$I_D = 5mA$$

$$R_S = -\frac{V_{GS}}{I_D}$$

$$R_S = \frac{-2V}{5mA}$$

$$R_S = 400\Omega$$

In sum, self bias is a minimal parts count circuit that offers modest stability. The stability can be improved with the addition of other components, as we shall see with the next bias configuration.

COMBINATION BIAS

The combination bias configuration (AKA source bias) is based on self bias but adds a negative power supply connected to ◇◇, hence its name. This will enhance the stability of ◇◇, ◇◇◇ and ◇◇. The combination bias prototype is shown in Figure 1.4.12.

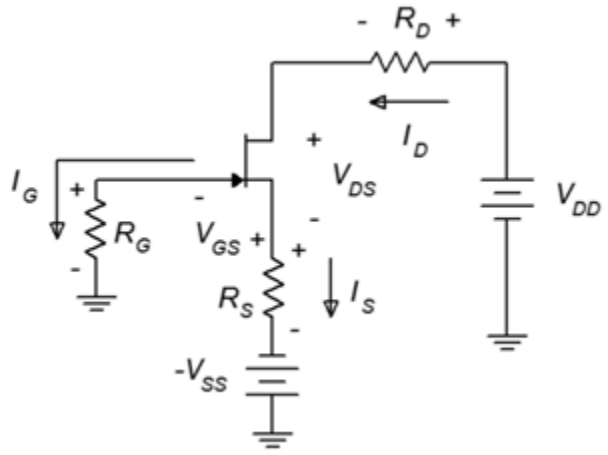


Figure 1.4.12: Combination bias prototype.

The analysis is similar to that of self bias but with one major twist: the source power supply increases the voltage drop across ◇◇. This stabilizes the voltage (and hence, the current) because it is no longer equal to –◇◇◇, but rather

$$V_{R_S} = I_D R_S = |V_{GS}| + |V_{SS}|$$

(1.4.3)

If ◇◇◇≫◇◇◇, then we can approximate ◇◇ as ◇◇◇/◇◇. As with self bias, an analytical solution for ◇◇ is possible. In order to do so, we would begin with the characteristic equation and Equation 1.4.3. The derivation is left as an exercise.

$$I_D = 2I_{DSS}\left(\frac{1 + g_{m0}R_S(1+k) - \sqrt{1 + 2g_{m0}R_S(1+k)}}{(g_{m0}R_S)^2}\right)$$

(1.4.4)

The formula is very similar to the self bias formula but with the addition of a factor, ◇. ◇ is a "swamping factor" and is defined as the ratio of ◇◇◇ to ◇◇◇(◇◇◇). If ◇=0, there is no source power supply and the formula reverts back to the simpler self bias formula. On the other hand, if ◇ is very large, ◇◇≈◇◇◇/◇◇.

As was the case with self bias, we can plot Equation 1.4.4 using the ◇◇0◇◇ factor. A series of three plots for ◇ = 2, 3 and 4 are rendered in Figure 1.4.13.[1]

1. We could add a third axis for ◇ and plot a surface, and while it might be pretty, a 3D plot like this rendered onto a 2D surface, such as a page in a textbook, is of marginal utility.

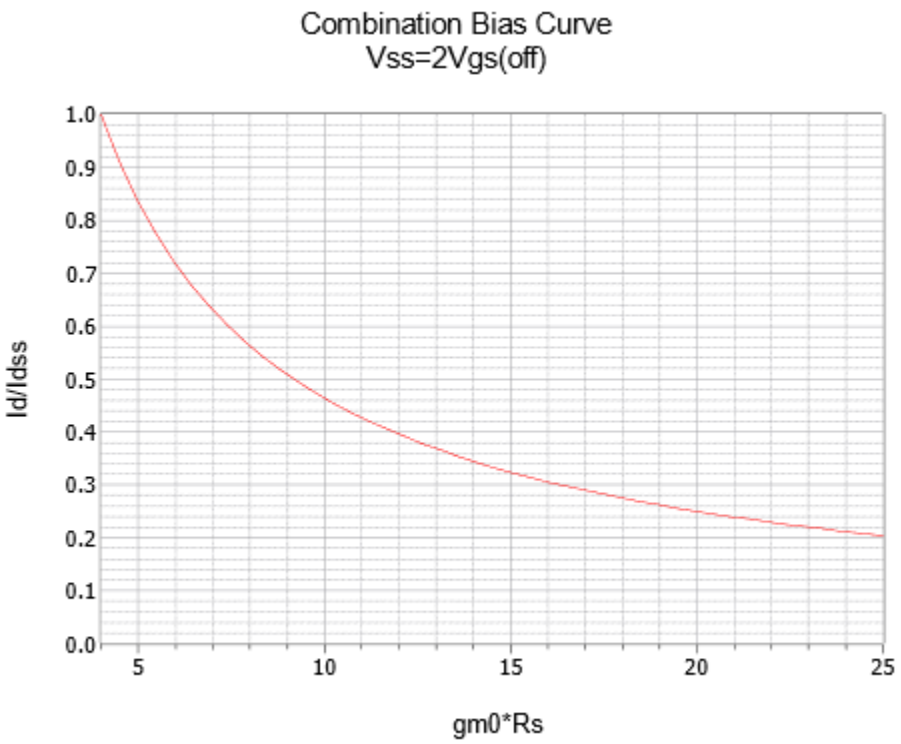


Figure 1.4.13◈: Combination bias curve, ◈=2.

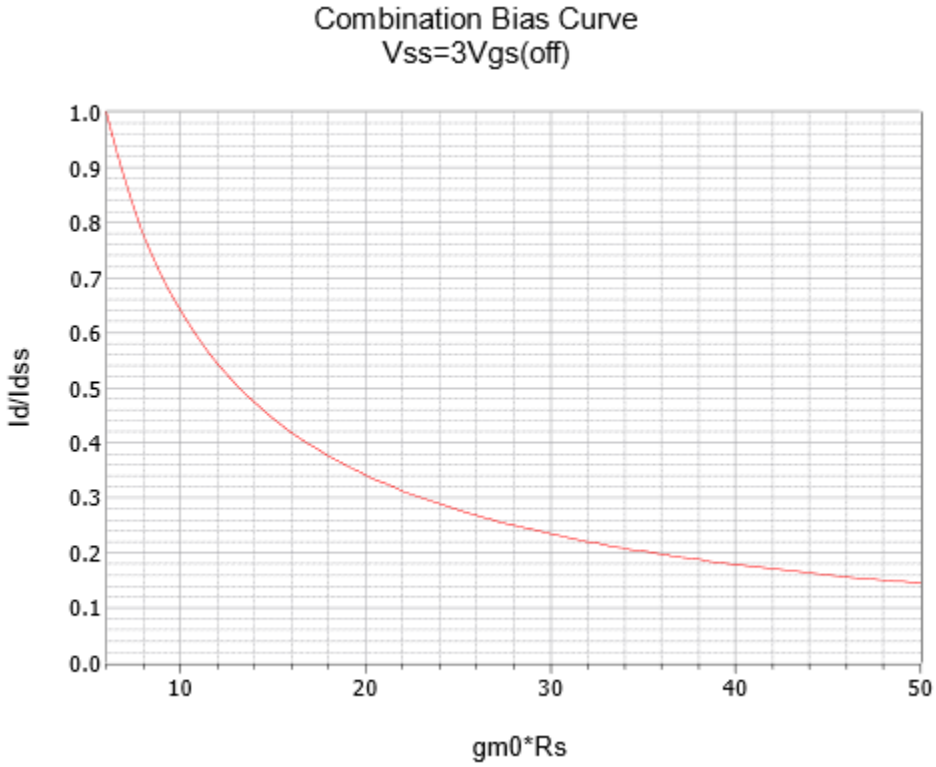


Figure 1.4.13◈: Combination bias curve, ◈=3.

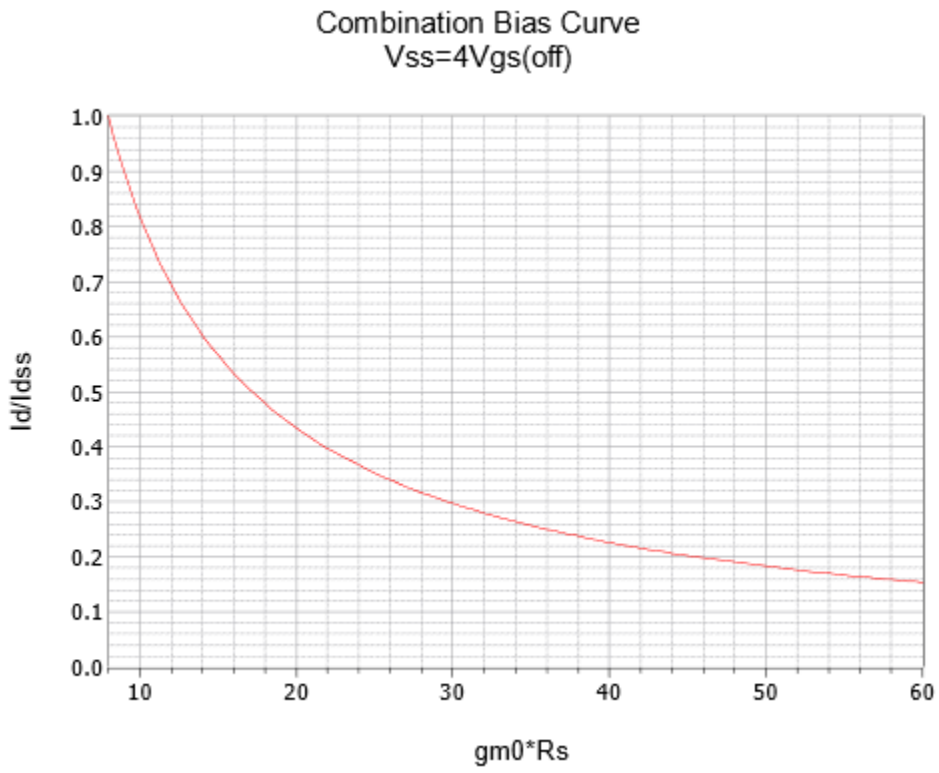


Figure 1.4.13 : Combination bias curve, =4.

Example 1.4.4

Determine and for the circuit shown in Figure 1.4.14. Assume = 12 mA and = –4 V.

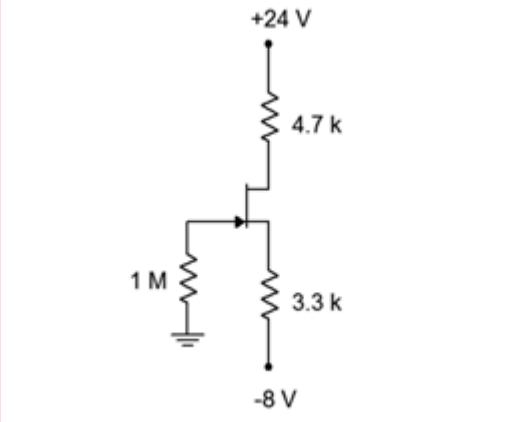


Figure 1.4.14: Schematic for Example 1.4.4.

Using the graphical method, first determine .

$$g_{m0} = -\frac{2I_{DSS}}{V_{GS(off)}}$$

$$g_{m0} = -\frac{2 \times 12mA}{-4V}$$

$$g_{m0} = 6mS$$

Therefore $\diamond_{\diamond 0 \diamond \diamond}$ = 6 mS · 3.3 k Ω = 19.8. The swamping ratio, $\diamond$, is $\diamond_{\diamond \diamond}/\diamond_{\diamond \diamond(\diamond \diamond \diamond)}$=–8/–4=2. This requires the graph in Figure 1.4.13$\diamond$. This graph yields approximately 0.25 for the normalized current ratio. Therefore

$$I_D = 0.25I_{DSS}$$
$$I_D = 0.25 \times 12mA$$
$$I_D = 3mA$$

Using Ohm's law and KVL

$$V_D = V_{DD} - I_D R_D$$
$$V_D = 24V - 3mA \times 4.7k\Omega$$
$$V_D = 9.9V$$
$$V_S = V_{SS} + I_D R_S$$
$$V_S = -8V + 3mA \times 3.3k\Omega$$
$$V_S = 1.9V$$
$$V_{DS} = V_D - V_S$$
$$V_{DS} = 9.9V - 1.9V$$
$$V_{DS} = 8V$$

As a crosscheck, using Equation 1.4.4 yields 3.028 mA for $\diamond_{\diamond}$. The deviation is no doubt due to inaccuracy in reading the graph. In any case, using this value of drain current we find $\diamond_{\diamond}$ to be 1.992 volts, a little higher than calculated above. This indicates that $\diamond_{\diamond \diamond}$ is –1.992 volts (because $\diamond_{\diamond} \approx 0$). If we plug this value of $\diamond_{\diamond \diamond}$ into Equation 1.2.1, $\diamond_{\diamond}$=3.024 mA; an excellent match with the deviation being due to accumulated rounding errors.

In order to show the increased Q point stability of the combination bias, we'll repeat the preceding problem using a JFET with a significantly lower $\diamond_{\diamond \diamond \diamond}$.

Example 1.4.5

Determine $\diamond_{\diamond}$ for the circuit shown in Figure 1.4.14. Assume $\diamond_{\diamond \diamond \diamond}$ = 8 mA and $\diamond_{\diamond \diamond(\diamond \diamond \diamond)}$ = –4 V.

For this version we'll use Equation 1.4.4. First determine $\diamond_{\diamond 0}\diamond_{\diamond}$.

$$g_{m0} = -\frac{2I_{DSS}}{V_{GS(off)}}$$

$$g_{m0} = -\frac{2 \times 8mA}{-4V}$$

$$g_{m0} = 4mS$$

Therefore $\diamond_{\diamond 0}\diamond_{\diamond}$ = 4 mS · 3.3 k Ω=13.2. The swamping ratio, $\diamond$, is $\diamond_{\diamond\diamond}/\diamond_{\diamond\diamond(\diamond\diamond\diamond)}$=–8/–4=2.

$$I_D = 2I_{DSS}\left(\frac{1 + g_{m0}R_S(1+k) - \sqrt{1 + 2g_{m0}R_S(1+k)}}{(g_{m0}R_S)^2}\right)$$

$$I_D = 2 \times 8mA\left(\frac{1 + 13.2(1+2) - \sqrt{1 + 2 \times 13.2(1+2)}}{(13.2)^2}\right)$$

$$I_D = 2.906mA$$

For the graphical method, a reasonable estimate for the normalized $\diamond_{\diamond}$ would be around 0.36, yielding a drain current of 2.88 mA. Stability is apparent because the drain current has dropped only a few percent in spite of the fact that $\diamond_{\diamond\diamond\diamond}$ decreased by 33%.

The graph of Figure 1.4.15 illustrates nicely the increased stability of the Q point. Once again, we plot two representative device curves in green and blue. As was the case with self bias, a plot line can be drawn, the slope of which is equal to the reciprocal of $\diamond_{\diamond}$. This plot line does not go though the origin, though. Instead, the $\diamond$x axis intercept is the voltage $|\diamond_{\diamond\diamond}|$. Thus, the red plot line is shifted along the $\diamond_{\diamond\diamond}$ axis.

As can be seen in the graph, the variation in $\diamond_{\diamond}$ is reduced (although at the expense of variation in $\diamond_{\diamond\diamond}$). For large values of $\diamond_{\diamond\diamond}$ with correspondingly large values of $\diamond_{\diamond}$, the bias plot line becomes nearly horizontal, indicating a very stable Q point. With two variables in play, this bias proves to be very flexible. It can also be realized by using a positive voltage divider at the gate and removing $\diamond_{\diamond\diamond}$ (returning $\diamond_{\diamond}$ to ground).

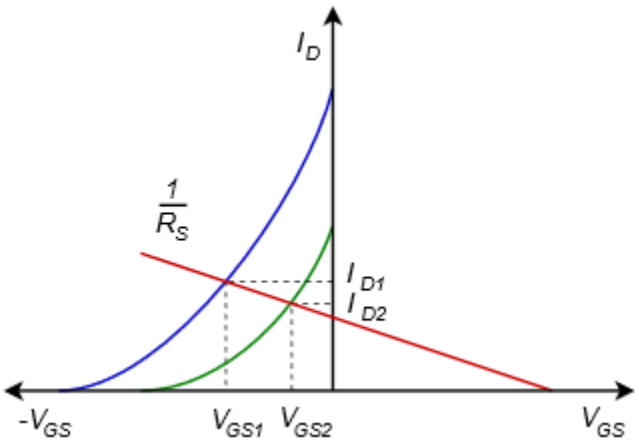


Figure 1.4.15 : Variation for combination bias.

CONSTANT CURRENT BIAS

The most stable bias for JFETs relies, oddly enough, on a current source made with a BJT. It is called constant current bias, yet another imaginative tag. Interestingly, although this will keep the Q point very stable, a fixed ◆ does not guarantee the most stable value of voltage gain. In fact, it might be easier to achieve that goal using combination bias. The prototype constant current bias circuit is shown in Figure 1.4.16. An NPN BJT is used for an N-channel JFET and a PNP would be used with a P-channel JFET, typically driven from above (i.e., circuit flipped top to bottom).

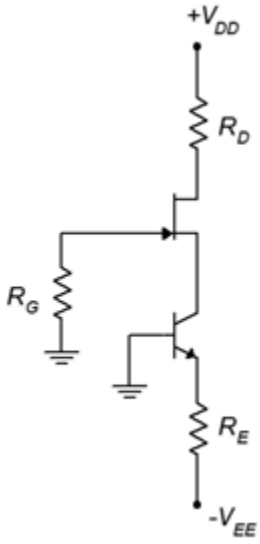


Figure 1.4.16 : Constant current bias prototype.

Ignoring the JFET for a moment, the BJT is configured as in two-supply emitter bias. In this case the base is tied directly to ground, leaving the emitter at about –0.7 VDC. The remainder of the ◆◆ supply drops across ◆, establishing the emitter current. As the collector is connected directly to the JFET's source terminal, this means that ◆≈◆. The source current winds up being just as stable as the emitter current, which we have already seen is very stable. The only requirement is that ◆ should not be programmed to be larger than ◆◆. This being true, ◆ will set up a corresponding ◆◆. This also establishes ◆ because ◆≈0. Therefore, the source terminal will be a small positive

voltage and this is precisely what the BJT needs in order to guarantee that its collector-base junction is reverse-biased.

Computation of circuit currents and voltages is straightforward and does not involve the use of graphical aides. The first step is to examine the BJT's emitter loop and determine ⍰⍰. Once this is found, ⍰⍰ and ⍰⍰ are known, and all remaining component potentials may be found using Ohm's law and KVL.

This technique does not involve the calculation of ⍰⍰⍰. In fact, because ⍰⍰ is very stable, ⍰⍰⍰ will show the widest variation of all biasing circuits when the JFET is changed. If ⍰⍰⍰ is needed, it can be determined via a little algebraic manipulation on Equation 1.2.1.

Example 1.4.6

Determine ⍰⍰, ⍰⍰⍰ and ⍰⍰⍰ in the circuit of Figure 1.4.17. ⍰⍰⍰⍰ = 15 mA and ⍰⍰⍰(⍰⍰⍰) = –3 V.

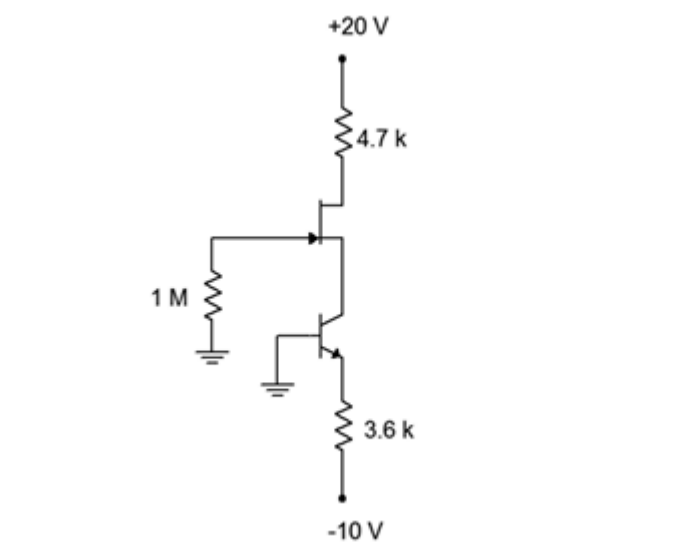


Figure 1.4.17: Schematic for Example 1.4.6 .

We begin by finding ⍰⍰.

$$I_E = \frac{|V_{EE}| - 0.7V}{R_E}$$

$$I_E = \frac{10V - 0.7V}{3.6k\Omega}$$

$$I_E = 2.58mA$$

⍰⍰ is the same as ⍰⍰ and ⍰⍰, therefore

$$V_D = V_{DD} - I_D R_D$$
$$V_D = 20V - 2.58mA \times 4.7k\Omega$$
$$V_D = 7.87V$$

To find ◈◈ we note that ◈◈=–◈◈◈ and rearrange Equation 1.2.1.

$$I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_{GS(off)}}\right)^2$$

$$V_{GS} = V_{GS(off)}\left(1 - \sqrt{\frac{I_D}{I_{DSS}}}\right)$$

$$V_{GS} = -3V\left(1 - \sqrt{\frac{2.58mA}{15mA}}\right)$$

$$V_{GS} = -1.24V$$

Therefore ◈◈=1.24V and

$$V_{DS} = V_D - V_S$$
$$V_{DS} = 7.87V - 1.24V$$
$$V_{DS} = 6.63V$$

We turn next to a computer simulation of a similar circuit to validate our methodology.

COMPUTER SIMULATION

A constant current bias circuit is entered into a simulator as shown in Figure 1.4.18.

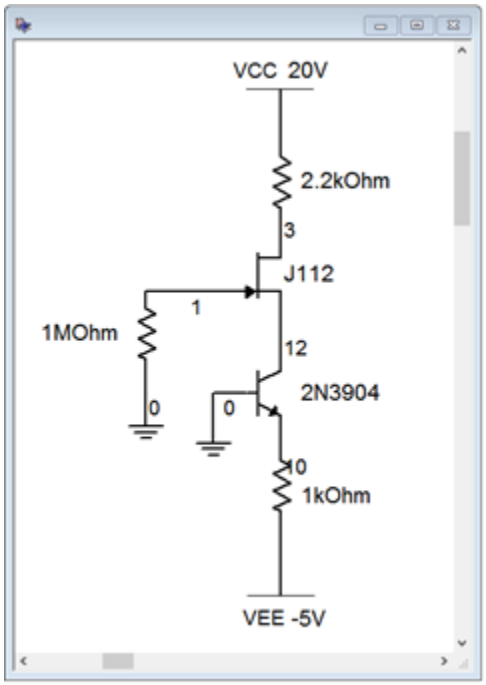


Figure 1.4.18: Constant current bias circuit in simulator.

A cursory estimate shows that ◈◈ and ◈◈ should be around 4.3 mA. Also, ◈◈ should be approximately 20 V – 4.3 mA · 2.2 kΩ, or about 1.54 volts. The results of a DC operating point analysis are shown in Figure 1.4.19.

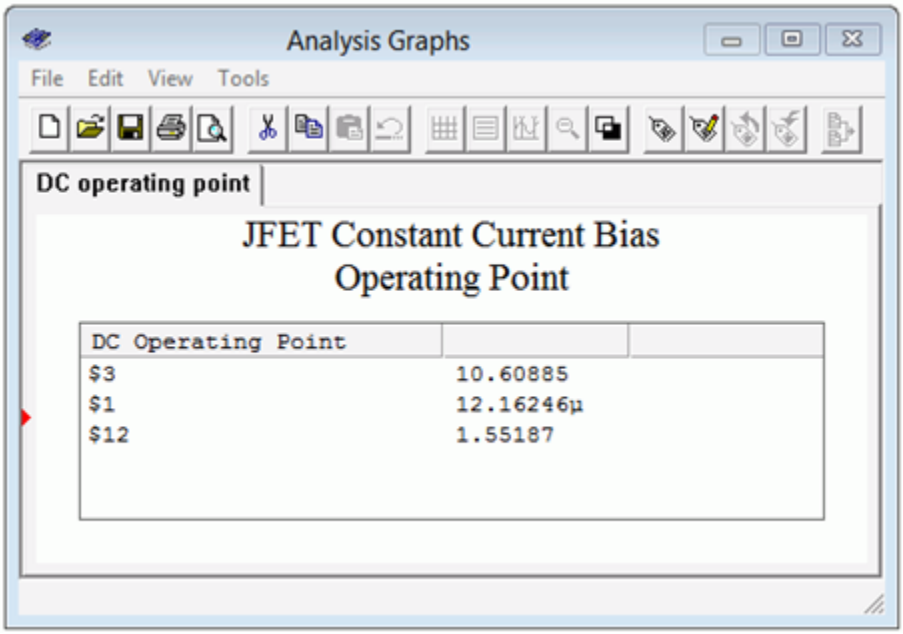


Figure 1.4.19 : Constant current bias DC operating point simulation results.

The drain voltage (node 3) is just over 1.6 volts, agreeing with our estimate. Also, note the minuscule gate voltage (node 1) of 12 ◈V which verifies our continuing assumption in these circuits that ◈◈≈0 V_{DC}. Finally, we see a modest potential of about 1.5 volts at the source terminal (node 12). This shows the proper reverse-biasing of both the gate-source and collector-base junctions.

Finally, we can examine the Q point variation using Figure 1.4.20. Here, the plot line is perfectly horizontal and all device variation is manifest in ⍰⍰⍰.

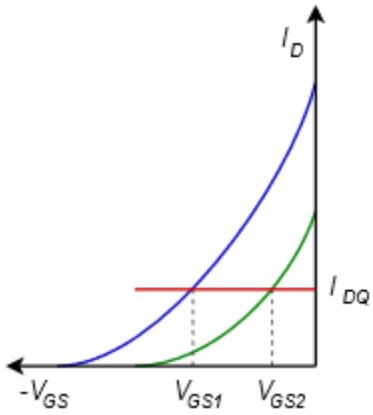


Figure 1.4.20: Variation for constant current bias.

1.5 SUMMARY

The junction field effect transistor is an altogether different device from the bipolar junction transistor. Instead of relying on a forward-biased PN junction to control current, the JFET utilizes a reverse-biased PN junction. Furthermore, the JFET uses voltage control rather than the BJT's current control. In spite of this, a family of JFET drain curves offers similarity to the BJT's collector curves, exhibiting three regions: ohmic, constant current and breakdown.

The DC model of a JFET includes a voltage-controlled current source in the drain and a very, very large resistance, ◇◇◇, from gate to source. This resistance models that of a reverse-biased PN junction. The characteristic equation of the JFET is square-law and is consequently much more gentle in slope than the corresponding equation for a BJT. The maximum current produced by a JFET is ◇◇◇◇ and occurs when ◇◇◇=0 V. ◇◇◇ must always be negative to ensure proper operation and all negative values will lead to a drain current less than ◇◇◇◇. Once the gate-source becomes negative enough (at ◇◇◇(◇◇◇)), drain current goes to zero.

There are several methods to bias JFETs. Perhaps the most simple method is to apply a fixed potential to the gate while grounding the source. This is called constant voltage bias and is the least stable bias in terms of Q point. Self bias uses a minimum of components and offers modest stability. It is a decent general-purpose bias. The addition of a negative power supply to the source resistor leads to the combination bias topology. This circuit offers improvements in stability over self bias. The most stable bias is the constant current bias. This form relies on a BJT to establish a very stable current.

Review Questions

1. Compare the operation of the JFET to the BJT.
2. Compare the regions of JFET drain curves to those of BJT collector curves.
3. Why is the JFET referred to as a square-law device?
4. Rank the biasing schemes presented in this chapter in terms of Q point stability.
5. What is pinch-off voltage?
6. How does the JFET DC biasing model differ from the BJT DC model?

1.6 EXERCISES

ANALYSIS PROBLEMS

1. For the circuit of Figure 1.6.1, determine [illegible] and [illegible]. [illegible] = 40 mA, [illegible] = –4 V, [illegible] = 26 V, [illegible] = –2 V, [illegible] = 220 kΩ, [illegible] = 1.2 kΩ.
2. For the circuit of Figure 1.6.1, determine [illegible] and [illegible]. [illegible] = 20 mA, [illegible] = –3 V, [illegible] = 22 V, [illegible] = –1 V, [illegible] = 390 kΩ, [illegible] = 1 kΩ.

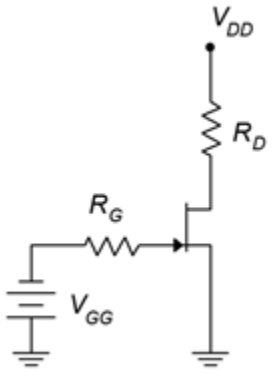
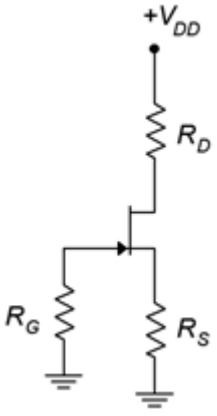


Figure 1.6.1

3. For the circuit of Figure 1.6.2, determine [illegible], [illegible] and [illegible]. [illegible] = 24 mA, [illegible] = –6 V, [illegible] = 36 V, [illegible] = 220 kΩ, [illegible] = 2 kΩ, [illegible] = 1.8 kΩ.
4. For the circuit of Figure 1.6.2, determine [illegible], [illegible] and [illegible]. [illegible] = 18 mA, [illegible] = –3 V, [illegible] = 30 V, [illegible] = 270 kΩ, [illegible] = 2.7 kΩ, [illegible] = 3.3 kΩ.



5. For Figure 1.6.3, determine [illegible], [illegible] and [illegible]. [illegible] = 16 mA, [illegible] = 25 V, [illegible] = –3 V, [illegible] = –6 V, [illegible] = 560 kΩ, [illegible] = 2 kΩ, [illegible] = 3.6 kΩ.
6. For Figure 1.6.3, determine [illegible], and [illegible]. [illegible] = 16 mA, [illegible] = 25 V, [illegible] = –3 V, [illegible] = –9 V, [illegible] = 680 kΩ, [illegible] = 2 kΩ, [illegible] = 2.7 kΩ.

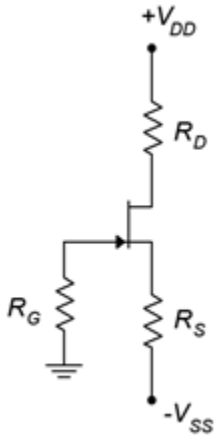
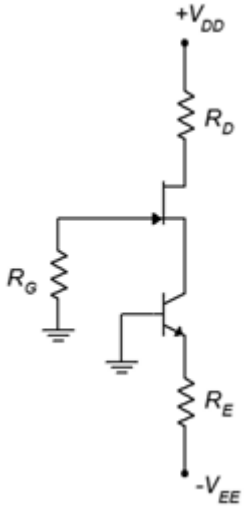


Figure 1.6.3

7. For Figure 1.6.4, determine [illegible], [illegible] and [illegible]. [illegible] = 16 mA, [illegible] = 25 V, [illegible] = –3 V, [illegible] = –9 V, [illegible] = 810 kΩ, [illegible] = 2 kΩ, [illegible] = 2.7 kΩ.
8. For the circuit of Figure 1.6.4, determine [illegible] and [illegible]. [illegible] = 40 mA, [illegible] = –4 V, [illegible] = 30 V, [illegible] = –6 V, [illegible] = 750 kΩ, [illegible] = 500 Ω, [illegible] = 1.8 kΩ.

DESIGN PROBLEMS

9. Using the circuit of Figure 1.6.2, determine a value for [illegible] to set [illegible] to 4 mA. [illegible] = 10 mA, [illegible] = –2 V, [illegible] = 20 V, [illegible] = 430 kΩ, [illegible] = 1.8 kΩ.
10. Using the circuit of Figure 1.6.1, determine a value for [illegible] to set [illegible] to 2 mA. [illegible] = 10 mA, [illegible] = –4 V, [illegible] = 28 V, [illegible] = 470 kΩ, [illegible] = 4.7 kΩ.
11. Using the circuit of Figure 10.6.410.6.4, determine a value for [illegible] to set [illegible] to 4 mA. [illegible] = 18 mA, [illegible] = –3 V, [illegible] = 25 V, [illegible] = –12 V, [illegible] = 330 kΩ, [illegible] = 2.2 kΩ.



12. Using the circuit of Figure 1.6.4, determine values for [illegible] and [illegible] to set [illegible] to 5 mA and [illegible] to 6 V. [illegible] = 20 mA, [illegible] = –4 V, [illegible] = 32 V, [illegible] = –10 V, [illegible] = 390 kΩ.

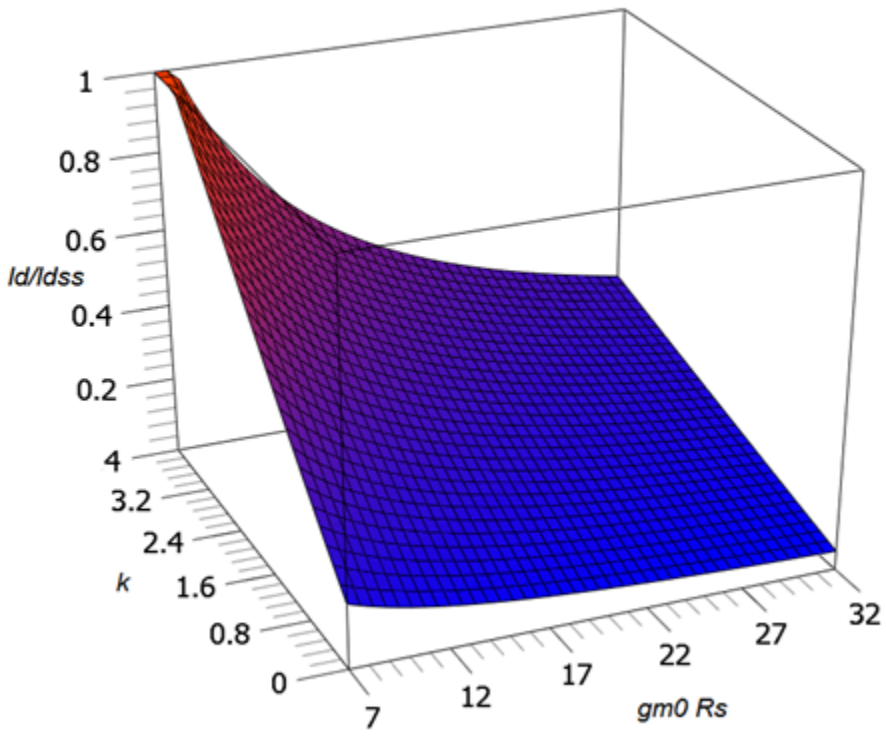
CHALLENGE PROBLEMS

13. Following the derivation of Equation 10.4.2, derive Equation 10.4.4.
14. Using the circuit of Figure 1.6.3, determine values for R_{D} and R_{SS} to set I_{D} to 4 mA. I_{DSS} = 16 mA, $V_{GS(OFF)}$ = –4 V, V_{DD} = 30 V, R_{D} = 680 kΩ, R_{S} = 2 kΩ.

COMPUTER SIMULATION PROBLEMS

15. Perform a DC operating point simulation on the circuit of Problem 7 to verify the results. The J111 will be sufficient.
16. Perform a DC operating point simulation on the circuit of Problem 10 to verify the results. The J111 will be sufficient.

DEPARTMENT OF MARGINAL UTILITY



The graphs of Figure 10.4.13 represent three slices from this surface.
Looks cool, but…

UNIT 2: JFET SMALL SIGNAL AMPLIFIERS

Learning Objectives

After completing this chapter, you should be able to:

- Determine the voltage gain, input impedance and output impedance of basic JFET amplifiers.
- Draw and explain a basic AC model of a JFET.
- Compare and analyze JFET voltage amplifiers and voltage followers.
- Discuss the advantages and disadvantages of JFET circuits with those of comparable BJT circuits.
- Analyze small signal combination BJT/JFET amplifier circuits.
- Discuss applications that make use of the JFET's ohmic region.

2.1 INTRODUCTION

The JFET can be used to create both voltage amplifiers and voltage followers. In comparison with the BJT, the JFET tends to have less voltage gain potential. On the other hand, JFET circuits offer the possibility of a much higher input impedance, lower noise and better high frequency performance. There are many other similarities with BJT amplifiers. For example, the possibility of swamping still exists as a means of lowering distortion at the expense of voltage gain. Also, the JFET voltage amplifier inverts the signal, just like the BJT version. When it comes to AC analysis, a key element for the BJT is ◈′◈. For the JFET, the comparable parameter is transconductance, ◈◈.

JFET amplifiers and followers can be used with their BJT cousins. Indeed, the combination of the two, each playing to their strengths, has the potential to outperform a design using only one type of device.

Alongside their use in amplifiers and followers, JFETs can also be used in their ohmic region. This includes applications as voltage-controlled resistors and analog switches. In this mode, the device no longer behaves as a constant current source. Instead, the channel resistance becomes a function of the gate-source voltage and can be used as a control element within a voltage divider. As such, it has the capability of changing resistance value much faster than a mechanical potentiometer.

2.2 SIMPLIFIED AC MODEL OF THE JFET

An AC model of the JFET is shown in Figure 11.2.1. This is essentially the same model as was used for DC analysis. Once again, we have a voltage-controlled current source situated in the drain. The reverse-biased junction shows up as a very large resistance, ◇◇◇.

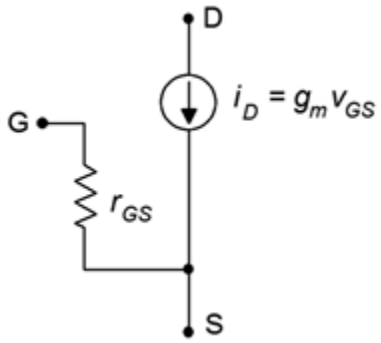


Figure 2.2.1: AC model of JFET.

It is worth mentioning that this model is suitable only for low frequencies. At higher frequencies, device capacitances can play a major role in the response of the amplifier. There are three device capacitances not shown in the Figure that shunt each pair of terminals: ◇◇◇, ◇◇◇ and ◇◇◇. On a data sheet, the "lumped" capacitances are often given. These are ◇◇◇◇, the capacitance looking into the gate with the source and drain shorted to ground: ◇◇◇◇=◇◇◇+◇◇◇; and ◇◇◇◇, the capacitance seen from the drain with the gate and source shorted to ground: ◇◇◇◇=◇◇◇+◇◇◇. As we shall see, these capacitances can have a sizable impact on amplifier characteristics such as ◇◇◇.

The value of transconductance, ◇◇, will prove to be of particular interest. It is roughly of equal importance to ◇′◇ in a BJT.[1]

1. In fact, we can say that 1/◇′◇ is ◇◇ for a BJT.

2.3 COMMON SOURCE AMPLIFIER

The common source amplifier is analogous to the common emitter amplifier. The prototype amplifier circuit with device model is shown in Figure 2.3.1 .

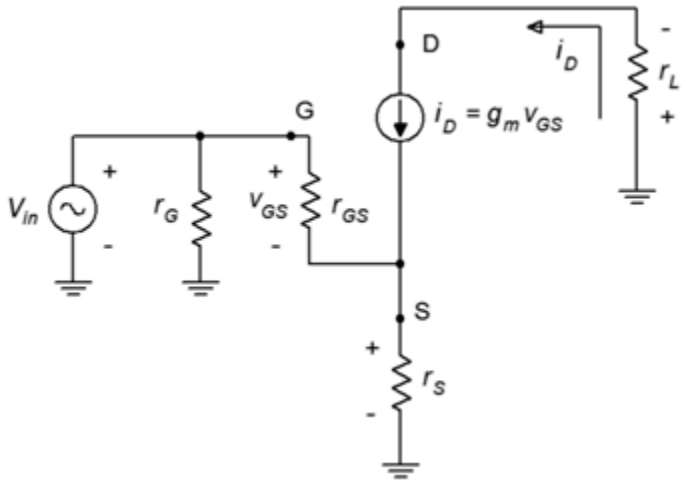


Figure 2.3.1 : Common source amplifier with model.

This circuit includes a swamping resistor, ◈◈ . The input signal is presented to the gate terminal while the output is taken from the drain.

VOLTAGE GAIN

An equation for the voltage gain, ◈◈, is developed as follows. First, we start with the fundamental definition, namely that voltage gain is the ratio of ◈◈◈◈ to ◈◈◈, and proceed by expressing these voltages in terms of their Ohm's law equivalents.

$$A_v = \frac{v_{out}}{v_{in}} = \frac{v_D}{v_G}$$

$$A_v = \frac{-i_D r_L}{i_D r_S + v_{GS}}$$

$$A_v = \frac{-g_m v_{GS} r_L}{g_m v_{GS} r_S + v_{GS}}$$

$$A_v = \frac{g_m r_L}{g_m r_S + 1}$$

(2.3.1)

If there is no swamping resistor, the first portion of the denominator drops out and the gain simplifies to –◈◈·◈◈ . The swamping resistor in the source, ◈◈, plays the same role here as it did in the BJT: it helps to stabilize the gain and reduce distortion. It does so at the expense of voltage gain.

INPUT IMPEDANCE

Referring back to Figure 2.3.1 , the input impedance of the amplifier will be ◈◈ in parallel with the impedance looking into the gate terminal, ◈◈◈(◈◈◈◈). For the non-swamped case, this will be ◈◈◈. At low frequencies ◈◈◈ is very large, well into the megohms. In most practical circuits, ◈◈ will be much lower, hence

Z_{in} = r_G || r_{GS} \approx r_G \](2.3.2)

Theoretically, for swamped amplifiers ◈◈◈(◈◈◈◈) will be higher than ◈◈◈ but this is a moot point. In either case, it is relatively easy to obtain a high input impedance, certainly much easier than it is for typical single-device BJT amplifiers.

It might be easy to become complacent and simply assume that ◈◈ sets the input impedance and that's the end of it. This would be a mistake. As mentioned earlier, with impedances this high, we cannot ignore items such as junction capacitance. For example, for a general purpose device a typical value for ◈◈◈ , the total input capacitance, may be in the vicinity of 5 to 10 pF. This capacitance appears in parallel with ◈◈. If this amplifier is used for ultrasonic signals, the capacitive reactance, ◈◈, would be as low as 160 k Ω at 100 kHz. Although this is high compared to typical BJT circuits, it's less than the ◈◈ values commonly used for biasing. At higher frequencies, the situation is even worse as ◈◈ decreases with frequency. Also, we are ignoring the Miller effect here which makes the situation even worse than even worse, so perhaps we can say that it's even worser, which is a claim we could also make regarding the grammar of this sentence.

OUTPUT IMPEDANCE

To investigate the output impedance, we'll refer to Figure 2.3.2 . This circuit is very similar to that of Figure 2.3.1 . The major difference is that the AC load equivalent has been split into its two components, the load itself, ◈◈, and the drain biasing resistor, ◈◈.

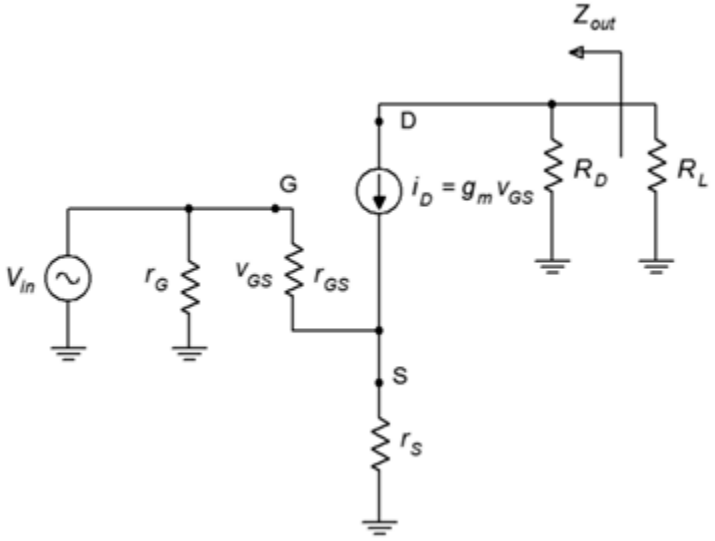


Figure 2.3.2 : Output impedance of common source amplifier.

From the vantage point of ◈◈, peering back into the amplifier we see ◈◈ in parallel with the impedance at the drain. At the drain we find the current source, ◈◈ . The internal impedance of this

equivalent current source is very high compared to typical values for ⍰⍰ (hundreds of Ω), therefore we can approximate the output impedance as

```
*** QuickLaTeX cannot compile formula:
\[Z_{out} \approx R_D \label\]

*** Error message:
Emergency stop.
```

(2.3.3)

It should be noted that all forms of DC bias discussed in the previous chapter are game here. There are a few limitations to be aware of, though. For example, when using constant voltage bias, swamping is not possible as that bias form does not use a source resistor. In contrast, self bias and combination bias include a source resistor so swamping is a possibility, however, ⍰⍰ may need to be split and partially bypassed to achieve the desired results. Finally, constant current bias is not well-positioned to use swamping as that would require some additional work to fit in a new ⍰⍰ along with the current source. More typically, the current source will just be bypassed with a capacitor to produce a non-swamped amplifier.

Example 2.3.1

Determine the voltage gain and input impedance for the circuit shown in Figure 2.3.3 . Assume ⍰⍰⍰=15 mA and ⍰⍰(⍰⍰)=–3 V.

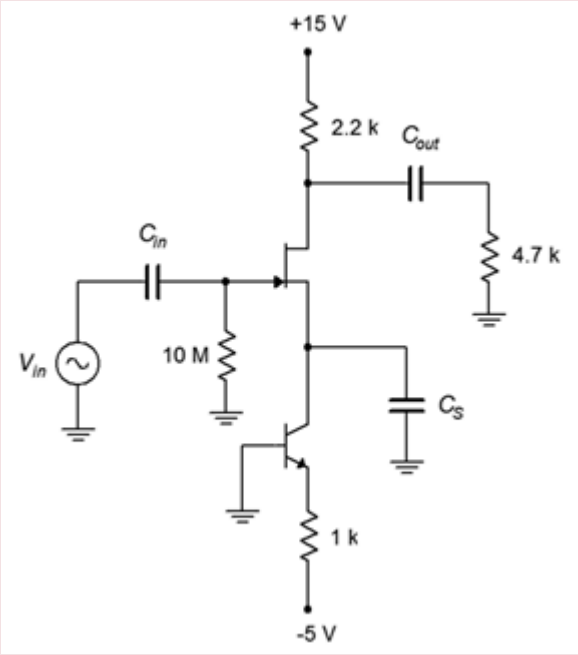


Figure 2.3.3 : Circuit for Example 2.3.1 .

This is an unswamped common source amplifier with constant current bias. We can determine ⍰⍰⍰ via inspection.

$$Z_{in} = Z_{in(gate)}||R_G$$

$$Z_{in} \approx 10M\Omega$$

To find the voltage gain, we'll first need to find $\diamondsuit_{\diamondsuit}$ and $\diamondsuit_{\diamondsuit 0}$ in order to find $\diamondsuit_{\diamondsuit}$.

$$I_D = \frac{|V_{EE}| - 0.7V}{R_E}$$

$$I_D = \frac{5V - 0.7V}{1k\Omega}$$

$$I_D = 4.3mA$$

$$g_{m0} = -\frac{2I_{DSS}}{V_{GS(off)}}$$

$$g_{m0} = -\frac{30mA}{-3V}$$

$$g_{m0} = 10mS$$

Knowing the current and maximum transconductance, we can find $\diamondsuit_{\diamondsuit}$ through the use of Equation 10.2.4.

$$\frac{g_m}{g_{m0}} = \sqrt{\frac{I_D}{I_{DSS}}}$$

$$g_m = g_{m0}\sqrt{\frac{I_D}{I_{DSS}}}$$

$$g_m = 10mS\sqrt{\frac{4.3mA}{15mA}}$$

$$g_m = 5.35mS$$

$$A_v = -\frac{g_m r_L}{g_m r_S + 1}$$

$$A_v = -\frac{5.35mS(2.2k\Omega||4.7k\Omega)}{5.35mS \times 0\Omega + 1}$$

$$A_v = -8.02$$

Example 2.3.2

Determine the voltage gain and input impedance for the circuit shown in Figure 2.3.3 . Assume I_{DSS}=24 mA and $V_{GS(off)}$=–4 V.

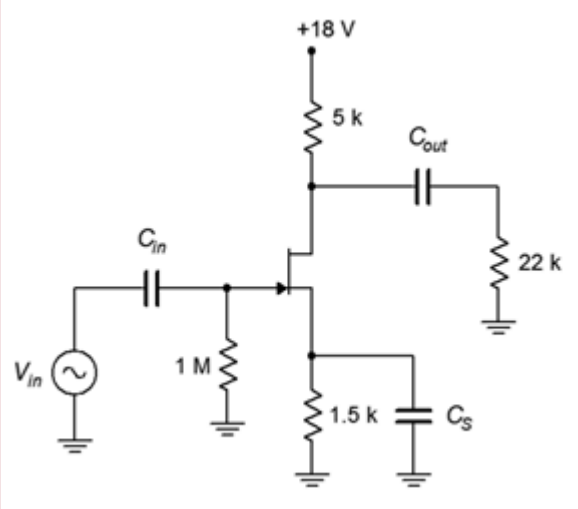


Figure 2.3.4 : Circuit for Example 2.3.2

This is an unswamped common source amplifier with self bias. Once again, we can determine $Z_{in(gate)}$ via inspection.

$$Z_{in} = Z_{in(gate)}||R_G$$

$$Z_{in} \approx 1M\Omega$$

To find the voltage gain, we'll first need to find g_{m0}. Also, we can take a shortcut and find the normalized drain current from the self bias graph instead of finding g_m itself.

$$g_{m0} = -\frac{2I_{DSS}}{V_{GS(off)}}$$

$$g_{m0} = -\frac{48mA}{-4V}$$

$$g_{m0} = 12mS$$

R_S is 1.5 kΩ, therefore $g_{m0}R_S$=18. From the self bias graph this produces a normalized drain current of 0.08.

$$g_m = g_{m0}\sqrt{\frac{I_D}{I_{DSS}}}$$

$$g_m = 12mS\sqrt{0.08}$$

$$g_m = 3.4mS$$

Again, there is no swamping so R_S=0. The gain formula reduces to

$$A_v = -g_m r_L$$

$$A_v = -3.4mS(22k\Omega||5k\Omega)$$

$$A_v = -13.9$$

We will now turn our attention to the effect of swamping. As in the BJT case, we expect to sacrifice gain and in return, see an improvement in distortion. We shall examine this through the use of a simulation.

COMPUTER SIMULATION

A common source amplifier using self bias is entered into the simulator as shown in Figure 2.3.5.

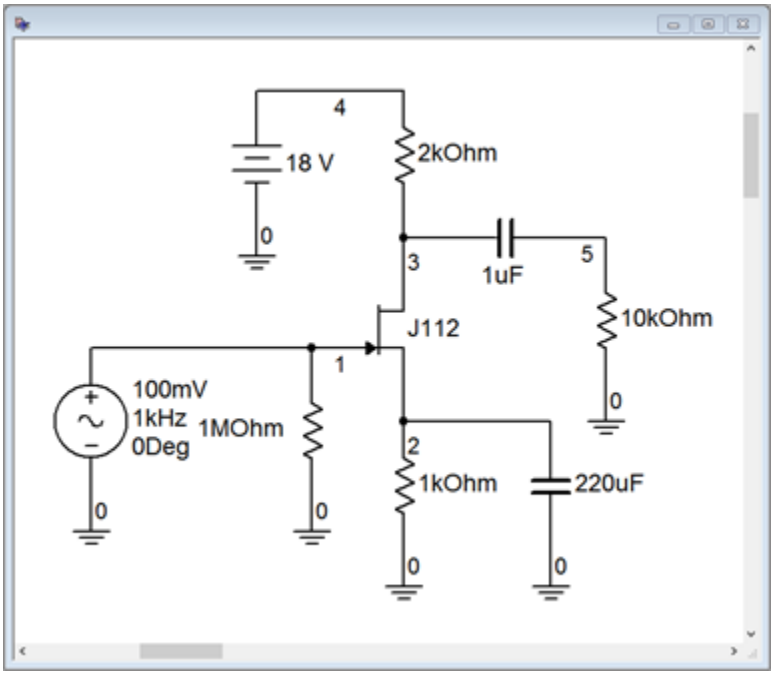


Figure 2.3.5: Common source amplifier in simulator.

Reasonable device values for this model are I_{DSS}=40 mA and $V_{GS(off)}$=−2.3 V. Based on these we find

$$g_{m0} = -\frac{2I_{DSS}}{V_{GS(off)}}$$

$$g_{m0} = -\frac{80mA}{-2.3V}$$

$$g_{m0} = 34.8mS$$

Given R_S=1 k Ω , the self-bias equation yields

$$I_D = 2I_{DSS}\left(\frac{1 + g_{m0}R_S - \sqrt{1 + 2g_{m0}R_S}}{(g_{m0}R_S)^2}\right)$$

$$I_D = 2I_{DSS}\left(\frac{1 + 34.8 - \sqrt{1 + 2 \times 34.8}}{(34.8)^2}\right)$$

$$I_D = 1.81mA$$
$$g_m = g_{m0}\sqrt{\frac{I_D}{I_{DSS}}}$$
$$g_m = 34.8mS\sqrt{\frac{1.81mA}{40mA}}$$
$$g_m = 7.4mS$$
$$A_v = -g_m r_L$$
$$A_v = -7.4mS(2k\Omega||10k\Omega)$$
$$A_v = -12.3$$

The results of a transient analysis are shown in Figure 2.3.6 for a 100 mV peak input signal.

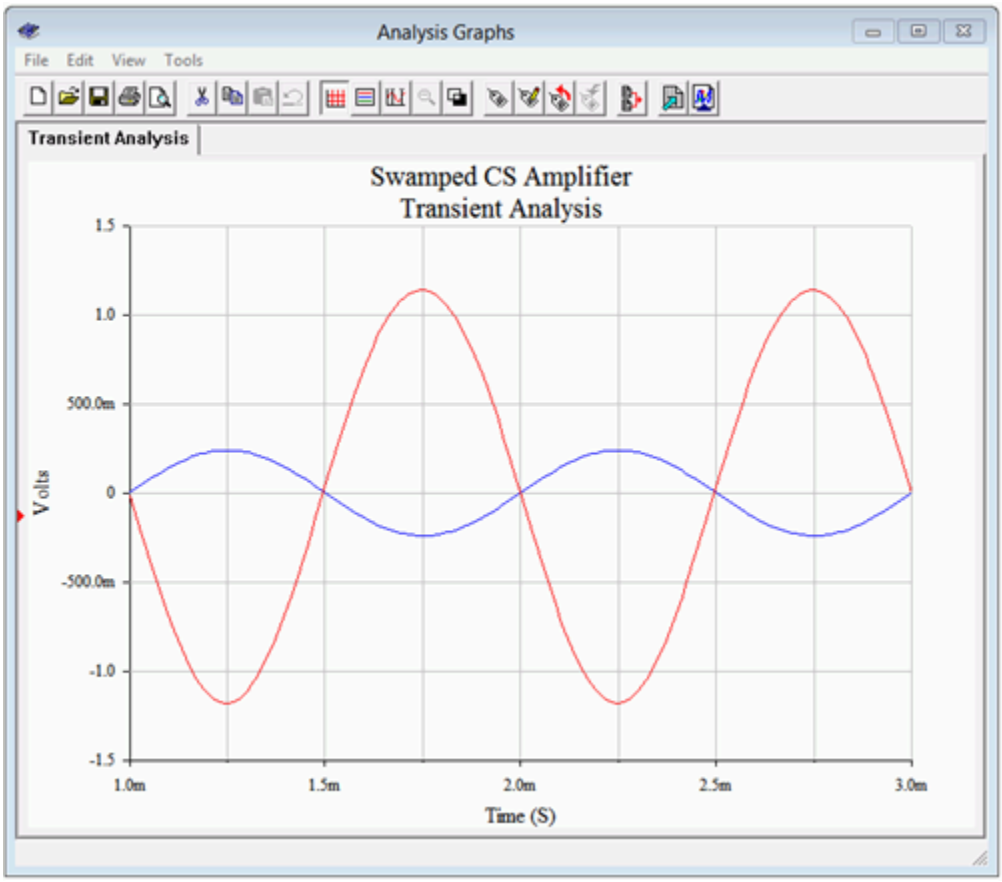


Figure 2.3.7: Transient analysis of the swamped common source amplifier.

The input signal was raised to 240 mV peak in order to keep the output signals of the two versions at the same amplitude. The symmetry appears to be better here and the gain works out to –4.85, just a few percent low.

Total harmonic distortion (THD) analysis is performed next. The results are shown in Figures 2.3.8 and 2.3.9 . To keep the comparison fair, the input levels are adjusted to maintain similar output voltages. The non-swamped results are seen in Figure 2.3.8 , and as expected based on the waveform asymmetry, the THD is relatively high at roughly 4%. The swamped version scores better at just over 1.6%, although this is still not stellar performance.

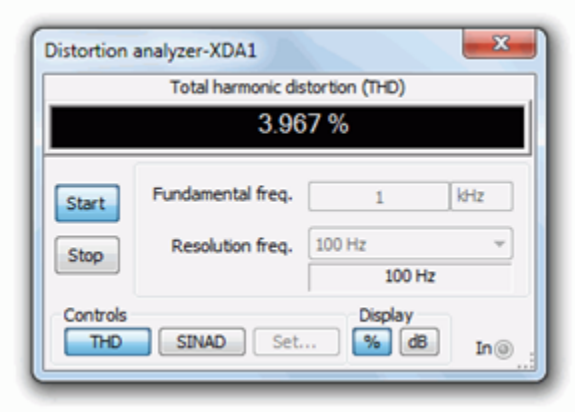


Figure 2.3.8: THD of common source amplifier.

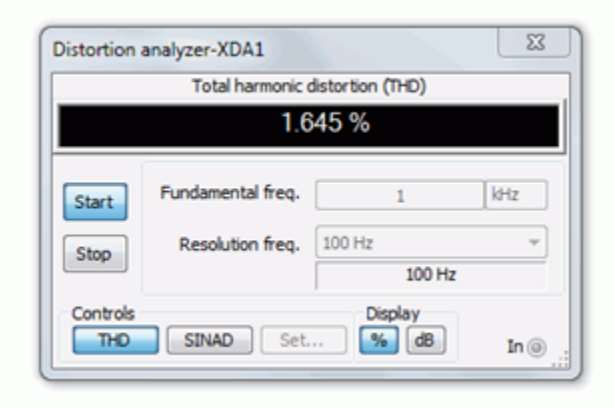


Figure 2.3.9: THD of swamped common source amplifier.

It is interesting to note that while the voltage gain of the swamped amplifier has dropped to 41% of the non-swamped gain, the THD has dropped to 41% of the nonswamped THD. In fact, given the square-law nature of the characteristic curve, we would expect the distortion to be lower if we used smaller signals. To verify this, the THD simulation is run again for the swamped amplifier, but now using an input signal ten times smaller at only 24 mV peak. The result is shown in Figure 2.3.10 .

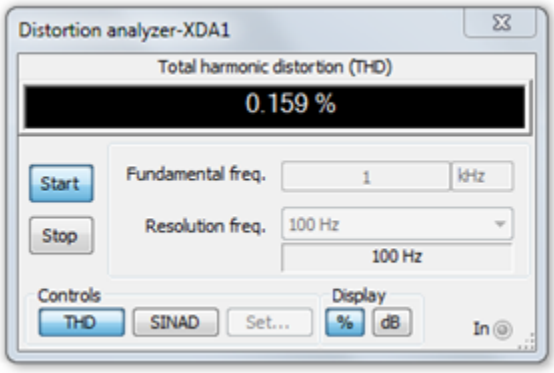


Figure 2.3.10: THD of swamped common source amplifier with reduced signal level.

The resulting THD is markedly lower for an order of magnitude improvement. We're now at least approaching "hi-fi" territory.

2.4 COMMON DRAIN AMPLIFIER

The common drain amplifier is analogous to the common collector emitter follower. The JFET version is also known as a source follower. The prototype amplifier circuit with device model is shown in Figure 2.4.1 . As with all voltage followers, we expect a non-inverting voltage gain close to unity, a high R_{in} and low R_{out}.

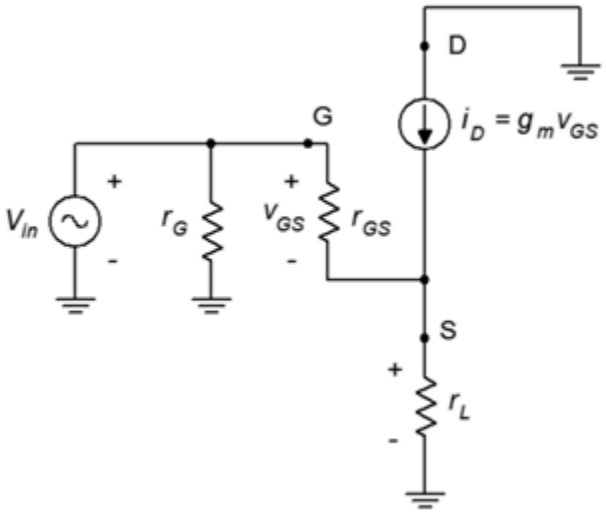


Figure 2.4.1: Common drain (source follower) prototype.

The input signal is presented to the gate terminal while the output is taken from the source. Many bias circuits may be used here as long as they do not have a grounded source terminal such as constant voltage bias.

VOLTAGE GAIN

In order to develop an equation for the voltage gain, A_v, we follow the same path we took with the common source amplifier earlier in this chapter. First, we start with the fundamental definition, namely that voltage gain is the ratio of v_{out} to v_{in}, and proceed by expressing these voltages in terms of their Ohm's law equivalents.

$$A_v = \frac{v_{out}}{v_{in}} = \frac{v_S}{v_G}$$

$$A_v = \frac{i_D r_L}{i_D r_L + v_{GS}}$$

$$A_v = \frac{g_m v_{GS} r_L}{g_m v_{GS} r_L + v_{GS}}]$$

$$A_v = \frac{g_m r_L}{g_m r_L + 1}$$

(2.4.1)

Equation 2.4.1 is very similar to the gain equation derived for the swamped common source amplifier; the notable changes being the lack of the minus sign indicating that this circuit does not invert the signal, and $\diamondsuit_{\diamondsuit}$ replacing $\diamondsuit_{\diamondsuit}$ in the denominator. It is worth remembering that $\diamondsuit_{\diamondsuit}$ here is the AC source resistance while in the common source amplifier $\diamondsuit_{\diamondsuit}$ is the AC drain resistance. To avoid potential confusion, this equation could also be written as

$$A_v = \frac{g_m r_S}{g_m r_S + 1}$$

(2.4.2)

In any event, the goal is to make sure that $\diamondsuit_{\diamondsuit}\diamondsuit_{\diamondsuit} \gg 1$. By doing so, the voltage gain will be very close to unity.

INPUT IMPEDANCE

The analysis for common drain input impedance is virtually identical to that for the swamped common source amplifier. The result is replicated here for convenience.

$$Z_{in} = r_G || r_{GS} \approx r_G$$

(2.4.3)

OUTPUT IMPEDANCE

In order to investigate the output impedance, we'll separate the load resistance from the source bias resistor, as shown in Figure 2.4.2.

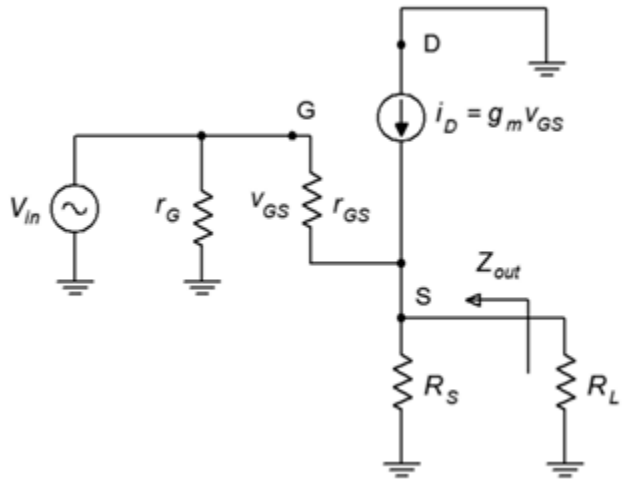


Figure 2.4.2: Common drain output impedance analysis.

From the position of $\diamondsuit_{\diamondsuit}$, looking back toward the source we find $\diamondsuit_{\diamondsuit}$ in parallel with the impedance looking back into the source terminal. The voltage at this node is $\diamondsuit_{\diamondsuit}\diamondsuit_{\diamondsuit}$ and the current entering this node is $\diamondsuit_{\diamondsuit}$. The ratio of the two must yield the impedance looking into the source.

$$Z_{source} = \frac{v_{GS}}{i_D}$$

$$Z_{source} = \frac{v_{GS}}{g_m v_{GS}}$$

$$Z_{source} = \frac{1}{g_m}$$

(2.4.4)

Therefore, the output impedance is

$$Z_{out} = R_S || \frac{1}{g_m}$$

(2.4.5)

We can expect this value to be much smaller than the output impedance of typical common source amplifiers.

Example 2.4.1

For the follower shown in Figure 2.4.3 , determine the input impedance and output voltage. Assume ◈◈◈=100 mV, ◈◈◈◈=30 mA, ◈◈◈(◈◈◈)=–2 V.

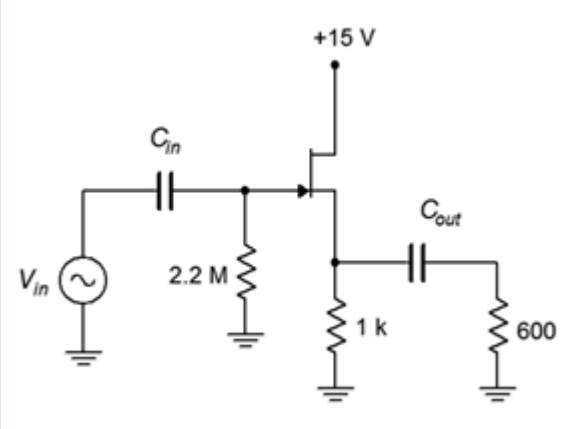


Figure 2.4.3: Circuit for Example 2.4.1 .

This is a follower using self bias. We'll find ◈◈ via the self bias graph.

$$g_{m0} = -\frac{2I_{DSS}}{V_{GS(off)}}$$

$$g_{m0} = -\frac{60mA}{-2V}$$

$$g_{m0} = 30mS$$

◈◈ is 1 k Ω , yielding 30 for ◈◈0◈◈ . The normalized drain current from the self bias graph is approximately 0.05.

$$g_m = g_{m0}\sqrt{\frac{I_D}{I_{DSS}}}$$

$$g_m = 30mS\sqrt{0.05}$$

$$g_m = 6.71mS$$

$$A_v = \frac{g_m r_S}{g_m r_S + 1}$$

$$A_v = \frac{6.71mS(1k\Omega||600\Omega)}{6.71mS(1k\Omega||600\Omega) + 1}$$

$$A_v = 0.716$$

Thus ◇◇◇◇ is 71.6 mV. By inspection, ◇◇◇ may be approximated as 2.2 M Ω .

Example 2.4.2

For the circuit shown in Figure 2.4.4 , determine the input impedance and output voltage. Assume ◇◇◇=100 mV, ◇◇◇◇=36 mA, ◇◇◇(◇◇◇)=3 V.

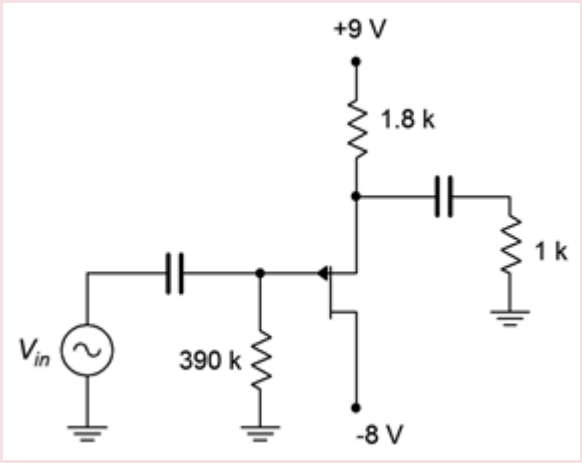


Figure 2.4.4: Circuit for Example 2.4.2 .

This follower uses combination bias with a P-channel JFET. Note that the source is at the top. We'll find ◇◇ via the combination bias graph for ◇=3 (◇=◇◇◇/◇◇◇(◇◇◇)) .

$$g_{m0} = \frac{2I_{DSS}}{V_{GS(off)}}$$

$$g_{m0} = \frac{72mA}{3V}$$

$$g_{m0} = 24mS$$

◈◇ is 1.8 kΩ, yielding 43.2 for ◈◇0◈◇. The normalized drain current from the ◈=3 combination bias graph is approximately 0.17.

$$g_m = g_{m0}\sqrt{\frac{I_D}{I_{DSS}}}$$

$$g_m = 24mS\sqrt{0.17}$$

$$g_m = 9.9mS$$

$$A_v = \frac{g_m r_S}{g_m r_S + 1}$$

$$A_v = \frac{9.9mS(1k\Omega||1.8k\Omega)}{9.9mS(1k\Omega||1.8k\Omega) + 1}$$

$$A_v = 0.864$$

Thus ◈◇◇◇ is 86.4 mV. By inspection, ◈◇◇ may be approximated as 390 k Ω .

2.5 MULTI-STAGE AND COMBINATION CIRCUITS

The rules for multi-stage circuits utilizing JFETs are the same as those discussed for BJTs: Steps must be taken to ensure that the bias of one stage does not adversely affect the bias of surrounding stages (typically by using coupling capacitors or going to a DC coupled system), the load for a given stage will be the input impedance of the following stage, the input impedance of the system will be the input impedance of the first stage, and the system gain will be the product of the individual stage gains.

Keeping those items in mind, there are no limits concerning mixing BJTs with JFETs, or mixing N-channel with P-channel devices. There are certain practical issues, however, that might dictate where certain devices are used. JFETs, due to their high input impedance and modest gain potential, tend to be used at the front end of amplifying systems. Their comparatively low self-noise is also a bonus at this location. BJTs, on the other hand, have high gain potential and tend to be used in the remaining stages. Their high distortion can be tamed through swamping.

To examine the possibilities, let's walk through the mixed, multi-stage amplifier presented in Figure 2.5.1 .

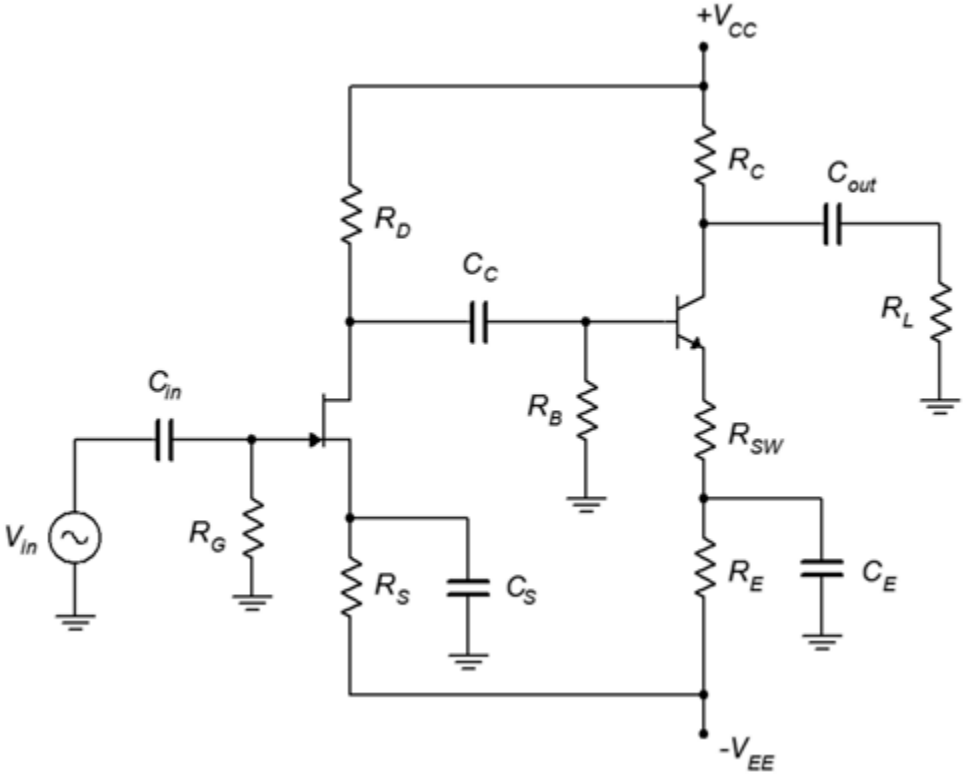


Figure 2.5.1: Two-stage JFET/BJT amplifier.

his amplifier uses a bipolar power supply which gives the designer a lot of flexibility. The first stage consists of a JFET common source amplifier. It utilizes combination bias (notice that ⍰◇ is connected to the shared negative supply, ⍰◇◇, that also serves as ⍰◇◇). ⍰◇ bypasses the source resistor so

this stage does not use swamping. Distortion should not be an issue unless the input signal is fairly large. The load for this stage is ◈◈ in parallel with the input impedance to the second stage (coupling capacitor ◈◈ will appear ideally as a short for signal frequencies).

The second stage utilizes an NPN BJT configured as a swamped common emitter amplifier. It utilizes two-supply emitter bias. It's input impedance is the parallel combination of ◈◈ and ◈◈◈(◈◈◈◈) . The base input impedance, in turn, is a function of ◈ and ◈◈◈ (◈′◈ will have only a small impact due to the swamping resistor). The load for this stage will be ◈◈ in parallel with ◈◈. That value divided by ◈◈◈ will give the approximate stage gain (again, ◈′◈ will have little impact). Although the second stage will be dealing with a larger signal, distortion will be mitigated by the swamping resistor.

The system gain will be the product of the two stage gains. As they both invert the signal, the inversion of the inversion will lead to an output signal that is in phase with the input signal. The system input impedance will depend on the JFET first stage and can be approximated to be equal to ◈◈, at least at low frequencies.

2.6 OHMIC REGION OPERATION

As noted in the previous chapter, the JFET's operational curves span three regions. Two have been discussed: the constant current region is where the normal amplifiers and followers are biased, and breakdown is a region to be avoided due to potential damage. The third region is known as the ohmic region, or triode region. It occurs in the area where V_{DS} is less than the pinch-off voltage, V_P. In this area, the device behaves more like a resistor than like a current source. If we were to examine a family of drain curves, like those of Figure 1.2.3, and magnify the area near the origin, we would see something like the plot in Figure 2.6.1 .

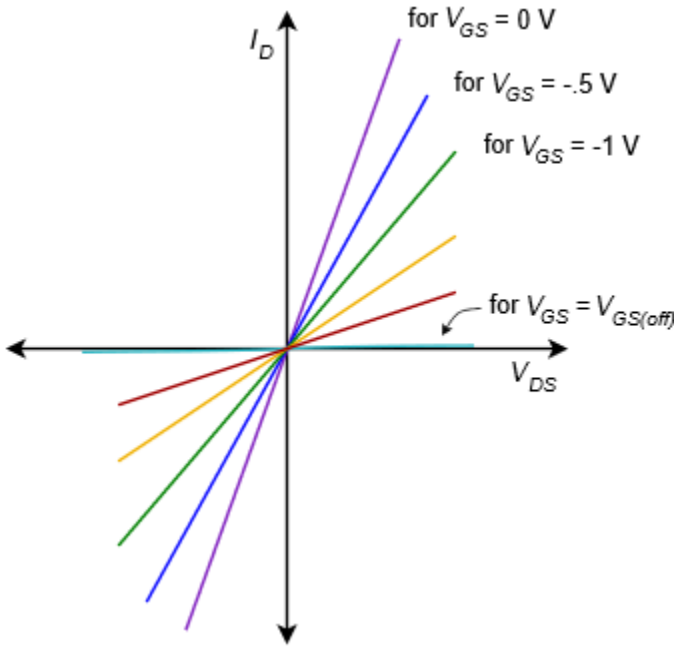


Figure 2.6.1: FET ohmic region.

If V_{DS} is a small value, typically less than 100 mV or so, each of the curves appears as a straight line. Further, the slope of that line is a function of the gate-source voltage, V_{GS}. The closer V_{GS} is to 0 V, the steeper the slope (violet line) and the closer V_{GS} is to $V_{GS(off)}$, the more shallow the slope (dark red line). Finally, if $V_{GS}=V_{GS(off)}$ the slope is nearly zero (aqua line). Because this is a plot of drain current versus drain-source voltage, the slope indicates the conductance of the channel. In somewhat more useful terms, we can say that the reciprocal of the slope indicates the resistance of the channel. Therefore, if V_{GS}=0 V, the channel resistance will be at its minimum, and when $V_{GS}=V_{GS(off)}$, the channel resistance will be at its maximum. The maximum channel resistance can be quite high, well into the hundreds of kilo-ohms. The minimum channel resistance varies considerably from device to device. It is found on a data sheet as $R_{DS(on)}$. $R_{DS(on)}$ can be as small as a few ohms for specialized JFETs and as large as hundreds of ohms for general purpose devices.[1] For example, the data sheet for the J111 series JFETs found in Figure 10.3.1 shows maximum

1. $R_{DS(on)}$ can be as little as a few milliohms for specialized high power MOSFETs (Chapters 12 and 13).

values of 30 Ω , 50 Ω and 100 Ω for the J111, J112 and J113, respectively. The channel resistance does not follow a linear relation with ⍰⍰.

To be more specific, in this region the drain current no longer follows the characteristic equation we used for biasing (Equation 1.2.1). The drain current equation in the ohmic region is:

$$I_D = \frac{V_{DS}}{V_P} 2I_{DSS} \left(\left(1 - \frac{V_{GS}}{V_P} \right) - \frac{V_{DS}}{V_P} \right)$$

(2.6.1)

Where ⍰⍰=|⍰⍰⍰(⍰⍰⍰)| and ⍰⍰⍰ is to be taken as an absolute value and lies between 0 and ⍰⍰.

Recalling that, in general, ⍰⍰⍰=⍰⍰⍰/⍰⍰, we can substitute Equation 2.6.1 for ⍰⍰ and, after including the definition of ⍰⍰0, arrive at an expression for ⍰⍰⍰:

```
*** QuickLaTeX cannot compile formula:
\[r_{DS} = \frac{V_P}{g_{m0} \left( V_P -V_{GS} - \frac{V_{DS}}{2} \right) \]

*** Error message:
File ended while scanning use of \frac .
Emergency stop.
```

For small values of ⍰⍰⍰, this reduces to a simple equation:

$$r_{DS} = \frac{V_P}{g_{m0}(V_P - V_{GS})}$$

(2.6.2)

What we have created here is voltage-controlled resistor. Equation 2.6.2 shows that the resistance of the channel is a function of the gate-source voltage: the channel resistance will be at its minimum (⍰⍰⍰(⍰⍰⍰)) when ⍰⍰⍰=0 V, and it approaches infinity when ⍰⍰⍰ equals ⍰⍰⍰(⍰⍰⍰⍰). Generally, there are two applications that make use of the ohmic region: an electronic rheostat/potentiometer and an analog switch. A simple circuit that can be used for either application is shown in Figure 2.6.2.

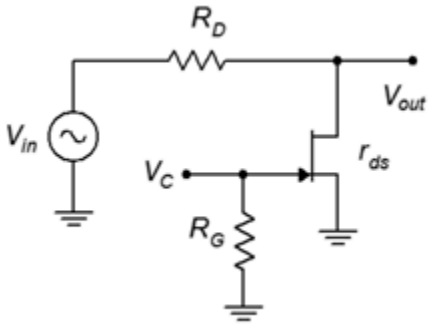


Figure 2.6.2: Using the JFET as a voltagecontrolled resistor or switch.

Note that no external bias is applied to the circuit. Instead, a control voltage, ⍰⍰, is applied to the gate and the input signal is applied to a resistor attached to the drain terminal. The output is taken across the JFET's drain-source.

The idea behind this circuit is the basic resistive voltage divider. The JFET's channel resistance, ⍰⍰⍰, forms a voltage divider along with ⍰⍰.

$$v_{out} = V_{in}\frac{r_{DS}}{r_{DS} + R_D}$$

If $r_{DS} \gg R_D$, v_{out} approaches V_{in}. Conversely, if $r_{DS} \ll R_D$, v_{out} approaches zero. Normally, R_D is set somewhere between the maximum and minimum channel resistances in order to obtain the widest range of operation.

As the control voltage V_C is V_{GS}, then V_C controls the size of r_{DS}. If we set V_C to 0 V, V_{GS} is very small and thus $r_{DS} \approx 0$. On the other hand, if V_C is set to a large negative potential (beyond $V_{GS(OFF)}$), then $r_{DS} \approx R_D$. If V_C is set between these extremes then r_{DS} will be somewhere in the middle range. If V_C is continuously variable, then the circuit behaves like a solid-state potentiometer. If, in contrast, V_C is only set at the limits, then the circuit behaves like a switch, either allowing or preventing the signal from transferring through. This is known as an analog switch.

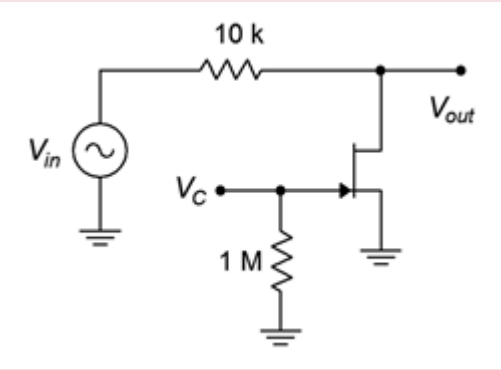
A single JFET/resistor combination as shown in Figure 2.6.2 will have limited isolation as an analog switch and only a modest range of adjustment when used as a voltage-controlled potentiometer. To improve performance, multiple circuits can be cascaded or other JFETs can be added to create a "pi" attenuator network.

This voltage-controlled resistor has a huge advantage over traditional electromechanical potentiometers and switches: speed. In this circuit, the resistance can be changed at very high rates, essentially, as fast as V_C can change. Consequently, it would be no big deal to switch the input signal on and off at rates well over 100,000 times per second. No mechanical switch or potentiometer can hope to perform anywhere near that speed, and any attempt to do so would lead to the devices burning up from the friction. In general, flaming potentiometers are frowned upon during the design and development process, although it would make a decent name for an indie rock band. Another advantage is that a switch can be thrown "remotely", that is, we only need to route the control voltage to the switch operator, not the signal itself. This can reduce system noise. It's also easier to implement if the switch is being "thrown" programmatically, such as via a microcontroller.

Example 2.6.1

For the circuit shown in Figure 2.6.3 , if the input signal is 50 mV, determine the output voltage for V_C=0 VDC and –6 VDC. Assume $V_{GS(OFF)}$=–5 V, $r_{DS(ON)}$=30 Ω and R_D=800 k Ω .

For V_C=0 VDC, the channel resistance will be at its minimum of $r_{DS(ON)}$.



2.6.3: Circuit for Example 2.6.1.

$$V_{out} = V_{in}\frac{r_{DS(on)}}{R_D + r_{DS(on)}}$$

$$V_{out} = 50mV\frac{30\Omega}{10k\Omega + 30\Omega}$$

$$V_{out} = 0.15mV$$

The signal has been reduced by a factor of over 330. That's not as good as a mechanical switch but if we cascaded two of these the overall reduction would be more than 100,000:1.

For ◈◈=–6 VDC, the channel resistance will be at its maximum of ◈◈◈(◈◈◈).

```
  *** QuickLaTeX cannot compile formula:
\[V_{out} = V_{i n} \frac{r_{DS (off )}{R_D+r_{DS (off )}}  \]

*** Error message:
File ended while scanning use of \frac .
Emergency stop.
```

$$V_{out} = 50mV\frac{800k\Omega}{10k\Omega + 800k\Omega}$$

$$V_{out} = 49.4mV$$

This represents nearly 99% of the input signal, so the signal is passed through cleanly.

COMPUTER SIMULATION

To verify the results of the preceding example, the circuit is entered into a simulator as shown in Figure 2.6.4 . A J111 JFET model is used which has parameters similar to those used in the example.

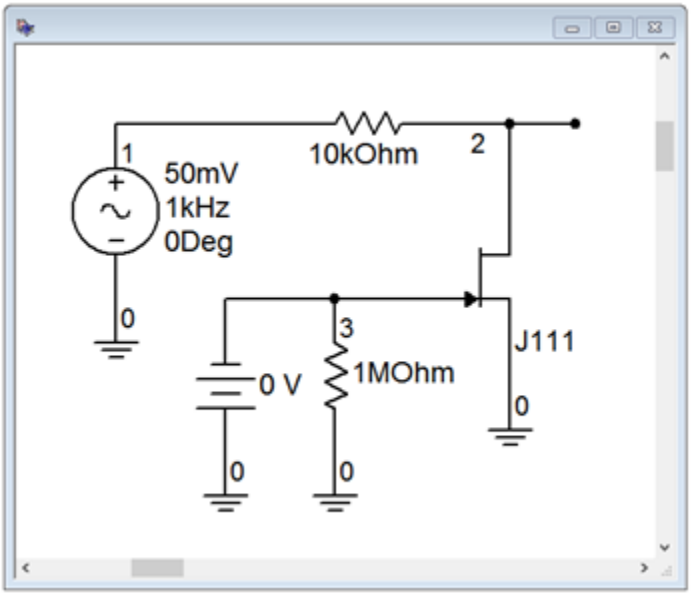


Figure 2.6.4: Analog switch in simulator.

A transient analysis is run twice; the first time is with a control voltage of 0 V and the second with a control voltage of –6 V. In the first case, the output should be just a small residual and in the second, we should see the full input signal. The results of the first trial are shown in Figure 2.6.5 while the second is shown in Figure 2.6.6 .

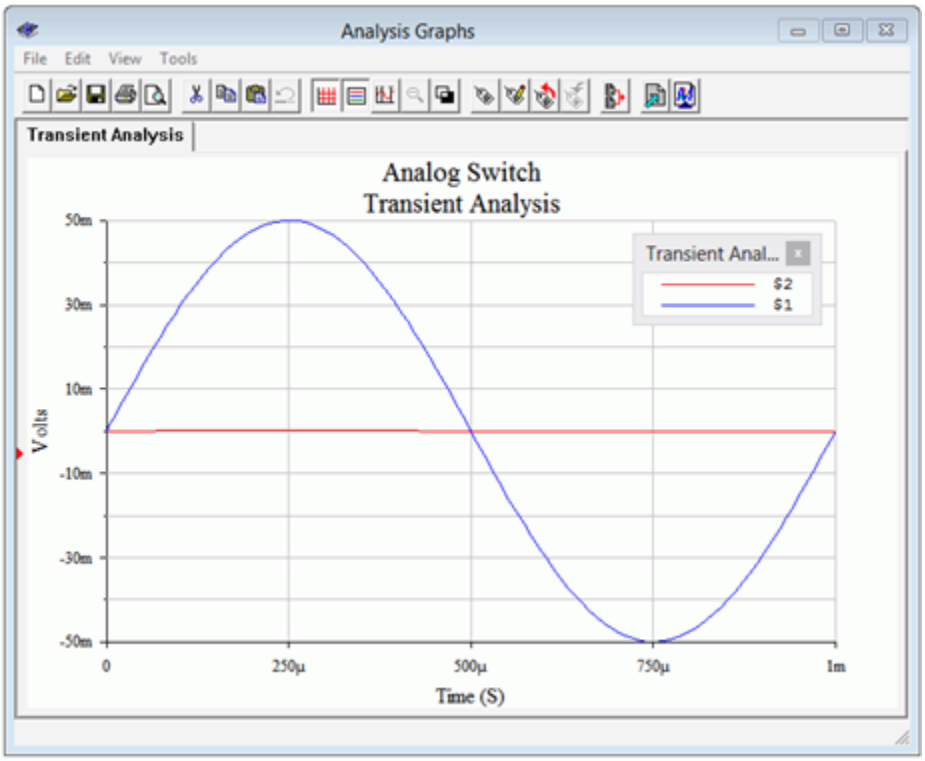


Figure 2.6.5: Transient analysis using ◇◇=0 V.

With ◇◇ at 0 V, the output trace (red, at node 2) is nearly flat. The precise value of its peak is 0.167 mV, not far from the value calculated in Example 2.6.1.

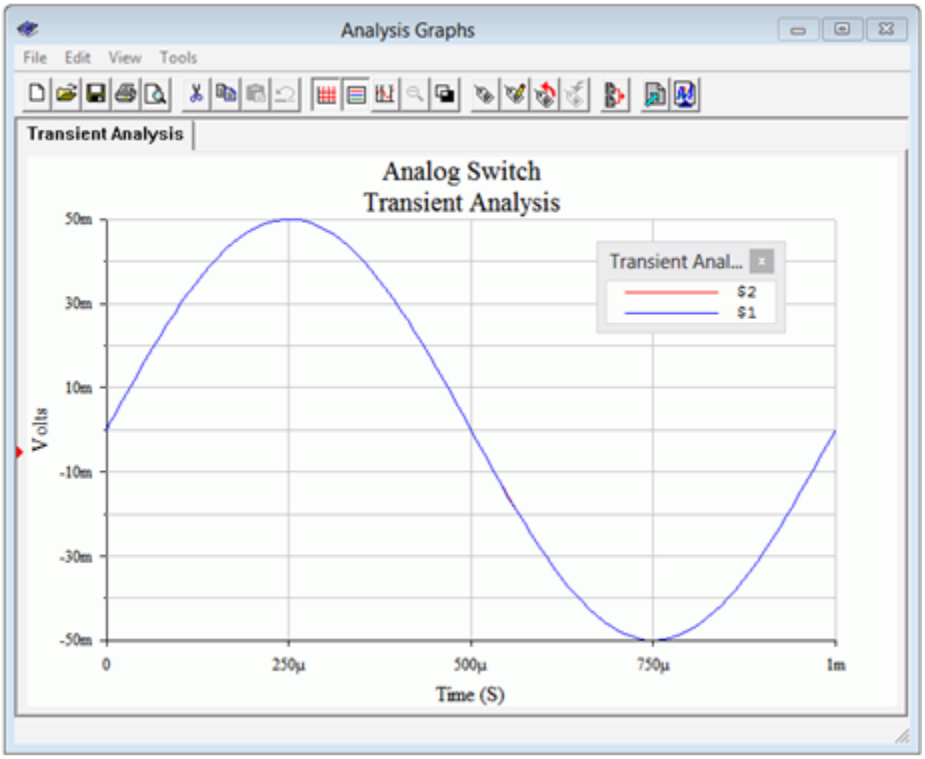


Figure 2.6.6: Transient analysis using ◇◇=−6 V.

In contrast, when ◇◇=−6 V the JFET is off, offering a high impedance and no loss of signal. At first glance, it may appears as though the output trace is missing but what has happened is that it is hidden behind the input trace (blue, node 1). The amplitudes are virtually identical so the blue trace completely obscures the red trace.

2.7 SUMMARY

JFETs can be used to create both voltage amplifiers and voltage followers. The common source configuration is similar to the BJT's common emitter configuration. It offers voltage gain with signal inversion. The amplifier can be built upon any of the bias schemes presented in the preceding chapter. Biasing circuits that made use of a source resistor, such as self bias and combination bias, may also use swamping. Swamping will decrease available voltage gain but reduce distortion.

The JFET voltage follower, or source follower, is similar to the BJT's emitter follower. It offers a voltage of gain of nearly unity without inversion, a high input impedance and a low output impedance.

In general, JFETs do not offer as high of a gain as BJTs. The parameter comparable to the BJT's β is the transconductance, g_m. Further, they tend to offer very high input impedance values compared to BJTs. This is due to using a reverse biased junction instead of a forward biased junction.

JFETs can also be used in their ohmic region to create voltage-controlled resistances and analog switches. A key parameter in these applications is the minimum channel resistance, $r_{DS(on)}$.

Review Questions

1. What are the functional differences between common source and common drain amplifiers?
2. Compare and contrast common source amplifiers to common emitter amplifiers.
3. Compare and contrast common drain followers to common collector followers.
4. What is the ohmic region?
5. What is an analog switch and how does it function?

2.8 EXERCISES

ANALYSIS PROBLEMS

1. For the amplifier of Figure 11.8.1 , determine $\square_{\square\square}$ and $\square_{\square}$. $\square_{\square\square\square}$ = 12 mA, $\square_{\square\square(\square\square\square)}$ = –2 V, $\square_{\square\square}$ = 15 V, $\square_{\square}$ = 220 k Ω , $\square_{\square}$ = 2 k Ω , $\square_{\square}$ = 3.3 k Ω , $\square_{\square}$ = 330 Ω.
2. For the amplifier of Figure 11.8.1 , determine $\square_{\square\square}$ and $\square_{\square\square\square}$. $\square_{\square\square}$ = 50 mV, $\square_{\square\square\square}$ = 15 mA, $\square_{\square\square(\square\square\square)}$ = –3 V, $\square_{\square\square}$ = 20 V, $\square_{\square}$ = 270 k Ω , $\square_{\square}$ = 2 k Ω , $\square_{\square}$ = 6.8 k Ω , $\square_{\square}$ = 270 Ω.

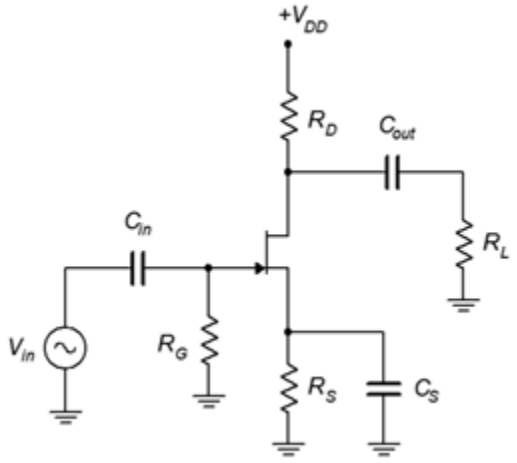


Figure 2.8.1

3. For the amplifier of Figure 11.8.2 , determine $\square_{\square\square}$ and $\square_{\square\square\square}$. $\square_{\square\square}$= 60 mV, $\square_{\square\square\square}$ = 10 mA, $\square_{\square\square(\square\square\square)}$ = –3 V, $\square_{\square\square}$ = 20 V, $\square_{\square\square}$ = –6 V, $\square_{\square}$ = 270 k Ω , $\square_{\square}$ = 2 k Ω , $\square_{\square}$ = 4 k Ω , $\square_{\square}$ = 1.8 k Ω , $\square_{\square\square}$ = 200 Ω .

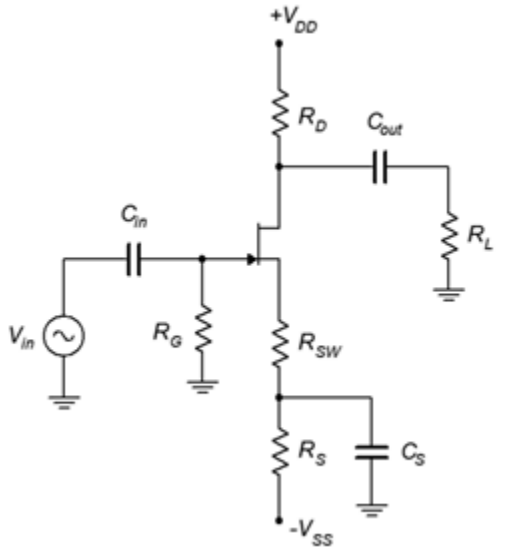


Figure 2.8.2

4. For the amplifier of Figure 11.8.2 , determine ◇◇ and ◇. ◇◇◇ = 12 mA, ◇◇(◇◇) = –2 V, ◇◇ = 18 V, ◇◇ = –4 V, ◇ = 330 k Ω , ◇ = 2.2 k Ω , ◇ = 10 k Ω , ◇ = 3 k Ω , ◇◇ = 100 Ω .

5. For the amplifier of Figure 11.8.3 , determine ◇◇ and ◇. ◇◇◇ = 12 mA, ◇◇(◇◇) = –2 V, ◇◇ = 18 V, ◇◇ = –4 V, ◇ = 390 k Ω , ◇ = 2.2 k Ω , ◇ = 1 k Ω , ◇ = 20 k Ω.

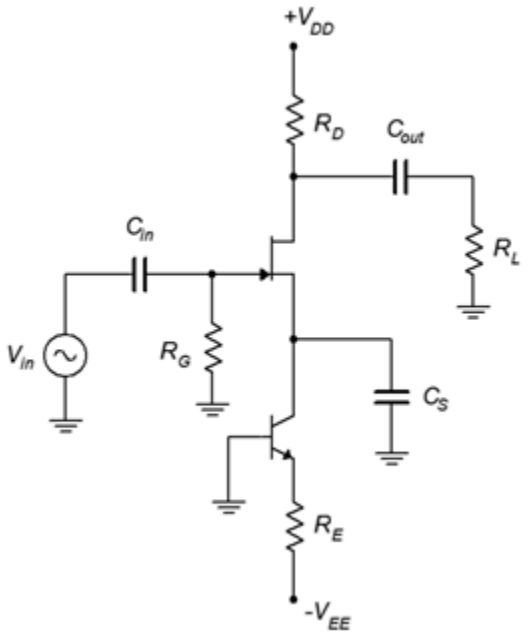


Figure 2.8.3

6. For the amplifier of Figure 11.8.3 , determine ◇◇ and ◇◇◇. ◇◇ = 70 mV, ◇◇◇ = 12 mA, ◇◇(◇◇◇) = –2 V, ◇◇ = 18 V, ◇◇ = –4 V, ◇ = 390 k Ω , ◇ = 2.2 k Ω , ◇ = 20 k Ω.

7. For the circuit of Figure 11.8.4 , determine ◇◇ and ◇. ◇◇◇ = 12 mA, ◇◇(◇◇◇) = –2 V, ◇◇ = 10 V, ◇ = 220 k Ω , ◇ = 3.3 k Ω , ◇ = 330 Ω .

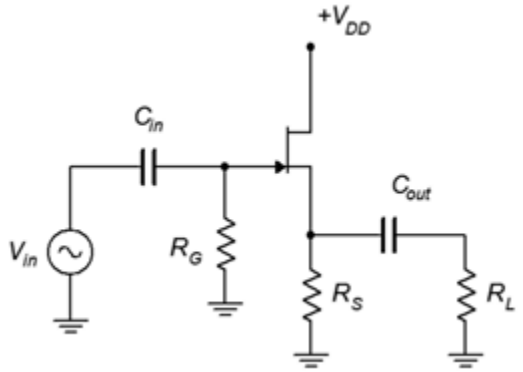


Figure 2.8.4

8. For the circuit of Figure 11.8.4 , determine ◆◆ and ◆◆◆. ◆◆ = 200 mV, ◆◆◆ = 15 mA, ◆◆(◆◆◆) = –3 V, ◆◆ = 12 V, ◆ = 270 k Ω , ◆ = 1.8 k Ω , ◆ = 270 Ω .
9. For the circuit of Figure 11.8.5 , determine ◆◆ and ◆◆◆. ◆◆ = 100 mV, ◆◆◆ = 10 mA, ◆◆(◆◆◆) = –3 V, ◆◆ = 15 V, ◆◆ = –6 V, ◆ = 470 k Ω , ◆ = 4 k Ω , ◆ = 1.8 k Ω.

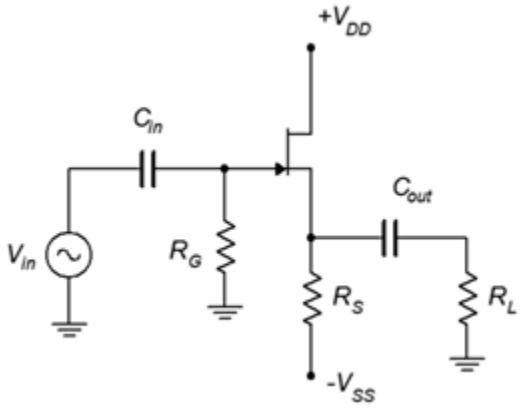


Figure 2.8.5

10. For the circuit of Figure 11.8.5 , determine ◆◆ and ◆. ◆◆◆ = 18 mA, ◆◆(◆◆◆) = –2 V, ◆◆ = 14 V, ◆◆ = –6 V, ◆ = 360 k Ω , ◆ = 10 k Ω , ◆ = 1 k Ω .
11. For the circuit of Figure 11.8.6 , determine ◆◆◆. ◆◆ = 100 mV, ◆◆(◆◆) = 50 Ω , ◆◆(◆◆◆) = 1 M Ω , ◆◆(◆◆◆) = –3 V, ◆ = –6 V, ◆ = 270 k Ω , ◆ = 6.8 k Ω .
12. For the circuit of Figure 11.8.6, determine ◆◆◆. ◆◆ = 100 mV, ◆◆(◆◆) = 75 Ω , ◆◆(◆◆◆) = 750 k Ω , ◆◆(◆◆◆) = –3 V, ◆ = 0 V, ◆ = 180 k Ω , ◆ = 5.1 k Ω .

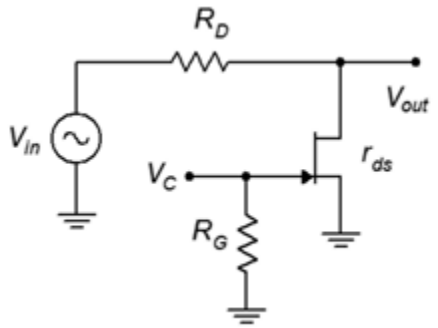


Figure 2.8.6

DESIGN CHALLENGE PROBLEMS

13. Following the circuit of Figure 11.8.2 , design an amplifier with a gain of at least 4 and an input impedance of at least 300 k Ω. ◆ = 10 k Ω . The JFET has the following parameters: ◆(◆) = –2 V, ◆ = 15 mA. Try to use standard resistor values.
14. Using the circuit of Figure 11.8.4 , design a follower with a gain of at least 0.7 and an input impedance of at least 500 k Ω. ◆ = 1 k Ω . The JFET has the following parameters: ◆(◆) = –3 V, ◆ = 20 mA. Try to use standard resistor values.

COMPUTER SIMULATION PROBLEMS

15. Utili ◆ g manufacturer's data sheets, find devices with the following specifications (typical) and verify them using the measurement techniques presented in the prior chapter. Device 1: ◆(◆) = –2 V, ◆ = 15 mA. Device 2: ◆(◆) = –3 V, ◆ = 20 mA.
16. Using the device model from the preceding problem, verify the design of Problem 13.
17. Using the device model from Problem 15, verify the design of Problem 14.

UNIT 3: METAL OXIDE SEMICONDUCTOR FETS (MOSFETS)

Learning Objectives

After completing this chapter, you should be able to:

- Discuss the functional differences between MOSFETs and JFETs.
- Draw and explain a basic DC bias model for a MOSFET.
- Graph the transconductance curves for both DE-MOS and E-MOS transistors, and describe their functional differences.
- Perform DC bias analysis on various MOSFET circuits.
- Explain necessary ESD precautions for MOS devices.

3.1 INTRODUCTION

The MOSFET shares many similarities with the JFET including very low gate current and being modeled as a voltage-controlled current source. It is also available in N- and P-channel varieties. Unlike the JFET, it has two variations: the depletion-enhancement mode variant, or DE-MOSFET; and the enhancement-only mode variant, or E-MOSFET. All of the bias types discussed for JFETs will work for DE-MOSFETs, plus a few others. EMOSFETs, on the other hand, require new biasing prototypes.

For AC analysis, both common source and common drain amplifier topologies may be realized with DE- and EMOSFETs. The equations for input impedance, voltage gain and the like are generally unchanged from the JFET. E-MOSFETs are also available as power devices. They have certain advantages over power BJTs, including higher speed and a negative temperature coefficient of transconductance which means they are less likely to suffer from thermal instabilities such as current hogging.

One item of practical importance is that MOSFETs are very susceptible to ESD (electrostatic discharge) and special precautions must be taken to prevent accidental damage to the device. Unlike both the JFET and the BJT, the MOSFET does not rely on a PN junction for its operation. Instead, it uses a charge-based system not unlike a capacitor. The gate is, in fact, insulated from the channel. For this reason it is sometimes referred to as an IGFET, which stands for Insulated Gate FET. This insulation layer will lead to very, very high input resistance due to extremely low gate current but also leads to the issue of ESD susceptibility.

3.2 THE DE-MOSFET

Like the JFET, the DE-MOSFET is based around the idea of modulating current flow through the drain-source channel by generating a depletion layer from a gate-source voltage. It achieves this through an entirely different process, though. To understand how the device is constructed, a simplified functional drawing of an N-channel DEMOSFET is shown in Figure 3.2.1 .

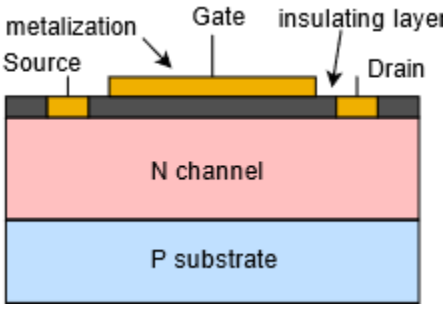


Figure 3.2.1: DE-MOSFET internal structure.

Here we see the N-type channel sitting on a P substrate. Drain and source leads are attached to either end. Above the channel is a very thin insulating layer (silicon dioxide is one possibility). Above this we have a metalization to which the gate terminal is attached. Note that there is no PN junction involved with the gate. To this we shall add external bias sources and limiting resistors, as shown in Figure 3.2.2 .

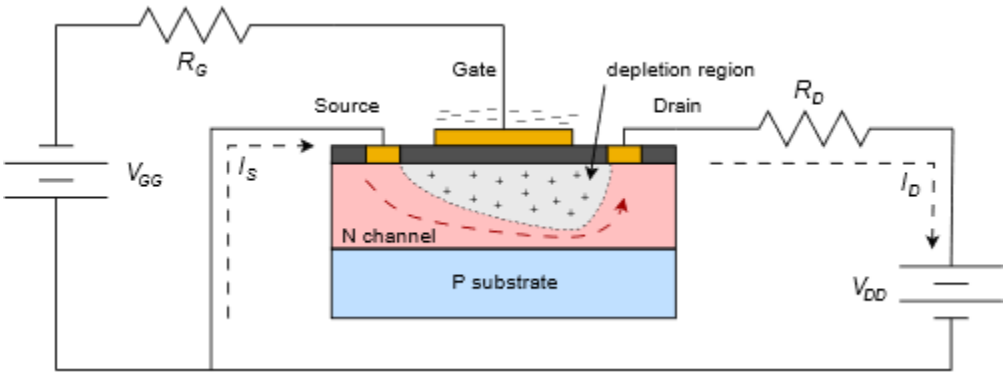


Figure 3.2.2: DE-MOSFET bias with electron flow.

The dashed lines represent electron current flow as in our previous device analyses. A positive supply, V_{DD}, is attached to the drain via a limiting resistor. A second supply, V_{GG}, is attached to the gate. Gate current can be approximated as zero, so $I_S = I_D$. If V_{GG} is zero, a certain amount of current will flow through the channel based on the channel's physical parameters and the applied drain-source potential. For relatively low values of V_{DD}, the channel will behave somewhat like a resistance. This is the same ohmic region as seen with the JFET. As V_{DD} increases, the channel will saturate and begin to behave like a constant current source. If V_{DD} is brought too high, the drain current increases sharply as the device enters the breakdown region. The general behavior mimics that of a JFET. Note

that the current moves laterally, across the device, so this type of construction is referred to as a lateral MOSFET.

If ◈◈◈ is set to a modest negative voltage, a depletion region will develop inside the channel. Basically, the gate is acting like one plate of a capacitor, the channel like the other plate, and the insulating layer is the dielectric. Just like a capacitor, the negative charge on the gate "plate" leads to an equivalent positive charge on the channel "plate". As the channel is made of N-type material, this action creates a region devoid of free charges, hence a depletion region. This depletion region will lead to pinch-off sooner, and thus a lower current in the saturation region. The more negative ◈◈◈ is made, the greater the depletion region and the lower the corresponding drain current. Eventually, if ◈◈◈ is brought negative enough, the channel will be blocked and no drain current will flow. This voltage is referred to as ◈◈◈(◈◈◈) (again). The current produced when ◈◈◈=0 V is likewise referred to as ◈◈◈◈. This mode of operation is referred to as depletion mode because of the depletion region that is created.

What makes the DE-MOSFET distinct from the JFET is what happens when ◈◈◈>0 volts. In a JFET, this would forward bias the junction and control would be lost. Here, however, a positive ◈◈◈ simply reverses the polarities associated with the gate and channel "plates". Thus, a positive ◈◈◈ enhances channel conductivity and drain current increases as ◈◈◈ is brought more positive. This mode of operation is called enhancement mode. This also means that ◈◈◈◈ is no longer the maximum drain current of which the device is capable. A characteristic curve is shown in Figure 3.2.3 , below.

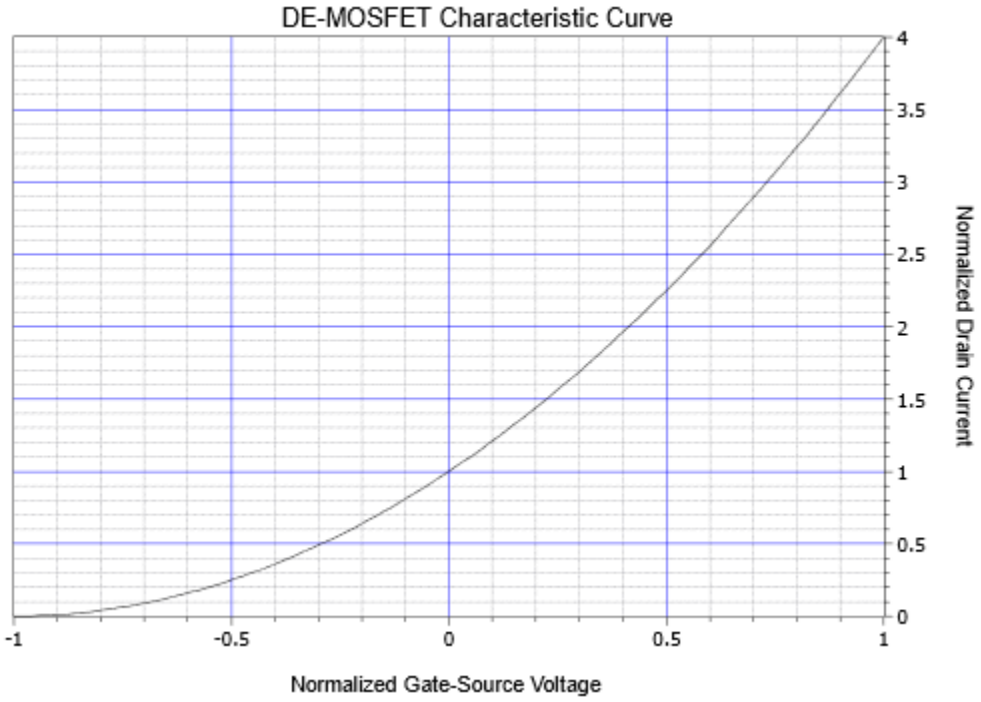


Figure 3.2.3: Characteristic curve for DEMOSFET (note: this uses –◈◈◈/◈◈◈(◈◈◈) for the normalized voltage so that the quadrants do not appear reversed compared to a typical device curve).

This curve is essentially the same curve as presented for the JFET with the exception that it extends into the first quadrant. This makes the DE-MOSFET a unique device in that it can operate in two different quadrants.

The device equation for operation when ◈◈◈>◈◈ is also the same, but with an extended range for ◈◈◈:

$$I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_{GS(off)}}\right)^2$$

(3.2.1)

Where

◈◈◈ is the gate-source voltage (◈◈◈(◈◈◈)≤◈◈◈≤◈◈◈(◈◈◈)) ,

◈◈ is the drain current,

◈◈◈◈ is the maximum current,

◈◈◈(◈◈◈) is the turn-off voltage.

◈◈◈(◈◈◈) may be found on a data sheet. Above this voltage the insulating layer will be damaged and the device will cease to function properly. A typical value for this might be in the range of 20 to 30 volts. The trick is that given the very small gate current, even a simple electrostatic discharge can damage the device. It is very easy to develop hundreds of volts static on the human body. In fact, it is generally not noticeable until the potential reaches a few thousand volts (as in body hair standing up). The consequence of this is that simply picking up the device could destroy it.[1]

There are a couple different ways of dealing with this issue. The first way is to add back-to-back Zener diodes across the device during its manufacture. The problem with this is that the diode leakage current will be greater than the gate current and this degrades performance. The other technique is to prevent the charge from getting to the device in the first place. For example, the MOSFET can be shipped in conductive plastic (not to be confused with ordinary plastic or polystyrene foam). Some devices are shipped with a metal shorting that encompasses all of the leads. Also, during manufacture or prototyping, environmental controls are established to minimize the creation of static charges, optimal humidity being important as one example. Workers who handle devices may work on special conductive mats or wear wrist straps that are attached to ground. These items are only mildly conductive, that is, of high resistance, as it would not be safe to electrically ground a human working in an electrical lab. The devices are conductive enough to bleed off static charge but not so conductive as to present a shock hazard. Once installed on the circuit board, normal ESD precautions apply. As the device's characteristic equation has not changed, many of the items derived for the JFET still apply to the DE-MOSFET. This includes the transconductance equation plot found in Chapter 1.

As the transconductance equation is unchanged with the exception of an extended range for ◈◈◈ , the definition for ◈◈0 is also unchanged.

$$g_{m0} = -\frac{2I_{DSS}}{V_{GS(off)}}$$

(3.2.2)

$$g_m = g_{m0}\left(1 - \frac{V_{GS}}{V_{GS(off)}}\right)$$

(3.2.3)

It is worth noting that ◈◈0 no longer represents the maximum device transconductance because

1. Which brings to mind the old question of what to store a universal solvent in.

I_{D} no longer represents the maximum drain current as seen in Figure 3.2.3 . To illustrate it another way,

$$g_m = g_{m0}\sqrt{\frac{I_D}{I_{DSS}}}$$

(3.2.4)

It is very important to watch the sign of V_{GS} in Equation 3.2.3 . In enhancement mode, a positive V_{GS} will lead to a I_D greater than I_{D0} due to the double negative.

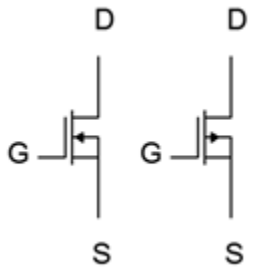


Figure 3.2.4: DE-MOSFET schematic symbols. N-channel (left) and P-channel (right).

The schematic symbols for the DE-MOSFET are shown in Figure 3.2.4 . As is the norm, the arrow points in the direction of N material, with the central vertical bar representing the channel. The arrow is attached to the substrate. In some devices this is brought out of the packaging as a fourth lead although in many it is simply tied back to the source terminal as shown here. Finally, note how the gate terminal is not drawn connected to the body of the device, emphasizing its isolated nature.

3.3 DE-MOSFET BIASING

As the characteristic equations of the JFET and DE-MOSFET are the same, the DC biasing model is the same. Consequently, the DE-MOSFET can be biased using any of the techniques used with the JFET including self bias, combination bias and current source bias as these are all second quadrant biasing schemes (i.e., have a negative ◇◇). The self bias and combination bias equations and plots from Chapter 10 may be used without modification. The DE-MOSFET also allows first quadrant operation so a couple of new biasing forms become available: zero bias and voltage divider bias. In reality, both are variations on constant voltage bias but which utilize the first quadrant.

ZERO BIAS

Zero bias is unique. In some ways it can be thought of as a cross between self bias and constant voltage bias. Like self bias, it does not require a second DC source for the gate or source terminal. Like constant voltage bias, there is no need for a source resistor, ◇. A prototype of zero bias is shown in Figure 3.3.1 . There is no question that it is a minimal parts-count circuit.

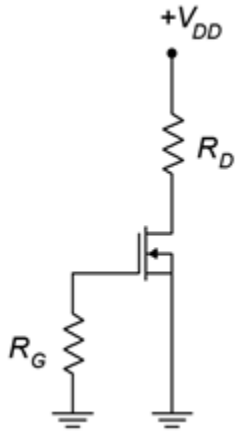


Figure 3.3.1: Zero bias prototype.

Zero bias is so named because it operates at ◇◇=0 V. Recall that the gate current is ideally zero, thus there is no drop across ◇◇ and ◇◇=0 V as a consequence. The source is tied directly to ground, therefore ◇◇ must equal 0 V. As ◇◇ doesn't change, this can be thought of as a form of constant voltage bias. The interesting bit is that when an AC signal is applied to the gate, its negative portion will pull the MOSFET down into depletion mode and the positive portion will push the operation into enhancement mode. Because the device can operate in this fashion, conducting current while straddling zero, so to speak, DE-MOSFETs are sometimes referred to as normally on devices.

Determining the operating point for zero bias is startlingly easy. Because ◇◇=0 V, ◇ must equal ◇◇ and ◇ must equal ◇0. Like all constant voltage biasing schemes, though, Q point stability is not very good. Another point to notice is that, as there is no source resistor, this bias is only applicable to non-swamped common source amplifiers. It cannot be used with a source follower or

swamped amplifier (if a small swamping resistor is inserted into the source, technically the circuit can be classified as self bias, although the AC signal may still push operation into enhancement mode).

Example 3.3.1

Determine $\diamondsuit_{\diamondsuit}$, $\diamondsuit_{\diamondsuit}$ and $\diamondsuit_{\diamondsuit 0}$ for the circuit shown in Figure 3.3.2 . Assume $\diamondsuit_{\diamondsuit\diamondsuit\diamondsuit}$=12 mA and $\diamondsuit_{\diamondsuit\diamondsuit(\diamondsuit\diamondsuit\diamondsuit)}$=−3 V.

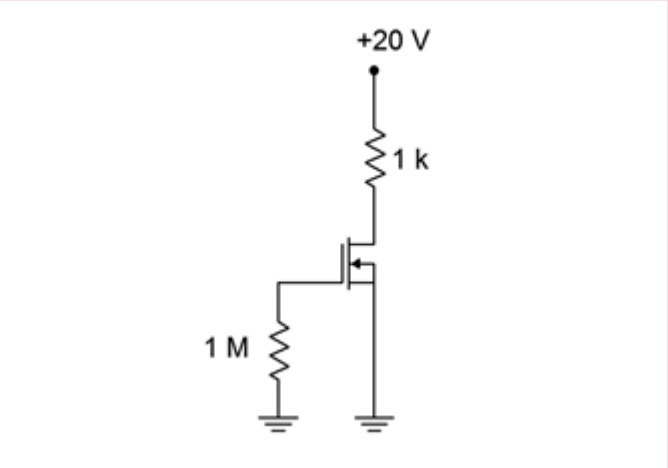


Figure 3.3.2: Circuit for Example 3.3.1 .

By inspection, as this is zero bias $\diamondsuit_{\diamondsuit}$=$\diamondsuit_{\diamondsuit\diamondsuit\diamondsuit}$, and therefore $\diamondsuit_{\diamondsuit}$=12 mA. Using KVL and Ohm's law we can find $\diamondsuit_{\diamondsuit}$.

$$V_D = V_{DD} - I_D R_D$$
$$V_D = 20V - 12mA \times 1k\Omega$$
$$V_D = 8V$$
$$g_{m0} = -\frac{2I_{DSS}}{V_{GS(off)}}$$
$$g_{m0} = -\frac{2 \times 12mA}{-3V}$$
$$g_{m0} = 8mS$$

VOLTAGE DIVIDER BIAS

Voltage divider bias is a form of constant voltage bias that operates in enhancement mode. A prototype circuit is shown in Figure 3.3.3 . Note that the source terminal is connected directly to ground. This is important. If this was not the case, this would be a form of combination bias (basically shifting the $\diamondsuit_{\diamondsuit\diamondsuit}$ supply up to ground and then shifting the gate voltage from ground up to a positive $\diamondsuit_{\diamondsuit\diamondsuit}$ to maintain the same differential voltage). As such, it would be operating in depletion mode.

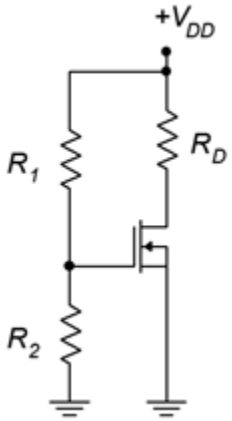


Figure 3.3.3: Voltage divider bias prototype.

The voltage divider comprised of R_1 and R_2 will establish a DC bias potential on the gate. As the source is at ground, $V_{GG}=V_G=V_{R_2}$. Given that V_{GG} must be positive, then V_{GS} must be positive, and enhancement mode operation is a given.

The most direct way to handle this is to determine the voltage divider potential and use either the characteristic equation (Equation 10.2.1) or associated graph to determine the drain current. Once V_G is found, the drain-source voltage may be found via the standard Ohm's law/KVL route.

Before continuing, note that the values of the divider resistors can be very high without creating biasing problems (unlike the BJT version of voltage divider bias). This is because the gate current is so small that even when using megohm values for the divider, the loading caused by the gate will not be noticeable.

Example 3.3.2

For the circuit of Figure 3.3.4 , determine V_G and V_D. Assume $I_D=2$ mA and $V_{GS}(V_T)=-6$ V.

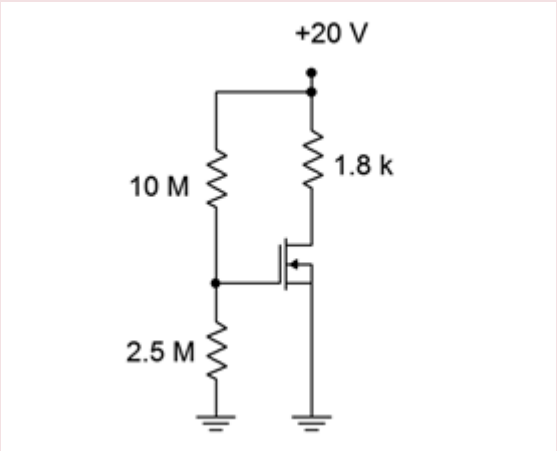


Figure 3.3.4: Circuit for Example 3.3.2 .

The voltage divider will yield V_{GS}.

$$V_{GS} = V_{DD}\left(\frac{R_2}{R_1 + R_2}\right)$$

$$V_{GS} = 20V\left(\frac{2.5M\Omega}{10M\Omega + 2.5M\Omega}\right)$$

$$V_{GS} = 4V$$

Use Equation 10.2.1 to find [illegible].

$$I_D = I_{DSS}\left(1 - \frac{V_{GS}}{V_{GS(off)}}\right)^2$$

$$I_D = 2mA\left(1\frac{4V}{-6V}\right)^2$$

$$I_D = 5.56mA$$

Use KVL and Ohm's law to find [illegible].

$$V_D = V_{DD} - I_D R_D$$
$$V_D = 20V - 5.56mA \times 1.8k\Omega$$
$$V_D = 9.99V$$

Alternately, using the curve of Figure 3.2.3, we would first find the normalized gate-source voltage which is 4 V/6 V or 0.667 (note that the curve plots –[illegible] so that the quadrants do not appear reversed). From this the normalized drain current, [illegible], may be determined to be approximately 2.8, yielding a drain current of 5.6 mA.

3.4 THE E-MOSFET

The E-MOSFET is available in both low power and high power versions. It operates in enhancement mode (first quadrant) only. The construction of the low power version is similar to that of the DE-MOSFET but with one important distinction. A simplified cross-section of an N-channel E-MOSFET is shown in Figure 3.4.1 .

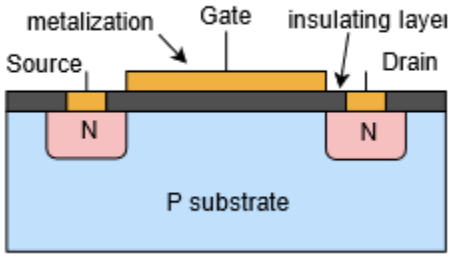


Figure 3.4.1: E-MOSFET internal structure.

In the E-MOSFET, the P material extends up through the channel and to the gate insulating layer. This has the effect of preventing current flow with negative gatesource voltages applied. Consequently, E-MOSFETs are sometimes referred to as normally off devices. In fact, the E-MOSFET will not conduct if ◈◈◈ is zero, or even for small positive values of ◈◈◈. The P and N materials functionally create energy hills or barriers that prevent current flow through the channel. This can be compared to an NPN BJT that has an open base terminal: no collector current would flow (unless the collector-emitter voltage exceeded the breakdown limit).

To understand how the E-MOSFET functions, refer to Figure 3.4.2 . This diagram shows the device with positive drain and gate supplies attached to it through limiting resistors. The dashed lines indicate electron current flow. As with the DE-MOSFET, the gate can be seen as one plate of a capacitor while the P material serves as the other plate. A positive voltage on the gate will lead to a negative charge on the P material side. If the charge is large enough, all of the holes in the P material can be filled leaving the portion of the material situated near the isolation layer neutral (neither P nor N). Any further increase in gate voltage injects more negative charge into this region, this making it behave like N material. This is called an N-type inversion layer and it allows a path for current to flow. The more positive we make the gate voltage, the greater the effect, and the greater the current.

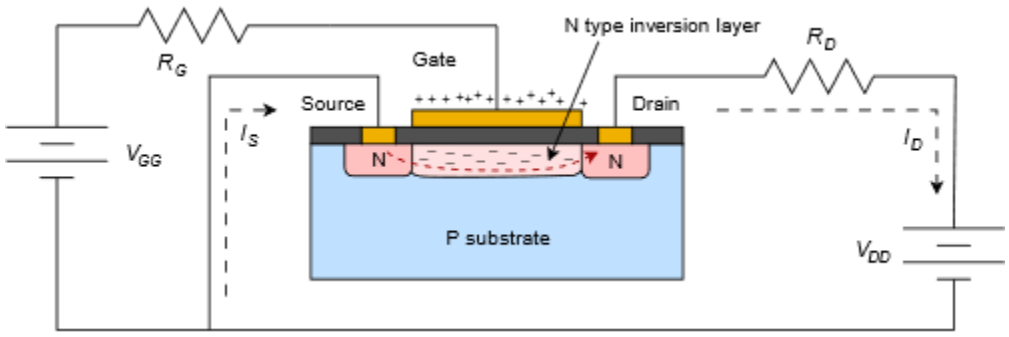


Figure 3.4.2: E-MOSFET bias with electron flow.

The voltage where the current begins to flow is called the threshold voltage and is usually denoted as $V_{GS(th)}$, although it is sometimes shortened to V_{th} or just V_t . Like both the JFET and DE-MOSFET, the E-MOSFET drain curve family exhibits three characteristic regions: the ohmic or triode region, the constant current or saturation region, and the breakdown region.

The characteristic equation for the E-MOSFET operating in its constant current region is given below. Like the other FETs examined, this is a square-law device.

$$I_D = k(V_{GS} - V_{GS(th)})^2$$

(3.4.1)

Where

I_D is the drain current,

V_{GS} is the gate-source voltage ($V_{GS(th)} \leq V_{GS} \leq V_{GS(max)}$),

$V_{GS(th)}$ is the the threshold voltage,

k is a device parameter (a constant, units of amps/volt 2 or siemens/volt).

The derivative of Equation 3.4.1 yields the transconductance.

$$g_m = \frac{dI_D}{dV_{GS}} = 2k(V_{GS} - V_{GS(th)})$$

(3.4.2)

Equation 3.4.1 is plotted in Figure 3.4.3 . The normalized gate-source voltage is $V_{GS}/V_{GS(th)}$ and the normalized drain current is the ratio of I_D to the current generated when V_{GS} is twice $V_{GS(th)}$. This curve is reminiscent of the characteristic curve of a BJT. First, they are both in the first quadrant. Second, both curves exhibit an increasing positive slope. Finally, the curves don't begin to "take off" until some specific turn-on voltage is reached. In the case of the BJT, that voltage is approximately 0.7 V for a silicon device. For the E-MOSFET, that voltage is $V_{GS(th)}$. Obviously though, the MOSFET curve does not increase as rapidly as the BJT curve.

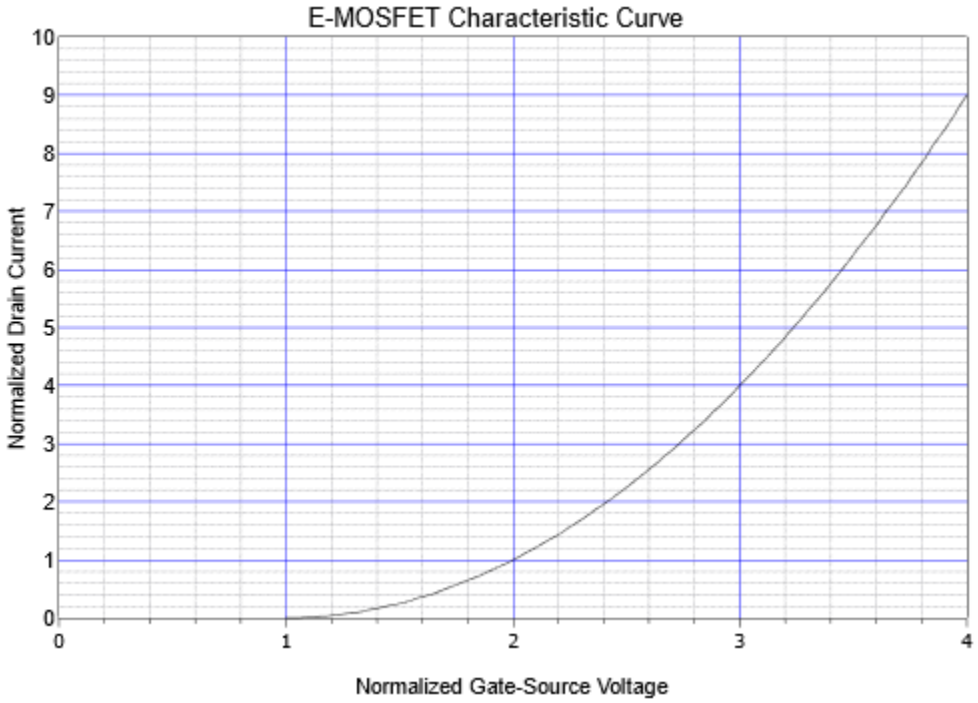


Figure 3.4.3: Characteristic curve for E-MOSFET.

The power E-MOSFET utilizes a different construction from low power MOSFETs and offers certain advantages over power BJTs including very fast switching speed and lower drive current demands. Consequently, they tend to be favored over BJTs in high power, high speed switching applications such as switching power supply regulators, DC-to-DC converters and class D amplifiers (Chapter 14). These devices also exhibit extremely low $R_{DS(on)}$ values, in some cases just a few milliohms. There are different methods of construction, the most recent being the trench style. A cutaway view is shown in Figure 3.4.4 .

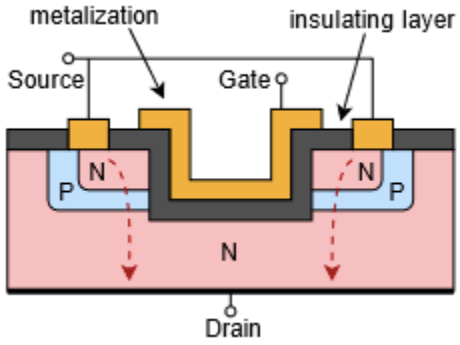


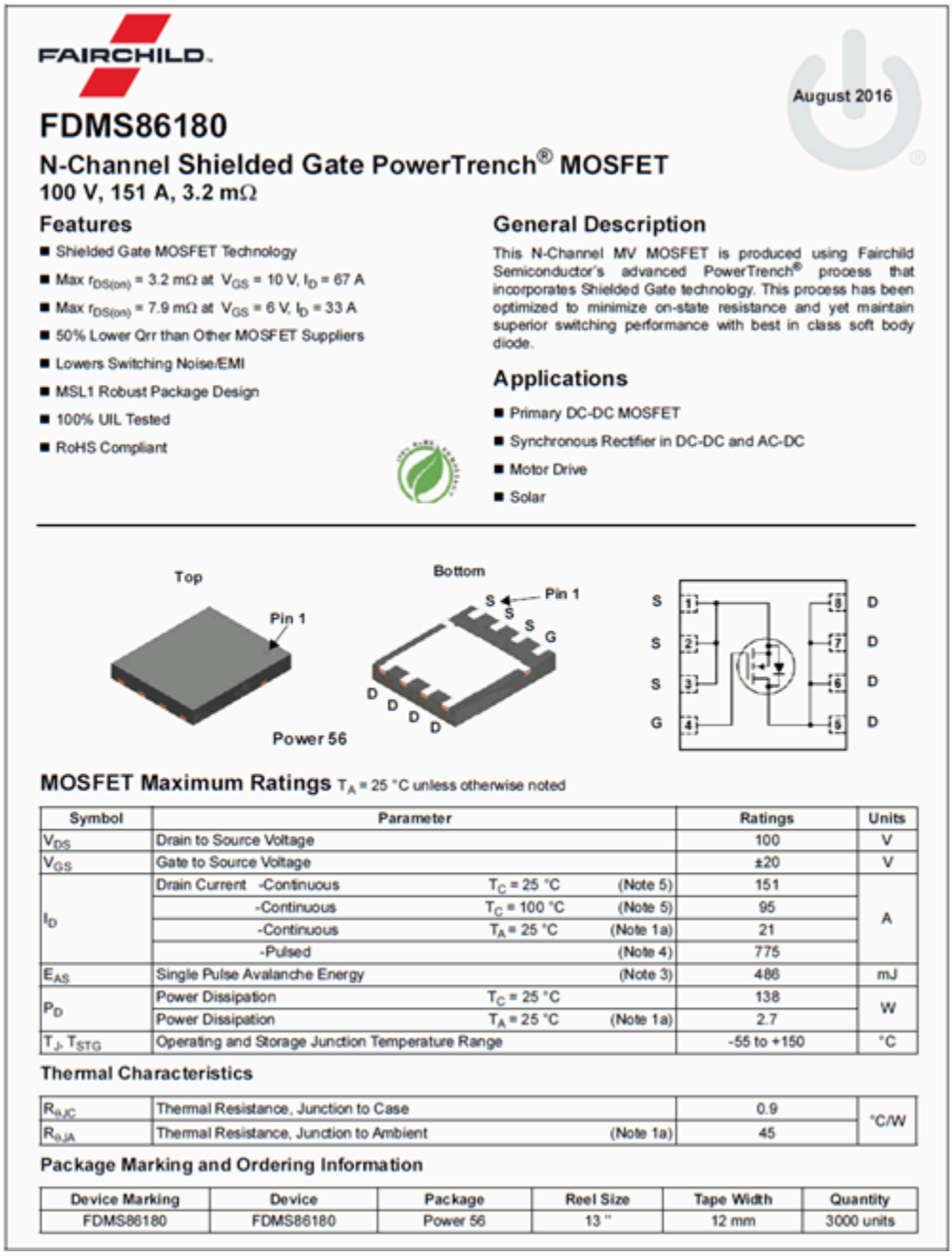
Figure 3.4.4: Construction of power Trench E-MOSFET.

A power E-MOSFET is made of a large number of cells, each featuring the U-shaped gate "trench" (an earlier style used a V-shaped trench). Note the location of the drain, now opposite of the gate and source. The advantage here is that the current flows vertically rather than laterally. This results in a much lower $R_{DS(on)}$ and considerably greater current capacity. The characteristic curve still echoes that of Figure 3.4.3 although it tends to be steeper when compared to low power devices.

3.5 E-MOSFET DATA SHEET INTERPRETATION

A data sheet for an E-MOSFET, the FDMS86180, is shown in Figure 3.5.1 . This is an N-channel, high power device using trench construction.

FDMS86180 data sheet



FAIRCHILD

August 2016

FDMS86180

N-Channel Shielded Gate PowerTrench® MOSFET

100 V, 151 A, 3.2 mΩ

Features

- Shielded Gate MOSFET Technology
- Max $r_{DS(on)}$ = 3.2 mΩ at V_{GS} = 10 V, I_D = 67 A
- Max $r_{DS(on)}$ = 7.9 mΩ at V_{GS} = 6 V, I_D = 33 A
- 50% Lower Qrr than Other MOSFET Suppliers
- Lowers Switching Noise/EMI
- MSL1 Robust Package Design
- 100% UIL Tested
- RoHS Compliant

General Description

This N-Channel MV MOSFET is produced using Fairchild Semiconductor's advanced PowerTrench® process that incorporates Shielded Gate technology. This process has been optimized to minimize on-state resistance and yet maintain superior switching performance with best in class soft body diode.

Applications

- Primary DC-DC MOSFET
- Synchronous Rectifier in DC-DC and AC-DC
- Motor Drive
- Solar

MOSFET Maximum Ratings T_A = 25 °C unless otherwise noted

Symbol	Parameter			Ratings	Units
V_{DS}	Drain to Source Voltage			100	V
V_{GS}	Gate to Source Voltage			±20	V
I_D	Drain Current -Continuous	T_C = 25 °C	(Note 5)	151	A
	-Continuous	T_C = 100 °C	(Note 5)	95	
	-Continuous	T_A = 25 °C	(Note 1a)	21	
	-Pulsed		(Note 4)	775	
E_{AS}	Single Pulse Avalanche Energy		(Note 3)	486	mJ
P_D	Power Dissipation	T_C = 25 °C		138	W
	Power Dissipation	T_A = 25 °C	(Note 1a)	2.7	
T_J, T_{STG}	Operating and Storage Junction Temperature Range			-55 to +150	°C

Thermal Characteristics

$R_{\theta JC}$	Thermal Resistance, Junction to Case	0.9	°C/W
$R_{\theta JA}$	Thermal Resistance, Junction to Ambient (Note 1a)	45	

Package Marking and Ordering Information

Device Marking	Device	Package	Reel Size	Tape Width	Quantity
FDMS86180	FDMS86180	Power 56	13 "	12 mm	3000 units

Figure 3.5.1 ⓒ: FDMS86180 data sheet. Used with permission from SCILLC dba ON Semiconductor.

One of the first things that might jump out is the "100% RoHS Compliant" green leaf logo in the upper

center, meaning that the device meets the Restriction of Hazardous Substances directive. The device comes in the flat, multi-pin Power 56 package and features an ◈◈◈(◈◈) of just a few milliohms. Continuous current capability at room temperature is 151 amps with a pulsed current maximum of 775 amps. In Figure 3.5.1◈ we find a breakdown voltage of 100 volts and an ◈◈◈◈ of only 1 ◈ A. Recall that this is a normally off device, and thus ◈◈◈◈ represents a leakage current. Continuing, ◈◈◈(◈*h*) varies between 2.0 and 4.0 volts, with 3.2 volts being typical. The forward transconductance, ◈◈ (here referred to as ◈◈◈) is 144 siemens at a drain current of 67 amps. This is orders of magnitude greater than what we might see with small signal devices. Turn-on and turn-off times are measured in the tens of nanoseconds, verifying the high speed switching ability of the device.

Electrical Characteristics T_J = 25 °C unless otherwise noted.

Symbol	Parameter	Test Conditions	Min.	Typ.	Max.	Units
Off Characteristics						
BV_{DSS}	Drain to Source Breakdown Voltage	I_D = 250 µA, V_{GS} = 0 V	100			V
$\frac{\Delta BV_{DSS}}{\Delta T_J}$	Breakdown Voltage Temperature Coefficient	I_D = 250 µA, referenced to 25 °C		73		mV/°C
I_{DSS}	Zero Gate Voltage Drain Current	V_{DS} = 80 V, V_{GS} = 0 V			1	µA
I_{GSS}	Gate to Source Leakage Current	V_{GS} = ±20 V, V_{DS} = 0 V			100	nA
On Characteristics						
$V_{GS(th)}$	Gate to Source Threshold Voltage	V_{GS} = V_{DS}, I_D = 370 µA	2.0	3.2	4.0	V
$\frac{\Delta V_{GS(th)}}{\Delta T_J}$	Gate to Source Threshold Voltage Temperature Coefficient	I_D = 370 µA, referenced to 25 °C		-8		mV/°C
$r_{DS(on)}$	Static Drain to Source On Resistance	V_{GS} = 10 V, I_D = 67 A		2.4	3.2	mΩ
		V_{GS} = 6 V, I_D = 33 A		3.8	7.9	
		V_{GS} = 10 V, I_D = 67 A, T_J = 125 °C		4.0	5.4	
g_{FS}	Forward Transconductance	V_{DS} = 5 V, I_D = 67 A		144		S
Dynamic Characteristics						
C_{iss}	Input Capacitance	V_{DS} = 50 V, V_{GS} = 0 V, f = 1 MHz		4439	6215	pF
C_{oss}	Output Capacitance			2663	3730	pF
C_{rss}	Reverse Transfer Capacitance			24	55	pF
R_g	Gate Resistance		0.1	0.8	1.6	Ω
Switching Characteristics						
$t_{d(on)}$	Turn-On Delay Time	V_{DD} = 50 V, I_D = 67 A, V_{GS} = 10 V, R_{GEN} = 6 Ω		24	39	ns
t_r	Rise Time			12	22	ns
$t_{d(off)}$	Turn-Off Delay Time			30	48	ns
t_f	Fall Time			7	14	ns
Q_g	Total Gate Charge	V_{GS} = 0 V to 10 V; V_{DD} = 50 V, I_D = 67 A		60	84	nC
Q_g	Total Gate Charge	V_{GS} = 0 V to 6 V; V_{DD} = 50 V, I_D = 67 A		38	54	nC
Q_{gs}	Gate to Source Charge	V_{DD} = 50 V, I_D = 67 A		20		nC
Q_{gd}	Gate to Drain "Miller" Charge	V_{DD} = 50 V, I_D = 67 A		12		nC
Q_{oss}	Output Charge	V_{DD} = 50 V, V_{GS} = 0 V		175		nC
Drain-Source Diode Characteristics						
V_{SD}	Source to Drain Diode Forward Voltage	V_{GS} = 0 V, I_S = 2.1 A (Note 2)		0.7	1.2	V
		V_{GS} = 0 V, I_S = 67 A (Note 2)		0.8	1.3	
t_{rr}	Reverse Recovery Time	I_F = 33 A, di/dt = 300 A/µs		44	71	ns
Q_{rr}	Reverse Recovery Charge			109	207	nC
t_{rr}	Reverse Recovery Time	I_F = 33 A, di/dt = 1000 A/µs		33	53	ns
Q_{rr}	Reverse Recovery Charge			235	376	nC

Figure 3.5.1◈: FDMS86180 data sheet (cont).

A series of performance graphs are found in Figure 3.5.1◈ . In the upper left is a section of drain curves showing the ohmic region through ◈◈◈=5 V. The plot directly below this shows the increase in ◈◈◈(◈◈) as temperature rises. There is about a three-fold variation across the temperature range. At lower left is the characteristic curve variation. Note that the curves are less steep as temperature increases, showing a decrease in ◈◈ and thus, verifying a negative temperature coefficient of transconductance.

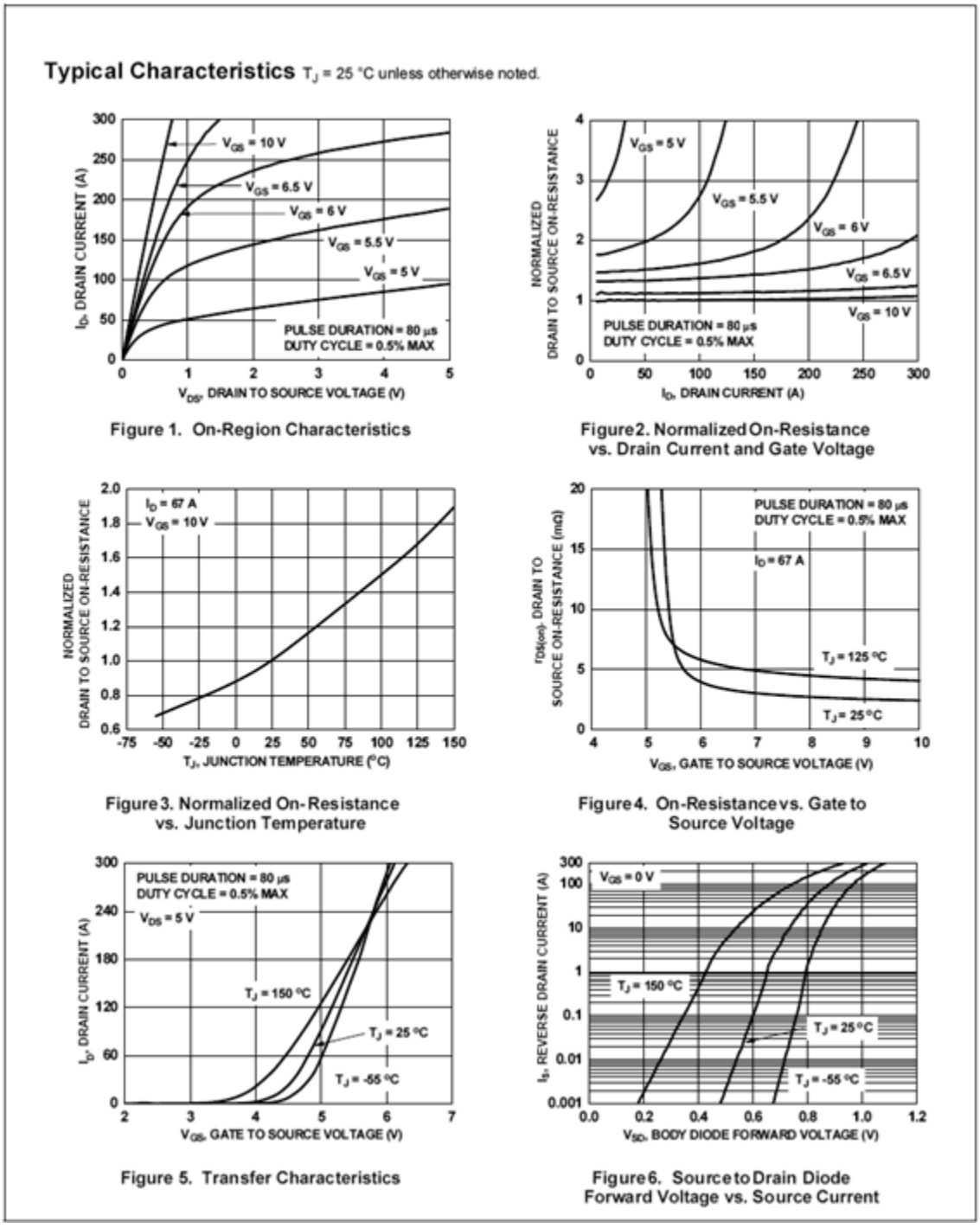


Figure 3.5.1◈: FDMS86180 data sheet (cont).

3.6 E-MOSFET BIASING

As the E-MOSFET operates only in the first quadrant, none of the biasing schemes used with JFETs will work with it. First, it should be noted that for large signal switching applications biasing is not much of an issue as we simply need to confirm that there is sufficient drive signal to turn the device on. For linear amplifiers we can use variations on constant voltage bias such as voltage divider bias, or on drain feedback bias.

VOLTAGE DIVIDER BIAS

Voltage divider bias is reminiscent of the divider circuit used with BJTs. Indeed, the N-channel E-MOSFET requires that its gate be higher than its source, just as the NPN BJT requires a base voltage higher than its emitter. The major differences between the two are that the E-MOSFET's input gate current is negligible compared to base current and that the gate-source voltage will be most likely higher than the 0.7 volt drop seen across the base-emitter junction. Also, the gate-source voltage will not be locked to a specific voltage but will vary depending on the remainder of the circuit.

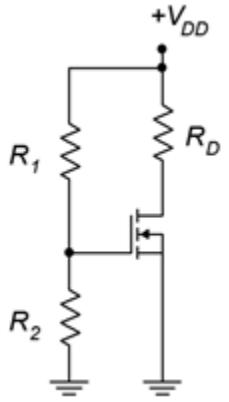


Figure 3.6.1: Voltage divider bias for E-MOSFET.

The prototype for the voltage divider bias is shown in Figure 3.6.1 . In general, the layout it is the same as the voltage divider bias used with the DE-MOSFET. The resistors R_1 and R_2 set up the divider to establish the gate voltage. As the source terminal is tied directly to ground, this means that $V_G = V_{R_2}$. The potential across R_2 needs to be set above $V_{GS(th)}$ for proper operation in accordance with Equation 3.4.1. Knowing the value of V_G, either the characteristic equation or the corresponding normalized drain current plot can be used to determine the drain current. The only factor missing is the device constant, K. This can be computed for any particular device based on the $I_{D(ON)}$, $V_{GS(ON)}$ coordinate pair specified in the data sheet (or measured in lab). An example is shown in Figure 3.6.2 .

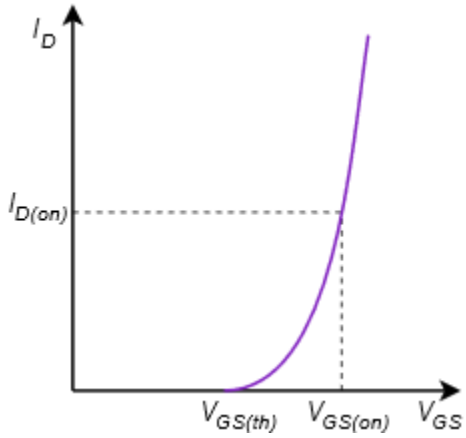


Figure 3.6.2 : Coordinate pair on E-MOSFET curve.

The constant ◈ is found via a rearrangement of Equation 3.4.1:

$$k = \frac{I_{D(on)}}{(V_{GS(on)} - V_{GS(th)})^2}$$

This value can then be used for other biasing points.

Example 3.6.1

For the circuit and matching device curve of Figure 3.6.3 , find ◈◈ and ◈◈◈ .

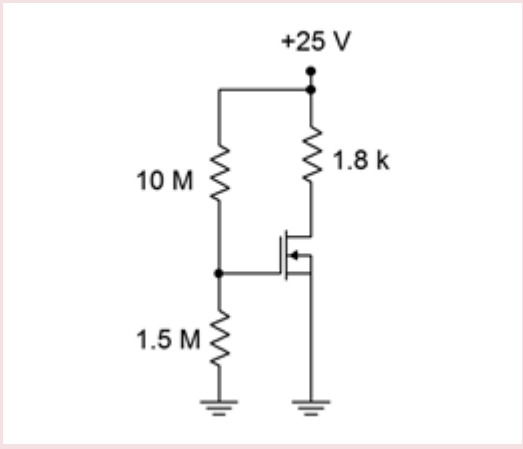


Figure 3.6.3◈: Circuit for Example 3.6.1 .

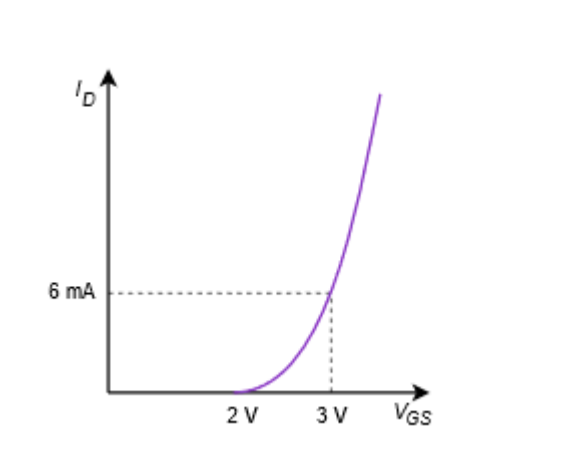


Figure 3.6.3 : Device curve for Example 3.6.1 .

First find the value of :

$$k = \frac{I_{D(on)}}{(V_{GS(on)} - V_{GS(th)})^2}$$

$$k = \frac{6mA}{(3V - 2V)^2}$$

$$k = 6mA/V^2$$

Now determine the gate voltage:

$$V_G = V_{DD}\frac{R_2}{R_1 + R_2}$$

$$V_G = 25V\frac{1.5M\Omega}{10M\Omega + 1.5M\Omega}$$

$$V_G = 3.26V$$

The source is grounded so .

$$I_D = k(V_{GS} - V_{GS(th)})^2$$

$$I_D = 6mA/V^2(3.26V - 2V)^2$$

$$I_D = 9.54mA$$

$$V_{DS} = V_{DD} - I_D R_D$$

$$V_{DS} = 25V - 9.54mA \times 1.8k\Omega$$

$$V_{DS} = 7.83V$$

In closing, note that it is possible to decouple the voltage divider using the same method employed

with BJTs in Figure 7.3.11. Very large value resistors are available in only a limited variety of sizes so this technique has an added benefit. The divider resistors can use more convenient sizes because $\diamond_1$ and $\diamond_2$ will not set the input impedance; it will be set by the decoupling resistor.

DRAIN FEEDBACK BIAS

Drain feedback bias utilizes the aforementioned "on" operating point from the characteristic curve. The idea is to establish a drain current via an appropriate selection of the drain resistor and power supply. The prototype of the drain feedback circuit is shown in Figure 3.6.4 .

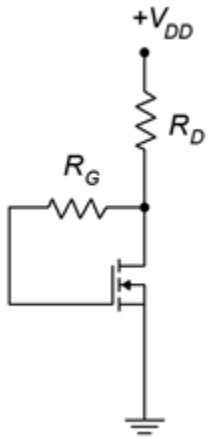


Figure 3.6.4: Drain feedback bias prototype.

This is relatively simple layout using few components. The key to understanding its operation is the KVL summation:

$$V_{DD} = V_{R_D} + V_{R_G} + V_{GS}$$
$$V_{DD} = I_D R_D + I_G R_G + V_{GS}$$

Gate current is negligible which means that

$$V_{DD} = I_D R_D + V_{GS}$$

and also

$$V_{DS} = V_{GS}$$

Therefore,

$$V_{GS} = V_{DS} = V_{DD} - I_D R_D$$

(3.6.1)

Equation 3.6.1 can be used as the basis for the design of the bias circuit.

Example 3.6.2

Utilizing the prototype of Figure 3.6.4 , determine values for $\diamond_{\diamond}$ and $\diamond_{\diamond}$ such that the drain current is 8 mA. Assume $\diamond_{\diamond\diamond}$=20 V, $\diamond_{\diamond(\diamond\diamond)}$=5 mA at $\diamond_{\diamond\diamond(\diamond\diamond)}$=4 V, and $\diamond_{\diamond\diamond(\diamond h)}$=2.5 V.

First find the value of $\diamond$:

$$k = \frac{I_{D(on)}}{(V_{GS(on)} - V_{GS(th)})^2}$$

$$k = \frac{5mA}{(4V - 2.5V)^2}$$

$$k = 2.22mA/V^2$$

Now determine the required ◈◈◈ to obtain 8 mA of drain current by rearranging Equation 3.4.1.

$$I_D = k(V_{GS} - V_{GS(th)})^2$$

$$V_{GS} = V_{GS(th)} + \sqrt{\frac{I_D}{k}}$$

$$V_{GS} = 2.5V + \sqrt{\frac{8mA}{2.22mA/V^2}}$$

$$V_{GS} = 4.4V$$

And finally,

$$V_{GS} = V_{DS} = V_{DD} - I_D R_D$$

$$R_D = \frac{V_{DD} - V_{GS}}{I_D}$$

$$R_D = \frac{20V - 4.4V}{8mA}$$

$$R_D = 1.95k\Omega$$

3.7 SUMMARY

There are two types of MOSFETs: the depletion-enhancement or DE-MOSFET and the enhancement-only or E-MOSFET. Both devices are constructed using an insulated gate instead of a PN junction and both devices exhibit a square-law characteristic curve. Like the JFET, MOSFETs are modeled as voltage-controlled current sources. Both devices show very, very small gate currents due to the insulated gate. They are static sensitive and precautions must be taken when handling them to avoid damage from ESD.

The DE-MOSFET exhibits the same characteristic curve as the JFET, however, the curve extends into the first quadrant (enhancement mode). Consequently, I_{DSS} is no longer the largest drain current possible, but rather, represents a middle ground. The DE-MOSFET can utilize all of the bias prototypes that are used with JFETs, including self bias, constant current bias and combination bias. Due to its dual quadrant capability, other biasing types are also possible including zero bias and voltage divider, both of which are variations on constant voltage bias.

The E-MOSFET operates in the first quadrant only (enhancement mode). Compared to the DE-MOSFET, its characteristic curve is shifted positive such that $V_{GS(off)}$ is now $V_{GS(th)}$, and $I_{D(off)}$ signifies the off-state leakage current. E-MOSFETs are available in both low power and high power variants. The high power versions utilize an alternate internal structure that allows drain current to flow vertically rather than horizontally. This results in very high current carrying ability and very low values for $R_{DS(on)}$.

Review Questions

1. What are the differences between JFETs and MOSFETs?
2. What are the differences between DE-MOSFETs and E-MOSFETs?
3. Why are MOSFETs sometimes referred to as "Insulated Gate" or IGFETs?
4. How does the DC bias model of the MOSFET compare to that of the BJT?
5. What biasing circuits are available for use with the DE-MOSFET?
6. What biasing circuits are available for use with the E-MOSFET?
7. What is "trench" construction and where is it used?
8. Explain typical precautions taken when handling MOSFETs and why they are necessary.

3.8 EXERCISES

ANALYSIS PROBLEMS

1. For the circuit of Figure 3.8.1 , determine □□, □□ and □□. □□□□ = 20 mA, □□□(□□□) = –6 V, □□□ = 15 V, □□ = 470 k Ω , □□ = 1.2 k Ω , □□ = 1.8 k Ω .
2. For the circuit of Figure 3.8.1 , determine □□, □□□ and □□. □□□□ = 20 mA, □□□(□□□) = –5 V, □□□ = 30 V, □□ = 560 k Ω , □□ = 420 Ω , □□ = 1.5 k Ω .

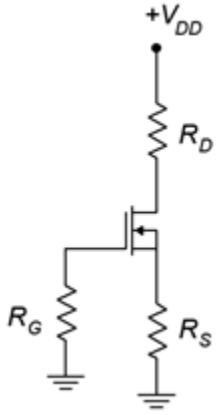


Figure 3.8.1

3. For Figure 3.8.2 , determine □□ , □□ and □□ . □□□□ = 15 mA, □□□ = 25 V, □□□(□□□) = –3 V, □□□ = –6 V, □□ = 820 k Ω , □□ = 2 k Ω , □□ = 3.6 k Ω .
4. For the circuit of Figure 3.8.2 , determine □□ , □□□ and □□ . □□□□ = 18 mA, □□□(□□□) = –3 V, □□□ = 30 V, □□□ = –9 V, □□ = 910 k Ω , □□ = 1.2 k Ω , □□ = 2.7 k Ω .
5. For the circuit of Figure 3.8.3 , determine □□ , □□ and □□ . □□□□ = 12 mA, □□□(□□□) = –4 V, □□□ = 35 V, □□ = 680 k Ω , □□ = 1.8 k Ω .

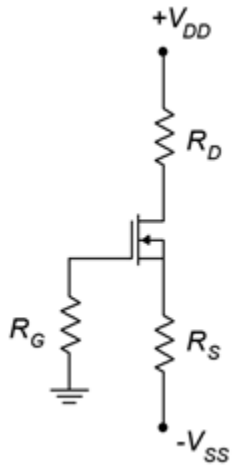


Figure 3.8.2

6. For the circuit of Figure 3.8.3 , determine ◇ , ◇◇ and ◇ . ◇◇◇ = 8 mA, ◇◇(◇◇) = –2 V, ◇◇ = 30 V, ◇ = 750 k Ω , ◇ = 2.7 k Ω .

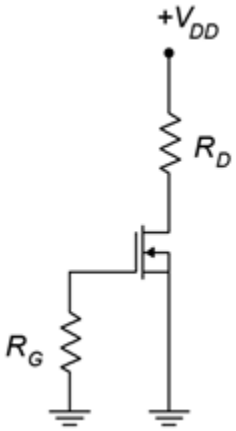


Figure 3.8.3

7. For the circuit of Figure 3.8.4 , determine ◇ , ◇ and ◇ . ◇◇◇ = 8 mA, ◇◇(◇◇) = –4 V, ◇◇ = 30 V, ◇1 = 2.7 M Ω , ◇2 = 110 k Ω , ◇ = 470 Ω .
8. For the circuit of Figure 3.8.4 , determine ◇ , ◇◇ and ◇ . ◇◇◇ = 12 mA, ◇◇(◇◇) = –6 V, ◇◇ = 20 V, ◇1 = 2 M Ω , ◇2 = 100 k Ω , ◇ = 680 Ω .
9. For the circuit of Figure 3.8.5 , determine ◇ , ◇ and ◇ . ◇(◇◇) = 8 mA, ◇◇(◇◇) = 5 V, ◇◇(◇*h*) = 3 V, ◇◇ = 30 V, ◇1 = 2 M Ω , ◇2 = 330 k Ω , ◇ = 1.2 k Ω .
10. For the circuit of Figure 3.8.5 , determine ◇ , ◇◇ and ◇ . ◇(◇◇) = 12 mA, ◇◇(◇◇) = 6 V, ◇◇(◇*h*) = 2.5 V, ◇◇ = 25 V, ◇1 = 1.5 M Ω , ◇2 = 470 k Ω , ◇ = 680 Ω .

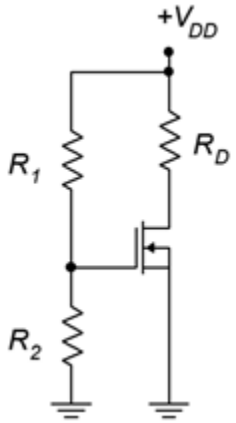


Figure 3.8.4

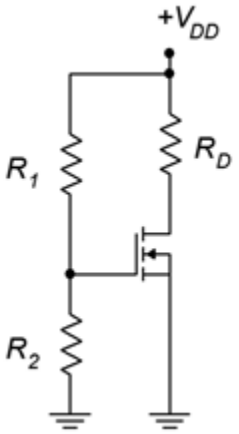


Figure 3.8.5

11. For the circuit of Figure 3.8.6 , determine ⍰⍰ , ⍰⍰ and ⍰⍰ . ⍰⍰⍰⍰ = 12 mA, ⍰⍰⍰(⍰⍰⍰) = 2 V, ⍰⍰⍰ = –25 V, ⍰⍰ = 470 k Ω , ⍰⍰ = 800 Ω , ⍰⍰ = 1.8 k Ω .
12. For the circuit of Figure 3.8.6 , determine ⍰⍰ and ⍰⍰ . ⍰⍰⍰⍰ = 10 mA, ⍰⍰⍰(⍰⍰⍰) = 2 V, ⍰⍰⍰ = –20 V, ⍰⍰ = 560 k Ω , ⍰⍰ = 680 Ω , ⍰⍰ = 1.5 k Ω .

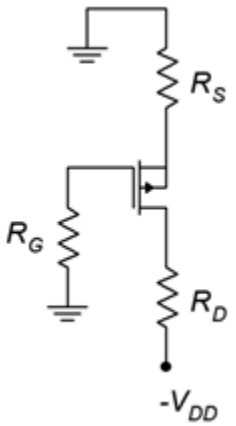


Figure 3.8.6

13. For the circuit of Figure 3.8.7 , determine ⍰⍰, ⍰⍰ and ⍰⍰ . ⍰⍰⍰⍰ = 14 mA, ⍰⍰⍰(⍰⍰⍰) = 3 V,

V_{GS} = –25 V, $V_{GS(off)}$ = 6 V, R_G = 780 kΩ, R_S = 2 kΩ, R_D = 3.3 kΩ.

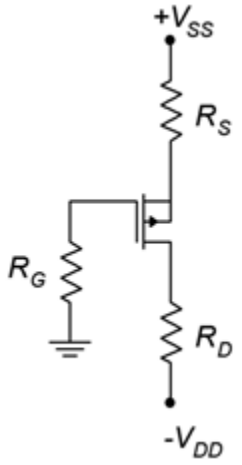


Figure 3.8.7

14. For the circuit of Figure 3.8.7, determine R_G and R_D. $I_{D(on)}$ = 16 mA, $V_{GS(th)}$ = 3.5 V, V_{GS} = –20 V, $V_{GS(off)}$ = 7 V, R_G = 1 MΩ, R_S = 1.5 kΩ, R_D = 2.2 kΩ.
15. For the circuit of Figure 3.8.8, determine R_G and R_D. $I_{D(on)}$ = 11 mA, $V_{GS(th)}$ = 2 V, V_{GS} = –24 V, R_S = 750 kΩ, R_D = 1.2 kΩ.
16. For the circuit of Figure 3.8.8, determine R_G and R_D. $I_{D(on)}$ = 9 mA, $V_{GS(th)}$ = 3 V, V_{GS} = –18 V, R_G = 430 kΩ, R_D = 910 Ω.

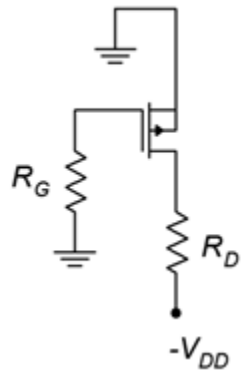


Figure 3.8.8

DESIGN PROBLEMS

17. Using the circuit of Figure 3.8.1, determine a value for R_G to set I_D to 4 mA. $I_{D(on)}$ = 10 mA, $V_{GS(th)}$ = –2 V, V_{GS} = 18 V, R_S = 470 kΩ, R_D = 1.5 kΩ.

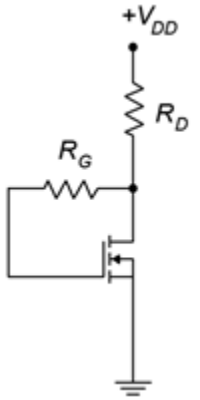


Figure 3.8.9

18. For the circuit of Figure 3.8.9 , determine ◇◇ and ◇◇ to set ◇◇ = 10 mA. ◇◇(◇◇) = 15 mA, ◇◇◇(◇◇) = 6 V, ◇◇◇(◇h) = 2 V, ◇◇◇ = 20 V.
19. For the circuit of Figure 3.8.9 , determine ◇◇ and ◇◇ to set ◇◇ = 15 mA. ◇◇(◇◇) = 10 mA, ◇◇◇(◇◇) = 5 V, ◇◇◇(◇h) = 2 V, ◇◇◇ = 25 V.

CHALLENGE PROBLEMS

20. Using the circuit of Figure 3.8.2 , determine values for ◇◇ , ◇◇ and ◇◇◇ to set ◇◇ to 5 mA and ◇◇ to 20 V. ◇◇◇◇ = 15 mA, ◇◇◇(◇◇◇) = –3 V, ◇◇◇ = 30 V, ◇◇ = 560 k Ω .
21. Using the circuit of Figure 3.8.10 , determine values for ◇◇ to set ◇◇ to 15 V. ◇◇◇◇ = 10 mA, ◇◇◇(◇◇◇) = 3 V, ◇◇◇ = 25 V, ◇◇ = 680 k Ω .

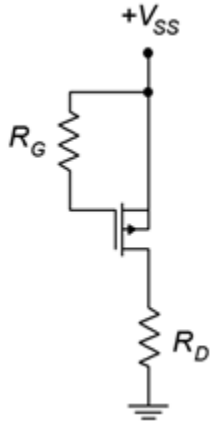


Figure 3.8.10

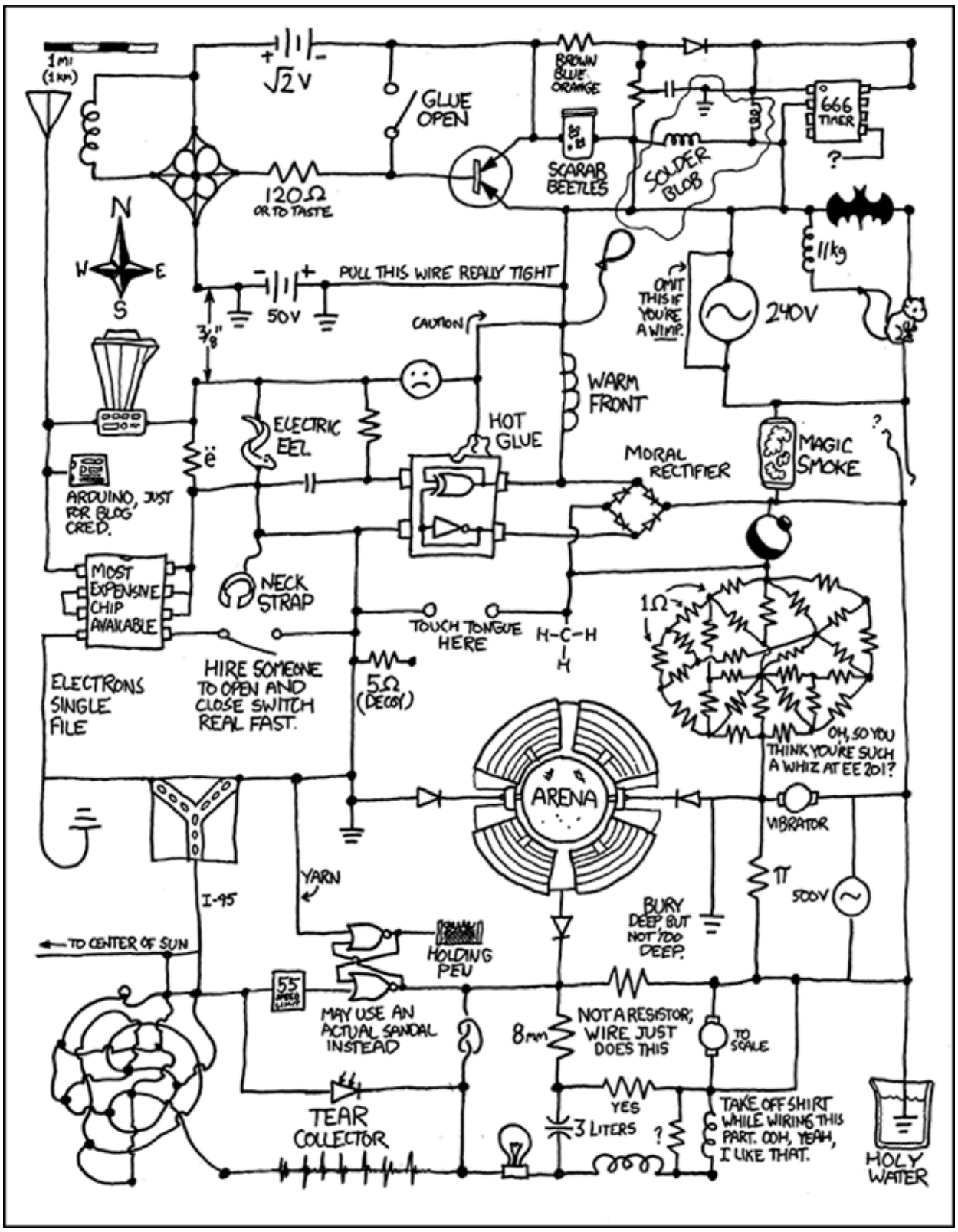


Figure 3.8.11 : Comic from xkcd.com

UNIT 4: MOSFET SMALL SIGNAL AMPLIFIERS

Learning Objectives

After completing this chapter, you should be able to:

- Draw and explain a basic AC model for a MOSFET.
- Analyze basic MOSFET amplifiers for voltage gain, input impedance and output impedance.
- Analyze basic MOSFET voltage followers for voltage gain, input impedance and output impedance.

4.1 INTRODUCTION

MOSFETs can be used to create both common source voltage amplifiers and common drain voltage followers (i.e., source followers). Both circuits offer the potential for very high input impedance due to the extremely low gate current MOSFETs provide. As with JFET amplifiers, at higher frequencies input capacitance dominates and reduces the input impedance. Not all bias prototypes lend themselves to all possible AC circuits. For example, zero bias for a DE-MOSFET is not suitable for followers or swamped amplifiers as it lacks a source resistor. The same is true for voltage divider biasing used with both DE- and E-MOSFETs. These biasing schemes are suitable for non-swamped amplifiers, though.

In general, MOSFET amplifiers tend to have good high frequency performance, offer low noise and exhibit low distortion with modestly sized input signals. Compared to BJTs, their voltage gain magnitude is lower.

A key parameter in determining gain is the device's transconductance, ⍰⍰ . Transconductance varies widely depending on the kind of MOSFET used. A small signal DE-MOSFET may exhibit a transconductance of just a few millisiemens. In contrast, a high power E-MOSFET may exhibit a transconductance of over 100 siemens.

4.2 MOSFET COMMON SOURCE AMPLIFIERS

Before we can examine the common source amplifier, an AC model is needed for both the DE- and E-MOSFET. A simplified model consists of a voltage-controlled current source and an input resistance, ◈◈◈. This model is shown in Figure 4.2.1 . The model is essentially the same as that used for the JFET. Technically, the gate-source resistance is higher in the MOSFET due to the insulated gate, and this is useful in specific applications such as in the design of electrometers, but for general purpose work it is a minor distinction. The impedance associated with the current source is not shown as it is typically large enough to ignore. Similarly, the device capacitances are not shown. It is worth noting that the capacitances associated with small signal devices might be just a few picofarads, however, a power device might exhibit values of a few nanofarads.

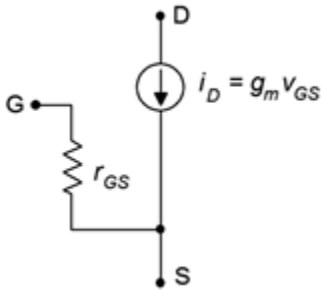


Figure 4.2.1: AC device model for MOSFETs.

As the device model is the same for both DE- and E-MOSFETs, the analysis of voltage gain, input impedance and output impedance will apply to both devices. The only practical differences will be how the transconductance is determined, and circuit variations due to the differing biasing requirements which will effect the input impedance. In fact, there will be a great uniformity between JFET-based circuits and DE-MOSFET circuits operating in depletion mode.

An AC equivalent of a swamped common source amplifier is shown in Figure 4.2.2 . This is a generic prototype and is suitable for any variation on device and bias type. Ultimately, all of the amplifiers can be reduced down to this equivalent, occasionally with some resistance values left out (either opened or shorted). For example, if the amplifier is not swamped then ◈◈=0 . Similarly, ◈◈ might correspond to a single gate biasing resistor or it might represent the equivalent of a pair of resistors that set up a gate voltage divider.

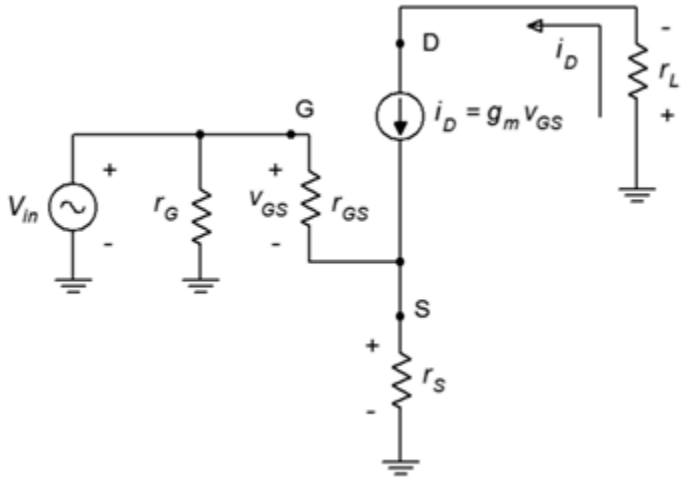


Figure 4.2.2: Generic common source amplifier equivalent.

VOLTAGE GAIN

In order to derive an equation for the voltage gain, we start with its definition, namely that voltage gain is the ratio of v_{out} to v_{in}. We then proceed by expressing these voltages in terms of their Ohm's law equivalents. Note that v_{in} can also be called v_G.

$$A_v = \frac{v_{out}}{v_{in}} = \frac{v_L}{v_G} = \frac{v_D}{v_G}$$

$$A_v = \frac{-i_D r_L}{i_D r_S + v_{GS}}$$

$$A_v = \frac{-g_m v_{GS} r_L}{g_m v_{GS} r_S + v_{GS}}$$

$$A_v = -\frac{g_m r_L}{g_m r_S + 1}$$

(4.2.1)

or, if preferred

$$A_v = -\frac{g_m r_D}{g_m r_S + 1}$$

(4.2.2)

This is the general equation for voltage gain. If the amplifier is not swamped then the first portion of the denominator drops out and the gain simplifies to

$$A_v = -g_m r_L$$

(4.2.3)

or alternately

$$A_v = -g_m r_D$$

(4.2.4)

The swamping resistor, r_S, plays the same role here as it did with both the BJT and JFET. Swamping helps to stabilize the gain and reduce distortion, but at the expense of voltage gain.

INPUT IMPEDANCE

Referring back to Figure 4.2.2 , the input impedance of the amplifier will be ⧫ in parallel with the impedance looking into the gate terminal, ⧫(⧫). At a minimum this will be ⧫ (it is somewhat higher when swamped but this can be ignored in most cases). At low frequencies ⧫ is very large, perhaps as high as 1012 ohms. In most practical circuits, ⧫ will be much lower, hence

$$Z_{in} = r_G || r_{GS} \approx r_G$$

(4.2.5)

It is important to reiterate that ⧫ is the equivalent resistance seen prior to the gate terminal that is seen from the vantage point of ⧫. In the case of self bias, combination bias, zero bias and constant current bias, this will be the single biasing resistor ⧫. For simple voltage divider biasing, ⧫ will be the parallel combination of the two divider resistors (i.e., ⧫1||⧫2). For decoupled voltage divider biasing, as shown in Figure 4.2.3 , ⧫ will be the decoupling resistor (i.e., ⧫3) that is connected between the divider and the gate. This is because the divider node is bypassed to ground via a capacitor. Finally, for drain feedback biasing, ⧫ is the Millerized ⧫ that bridges the drain and gate.

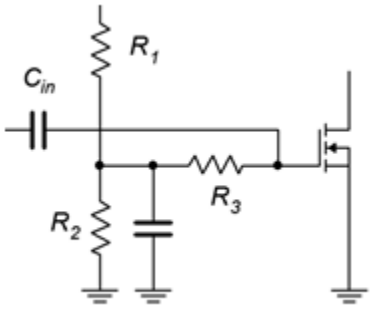


Figure 4.2.3: Decoupled voltage divider.

OUTPUT IMPEDANCE

The derivation of output impedance is unchanged from the JFET case. From the perspective of the load, the output impedance will be the drain biasing resistor, ⧫, in parallel with the internal impedance of the current source within the device model. ⧫ tends to be much lower than this, and thus, the output impedance can be approximated as ⧫.

Therefore we may state

$$Z_{out} = r_{model} || R_D \approx R_D$$

(4.2.6)

At this point, a variety of examples are in order to illustrate some of the myriad combinations.

Example 4.2.1

For the amplifier in Figure 4.2.4 , determine the input impedance and load voltage. ⧫ = 20 mV, ⧫ =

20 V, $\diamondsuit_{\diamondsuit}$ = 1 M Ω , $\diamondsuit_{\diamondsuit}$ = 1.8 k Ω , $\diamondsuit_{\diamondsuit\diamondsuit}$ = 20 Ω , $\diamondsuit_{\diamondsuit}$ = 400 Ω , $\diamondsuit_{\diamondsuit}$ = 12 k Ω , $\diamondsuit_{\diamondsuit\diamondsuit\diamondsuit}$ = 40 mA, $\diamondsuit_{\diamondsuit\diamondsuit(\diamondsuit\diamondsuit\diamondsuit)}$ = –1 V.

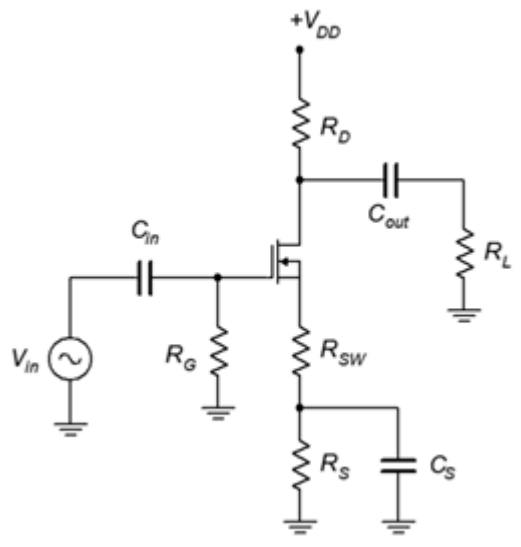


Figure 4.2.4 : Circuit for Example 4.2.1 .

This is a swamped common drain amplifier utilizing self bias. $\diamondsuit_{\diamondsuit\diamondsuit}$ can be determined via inspection.

$$Z_{in} = Z_{in(gate)} || R_G$$

$$Z_{in} \approx 1M\Omega$$

To find the load voltage we'll need the voltage gain, and to find the gain we'll first need to find $\diamondsuit_{\diamondsuit 0}$.

$$g_{m0} = -\frac{2I_{DSS}}{V_{GS(off)}}$$

$$g_{m0} = -\frac{80mA}{-1V}$$

$$g_{m0} = 80mS$$

The combined DC value of $\diamondsuit_{\diamondsuit}$ is 420 Ω , therefore $\diamondsuit_{\diamondsuit 0}\diamondsuit_{\diamondsuit}$ = 33.6. From the self bias equation or graph this produces a drain current of 1.867 mA.

$$g_m = g_{m0}\sqrt{\frac{I_D}{I_{DSS}}}$$

$$g_m = 80mS\sqrt{\frac{1.867mA}{40mA}}$$

$$g_m = 17.3mS$$

The swamping resistor, $\diamondsuit_{\diamondsuit}$, is 20 Ω . The voltage gain is

$$A_v = -\frac{g_m r_L}{g_m r_S + 1}$$

$$A_v = -\frac{17.3mS(1.8k\Omega||12k\Omega)}{17.3mS \times 20\Omega + 1}$$

$$A_v = -20.1$$

And finally

$$V_{load} = A_v V_{in}$$
$$V_{load} = -20.1 \times 20mV$$
$$V_{load} = 402mV$$

COMPUTER SIMULATION

The amplifier of Example 4.2.1 is simulated to verify the results. The circuit is entered into the simulator as shown in Figure 4.2.5 . One issue is finding an appropriate DE-MOS device to match the parameters used in the example. The BSS229 proves to be reasonably close. This device model was tested for ◈◈◈◈ by applying a 20 volt source to the drain and shorting the source and gate terminals to ground in the simulator. The current was just under the 40 mA target. Similarly, a negative voltage was attached to the gate and adjusted until the drain current dropped to nearly zero in order to determine ◈◈◈(◈◈◈) . The model's value was just under the desired –1 volt. Consequently, we can expect the simulation results to be close to those predicted, although not identical.

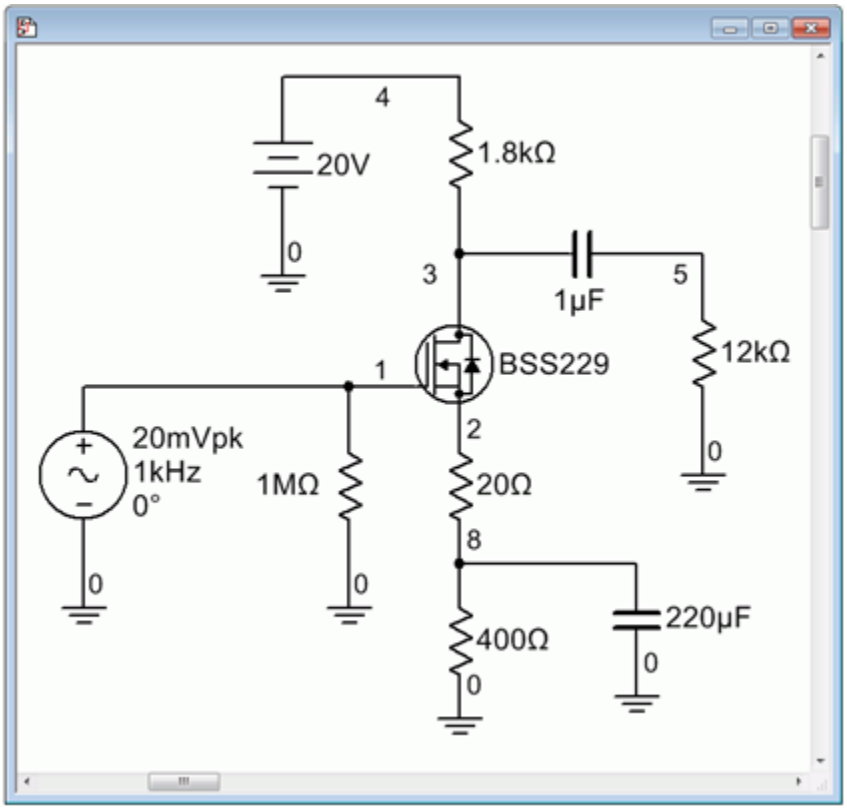


Figure 4.2.5: The circuit of Example 4.2.1 in the simulator.

The transient analysis is run next and is shown in Figure 4.2.6 . The expected signal inversion is obvious. The peak amplitude is 417 mV, just a few percent higher than the calculated value. At least some of this deviation is due to the model's variation from the assumed device parameter values.

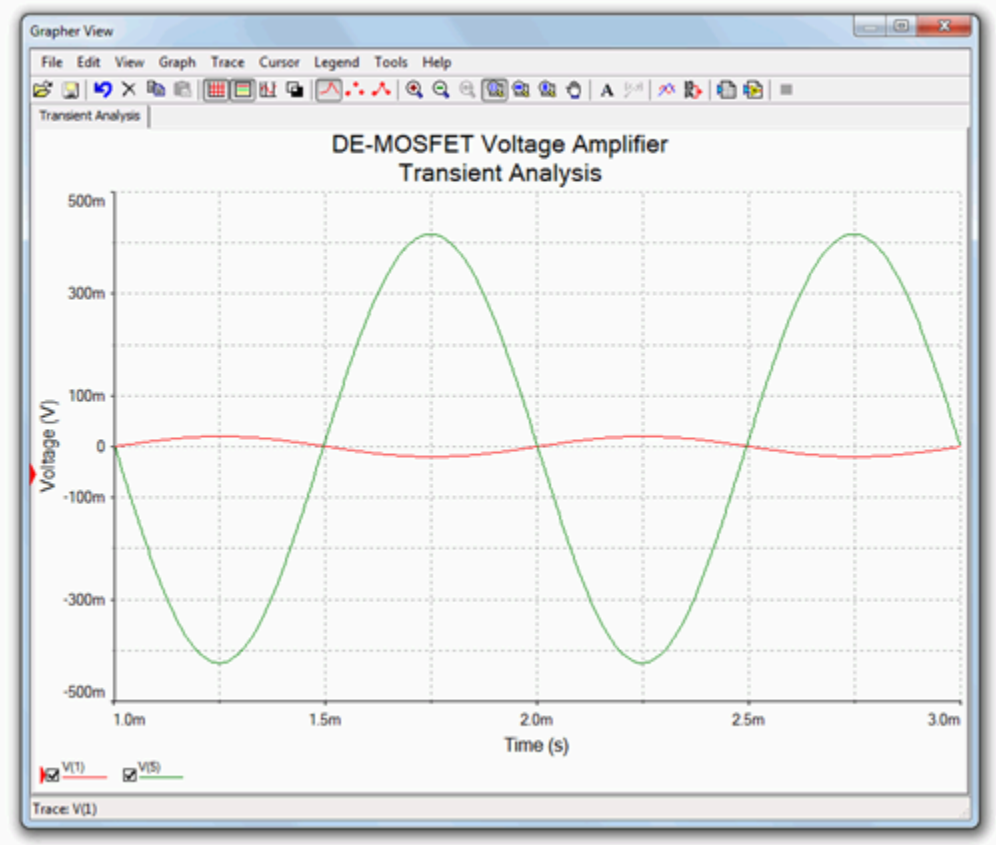


Figure 4.2.6 : Transient analysis simulation for the circuit of Example 4.2.1 .

A DC bias check is also performed. The drain current was calculated to be 1.867 mA. This yields an ⍰ voltage of a little over 3 volts, thus we expect to see a drain voltage of about 17 volts. Similarly, we would expect the source terminal to be sitting at around 700 to 800 mV and the gate at about 0 V.

The results of the DC operating point simulation are shown in Figure 4.2.7 . The agreement with the predicted values is quite good, especially considering that the device model is not a perfect match.

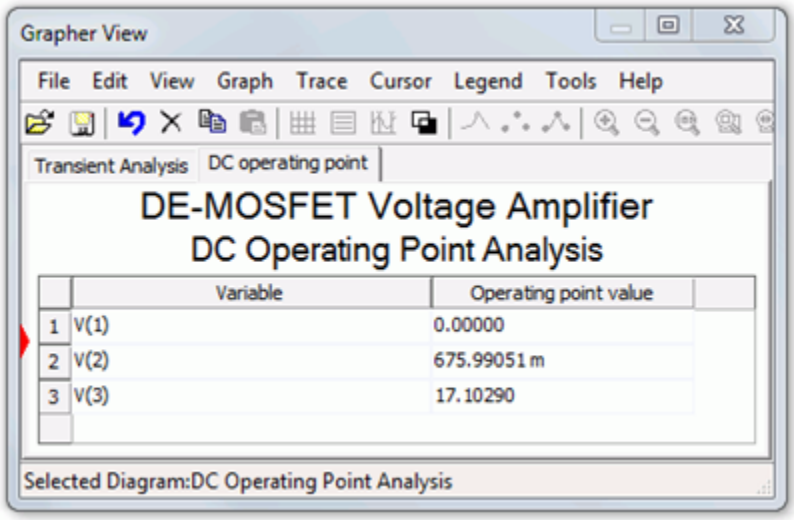


Figure 4.2.7 : DC bias simulation for the circuit of Example 4.2.1 .

Example 4.2.2

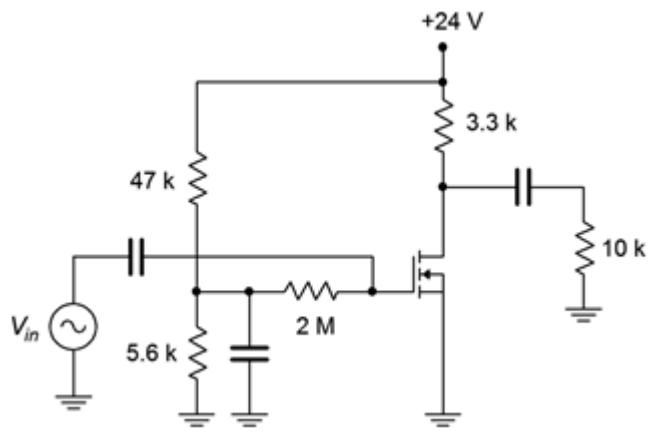


Figure 4.2.8 : Circuit for Example 4.2.2 .

For the circuit of Figure 4.2.8 , determine the voltage gain and input impedance. Assume ◈◈◈(◈*h*) = 2 V, ◈◈(◈◈) = 50 mA at ◈◈◈(◈◈) = 5 V.

First find the value of ◈ :

$$k = \frac{I_{D(on)}}{(V_{GS(on)} - V_{GS(th)})^2}$$

$$k = \frac{50mA}{(5V - 2V)^2}$$

$$k = 5.56mA/V^2$$

This circuit uses power supply decoupling. The voltage drop across the 2 M Ω resistor is small enough to ignore as the current passing through it is gate current. Therefore the gate voltage is determined by the divider. Also, as the left end of the 2 M Ω resistor is tied to an AC ground due to the bypass capacitor, it represents the input impedance.

$$Z_{in} = 2M\Omega || Z_{in(gate)} \approx 2M\Omega$$

$$V_G = V_{DD}\frac{R_2}{R_1 + R_2}$$

$$V_G = 24V\frac{5.6k\Omega}{47k\Omega + 5.6k\Omega}$$

$$V_G = 2.56V$$

The source is grounded so ◈◈◈=◈◈ .

$$I_D = k(V_{GS}V_{GS(th)})^2$$

$$I_D = 5.56mA/V^2(2.56V - 2V)^2$$

$$I_D = 1.74mA$$

$$g_m = 2k(V_{GS} - V_{GS(th)})$$

$$g_m = 2 \times 5.56mA/V^2(2.56V - 2V)$$

$$g_m = 6.23mS$$

This amplifier is not swamped so the simplified gain equation may be used.

$$A_v = -g_m r_D$$

$$A_v = -6.23mS(3.3k\Omega||10k\Omega)$$

$$A_v = -15.5$$

Example 4.2.3

For the circuit of Figure 4.2.9 , determine the voltage gain and input impedance. Assume $V_{GS(off)}$ = –0.75 V and I_{DSS} = 6 mA.

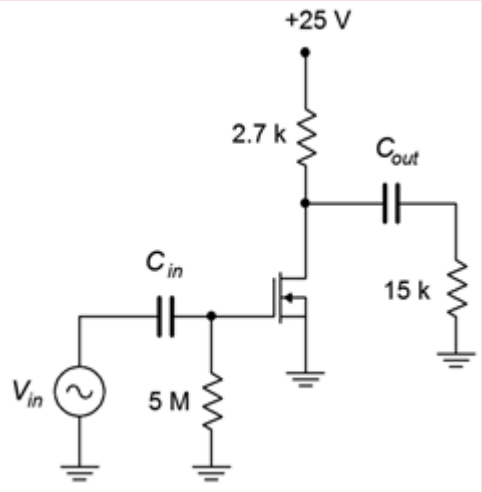


Figure 4.2.9 : Circuit for Example 4.2.3 .

This amplifier uses zero bias, therefore g_m=g_{m0} and g_m=g_{m0} .

$$g_{m0} = -\frac{2I_{DSS}}{V_{GS(off)}}$$

$$g_{m0} = -\frac{2 \times 6mA}{-0.75V}$$

$$g_{m0} = 16mS$$

This amplifier is not swamped so we may use the simplified equation for voltage gain.

$$A_v = -g_m r_D$$

$$A_v = -16mS(2.7k\Omega||15k\Omega)$$

$$A_v = -36.6$$

Finally, for the input impedance we have

$$Z_{in} = 5M\Omega||Z_{in(gate)} \approx 5M\Omega$$

4.3 MOSFET COMMON DRAIN FOLLOWERS

As discussed under the section on JFETs, the common drain amplifier is also known as the source follower. The prototype amplifier circuit with device model is shown in Figure 4.3.1 . As with all voltage followers, we expect a non-inverting voltage gain close to unity with a high ◈◈◈ and a low ◈◈◈◈ .

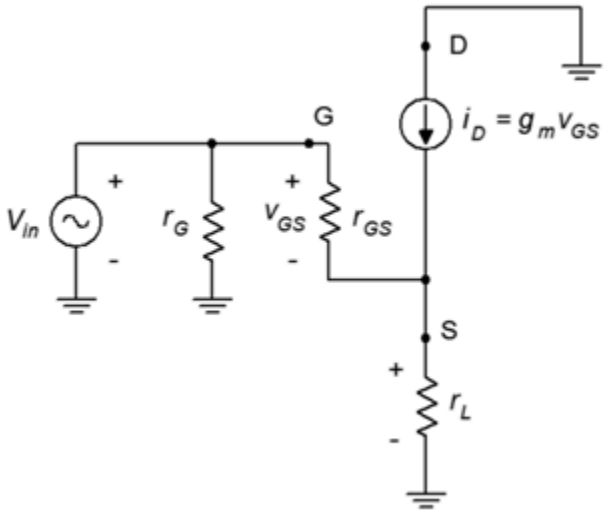


Figure 4.3.1 : Common drain (source follower) prototype.

As is usual, the input signal is applied to the gate terminal and the output is taken from the source. Because the output is at the source, biasing schemes that have the source terminal grounded, such as zero bias and voltage divider bias, cannot be used.

VOLTAGE GAIN

The voltage gain equation for the common drain follower is developed as follows: We begin with the fundamental definition that voltage gain is the ratio of ◈◈◈◈ to ◈◈◈ , and proceed by expressing these voltages in terms of their Ohm's law equivalents. The load is now located at the MOSFET's source, and thus can be referred to as either ◈◈ or ◈◈ .

$$A_v = \frac{v_{out}}{v_{in}} = \frac{v_S}{v_G} = \frac{v_L}{v_G}$$

$$A_v = \frac{i_D r_L}{i_D r_L + v_{GS}}$$

$$A_v = \frac{g_m v_{GS} r_L}{g_m v_{GS} r_L + v_{GS}}$$

$$A_v = \frac{g_m r_L}{g_m r_L + 1}$$

(4.3.1)

or, if preferred

$$A_v = \frac{g_m r_S}{g_m r_S + 1}$$

If ◆◆◆◆≫1 , the voltage gain will be very close to unity; a desired outcome.

INPUT IMPEDANCE

The analysis for source follower's input impedance is virtually identical to that for the common source amplifier. The same commentary applies regarding the simplification of gate biasing resistors to arrive at the value of ◆◆ .

$$Z_{in} = r_G || r_{GS} \approx r_G$$

(4.3.3)

OUTPUT IMPEDANCE

In order to determine the output impedance, we modify the circuit of Figure 4.3.1 by separating the load resistance from the source bias resistor. This is shown in Figure 4.3.2 .

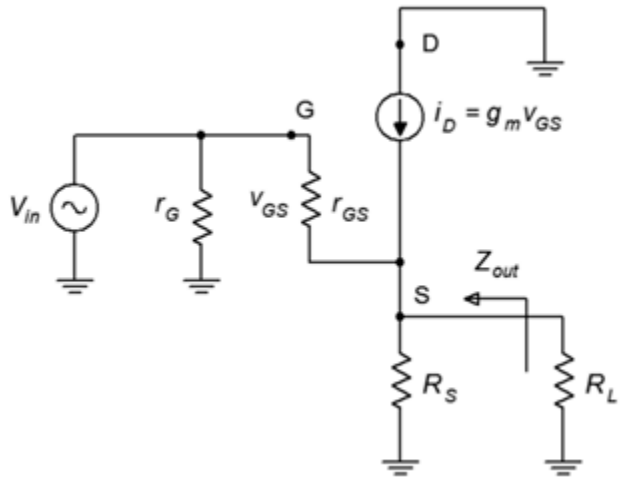


Figure 4.3.2 : Source follower output impedance analysis.

Looking back into the source from the perspective of the load we find that the source biasing resistor, ◆◆, is in parallel with the impedance looking back into the source terminal.

$$Z_{out} = R_S || Z_{source}$$

To find ◆◆◆◆◆◆◆◆ , note that the voltage at the source is ◆◆◆ and the current entering this node is ◆◆ . The ratio of the two will yield the impedance looking back into the source.

$$Z_{source} = \frac{v_{GS}}{i_D}$$

$$Z_{source} = \frac{v_{GS}}{g_m v_{GS}}$$

$$Z_{source} = \frac{1}{g_m}$$

(4.3.4)

Therefore, the output impedance is

$$Z_{out} = R_S || \frac{1}{g_m}$$

(4.3.5)

Looking at Equation 4.3.5 it is obvious that the higher the transconductance, the lower the output impedance. As noted earlier, a large transconductance also means that the voltage gain will be close to unity. As a general rule then, a large transconductance is desired for the source follower.

Time for a few illustrative examples.

Example 4.3.1

For the circuit of Figure 4.3.3 , determine the voltage gain and input impedance. Assume $\diamondsuit_{\diamondsuit\diamondsuit(\diamondsuit\diamondsuit\diamondsuit)}$ = –0.8 V and $\diamondsuit_{\diamondsuit\diamondsuit\diamondsuit}$ = 30 mA.

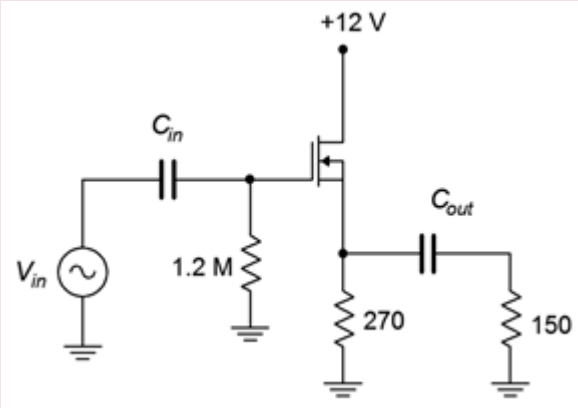


Figure 4.3.3 : Circuit for Example 4.3.1 .

This amplifier uses self bias so we need to determine $\diamondsuit_{\diamondsuit 0}\diamondsuit_{\diamondsuit}$.

$$g_{m0} = -\frac{2I_{DSS}}{V_{GS(off)}}$$

$$g_{m0} = -\frac{2 \times 30mA}{-0.8V}$$

$$g_{m0} = 75mS$$

The DC source resistance is the 270 Ω biasing resistor resulting in $\diamondsuit_{\diamondsuit 0}\diamondsuit_{\diamondsuit}$ = 16.2. From the self bias equation or graph this produces a drain current of 2.61 mA.

$$g_m = g_{m0}\sqrt{\frac{I_D}{I_{DSS}}}$$

$$g_m = 75mS\sqrt{\frac{2.61mA}{30mA}}$$

$$g_m = 22.1mS$$

The voltage gain is

$$A_v = \frac{g_m r_S}{g_m r_S + 1}$$

$$A_v = \frac{22.1mS(270\Omega||150\Omega)}{22.1mS \times (270\Omega||150\Omega) + 1}$$

$$A_v = 0.68$$

Finally, for the input impedance we have

$$Z_{in} = 1.2M\Omega||Z_{in(gate)} \approx 1.2M\Omega$$

Example 4.3.2

For the circuit of Figure 4.3.4 , determine the voltage gain and input impedance. Assume $V_{GS(off)}$ = –2.5 V and I_{DSS} = 80 mA.

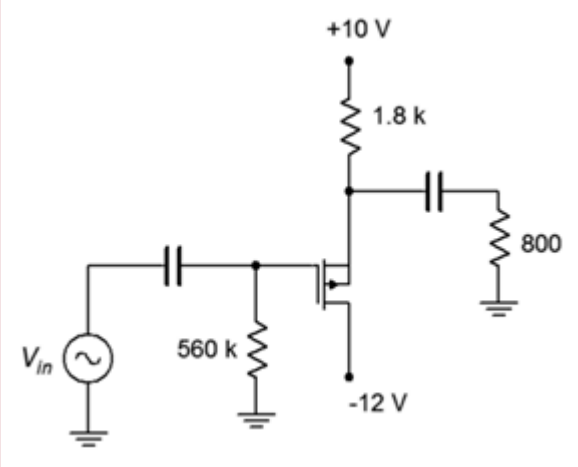


Figure 4.3.4 : Circuit for Example 4.3.2 .

This follower uses a P-channel device with combination bias. Note that the source terminal is toward the top of the schematic. First, determine I_{D0} V_{GS} and the bias factor, K . Then the combination bias equation can be used to determine the drain current.

$$g_{m0} = -\frac{2I_{DSS}}{V_{GS(off)}}$$

$$g_{m0} = -\frac{2 \times 80mA}{-2.5V}$$

$$g_{m0} = 64mS$$

The DC source resistance is the 1.8 k Ω biasing resistor resulting in $\diamondsuit_{\diamondsuit 0}\diamondsuit_{\diamondsuit}$ = 115.2. The bias factor is $\diamondsuit_{\diamondsuit\diamondsuit}/\diamondsuit_{\diamondsuit\diamondsuit(\diamondsuit\diamondsuit\diamondsuit)}$, or 4. The combination bias equation (Equation 10.9) yields $\diamondsuit_{\diamondsuit}$ = 6.67 mA.

We can now find the transconductance and voltage gain.

$$g_m = g_{m0}\sqrt{\frac{I_D}{I_{DSS}}}$$

$$g_m = 64mS\sqrt{\frac{6.67mA}{80mA}}$$

$$g_m = 18.5mS$$

The voltage gain is

$$A_v = \frac{g_m r_S}{g_m r_S + 1}$$

$$A_v = \frac{18.5mS(1.8k\Omega||800\Omega)}{18.5mS \times (1.8k\Omega||800\Omega) + 1}$$

$$A_v = 0.91$$

Lastly, the input impedance is

$$Z_{in} = 560k\Omega||Z_{in(gate)} \approx 560k\Omega$$

4.4 SUMMARY

DE- and E-MOSFET devices may be used to create both common source voltage amplifiers and common drain voltage followers. The common source amplifiers may be swamped or non-swamped, depending on the bias form used. If the bias type does not utilize a source resistor, swamping is not available. This includes zero bias for the DE-MOSFET and voltage divider bias for both the DE- and E-MOSFET.

Like their JFET counterparts, MOSFET common source amplifiers exhibit moderate inverting voltage gain, very high input impedance and moderate output impedance. The input impedance is a function of the biasing resistor configuration situated in front of the gate as the impedance looking into the gate itself is very, very high at low frequencies.

The MOSFET common drain followers also behave similarly to the JFET version. Again we see a non-inverting voltage gain approaching unity, a very high input impedance and a low output impedance. The higher the transconductance is, the closer the gain will be to unity and the lower the output impedance will be.

Review Questions

1. How well does the MOSFET voltage amplifier compare to its JFET counterpart?
2. How well does the MOSFET source follower compare to its JFET counterpart?
3. What are the practical differences between a voltage amplifier using a DEMOSFET versus using an E-MOSFET ?
4. It has been stated that a source follower cannot be made using a standard zero biased DE-MOSFET. Why is this?
5. It has been stated that a swamped voltage amplifier cannot be made using a standard voltage divider biased E-MOSFET. Why is this?

4.5 EXERCISES

ANALYSIS PROBLEMS

1. For the amplifier of Figure 4.5.1 , determine $\square_{\square\square}$ and $\square_{\square}$. $\square_{\square\square}$ = 20 mV, $\square_{\square\square\square}$ = 10 mA, $\square_{\square\square(\square\square\square)}$ = –2 V, $\square_{\square\square}$ = 20 V, $\square_{\square}$ = 750 k Ω , $\square_{\square}$ = 2 k Ω , $\square_{\square}$ = 4 k Ω , $\square_{\square}$ = 1 k Ω , $\square_{\square\square}$ = 200 Ω .
2. For the amplifier of Figure 4.5.1 , determine $\square_{\square\square}$ and $\square_{\square\square\square}$. $\square_{\square\square}$ = 25 mV, $\square_{\square\square\square}$ = 15 mA, $\square_{\square\square(\square\square\square)}$ = –2 V, $\square_{\square\square}$ = 22 V, $\square_{\square}$ = 330 k Ω , $\square_{\square}$ = 2 k Ω , $\square_{\square}$ = 6 k Ω , $\square_{\square}$ = 510 Ω , $\square_{\square\square}$ = 220 Ω .

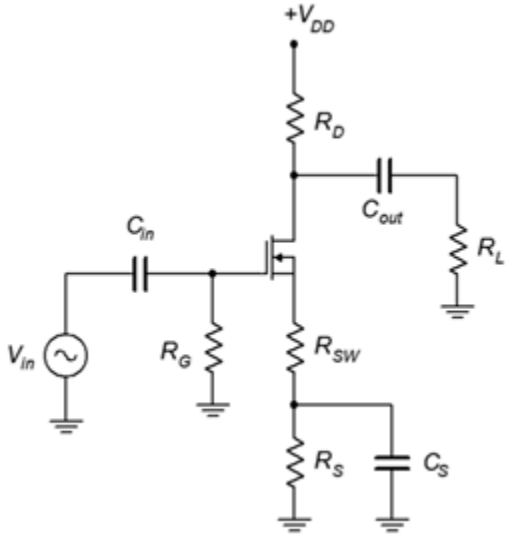


Figure 4.5.1

3. For the circuit of Figure 4.5.2 , determine $\square_{\square\square}$ and $\square_{\square}$. $\square_{\square\square}$ = 10 mV, $\square_{\square\square\square}$ = 12 mA, $\square_{\square\square(\square\square\square)}$ = –2.5 V, $\square_{\square\square}$ = 26 V, $\square_{\square}$ = 510 k Ω , $\square_{\square}$ = 1.2 k Ω , $\square_{\square}$ = 25 k Ω .
4. For the circuit of Figure 4.5.2 , determine $\square_{\square\square}$ and $\square_{\square\square\square}$. $\square_{\square\square}$ = 25 mV, $\square_{\square\square\square}$ = 15 mA, $\square_{\square\square(\square\square\square)}$ = –1.5 V, $\square_{\square\square}$ = 24 V, $\square_{\square}$ = 820 k Ω , $\square_{\square}$ = 1 k Ω , $\square_{\square}$ = 12 k Ω .

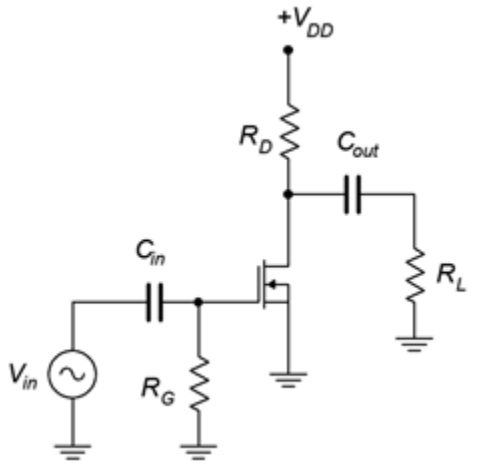


Figure 4.5.2

5. For the circuit of Figure 4.5.3 , determine $\square_{\square\square}$ and $\square_{\square\square\square}$. $\square_{\square\square}$= 25 mV, $\square_{\square\square\square}$ = 8 mA, $\square_{\square\square(\square\square\square)}$ = –3.5 V, $\square_{\square\square}$ = 24 V, $\square_1$ = 1 M Ω , $\square_2$ = 100 k Ω , $\square_{\square}$ = 800 Ω , $\square_{\square}$ = 10 k Ω .
6. For the circuit of Figure 4.5.3 , determine $\square_{\square\square}$ and $\square_{\square}$. $\square_{\square\square}$ = 10 mV, $\square_{\square\square\square}$ = 6 mA, $\square_{\square\square(\square\square\square)}$ = –4 V, $\square_{\square\square}$ = 26 V, $\square_1$ = 2 M Ω , $\square_2$ = 120 k Ω , $\square_{\square}$ = 1.2 k Ω , $\square_{\square}$ = 15 k Ω .

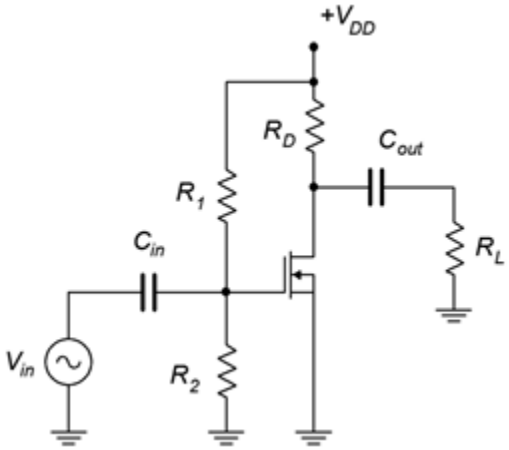


Figure 4.5.3

7. For the circuit of Figure 4.5.4 , determine $\square_{\square\square}$ and $\square_{\square\square\square}$. $\square_{\square\square}$ = 20 mV, $\square_{\square(\square\square)}$ = 6 mA at $\square_{\square\square(\square\square)}$ = 3 V, $\square_{\square\square(\square h)}$ = 2.5 V, $\square_{\square\square}$ = 34 V, $\square_1$ = 1 M Ω , $\square_2$ = 100 k Ω , $\square_{\square}$ = 1 k Ω , $\square_{\square}$ = 10 k Ω .
8. For the circuit of Figure 4.5.4 , determine $\square_{\square\square}$ and $\square_{\square}$. $\square_{\square\square}$ = 15 mV, $\square_{\square(\square\square)}$ = 10 mA at $\square_{\square\square(\square\square)}$ = 4 V, $\square_{\square\square(\square h)}$ = 2 V, $\square_{\square\square}$ = 30 V, $\square_1$ = 2 M Ω , $\square_2$ = 180 k Ω , $\square_{\square}$ = 1.2 k Ω , $\square_{\square}$ = 15 k Ω .

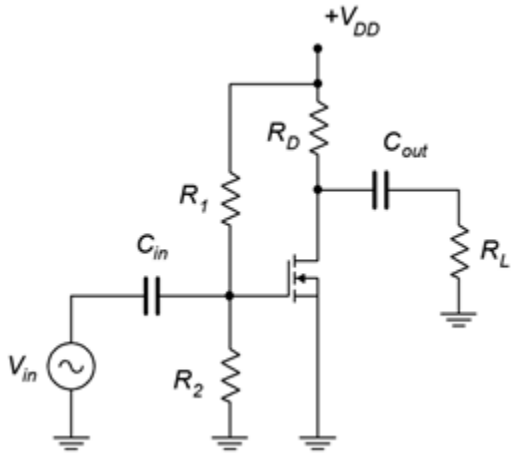


Figure 4.5.4

9. For the circuit of Figure 4.5.5 , determine ? and ? . ?= 200 mV, ? = 15 mA, ?(?) = –3 V, ? = 15 V, ? = 910 k Ω , ? = 10 k Ω , ? = 330 Ω .
10. For the circuit of Figure 4.5.5 , determine ? and ? . ? = 200 mV, ? = 20 mA, ?(?) = –2 V, ? = 12 V, ? = 1 M Ω , ? = 1.8 k Ω , ? = 220 Ω .

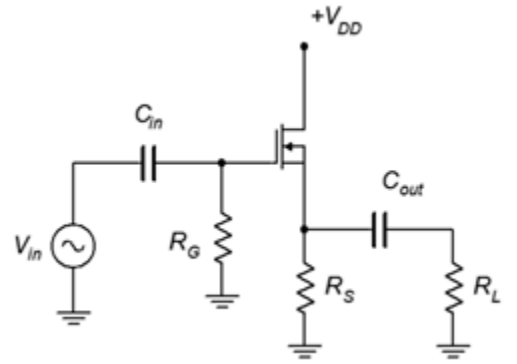


Figure 4.5.5

11. For the circuit of Figure 4.5.6 , determine ? and ? . ? = 18 mA, ?(?) = –2 V, ? = 12 V, ? = –4 V, ? = 680 k Ω , ? = 10 k Ω , ? = 1 k Ω .
12. For the circuit of Figure 4.5.6 , determine ? and ?. ? = 20 mA, ?(?) = –2 V, ? = 10 V, ? = –6 V, ? = 2.2 M Ω , ? = 5 k Ω , ? = 510 Ω .

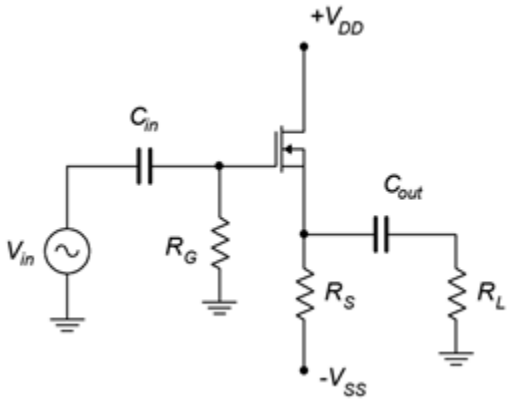


Figure 4.5.6

DESIGN PROBLEMS

13. Following the circuit of Figure 4.5.1 , design an amplifier with a gain of at least 5 and an input impedance of at least 500 k Ω . ◇ = 10 k Ω . The MOSFET has the following parameters: ◇◇(◇◇) = –2 V, ◇◇ = 25 mA. Try to use standard resistor values.
14. Using the circuit of Figure 4.5.5 , design a follower with a gain of at least .75 and an input impedance of at least 1 M Ω . ◇ = 2 k Ω . The MOSFET has the following parameters: ◇◇(◇◇) = –1.5 V, ◇◇ = 40 mA. Try to use standard resistor values.

CHALLENGE PROBLEMS

15. For the circuit of Figure 4.5.7 , determine ◇◇ and ◇ . ◇◇ = 15 mA, ◇◇(◇◇) = –2 V.

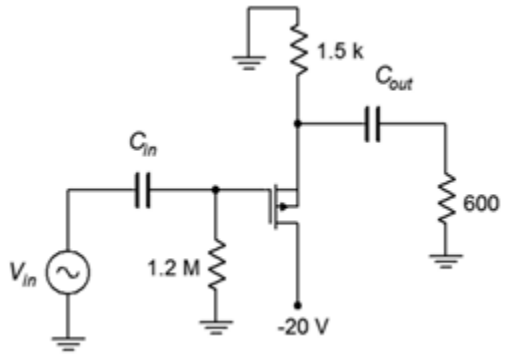


Figure 4.5.7

16. For the circuit of Figure 4.5.8 , determine ◇◇ and ◇ . ◇◇ = 12 mA, ◇◇(◇◇) = –1.5 V.

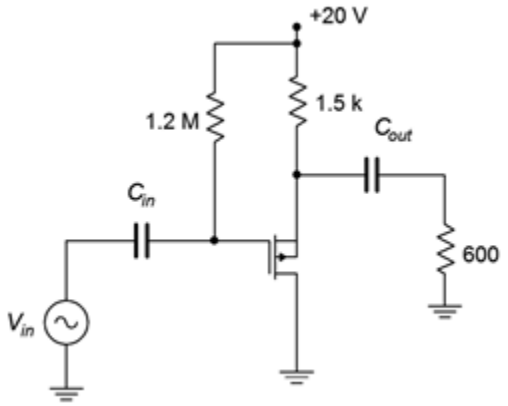


Figure 4.5.8

COMPUTER SIMULATION PROBLEMS

17. Utili ◈◈◈ g manufacturer's data sheets, find devices with the following specifications (typical) and verify them using the measurement techniques presented in the prior chapter.
    a. Device 1: ◈◈◈(◈◈◈) = –2 V, ◈◈◈◈ = 25 mA.
    b. Device 2: ◈◈◈(◈◈◈) = –1.5 V, ◈◈◈◈ = 40 mA.
18. Using the device model from the preceding problem, verify the design of Problem 4.
19. Using the device model from Problem 17, verify the design of Problem 14.

UNIT 5: CLASS D POWER AMPLIFIERS

Learning Objectives

After completing this chapter, you should be able to:

- Outline the building blocks of a class D amplifier.
- Discuss the advantages and disadvantages of the class D amplifier compared to class A and class B amplifiers.
- Explain the concepts of pulse width modulation, shoot-through and dead time.

5.1 INTRODUCTION

Much has been written in this text regarding the efficiency of various power amplifier topologies. While class A is known for its circuit layout simplicity, it is also known for its very low efficiency. Class B and class AB, while more complex than class A, present serious improvements in efficiency. In spite of these improvements, the family of class B amplifiers can hardly be considered as exhibiting high efficiency. Although not explicitly covered in this text, class G and H topologies are variations on class B and attempt to increase efficiency through the use of multiple sets of power supply rails or output devices, and in the process, tick the complexity up to another level.

The class D amplifier is perhaps the last word in amplifier efficiency. Theoretically with ideal devices, the efficiency of the output stage approaches 100%. Unlike the other amplifier forms, the transistors used in class D amplifiers never operate in the linear region; the output devices only operate as a switch, in either saturation or cutoff. High switching speed turns out to be a huge plus as it plays a major role in efficiency.

The increase in efficiency comes at a considerable increase in circuit complexity, however, for some applications this turns out to be a very good trade-off. As odd as it might at first seem, the two areas where class D topologies have taken root are at the opposite ends of the power output spectrum. The first area is perhaps the most obvious, mainly, very high output power amplifiers. An example might be an amplifier used as part of a large public address system and capable of delivering in excess of 1000 watts into a loudspeaker. High efficiency does two things here: First, it reduces the waste heat in the amplifier itself, and second, it reduces the current draw from the AC mains. Both of these are serious issues in a PA system used to fill a stadium or large concert hall as there may be dozens of such amplifiers comprising the system. As a bonus, improved efficiency also leads to a lighter and small enclosure because the need for heat sink area and mass will be reduced, as will the size of the AC power supply transformer. These traits will also reduce production costs and help offset the design complexity cost. The advantages have become so great that, in recent years, class D designs dominate the high end professional audio power amplifier market as well as the very high power automotive audio market (here there is another system limitation working in favor of class D, and that's the limited current capacity of the vehicle's alternator to deliver current).

The second area where class D has found acceptance is for low power portable devices. Examples include personal music devices, cell phones and hearing aids. Output powers for these applications might range from tens of milliwatts up to a few watts, so excess heat is generally not a big problem except in the most compact of enclosures. What is a problem, though, is the energy budget. Unlike a large PA amplifier that might pump out in excess of two horsepower, these portable devices do not have the luxury of running off of the AC mains with tens or even hundreds of amps of current capacity. Instead, these devices are restricted to battery power and batteries can only store so much energy. For a given battery capacity, higher efficiency directly translates into longer battery life. Another way of thinking about this is that, given a higher efficiency, a smaller battery can be used to achieve the same battery life, and this means that the unit can be both smaller and less expensive. Of course, nothing says that we can't opt for a little of each.

5.2 CLASS D BASICS

The key to the high efficiency of class D operation is to only operate the output devices as switches. That is, they are operated at the extreme ends of the load line, in either cutoff or saturation. The only exception to this rule is when the output device is transitioning from one state to another. Either BJTs or E-MOSFETs can be used, although for reasons that we shall examine, E-MOSFETs tend to be preferred in many applications and therefore we shall use them here as a general rule.

To understand the power advantage of class D switching, consider the circuit of Figure 5.2.1 . Here we have an E-MOSFET being driven by a square wave at its gate terminal. The square wave runs from zero to some voltage well above ◈◈◈(◈*h*) , sufficient to fully turn-on the MOSFET.

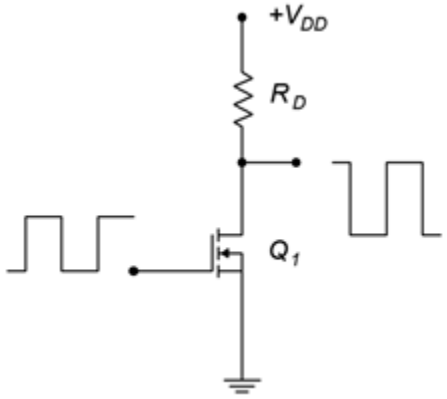


Figure 5.2.1: A simple MOSFET switch.

At the drain, an inverted square wave appears that ideally runs from +◈◈◈ down to zero volts. When the MOSFET is off, no drain current flows. Therefore, the drop across the drain resistor, ◈◈ , is zero. Consequently, by KVL, ◈◈◈ must equal ◈◈◈ . When the gate signal goes high, it turns on the MOSFET causing a large current. Ideally, this current equals ◈◈◈/◈◈ and ◈◈◈=0 V. In reality, the MOSFET will present some resistance (◈◈◈(◈◈)) and this essentially creates a voltage divider between the device and ◈◈. Obviously, if ◈◈◈(◈◈)≪◈◈ then we can approximate the low state as zero volts.

In the ideal case, the transistor dissipates no power. Here's why: When the gate is low, the device is off so no current flows. Although the device voltage is very high, the product of the device's voltage and current is zero. When the gate is high, the transistor turns on and conducts maximum current, however, the voltage across the device is zero, and the resulting product is zero once again. Therefore, the device dissipates no power.

The reality of the situation reveals that some power is indeed wasted by the output device. There are two chief culprits: non-zero turn-on voltage (caused by ◈◈◈(◈◈) for example) and state transitions that are not instantaneous (i.e., the rise and fall times of the waveforms are not zero). These effects are illustrated in Figure 5.2.2 .

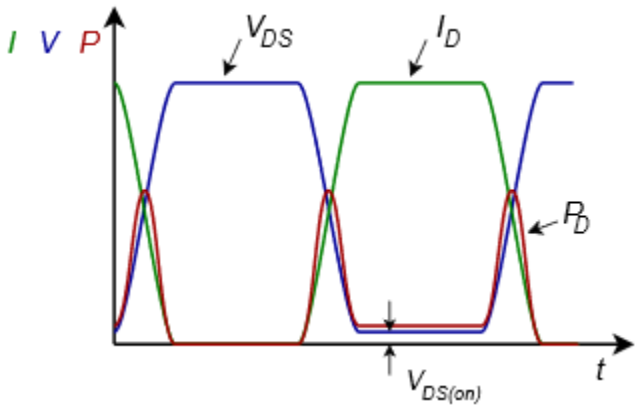


Figure 5.2.2 : Switching losses.

The rise and fall times are exaggerated for clarity as is the voltage drop across the transistor. The green curve represents the device current while the blue curve indicates the device voltage. Note that during the on-state the transistor voltage is not zero, although it is very small. The product of these two curves is the power dissipated by the transistor and is shown in red. In the off-state, no power is dissipated. In the on-state, a small amount of power is dissipated due to the device's on-voltage. Power spikes also occur at the state transitions as neither the current nor the voltage are at zero.

The area under the red curve represents the total power dissipated by the output transistor. In order to minimize this loss and maximize the efficiency, we would like as low of an $R_{DS(on)}$ as possible along with very fast switching speeds. As power EMOSFETs tend to switch faster than their BJT counterparts, this presents a compelling argument for their use. Further, large power E-MOSFETs may exhibit $R_{DS(on)}$ values of just a few milliohms, guaranteeing relatively small on-state losses (compared to load power) even with drain currents of several tens of amps.

5.3 PULSE WIDTH MODULATION

Clearly, class D presents the possibility of minimal wasted power and high efficiency. We are now left with the problem of how to turn a series of pulses into a continuous, smoothly varying waveform, such as a voice or music signal. There are a few ways to accomplish this; it's a matter of encoding the amplitude of the original signal into the pulse train that drives the output devices. Theoretically, as long as the "area under the curve" for a segment of input signal is identical to the area represented by the pulse train, we will have successfully encoded the signal and we then should be able to decode it, turning it back into a smoothly varying output signal. For this to work properly, the pulse train will have to be at much higher frequency than the input signal in order to follow its changes over time. One way to do this is through pulse density modulation, or PDM. The idea is to produce a number of narrow pulses to represent the area. If the input amplitude is large, we create a large number of pulses and if the amplitude is small, we produce a small number of pulses. While this technique can work, it is somewhat challenging to turn this pulse train back into the desired signal at the load.

Another technique to encode the input is pulse width modulation, or PWM. Instead of altering the number of pulses in a given period of time, we keep the frequency constant and adjust the width of the pulses. If the input amplitude is high, the width of the corresponding pulse will be wide and if the amplitude is low, the pulse width will be narrow. The decoding of PWM is easier than that of PDM and is generally the preferred route.

Generating PWM is a relatively straightforward affair. All we need is a triangle wave and a comparator. The comparator has two input terminals: the signal to be encoded (input signal) and the reference wave (triangle wave). It has a two-state, logical output. The output will be high if the signal is more positive than the reference and it will be low if the reference is more positive than the signal. This is shown in block diagram form in Figure 5.3.1 .

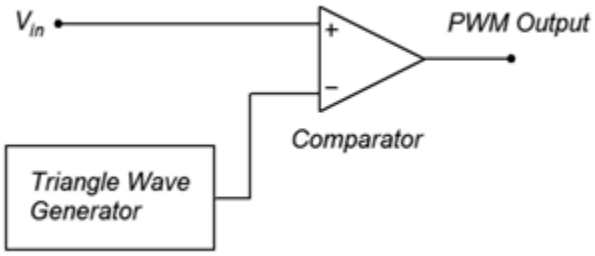


Figure 5.3.1 : PWM encoder.

As mentioned, the triangle wave needs to be at a much higher frequency than the signal that is being encoded. As a general rule, this frequency should be at least ten times the highest input signal frequency.

$$f_{triangle} \geq 10 f_2$$

(5.3.1)

A simulation showing the PWM waveforms is presented in Figure 5.3.2 .

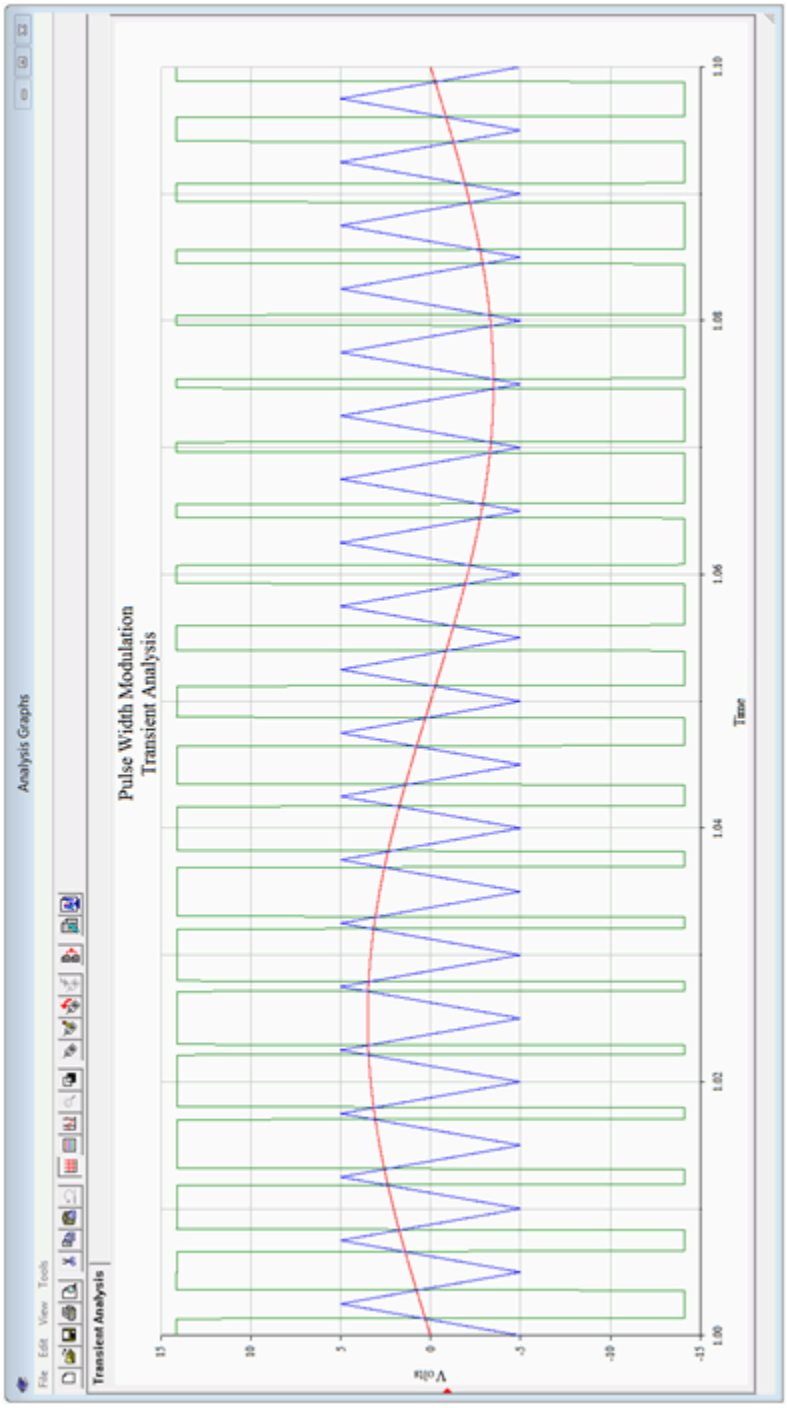


Figure 5.3.2 : PWM waveforms.

The input signal is the red sine wave. The blue triangle wave is the reference and is approximately 20 times higher in frequency. The green wave is the PWM output. Note that when the red input signal climbs above the blue reference triangle, the green output goes high, otherwise, the output is low. Thus, the duty cycle of the pulses correlates with the input signal amplitude. The input signal should not exceed the amplitude of the triangle wave otherwise accuracy will impaired. Also, the accuracy of the encoding process is dependent on the linearity of the triangle wave, so a high quality triangle wave generator is needed. Lastly, for accuracy and ease of decoding, the output pulses should not be

allowed to become too thin, so the input signal should be limited to about 75% of the amplitude of the triangle wave.

RECONSTITUTING THE OUTPUT

Although we have successfully encoded the input signal into a series of pulses, we still need to decode them, that is, reconstitute the original continuously variable input signal. Mathematically, the PWM signal contains all of the original input signal frequency components and amplitudes, it just has added a large of number of new frequency components. These new components are multiples (harmonics) of the fundamental PWM frequency, and consequently, they are all higher than the input signal frequencies. As such, they may be removed with a low-pass filter. This will effectively reconstitute the original signal (but at a much higher amplitude, of course). A simple passive ⍰⍰ filter would be appropriate in this instance as it must pass large currents and voltages. An example is shown in Figure 5.3.3 .

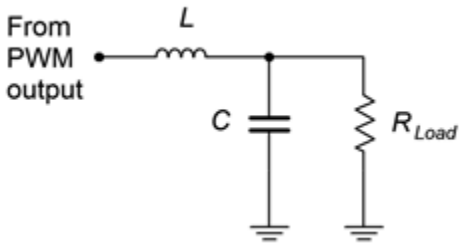


Figure 14.3.3: Passive low-pass filter.

At low frequencies, ⍰⍰ will be very small while ⍰⍰ will be very high, thus virtually all of the input signal frequencies will pass on to the load. At high frequencies, such as the harmonics of the PWM pulses, the situation is reversed: ⍰⍰ is large and ⍰⍰ is small, creating a large loss so that these components do not reach the load.[1] The critical frequency of the network is set to the highest input signal frequency (e.g., for high fidelity audio, slightly above 20 kHz).

We now have a complete outline for the class D amplifier, as shown in Figure 14.3.4 .

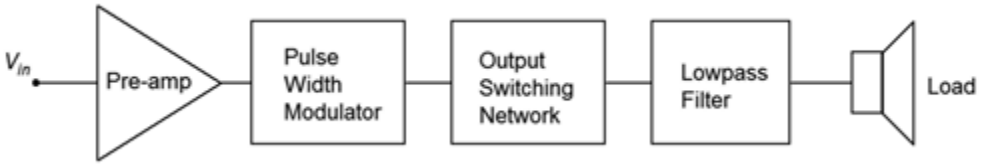


Figure 14.3.4 : Class D amplifier block diagram.

The pre-amp can be comprised of any of the linear amplifier outlines presented in earlier chapters, whether they use BJTs or FETs. What remains then, is further investigation into the output switching network.

1. For an audio amplifier, it is important that these components do not reach the loudspeaker. Even though they are beyond the range of human hearing, they can damage loudspeaker sub-components and, at the very least, present an extra power dissipation burden to them. Other kinds of loads may not be effected by the harmonics and filtering may not be needed.

If we apply the switching concept to a dual supply, push-pull topology, we arrive at the generic circuit of Figure 5.4.1 .

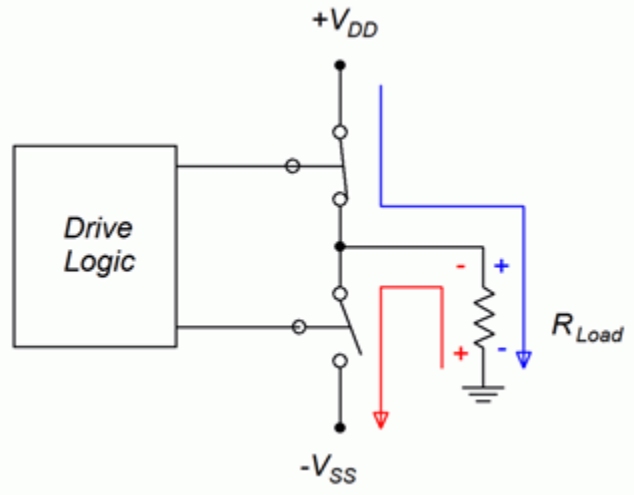


Figure 5.4.1 : Generic push-pull switching.

The two output devices are alternately switched on and off. When the upper device is on, the lower device is off, and current flows from the positive supply to the load (blue path). Alternately, when the lower device is on, the upper device is off, and current flows through the load via ◈◈◈ (red path). We could use either BJTs or EMOSFETs for these devices.

Two obvious variations exist of the generic output circuit. The first version, shown in Figure 5.4.2, appears to be a direct take-off of a class B output. It is shown with EMOSFETs but could be made with BJTs. Biasing details are not shown, instead a generic "driver" circuit block will prove sufficient for our discussion.

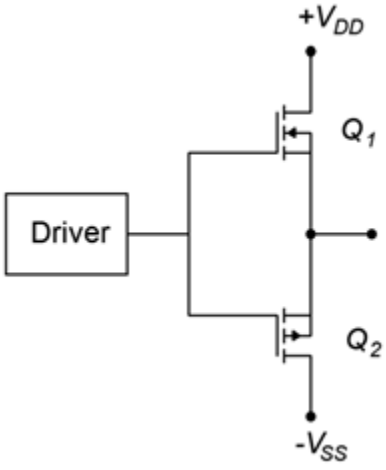
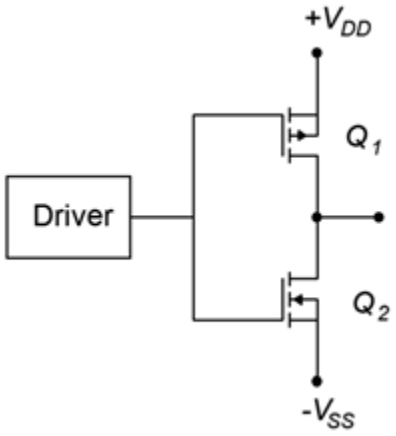


Figure 5.4.2: Basic push-pull switching.

In this circuit, the driver produces a bipolar pulse train that swings from negative to positive rather than from ground to positive. A positive level from the driver will turn on the upper N-channel device, allowing current flow to the load. In contrast, a negative level will turn on the lower P-channel device, allowing load current flow via ◈◈◈ .[1] It is worth noting that the gate drive signal must swing higher and lower than the two power supplies. This is because when a device is on, ◈◈◈ will be nearly zero, meaning that the source will be at the power rail. As ◈◈◈ must be greater than ◈◈◈(◈h) , this means that ◈◈ must be greater than the power supply.

An alternate connection scheme is shown in Figure 5.4.3 . Here, the N- and P-channel devices have switched positions.



The logic here is reversed: the negative pulse turns on the upper device and the positive pulse turns on the lower device. The gate swing is lessened a little compared to the earlier circuit but it suffers from a flaw common to both configurations, namely, asymmetry between the N- and P-channel device characteristics. This includes variations between internal device capacitances and ◈◈◈(◈◈) values. For the best possible matching, and thus the lowest distortion and highest performance, it would be better to configure the output using identical devices. This is shown in Figure 5.4.4 .

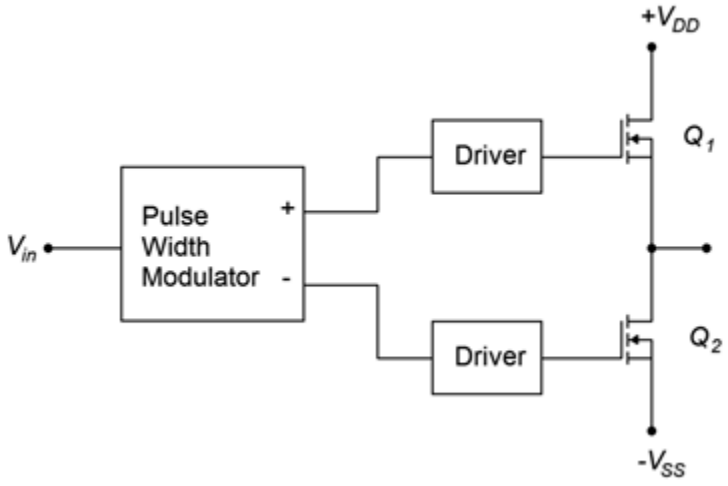


Figure 5.4.4: Output switching using identical devices.

This configuration complicates the drive signal in that we can no longer drive both gates with the

1. Yes, it is labeled ◈◈◈ in spite of the fact that it's connected to the drain of the P-channel device. It's a matter of consistency with other circuits. "A rose by any other name" and all that...

same signal; instead, unique signals must be presented to each gate terminal. This circuit is known as a half bridge. Our final step will be to drive both ends of the load in differential fashion using a full bridge, or H bridge as it is sometimes known. This is shown in Figure 5.4.5 .

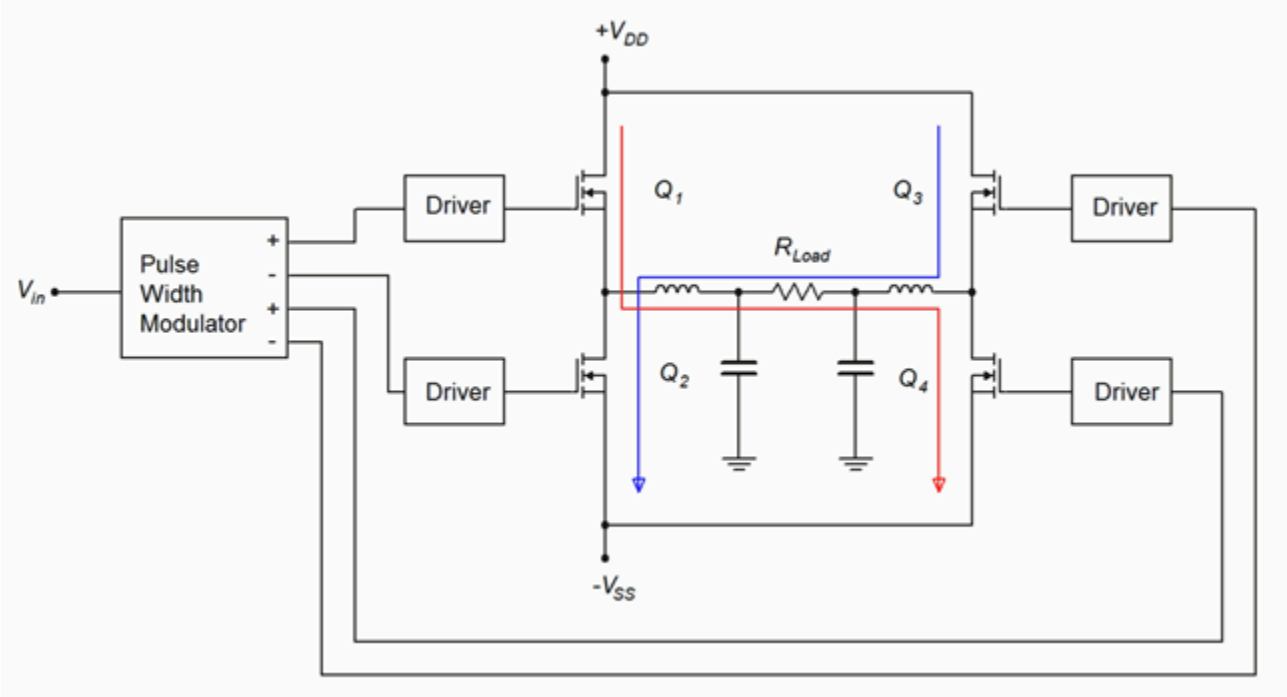


Figure 5.4.5 : Full bridge output.

The output devices are controlled as diagonal pairs. When $\diamondsuit_1$ is on, $\diamondsuit_4$ is on, creating a path for load current from left to right (red trace). In contrast, when $\diamondsuit_2$ is on, $\diamondsuit_3$ will also be on, thus creating a load current path from right to left (blue trace). This effectively doubles the current amplitude which quadruples the load power (because power varies as the square of current). This is the same technique discussed in Chapter 9 with class B amplifiers. A dual $\diamondsuit_\diamondsuit$ filter is included in this diagram to remove unwanted frequency components.

Example 5.4.1

A pair of E-MOSFETs are configured to drive an 8 Ω load as in Figure 5.4.4 . Assuming that ± 50 volt sources are used and that each device has an $\diamondsuit_{\diamondsuit\diamondsuit(\diamondsuit\diamondsuit)}$ of 0.03 Ω , determine the peak load current and $\diamondsuit_{\diamondsuit\diamondsuit}$ for the MOSFETs.

At any given time, one MOSFET will be on, creating a path between one supply, itself, the load and ground. The total resistance to limit the current will be the load plus $\diamondsuit_{\diamondsuit\diamondsuit(\diamondsuit\diamondsuit)}$.

$$i_{load} = \frac{V_{DD}}{r_{load} + r_{DS(on)}}$$

$$i_{load} = \frac{50V}{8\Omega + 0.03\Omega}$$

$$i_{load} = 6.227A$$

The device voltage is found via Ohms law as load and drain current are identical.

$$v_{DS} = i_{load} r_{DS(on)}$$

$$v_{DS} = 6.227A \times 0.03\Omega$$

$$v_{DS} = 0.19V$$

PRACTICAL CONCERNS

There are a few details left that should not be overlooked. Two of them are related to the edge transition areas, another concerns the complexity of the drive circuits, and the final issue deals with the power supplies themselves

The first item of concern is precisely what happens during the transition. All of the output forms we have examined utilize two active devices configured in series between two power sources. There is nothing in that path to limit current. If both devices were to be simultaneously triggered to the on-state, a huge and possibly damaging current would flow. While it would be foolish to turn both devices on intentionally, the rise and fall times of the pulses effectively do this. As one device is turning on and the other is turning off, both devices are in a conducting state, even if it's not maximum conduction. Essentially, we have two low impedance devices in series between two sources. This results in a large current pulse known as shoot-through. This situation is depicted in Figure 5.4.6 .

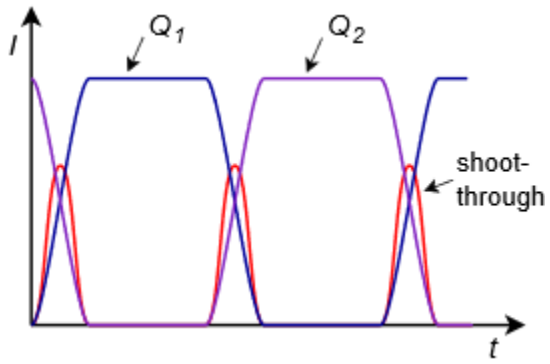


Figure 5.4.6: Shoot-through.

The current pulses for the two devices are shown in blue and violet. The maximal currents are directed to the load, but during the transition, a pulse of current, shown in red, "shoots through" the two devices, from one power supply directly to the other.

The solution to shoot-through is to adjust the turn-on and turn-off pulse timing so as to create a dead time, that is, a time span when neither device is directed to turn on. This is illustrated in Figure 5.4.7 .

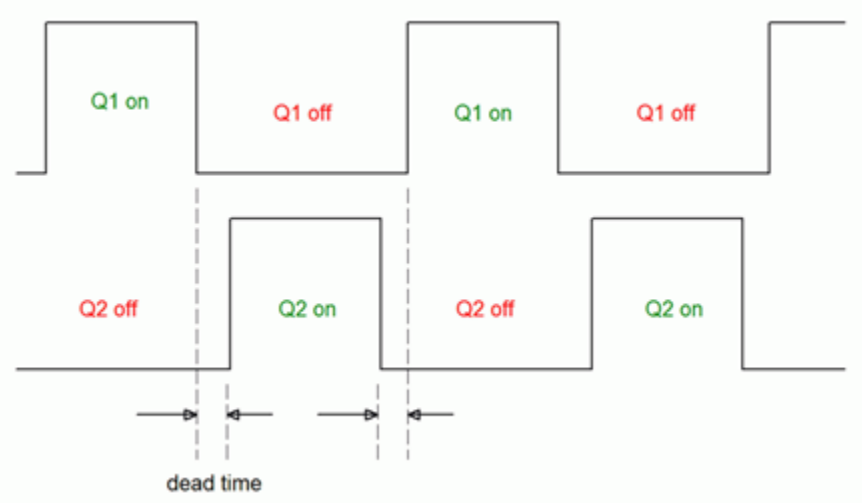


Figure 5.4.7 : Dead time.

Dead time is adjusted to correspond with the rise and fall times of the output devices. Basically, a device is not allowed to turn on until the other device is, indeed, fully off. The inclusion of dead time alters the width of the pulse and consequently can introduce waveform distortion. A minimum amount of dead time should be used to avoid this. This is another reason to use very fast output devices as they will require shorter dead times.

The second issue regarding timing is one of device capacitance. Power MOSFETs exhibit relatively high device capacitances. For example, the FDMS86180 examined in Chapter 12 exhibits input and output capacitances of roughly 4.4 nF and 2.7 nF, respectively. Although the extremely high gate input resistance might seem to indicate that very little drive current is needed to turn these devices on, the capacitance tells a different story.

The rate of change of voltage across a capacitor is a function of the capacitance and the current driving it:

$$\frac{dv_C}{dt} = \frac{i_C}{C}$$

The larger the current, or the smaller the capacitor, the greater the rate of change of voltage. This can place a serious limit on how quickly a device may be controlled. For example, suppose the drive circuit can pump out up to 10 mA. At first glance that may seem like an enormous amount of current to drive a MOSFET. Now, consider what happens if the input capacitance is 2 nF:

$$\frac{dv_C}{dt} = \frac{i_C}{C}$$

$$\frac{dv_C}{dt} = \frac{10mA}{2nF}$$

$$\frac{dv_C}{dt} = 5E6V/s$$

While a 5 million volt-per-second slope might sound fast, it's only 5 volts per microsecond. Compared to the requirements of, say, a 200 kHz to 300 kHz switching frequency, that is horribly slow.

COMPUTER SIMULATION

To see the effect of input capacitance, a two-stage amplifier is captured in a simulator, as shown in Figure 5.4.8 . The circuit consists of a relatively standard small signal amplifier feeding a medium power E-MOSFET, the IRF7201. A 10 kHz square wave is used to drive the circuit. The input capacitance of the MOSFET is 550 pF, certainly larger than a small signal FET but not an extremely large value. A single capacitor is placed across the gate that will be used to simulate a much larger device and the associated increased input capacitance.

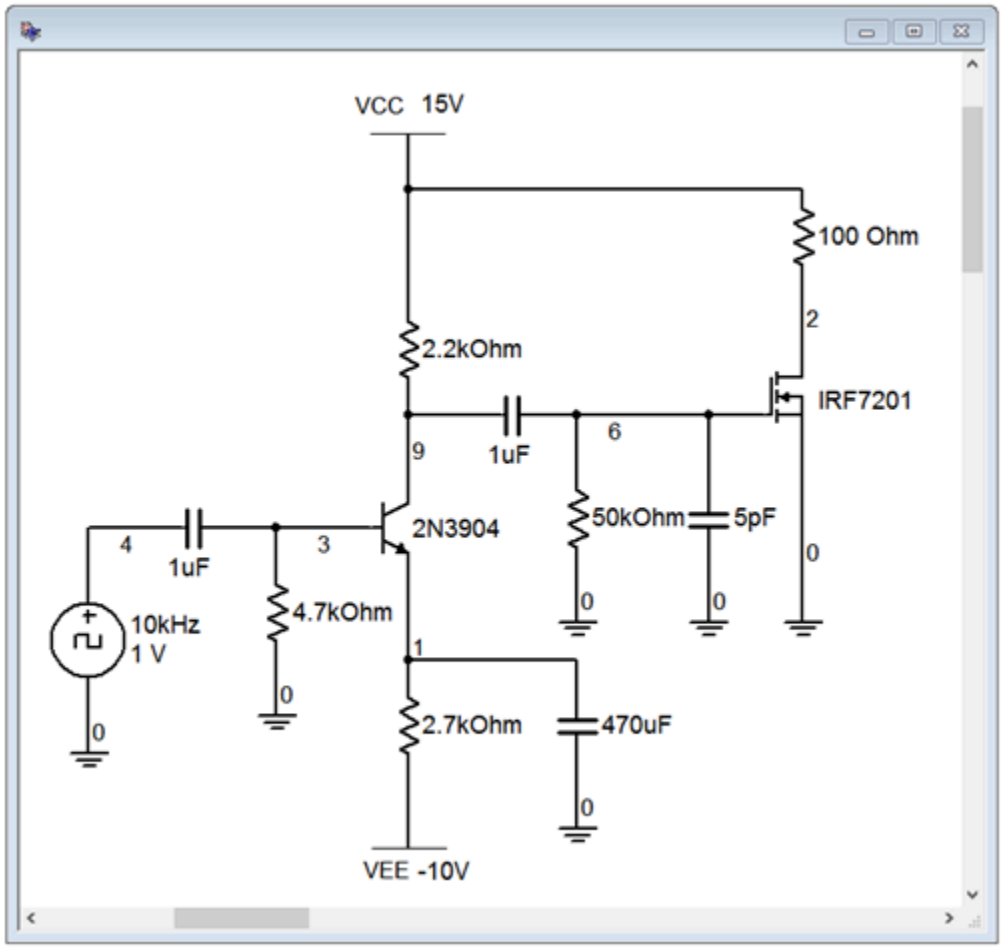


Figure 5.4.8: Circuit for input capacitance testing.

The initial transient analysis is run using a 5 pF gate capacitance which has no appreciable effect on the outcome. The result of the simulation is shown in Figure 5.4.9 .

The red trace is the gate voltage at node 6 while the blue trace is the final output at node 2. The gate drive signal is suffering somewhat at the upper portion and the rise and fall times are evident. The output signal is swinging from the power supply of +15 volts down to ground, as expected. The rise time is somewhat quicker than the fall time but, in general, the output presents a decent square pulse at close to 50% duty cycle. The simulation is repeated but this time the gate capacitance is increased from 5 pF to 5 nF, a value more typical of a large power FET. The result is shown in Figure 5.4.10 . The red gate drive signal has taken a serious hit and is no longer square in shape. The output pulse in blue still runs to and from the expected voltage levels, but the rising and falling edges are noticeably

slowed. Also, the positive pulse width has been stretched due to the slowing of the gate signal which retards the turn on of the MOSFET. The result is a duty cycle that is greater than 50%.

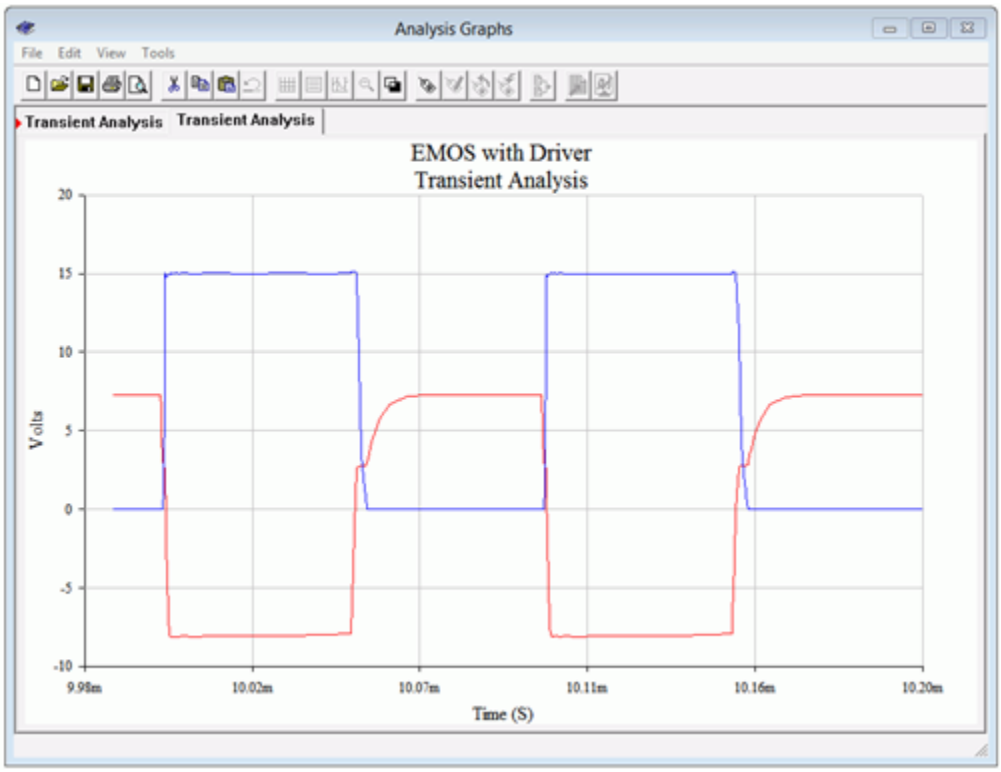


Figure 5.4.9: Waveforms for normal circuit.

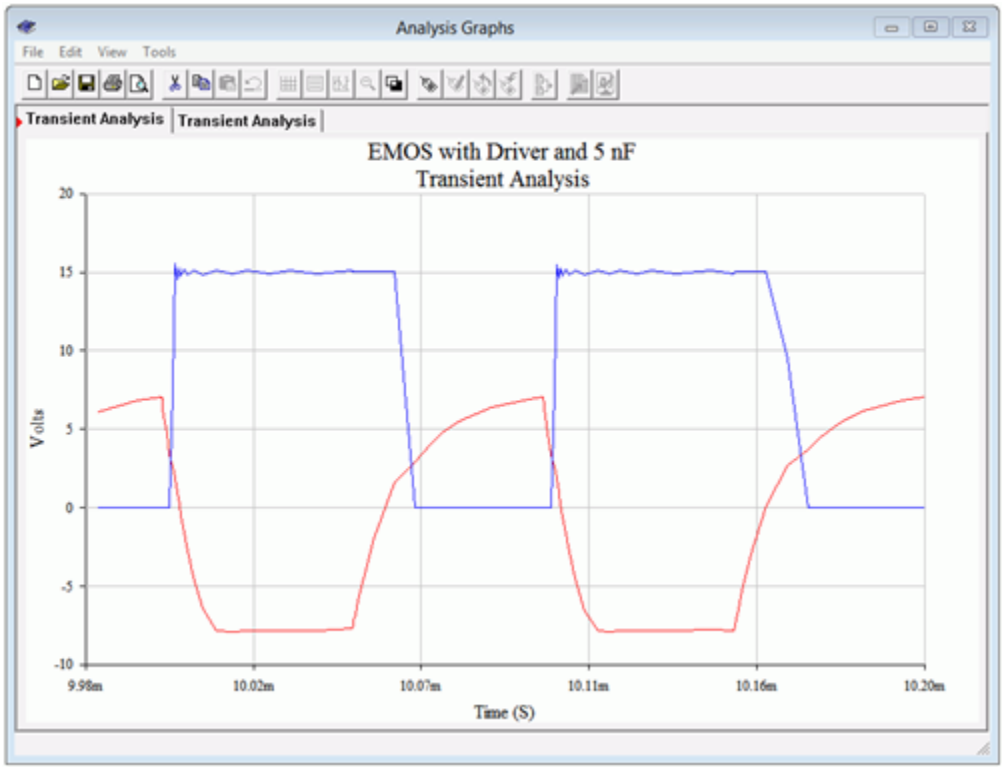


Figure 5.4.10: Waveforms for circuit with increased gate capacitance.

The bottom line is that, in order maximize speed, care must be taken to minimize capacitance,

decrease the output impedance of the driver circuit and increase the drive current. Fortunately, this issue has been largely solved by IC manufacturers who offer FET driver integrated circuits designed specifically for these applications.

The final item of practical concern is the power supply itself, or more precisely, the quality of the supply voltage. Remember, the output device is being used as a switch. When the device is on, the power supply is directly connected to the output (with the exception of the small voltage drop across the output device). This means that any noise or ripple on the power supply will make its way to the output filter. Whatever noise components fall within the desired input signal range will not be filtered out, and thus are delivered to the load. In short, the output devices will "leak" the power supply noise into the output, so care must be taken to have as clean of a power supply voltage as possible.

5.5 SUMMARY

The class D amplifier boasts very high efficiency, theoretically approaching 100%. The amplifier operates its output devices as switches; they are either fully on or fully off. The power losses are mostly relegated to switching edge losses so it is important to not switch the output devices at too high of a frequency.

The input signal is encoded as a series of pulses, typically via pulse width modulation. The pulse frequency is much higher than the highest input signal frequency, typically by an order of magnitude. The width of the pulses is a function of the amplitude of the input signal. That is, the higher the input signal, the greater the pulse width. The pulses are amplified by applying them to the output devices which then act as switches to alternately connect and disconnect the power supplies to the output terminal. The sequence of much larger amplitude pulses are then fed to a low-pass filter, typically, an ⍰⍰ filter, to remove the high frequency components of the pulse train. The reconstituted input signal is what remains, but at a much higher amplitude.

The output can be configured using either two-device half-bridge or four-device fullbridge arrangements. The full-bridge is preferred for higher performance. The input capacitance of the output devices can be relatively high, so care must be taken to ensure that sufficient capacity is available from the driver circuit.

Review Questions

1. What is PWM?
2. How does PWM differ from PDM (pulse density modulation)?
3. What is shoot-through? What causes it?
4. What is dead time? What is its purpose?
5. What is the function of the output ⍰⍰ filter?
6. What effect does device input (e.g., gate) capacitance have on the operation of the amplifier?
7. Sketch a half-bridge configuration.
8. Sketch a full-bridge configuration.
9. What is the effect of ⍰⍰⍰(⍰⍰) regarding load current and power losses?
10. What is the effect of drive current capacity on the output devices?
11. What effect does power supply noise have on the output?

5.6 EXERCISES

ANALYSIS PROBLEMS

1. A telephony system has a frequency range from 200 Hz to 3.5 kHz. Determine the minimum acceptable PWM frequency.
2. A background music and paging system has a frequency range from 50 Hz to 10 kHz. Determine the minimum acceptable PWM frequency.
3. Determine the maximum rate of change of input voltage for a driver circuit capable of producing 50 mA with a load consisting of a 1.5 nF gate capacitance.
4. Determine the maximum rate of change of input voltage for a driver circuit capable of producing 60 mA with a load consisting of a 3.5 nF gate capacitance.
5. A power E-MOSFET has an $R_{DS(on)}$ of 0.012 Ω and switches a 100 volt source to an 8 Ω load. Determine the maximum load current and P_D .
6. Four power E-MOSFETs drive a 4 Ω load via a full bridge network. If each device has an $R_{DS(on)}$ of 0.02 Ω and they switch ± 75 volt sources to the load, determine the load current and P_D of the devices.

UNIT 6 INSULATED GATE BIPOLAR TRANSISTORS (IGBTS)

Learning Objectives

After completing this chapter, you should be able to:

- Discuss the basic operation of the IGBT.
- Describe the internal structure of the IGBT.
- Describe the differences between PT and NPT IGBTs.
- Compare and contrast the IGBT to the power BJT.
- Compare and contrast the IGBT to the power MOSFET.
- Interpret important parameters found on an IGBT data sheet.
- Describe basic circuits using the IGBT as a power control device.

6.1 INTRODUCTION

The Insulated Gate Bipolar Transistor, or IGBT, is a power semiconductor that first became available to the commercial market during the 1980s. Initial devices had certain performance issues but these problems largely have been taken care of with subsequent generations. Today, the IGBT is in wide use, competing with power BJTs and power E-MOSFETs across a range of applications. The IGBT is designed to be used as a high voltage/high current switch and typically is not used for linear applications such as an audio class B power amplifier. The IGBT has also overtaken the older thyristor devices (e.g., SCR) in many areas due to its speed and the relative simplicity of the driving circuits used to control it.

The IGBT offers a mix of performance characteristics of both the power BJT and the power E-MOSFET. Like the BJT, the IGBT offers low on-state power loss and the ability to handle large currents and voltages. Like the power E-MOSFET, it is relatively easy to drive, being a voltage-controlled device rather than a current-controlled device. On the down-side, it is not as fast as the current generation of power E-MOSFETs and tends to be more costly than both the power BJT and power E-MOSFET. Consequently, the choice of which of these three devices should be used for a given power switching application will depend on the specifics of the design. For example, a medium to high power design that focuses on lowest cost may favor the BJT, a low to medium power application that requires very high switching speeds may be best solved with a power E-MOSFET, while an IGBT might be ideal for a very high power application utilizing low to medium speed clocking. This chapter will examine a number of power switching applications, and while they will all revolve around using the IGBT, please bear in mind that, depending on the specifics, power BJTs and E-MOSFETs might also be used.

The IGBT is available in two variants: the punch through, or PT; and the non-punch through, or NPT, versions. We shall look at both.

6.2 IGBT INTERNALS

The IGBT is a multi-layer device. The cutaway shown in Figure 6.2.1 uses an Nchannel, although P-channel is possible. This device has many features in common with the power E-MOSFET discussed in Chapter 12.

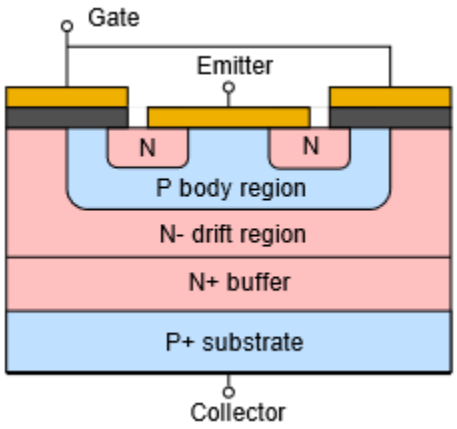


Figure 6.2.1: Cutaway view of IGBT.

To begin, note the labeling of the three terminals: gate, emitter and collector. The gate is isolated just as it is in MOSFETs. When a positive gate-emitter potential is applied, an N-type inversion layer will develop in the the P body region, allowing current to flow. The middle N region is split into two sections, the main N– drift region and the N+ buffer layer (the +/– signs indicate heavy and light doping, respectively). The device shown is the punch through (PT) variety. The non-punch through (NPT) omits the buffer layer. The functional differences between the two are that splitting the N material to include the buffer layer improves speed and lowers on-state voltage. Consequently, the PT version exhibits lower switching and conduction losses.

The upper three layers of Figure 6.2.1 form an E-MOSFET while the lower section (P body, N drift/buffer and P substrate) form a PNP transistor. Thus, we can make a simplified model of the IGBT using these other devices, as shown in Figure 3.2.2.

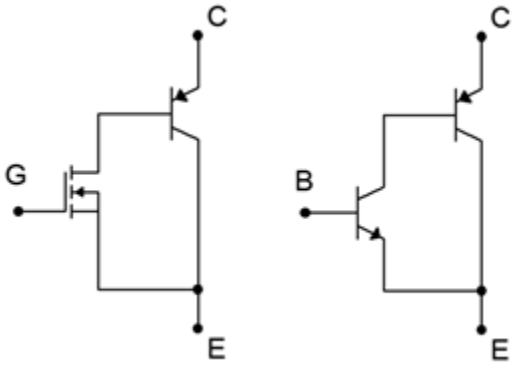


Figure 6.2.2: Simple model of an IGBT (left) compared to a Sziklai pair (right).

This simplified model is reminiscent of the NPN version of the Sziklai pair examined in Chapter 9. The input device has been replaced by an E-MOSFET. Therefore we expect the very small gate current and relatively simple drive requirements of the EMOSFET with the power handling of a BJT.

The operational device curves are, unsurprisingly, also reminiscent of these two components. A set of collector curves is presented in Figure 6.2.3 using representative values for voltage and current.

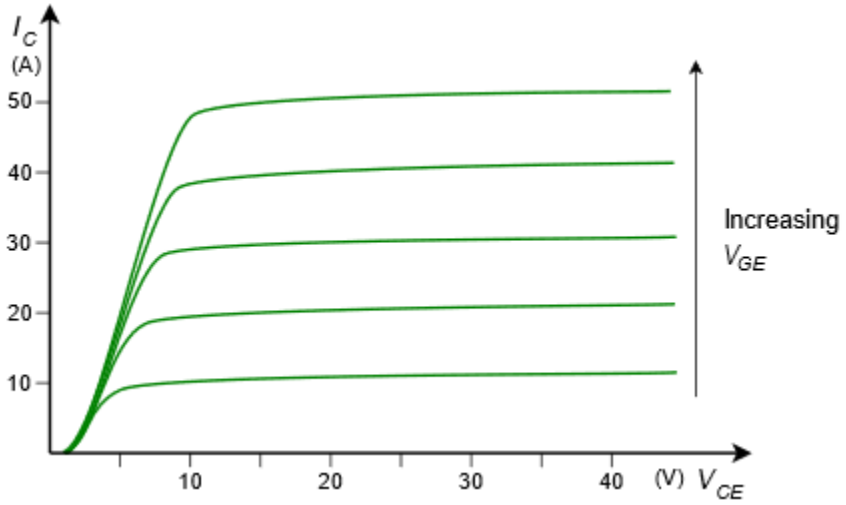


Figure 6.2.3: IGBT collector curve family.

This set of curves appears as a cross between the collector curve family seen in Chapter 4 and the drain curve family seen in Chapter 10, with one slight twist: the entire set of curves is displaced positively from the origin by about one volt. As with any E-MOSFET, channel current does not begin to flow until the gate threshold voltage is reached, here referred to as $V_{GE(th)}$. Increasing values of V_{GE} cause increased conduction and current flow. Finally, note that the curves do not begin to "flatten" until V_{GE} has reached several volts, unlike the saturation voltage of a single BJT which might be only tenths of a volt.

The forward current-voltage characteristic curve reflects the E-MOSFET portion of the model. This is shown in Figure 6.2.4 .

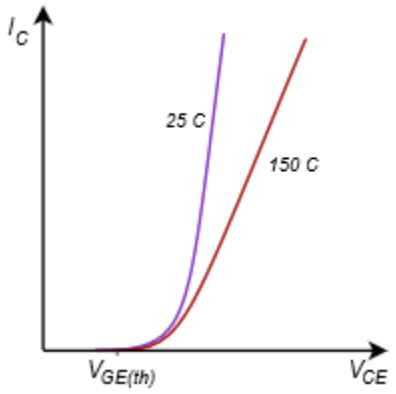


Figure 6.2.4: IGBT characteristic curve and variation with temperature.

Two traces are plotted for two different temperatures. The curves are essentially the same as the characteristic curve of the E-MOSFET, as seen in the introductory MOSFET chapter, Figure 3.4.3 (here the current axis has been expanded, making the traces appear steeper, in order to more clearly

show the variation with temperature). Conduction begins at the threshold voltage, V_{th}, and then rises rapidly, following a square law trajectory. Once a sufficient current level is reached, the curve can be approximated as a straight line.

Of particular interest here is how the curve varies with temperature. As temperature increases (red trace), the slope decreases. Recalling that the slope of the current-voltage characteristic curve is the device transconductance, this means that the transconductance decreases with temperature. In other words, the IGBT exhibits a negative temperature coefficient of transconductance, just like power E-MOSFETs, and thus are also less inclined to suffer from current hogging and thermal runaway problems than BJTs.

Unfortunately, there isn't a single, standardized schematic symbol for the IGBT. Two versions are shown in Figure 6.2.5 .

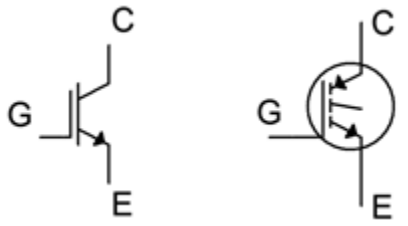


Figure 6.2.5: IGBT schematic symbols.

Both symbols attempt to reflect the dual E-MOSFET/BJT character of the device. The symbol on the left appears to have wider use currently, although there is also a third variation based on it where the gate connection is drawn toward the center rather than closer to the emitter.

6.3 IGBT DATA SHEET INTERPRETATION

A portion of the data sheet for the Fairchild/ON Semiconductor FGH50T65SQD IGBT is shown in Figure 6.3.1 . This is a fourth generation IGBT featuring trench construction. It is rated for 650 volts and 50 amps. The device includes an antiparallel diode. This is useful for bridge applications that drive inductive loads (recall that the current through an inductor cannot change instantaneously, thus, when devices are switched on/off in a bridge, the diode serves to create a path around the IGBT for this current).

FAIRCHILD

April 2016

FGH50T65SQD

650 V, 50 A Field Stop Trench IGBT

Features

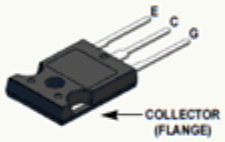
- Maximum Junction Temperature : $T_J = 175^oC$
- Positive Temperaure Co-efficient for Easy Parallel Operating
- High Current Capability
- Low Saturation Voltage: $V_{CE(sat)} = 1.6$ V(Typ.) @ $I_C = 50$ A
- 100% of the Parts Tested for $I_{LM}(1)$
- High Input Impedance
- Fast Switching
- Tighten Parameter Distribution
- RoHS Compliant

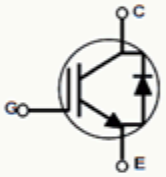
General Description

Using novel field stop IGBT technology, Fairchild's new series of field stop 4th generation IGBTs offer the optimum performance for solar inverter, UPS, welder, telecom, ESS and PFC applications where low conduction and switching losses are essential.

Applications

- Solar Inverter, UPS, Welder, Telecom, ESS, PFC





Absolute Maximum Ratings

Symbol	Description		FGH50T65SQD_F155	Unit
V_{CES}	Collector to Emitter Voltage		650	V
V_{GES}	Gate to Emitter Voltage		±20	V
	Transient Gate to Emitter Voltage		± 30	V
I_C	Collector Current	@ $T_C = 25^oC$	100	A
	Collector Current	@ $T_C = 100^oC$	50	A
$I_{LM\ (1)}$	Pulsed Collector Current	@ $T_C = 25^oC$	200	A
$I_{CM\ (2)}$	Pulsed Collector Current		200	A
I_F	Diode Forward Current	@ $T_C = 25^oC$	50	A
	Diode Forward Current	@ $T_C = 100^oC$	30	A
I_{FM}	Pulsed Diode Maximum Forward Current		200	A
P_D	Maximum Power Dissipation	@ $T_C = 25^oC$	268	W
	Maximum Power Dissipation	@ $T_C = 100^oC$	134	W
T_J	Operating Junction Temperature		-55 to +175	^oC
T_{stg}	Storage Temperature Range		-55 to +175	^oC
T_L	Maximum Lead Temp. for soldering Purposes, 1/8" from case for 5 seconds		300	^oC

Notes:
1. $V_{CC} = 400$ V, $V_{GE} = 15$ V, $I_C = 200$ A, $R_G = 3\ \Omega$, Inductive Load
2. Repetitive rating: Pulse width limited by max. junction temperature

Figure 6.3.1 ⟐: FGH50T65SQD data sheet. Used with permission from SCILLC dba ON Semiconductor.

Thermal Characteristics

Symbol	Parameter	FGH50T65SQD_F155	Unit
$R_{\theta JC}$(IGBT)	Thermal Resistance, Junction to Case, Max.	0.56	°C/W
$R_{\theta JC}$(Diode)	Thermal Resistance, Junction to Case, Max.	1.25	°C/W
$R_{\theta JA}$	Thermal Resistance, Junction to Ambient, Max.	40	°C/W

Package Marking and Ordering Information

Part Number	Top Mark	Package	Packing Method	Reel Size	Tape Width	Qty per Tube
FGH50T65SQD_F155	FGH50T65SQD	TO-247 G03	Tube	-	-	30

Electrical Characteristics of the IGBT T_C = 25°C unless otherwise noted

Symbol	Parameter	Test Conditions	Min.	Typ.	Max.	Unit
Off Characteristics						
BV_{CES}	Collector to Emitter Breakdown Voltage	V_{GE} = 0V, I_C = 1 mA	650	-	-	V
ΔBV_{CES} / ΔT_J	Temperature Coefficient of Breakdown Voltage	I_C = 1 mA, Reference to 25°C	-	0.6	-	V/°C
I_{CES}	Collector Cut-Off Current	V_{CE} = V_{CES}, V_{GE} = 0 V	-	-	250	µA
I_{GES}	G-E Leakage Current	V_{GE} = V_{GES}, V_{CE} = 0 V	-	-	±400	nA
On Characteristics						
$V_{GE(th)}$	G-E Threshold Voltage	I_C = 50 mA, V_{CE} = V_{GE}	2.6	4.5	6.4	V
$V_{CE(sat)}$	Collector to Emitter Saturation Voltage	I_C = 50 A, V_{GE} = 15 V, T_C = 25°C	-	1.6	2.1	V
		I_C = 50 A, V_{GE} = 15 V, T_C = 175°C	-	1.92	-	V
Dynamic Characteristics						
C_{ies}	Input Capacitance	V_{CE} = 30 V, V_{GE} = 0 V, f = 1MHz	-	3275	-	pF
C_{oes}	Output Capacitance		-	84	-	pF
C_{res}	Reverse Transfer Capacitance		-	12	-	pF
Switching Characteristics						
$t_{d(on)}$	Turn-On Delay Time	V_{CC} = 400 V, I_C = 12.5 A, R_G = 4.7 Ω, V_{GE} = 15 V, Inductive Load, T_C = 25°C	-	22	-	ns
t_r	Rise Time		-	8.7	-	ns
$t_{d(off)}$	Turn-Off Delay Time		-	105	-	ns
t_f	Fall Time		-	2.5	-	ns
E_{on}	Turn-On Switching Loss		-	180	-	uJ
E_{off}	Turn-Off Switching Loss		-	45	-	uJ
E_{ts}	Total Switching Loss		-	225	-	uJ
$t_{d(on)}$	Turn-On Delay Time	V_{CC} = 400 V, I_C = 25 A, R_G = 4.7 Ω, V_{GE} = 15 V, Inductive Load, T_C = 25°C	-	19	-	ns
t_r	Rise Time		-	13	-	ns
$t_{d(off)}$	Turn-Off Delay Time		-	93	-	ns
t_f	Fall Time		-	6.4	-	ns
E_{on}	Turn-On Switching Loss		-	410	-	uJ
E_{off}	Turn-Off Switching Loss		-	88	-	uJ
E_{ts}	Total Switching Loss		-	498	-	uJ

Figure 6.3.1◈: FGH50T65SQD data sheet (cont).

The main features are the 650 volt rating for ◈◈◈ , 100 amp continuous collector current and 268 watt dissipation at 25°C. The current and power ratings are essentially halved at the more practical temperature of 100°C. The threshold voltage, ◈◈◈(◈*h*) , is specified as 4.5 volts with a ± 1.9 volt spread. The saturation voltage typically is 1.6 volts at room temperature with 50 amps of collector current. This compares favorably to basic power BJTs. Like power MOSFETs, the input capacitance is relatively high at 3275 pF, so the same gate drive precautions must be followed. Finally, note the

asymmetry in switching times. At room temperature and 12.5 amps of collector current, the turn-on delay plus rise time is specified as approximately 31 nanoseconds while the turn-off delay and fall time is nearly 110 nanoseconds. This relative slowing of the off-state transition is typical of IGBTs. Further, as both current and temperature increase (Figure 6.3.1◈), these times increase by a few percent. By comparison, the FDMS86180 power E-MOSFET examined in Chapter 12 exhibited symmetrical values in the mid-30 nanosecond region at a drain current of 67 amps.

Electrical Characteristics of the IGBT (Continued)

Symbol	Parameter	Test Conditions	Min.	Typ.	Max	Unit
$t_{d(on)}$	Turn-On Delay Time	V_{CC} = 400 V, I_C = 12.5 A, R_G = 4.7 Ω, V_{GE} = 15 V, Inductive Load, T_C = 175°C	-	20	-	ns
t_r	Rise Time		-	9.8	-	ns
$t_{d(off)}$	Turn-Off Delay Time		-	116	-	ns
t_f	Fall Time		-	3.5	-	ns
E_{on}	Turn-On Switching Loss		-	402	-	uJ
E_{off}	Turn-Off Switching Loss		-	110	-	uJ
E_{ts}	Total Switching Loss		-	512	-	uJ
$t_{d(on)}$	Turn-On Delay Time	V_{CC} = 400 V, I_C = 25 A, R_G = 4.7 Ω, V_{GE} = 15 V, Inductive Load, T_C = 175°C	-	18	-	ns
t_r	Rise Time		-	15	-	ns
$t_{d(off)}$	Turn-Off Delay Time		-	102	-	ns
t_f	Fall Time		-	8	-	ns
E_{on}	Turn-On Switching Loss		-	641	-	uJ
E_{off}	Turn-Off Switching Loss		-	203	-	uJ
E_{ts}	Total Switching Loss		-	844	-	uJ
Q_g	Total Gate Charge	V_{CE} = 400 V, I_C = 50 A, V_{GE} = 15 V	-	99	-	nC
Q_{ge}	Gate to Emitter Charge		-	17	-	nC
Q_{gc}	Gate to Collector Charge		-	23	-	nC

Electrical Characteristics of the Diode T_C = 25°C unless otherwise noted

Symbol	Parameter	Test Conditions		Min.	Typ.	Max	Unit
V_{FM}	Diode Forward Voltage	I_F = 30 A	T_C = 25°C	-	2.2	2.6	V
			T_C = 175°C	-	1.9	-	
E_{rec}	Reverse Recovery Energy	I_F =30 A, dI_F/dt = 200 A/µs	T_C = 175°C	-	40	-	uJ
t_{rr}	Diode Reverse Recovery Time		T_C = 25°C	-	31	-	ns
			T_C = 175°C	-	207	-	
Q_{rr}	Diode Reverse Recovery Charge		T_C = 25°C	-	48	-	nC
			T_C = 175°C	-	820	-	

Figure 6.3.1◈: FGH50T65SQD data sheet (cont).

Finally, consider the graphical data presented in Figure 6.3.1◈ .

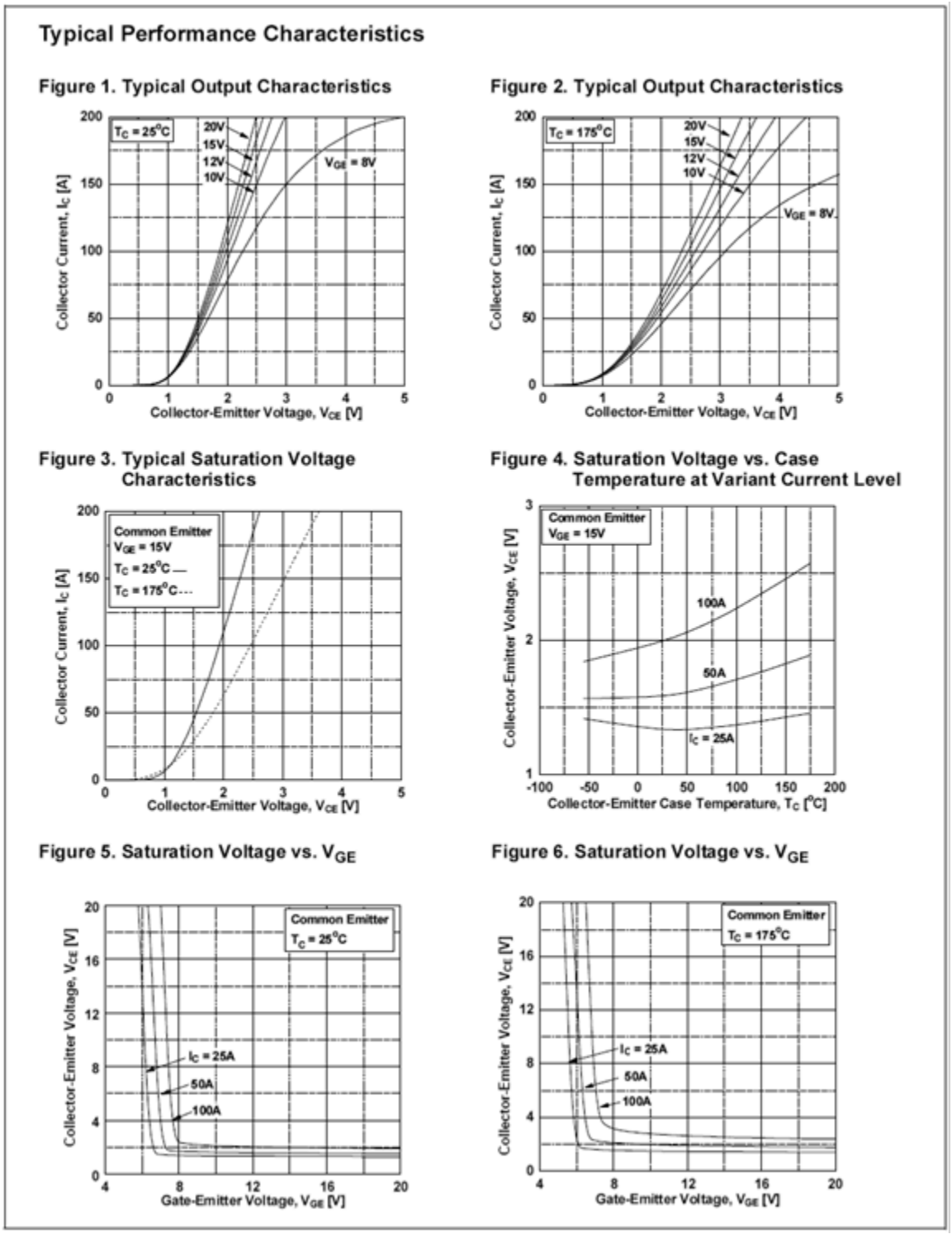


Figure 6.3.1 ⟐: FGH50T65SQD data sheet (cont).

The top two graphs show the collector curves in the low voltage region. Compare these with Figure 6.2.3. Similarly, the middle-left graph shows the current-voltage characteristic with temperature variation. This reflects Figure 6.2.4. Finally, the bottom two graphs show the collector-emitter saturation voltage with respect to gateemitter drive voltage for three different collector currents. Note that for the highest current at room temperature, saturation voltage is around 2 volts for a gate drive of at least 8 volts. This rises to about 3 volts at 175°C.

COMPUTER SIMULATION

To highlight the performance of the IGBT, a simple series load switch is simulated. The circuit is shown in Figure 6.3.2 .

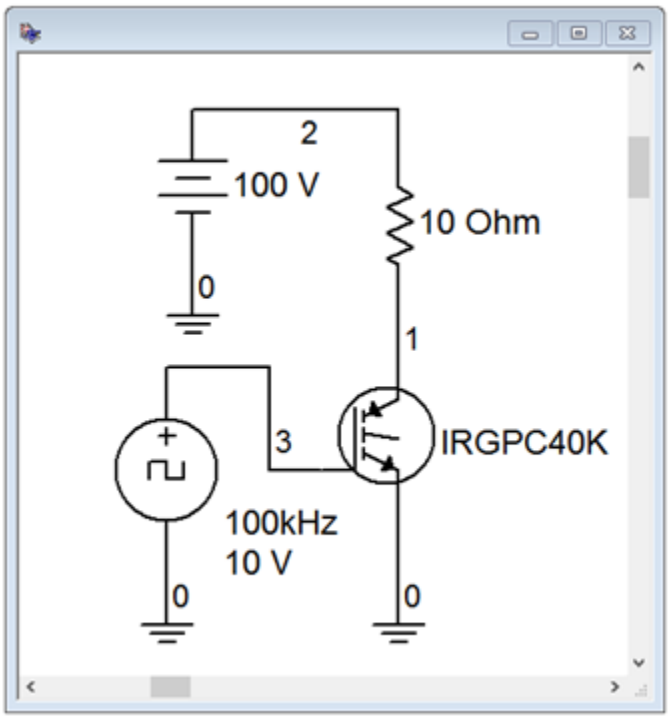


Figure 6.3.2: IGBT simulation schematic.

A 10 ohm load is switched from a 100 volt DC power supply via an International Rectifier IRGPC40K IGBT. The gate is driven from a 10 volt peak square wave running at 100 kHz.

Ideally, if the IGBT produced no losses, the full 100 volt source would drop across the load, producing 10 amp current pulses. According to the device data sheet, collector-emitter saturation voltage typically is 2.1 volts (3.2 volts maximum) with a 25 amp collector current. The total turn-on/rise time and turn-off/fall time values are 62 nanoseconds and 290 nanoseconds typically, at 25 amps and room temperature.

A transient analysis is run, with the results shown in Figure 6.3.3 .

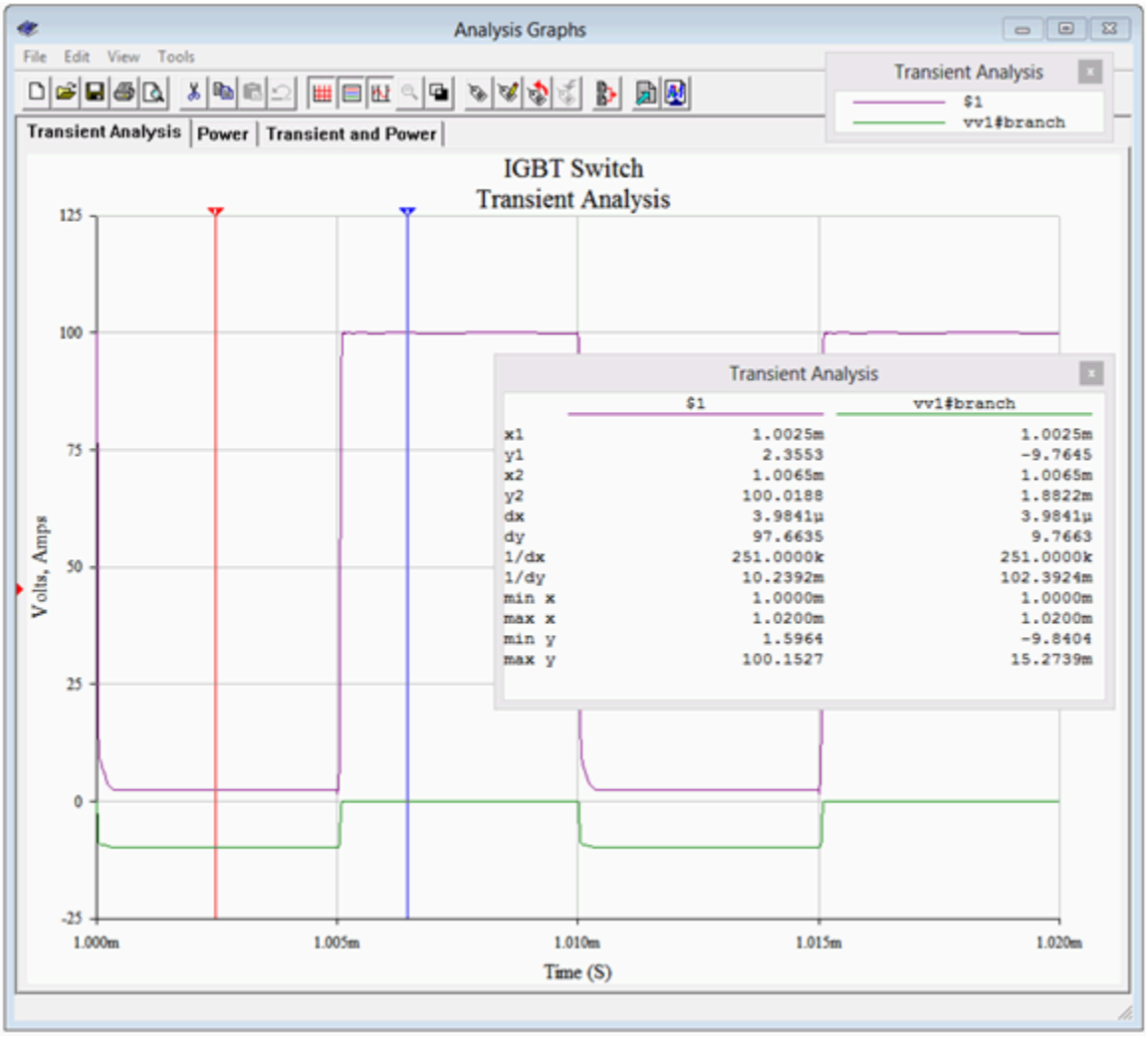


Figure 6.3.3: Transient analysis of simple IGBT switch.

The drain voltage (node 1) is shown as the purple trace and the drain current is shown in green (reversed in polarity to see it more clearly). The two cursors (red is #1, blue is #2) show the various levels. As expected, the drain voltage peaks at 100 volts in the off-state, at which time the current is nil. During the on-state, the drain voltage drops to about 2.36 volts, very close to the data sheet's typical saturation potential. This small drop reduces the voltage across the load to about 97.64 volts. This is verified by cursor #1 showing a current of 9.765 amps. Also, although it is not possible to determine the edge timings with great precision from this plot, the asymmetry between rise and fall is apparent, and the edges are generally consistent with the numeric values from the data sheet.

Figure 6.3.4 : shows the result of multiplying the current and voltage waveforms. This new waveform represents the power dissipation of the IGBT.

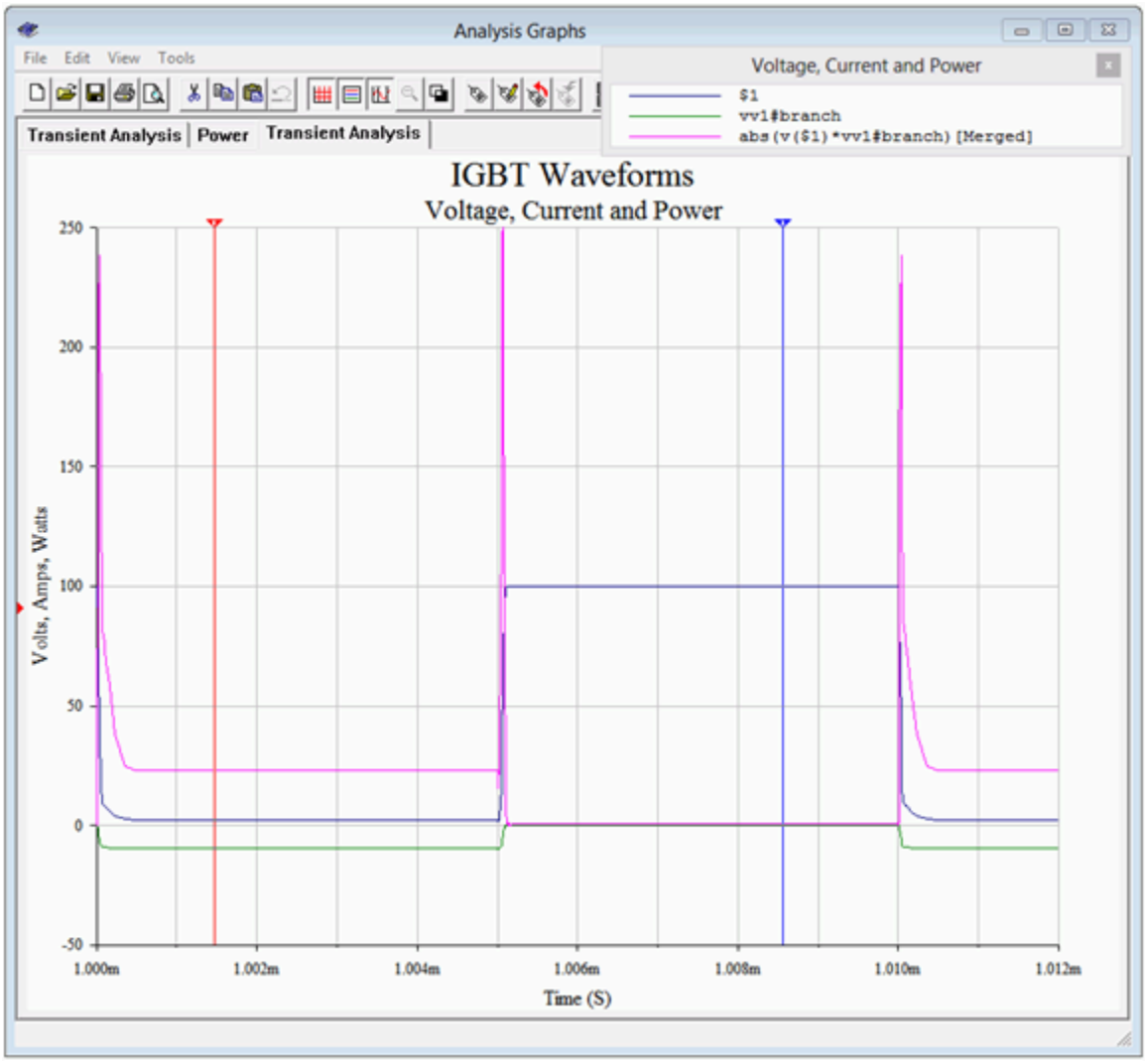
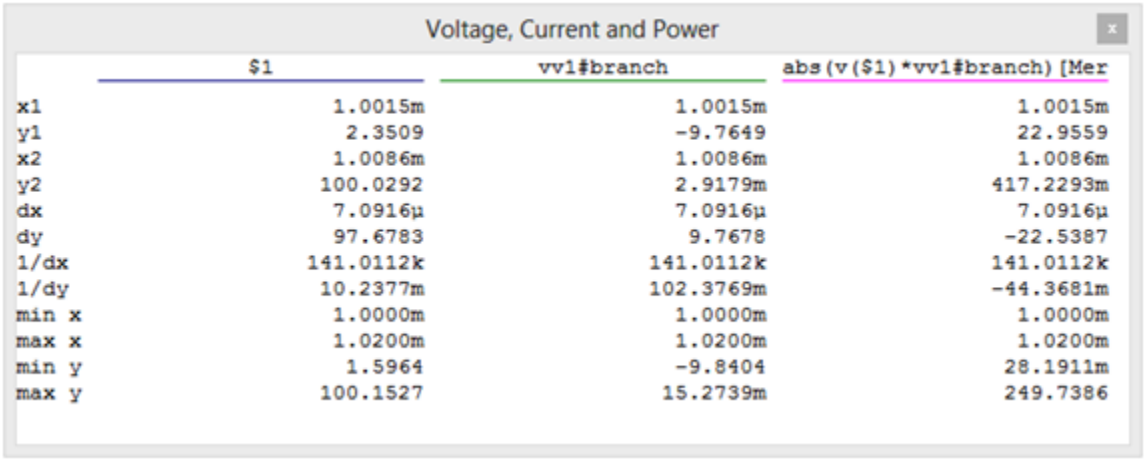


Figure 6.3.4: Power waveform of simple IGBT switch.

The power trace is shown in fuchsia, or magenta, or screaming purple-pink, or whatever-you-want-to-call-it. The edge spikes dominate but the on-state power is apparent as well. Remember, during the on-state, the load is dissipating close to 1000 watts. The cursor output window is shown separately in Figure 6.3.5 . The on-state power dissipation is approximately 23 watts which represents less than 2.5 percent of the load power. In contrast, the edge spikes are maxing out at around 250 watts. Of course, the time duration is very short, being only a few percent of the cycle period, but it cannot be ignored.



	$1	vv1#branch	abs(v($1)*vv1#branch) [Mer
x1	1.0015m	1.0015m	1.0015m
y1	2.3509	-9.7649	22.9559
x2	1.0086m	1.0086m	1.0086m
y2	100.0292	2.9179m	417.2293m
dx	7.0916µ	7.0916µ	7.0916µ
dy	97.6783	9.7678	-22.5387
1/dx	141.0112k	141.0112k	141.0112k
1/dy	10.2377m	102.3769m	-44.3681m
min x	1.0000m	1.0000m	1.0000m
max x	1.0200m	1.0200m	1.0200m
min y	1.5964	-9.8404	28.1911m
max y	100.1527	15.2739m	249.7386

Figure 6.3.5 : Numeric values at waveform cursors for the IGBT switch.

6.4 IGBT APPLICATIONS

IGBTs lend themselves to a variety of high power switching applications. In this section, we shall look at four of them. Bear in mind that power BJTs and power EMOSFETs might also be used for these applications, depending on the specifics of the design. In general, power E-MOSFETs will be preferred when using high switching frequencies at medium to low powers and voltages, while IGBTs are favored at higher voltages, currents and powers.

INDUCTION HEATING

Although thermal conduction is the first method commonly thought of when it comes to heating something, magnetic induction may also be used. Magnetic induction creates heat through the Joule effect and can be used for large scale industrial processes, such as creating metal alloys via an induction furnace, to much smaller scale consumer applications, such as an inductive cooktop. Induction heating is efficient because the vessel itself is heated directly and less heat is lost to the immediate environment. Also, control of heating can be very precise. The basic idea is to create a rapidly changing magnetic field placed next to the container to be heated. If this vessel is ferromagnetic, eddy currents will be induced in the vessel, creating heat. Thus, if we were to place a cast iron pot within the field, the pot itself would heat up as eddy currents are induced within it, thus heating the pot's contents. There are no open flames or surface heating elements involved. The only downside to this process is that the vessel must be made of ferromagnetic material. For a cooktop, that means that pots and pans must be made of iron or certain steel alloys. An aluminum sauté pan or ceramic container will not work with this system.

As an example, let's consider an inductive cooktop. A sophisticated design could feature a full- or half-bridge arrangement of IGBTs, but for illustrative purposes we'll focus on a simple single-ended system using just one IGBT.

The system features four main components: the rectifier/EMI filter, the control/drive circuit, the IGBT switch and the ⧀⧁ resonant tank sub-circuit that generates the field. This is illustrated in Figure 6.4.1 .

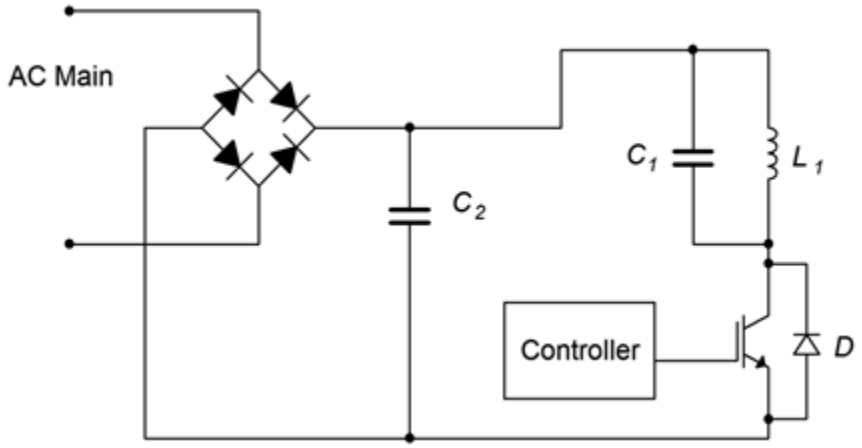


Figure 6.4.1 : Simple inductive heater.

The rectifier produces full-wave pulsating DC, and the associated capacitor, ◈2 , is used to help minimize EMI (electromagnetic interference) and also provide a return path for the tank. The control circuit produces a variable duty cycle pulse train to drive the gate of the IGBT. The greater the duty cycle, the longer the on-state of the IGBT, and ultimately, the greater the heating. Between the IGBT and the rectified power signal is a parallel resonant tank circuit comprised of ◈1 and ◈1 . The inductor is comprised of a series of loops of large gauge wiring or copper tubing embedded in the cook surface, typically under a glass or ceramic top. The resonant frequency of the tank is tuned to the frequency of the controlling pulse train. This will maximize the tank current and thus produce a more powerful magnetic field. The switching frequency is usually placed just above the range of human hearing to avoid audible microphonics. Values in the range of 20 kHz to 30 kHz are typical, and the base frequency may change as the heat demand changes. For example, to minimize switching losses, the controlling frequency might start at 30 kHz for modest heating and decrease to 20 kHz for maximum heating.

From the cook's perspective there is no change between using the inductive cooktop and an ordinary electric cooktop using resistive heating elements: The cook places the pot or pan on the surface, under which lies the coil. A heat level control knob is provided for them to adjust the heat intensity. To their advantage, when they remove the pot or pan, the cooking surface itself will not be as hot as an ordinary cooktop.

From the designer's perspective, the heat control knob simply changes the duty cycle of the controlling pulse train (and optionally, its frequency, as mentioned previously). Other refinements might include sensing whether or not a vessel is on the cooktop and throttling back control if nothing is detected. Finally, an even simpler system could switch the IGBT on and off at a much slower rate (think in terms of seconds) to greatly reduce switching losses but this runs the risk of heat cycling if the pots and pans used are of very light gauge construction (i.e., their thermal time constant will be faster).

DC-TO-AC INVERSION

There are many instances where we wish to derive an AC voltage from an existing DC voltage. Examples include an uninterruptible power supply (UPS) that would draw current from a battery and deliver standard AC power when there is a disruption in the power grid, and the need to operate

electronic devices designed for the home in a remote location. This process is known as DC-to-AC inversion.

The simplest method to create AC from DC is to just "chop" the DC at the desired line frequency and then scale it, that is, feed the DC into a simple IGBT switch which will produce a square wave and then feed the square wave into a transformer to arrive at the desired voltage. The obvious problem with this technique is that the AC signal will not be a nice, smooth sine wave, but rather, a distorted square wave. Unfortunately, for many electronic components this will present a challenge to their power supply circuits. A possible refinement involves making a step-wise approximation of a sine wave but this is still not ideal.

A more accurate scheme involves synthesizing a sine wave via PWM. We still chop the DC, but now the action is performed at a higher frequency and with a varying duty cycle such that, when the output is averaged, we arrive at a sine wave. A block diagram of this scheme is shown in Figure 6.4.2 .

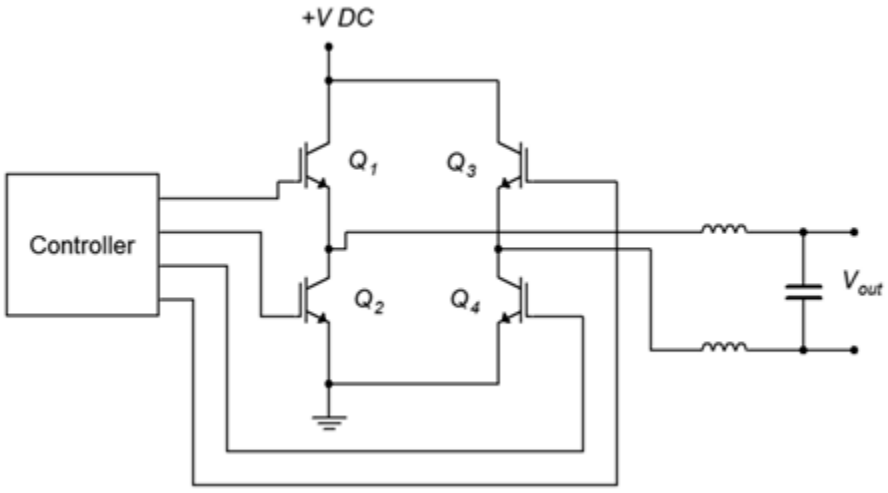


Figure 6.4.2 : DC-to-AC inverter (antiparallel diodes not shown).

The controller generates a set of PWM signals to drive the four IGBTs configured as a full-bridge. The bridge output is then fed to a balanced LC filter that removes the high frequency PWM components, leaving a smoothed sine wave. This signal can be used as is, or fed into a step-up transformer if the desired AC voltage is greater than the starting DC voltage. An example of this would be the need to supply a device designed to run on 120 VAC from the nominal 12 VDC system found in a car.

There is a side item worthy of mention here. Some devices derive timing signals from the line frequency (a classic example is an electronic alarm clock/radio). This is possible because the power generation utility monitors this frequency with great accuracy. If the controller shown in Figure 6.4.2 does not produce an accurate base frequency, then that clock/radio will not tell time accurately.

MOTOR CONTROL

IGBTs can be used to control the speed of electric motors. The configuration of the control circuit will depend on the kind of motor being controlled. In simplest terms, the speed of a DC motor is controlled by the voltage applied to it: the higher the voltage, the higher the speed. In contrast, the speed of an AC motor depends on the frequency of the applied source (they are proportional).

Controlling a DC motor is a straightforward situation. If all we need to do is start and stop the motor, the IGBT can be inserted in series with the motor and used as a switch to open and close

the circuit. Being solid-state, the IGBT has numerous advantages over a mechanical switch or relay including long-term reliability and simplicity of the control circuit. To alter the speed, the IGBT can be controlled via PWM. This is illustrated in Figure 6.4.3 .

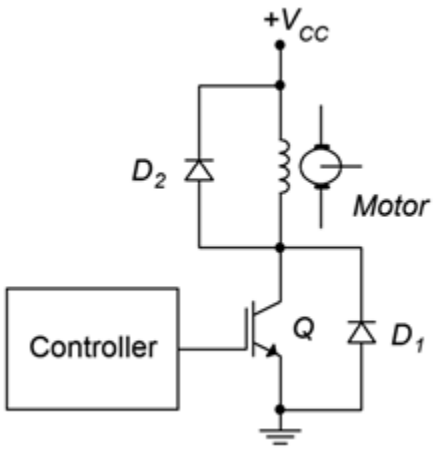


Figure 6.4.3 : DC motor control via IGBT.

◈1 and ◈2 are flyback or snubber diodes used for protection from inductive current spikes caused by the motor's current being switched on and off. As noted earlier, some IGBTs are co-packaged with an anti-parallel diode (◈1).

To vary the motor's speed, the controller produces a PWM drive signal. The smaller the duty cycle of this pulse train, the lower the average applied voltage to the motor, and therefore the slower its speed. The base frequency of the PWM signal does not have to be particularly high in this scenario; a few hundred hertz may prove sufficient. This will help minimize switching losses.

For an AC motor the situation is a little more complex. One approach is to use the PWM technique explained under the DC-to-AC converter section. The difference is that the power source would not be DC, but rather, AC. Consequently, we would need to transform the AC power source into a more usable signal and then apply the circuit depicted in Figure 6.4.2 to power the motor. The controller itself will need to be considerably more sophisticated. In Figure 6.4.2, the duty cycle is continuously changed such that the "area under the curve" approximates a sine wave. Eventually, the pattern will repeat itself for subsequent cycles of the sine wave. In other words, the rate at which the pattern repeats is the sine wave's period. In the DC-to-AC inverter application, this rate never changes because we need a constant output frequency (e.g., 60 Hz). In the AC motor control application, such is not the case. This repetition rate needs to be adjustable because that's what controls the motor's speed. One way to do this is to simply increase the base frequency of the PWM pulse train. This method is simple and direct but has the disadvantage of creating more transient edges per unit time and therefore tends to increase switching losses. An alternate approach is to keep the base frequency constant and instead alter the duty cycle pattern. This helps minimize the switching loss issues but has the disadvantage of requiring a more complex control circuit and possibly producing a lower quality sine wave at higher output frequencies.

DC-TO-DC CONVERSION

Our final application is DC-to-DC up-conversion, that is, producing a new DC voltage that is higher than the original source and also capable of high output current (voltage doublers and triplers can be

made from diode/capacitor lattices but they are not designed to deliver high, continuous currents). Applications requiring up-conversion include photovoltaic systems (i.e, combining the outputs of several solar panels and tying them into the power system) and high output car audio systems. Specifically, the nominal 12 volt automotive power system is insufficient to supply an amplifier intended to deliver hundreds or even thousands of watts to a subwoofer. The 12 volt source will need to be increased, perhaps by a factor of ten, to achieve the desired output levels.[1]

As we have already seen how DC can be translated into AC via an inverter, it is possible to simply rectify and filter the new AC, producing a higher DC level. This scheme is illustrated in Figure 6.4.4 .

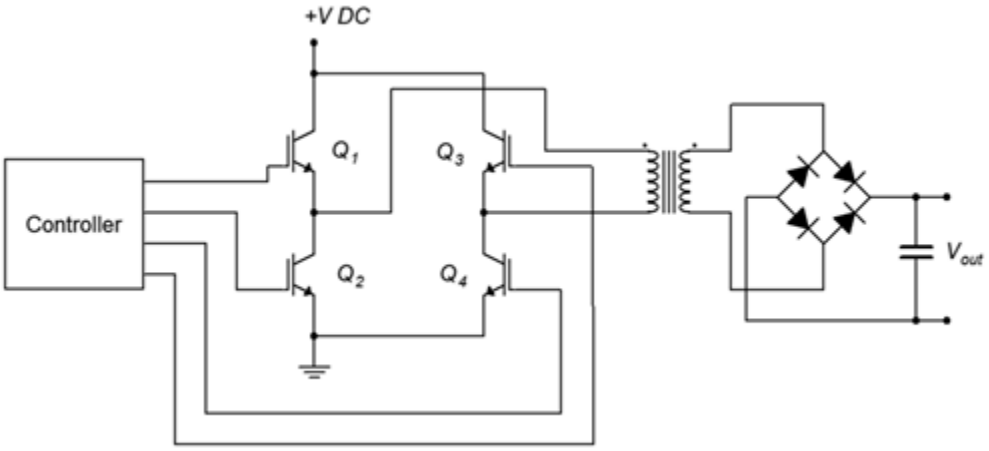


Figure 6.4.4 : DC-to-DC conversion via transformer (anti-parallel diodes not shown).

The transformer pictured here will need to be a step-up variety in order to achieve the desired output voltage level. Note that the AC generation side does not have to produce a particularly nice sine wave nor does it have to be at the usual line frequency. In fact, increasing the frequency will likely result in reduced sizes for the transformer and filter capacitors.

A completely different approach is to use a switching regulator. Switching regulators use a feedback control system to generate a very stable output voltage by comparing it to a reference voltage. They can be configured in step-down, step-up or polarity inversion forms.[2] In this case, we can use the step-up, or boost, form. An example is shown in Figure 6.4.5.

1. In the process, the current demand will be increased greatly as well, perhaps beyond the capabilities of the vehicle's alternator (which will also require upgrading), but these are the prices one must pay if one desires very high sound pressure levels in what is arguably the worst acoustical environment in which to listen to music. Of course, we should also admit that the act of critically listening to and enjoying music may not be the point of such an exercise.
2. For details on switching regulators, see Fiore, J, Operational Amplifiers and Linear Integrated Circuits: Theory and Application, another free OER text.

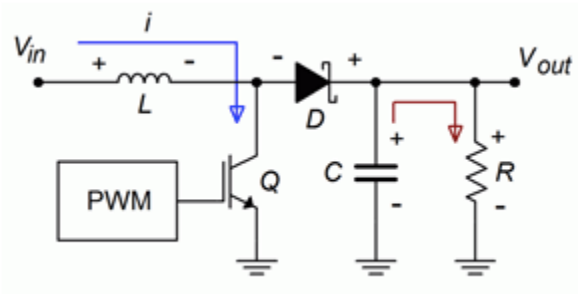


Figure 6.4.5 : Step-up switching regulator, on-state.

As with the other applications presented so far, the IGBT is being used as a switch. Figure 6.4.5 illustrates the on-state of the IGBT. During this phase, current is drawn through the inductor, , storing energy in the associated magnetic field. The reverse-biased Schottky diode, , isolates this section from the output section, where is delivering the load voltage and current.

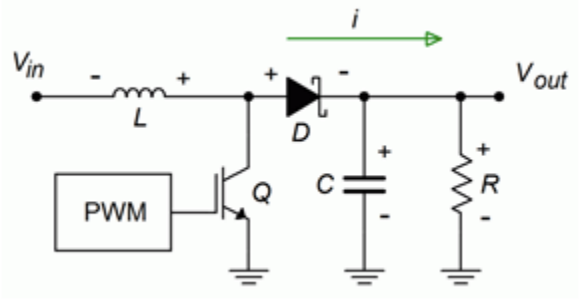


Figure 6.4.5 : Step-up switching regulator, off-state.

The off-state is depicted in Figure 6.4.5 . During this portion, the inductor discharges and appears as a source. As it is in series with the input voltage, the voltage at the output will equal the inductor's voltage plus the input voltage. Also, during this phase the capacitor is recharged, ready for the next on-state of the IGBT. Note that a Schottky diode is used here because it exhibits fast switching times and a low forward voltage drop. Switching frequencies for these circuits tend to be high (100 kHz and up are common) as that minimizes the sizes of the inductor and capacitor.

6.5 SUMMARY

The Insulated Gate Bipolar Transistor, or IGBT, can be thought of as a combination of a power BJT and a power E-MOSFET. As such, it combines the low on-state conduction losses of the BJT with the relatively easy drive requirements of the EMOSFET. The IGBT is available in two variants; the PT, or punch through, and the NPT, or non-punch through types. The PT type includes an N+ buffer layer in its construction and this endows the device with faster switching speed and lower onstate losses.

The IGBT's characteristic curves tend to echo that of the E-MOSFET. Conduction does not begin until the gate-emitter voltage exceeds a threshold voltage, ◇◇◇(◇*h*) . From there, the current-voltage characteristic follows a square-law trajectory, and at sufficiently high current levels it can be approximated as a straight line. The IGBT exhibits a negative temperature coefficient of transconductance, like the MOSFET, making it less prone to thermal runaway and current hogging issues. A family of collector curves (i.e., ◇◇◇ vs. ◇◇) shares attributes with BJT collector curves and MOSFET drain curves. The curves echo the same overall shape, starting with a section where current rises rapidly compared to voltage, and then leveling out into a constant current region. The initial region of rapid change is somewhat drawn out as it is in the MOSFET. Also, the entire set of curves is displaced positively by about a volt, rather than current increasing immediately from the origin.

In general, the IGBT offers higher voltage, current and power capability than the power E-MOSFET although it lags behind in switching speed. Further, switching times for the on- and off-state are asymmetrical. Compared to the power BJT, the IGBT tends to be more expensive. Consequently, power E-MOSFETs tend to be favored at low and moderate power levels when high switching speeds are needed and BJTs tend to be preferred when cost is a major component in more modest designs. As such, IGBTs find use across a range of applications including power inverters, uninterruptible power supplies, induction heaters, solar power systems, motor controllers and so forth.

Review Questions

1. What are the advantages and disadvantages of IGBTs compared to power BJTs?
2. What are the advantages and disadvantages of IGBTs compared to power E-MOSFETs?
3. What are the differences between NPT and PT IGBTs
4. Compare the simple IGBT model to that of a Sziklai pair.
5. Explain the basic operation of an induction heater and how an IGBT is used to control the generation of heat.
6. Explain how pulse width modulation might be used to control the speed of a DC motor via an IGBT.

7. Explain how an IGBT can be used to translate a DC power source into an AC power source.

6.6 EXERCISES

ANALYSIS PROBLEMS

1. From the FGH50T65SQD data sheet, determine the collector-emitter saturation voltage at 25°C for V_{GE} = 7 volts and I_C = 50 amps.
2. From the FGH50T65SQD data sheet, determine the change in collector-emitter saturation voltage from 25°C to 175°C for V_{GE} = 8 volts and I_C = 100 amps.
3. From the FGH50T65SQD data sheet, determine the rise time for a 50 amp collector current at 25°C for V_{GE} = 15 volts.
4. From the IRGPC40K data sheet, determine the combined turn-on and rise time at 150°C.
5. From the IRGPC40K data sheet, determine the combined turn-off and fall time at 150°C.

COMPUTER SIMULATION PROBLEMS

7. Repeat the simple switch simulation from this chapter using the FGH50T65SQD IGBT and compare the resulting voltage, current and power waveforms.
8. Repeat the simple switch simulation from this chapter using the FDMS86180 power E-MOSFET from Chapter 3 and compare the resulting voltage, current and power waveforms.

UNIT 7 - INTRODUCTORY CONCEPTS AND FUNDAMENTALS

Learning Objectives

After completing this chapter, you should be able to:

- Convert between ordinary and decibel based power and voltage gains.
- Utilize decibel-based voltage and power measurements during circuit analysis.
- Define and graph a general Bode plot.
- Detail the differences between lead and lag networks, and graph Bode plots for each.
- Combine the effects of several lead and lag networks together in order to determine a system Bode plot.
- Describe the use of digital computers in the area of circuit simulation.
- Analyze differential amplifiers for a variety of AC characteristics, including single ended and differential voltage gains.
- Define *common-mode gain* and *common-mode rejection*.
- Describe a current mirror and note typical uses for it.

INTRO TO OPERATIONAL AMPLIFIERS

Before we can begin our study of the operational amplifier, it is very important that certain background elements be in place. The purpose of this chapter is to present the very useful analysis concepts and tools associated with the decibel measurement scheme and the frequency domain. We will also be examining the differential amplifier that serves as the heart of most operational amplifiers. With a thorough working knowledge of these items, you will find that circuit design and analysis will proceed at a much quicker and more efficient pace. Consider this chapter as an investment in time, and treat it appropriately.

The decibel measurement scheme is in very wide use, particularly in the field of communications. We will be examining its advantages over the ordinary system of measurement, and how to convert values of one form into the other. One of the more important parameters of a circuit is its frequency response. To this end, we will be looking at the general frequency domain representations of a circuit's gain and phase. This will include both manual and computer generated analysis and graphing techniques. While our main emphasis will eventually concentrate on application with operational amplifiers, the techniques explored can be applied equally to discrete circuits. Indeed, our initial examples will use simple discrete or black box circuits exclusively.

Finally, we will examine the DC and AC operation of the differential amplifier. This is an amplifier that utilizes two active devices and offers dual inputs. It offers certain features that make it suitable as the first section of most operational amplifiers.

VARIABLE NAMING CONVENTION

One item that often confuses students of almost any subject is nomenclature. It is important, then, that we decide upon a consistent naming convention at the outset. Throughout this text, we will be examining numerous circuits containing several passive and active components. We will be interested in a variety of parameters and signals. Although we will utilize the standard conventions, such as fc for critical frequency and Xc for capacitive reactance, a great number of other possibilities exist. In order to keep confusion to a minimum, we will use the following conventions in our equations for naming devices and signals that haven't been standardized.

◈ Resistor (DC, or actual circuit component)
◈ (AC equivalent, where phase is 0 or ignored)
◈ Capacitor
◈ Inductor
◈ Transistor (Bipolar or FET)
◈ Diode
◈ Voltage (DC)
◈ Voltage (AC)
◈ Current (DC)
◈ **Current (AC)**

Table 7.1.1

Resistors, capacitors and inductors are differentiated via a subscript that usually refers to the active device it is connected to. For example, R_E is a DC bias resistor connected to the emitter of a transistor, while r_C refers to the AC equivalent resistance seen at a transistor's collector. C_E refers to a capacitor connected to a transistor's emitter lead (most likely a bypass or coupling capacitor). Note that the device related subscripts are always shown in upper case, with one exception: If the resistance or capacitance is part of the device model, the subscript will be shown in lower case to distinguish it from the external circuit components. For example, the AC dynamic resistance of a diode would be called rd. If no active devices are present, or if several items exist in the circuit, a simple numbering scheme is used, such as R_1. In very complex circuits, a specific name will be given to particularly important components, as in R_{source}.

Voltages are normally given a two-letter subscript indicating the nodes at which it is measured. V_{CE} is the DC potential from the collector to the emitter of a transistor, while v_{BE} indicates the AC signal appearing across a transistor's base-emitter junction. A single-letter subscript, as in V_B, indicates a potential relative to ground (in this case, base to ground potential). The exceptions to this rule are power supplies, that are given a double letter subscript indicating the connection point (V_{CC} is the collector power supply), and particularly important potentials that are directly named, as in vin (AC input voltage) and V_{R1} (DC voltage appearing across R_1). I_f an equation for a specific potential is valid for both the AC and DC equivalent circuits, the uppercase form is preferred (this makes things much more consistent with the vast majority of op amp circuits that are directly coupled, and thus can amplify both AC and DC signals). Currents are named in a similar way, but generally use a single subscript referring to the measurement node (IC is the DC collector current). All other items are directly named. By using this scheme, you will always be able to determine whether the item expressed in an equation is a DC or AC equivalent, its approximate circuit location, and other factors about it.

Most people are familiar with the term "decibel" in reference to sound pressure. It's not uncommon to hear someone say something such as "It was 110 decibels at the concert last night, and my ears are still ringing." This popular use is somewhat inaccurate, but does show that decibels indicate some sort of quantity — in this case, sound pressure level.

DECIBEL REPRESENTATION OF POWER AND VOLTAGE GAINS

In its simplest form, the decibel is used to measure some sort of gain, such as power or voltage gain. Unlike the ordinary gain measurements that you may be familiar with, the decibel form is logarithmic. Because of this, it can be very useful for showing ratios of change, as well as absolute change. The base unit is the Bel. To convert an ordinary gain to its Bel counterpart, just take the common log (base 10) of the gain. In equation form:

$$\text{Bel gain } = \log_{10}(ordinary\ gain)$$

Note that on most hand calculators common log is denoted as "log" while the natural log is given as "ln". Unfortunately, some programming languages use "log" to indicate natural log and "log10" for common log. More than one student has been bitten by this bug, so be forewarned! As an example, if a circuit produces an output power of 200 milliwatts for an input of 10 milliwatts, we would normally say that it has a power gain of:

$$G = \frac{^{P}out}{^{P}in}$$

$$G = \frac{200mW}{10mW}$$

$$G = 20$$

For the Bel version, just take the log of this result.

$$G^{'} = \log_{10}\ GG^{'} = \log_{10}\ 20G^{'} = 1.301$$

$$G^{'} = \log_{10}\ GG^{'} = \log_{10}\ 20G^{'} = 1.301$$

$$G^{'} = \log_{10}\ GG^{'} = \log_{10}\ 20G^{'} = 1.301$$

The Bel gain is 1.3 Bels. The term "Bels" is not a unit in the strict sense of the word (as in "watts"), but is simply used to indicate that this is not an ordinary gain. In contrast, ordinary power and voltage gains are sometimes given units of W/W and V/V to distinguish them from Bel gains. Also, note that the symbol for Bel power gain is G' and not G. All Bel gains are denoted with the following prime (') notation to avoid confusion. Because Bels tend to be rather large, we typically use one-tenth of a Bel as the norm. The result is the decibel (one-tenth Bel). To convert to decibels, simply multiply the number of Bels by 10. Our gain of 1.3 Bels is equivalent to 13 decibels. The units are commonly shortened to dB. Consequently, we may say:

$$G^{'} = 10 \log_{10} G$$

(7.2.1)

Where the result is in dB.

At this point, you may be wondering what the big advantage of the decibel system is. To answer this, recall a few log identities. Normal multiplication becomes addition in the log system, and division becomes subtraction. Likewise, powers and roots become multiplication and division. Because of this, two important things show up. First, ratios of change become constant offsets in the decibel system, and second, the entire range of values diminishes in size. The result is that a very wide range of gains may be represented within a fairly small scope of values, and the corresponding calculations can become quicker. There are a couple of dB values that are useful to remember. With the aid of your hand calculator, it is very easy to show the following:

Factor	dB Value using ◈′=10 log $_{10}$◈
1	0 dB
2	3.01 dB
4	6.02 dB
8	9.03 dB
10	10 dB

Table 7.2.1

We can also look at fractional factors (i.e., losses instead of gains.)

Factor	dB Value
0.5	–3.01 dB
0.25	–6.02 dB
0.125	–9.03 dB
0.1	–10 dB

Table 7.2.2

If you look carefully, you will notice that a doubling is represented by an increase of approximately 3 dB. A factor of 4 is in essence, two doublings. Therefore, it is equivalent to 3 dB + 3 dB, or 6 dB. Remember, because we are using logs, multiplication turns into simple addition. In a similar manner, a halving is represented by approximately –3 dB. The negative sign indicates a reduction. To simplify things a bit, think of factors of 2 as ±3 dB, the sign indicating whether you are increasing (multiplying), or decreasing (dividing). As you can see, factors of 10 work out to a very convenient 10 dB. By remembering these two factors, you can often estimate a dB conversion without the use of your calculator. For instance, we could rework our initial conversion problem as follows:

- The amplifier has a gain of 20
- 20 can be written as 2 times 10
- The factor of 2 is 3 dB, the factor of 10 is 10 dB
- The answer must be 3 dB + 10 dB, or 13 dB
- This verifies our earlier result

Time for a few examples.

Example 7.2.1

An amplifier has a power gain of 800. What is the decibel power gain?

$$G' = 10 \ \log_{10} \ G$$

$$G' = 10 \ \log_{10} \ 800$$

$$G' = 10 \times 2.903 \ G' = 29.03 \ dB$$

We could also use our estimation technique:

- G = 800 = $8 \cdot 10^2$
- 8 is equivalent to 3 factors of 2, or 2·2·2, and can be expressed as 3 dB + 3 dB + 3 dB, which is, of course, 9 dB
- 10^2 is equivalent to 2 factors of 10, or 10 dB + 10 dB = 20 dB
- The result is 9 dB + 20 dB, or 29 dB

Note that if the leading digit is not a power of 2, the estimation will not be as precise. For example, if the gain is 850, you know that the decibel gain is just a bit over 29 dB. You also know that it must be less than 30 dB (1000=103 which is 3 factors of 10, or 30 dB.) As you can see, by using the dB form, you tend to concentrate on the magnitude of gain, and not so much on trailing digits.

Example 7.2.2

An attenuator reduces signal power by a factor of 10,000. What is this loss expressed in dB?

$$G' = 10 \ log_{10} \ ]frac110000G' = 10 \times (-4)$$

$$G' = -40 \ dB$$

By using the approximation, we can say,

$$\frac{1}{10000} = 10^{-4}$$

The negative exponent tells us we have a loss (negative dB value), and 4 factors of 10.

$$G' = -10 \ dB - 10 \ dB - 10 \ dB - 10 \ dB$$

$$G' = -40 \ dB$$

Remember, if an increase in signal is produced, the result will be a positive dB value. A decrease in signal will always result in a negative dB value. A signal that is unchanged indicates a gain of unity, or 0 dB.

To convert from dB to ordinary form, just invert the steps; that is, divide by ten and then take the antilog.

$$G = \log_{10}^{-1} \frac{G'}{10}$$

On most hand calculators, base 10 antilog is denoted as 10x. In most computer languages, you just raise 10 to the appropriate power, as in G = 10.0^(Gprime / 10.0) (BASIC), or use an exponent function, as in pow(10.0, Gprime / 10.0) (C or Python).

Example 7.2.3

An amplifier has a power gain of 23 dB. If the input is 1 mW, what is the output?

In order find the output power, we need to find the ordinary power gain, G.

$$G = \log_{10}^{-1} \frac{G'}{10}$$

$$G = \log_{10}^{-1} \frac{23}{10}$$

$$G = 199.5$$

Therefore, P_{out} =199.5 · 1 mW, or 199.5 mW

You could also use the approximation technique in reverse. To do this, break up the dB gain in 10 dB and 3 dB chunks:

$$23\ dB = 3\ dB + 10\ dB + 10\ dB$$

Now replace each chunk with the appropriate factor, and multiply them together (remember, when going from log to ordinary form, addition turns into multiplication.)

$$3dB = 2X\ 10\ dB = 10X,\ \text{so,}$$

$$G = 2 \times 10 \times 10G = 200$$

While the approximation technique appears to be slower than the calculator, practice will show otherwise. Being able to quickly estimate dB values can prove to be a very handy skill in the electronics field. This is particularly true in larger, multi-stage designs.

Example 7.2.4

A three-stage amplifier has gains of 10 dB, 16 dB, and 14 dB per section. What is the total dB gain?

Because dB gains are a log form, just add the individual stage gains to arrive at the system gain.

$$G'_{total} = G'_1 + G'_2 + G'_3$$

$$G'_{total} = 10\ dB + 16\ dB + 14\ dB$$

$$G'_{total} = 40\ dB$$

As you may have noticed, all of the examples up to this point have used power gain and not voltage gain. You may be tempted to use the same equations for voltage gain. In a word, don't. If you think back for a moment, you will recall that power varies as the square of voltage. In other words, a doubling of voltage will produce a quadrupling of power. If you were to use the same dB conversions, a doubling of voltage would be 3 dB, yet, because the power has quadrupled, this would indicate a 6 dB rise. Consequently, voltage gain (and current gain as well) are treated in a slightly different fashion. We would rather have our doubling of voltage work out to 6 dB, so that it matches the power calculation. The correction factor is very simple. Because power varies as the second power of voltage, the dB form should be twice as large for voltage (remember, exponentiation turns into multiplication when using logs). Applying this factor to equation 1.1 yields:

$$A'_v = 20\ \log_{10}\ A_v$$

(7.2.2

Be careful though, the Bel voltage gain only equals the Bel power gain if the input and output impedances of the system are matched (you may recall from your earlier work that it is quite possible to design a circuit with vastly different voltage and power gains. A voltage follower, for example, exhibits moderate power gain with a voltage gain of unity. It is quite likely that the follower will not exhibit matched impedances.) If we were to recalculate our earlier table of common factors, we would find that a doubling of voltage gain is equivalent to a 6 dB rise, and a ten fold increase is equivalent to a 20 dB rise, twice the size of their power gain counterparts.

Note that current gain may be treated in the same manner as voltage gain (although this is less commonly done in practice).

Example 7.2.5

An amplifier has an output signal of 2 V for an input of 50 mV. What is A'$_v$? First find the ordinary gain.

$$A_v = \frac{2}{0.05} = 40$$

Now convert to dB form.

$$A'_v = 20\ \log_{10}\ 40$$

$$A'_v = 20 \times 1.602$$

$$A'_v = 32.04\ dB$$

The approximation technique yields 40 = 2·2·10, or 6 dB + 6 dB + 20 dB = 32 dB

To convert A'$_v$ to A, reverse the process.

$$A_v = \log_{10}^{-1} \frac{A_v^{'}}{20}$$

Example 7.2.6

An amplifier has a gain of 26 dB. If the input signal is 10 mV, what is the output?

$$A_v = \log_{10}^{-1} \frac{A_v^{'}}{20}$$

$$A_v = \log_{10}^{-1} \frac{26}{20}$$

$$A_v = 19.95$$

$$V_{out} = A_v \ V_{in}$$

$$V_{out} = 19.95 \times 10 \ mV$$

The final point to note in this section is that, as in the case of power gain, a negative dB value indicates a loss. Therefore, a 2:1 voltage divider would have a gain of –6 dB.

SIGNAL REPRESENTATION IN DBW AND DBV

As you can see from the preceding section, it is possible to spend considerable time converting between decibel gains and ordinary voltages and powers. Because the decibel form does offer advantages for gain measurement, it would make sense to use a decibel form for power and voltage levels as well. This is a relatively straightforward process. There is no reason why we can't express a power or voltage in a logarithmic form. Because a dB value just indicates a ratio, all we need to do is decide on a reference (i.e., a comparative base for the ratio). For power measurements, a likely choice would be 1 watt. In other words, we can describe a power as being a certain number of dB above or below 1 watt. Positive values will indicate powers greater than 1 watt, while negative values will indicate powers less than 1 watt. In general equation form:

$$P^{'} = 10 \ \log_{10} \frac{P}{reference}$$

(7.2.3)

The answer will have units of dBW, that is, decibels relative to 1 watt.

Example 7.2.7

A power amplifier has a maximum output of 120 W. What is this power in dBW?

$$P^{'} = 10 \ \log_{10} \frac{P}{1 \ Watt}$$

$$P^{'} = 10 \ \log_{10} \frac{120 \ W}{1 \ W}$$

$$P^{'} = 20.8 \ dBW$$

There is nothing sacred about the 1 watt reference, short of its convenience. We could just as easily choose a different reference. Other common reference points are 1 milliwatt (dBm) and 1 femtowatt (dBf). Obviously, dBf is used for very low signal levels, such as those coming from an antenna. dBm is in very wide use in the communications industry. To use these other references, just divide the given power by the new reference.

Example 7.2.8

A small personal music player delivers 200 mW to its headphones. What is this output power in dBW, and in dBm?

For an answer in units of dBW, use the 1 watt reference

$$P^{'} = 10 \ \log_{10} \frac{P}{1 \ Watt}$$

$$P^{'} = 10 \ \log_{10} \frac{200 \ mW}{1 \ W}$$

$$P^{'} = -7 \ dBW$$

For units of dBm, use a 1 milliwatt reference

$$P^{'} = 10 \ \log_{10} \frac{P}{1 \ Watt}$$

$$P^{'} = 10 \ \log_{10} \frac{200 \ mW}{1 \ mW}$$

$$P^{'} = 23 \ dBW$$

200 mW, –7 dBW, and 23 dBm are three ways of saying the same thing. Note that the dBW and dBm values are 30 dB apart. This will always be true, because the references are a factor of 1000 (30 dB) apart.

In order to transfer a dBW or similar value into watts, reverse the process.

$$P = \log_{10}^{-1} \frac{P'}{10} \times reference$$

Example 7.2.9

A studio microphone produces a 12 dBm signal while recording normal speech. What is the output power in watts?

$$P' = \log_{10}^{-1} \frac{P'}{10} \times \ reference$$

$$P' = \log_{10}^{-1} \frac{12\ dBm}{10} \times \ 1\ mW$$

$$P = 15.8\ mW = 0.0158\ W$$

For voltages, we can use a similar system. A logical reference is 1 V, with the resulting units being dBV. As before, these voltage measurements will use a multiplier of 20 instead of 10.

$$V' = 20\ \log_{10} \frac{V}{reference}$$

(7.2.4)

Example 7.2.10

A test oscillator produces a 2 V signal. What is this value in dBV?

$$V' = 20\ \log_{10} \frac{V}{reference}$$

$$V' = 20\ \log_{10} \frac{2\ V}{1\ V}$$

$$V' = 6.02\ dB$$

When both circuit gains and signal levels are specified in dB form, analysis can be very quick. Given an input level, simply add the gain to it in order to find the output level. Given input and output levels, subtract them in order to find the gain.

Example 7.2.11

A computer hard drive read/write amplifier exhibits a gain of 35 dB. If the input signal is –42 dBV, what is the output signal?

$$V_{out}^{'} = V_{in}^{'} + A_{v}^{'}$$

$$V_{out}^{'} = -42\ dBV + 35\ dB$$

$$V_{out}^{'} = -7\ dBV$$

Note that the final units are dBV and not dB, thus indicating a voltage and not merely a gain.

Example 7.2.12

A guitar power amp needs an input of 20 dBm to achieve an output of 25 dBW. What is the gain of the amplifier in dB?

First, it is necessary to convert the power readings so that they share the same reference unit. Because dBm represents a reference 30 dB smaller than the dBW reference, just subtract 30 dB to compensate.

$$20\ dBm = -10\ dBW$$

$$G^{'} = P_{out}^{'} - P_{in}^{'}$$

$$G^{'} = 25\ dBW - (-10\ dBW)$$

$$G^{'} = 35\ dB$$

Note that the units are dB and not dBW. This is very important! Saying that the gain is "so many" dBW is the same as saying the gain is "so many" watts. Obviously, gains are "pure" numbers and do not carry units such as watts or volts.

The usage of a dB-based system is shown graphically in Figure 1.1. Note how the stage gains are added to the input signal to form the output. Even large circuits can be quickly analyzed in this form. To make life in the lab even easier, it is possible to take measurements directly in dB form. By doing this, you need never convert while troubleshooting a design. For general-purpose work, voltage measurements are the norm, and therefore a dBV scale is often used.

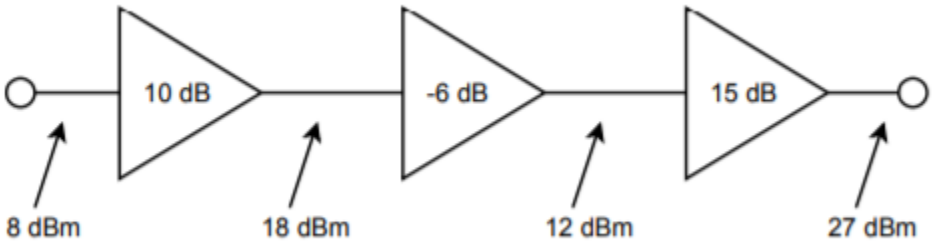


Figure 7.2.1 : Multistage dB application.

ITEMS OF INTEREST IN THE LABORATORY

When using a digital meter on a dBV scale it is possible to "underflow" the meter if the signal is too weak. This will happen if you try to measure around zero volts, for example. If you attempt to calculate the corresponding dBV value, your calculator will probably show "error". The effective value is negative infinite dBV. The meter will certainly have a hard time showing this value! Another item of interest revolves around the use of dBm measurements. It is common to use a voltmeter to make dBm measurements, in lieu of a wattmeter. While the connections are considerably simpler, a voltmeter cannot measure power. How is this accomplished then? Well, as long as the circuit impedance is known, power can be derived from a voltage measurement. A common impedance in communication systems (such as recording studios) is 600 Ω, so a meter can be calibrated to give correct dBm readings by using Power Law. If this meter is used on a non-600 Ω circuit, the readings will no longer reflect accurate dBm values (but will still properly reflect relative changes in dB).

7.3 BODE PLOTS

The Bode plot is a graphical response prediction technique that is useful for both circuit design and analysis. It is named after Hendrik Wade Bode, an American engineer known for his work in control systems theory and telecommunications. A Bode plot is, in actuality, a pair of plots: One graphs the gain of a system versus frequency, while the other details the circuit phase versus frequency. Both of these items are very important in the design of well-behaved, optimal operational amplifier circuits.

Generally, Bode plots are drawn with logarithmic frequency axes, a decibel gain axis, and a phase axis in degrees. First, let's take a look at the gain plot. A typical gain plot is shown Figure 7.3.1 .

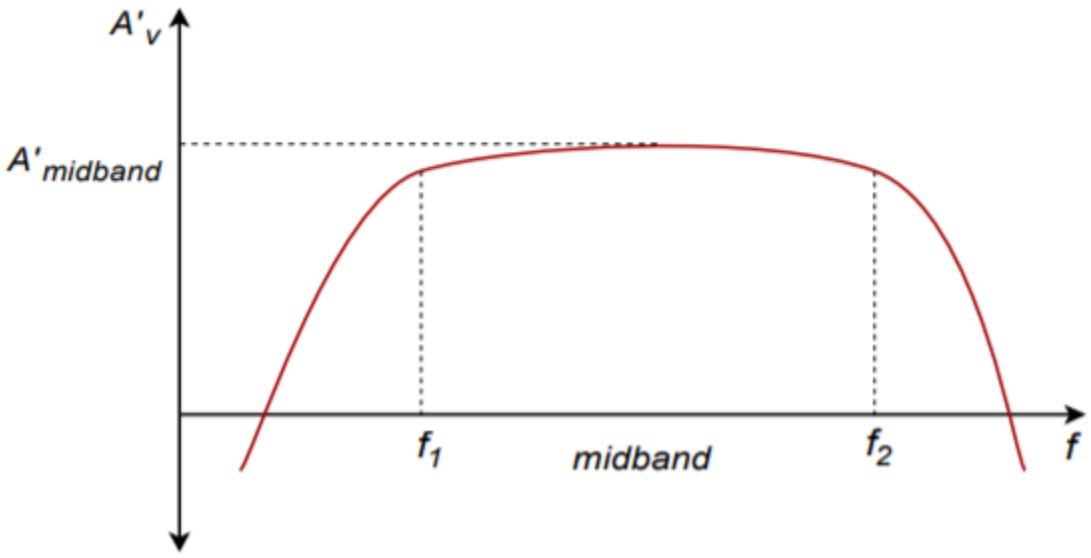


Figure 7.3.1 : Gain plot.

Note how the plot is relatively flat in the middle, or midband, region. The gain value in this region is known as the midband gain. At either extreme of the midband region, the gain begins to decrease. The gain plot shows two important frequencies, f_1 and f_2. f_1 is the lower break frequency while f_2 is the upper break frequency. The gain at the break frequencies is 3 dB less than the midband gain. These frequencies are also known as the half-power points, or corner frequencies. Normally, amplifiers are only used for signals between f_1 and f_2. The exact shape of the rolloff regions will depend on the design of the circuit. It is possible to design amplifiers with no lower break frequency (i.e., a DC amplifier), however, all amplifiers will exhibit an upper break. The break points are caused by the presence of circuit reactances, typically coupling and stray capacitances. The gain plot is a summation of the midband response with the upper and lower frequency limiting networks. Let's take a look at the lower break, f_1 .

LEAD NETWORK GAIN RESPONSE

Reduction in low frequency gain is caused by lead networks. A generic lead network is shown in Figure 7.3.2 . It gets its name from the fact that the output voltage developed across ⍰ leads the input. At very high frequencies the circuit will be essentially resistive. Conceptually, think of this as a simple voltage divider. The divider ratio depends on the reactance of ⍰ . As the input frequency drops, ⍰⍰ increases. This makes ⍰⍰⍰⍰ decrease. At very high frequencies, where ⍰⍰≪⍰ , ⍰⍰⍰⍰ is approximately equal to ⍰⍰⍰ . This can be seen graphically in Figure 7.3.3 . The break frequency (i.e., the frequency at which the signal has decreased by 3 dB) is found via the standard equation

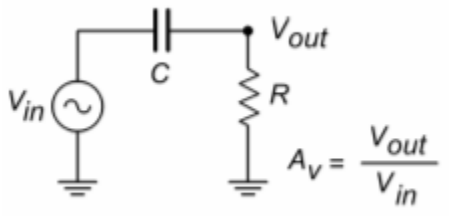


Figure 7.3.2 : Lead network.

$$f_c = \frac{1}{2\pi RC}$$

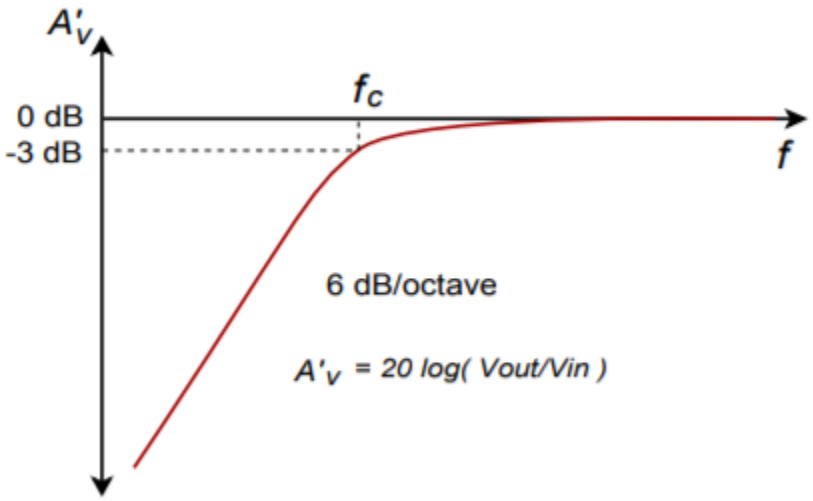


Figure 7.3.3 : Lead gain plot (exact).

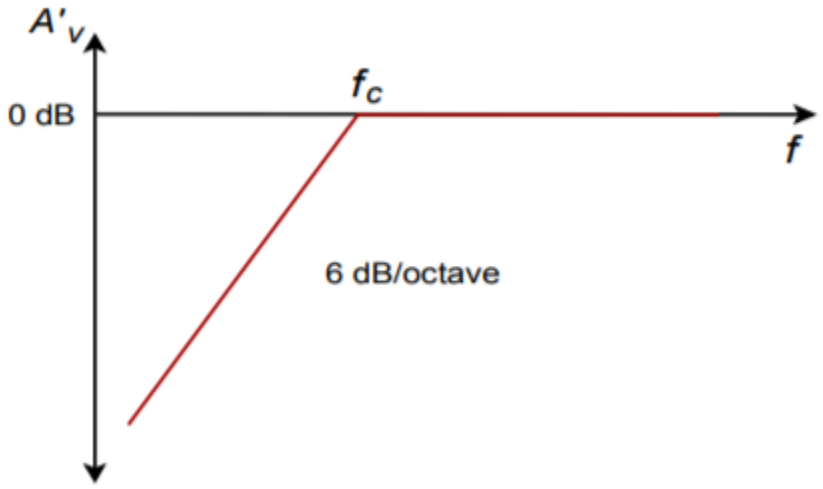


Figure 7.3.4 : Lead gain plot (approximate).

The response below ⟐⟐ will be a straight line if a decibel gain axis and a logarithmic frequency axis are used. This makes for very quick and convenient sketching of circuit response. The slope of this line is 6 dB per octave (an octave is a doubling or halving of frequency, e.g., 800 Hz is 3 octaves above 100 Hz).[1] This range covers a factor of two in frequency. This slope may also be expressed as 20 dB per decade, where a decade is a factor of 10 in frequency. With reasonable accuracy, this curve may be approximated as two line segments, called asymptotes, as shown in Figure 7.3.4 . The shape of this curve is the same for any lead network. Because of this, it is very easy to find the approximate gain at any given frequency as long as ⟐⟐ is known. It is not necessary to go through reactance and phasor calculations. To create a general response equation, start with the voltage divider rule to find the gain:

$$\frac{V_{out}}{V_{in}} = \frac{R}{R - j\ X_c}$$

$$\frac{V_{out}}{V_{in}} = \frac{R\angle 0}{\sqrt{R^2 + X_c^2}\angle - \arctan\frac{X_c}{R}}$$

The magnitude of this is,

$$|A_v| = \frac{R}{\sqrt{R^2 + X_c^2}}$$

$$|A_v| = \frac{1}{\sqrt{1 + \frac{X_c^2}{R^2}}}$$

(7.3.1)

Recalling that,

$$f_c = \frac{1}{2\pi R\ C}$$

1. The term octave is borrowed from the field of music. It gets its name from the fact that there are eight notes in the standard western scale: do-re-me-fa-so-la-ti-do.

we may say,

$$R = \frac{1}{2\pi f_c\ C}$$

For any frequency of interest, ◈ ,

$$X_c = \frac{1}{2\pi f_c\ C}$$

Equating the two preceding equations yields,

$$\frac{f_c}{f} = \frac{X_c}{R} l](7.3.2) Substituting Equation 7.3.2 in Equation 7.3.1 gives, A_v = \frac{1}{\sqrt{1 + \frac{f_c^2}{f^2}}}$$

(7.3.3)

To express ◈◈ in dB, substitute Equation 7.3.3 into Equation 1.2.2

$$A_v^{'} = 20 \log_{10} \frac{1}{\sqrt{1 + \frac{f_c^2}{f^2}}}$$

After simplification, the final result is:

$$A_v^{'} = -10 \log_{10} \left(1 + \frac{f_c^2}{f^2}\right)$$

(7.3.4)

Where

◈◈ is the critical frequency,

◈ is the frequency of interest,

◈′◈ is the decibel gain at the frequency of interest.

Example 7.3.1

An amplifier has a lower break frequency of 40 Hz. How much gain is lost at 10 Hz?

$$A_v^{'} = -10 \log_{10} \left(1 + \frac{f_c^2}{f^2}\right)$$

$$A_v^{'} = -10 \log_{10} \left(1 + \frac{40^2}{10^2}\right)$$

$$A_v^{'} = -12.3\ dB$$

In other words, the gain is 12.3 dB lower than it is in the midband. Note that 10 Hz is 2 octaves below the break frequency. Because the cutoff slope is 6 dB per octave, each octave loses 6 dB. Therefore, the

approximate result is -12 dB, which double-checks the exact result. Without the lead network, the gain would stay at 0 dB all the way down to DC (0 Hz.)

LEAD NETWORK PHASE RESPONSE

At very low frequencies, the circuit of Figure 7.3.2 is largely capacitive. Because of this, the output voltage developed across ◈ leads by 90 degrees. At very high frequencies the circuit will be largely resistive. At this point ◈◈◈◈ will be in phase with ◈◈◈ . At the critical frequency, ◈◈◈◈ will lead by 45 degrees. A general phase graph is shown in Figure 7.3.5 . As with the gain plot, the phase plot shape is the same for any lead network. The general phase Equation may be obtained from the voltage divider:

$$\frac{V_{out}}{V_{in}} = \frac{R}{R - j\ X_c}$$

$$\frac{V_{out}}{V_{in}} = \frac{R\angle 0}{\sqrt{R^2 + X_c^2}\angle - \arctan\frac{X_c}{R}}$$

The phase portion of this is,

$$\theta = \arctan\frac{X_c}{R}$$

By using Equation 7.3.2 , this simplifies to,

$$\theta = \arctan\frac{f_c}{f}$$

(7.3.5)

Where

◈◈ is the critical frequency,

◈ is the frequency of interest,

◈ is the phase angle at the frequency of interest.

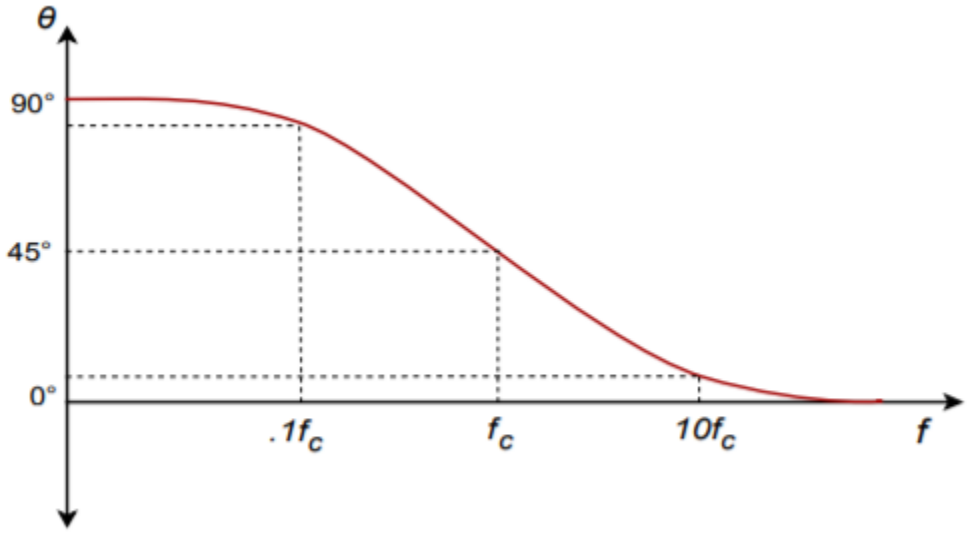


Figure 7.3.5 : Lead phase (exact).

Often, an approximation such as Figure 7.3.6 used in place of 7.3.5 . By using Equation 7.3.5 , you can show that the approximation is off by no more than 6 degrees at the corners.

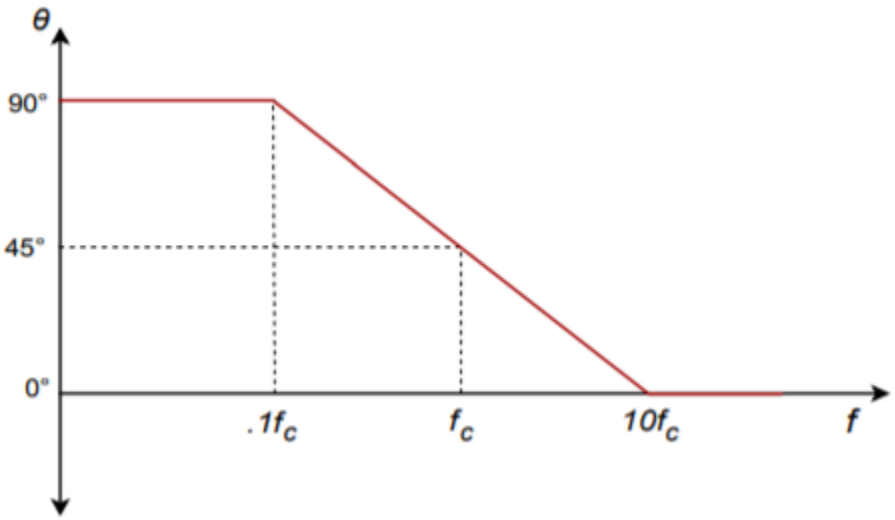


Figure 7.3.6 : Lead phase (approximate).

Example 7.3.2

A telephone amplifier has a lower break frequency of 120 Hz. What is the phase response one decade below and one decade above?

One decade below 120 Hz is 12 Hz, while one decade above is 1.2 kHz.

$$\theta = \arctan \frac{f_c}{f}$$

$$\theta = \arctan \frac{120\ Hz}{12\ Hz}$$

◈=84.3 degrees one decade below ◈◈ (i.e, approaching 90 degrees)

$$\theta = \arctan \frac{120}{1.2\ kHz}$$

◈=5.71 degrees one decade above ◈◈ (i.e., approaching 0 degrees)

Remember, if an amplifier is direct-coupled, and has no lead networks, the phase will remain at 0 degrees right back to 0 Hz (DC).

LAG NETWORK RESPONSE

Unlike its lead network counterpart, all amplifiers will contain lag networks. In essence, it's little more than an inverted lead network. As you can see from Figure 7.3.7 , it simply transposes the ◈ and ◈ locations. Because of this, the response tends to be inverted as well. In terms of gain, ◈◈ is very large at low frequencies, and thus ◈◈◈◈equals ◈◈◈. At high frequencies, ◈◈ decreases, and ◈◈◈◈falls. The break point occurs when ◈◈ equals ◈ . The general gain plot is shown in Figure 7.3.8 . Like the lead network response, the slope of this curve is -6 dB per octave (or -20 dB per decade.) Note that the slope is negative instead of positive. A straight-line approximation is shown in Figure 7.3.9 . We can derive a general gain Equation for this circuit in virtually the same manner as we did for the lead network. The derivation is left as an exercise.

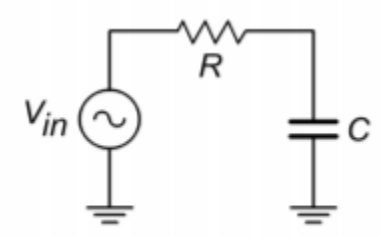


Figure 7.3.7 : Lag network.

$$A_v' = -10 \log_{10} \left(1 + \frac{f^2}{f_c^2}\right)$$

(7.3.6)
Where
◈◈ is the critical frequency,
◈ is the frequency of interest,
◈′◈ is the decibel gain at the frequency of interest.

Note that this Equation is almost the same as Equation 7.3.4 . The only difference is that ⍰ and ⍰⍰ have been transposed.

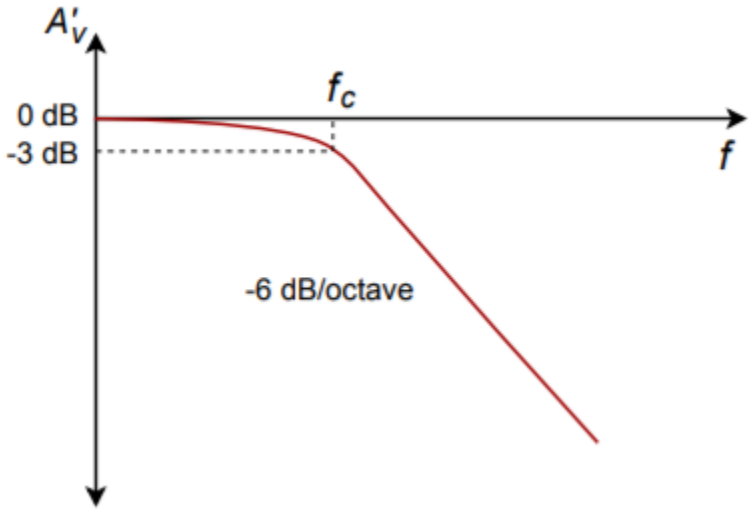


Figure 7.3.8 : Lag gain (exact).

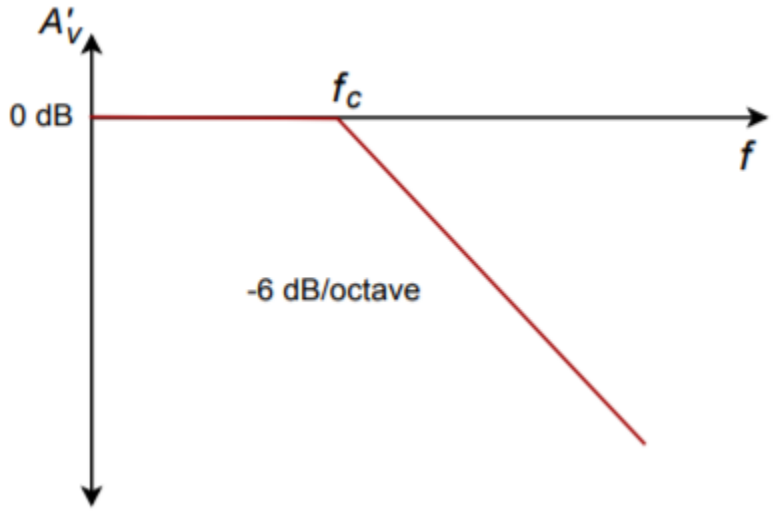


Figure 7.3.9 : Lag gain (approximate).

In a similar vein, we may examine the phase response. At very low frequencies, the circuit is basically capacitive. Because the output is taken across ⍰ , ⍰⍰⍰⍰will be in phase with ⍰⍰⍰. At very high frequencies, the circuit is essentially resistive. Consequently, the output voltage across ⍰ will lag by 90 degrees. At the break frequency the phase will be -45 degrees. A general phase plot is shown in Figure 7.3.10 , with the approximate response detailed in Figure 7.3.11 . As with the lead network,we may derive a phase equation. Again, the exact steps are very similar, and left as an exercise.

$$\theta = -90 + \arctan \frac{f_c}{f}$$

(7.3.7)

Where ◈◈ is the critical frequency,
◈ is the frequency of interest,
◈ is the phase angle at the frequency of interest.

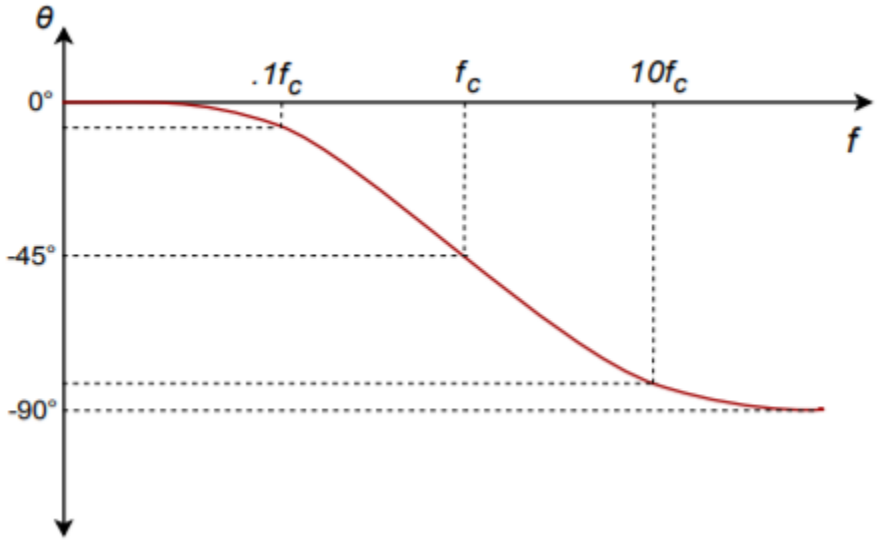


Figure 7.3.10 : Lag phase (exact).

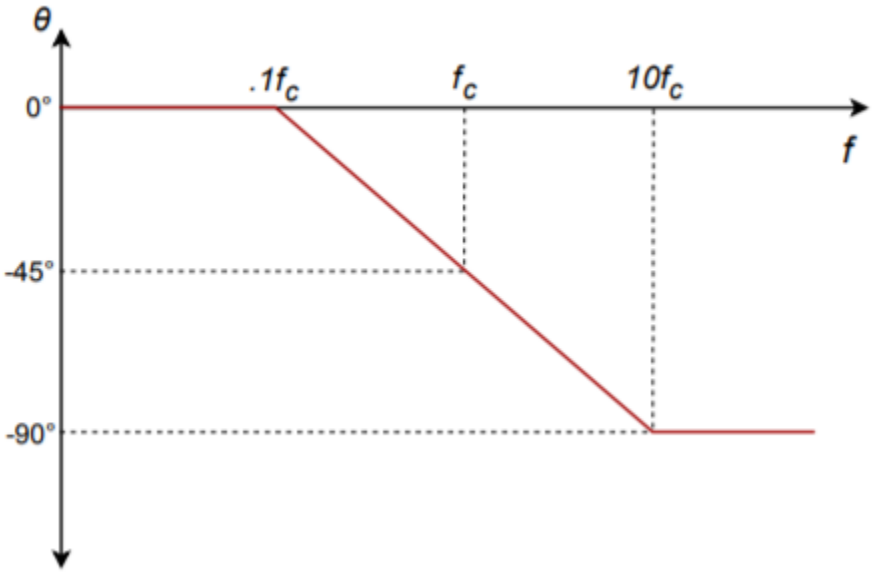


Figure 7.3.11 : Lag phase (approximate).

Example 7.3.3

A medical ultra sound transducer feeds a lag network with an upper break frequency of 150 kHz. What are the gain and phase values at 1.6 MHz? Because this represents a little more than a 1 decade increase, the

approximate values are -20 dB and -90 degrees, from Figures 7.3.9 and 7.3.11, respectively. The exact values are:

$$A_v' = -10 \log_{10} \left(1 + \frac{f^2}{f_c^2}\right)$$

$$A_v' = -10 \log_{10} \left(1 + \frac{1.6\ MHz^2}{150\ MHz^2}\right)$$

$$A_v' = -20.6\ dB$$

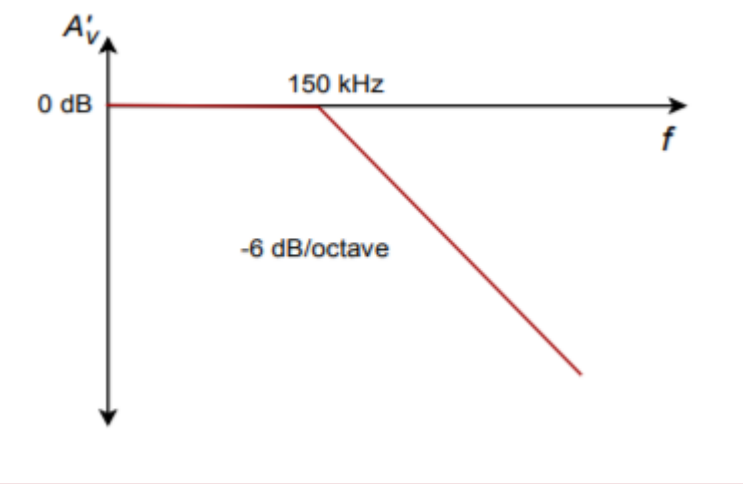


Figure 7.3.12 : Bode plot for 150 kHz lag.

$$\theta = -90 + \arctan \frac{f_c}{f}$$

$$\theta = -90 + \arctan \frac{150\ kHz}{1.6\ MHz}$$

$$\theta = -84.6\ degrees$$

The complete Bode plot for this network is shown in Figure 7.3.12 . It is very useful to examine both plots simultaneously. In this manner you can find the exact phase change for a given gain quite easily. This information is very important when the application of negative feedback is considered (Chapter Three). For example, if you look carefully at the plots of Figure 7.3.12 , you will note that at the critical frequency of 150 kHz, the total phase change is -45 degrees. Because this circuit involved the use of a single lag network, this is exactly what you would expect.

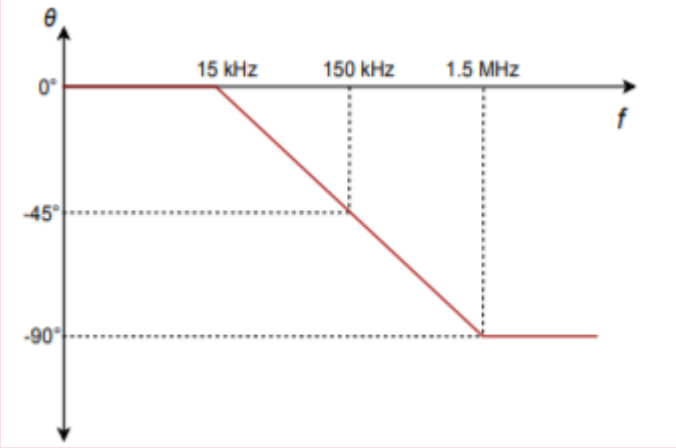


Figure 7.3.13 : (continued) Bode plot for 150 kHz lag.

RISE TIME VERSUS BANDWIDTH

For pulse-type signals, the speed of an amplifier is often expressed in terms of its rise time. If a square pulse such as Figure 7.3.13 is passed into a simple lag network, the capacitor charging effect will produce a rounded variation, as seen in Figure 7.3.13 . This effect places an upper limit on the duration of pulses that a given amplifier can handle without producing excessive distortion.

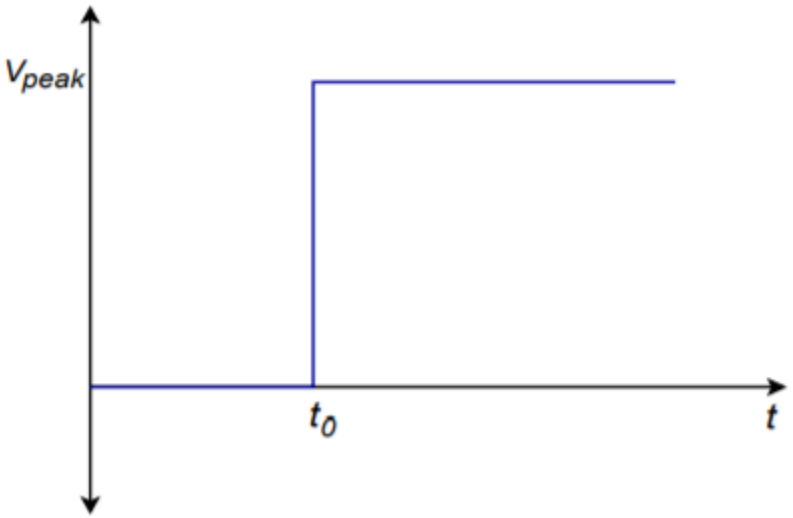


Figure 7.3.14 : Pulse rise time effect: Input to network.

By definition, rise time is the amount of time it takes for the signal to traverse from 10% to 90% of the peak value of the pulse. The shape of this pulse is defined by the standard capacitor charge Equation examined in earlier course work, and is valid for any system with a single clearly dominant lag network.

$$V_{out} = V_{peak}\left(1 - \epsilon^{\frac{-t}{RC}}\right)$$

(7.3.12

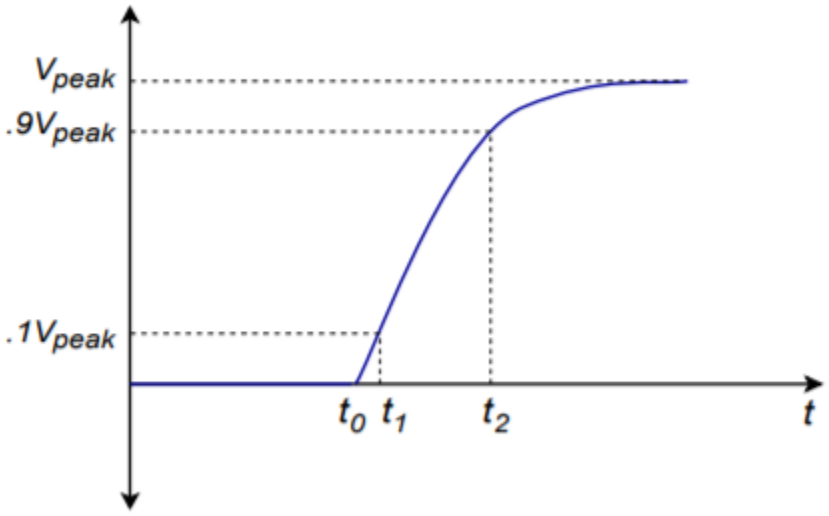


Figure 7.3.14 : Pulse rise time effect: Output of network.

In order to find the time internal from the initial starting point to the 10% point, set to 0.1 in Equation 1.12 and solve for 1 .

$$0.1V_{peak} = V_{peak}\left(1 - \epsilon^{\frac{-t_1}{RC}}\right)$$

$$0.1V_{peak} = V_{peak} - V_{peak}\ \epsilon^{\frac{-t_1}{RC}}$$

$$0.9V_{peak} = V_{peak}\ \epsilon^{\frac{-t_1}{RC}}$$

$$0.9 = \epsilon^{\frac{-t_1}{RC}}$$

$$\log\ 0.9 = \frac{-t_1}{RC}$$

$$t_1 = 0.105\ RC$$

(7.3.8)

To find the interval up to the 90% point, follow the same technique using 0.9 . Doing so yields

$$t_2 = 2.303\ RC$$

(7.3.9)

The rise time, , is the difference between 1 and 2

$$T_r = t_1 - t_2$$
$$T_r = 2.303\ RC - 0.105\ RC$$
$$T_r \approx 2.2\ RC$$

(7.3.10)

Equation 7.3.10 ties the rise time to the lag network's and values. These same values also set the critical frequency 2 . By combining Equation 7.3.10 with basic critical frequency relationship, we can derive an Equation relating 2 to .

$$f_2 = \frac{1}{2\pi RC}$$

Solving 7.3.10 in terms of ⍰⍰ , and substituting yields

$$f_2 = \frac{2.2}{2\pi T_r}$$

$$f_2 = \frac{0.35}{T_r}$$

(7.3.11)

Where ⍰2 is the upper critical frequency,

⍰⍰ is the rise time of the output pulse.

Example 7.3.4

Determine the rise time for a lag network critical at 100 kHz.

$$f_2 = \frac{0.35}{T_r}$$

\

$$T_r = \frac{0.35}{f_2}$$

\

$$T_r = \frac{0.35}{100\ kHz}$$

$$T_r = 3.5\ \mu s$$

7.4 COMBINING THE ELEMENTS - MULTI-STAGE EFFECTS

A complete gain or phase plot combines three elements: (1) the midband response, (2) the lead response, and (3) the lag response. Normally, a particular design will contain multiple lead and lag networks. The complete response is the summation of the individual responses. For this reason, it is useful to find the dominant lead and lag networks. These are the networks that affect the midband response first. For lead networks, the dominant one will be the one with the highest ω_0 . Conversely, the dominant lag network will be the one with the lowest ω_0 . It is very common to approximate the complete system response by drawing straight-line segments such as those given in Figures 1.3.4 and 1.3.9. The process goes something like this:

- Locate all ω_0 s on the frequency axis.
- Draw a straight line between the dominant lag and lead ω_0 s at the midband gain. If the system does not contain any lead networks, continue the midband gain line down to DC.
- Draw a 6 dB per octave slope between the dominant lead and the next lower lead network.
- Because the effects of the networks are cumulative, draw a 12 dB per octave slope between the second lead ω_0 and the third ω_0 . After the third ω_0 , the slope should be 18 dB per octave, after the fourth, 24 dB per octave, and so on.
- Draw a -6 dB per octave slope between the dominant lag ω_0 and the next highest ω_0 . Again, the effects are cumulative, so increase the slope by -6 dB at every new ω_0 .

Example 7.4.1

Draw the Bode gain plot for the following amplifier: ω_0' midband = 26 dB, one lead network critical at 200 Hz, one lag network critical at 10 kHz, and another lag network critical at 30 kHz.

The dominant lag network is 10 kHz. There is only one lead network, so it's dominant by default.

- Draw a straight line between 200 Hz and 10 kHz at an amplitude of 26 dB.
- Draw a 6 dB per octave slope below 200 Hz. To do this, drop down one octave (100 Hz) and subtract 6 dB from the present gain (26 dB – 6dB = 20 dB.) The line will start at the point 200 Hz/26 dB, and pass through the point 100 Hz/20 dB. Because there are no other lead networks, this line may be extended to the left edge of the graph.
- Draw a -6 dB per octave slope between 10 kHz and 30 kHz. The construction point will be 20 kHz/20 dB. Continue this line to 30 kHz. The gain at the 30 kHz intersection should be around 16 dB. The slope above this second ω_0 will be -12 dB per octave. Therefore, the second construction point should be at 60 kHz/4 dB (one octave above 30 kHz, and 12 dB down from the 30 kHz gain). Because this is the final lag network, this line may be extended to the right edge of the graph.

A completed graph is shown in Figure 1.4.1 .

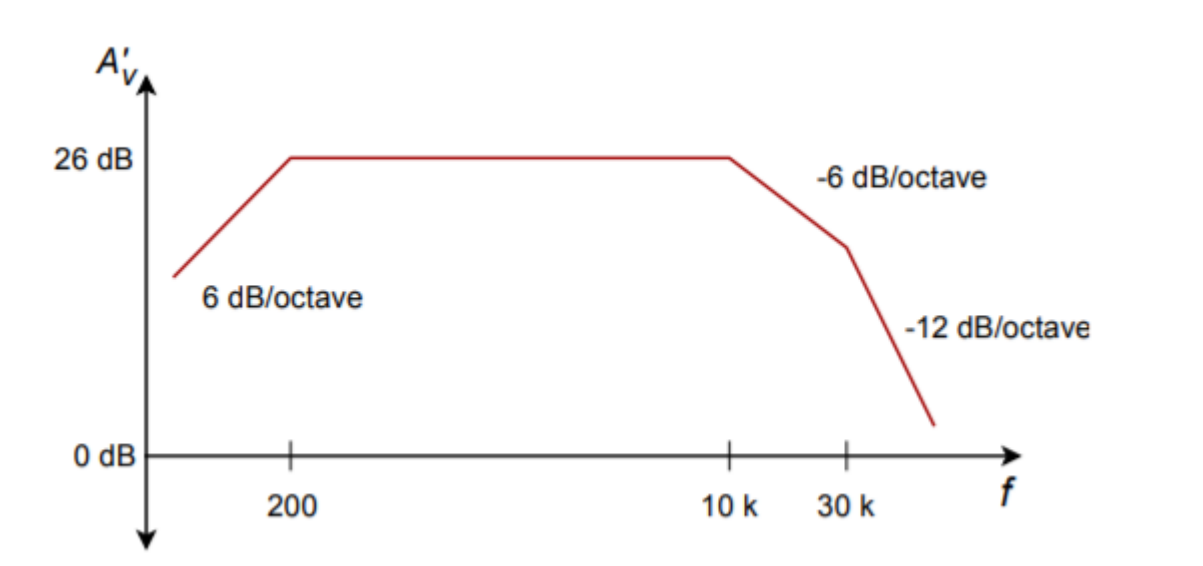


Figure 1.4.1 : Gain plot of complete amplifier.

There is one item that should be noted before we leave this section, and that is the concept of narrowing. Narrowing occurs when two or more networks share similar critical frequencies, and one of them is a dominant network. The result is that the true -3 dB breakpoints may be altered. Here is an extreme example. Assume that a circuit has two lag networks, both critical at 1 MHz. A Bode plot would indicate that the breakpoint is 1 MHz. This is not really true. Remember, the effects of lead and lag networks are cumulative. Because each network produces a 3 dB loss at 1 MHz, the net loss at this frequency is actually 6 dB. The true -3 dB point will have been shifted. The Bode plot only gives you the approximate shape of the response.

7.5 CIRCUIT SIMULATIONS USING COMPUTERS

With the advent of low cost personal computers, there are many alternatives to hand sketching of plots. One method involves the use of commercial or public domain software packages designed for circuit analysis. One common package is the public domain program SPICE. SPICE is an acronym that stands for Simulation Program with Integrated Circuit Emphasis, and was originally written in the mid- 1970's by Dr. Laurence Nagel of the University of California. This program is available for many different computing platforms for minimal cost. SPICE also serves as the core for a number of commercial packages. The commercial versions generally add features such as schematic capture (the ability to "draw" circuits using the computer's mouse), graphical input and output of data, interactive analysis, analog-digital mixed signal analysis and large device libraries. Examples of popular simulation packages include the commercial offerings OrCAD PSpice and Electronics Workbench Multisim, while freeware packages include Linear Technology LTspice and Texas Instruments TINA-TI. The choice of a specific simulation tool depends on the needs of the user, available facilities, costs, and so forth. Of course, everyone has their own working style, so personal preference also plays a role. Generally, any quality SPICE-based simulator will be sufficient for the circuits presented in this book.

In order to use a simulator, a circuit is "described" with a special data file. For non-graphical simulators, the data file can be created using an ordinary text editor. The data file is then used by the simulation program to estimate the circuit response. Simulation results can include items such as Bode gain and phase plots.

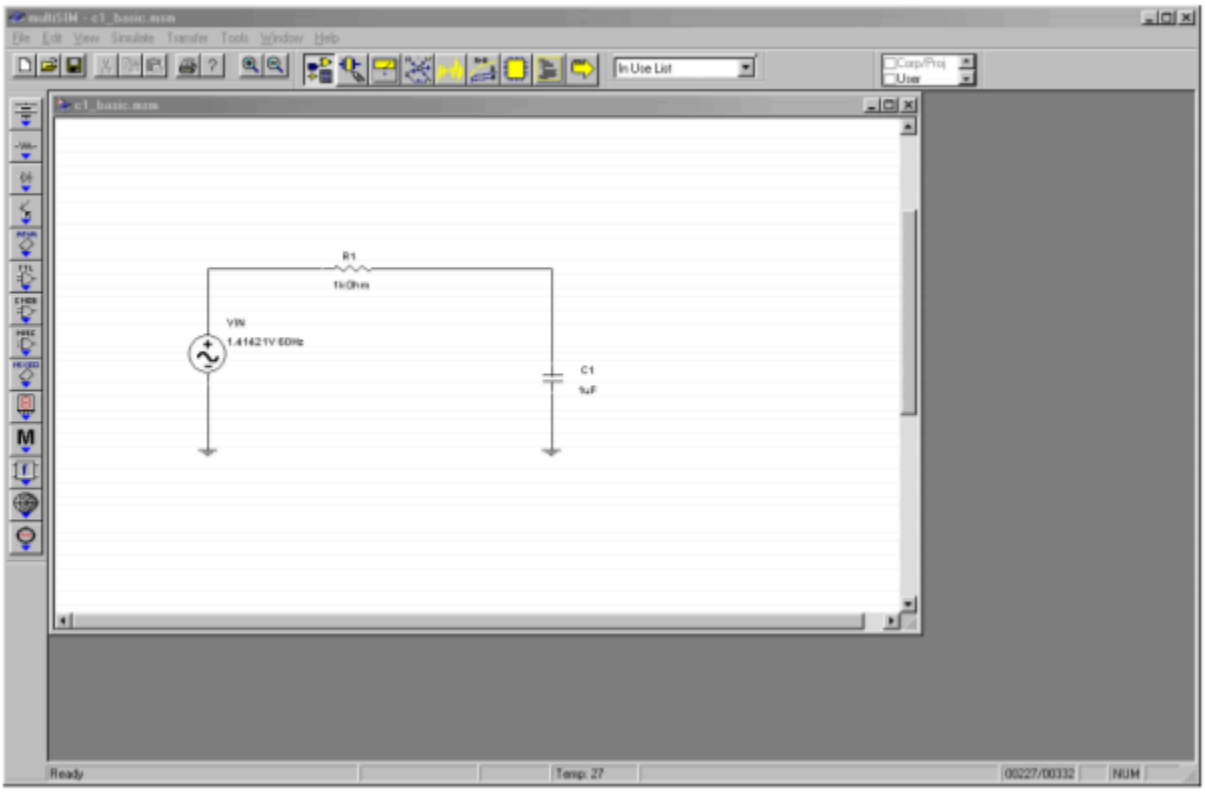


Figure 7.5.1 : Multisim schematic for a simple lag network.

With graphical input (AKA schematic capture), components are usually dragged onto the work area and interconnected through the use of a mouse or other pointing device. Consequently, text-based input files are not needed, although many programs can import them and create the circuit from there. Similarly, simulation results are generally shown in graphical form using a plotting window or virtual oscilloscope instead of using a text-based output file. Graphical input and output using Multisim is shown in Figure 7.5.1 . Here a simple lag network is drawn on the worksheet. A Bode plot is then generated by selecting the AC Analysis option of Simulate.

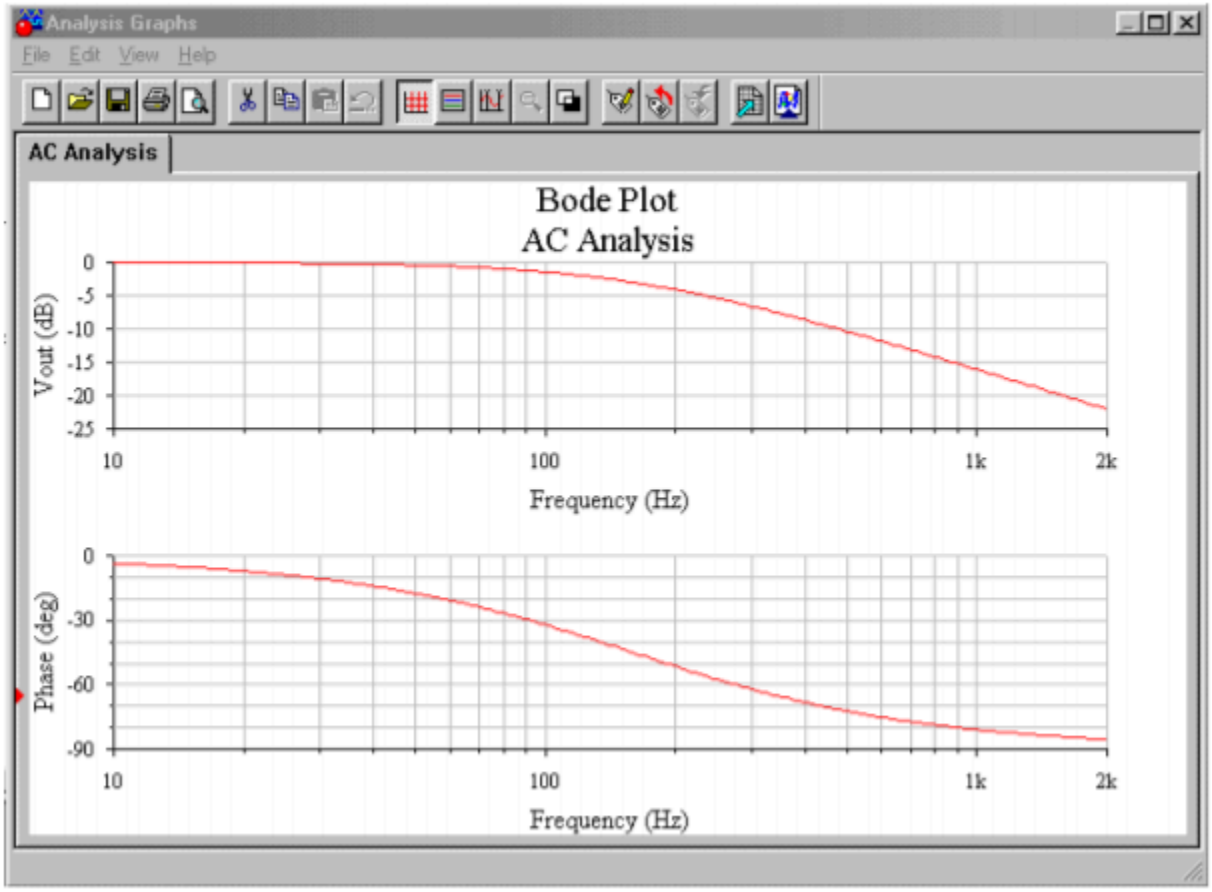


Figure 7.5.1 : Multisim gain and phase output graphs.

Note that accurate models for specific op amps or other devices that are not included in the simulator's libraries may be obtained from their manufacturer. This is often as simple as downloading them from the manufacturer's Web site. A listing of manufacturer's Web sites may be found in the Appendix.

Simulators are by no means small or trivial programs. They have many features and options, not to mention the variations produced by the many commercial versions. This text does not attempt to teach all of the intricacies of SPICE-based simulators. For that, you should consult your simulator user's manual, or one of the books available on the subject. The examples in this book assume that you already have some familiarity with computer circuit simulators.

We will be using simulations in the following chapters for various purposes. One thing that you should always bear in mind is that simulation tools should not be used in place of a normal "human" analysis. Doing so can cause no end of grief. Simulations are only as good as the models used with them. It is easy to see that if the description of the circuit or the components within the circuit is not accurate, the simulation will not be accurate. Simulation tools are best used as a form of double checking a design, not as a substitute for proper analysis.

7.6 THE DIFFERENTIAL AMPLIFIER

Most modern operational amplifiers utilize a differential amplifier front end. In other words, the first stage of the operational amplifier is a differential amplifier. This circuit is commonly referred to as a diff amp or as a long-tailed pair. A diff amp utilizes a minimum of 2 active devices, although 4 or more may be used in more complex designs. Our purpose here is to examine the basics of the diff amp so that we can understand how it relates to the larger operational amplifier. Therefore, we will not be investigating the more esoteric designs. To approach this in an orderly fashion, we will examine the DC analysis first, and then follow with the AC small signal analysis.

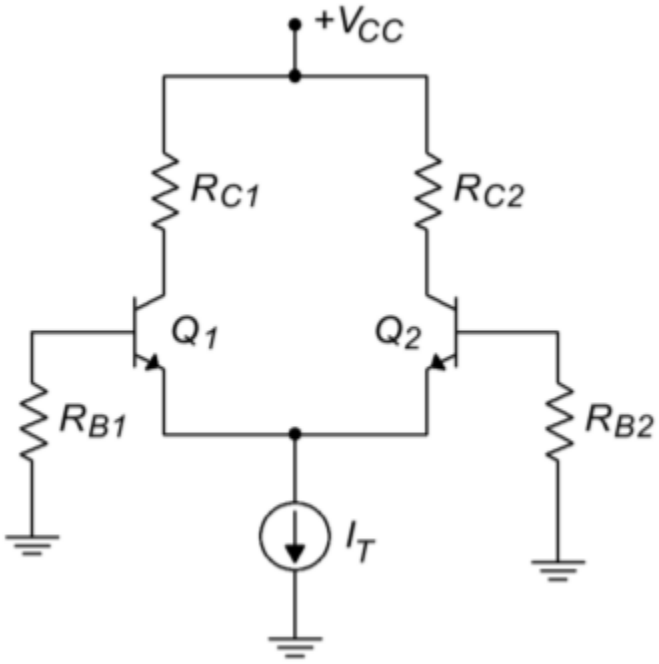


Figure 7.6.1 : Simplified diff amp.

DC ANALYSIS

A simplified diff amp is shown in Figure 7.6.1 . This circuit utilizes a pair of NPN bipolar transistors, although the circuit could just as easily be built with PNPs or FETs. Note the inherent symmetry of the circuit. If you were to slice the circuit in half vertically, all of the components on the left half would have a corresponding component on the right half. Indeed, for optimal performance, we will see that these component pairs should have identical values. For critical applications, a matched pair of transistors would be used. In this case, the transistor parameters, such as ◈ , would be very closely matched for the two devices.

In Figure 7.6.2 , the circuit currents are noted, and the generalized current source has been replaced with a resistor/negative power supply combination. This is in essence, an emitter bias technique. Assuming that the base voltages are negligible and that ◈◈◈ is equal to 0.7 V, we can see that the

emitter of each device is at approximately -0.7 V. Kirchhoff's Voltage Law indicates that the bulk of the negative supply potential must drop across ◈ .

$$V_{RT} = |V_{EE}| - .7V$$

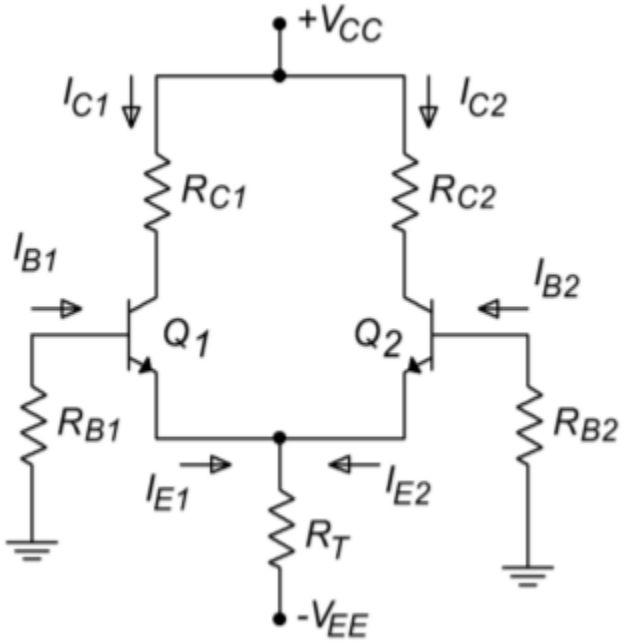


Figure 7.6.2 : Diff amp analysis of Figure 7.6.1 .

Knowing this, we may find the current through ◈ , which is known as the tail current, ◈ .

$$I_T = \frac{|V_{EE}| - 0.7V}{R_T}$$

If the two halves of the circuit are well matched, the tail current will split equally into two portions, ◈1 and ◈2 . Given identical emitter currents, it follows that the remaining currents and voltages in the two halves must be identical as well. These potentials and currents are found through the application of Kirchhoff's Voltage and Current Laws just as in any other transistor bias analysis.

Example 7.6.1

Find the tail current, the two emitter currents, and the two collector to ground voltages in the circuit of Figure 7.6.3 . You may assume that the two transistors are very closely matched.

The first step is to find the tail current:

$$I_T = \frac{|V_{EE}| - 0.7V}{R_T}$$

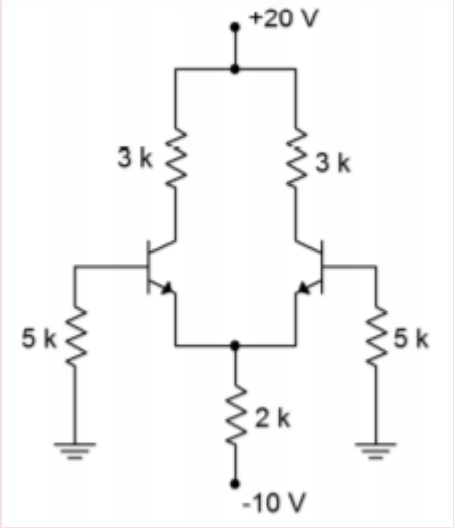


Figure 7.6.3 : Diff amp for example 7.6.1 .

The tail current is the combination of the two equal emitter currents, so

$$I_{EI} = I_{E2} = I_{T2}$$

$$I_{EI} = I_{E2} = \frac{4.65\ mA}{2}$$

$$I_{EI} = I_{E2} = 2.325\ mA$$

If we make the approximation that collector and emitter currents are equal, we may find the collector voltage by calculating the voltage drop across the collector resistor, and subtracting the result from the positive power supply.

$$V_c = V_{cc} - I_c\ R_c$$

$$V_c = 20\ V - 2.325\ mA \times 3k\Omega$$

$$V_c = 20\ V - 6.975\ V$$

$$V_c = 13.025\ V$$

Again, because we have identical values for both halves of the circuit, $\beta_1 = \beta_2$. If we continue with this and assume a typical β of 100, we find that the two base currents are identical as well.

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

$$I_B = \frac{2.325\ mA}{100}$$

$$I_B = 23.25\ \mu A$$

Noting that the base currents flow through the 5 k Ω base resistors, we may find the base voltages. Note that this is a negative potential because the base current is flowing from ground into the transistor's base.

$$V_B = -I_b\ R_B$$

$$V_B = 23.25\ \mu A \times 5k\Omega$$

$$V_B = 116.25\ mV$$

This result indicates that the actual emitter voltage is closer to -0.8 V than -0.7 V, and thus, the tail current is actually a little less than our approximation of 4.65 mA. This error is probably within the error we can expect by using the 0.7 V junction potential approximation.

INPUT OFFSET CURRENT AND VOLTAGE

As you have no doubt guessed, it is impossible to make both halves of the circuit identical, and thus, the currents and voltages will never be exactly the same. Even a small resistor tolerance variation will cause an upset. If the base resistors are mismatched, this will cause a direct change in the two base potentials. A variation in collector resistance will cause a mismatch in the collector potentials. A simple ◈ or ◈◈◈ mismatch can cause variations in the base currents and base voltages, as well as smaller changes in emitter currents and collector potentials. It is desirable then to quantify the circuit's performance so that we can see just how well balanced it is. We can judge a diff amp's DC performance by measuring its input offset current and its input and output offset voltages. In simple terms, the difference between the two base currents is the input offset current. The difference between the two collector voltages is the output offset voltage. The DC potential required at one of the bases to counteract the output offset voltage is called the input offset voltage (this is little more than the output offset voltage divided by the DC gain of the amplifier). In an ideal diff amp all three of these factors are equal to 0. We will take a much closer look at these parameters and how they relate to operational amplifiers in later chapters. For now, it is only important that you understand that these inaccuracies exist, and what can cause them.

AC ANALYSIS

Figure 7.6.4 shows a typical circuit with input and output connections. In order to minimize confusion with the DC circuit, AC equivalent values will be shown in lower case. Small emitter degeneration resistors, ◈1 and ◈2, have been added to this

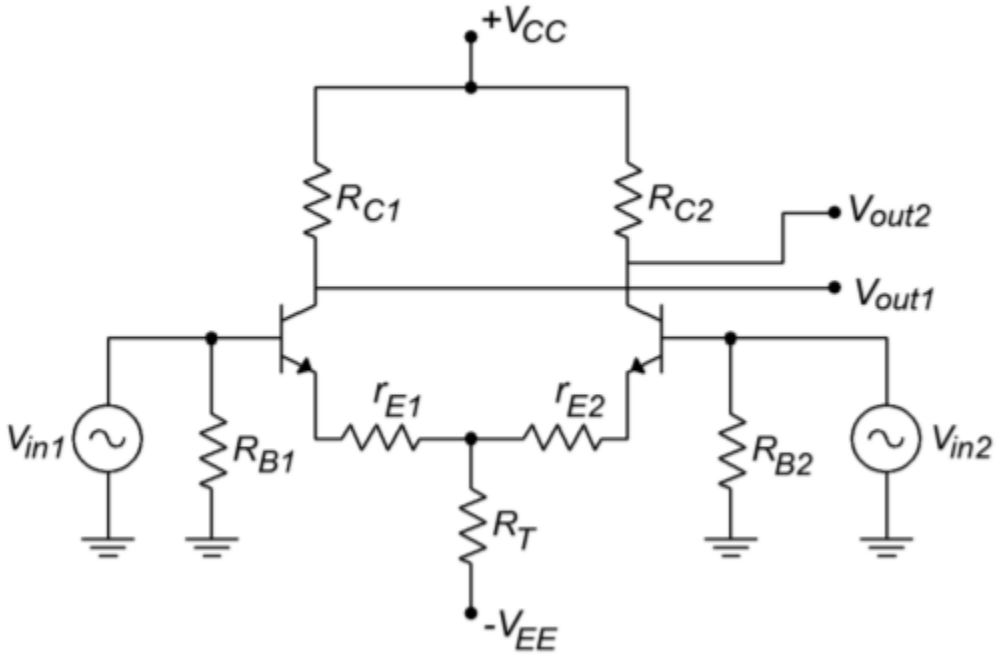
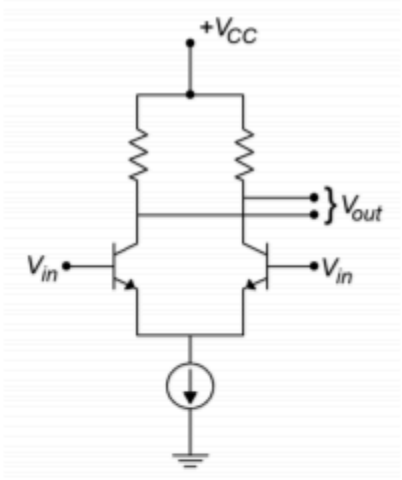


Figure 7.6.4 : A typical diff amp with input and output connections.

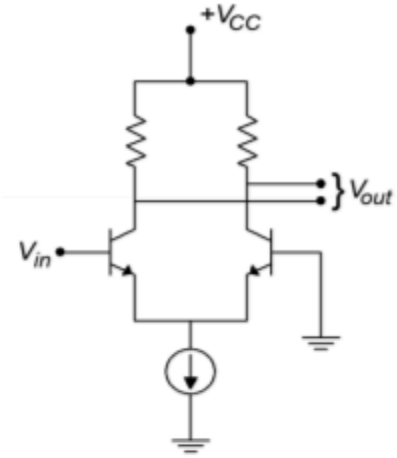
diff amp. This circuit has two signal inputs and two signal outputs. It is possible to configure a diff amp so that only a single input and/or output is used. This means that there are four variations on the theme:

- Differential (also called dual- or double-ended) input, differential output.
- Differential input, single-ended output.
- Single-ended input, differential output.
- Single-ended input, single-ended output.

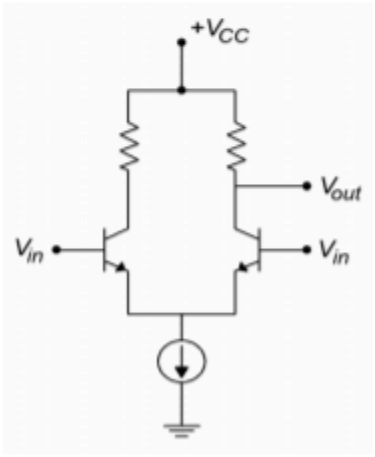
These variations are shown in Figure 7.6.5 . For use in operational amplifiers, the differential input/single-ended output variation is the most common. We will examine the most general case, the differential input/differential output version.



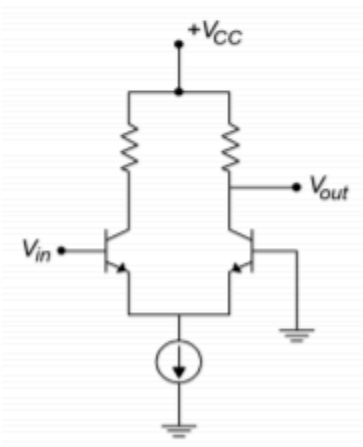
a. Differential input and output



b. Single-ended input and differential output



c. Differential input and single-ended output



d. Single-ended input and output

Because the diff amp is a linear circuit, we can use the principle of Superposition to independently determine the output contribution from each of the inputs. Utilizing the circuit of Figure 7.6.4 , we will first determine the gain Equation from ◈◈1 to either output. To do this, we replace ◈◈2 with a short circuit. The AC equivalent circuit is shown in Figure 7.6.6 .

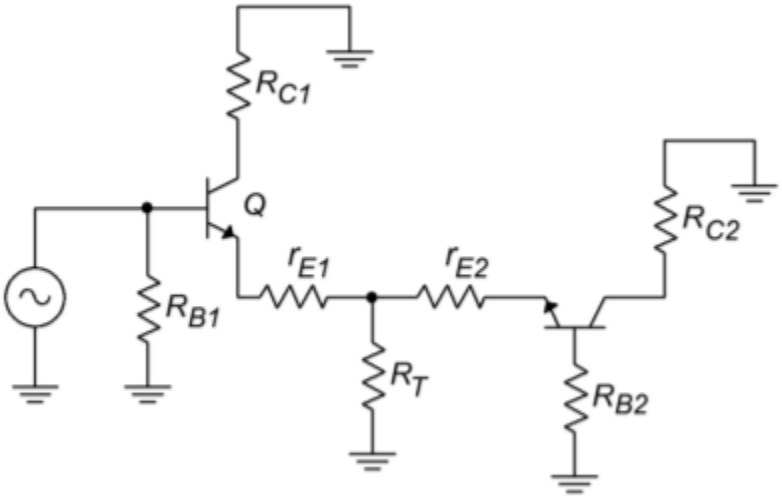


Figure 7.6.6 : The circuit of Figure 7.6.4 redrawn for AC analysis.

For the output on collector 1, transistor 1 forms the basis of a common emitter amplifier. The voltage across ◈◈1 is found via Ohm's Law.

$$v_{r_{C1}} = -i_{C1}\ r_{C1}$$

The negative sign comes from the fact that AC ground is used as our reference. (i.e., for a positive input, current flows from AC ground down through ◈◈1 , and into the collector.) To a reasonable approximation, we can say that the collector and emitter currents are identical.

$$v_{r_{cl}} = -i_{EI}\ r_{C1}$$

We must now determine the AC emitter current in relation to ◈◈◈1 . In order to better visualize the process, the circuit of Figure 7.6.6 is altered to include simplified transistor models, as shown in Figure 7.6.7 .

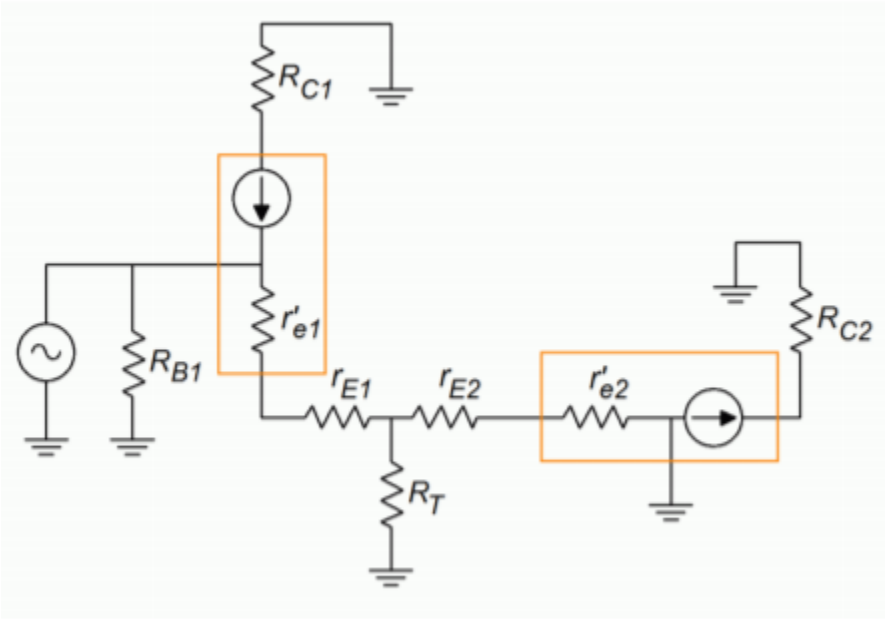


Figure 7.6.7 : AC analysis.

◈′◈ is the dynamic resistance of the base emitter junctions and is inversely proportional to the DC emitter current. You may recall the following Equation from your prior course work:

$$r_e^{'} = \frac{26\ mV}{I_E}$$

Where

◈′◈ is the dynamic base-emitter junction resistance,

◈◈ is the DC emitter current.

For typical circuits, the values of ◈′◈ and ◈◈ are much smaller than the tail current biasing resistor, ◈◈ . Because of its large size, we can ignore the parallel effect of ◈◈ . By definition, the AC emitter current must equal the AC emitter potential divided by the AC resistance in the emitter section. If you trace the signal flow from the base of transistor 1 to ground, you find that it passes through ◈′◈1 , ◈◈1 , ◈′◈2 and ◈◈2 . You will also notice that the magnitude of ◈◈1 is the same as ◈◈2 , although they are out of phase.

$$i_E = \frac{v_{in1}}{r_{e1}^{'} + r_{E1} + r_{e2}^{'} + r_{E2}}$$

Because the circuit values should be symmetrical for best performance, this Equation may be simplified to

$$i_E = \frac{v_{in}}{2(r_e^{'} + r_E)}$$

If we now solve for voltage gain,

$$A_v = -\frac{v_{out}}{v_{in}}$$

$$A_v = \frac{-i_E\ r_c}{v_{in}}$$

$$A_v = \frac{\frac{v_{in}}{2(r_e^{'}+r_E)} r_c}{v_{in}}$$

$$A_v = \frac{-r_c}{2(r_e^{'} + r_E)}$$

Where

◈◈ is the voltage gain,

◈◈ is the AC equivalent collector resistance,

◈◈ is the AC equivalent emitter resistance,

◈′◈ is the dynamic base-emitter junction resistance.

The final negative sign indicates that the collector voltage at transistor number 1 is 180 degrees out of phase with the input signal. Earlier, we noted that ◈◈2 is the same magnitude as ◈◈1 , the only difference being that it is out of phase. Because of this, the magnitude of the collector voltage at transistor number 2 will be the same as that on the first transistor. Because the second current is out of phase with the first, it follows that the second collector voltage must be out of phase with the first. This means that the voltage at the second collector is in phase with the first input signal. Its gain Equation is

$$A_v = \frac{r_c}{2(r_e' + r_E)}$$

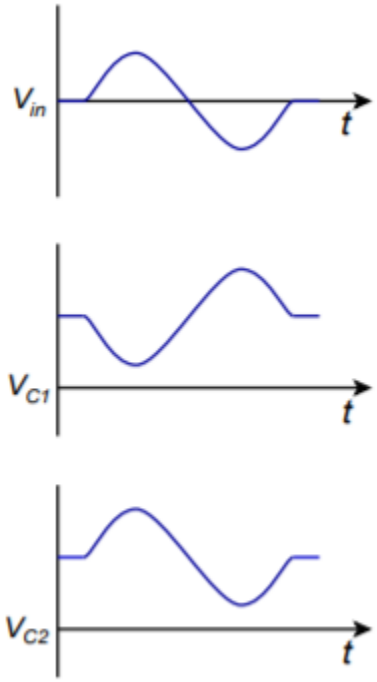


Figure 7.6.8 : Waveforms for a single input.

The various waveforms are depicted in Figure 7.6.8 . The preceding Equation is often referred to as the single-ended input/single-ended output gain Equation because it describes the single change from one input to one output. The output signal will be in phase if we are examining the opposite transistor, and out of phase if we are looking at the input transistor. Because the circuit is symmetrical, we will get similar results when we examine the second input. The voltage between the two collectors is 180 degrees apart. If we were to use a differential output, that is, derive the output from collector to collector rather than from one collector to ground, we would see an effective doubling of the output signal. If the reason for this is not clear to you, consider the following. Assume that each collector has a 1 V peak sine wave riding on it. When collector 1 is at +1 V, collector 2 is at -1 V, making +2 V total. Likewise, when collector 1 is at its negative peak, collector 2 is at its positive peak, producing a total of -2 V. The single ended input/differential output gain therefore is

$$A_v = \frac{r_c}{r_e' + r_E}$$

Example 7.6.2

Using the circuit of Figure 7.6.4 , determine the single-ended input/differential output and single-ended input/single-ended output voltage gains. Use the following component values: ⍰⍰⍰=15⍰ , ⍰⍰⍰=−8⍰ , ⍰⍰=10⍰Ω , ⍰⍰=8⍰Ω , ⍰⍰=30Ω . In order to find ⍰′⍰ we must find the DC current.

$$I_T = \frac{|V_{EE}| - 0.7V}{R_T}$$

$$I_T = \frac{7.3\ V}{10\ k\Omega}$$

$$I_T = 730\ \mu A$$

$$I_E = \frac{I_T}{2}$$

$$I_E = \frac{730\ \mu A}{2}$$

$$I_E = 365\ \mu A$$

$$r_e^{'} = \frac{26\ mV}{I_E}$$

$$r_e^{'} = \frac{26\ mV}{365\ \mu A}$$

$$r_e^{'} = 71.2\ \Omega$$

For the single ended output gain,

$$A_v = \frac{r_c}{2(r_e^{'} + r_E)}$$

$$A_v = \frac{8\ k\Omega}{2(71.2\ \Omega + 30\ \Omega)}$$

$$A_v = \frac{8\ k\Omega}{202.4\ \Omega}$$

$$A_v = 39.5$$

The differential output gain is twice this value, or 79.

Because it is possible to drive a diff amp with two distinct inputs, a wide variety of outputs may be obtained. It is useful to investigate two specific cases:

- Two identical inputs in both phase and magnitude.
- Two inputs with identical magnitude, but 180 degrees out of phase.

Let's consider the collector potentials for the first case. Assume that a diff amp has a single-ended input/single-ended output gain of 100 and a 10 mV signal is applied to both bases. Using

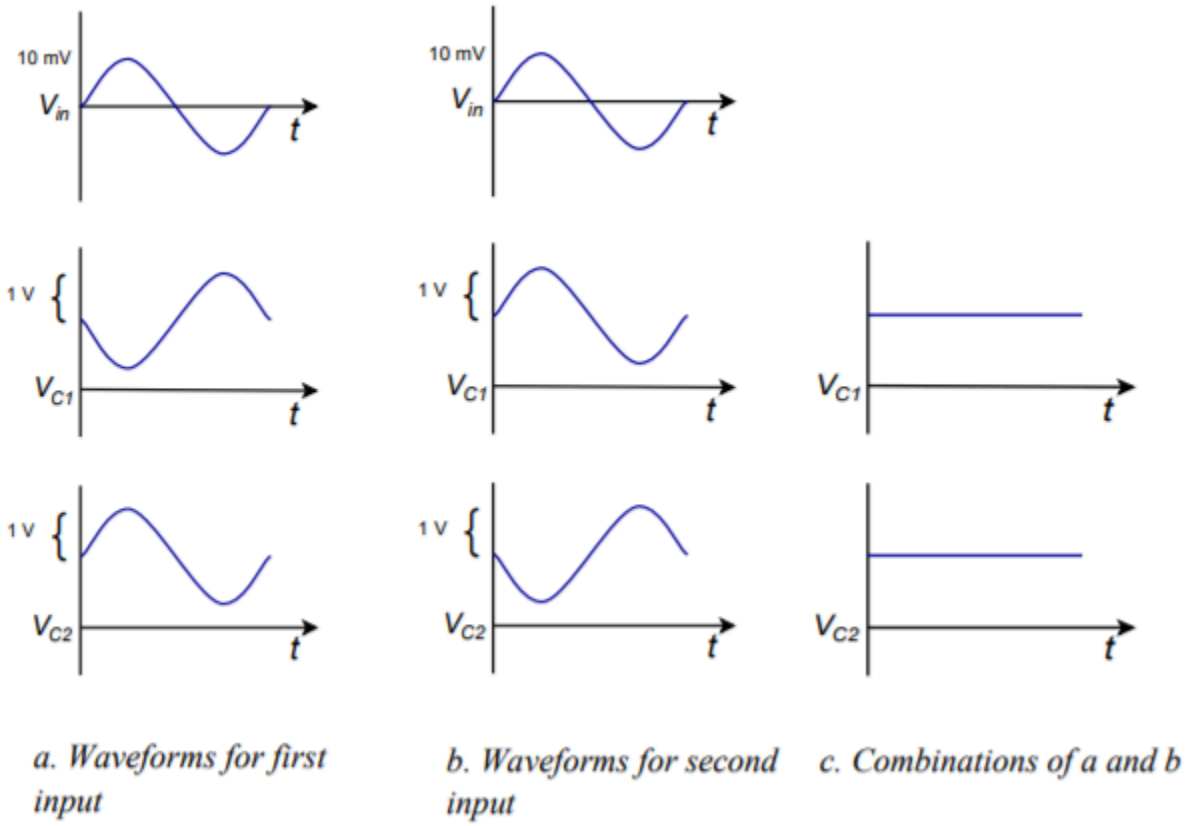
Superposition, we find that the outputs due to each input are 100 times 10 mV, or 1 V in magnitude. For the first input, the voltages are sketched in Figure 7.6.9◈ (following page). For the second input, the voltages are sketched in Figure 7.6.9◈ . Note that each collector sees both a sine wave and an inverted sine wave, both of equal amplitude. When these two signals are added, the result is zero, as seen in Figure 7.6.9◈. In Equation form,

$$v_{C1} = v_{in1}(-A_v) + v_{in2}\ A_v$$
$$v_{C1} = A_v(v_{in2} - v_{in1})$$

Because ◈◈◈1 and ◈◈◈2 are identical, the output is ideally zero given a perfectly matched and biased diff amp. The exact same effect is seen on the opposite collector. This last Equation is very important. It says that the output voltage is equal to the gain times the difference between the two inputs. This is how the differential amplifier got its name. In this case, the two inputs are identical, and thus their difference is zero. On the other hand, if we were to invert one of the input signals (case 2), we find a completely different result.

$$v_{in1} = -v_{in2}$$
$$v_{C1} = A_v(v_{in2} - v_{in1})$$
$$v_{C1} = A_v(v_{in2} - (-v_{in2}))$$
$$v_{C1} = 2\ A_v\ v_{in2}$$

Thus, if one input is inverted, the net result is a doubling of gain. This effect is shown graphically in Figures 7.6.9◈ through 7.6.9◈ . In short, a differential amplifier suppresses in phase signals while simultaneously boosting out of phase signals. This can be a very useful attribute, particularly in the area of noise reduction.



a. Waveforms for first input

b. Waveforms for second input

c. Combinations of a and b

Figure 7.6.9 : Input-output waveforms for common mode.

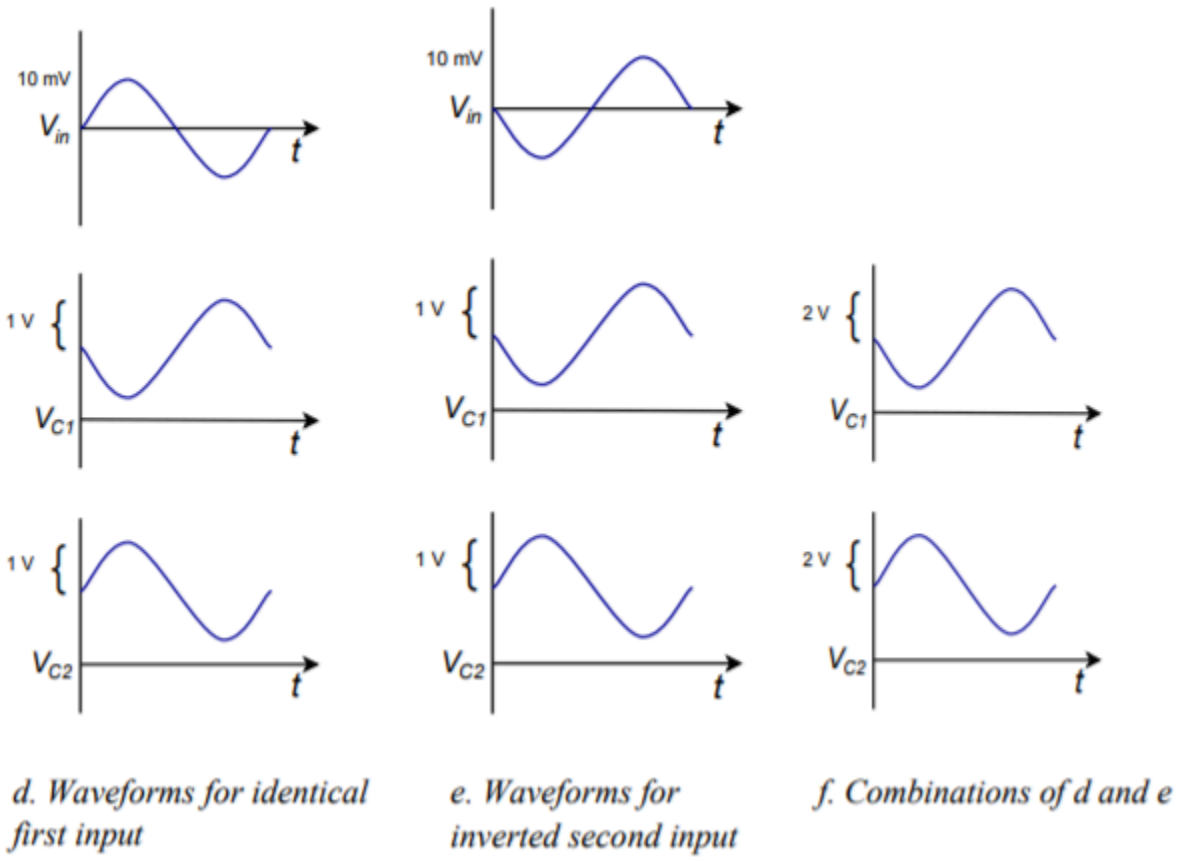


Figure 7.6.9 : (continued) Input-output waveforms for common mode.

COMMON MODE REJECTION

By convention, in phase signals are known as common-mode signals. An ideal differential amplifier will perfectly suppress these common-mode signals, and thus, its common-mode gain is said to be zero. In the real world, a diff amp will never exhibit perfect common-mode rejection. The common-mode gain may be made very small, but it is never zero. For a common-mode gain of zero, the two halves of the circuit have to be perfectly matched, and all circuit elements must be ideal. This is impossible to achieve as errors may arise from several sources. The most obvious error sources are resistor tolerance variations and transistor parameter spreads. The basic design of the circuit will also affect the common-mode gain. With some circuit rearrangements, it is possible to determine a common-mode gain for the circuits we have been using. The circuit of Figure 7.6.4 has been redrawn in Figure 7.6.10 in order to emphasize its parallel symmetry.

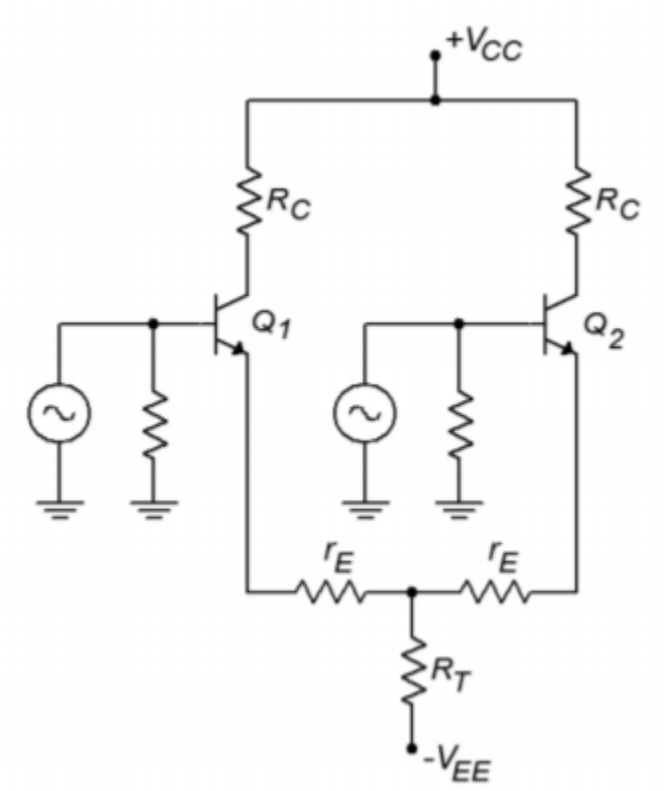


Figure 7.6.10 : The circuit of Figure 7.6.4 redrawn for common mode rejection ratio (CMRR) analysis.

Because the DC potentials are identical in both halves, and identical signals drive both inputs, we can combine resistors in parallel in order to arrive at the circuit of Figure 7.6.11 .

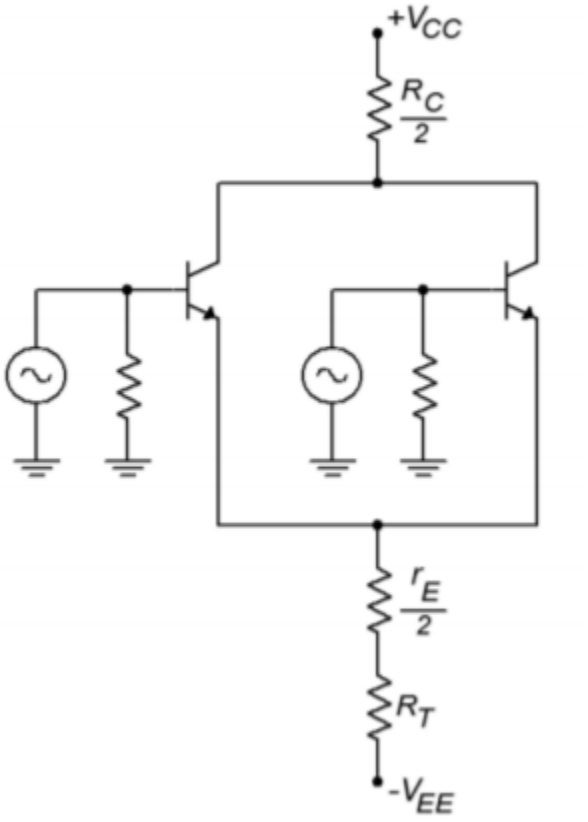


Figure 7.6.11 : Common mode gain analysis.

Although it is not shown explicitly on the diagram, the internal dynamic resistances (r'_e) may also be combined ($r'_e/2$). This circuit has been effectively reduced to a simple common emitter stage. Based on our earlier work, the gain for this circuit is

$$A_{v(cm)} = \frac{\frac{r_C}{2}}{R_T + \frac{r'_e}{2} + \frac{r_E}{2}}$$

This is the common-mode voltage gain. If r_e is considerably larger than r_E , then this circuit will exhibit good common mode rejection (assuming that the other parts are matched, naturally). r_E is the effective resistance of the tail current source. A very high internal resistance (i.e., an ideal current source) is desirable. There are many ways of creating a more ideal current source. One way is to use a third bipolar transistor as shown in Figure 7.6.12 .

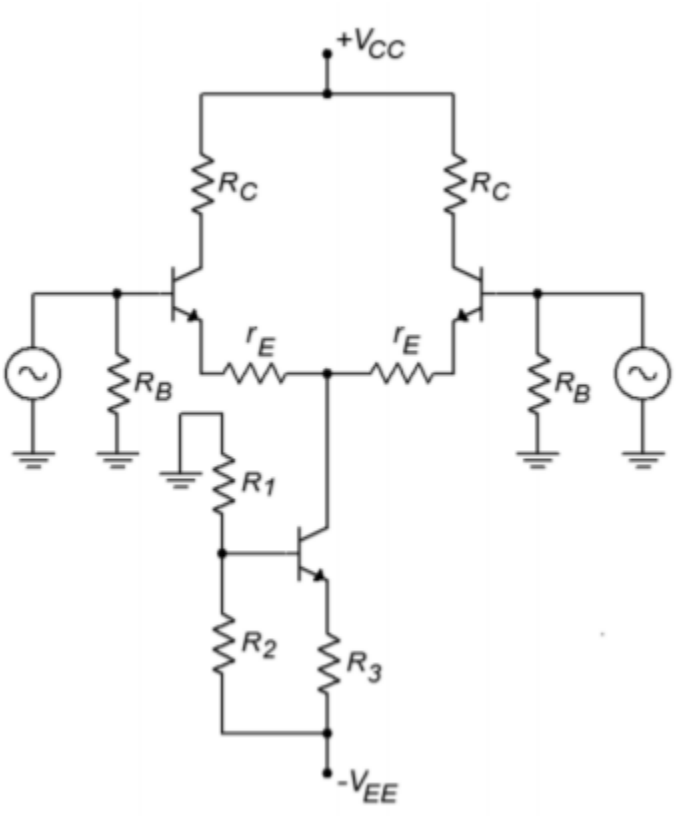


Figure 7.6.12 : Improved current source.

The tail current is found by determining the potential across ◈2 and subtracting the 0.7 V ◈◈◈ drop. The remaining potential appears across ◈3 . Given the voltage and resistance, Ohm's Law will let you find the tail current. In this circuit, ◈2 is sometimes replaced with a Zener diode. This can help to reduce temperature induced current fluctuations. In any case, the effective resistance of this current source is considerably larger than the simple tail resistor variation. It is largely dependent on the characteristics of the tail current transistor, and can easily be in the megohm region.

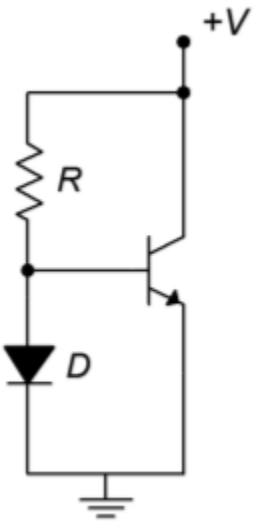


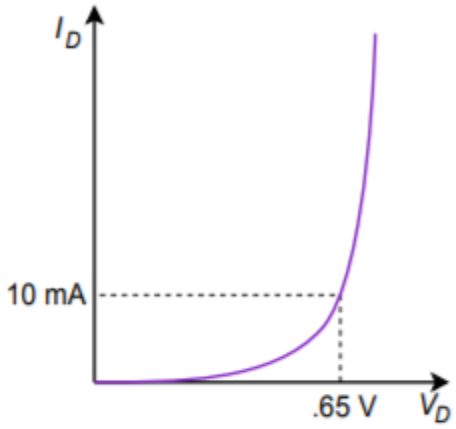
Figure 7.6.13 : Current mirror.

CURRENT MIRROR

A very popular biasing technique in integrated circuits involves the current mirror. Current mirrors are also employed as active loads in order to optimize a circuit's gain. A simple current mirror is shown in Figure 7.6.13 . This circuit requires that the transconductance curves of the diode and the transistor be very closely matched. One way to guarantee this is to use two transistors, and form one of them into a diode by shorting its collector to its base. If we use an approximate forward bias potential of 0.7 V and ignore the small base current, the current through the diode is

$$I_D = \frac{V_{CC} - 0.7\ V}{R}$$

In reality, the diode potential will probably not be exactly 0.7 V. This will have little effect on ◈◈ though. Because the diode is in parallel with the transistor's base-emitter junction, we know that ◈◈=◈◈◈ . If the two devices have identical transconductance curves, the transistor's emitter current will equal the diode current. You can think of the transistor as mirroring the diode's current, hence the circuit's name. If the two device curves are slightly askew, then the two currents will not be identical. This is shown graphically in Figure 7.6.14 .



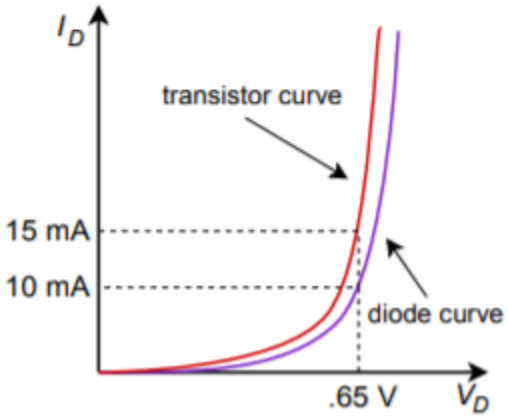


Figure 7.6.14 : Transfer curve mismatch.

Example 7.6.3

A current mirror could be used in the circuit of Figure 7.6.12 . The result is shown in Figure 7.6.15 . If the positive power supply is 15 V, the negative supply is -10 V, and ◈ is 10 k\Omega, the tail current will be

$$I_D = \frac{V_{CC} - V_{EE} - V_D}{R}$$

$$I_D = \frac{15\ V - (-10\ V) - 0.7\ V}{10k\Omega}$$

$$I_D = 2.43\ mA$$

Because the tail current is the mirror current,

$$I_T = I_D$$
$$I_T = 2.43\ mA$$

Biasing of this type is very popular in operational amplifiers. Another use for current mirrors is in the application of active loads. Instead of using simple resistors for the collector loads, a current mirror may be used instead. A PNP based current mirror suitable for use as an active load in our previous circuits is shown in Figure 7.6.16 .

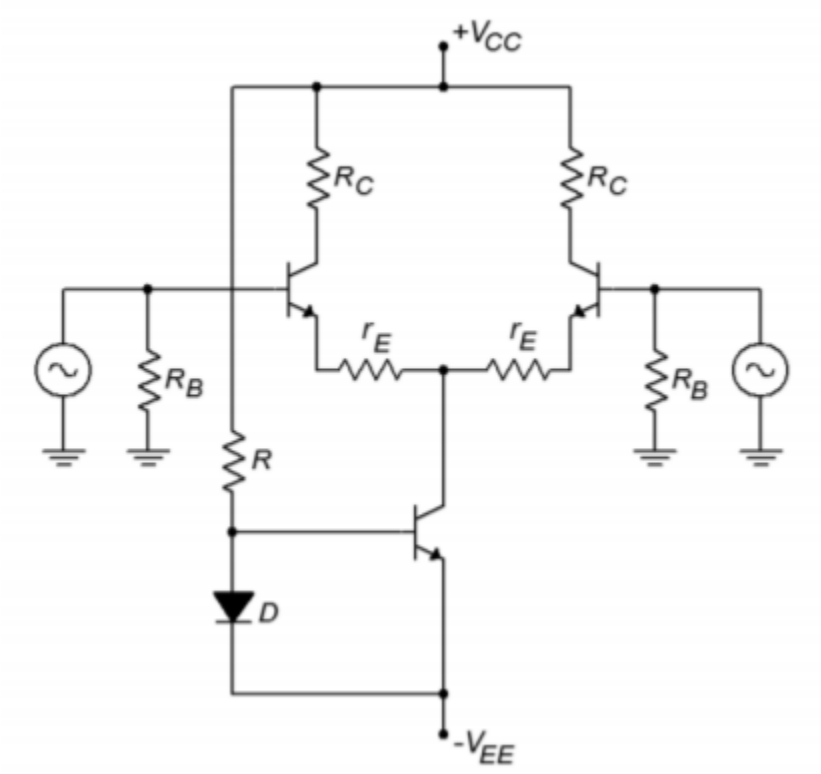


Figure 7.6.15 : Current mirror bias.

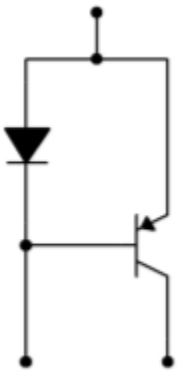


Figure 7.6.16 : Active load current mirror.

To use this, we simply remove the two collector resistors from a circuit such as Figure 7.6.15 , and drop in the current mirror. The result of this operation is shown in Figure 7.6.17 . The current mirror active load produces a very high internal impedance, thus contributing to a very high differential gain. In effect, by using a constant current source in the collectors, all AC current is forced into the following stage. You may also note that the number of resistors used in the circuit has decreased considerably.

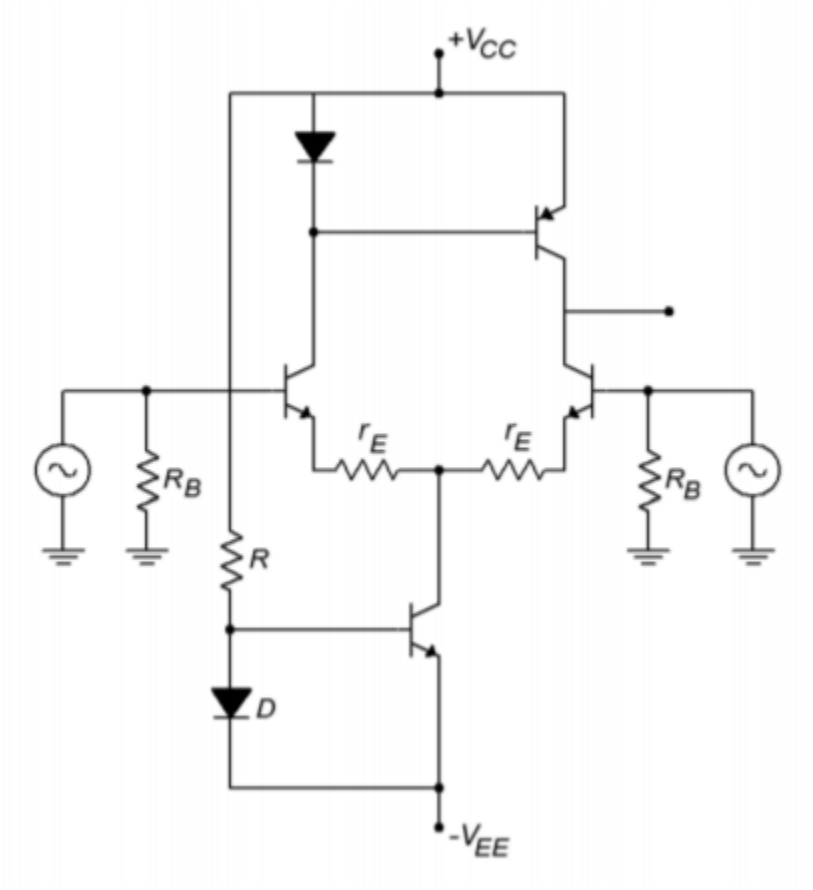


Figure 7.6.17 : Current mirrors for bias and active load.

7.7 SUMMARY

We have seen how to convert gains and signals into a decibel form for both powers and voltages. This is convenient because what would require multiplication and division under the ordinary scheme only requires simple addition and subtraction in the dB scheme. Along with this, dB measurement is used almost exclusively for Bode gain plots. A Bode plot details a system's gain magnitude and phase response. For gain, the amplitude is measured in dB, while the frequency is normally presented in log form. For a phase plot, phase is measured in degrees, and again, the frequency axis is logarithmic. The changes in gain and phase at the frequency extremes are caused by lead and lag networks. Lead networks cause the low frequency gain to roll off. The roll off rate is 6 dB per octave per network. The phase will change from +90 degrees to 0 degrees per network. Lag networks cause the high frequency gain to roll off at a rate of -6 dB per octave per network. The phase change per lag network is from 0 degrees to -90 degrees. It was noted that computers may be used to quickly tabulate the response of complex networks such as these. Many computer circuit simulators are based on the original SPICE program. Among the popular commercial simulation programs are PSpice and Multisim. Freeware versions include LTspice and TINA-TI. Packages such as these offer graphical schematic capture tools and large component libraries.

Differential amplifiers are symmetrical circuits, employing a minimum of two active devices. They may be configured with single or dual inputs, and single or dual outputs. Diff amps are commonly used as the first stage of an operational amplifier. They tend to amplify the difference in the input signals while simultaneously suppressing in-phase, or common-mode, signals. Current mirrors are widely used for biasing purposes and as active loads. Active loads offer the advantage of producing higher gains than ordinary resistive loads.

7.8 PROBLEMS

REVIEW QUESTIONS

1. What are the advantages of using decibels over the ordinary scheme?
2. How do decibel power and voltage gain calculations differ?
3. What does the third letter in a decibel-based signal measurement indicate (as in dBV or dBm)?
4. What is a Bode plot?
5. What is a lead network? What general response does it yield?
6. What is a lag network? What general response does it yield?
7. What do the terms $\diamondsuit_1$ and $\diamondsuit_2$ indicate about a system's response?
8. What are the rolloff slopes for lead and lag networks?
9. What are the phase changes produced by individual lead and lag networks?
10. How is rise time related to upper break frequency?
11. How do multiple lead or lag networks interact to form an overall system response?
12. What is SPICE?
13. What is common-mode rejection?
14. What is a current mirror?
15. What is the advantage of using an active load?

PROBLEMS

Analysis Problems – dB emphasis

1. Convert the following power gains into dB form:
    a. 10
    b. 80
    c. 500
    d. 1
    e. 0.2
    f. 0.03.
2. Convert the following dB power gains into ordinary form:
    a. 0 dB

b. 12 dB

c. 33.1 dB

d. 0.2 dB

e. -5.4 dB

f. -20 dB

3. An amplifier has an input signal of 1 mW, and produces a 2 W output. What is the power gain in dB?
4. A hi-fi power amplifier has a maximum output of 50 W and a power gain of 19 dB. What is the maximum input signal power?
5. An amplifier with a power gain of 27 dB is driven by a 25 mW source. Assuming the amplifier doesn't clip, what is the output signal in watts?
6. Convert the following voltage gains into dB form:
    a. 10
    b. 40
    c. 250
    d. 1
    e. 0.5
    f. 0.004
7. Convert the following dB voltage gains into ordinary form:
    a. 0.5 dB
    b. 0 dB
    c. 46 dB
    d. 10.7 dB
    e. -8 dB
    f. -14.5 dB
8. A guitar pre-amp has a gain of 44 dB. If the input signal is 12 mV, what is the output signal?
9. A video amplifier has a 140 mV input and a 1.2 V output. What is the voltage gain in dB?
10. The pre-amp in a particular tape deck can output a maximum signal of 4 V. If this amplifier has a gain of 18 dB, what is the maximum input signal?
11. Convert the following powers into dBW:
    a. 1 W
    b. 23 W
    c. 6.5 W
    d. 0.2 W

e. 2.3 mW

f. 1.2 kW

g. 0.045 mW

h. 0.3μW

i. 5.6E-18W

12. Repeat Problem 11 for units of dBm.
13. Repeat Problem 11 for units of dBf.
14. Convert the following voltages into dBV:

    a. 12.4 V

    b. 1 V

    c. 0.25 V

    d. 1.414 V

    e. 0.1 V

    f. 10.6 kV

    g. 13 mV

    h. 2.78 μV

15. A two stage power amplifier has power gains of 12 dB and 16 dB. What is the total gain in dB and in ordinary form?
16. If the amplifier of Problem 15 has an input of –18 dBW, what is the final output in dBW? in dBm? in watts?
17. Referring to Figure 1.2.1, what are the various stages' outputs if the input is changed to –4 dBm? to –34 dBW?
18. Which amplifier has the greatest power output?

    a. 50 watts

    b. 18 dBW

    c. 50 dBm

19. Which amplifier has the greatest power output?

    a. 200 mW

    b. -10 dBW

    c. 22 dBm

20. A three stage amplifier has voltage gains of 20 dB, 5 dB, and 12 dB respectively. What is the total voltage gain in dB and in ordinary form?
21. If the circuit of Problem 20 has an input voltage of –16 dBV, what are the outputs of the various stages in dBV? In volts?

22. Repeat Problem 21 for an input of 12 mV.
23. Which amplifier produces the largest output voltage?
    a. 15 V
    b. 16 dBV

Analysis Problems – Bode plot emphasis

24. Given a lead network critical at 3 kHz, what are the gain and phase values at 100 Hz, 3 kHz, and 40 kHz?
25. Given a lag network tuned to 700 kHz, what are the gain and phase values at 50 kHz, 700 kHz, and 10 MHz? What is the rise time?
26. A noninverting amplifier has a midband voltage gain of 18 dB and a single lag network at 200 kHz. What are the gain and phase values at 30 kHz, 200 kHz, and 1 MHz. What is the rise time?
27. Repeat Problem 26 for an inverting (–180 degrees) amplifier.
28. Draw the Bode plot for the circuit of Problem 26.
29. Draw the Bode plot for the circuit of Problem 27.
30. An inverting (–180 degrees) amplifier has a midband gain of 32 dB and a single lead network critical at 20 Hz (assume the lag network ◈◈ is high enough to ignore for low frequency calculations). What are the gain and phase values at 4 Hz, 20Hz, and 100 Hz?
31. Repeat Problem 29 with a noninverting amplifier.
32. Draw the Bode plot for the circuit of Problem 30.
33. Draw the Bode plot for the circuit of Problem 31.
34. A noninverting amplifier used for ultrasonic applications has a midband gain of 41 dB, a lag network critical at 250 kHz, and a lead network critical at 30 kHz. Draw its gain Bode plot.
35. Find the gain and phase at 20 kHz, 100 kHz, and 800 kHz for the circuit of Problem 34.
36. If the circuit of Problem 34 has a second lag network added at 300 kHz, What are the new gain and phase values at 20 kHz, 100 kHz, and 800 kHz?
37. Draw the gain Bode plot for the circuit of Problem 36.
38. What are the maximum and minimum phase shifts across the entire frequency spectrum for the circuit of Problem 36?
39. A noninverting DC amplifier has a midband gain of 36 dB, and lag networks at 100 kHz, 750 kHz, and 1.2 MHz. Draw its gain Bode plot.
40. What are the maximum and minimum phase shifts across the entire frequency spectrum for the circuit of Problem 39?
41. What is the maximum rate of high frequency attenuation for the circuit of Problem 39 in dB/decade?
42. If an amplifier has two lead networks, what is the maximum rate of low frequency attenuation

in dB/octave?

Analysis Problems – Differential amplifier emphasis

43. Given the circuit of Figure 1.6.4, determine the single-ended input/singleended output gain for the following values: ◇◇=5◇Ω, ◇◇=7.5◇Ω, ◇◇=12◇Ω, ◇◇◇=25◇, ◇◇◇=–9◇, ◇◇=50Ω .
44. Determine the differential voltage gain in the circuit of Figure 1.6.13 if ◇◇=15◇Ω, ◇1=5◇Ω, ◇2=7◇Ω, ◇3=10◇Ω,◇◇=20◇Ω, ◇◇◇=22◇, ◇◇◇=–12◇, ◇◇=75Ω .
45. For the circuit of Problem 44, determine the output at collector 2 if ◇◇◇1(◇)=0.001 sin2◇1000◇ and ◇◇◇2(◇)=–0.001 sin2◇1000◇ .
46. Determine the differential voltage gain in the circuit of Figure 1.6.15 if ◇◇=8◇Ω, ◇◇◇◇◇◇◇=22◇Ω, ◇◇=10◇Ω, ◇◇◇=18◇, ◇◇◇=–15◇, ◇◇=25Ω .
47. For the circuit of Problem 46, determine the output at collector 1 if ◇◇◇1(◇)=–0.005 sin2◇2000◇ and ◇◇◇2(◇)=0.005 sin2◇2000◇ .
48. Determine the tail and emitter currents in the circuit of Figure 1.6.17 if ◇◇=6◇Ω, ◇◇◇◇◇◇◇=50◇Ω, ◇◇◇=15◇, ◇◇◇=–15◇, ◇◇=0Ω .

Challenge Problems

49. You would like to use a voltmeter to take dBm readings in a 600 Ω system. What voltage should produce 0 dBm?
50. Assuming that it takes about an 8 dB increase in sound pressure level in order to produce a sound that is subjectively "twice as loud" to the human ear, can a hi-fi using a 100 W amplifier sound twice as loud as one with a 40 W amplifier (assuming the same loudspeakers)?
51. Hi-fi amplifiers are often rated with a "headroom factor" in dB. This indicates how much extra power the amplifier can produce for short periods of time, over and above its nominal rating. What is the maximum output power of a 250 W amplifier with 1.6 dB headroom?
52. If the amplifier of Problem 34 picks up an extraneous signal that is a –10 dBV sine wave at 15 kHz, what is the output?
53. If the amplifier of Problem 39 picks up a high frequency interference signal at 30 MHz, how much is it attenuated over a normal signal? If this input signal is measured at 2 dBV, what should the output be?
54. If an amplifier has two lag networks, and both are critical at 2 MHz, is the resulting ◇2 less than, equal to, or greater than 2 MHz?
55. If an amplifier has two lead networks, and both are critical at 30 Hz, is the resulting ◇1 less than, equal to, or greater than 30 Hz?

Computer Simulation Problems

1. Use a simulator to plot the Bode gain response of the circuit in Problem 39.
2. Use a simulator to plot the Bode phase response of the circuit in Problem 34.

3. Use a simulation program to generate a Bode plot for a lead network comprised of a 1 k Ω resistor and a 100 nF capacitor.

UNIT 8: OPERATIONAL AMPLIFIER INTERNALS

Learning Objectives

After completing this chapter, you should be able to:

- Describe the internal layout of a typical op amp.
- Describe a simple op amp computer simulation model.
- Determine fundamental parameters from an op amp data sheet.
- Describe an op amp based comparator, and note where it might be used.
- Describe how integrated circuits are constructed.
- Define monolithic planar construction.
- Define hybrid construction.

8.1 INTRODUCTION TO OPERATIONAL AMPLIFIERS

In this chapter we introduce the fundamentals of the operational amplifier, or op amp, as it is commonly known. We will investigate the construction and usage of a typical general-purpose op amp. Specific along with generalized internal circuits and associated block diagrams are examined. An initial op amp data sheet interpretation is given as well. Toward the middle of the chapter, the first op amp circuit examples are presented. The chapter finishes with an explanation of semiconductor integration and construction techniques. After finishing this chapter, you should be familiar with the concepts of what an op amp is composed of, how it is manufactured, and a beginning idea of how it might be used in application circuit design.

8.2 WHAT IS AN OP AMP?

An operational amplifier is, in essence, a multi-stage high gain amplifier treated as a single entity. Normally, op amps have a differential input and a single-ended output. In other words, one input produces an inverted output signal, and the other input produces a noninverted output signal. Often, the op amp is driven from a bipolar power supply (i.e., two supplies, one positive and one negative). Just about any sort of active amplifying device may be used for the individual stages. Op amps can be made entirely from vacuum tubes or discrete bipolar transistors (and of course, they were made that way some years ago). The advances in semiconductor manufacture in the late 1960's and early 1970's eventually made it possible to miniaturize the required components and place the whole affair on a single silicon chip (hence the term, integrated circuit). Through common use, this is what is generally meant by the term op amp today.

As seen in Figure 8.2.1 , a typical op amp has at least five distinct connections; an inverting input (labeled "-"), a noninverting input (labeled "+"), an output, and positive and negative power supply inputs. These power supply connections are sometimes referred to as supply rails. Note that a ground connection is not directly given. Rather, a ground connection is implied through the other connections. This symbol and its associated connections are typical, but by no means absolute. There are a wide variety of devices available to the designer that offer such features as differential outputs or unipolar power supply operation. In any case, some form of triangle will be used for the schematic symbol.

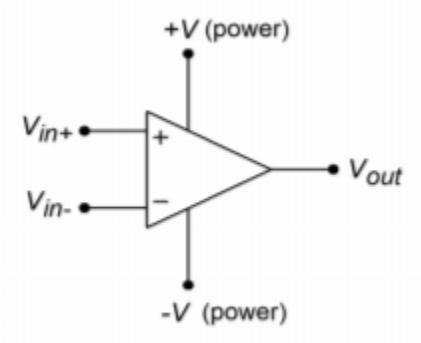


Figure 8.2.1 : General op amp symbol.

It is best to think of op amps as general-purpose building blocks. With them, you can create a wide variety of useful circuits. For general-purpose work, designing with op amps is usually much quicker and more economical than an all-discrete approach. Likewise, trouble shooting and packaging constraints may be lessened. For more demanding applications, such as those requiring very low noise, high output current and/or voltage, or wide bandwidth, manufacturers have created specialized op amps. The final word in performance today is still dominated by discrete circuit designs though. Also, it is very common to see a mix of discrete and integrated devices in a given circuit. There is certainly no law that states that op amps can only be used with other op amps. Often, a judicious mix

of discrete devices and op amps can produce a circuit superior to one made entirely of discretes or op amps alone.

Where might you find op amp circuits? In a word, anywhere. They're probably in use in your home stereo or TV where they help capture incoming signals, in electronic musical instruments where they can be used to create and modify tones, in a camera in conjunction with a light metering system, or in medical instruments where they might be used along with various bio-sensing devices. The possibilities are almost endless.

BLOCK DIAGRAM OF AN OP AMP

At this point you may be asking yourself, "what's inside of the op amp?" The generic op amp consists of three main functional stages. A real op amp may contain more than three distinct stages, but can be reduced to this level for analysis. A generalized discrete representation is given in Figure 8.2.2 . Since the op amp requires a differential input scheme, the first stage is most often a differential amplifier. As seen here, ◈1 and ◈2 comprise a PNP-based differential amplifier. The output of one collector (◈2 here) is then fed to a high-gain second stage. This stage usually includes a lag network capacitor that plays a major roll in setting the op amp's AC characteristics (this is examined further in Chapter Five). ◈3 makes up the second stage in the example. It is set in a common emitter configuration for both current and voltage gain. The a-fore mentioned lag capacitor is positioned across ◈3 's base-collector junction in order to take advantage of the Miller effect. The third and final section is a class B or class AB follower for the most effective load drive. ◈4 and ◈5 make up the final stage. The twin diodes compensate for the ◈4 and ◈5 ◈◈ drops, and produce a trickle bias current that minimizes distortion. This is a relatively standard class AB stage. Note that the entire circuit is direct-coupled. There are no lead networks, and thus the op amp can amplify down to zero Hertz (DC). There are many possible changes that may be seen in a real world circuit, including the use of Darlington pairs or FETs for the differential amplifier, multiple high gain stages, and output current limiting for the class B section.

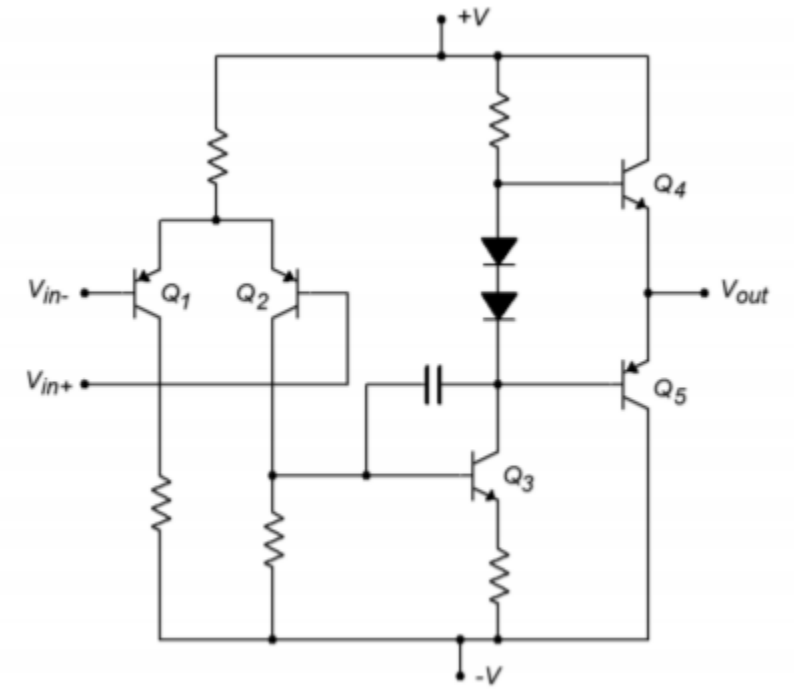


Figure 8.2.2 : General op amp schematic.

The example discrete circuit uses only five transistors and two diodes. In contrast, an integrated version may use two to three dozen active devices. Because of the excellent device matching abilities of single chip integration, certain techniques are used in favor of standard discrete designs. Internal IC current sources are normally made through the use of current mirrors. Current mirror configurations are also employed to create active loads, in order to achieve maximum circuit gain.

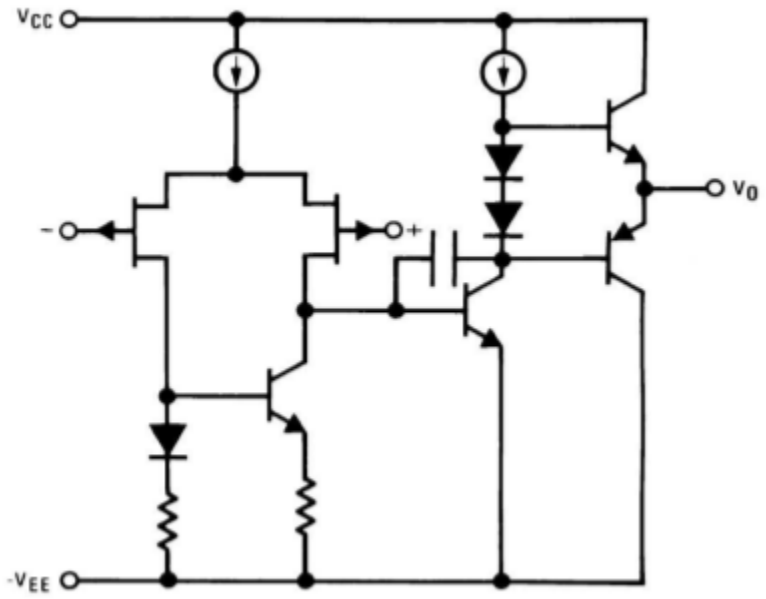


Figure 8.2.3 : LF411 simplified schematic. Reprinted courtesy of Texas Instruments

A typical integrated op amp will contain very few resistors, and usually only one or two lag network

capacitors. Due to size limitations and other factors, inductors are virtually unseen in these circuits. A simplified equivalent circuit of the LF411 op amp is shown in Figure 8.2.3 . Note that this device uses JFETs for the diff amp with an active load. The diff amp tail current source and the class AB trickle bias source are shown as simple current sources. In reality, they are a bit more complex, utilizing current mirror arrangements.

One of the most popular op amps over the years has been the 741. The specifications of this device seem rather lackluster by today's standards, but it was one of the first easy-to-use devices produced. As a result, it has found its way into a large number of designs. Indeed, it is still a wise choice for less demanding applications, or where parts costs are a major consideration. A complete schematic of the ◈ A 741 is shown in Figure 8.2.4 . Several different manufacturers make the 741. This version is manufactured by Signetics, and may be somewhat different than a 741 made by another company.[1] The circuit contains 20 active devices and about one dozen resistors.

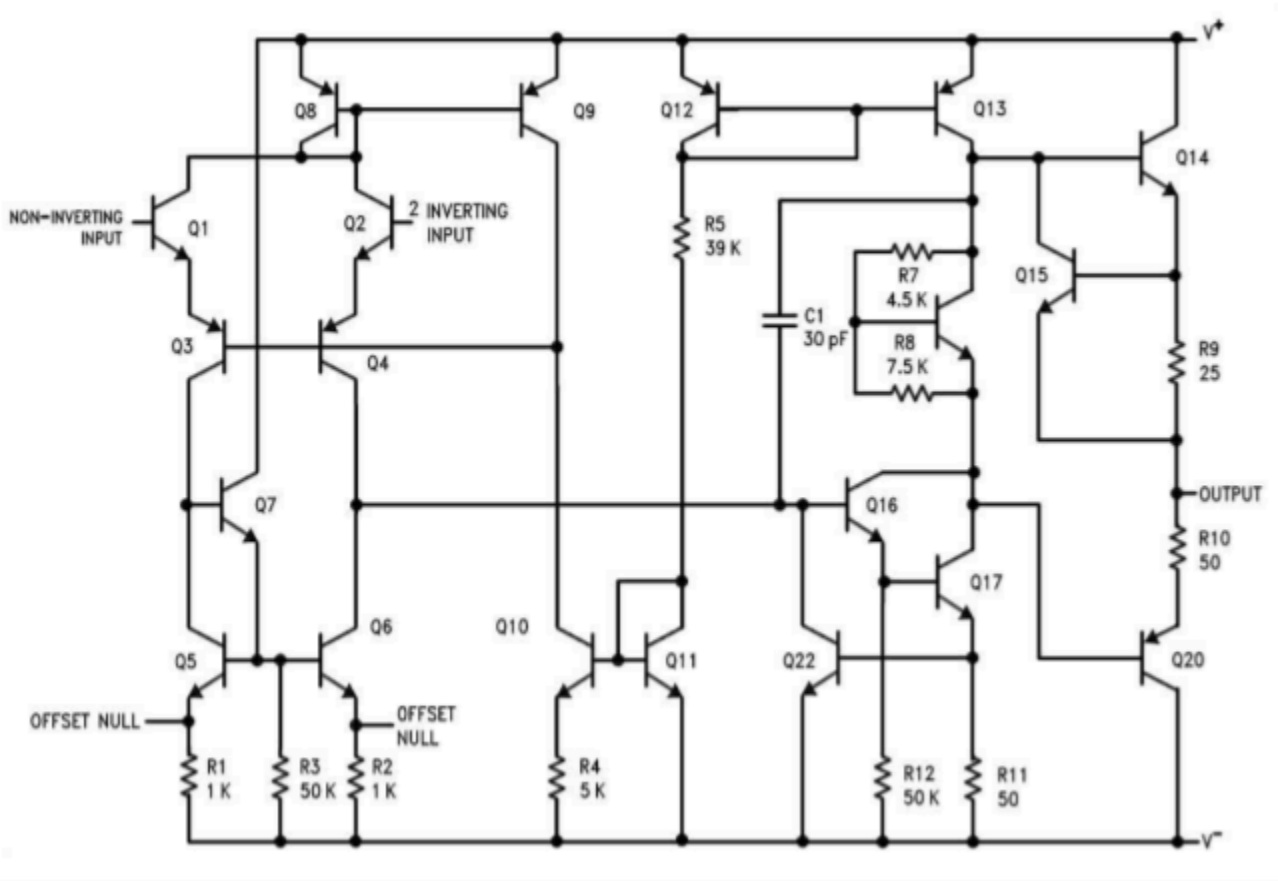


Figure 8.2.4 : Schematic of ◈ A741. Reprinted courtesy of Philips Semiconductors

At first glance, this circuit may look hopelessly confusing. A closer look reveals many familiar circuit blocks. First off, you will notice that a number of devices show a shorting connection between their base and collector terminals, such as ◈$_8$, ◈$_{11}$, and ◈$_{12}$. In essence, these are diodes (they are drawn this way because they are manufactured as transistor junctions. It is actually easier to make diodes in this fashion). For the most part, these diodes are part of current mirror biasing networks. The bias setup is found in the very center of the schematic, and revolves around ◈$_9$ through ◈$_{12}$. The setup

1. While the exact internal circuitry may be altered, the various manufacturers versions will have the same pinouts, and very similar performance specifications.

current is found by subtracting two diode drops (◈11,\(◈12) from the total power supply potential (◈+–◈–), and dividing the result by ◈5 . For a standard ±15 V power supply, this works out to

$$I_{bias} = \frac{V_+ - V_- - V_{BE-Q11} - V_{BE-Q12}}{R_5}$$

$$I_{bias} = \frac{30\ V - 1.4\ V}{39\ k\Omega}$$

$$I_{bias} = 733\mu A$$

This current is reflected into ◈13 . A close look at ◈10 and ◈11 reveals that this portion is not a simple current mirror. By including ◈4 , the voltage drop across the base emitter of ◈10 is decreased, thus producing a current less than 733◈◈ . This configuration is known as a Widlar current source. The derivation of the exact current Equation is rather involved, and beyond the scope of this chapter.[2] This current is reflected into ◈8 via ◈9 , and establishes the tail current for the differential amplifier. The diff amp stage uses a total of four amplifying transistors in a common-collector/common-base configuration (◈1 through ◈4). In essence, ◈1 and ◈2 are configured as emitter followers, thus producing high input impedance and reasonable current gain. ◈3 and ◈4 are configured as common base amplifiers, and as such, produce a large voltage gain. The gain is maximized by the active load comprised of ◈5 through ◈7. The output signal at the collector of ◈4 passes on to a dual transistor high-gain stage (◈16 and ◈17). ◈16 is configured as an emitter follower and buffers ◈17 , which is set as a common emitter voltage amplifier. Resistor ◈11 serves to stabilize both the bias and gain of this stage (i.e., it is an emitter degeneration or swamping resistor). ◈17 is directly coupled to the class AB output stage (◈14 and ◈20). Note the use of a◈◈◈ multiplier to bias the output transistors. The ◈◈◈multiplier is formed from ◈18 and resistors ◈7 and ◈8 . Note that this section receives its bias current from ◈13 which is part of the central current mirror complex.

Some transistors in this circuit are used solely for protection from overloads. A good example of this is ◈15. As the output current increases, the voltage across ◈9 will increase proportionally. Note that this resistor is in parallel with the base emitter junction of ◈15 . If this potential gets high enough, ◈15 will turn on, shunting base drive current around the output device (◈14). In this manner, current gain is reduced, and the maximum output current is limited to a safe value. This limiting value may be found via Ohm's Law:

$$I_{limit} = \frac{V_{BE}}{R_9}$$

$$I_{limit} = \frac{0.7\ V}{25\ \Omega}$$

$$I_{limit} = 28\ mA$$

In a similar fashion, ◈20 is protected by ◈10 , ◈11 , and ◈22 . If the output tries to sink too large of a current, ◈22 will turn on, shunting current away from the base of ◈16 . While individual op amp schematics will vary widely, they generally hold to the basic four-part theme presented here:

- A central current source/current mirror section to establish proper bias.
- A differential amplifier input stage with active load.

2. A complete derivation of the Widlar current source may be found in Principles of Electronic Circuits, by S.G.Burns and P.R.Bond, 1987, West Publishing Company.

- A high voltage gain intermediate stage.
- A class B or AB follower output section.

Fortunately for the designer or repair technician, intimate knowledge of a particular op amp's internal structure is usually not required for successful application of the device. In fact, a few simple models can be used for the majority of cases. One very useful model is given in Figure 8.2.5 .

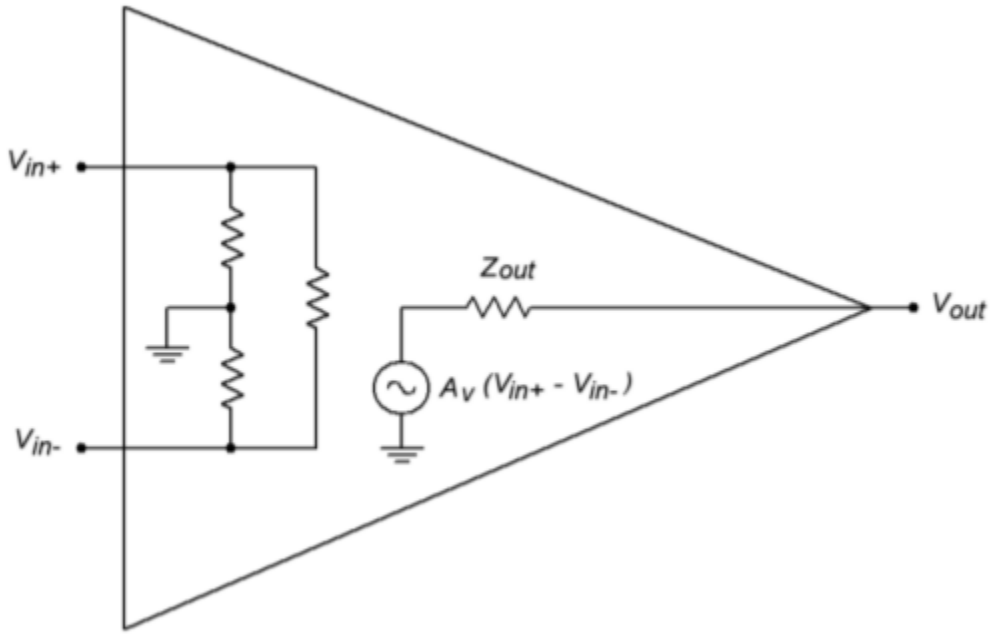


Figure 8.2.5 : Simplified model.

Here the entire multistage op amp is modeled with a simple resistive input network, and a voltage source output. This output source is a dependent source. Specifically, it is a voltage-controlled voltage source. The value of this source is

$$E_{out} = A_v\ (V_{in+} - V_{in-})$$

The input network is specified as a resistance from each input to ground, as well as an input-to-input isolation resistance. For typical op amps these values are normally hundreds of kilo-ohms or more at low frequencies. Due to the differential input stage, the difference between the two inputs is multiplied by the system gain. This signal is presented to the output terminal through the final stage's output impedance. The output impedance will most likely be less than 100 Ω . System voltage gains in excess of 80 dB (10,000) are the norm.

A SIMPLE OP AMP SIMULATION MODEL

It is possible to create a great variety of simulation models for any given op amp. Generally speaking, the more accurate the model is, the more likely it is to be complex. Due to the nature of most simulators, a more complex model requires a greater amount of time for an analysis to be completed. There is always a trade-off between model complexity and computation time. We can create a very simple model based on the previous section. This model is shown in Figure 8.2.6 .

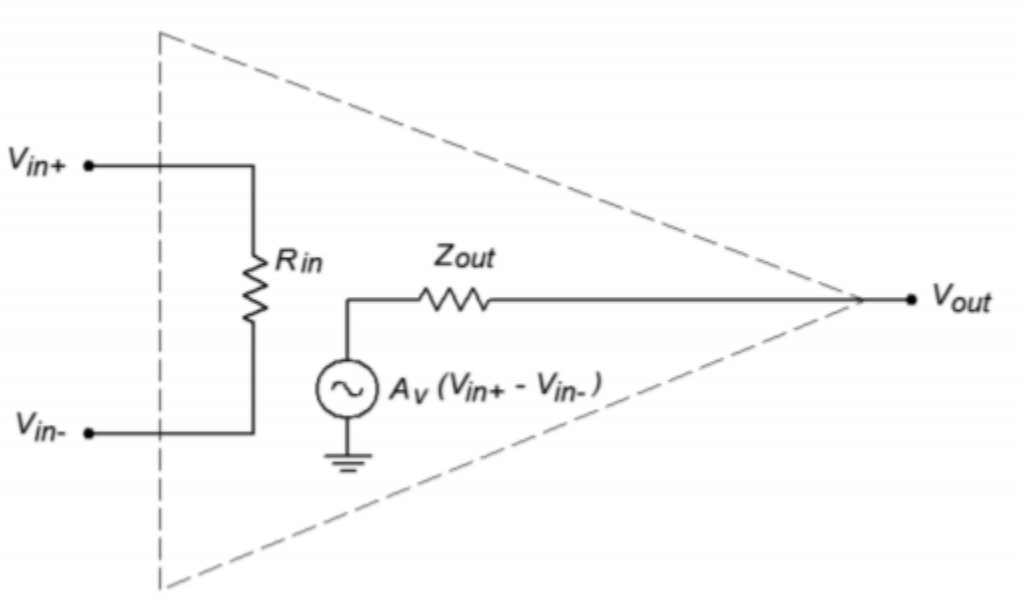


Figure 8.2.6 : Simplified SPICE model.

It consists of just five nodes. The input section is modeled as a single resistor, ◈◈◈ , between nodes 1 and 8. These two nodes are the noninverting and inverting inputs of the op amp, respectively. The second half of the model consists of a voltage controlled voltage source and an output resistor. The value of this dependent source is a function of the differential input voltage and the voltage gain. With a minimum of components, the simulation time for this model is very low. In order to use this model, you need only set three parameters, the input resistance, the output resistance, and the voltage gain. An example is shown in Figure 8.2.7 using Multisim.

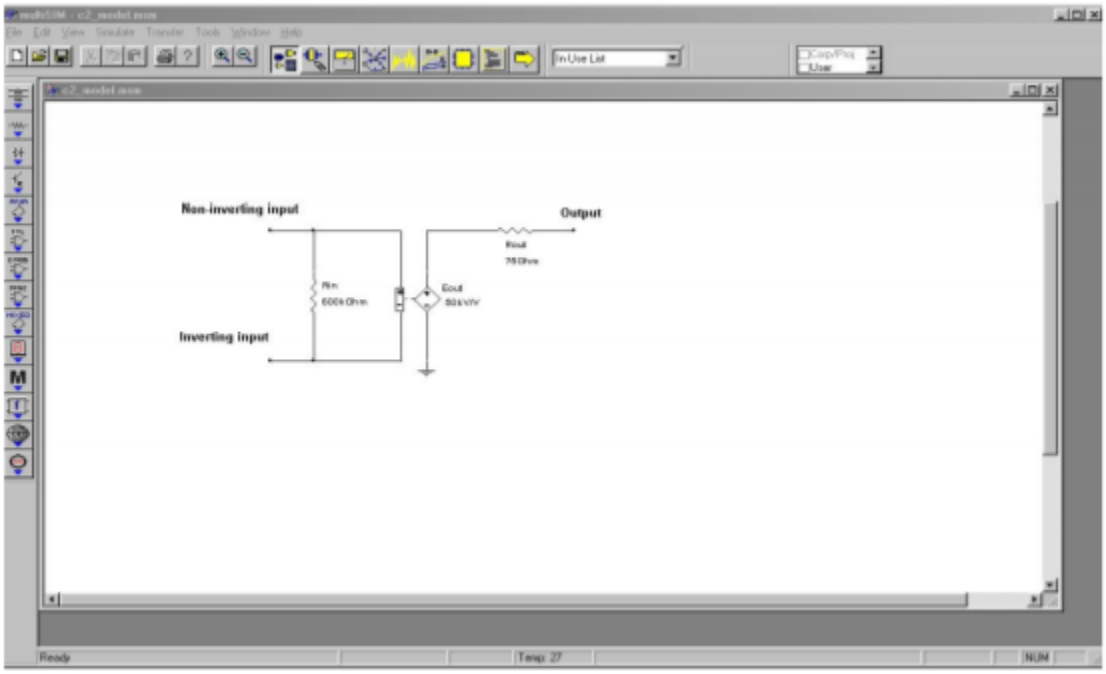


Figure 8.2.7 : Simple op amp model in Multisim.

This model must be used with great care because it is so simplistic. It is useful as a learning tool for

investigating the general operation of an op amp, but should never be considered as part of a true-to-life simulation. This model makes no attempt to consider the many limitations of the op amp. Because this model in no way imposes output signal swing limits, the effects of saturation will go unnoticed. Similarly, no attempt has been made at modeling the frequency response of the op amp. This is of great concern and we will spend considerable time on this subject in later chapters. Many other effects are also ignored. With so many limitations, you might wonder just where such a model may be used. This model is useful for non-critical simulations given low frequency inputs. You must also recognize the onset of saturation (clipping) yourself. Its primary advantage is that the circuit model is small, and thus computationally fast. Because of this, it is very efficient for students who are new to both op amps and circuit simulation. Perhaps of equal importance, is the fact that this model points out the fact that your simulation results can only be as good as the models you use. Many people fall into the trap that "because the simulation came from a computer, it must be correct". Nothing could be further from the truth. Always remember the old axiom: GIGO (Garbage In = Garbage Out). It can be very instructive to simulate a circuit using differing levels of accuracy and complexity, and then noting how closely the results match the same circuit built in the laboratory.

OP AMP DATA SHEET AND INTERPRETATION

Different manufacturers often use special codes and naming conventions to delineate their products from those of other manufacturers, as well as providing quality level and manufacturing information. A manufacturer's code is usually a letter prefix, while a quality or construction code is a suffix. Common prefix codes include ⋄ A (Fairchild), AD (Analog Devices), CS (Crystal), LM, LH, and LF (National Semiconductor, now owned by Texas Instruments, with M indicating monolithic construction, H indicating hybrid construction, and F indicating an FET device), LT (Linear Technology), MC (Motorola), NE and SE (Signetics), OPA (Burr-Brown), and TL (Texas Instruments).

Many manufacturers make a host of standard parts such as the 741. For example, Texas Instruments makes the LM741, while Fairchild makes the ⋄ A741. These parts are generally considered to be interchangeable, although they may vary in some ways. Some manufacturers will use the prefix code of the original developer of a part, and reserve their prefix for their own designs. As an example, Signetics produces their version of the 741, that they call ⋄ A741 because this op amp was first developed by Fairchild. (Signetics is then referred to as a second source for the ⋄ A741).

Suffix codes vary widely between manufacturers. Typical designations for consumer grade parts are C, and CN. The suffix N often means Not Graded. Interestingly, the lack of a final suffix often indicates a very high quality part, usually with an extended temperature range. Suffix codes are also used to indicate package styles. This practice is particularly popular among voltage regulators and other high-power linear ICs.

Finally, some manufacturers will use a "parallel" numbering system for high-grade parts. For example, the commercial grade device may have a "300 series" part number, with industrial grade given a "200 series" designation, and a military grade part given a "100 series" number. One possible example is the LM318 commercial grade op amp versus its high-grade counterpart, the LM118. Generally, militaryspecified parts will have a very wide temperature range, with industrial and commercial grades offering progressively less.

The data sheet for the LF411 op amp is shown in Figure 8.2.8 . Let's take a look at some of the basic parameters and descriptions. The values given are typical of a modern op amp. A complete

investigation of all parameters will be given in Chapter Five, once we've gotten a little more familiar with the device.

6 Specifications

6.1 Absolute Maximum Ratings (1)

	LF411A		LF411		UNIT
	MIN	MAX	MIN	MAX	
Supply Voltage		±22		±18	V
Differential Input Voltage		±38		±30	V
Input Voltage Range (2)		±19		±15	V
Output Short Circuit Duration		Continuous		Continuous	

	TO Package		PDIP Package		UNIT
	MIN	MAX	MIN	MAX	
Power Dissipation (3) (4)		670		670	mW
$T_j$max		150		115	°C
Operating Temperature		See (5)		See (5)	
Lead Temperature (Soldering, 10 s)		260		260	°C

(1) "Absolute Maximum Ratings" indicate limits beyond which damage to the device may occur. Operating Ratings indicate conditions for which the device is functional, but do not ensure specific performance limits.
(2) Unless otherwise specified the absolute maximum negative input voltage is equal to the negative power supply voltage.
(3) For operating at elevated temperature, these devices must be derated based on a thermal resistance of θ_{JA}.
(4) Max. Power Dissipation is defined by the package characteristics. Operating the part near the Max. Power Dissipation may cause the part to operate outside specified limits.
(5) These devices are available in both the commercial temperature range $0°C \le T_A \le 70°C$ and the military temperature range $-55°C \le T_A \le 125°C$. The temperature range is designated by the position just before the package type in the device number. A "C" indicates the commercial temperature range and an "M" indicates the military temperature range. The military temperature range is available in the TO package only.

6.2 Handling Ratings

		MIN	MAX	UNIT
T_{stg}	Storage temperature range	−65	150	°C

6.3 Recommended Operating Conditions

over operating free-air temperature range (unless otherwise noted)

	MIN	NOM	MAX	UNIT
Supply Voltage LF411A			±20	V
Supply Voltage LF411			±15	V

6.4 Thermal Information

THERMAL METRIC (1)			TO	PDIP	UNIT
			8 PINS	8 PINS	
$R_{\theta JA}$	Junction-to-ambient thermal resistance	Still Air	162		°C/W
		400 LF/min Air Flow	65		
$R_{\theta JA}$	Junction-to-ambient thermal resistance			120	
$R_{\theta JC}$	Junction-to-case (top) thermal resistance		20		

(1) For more information about traditional and new thermal metrics, see the *IC Package Thermal Metrics* application report, SPRA953.

Figure 8.2.8 : Data sheet for the LF411. Credit: Texas Instruments

6.5 DC Electrical Characteristics (1)(2)

PARAMETER		TEST CONDITIONS		LF411A			LF411			UNIT
				MIN	TYP	MAX	MIN	TYP	MAX	
V_{OS}	Input Offset Voltage	R_S=10 kΩ, T_A=25°C			0.3	0.5		0.8	2.0	mV
$\Delta V_{OS}/\Delta T$	Average TC of Input Offset Voltage	R_S=10 kΩ			7			7		µV/°C
I_{OS}	Input Offset Current	V_S=±15V (2) (3)	T_j=25°C		25	100		25	100	pA
			T_j=70°C			2			2	nA
			T_j=125°C			25			25	nA
I_B	Input Bias Current	V_S=±15V (2) (3)	T_j=25°C		50	200		50	200	pA
			T_j=70°C			4			4	nA
			T_j=125°C			50			50	nA
R_{IN}	Input Resistance	T_j=25°C			10^{12}			10^{12}		Ω
A_{VOL}	Large Signal Voltage Gain	V_S=±15V, V_O=±10V, R_L=2k, T_A=25°C		50	200		25	200		V/mV
		Over Temperature		25	200		15	200		V/mV
V_O	Output Voltage Swing	V_S=±15V, R_L=10k		±12	±13.5		±12	±13.5		V
V_{CM}	Input Common-Mode Voltage Range			±16	+19.5		±11	+14.5		V
					−16.5			−11.5		V
CMRR	Common-Mode Rejection Ratio	R_S≤10k		80	100		70	100		dB
PSRR	Supply Voltage Rejection Ratio	See (4)		80	100		70	100		dB
I_S	Supply Current				1.8	2.8		1.8	3.4	mA

(1) RETS 411X for LF411MH and LF411MJ military specifications.
(2) Unless otherwise specified, the specifications apply over the full temperature range and for V_S=±20V for the LF411A and for V_S=±15V for the LF411. V_{OS}, I_B, and I_{OS} are measured at V_{CM}=0.
(3) The input bias currents are junction leakage currents which approximately double for every 10°C increase in the junction temperature, T_j. Due to limited production test time, the input bias currents measured are correlated to junction temperature. In normal operation the junction temperature rises above the ambient temperature as a result of internal power dissipation, P_D. T_j=T_A+θ_{jA} P_D where θ_{jA} is the thermal resistance from junction to ambient. Use of a heat sink is recommended if input bias current is to be kept to a minimum.
(4) Supply voltage rejection ratio is measured for both supply magnitudes increasing or decreasing simultaneously in accordance with common practice, from ±15V to ±5V for the LF411 and from ±20V to ±5V for the LF411A.

6.6 AC Electrical Characteristics

PARAMETER(1)(2)		TEST CONDITIONS	LF411A			LF411			UNIT
			MIN	TYP	MAX	MIN	TYP	MAX	
SR	Slew Rate	V_S=±15V, T_A=25°C	10	15		8	15		V/µs
GBW	Gain-Bandwidth Product	V_S=±15V, T_A=25°C	3	4		2.7	4		MHz
e_n	Equivalent Input Noise Voltage	T_A=25°C, R_S=100Ω, f=1 kHz		25			25		nV / √Hz
i_n	Equivalent Input Noise Current	T_A=25°C, f=1 kHz		0.01			0.01		pA / √Hz
THD	Total Harmonic Distortion	A_V=+10, R_L=10k, V_O=20 Vp-p, BW=20 Hz–20 kHz		<0.02 %			<0.02 %		

(1) Unless otherwise specified, the specifications apply over the full temperature range and for V_S=±20V for the LF411A and for V_S=±15V for the LF411. V_{OS}, I_B, and I_{OS} are measured at V_{CM}=0.
(2) RETS 411X for LF411MH and LF411MJ military specifications.

Figure 8.2.8 (cont.): Data sheet for the LF411. Credit: Texas Instruments

First, note that two versions of the IC are given. We will examine the LF411 rather than the high-grade LF411A. At the very top of the data sheet is a listing of the absolute maximum ratings. The op amp should never operate at values greater than those presented as doing so may permanently damage it. Like most general-purpose op amps, the LF411 is powered by a bipolar power supply. The supply rails should never exceed ± 18 V DC. Normally, op amps will be used with ± 15 V supplies. Maximum power dissipation is given as 670 mW. Obviously then, this is a small signal device. In keeping with this, the operating temperature range and maximum junction temperatures are relatively low. We also see that the device can withstand differential input signals of up to 30 V, and single-ended inputs of up to 15 V, without damage. On the output, the LF411 is capable of withstanding a shorted load

condition continuously. This makes the op amp a bit more "bullet-proof". The remainder of this section details soldering conditions. Excessive heat during soldering may damage the device.

The second section of the data sheet lists the DC characteristics of the op amp. This table is broken down into five major sections:

- The parameter symbol.
- The parameter name.
- The conditions under which the parameter is measured.
- The parameter values, either typical or min/max.
- The parameter units.

We will examine a few of these parameters right now. The fourth parameter given is I_B, the input bias current. I_B is the current drawn by the bases (or gates) of the input differential amplifier stage. Because the LF411 utilizes a JFET diff amp, we expect this value to be rather small. For an operating temperature of 25°C, a typical LF411 will draw 50 pA, and a worst-case LF411 no more than 200 pA. If we extend the temperature range out a bit, I_B can extend out to 50 nA. This is a sizable jump, but even 50 nA is a very small value for general-purpose work. Because larger bias currents are normally seen as undesirable, the maximum I_B is the worst case scenario, hence a minimum I_B is not reported. Along with this, we see a very high value for input resistance, some $10^{12}\Omega$, typically. Op amps utilizing bipolar input devices will show much higher values for I_B, and much lower values for I_{IO}.

Next in line comes A_{VOL}. This is the DC voltage gain. Note the test conditions. The power supply is set to ± 15 V, the load is 2 k Ω, and A_{VOL} is 10 V peak. Normally, we desire as much gain as possible, so the worst case scenario is the minimum A_{VOL}. For 25° C operation this is specified as 25 V/mV, or 25,000. The average device will produce a gain of 200,000. As is typical, once the temperature range is expanded performance degrades. Over the operating temperature range, the minimum gain may drop to 15,000.

Because the op amp uses a class AB follower for its output stage, we should expect the output compliance to be very close to the power supply rails. The output voltage swing is specified for ± 15 V supplies with a 10 k Ω load. The typical device can swing out to ± 13.5 V, with a worst case swing of ± 12 V. A reduction in power supply value will naturally cause the maximum output swing to drop. A sizable reduction in the load resistance will also cause a drop in I_B, as we will see a bit later. These maximum output values are caused by the internal stages reaching their saturation limits. When this happens, the op amp is said to be clipping or in saturation. As a general rule of thumb, saturation may be approximated as 1.5 V less than the magnitude of the power supplies.

The last item in the list is the standby current draw, I_D. Note how small this is, only 1.8 mA, 3.4 mA worst case. This is the current the op amp draws from the supply under no signal conditions. When producing output signals, the current draw will rise.

The final section of the data sheet lists certain AC characteristics of the op amp that will be of great concern to us in later sections. Many device parameters change a great deal with frequency, temperature, supply voltage, or other factors. Because of this, data sheets also include a large number of graphs that further detail the op amp's performance. Finally, application hints and typical circuits may round out the basic data sheet.

8.3 SIMPLE OP AMP COMPARATOR

Now that you have a feel for what an op amp is and what some typical parameters are, let's take a look at an application. The one thing that jumps to most people's attention is the very high gain of the average op amp. The typical LF411 showed ◈◈◈◈ at approximately 200,000. With gains this high, it is obvious that even very small input signals can force the output into saturation (clipping). Take a look at Figure 8.3.1 . Here an op amp is being supplied by ± 15 V, and is driving a 10 k Ω load. As seen in our model of Figure 8.2.5, ◈◇◈◇◈should equal the differential input voltage times the op amp's gain, ◈◈◈◈ .

$$V_{out} = A_{vol}(V_{in+} - V_{in-})$$
$$V_{out} = 200,000 \times (0.1\ V0\ V)$$
$$V_{out} = 20,000\ V$$

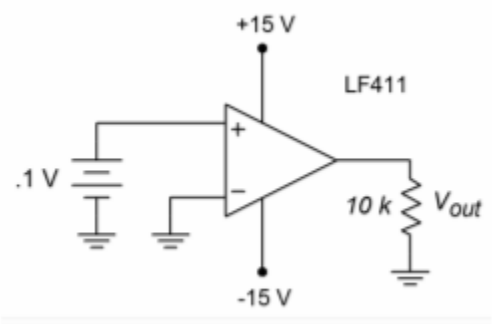


Figure 8.3.1 : Comparator (single input).

The op amp cannot produce 20,000 V. The data sheet lists a maximum output swing of only ± 13.5 V when using ± 15 V supplies. The output will be truncated at 13.5 V. If the input signal is reduced to only 1 mV, the output will still be clipped at 13.5 V. This holds true even if we apply a signal to the inverting input, as in Figure 8.3.2 .

$$V_{out} = A_{vol}(V_{in+} - V_{in-})$$
$$V_{out} = 200,000 \times (0.5\ V - 0.3\ V)$$
$$V_{out} = 40,000\ V$$
$$V_{out} = 13.5\ V, \text{due to clipping}$$

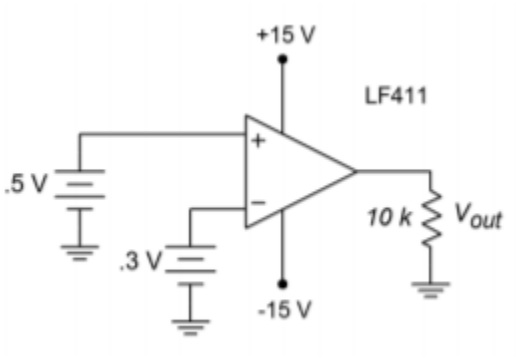


Figure 8.3.2 : Comparator (dual input).

COMPUTER SIMULATION

A simulation of Figure 8.3.2 using Multisim is shown in Figure 8.3.3 . The LF411 op amp is selected from the component library, and you need not concern yourself with the model's internal make-up just yet. This particular model includes the effects of power supply limitations (i.e., output saturation) that the very simple dependent-source model presented earlier does not. Individual DC sources are used for the input signals. Although no AC signals are applied, it is perfectly valid to run a transient-mode simulation. The first millisecond of the output voltage is shown. It verifies the manual calculation, indicating a DC level of slightly more than 13.5 V.

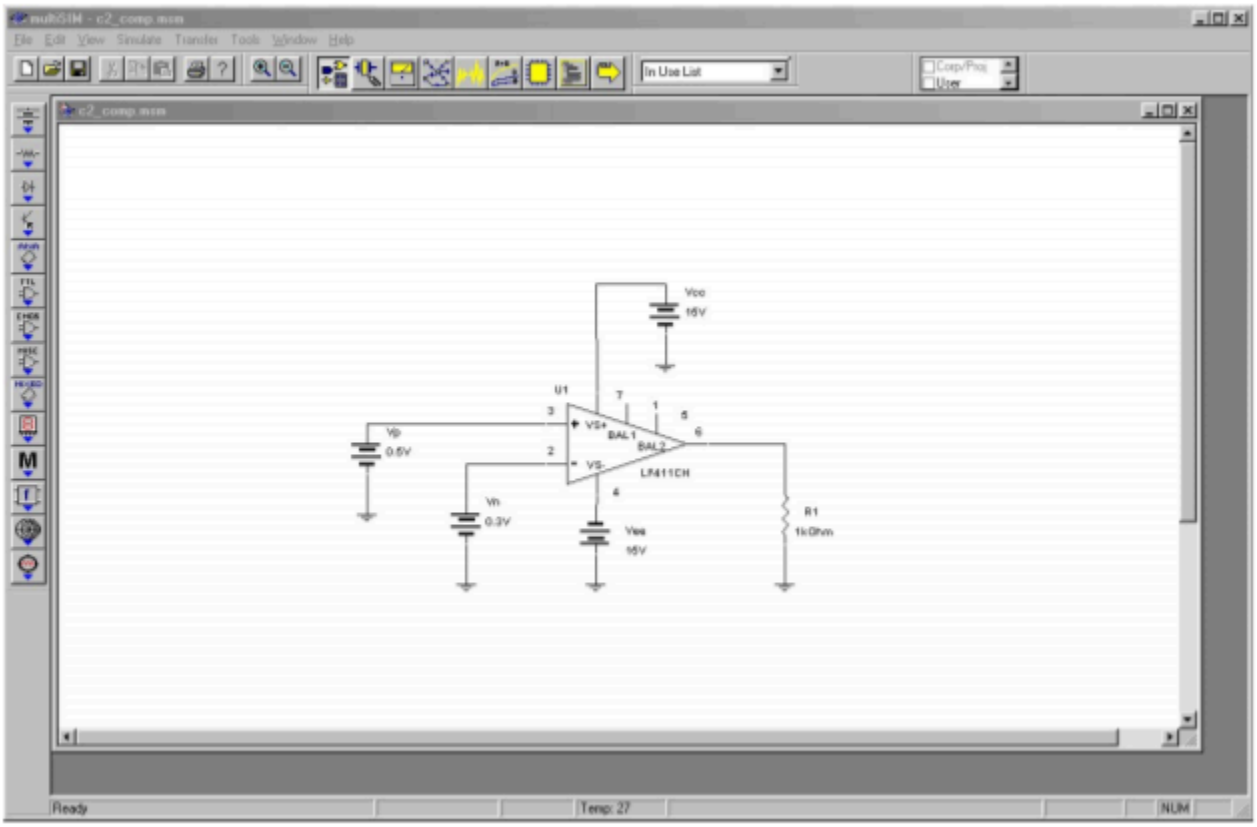


Figure 8.3.3 : Comparator example in Multisim.

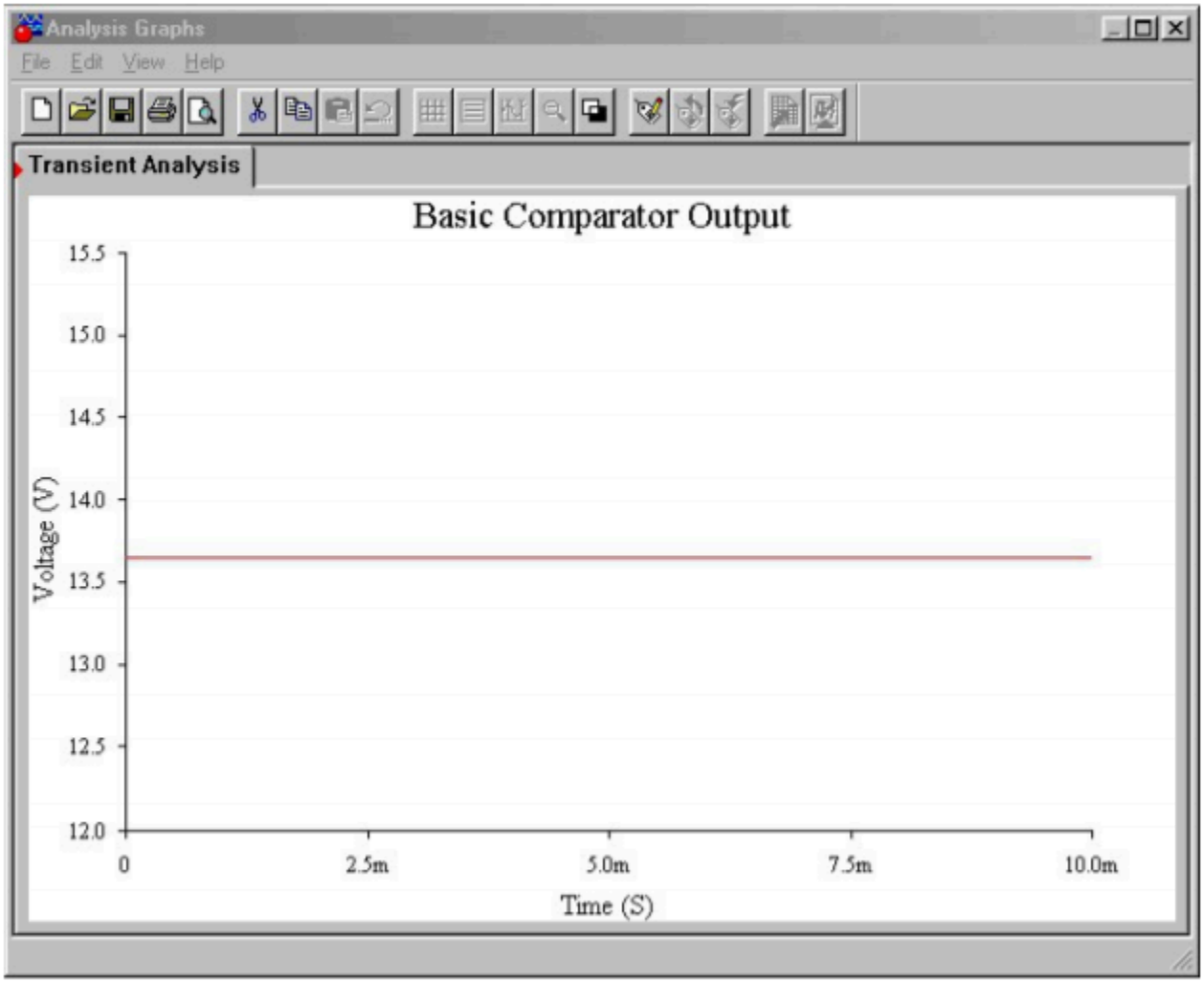


Figure 8.3.3 : Comparator output.

For any reasonable set of inputs, as long as the noninverting signal is greater than the inverting signal, the output will be positive saturation. If you trade the input signals so that the inverting signal is the larger, the converse will be true. As long as the inverting signal is greater than the noninverting signal, the output will be negative saturation. If the inverting and noninverting signals are identical, should be 0 V. In the real world, this will not happen. Due to minute discrepancies and offsets in the diff amp stage, either positive or negative saturation will result. You have no quick way of knowing in which direction it will go. It is for this reason that it is impractical to amplify a very small signal, say around 10 V. You might then wonder, "What is the use of this amplifier if it always clips? How can I get it to amplify a simple signal?" Well, for normal amplification uses, we will have to add on some extra components, and through the use of negative feedback (next chapter), we will create some very well-controlled, useful amplifiers. This is not to say that our barren op amp circuit is useless. Quite to the contrary, we have just created a comparator.

A comparator has two output states: High and low. In other words, it is a digital, logical output. Our comparator has a high state potential of 13.5 V, and a low state potential of -13.5 V. The input signals, in contrast, are continuously variable analog potentials. A comparator, then, is an interface between analog and digital circuitry. One input will be considered the reference, while the other input will be considered the sensing line. Note that the differential input signal is the difference between

the sensing input and the reference input. When the polarity of the differential input signal changes, the logical output of the comparator changes state.

Example 8.3.1

Figure 2.3.4 shows a light leakage detector that might be used in a photographer's dark room. This circuit utilizes a Cadmium Sulfide (CdS) cell that is used as a light sensitive resistor. The inverting input of the op amp is being used as the reference input, with a 1 V DC level. The noninverting input is being used as the sensing input. Under normal (no light) conditions, the CdS cell acts as a very high resistance, perhaps 1 M Ω . Under these conditions, a voltage divider is set up with the 10 k Ω resistor, producing about 150 mV at the noninverting input. Remember, no loading of the divider occurs because the LF411 utilizes a JFET input. Because the noninverting input is less than the inverting input, the comparator's output is negative saturation, or approximately -13.5 V. If the ambient light level rises, the resistance of the CdS cell drops, thus raising the signal applied to the noninverting input. Eventually, if the light level is high enough, the noninverting input signal will exceed the 1 V reference, and the comparator's output will move to positive saturation, about +13.5 V. This signal could then be used to trigger some form of audible alarm. A real world circuit would need the flexibility of an adjustable reference in place of the fixed 1 V reference. By swapping the CdS cell and the 10 k Ω resistor, and adjusting the reference, an inverse circuit (i.e., an alarm that senses darkness) may be produced.

Circuits of this type can be used to sense a variety of over level/under level conditions, including temperature and pressure. All that is needed is an appropriate sensing device. Comparators can also be used with AC input signals.

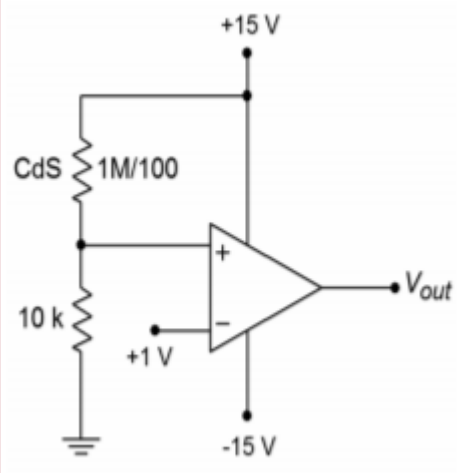


Figure 2.3.4 : Light alarm.

Example 8.3.2

Sometimes, it is necessary to square up an AC signal for further processing. That is, we must turn it into an equivalent pulse waveform. One example of this might be a frequency counter. A frequency counter works by tallying the number of high-to-low or low-to-high transitions in the input signal over a specific

length of time. For accurate counts, good edge transitions are required. Because a simple sine wave changes relatively slowly compared to an equal frequency square wave, some inaccuracy may creep into the readings. We can turn the input into a pulse-type output by running it through the comparator of Figure 2.3.5 . Note that the reference signal is adjustable from –15 to +15 V. Normally, the reference is set for 0 V. Whenever the input is greater than the reference, the output will be positive saturation. When the input is less than the reference, the output will be negative saturation. By making the reference adjustable, we have control over the output duty cycle, and can also compensate for DC offsets on the input signal. A typical input/output signal set is given in Figure 2.3.6 .

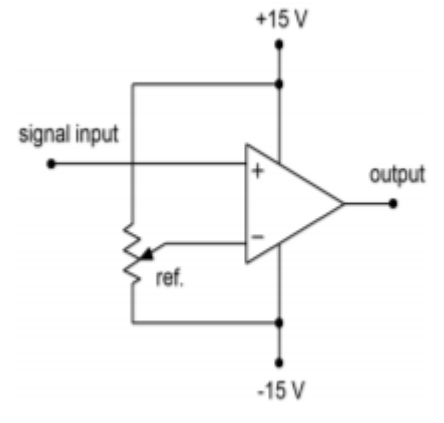


Figure 2.3.5 : "Square-up" circuit.

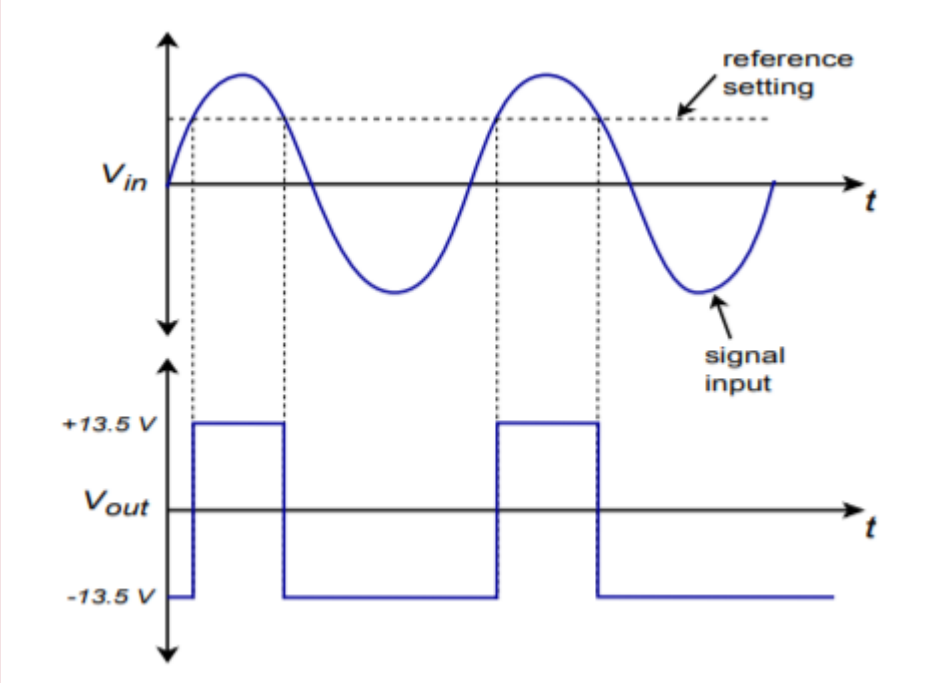


Figure 2.3.6 : Output of "square-up" circuit.

There are a few limitations with our simple op amp comparator. For very fast signal changes, a typical op amp will not be able to accurately track its output. Also, the output signal range is rather wide and is bipolar. It is not at all compatible with normal TTL logic circuits. Extra limiting circuitry is required for proper interfacing. To help reduce these problems, a number of circuits have been specially optimized for comparator purposes. We will take a closer look at a few of them in Chapter Seven.

8.4 OP AMP MANUFACTURE

Op amps and other linear integrated circuits are generally manufactured in one of two ways: The device is either a hybrid, or is monolithic. In either case, the circuit can contain hundreds of components. The resulting device will be packaged in a variety of styles, including plastic and ceramic dual in-line and single in-line types, multi-lead cans, flat packs, and surface mount forms. Some examples are shown in Figure 2.4.1 . In each type, the circuitry is completely encased and not accessible to the designer or technician. If one of the components should fail, the entire op amp is replaced. The design and layout of the integrated circuit itself is normally carried out with the use of special computer workstations and software tools. These allow the designers to simulate portions of the circuit, and create the outlines and interconnections for the various components to be formed.

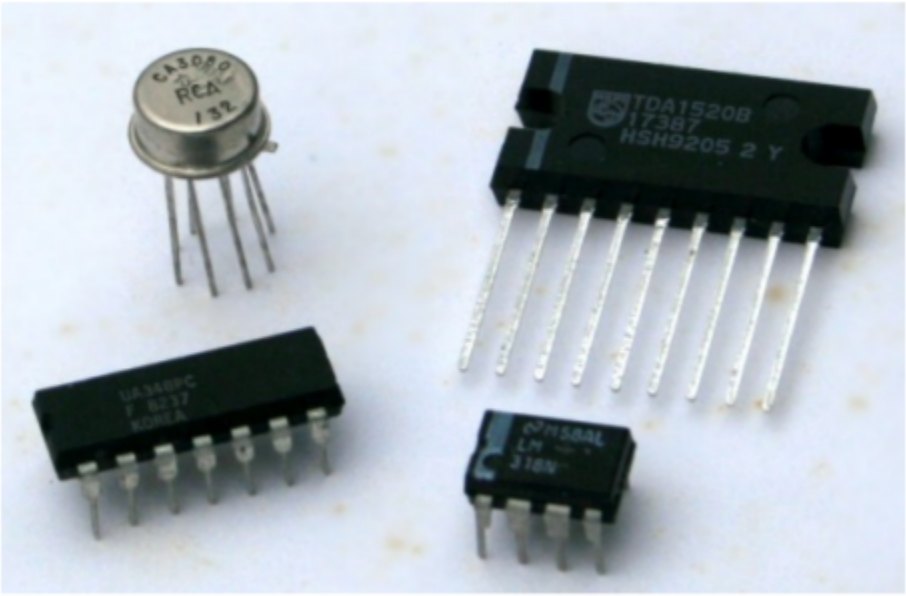


Figure 2.4.1 : Package styles (through hole) clockwise from top-left: can, single in-line, mini dual in-line, dual in-line.

MONOLITHIC CONSTRUCTION

The term monolithic is from the Greek, meaning literally "single stone". In this process, all circuit elements are created and interconnected using a single slab of silicon (or other suitable material). Normally, several op amps are made from a single silicon wafer. Each wafer may be a few inches in diameter, with each op amp circuit chip comprising perhaps a square 1 millimeter by 1 millimeter in area. A single transistor can easily be smaller than 15 micrometers by 20 micrometers. Because the scale of construction is so small, special clean rooms are required in order to remove tiny air-borne particles of dust and grit that could interfere with the production of these super-small components. Workers in clean rooms are required to wear special suits as well.

Figure 2.4.2 outlines the major steps in the chip manufacturing process. This process starts with

the preparation of a P-type silicon wafer. This is referred to as the substrate. After it has been cleaned and polished, an N-type epitaxial region is diffused into the P-type base. Epitaxial is from the Greek, roughly meaning "to arrange upon". It is within this thin epitaxial region that the circuit elements will be formed, the remainder of the substrate lending mechanical support to the structure. The term diffusion refers to the manner in which the semiconductor material becomes doped. In essence, the base material is surrounded by a high concentration of doping material, usually gaseous, along with the application of heat. The low concentration wafer material will be infiltrated by the high concentration doping material. Diffusion is a relatively accurate and inexpensive means of controlling the semiconductor's properties.

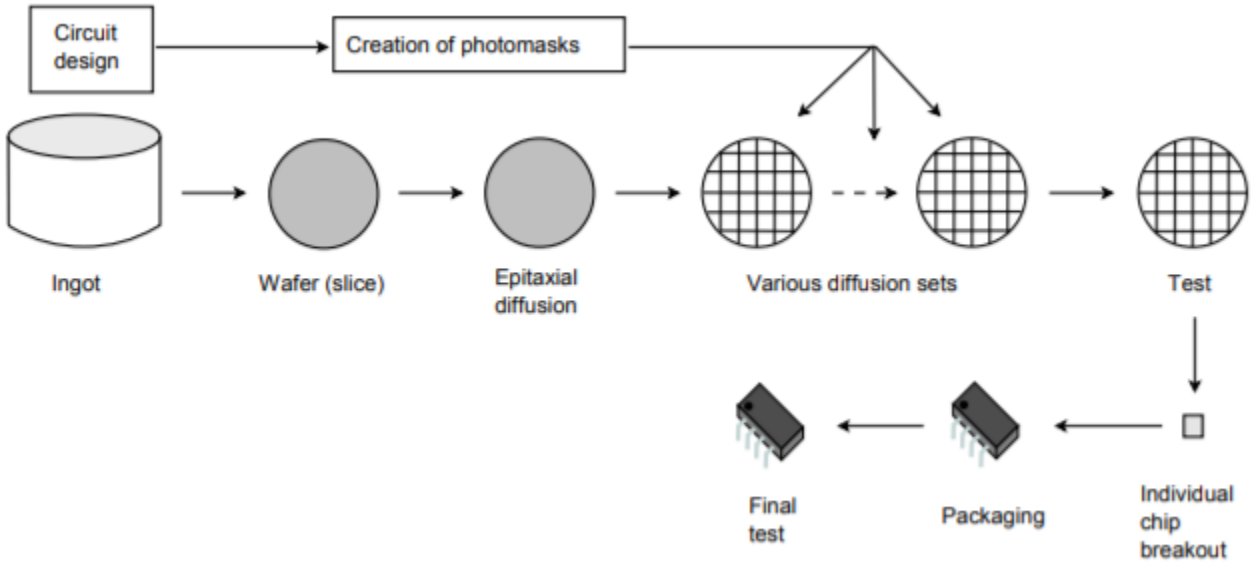


Figure 2.4.2 : Chip manufacturing process.

Once the N-type region is produced, the wafer will undergo an oxidizing process that will leave the top surface covered with silicon dioxide. This layer prevents impurities from entering the N-type region. At this point, a series of steps will be used to create wells or deposits of alternate p and n material. These deposits will form the various active and passive components. Normally, this is done through a photolithographic process. This involves the use of light sensitive materials and masks. Conceptually, the process is not very much different from the way printed circuit boards are often made. In essence, specific areas of the silicon dioxide layer will be stripped away, thus exposing the epitaxial region, and allowing diffusion of other acceptor/donor impurities to take place. Because the silicon dioxide serves as an effective barrier to diffusion, only areas cleared of silicon dioxide will be effected by the diffusion process. In this manner, specific areas can be singled out, and selectively doped to create specific components. This is detailed below and in Figure 2.4.3 .

In order to selectively remove the silicon dioxide, the top surface is coated with a light sensitive material called photoresist. Above this is placed a mask. This mask is much like a black and white negative; some areas are clear, and some areas are opaque. The resulting sandwich is then exposed to ultraviolet light. The clear areas of the mask will allow the light to pass through and cause a chemical change in the photoresist. A solution is then used to wash away the unexposed photoresist. At this point, a second solution is used to wash away the silicon dioxide. This solution will not effect the exposed photoresist, and thus, the silicon dioxide beneath it is unaffected. After this protecting layer

of photoresist is removed, all that remains on the top surface of the wafer are alternating patches of silicon dioxide. The wafer can now be lead through another diffusion process.

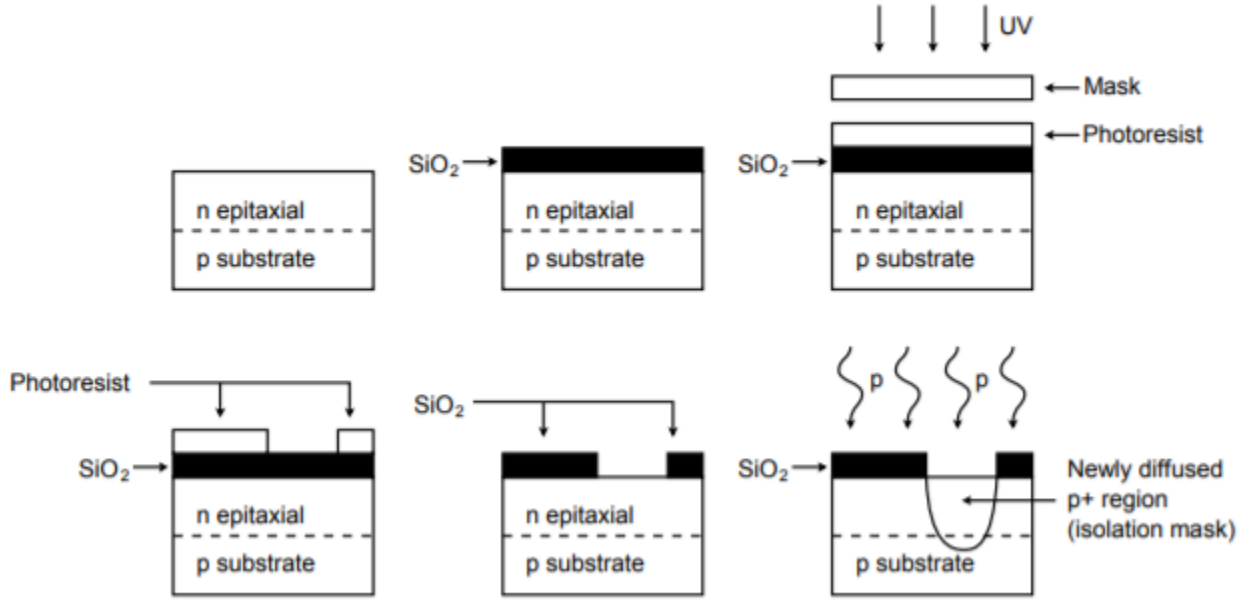


Figure 2.4.3 : Diffusion process (one run).

The process of oxidizing, masking, and diffusing will be repeated several times. The initial run will be produced with an isolation mask. This is used to separate the various components. Normally, a base mask will be used next, followed by the emitter mask. The final masks will be used for contacts and interconnections. In this way, N-type material can be placed next to, or completely within, P-type material. The adjoining areas are, of course, PN junctions. Because all circuit elements are laid out length-wise on a thin strip, this form of manufacture is referred to as a planar process.

Once the final mask is completed, the wafer will be inspected. The individual chips will then be broken out of the wafer and mounted into the desired package. Leads will be connected to the chip with fine angel hair wire, and then the package will be sealed. It is now ready for final test and inspection. Part numbers and date codes will also be imprinted.

Virtually all general-purpose op amps today use a planar monolithic process. Some of the advantages of monolithic construction are its relative simplicity and low per-part cost.

HYBRID CONSTRUCTION

Hybrids are usually used where a complete monolithic solution is impractical. This is usually the case for special purpose devices, such as those requiring very high output current, very wide bandwidth, or that are very complex or sensitive. Hybrids, as the name suggests, are a collection of smaller circuit elements interconnected. A typical hybrid may contain two or three smaller monolithic chips and assorted miniaturized passive and/or power components. Passive components may be further integrated by using either a thin or thick film chip process. (A discussion of thin and thick film chip techniques is beyond the scope of this text). Due to the complexity of a hybrid chip, it is normally more expensive than its monolithic cousins. Although the IC itself may be more expensive, the complete application may very well wind up being less costly to produce because the cost of other components are effectively absorbed within the hybrid IC. One place where hybrids are often used is in consumer stereo music systems. A hybrid power amplifier IC offers the convenience of a single IC solution with

the capabilities of a discrete transistor approach. As an op amp user, it makes little difference whether the device is hybrid or monolithic when it comes to circuit analysis or design.

8.5 SUMMARY

Op amps are presently in wide use in just about every aspect of linear electronics. An op amp is a multi-stage amplifier treated as a single entity. The first stage usually utilizes a differential amplifier that can be made with either bipolar or FET devices. The following stage(s) create a large voltage gain. The final stage is a class B voltage follower. The resulting op amp typically has a high input impedance, a low output impedance, and voltage gains in excess of 10,000. The op amp operates from a bipolar power supply, usually around ±15 V. Externally, it has connections for the inverting and noninverting inputs, the single-ended output, and the power supplies. The op amp may be packaged in a variety of forms, including DIPs, SIPs, cans, flat packs and surface mount.

The general-purpose op amp is manufactured using a monolithic structure and a photolithographic process. Several chips are created from a single master wafer. Creation is a multi-step process involving the selective doping of specific areas on the chip through diffusion. The monolithic technique is relatively inexpensive and accurate. Because integration allows for very tight part matching and consistency, certain circuit design techniques are favored, including the use of current mirrors and active loads.

Finally, with very little supporting circuitry, simple op amps can make effective comparator circuits. A comparator is in essence, a bridge between the analog and digital worlds.

8.6 PROBLEMS

REVIEW QUESTIONS

1. What is an op amp?
2. Give several examples of where op amps might be used.
3. What is the typical stage layout of an op amp?
4. What comprises the first stage of a typical op amp?
5. What comprises the final stage of a typical op amp?
6. How does integrated circuit design differ from discrete design?
7. What is a comparator, and how might it be used?
8. What is meant by monolithic planar construction?
9. What is a mask, and how is it used in the construction of IC op amps?
10. What is the process of diffusion and how does it relate to the construction of IC op amps?
11. What are the advantages of monolithic IC construction?
12. How does hybrid construction differ from monolithic construction?

PROBLEMS

1. For the circuit in Figure 2.6.1 , find ⍰⍰⍰⍰ for the following inputs:
    a. 0 V
    b. -1 V
    c. +2 V

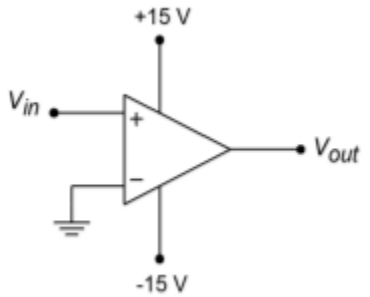


Figure 2.6.1

2. For the circuit in Figure 2.6.2 , find ⍰⍰⍰⍰ for the following inputs:
    a. 0 V

b. -4 V

c. +5 V

d. -0.5 V

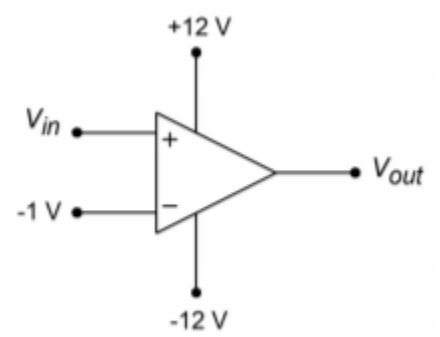


Figure 2.6.2

3. For the circuit in Figure 2.6.3, find ◈◈◈◈ for the following inputs:

   a. 0 V

   b. -2 V

   c. +1 V

   d. -0.5 V

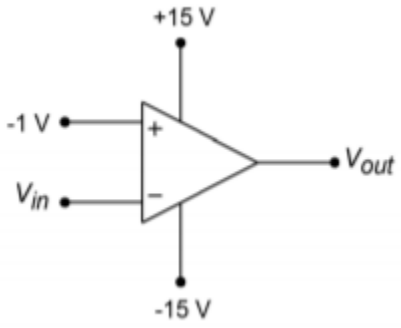


Figure 2.6.3

4. Sketch ◈◈◈◈if ◈◈◈(◈)=1 sin 60◈ in Figure 2.6.1 .

5. Sketch ◈◈◈◈if ◈◈◈(◈)=2 sin 20◈ in Figure 2.6.2 .

6. Sketch ◈◈◈◈if ◈◈◈(◈)=3 sin 10◈ in Figure 2.6.3 .

7. A Temperature Dependent Resistor is used in the comparator of Figure 2.6.4 . At what temperature will the comparator change state?

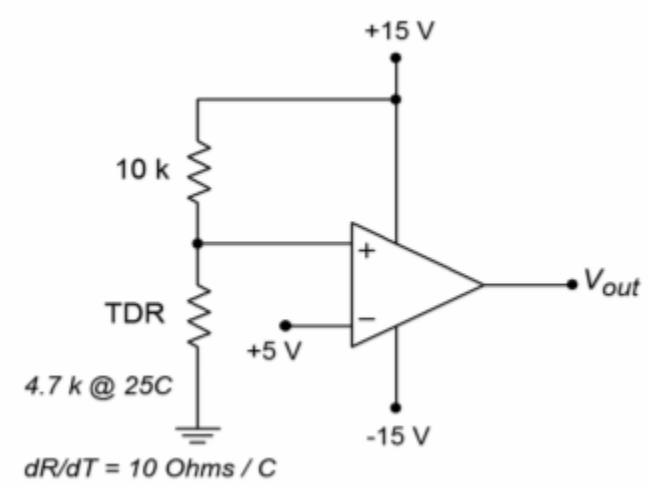


Figure 2.6.4

8. What is the value of the Light Dependent Resistor at the comparator trip point in Figure 2.6.5 ?

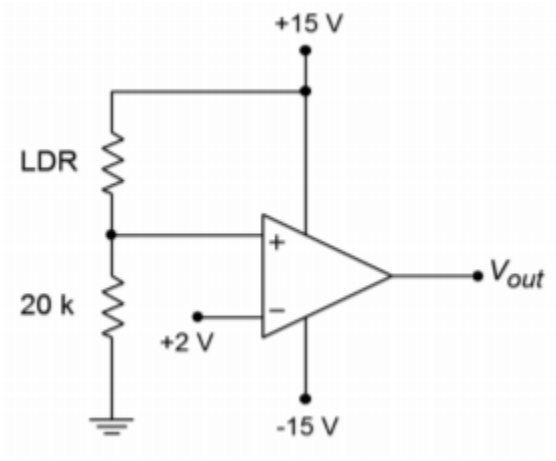


Figure 2.6.5

9. What reference voltage is required for a 50°C trip point in Figure 2.6.4 ?

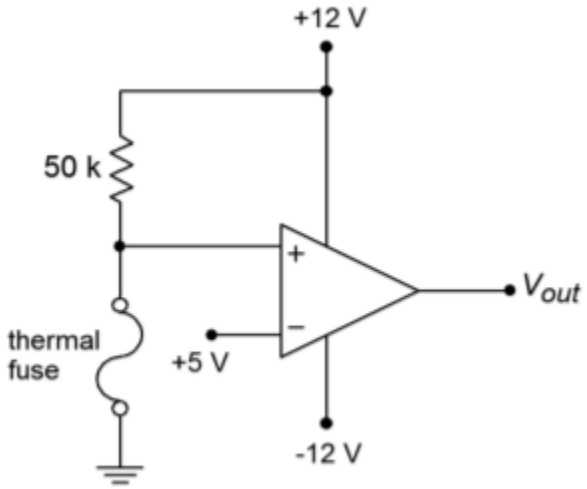


Figure 2.6.6

10. A thermal fuse is device found in such common items as coffee makers. Normally, its resistance is very low, ideally 0 Ω . With the application of excessive heat, the fuse opens, presenting a very high resistance, ideally infinite. Explain the operation of the circuit in Figure 2.6.6 . Is the output voltage normally high or low?
11. A strain gauge is a device that can be used to measure the amount of bend or deflection in a part that it is attached to. It can be connected so that its resistance rises as the bend increases. What resistance is required to trip the comparator of Figure 2.6.7 ?

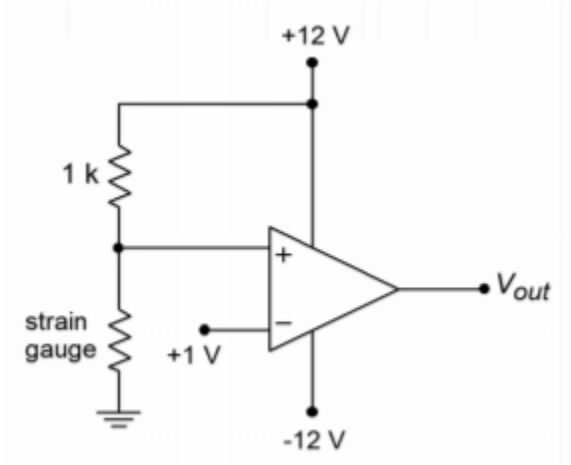


Figure 2.6.7

12. Determine the current flowing out of the collector of ◈13 in Figure 2.2.4.

COMPUTER SIMULATION PROBLEM

13. Alter the simple simulation model presented in the chapter to include saturation effects. (Hint: Consider a device that limits voltage.)

UNIT 9: NEGATIVE FEEDBACK

Learning Objectives

After completing this chapter, you should be able to:

- Give examples of how negative feedback is used in everyday life.
- Discuss the four basic feedback connections, detailing their similarities and differences.
- Detail which circuit parameters negative feedback will alter, and how.
- Discuss which circuit parameters are not altered by negative feedback.
- Define the terms sacrifice factor, gain margin, and phase margin, and relate them to a Bode plot.
- Discuss in general, the limits of negative feedback in practical amplifiers.

9.1 INTRODUCTION TO NEGATIVE FEEDBACK

As we saw in the last chapter, op amps are very useful devices. However, in many applications, the device's gain is simply too large and its bandwidth too narrow, for effective use. In this chapter we will explore the concept of negative feedback. This concept is realized by feeding a portion of the output signal back to the input of the system. The proper use of negative feedback will allow us to exercise fine control over the performance of electronic circuits. As a matter of fact, negative feedback is so useful to us that we will seldom use op amps without it. Negative feedback is not tied solely to op amps though, as almost any electronic circuit may benefit from its application. As with most things, there are disadvantages as well. A successful design will minimize the disadvantages and capitalize on the positive aspects. We will begin with the basic concepts of what negative feedback is and does, and then fine-tune our viewing by examining its four specific variants. We will look at specific examples of how negative feedback is applied to op amps, and finish off with a discussion of its practical limits.

9.2 WHAT NEGATIVE FEEDBACK IS AND WHY WE USE IT

People use negative feedback every day of their lives. In fact, we probably couldn't get along without it. Simply put, negative feedback is a very rudimentary part of intelligence. In essence, negative feedback lets something correct for mistakes. It tends to stabilize operations and reduce change. Negative feedback relies on a loop concept. In human terms, it is akin to knowing what you are doing and being able to correct for mistakes as they happen. You are constantly evaluating and correcting your actions in order to achieve a desired goal. This may be stated as letting the input know what the output is doing. A good example of this is your ability to maintain a constant speed while driving along the highway. You have a desired result, or set-point, in mind, say 60 MPH. As you drive, you constantly monitor the speedometer. If you glance down and see that you're zipping along at 70 MPH, you think "Oops, I'm going a bit too fast" and lift your foot slightly off of the gas pedal. On the other hand, if you're only going 40 MPH, you will depress the gas pedal further. The faster and more accurate your updates are, the better you will be at maintaining an exact speed.

In contrast to negative feedback is positive feedback, which reinforces change. If you were to correct your speed by saying "Hmm, I'm going 70 MPH, I'd better step on the gas", you'd be using positive feedback. Other examples of positive feedback include the "acoustic squeal" often heard over public address systems, and thermal runaway effects seen in discrete devices. When positive feedback is applied to normal amplifiers, they oscillate. That is, they produce their own signals without any input applied.

9.3 BASIC CONCEPTS

Seeing the usefulness of negative feedback, it would be nice if we could apply the concept to our electronic circuits. The basic idea is quite simple, really. What we will do is sample a piece of the output signal, and then add it to the input signal out of phase (i.e., subtract it). By doing so, the circuit will see the difference between the input and the output. If the output signal is too large, the difference will be negative. Conversely, the difference signal will be positive if the output is too small. This signal is then multiplied by the circuit gain and cancels the output error. Thus, the circuit will be presented with the undesired errors in a way that will force the output to compensate (move in the opposite direction). This process is done continuously; the only time lags involved are the propagation delays of the circuits used.

Because the sampled output signal is effectively subtracted from the input signal, negative feedback is sometimes referred to as degenerative or destructive feedback. This subtraction can be achieved in a variety of ways. A differential amplifier is tailor made for this task because it has one inverting input and one noninverting input. (Note: If the error is presented in phase, the circuit magnifies the errors and positive feedback results.)

To see an example of how this works, refer to Figure 3.3.1 . The triangle represents an amplifying circuit. It has a gain of ◈ . The output signal is also presented to the input of the feedback network represented by the box. This network scales the output signal by a factor, ◈ . The feedback network ranges from very simple to complex. It may contain several resistors, capacitors, diodes and what not, or it may be as simple as a single piece of wire. In any case, this scaled output signal is referred to as the feedback signal and is effectively subtracted from the input signal. This combination, called the error signal, is then fed to the amplifier where it is boosted and appears at the output. The process repeats like this forever (or at least until the power is switched off).

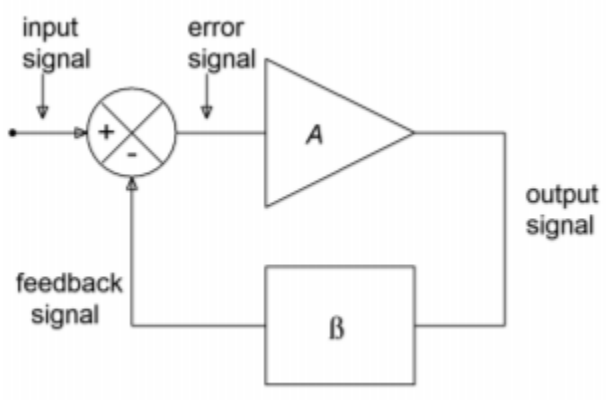


Figure 3.3.1 : Negative feedback.

Let's assume that for some reason (perhaps a temperature change) the amplifier's gain were to rise. This should make the output signal increase by a similar percentage, but it doesn't. Here's why: As the output signal tries to rise, the feedback signal tracks with it. Now that there is a larger feedback signal, the error signal will become smaller (remember, error = input – feedback). This smaller signal is

multiplied by the gain of the amplifier, thus producing a smaller output signal that almost completely offsets the original positive change. Note that if the output signal were too small, the error signal would increase, thus bringing the output back up to a normal level. When everything is working right, the feedback and input signals are almost the same size. (Actually, the feedback signal is somewhat smaller in magnitude.)

THE EFFECTS OF NEGATIVE FEEDBACK

Besides smoothing out gain anomalies, negative feedback can reduce the effect of device non-linearities, thus producing a reduction in static forms of distortion such as THD (Total Harmonic Distortion). Basically, these non-linearities can be viewed as a string of small gain errors. As such, they produce appropriate error signals and are compensated for in the above manner. Negative feedback can also increase the bandwidth of the system. It can increase the upper cutoff frequency f_2 and decrease the lower cutoff frequency f_1 (assuming the system has one). Also, we can exercise control over the input and output impedances of the circuit. It is possible to increase or decrease the impedances. As you might have guessed, we don't receive these benefits for nothing. The down side to negative feedback is that you lose gain. Effectively, you get to trade off gain for an increase in bandwidth, a decrease in distortion, and control over impedances. The more gain you trade off, the greater your rewards in the other three areas. In the case of our op amp, this is a wise tradeoff because we already have more gain than we need for typical applications. This give-and-take is a very important idea, so remember "BIG D". That stands for Bandwidth, Impedance, Gain, and Distortion

At this point, we need to define a few terms. **Closed loop** refers to the characteristics of the system when feedback exists. For example, closed-loop gain is the gain of the system with feedback, whereas closed-loop frequency response} refers to the new system break points. Generalized closed-loop quantities will be shown with the subscript " ff ". Similarly, we will denote impedances, gain and the like for specific feedback variants with a two- letter subscript abbreviating the exact feedback configuration. One possibility for closed-loop gain would be A_{fff} . Open loop refers to the characteristics of the amplifier itself. To remember this, think of disconnecting or opening the path through the feedback network. Once the path is broken, the amplifier is on its own. **Open-loop gain** then, refers to the gain of the amplifier by itself, with no feedback. All open-loop quantities will be shown with the subscript " oo ". The symbol for open-loop gain would be A_{ooo} . The term loop gain refers to the ratio between the open- and closed-loop gains (it may also be computed from their difference in decibels on a Bode plot). Loop gain indicates how much gain we have given up or sacrificed in order to enhance the operation of the system. Consequently, loop gain is often called sacrifice factor, and given the symbol f . Generally, tradeoffs are proportional to the sacrifice factor. For example, if we cut the gain in half, we will generally double the bandwidth, and halve the distortion. An example that illustrates this can be seen in Figure 3.3.2 . Note how the loop gain decreases with increasing frequency. This means that the effects of feedback at higher frequencies are not as great.

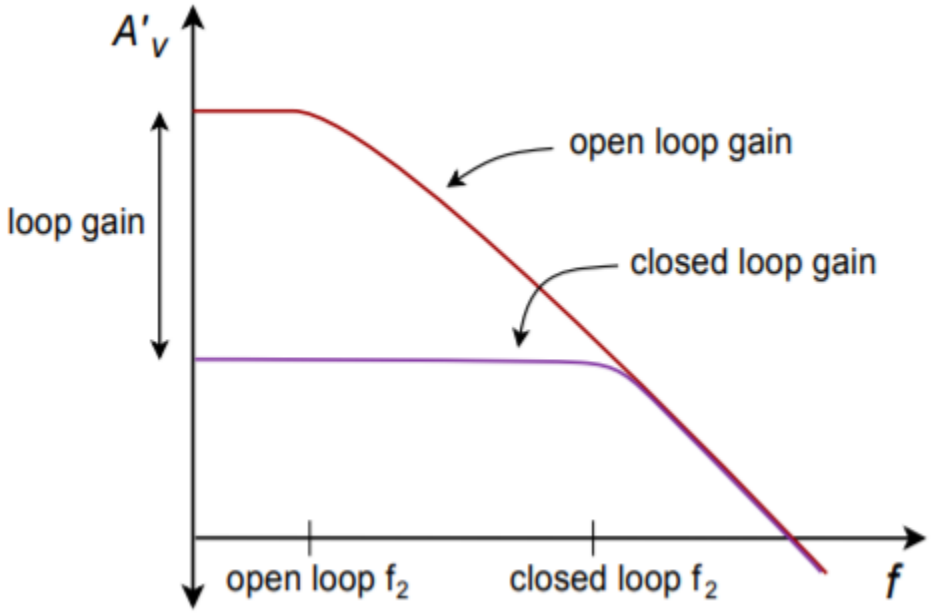


Figure 3.3.2 : Response with and without feedback.

Up to this point we have made one simple assumption about our system; that it exhibits no "extra" phase change beyond the desired inversion. As we saw in the first chapter, however, all circuits do produce phase changes as the input frequency is increased. If this extra phase change were to reach -180° while the gain was greater than unity (0dB), our negative feedback will turn into positive feedback (the inversion = -180° , plus this extra -180°, places us at -360° . The net result is an inphase signal). If this were to happen, our amplifier would no longer be stable. In fact, it may very well turn into a high frequency oscillator. (You will see how to do this on purpose when you cover Chapter Nine). As the input frequency is raised, the phase will eventually exceed -180° and the gain will drop to a fraction (< 0dB). The real key here is making sure that the phase never reaches or exceeds -180° when the gain falls to 1. Stated another way, when the phase hits -180°, the gain should be a fraction. Generally, the farther you are from this "danger zone", the better. In other words, you're better off if the extra phase at the unity gain point is -90° rather than -170°. Both are stable, but the first one gives you some breathing room. Two measures of circuit safety are the gain and phase margins. Phase margin indicates the difference between the actual phase at unity gain and -180°. In the example above the first circuit would have a phase margin of 90° and the second would have a margin of 10°. Gain margin is the difference between the actual gain in dB at the -180° phase point, and 0dB. If our gain was -9 dB at -180°, the gain margin would be 9 dB (i.e., we have 9 dB "to spare"). Reasonable values for gain and phase margin are > 6 dB and > 45°. Gain and phase margins are depicted in Figure 3.3.3 . It is possible to guarantee safe margins if the amplifier's open loop response maintains a 20 dB-per-decade rolloff up to the unity gain frequency, ◇◇◇◇◇◇ . This means that there is only one dominating lag network that will add a maximum phase shift of -90° . Even if the second network coincided with ◇◇◇◇◇◇ , it would add -45° at most. This would still leave us with a 45 ° phase margin. (Note that if we had several secondary networks critical at ◇◇◇◇◇◇ the phase could exceed -180°, however the slope would no longer be 20 dB-per-decade in reality.) It is for this reason that the general-purpose op amps examined in Chapter Two included a compensating capacitor. No matter how much feedback we wish to use, our circuits will always end up being stable. For the best circuit

performance, it is possible to use amplifiers that do not have the “constant rolloff” characteristic. The possibility exists that they may go into oscillation or become unstable if you are not careful and ignore the margins.

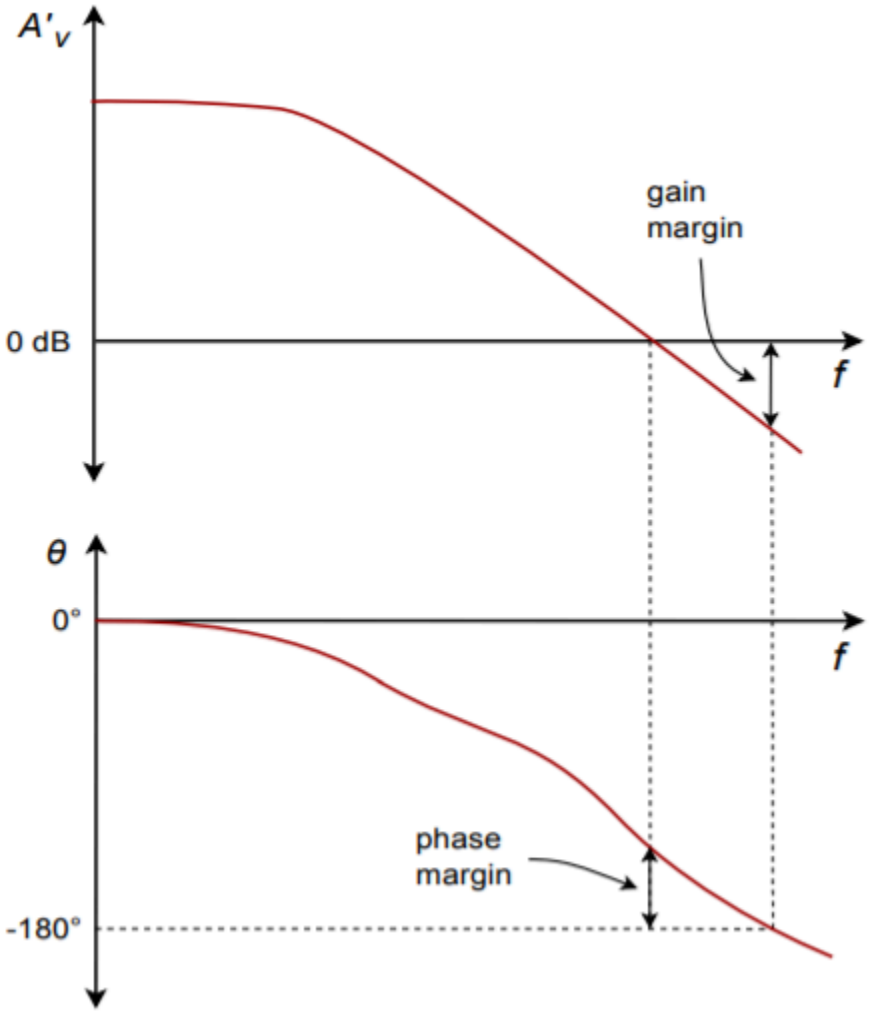


Figure 3.3.3 : Gain and phase margin graphically determined from Bode plot.

9.4 THE FOUR VARIANTS OF NEGATIVE FEEDBACK

Negative feedback can be achieved via four different forms. They differ in how the input and output impedances are changed. We have basically two choices when it comes to connecting the input and output of the amplifier to the output and input of the feedback network. We may produce either a series connection or a parallel connection. This yields four possibilities total. Each connection will produce a specific effect on the input or output impedance of the system. As you might guess, parallel connections decrease the impedance and series connections increase it. A high input impedance is desirable for maximum voltage transfer, whereas a low impedance is required for maximum current transfer. As a memory aid, think of volt meters and ammeters. For the smallest loading effect, volt meters should exhibit a high impedance and ammeters a very low one. In the case of the output connection, a low source impedance is required for the best voltage transfer, and a high source impedance for current transfer. You should think of the ideal voltage source (zero) and the ideal current source (infinite) here. Consequently, if we were to connect our feedback network in series with the amplifier's input, and in parallel with its output, we would have an increase in and a decrease in . This means that our system would be very good at sensing an input voltage, and ideal for producing a voltage. We will have created a voltage-controlled voltage source (VCVS), the ideal voltage amplifier. So that you can get a good idea of the possibilities, all four types are summarized in the following table.

Type (in-out)			**Model**	**Idealization**	**Transfer Ratio**
Series-Parallel	High	Low	VCVS	Voltage Amplifier	/ Voltage gain
Series-Series	High	High	VCCS	Voltage to Current Transducer	/ Transconductance
Parallel-Parallel	Low	Low	CCVS	Current to Voltage Transducer	/ Transresistance
Parallel-Series	Low	High	CCCS	Current Amplifier	/ Current Gain

Table 9.4.1

Generally speaking, the input and output impedances will be raised or lowered from the non-feedback value by the sacrifice factor. Note that by using the proper form of feedback, we can achieve any of the possible models. This greatly enhances our ability to deal with specific applications. Much work in our field relies on optimal voltage transfer, therefore Series-Parallel (SP) is often used. A variation on Parallel Parallel (PP) is also used frequently, as we shall see.

SERIES-PARALLEL (SP)

The SP connection makes for the ideal voltage amplifier. A generalized block diagram is shown in Figure 3.4.1 . You can tell that it has a series input because there is no input current node. Contrast this with the output: Note that the op amp's output current splits into two paths, one through the load and the other into the feedback network. This output node clearly denotes the feedback's parallel output connection.

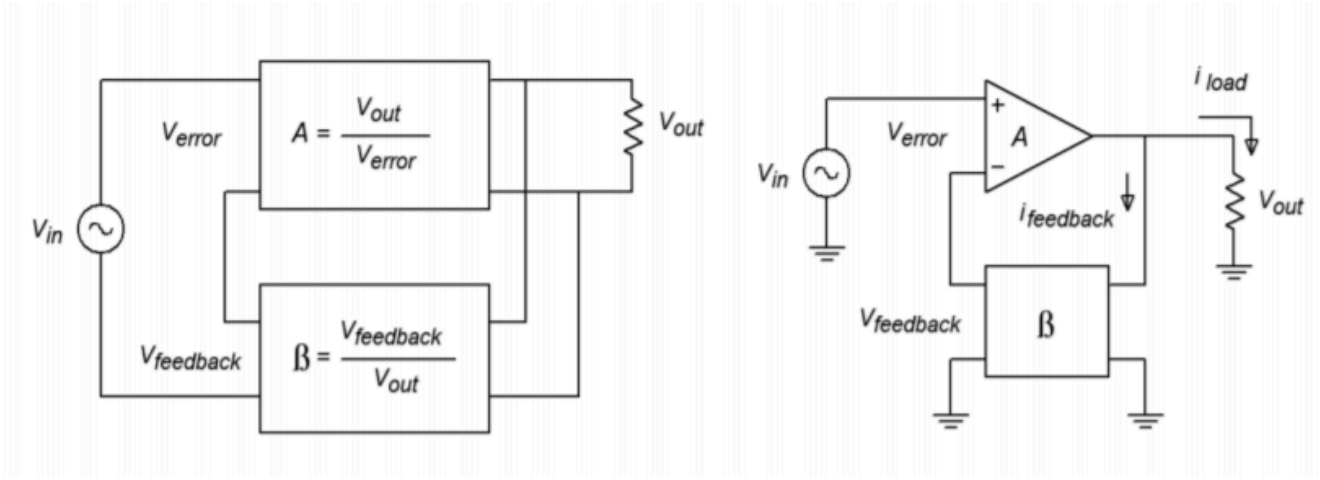


Figure 3.4.1 : Series-parallel connection.

Let's take a look at exactly how SP negative feedback alters the system gain, impedances, and frequency response. For starters, let's examine the closed-loop gain (◈◈◈). Our amplifier block produces a gain ◈ , and could be a diff amp, op amp, or other multi-stage possibility. The feedback network is typically a voltage divider and produces a loss, ◈ . The signal presented to the inverting input of the amp is the feedback signal and is equal to ◈◈◈◈◈ . Note that the source's signal, ◈◈◈, is applied to the non inverting input. Therefore, the differential input voltage (usually referred to as ◈◈◈◈◈◈), equals ◈◈◈–◈◈◈◈◈◈◈◈◈ . We also know from previous work with differential amplifiers that that ◈◈◈◈=◈◈◈◈◈◈◈◈◈ (◈◈◈ is the amplifiers open-loop gain). In other words we know:

$$V_{in} = V_{error} + V_{feedback}$$

(9.4.1)

$$V_{feedback} = V_{out}\beta$$

(9.4.2)

$$V_{out} = V_{error}A_{ol}$$

(9.4.3)

and by definition,

$$A_{sp} = \frac{V_{out}}{V_{in}}$$

(9.4.4)

Substituting Equation 3.4.3 into Equation 3.4.2 ,

$$V_{feedback} = V_{error}\beta A_{ol}$$

(9.4.5)

Substituting Equation 3.4.5 into Equation 3.4.1 , and simplifying,

$$V_{in} = V_{error}(1 + \beta A_{ol})$$

(9.4.6)

Finally, substituting Equation 3.4.6 and Equation 3.4.3 into Equation 3.4.4 and simplifying, yields,

$$A_{sp} = \frac{A_{ol}}{1 + \beta A_{ol}}$$

(3.4.7)

Since the fundamental definition of sacrifice factor, ◈ , is ◈◈◈/◈◈◈ , we may also say ◈◈◈=◈◈◈/◈ and therefore, for SP,

$$S = 1 + \beta A_{ol}$$

(3.4.8)

Equation 3.4.7 is our general gain Equation but, if we can make ◈◈◈◈≫1 we may ignore the "+1" in the denominator and further simplify this as,

$$A_{sp} = \frac{1}{\beta}$$

(9.4.9)

This seemingly innocent Equation packs a rather hefty punch. What it is telling us is that the open-loop gain of the amplifier does not} play a role in setting the system gain, as long as the open-loop gain is very large. In other words, the system gain is controlled solely by the feedback network. Consequently, our amplifier can exhibit large gain changes in its open-loop response, but the closed-loop response will remain essentially constant. For this reason we will achieve identical closed-loop gains for op amps that exhibit sizable differences in their open-loop gains. Because signal distortion is produced by non-linearities that can be viewed as dynamic gain changes, our closed-loop distortion drops as well. Also, it is this very effect that extends our closed-loop frequency response. Imagine that our amplifier exhibits a gain of 10,000 at its upper break frequency of 100 Hz. If the feedback factor is equal to 0.1, our exact gain is:

$$A_{sp} = \frac{10,000}{1 + 0.1 \times 10,000} = 9.99$$

If we were to measure the amplifier's open-loop gain one decade up, at 1 kHz, it should be around 1,000 (assuming 20 dB/decade loss). The closed loop gain now equals:

$$A_{sp} = \frac{1,000}{1 + 0.1 \times 1,000} = 9.9$$

As you can see, the closed-loop gain changed only about 1 despite the fact that the open-loop gain dropped by a factor of 10. If we continue to raise the frequency, Asp would equal 9.09 at 10 kHz. Finally, at 100 kHz a sizable drop is seen because the gain falls to 5. At this point, our assumption of ◈◈◈◈>>1 falls apart. Note however, that our loss relative to the midband gain is only a few dB. We have effectively stretched out the bandwidth of the system. Actually, this calculation is somewhat over-simplified as we have ignored the extra phase lag produced by the amplifier above the open-loop break frequency. If we assume that the open-loop response is dominated by a single lag network (and it should be, in order to guarantee stability, remember?), a phase sensitive version of Equation 3.4.7 would be:

$$A_{sp} = \frac{-jA_{ol}}{1 - jA_{ol}\beta}$$

This extra phase will reach its maximum of -90 degrees approximately one decade above the open-loop break frequency. Consequently, when we find the magnitude of gain at 100 kHz, it's not

$$A_{sp} = \frac{10}{1 + 1}$$

but rather

$$A_{sp} = \frac{10}{\sqrt{1^2 + 1^2}}$$

which equals 7.07, for a -3 dB relative loss

A simpler way of stating all of this is: The new upper break frequency is equal to the open-loop upper break times the sacrifice factor, ◈ . Because ◈ is the loop gain, it is equal to ◈◈◈/◈◈◈ . Note that our low frequency ◈ = 10,000/10, or 1,000. Therefore, our closed-loop break equals 1,000 times 100 Hz, or 100 kHz. A very important item to notice here is that there is an inverse relation between closed-loop gain and frequency response. Systems with low gains will have high upper-breaks, while high gain systems will suffer from low upper-breaks. This sort of trade-off is very common. Although most diff amps and op amps do not have lower break frequencies, circuits that do will see an extension of their lower response in a similar manner. (i.e., the lower break will be reduced by ◈). In order to achieve both high gain and wide bandwidth, it may be necessary to cascade multiple low gain stages.

Example 9.4.1

Assume that you have an amplifier connected as in Figure 3.4.1 . The open loop gain (◈◈◈) of the amp is 200 and its open loop upper break frequency (◈2–◈◈) is 10 kHz. If the feedback factor (◈) is 0.04, what are the closed loop gain (◈◈◈) and break frequency (◈2–◈◈)

For ◈◈◈ ,

$$A_{sp} = \frac{A_{ol}}{1 + \beta A_{ol}}$$

$$A_{sp} = \frac{200}{1 + 0.04 \times 200}$$

$$A_{sp} = 22.22$$

The approximation says,

$$A_{sp} = \frac{1}{\beta}$$

$$A_{sp} = \frac{1}{0.04}$$

$$A_{sp} = 25$$

That is reasonably close to the general equation's answer (note that there is no need to include phase effects as we are looking for the midband gain). The approximation is more accurate when ◈◈◈ is larger.

For ◈2–◈◈ , first find the sacrifice factor, ◈

$$S = \frac{A_{ol}}{A_{sp}}$$

$$S = \frac{200}{22.22}$$

$$S = 9$$
$$f_{2-sp} = f_{2-ol}S$$
$$f_{2-sp} = 10kHz \times 9$$
$$f_{2-sp} = 90kHz$$

One interesting thing to note is that the product of the gain and upper break frequency will always equal a constant value, assuming a 20 dB per decade roll off. Our open loop product is 200 times 10 kHz, or 2 MHz. Our closed loop product is 22.22 times 90 kHz, which is 2 MHz. If we choose any other feedback factor, the resulting ◈◈◈ and ◈2–◈◈ will also produce a product of 2 MHz (try it and see). The reason for this is simple. A 20 dB per decade rolloff means that the gain drops by a factor of 10 when the frequency is increased by a factor of 10. There is a perfect 1:1 inverse relationship between the two parameters. No matter how much you increase one parameter, the other one will decrease by the same proportion. Thus, the product is a constant.

At this point you may be asking yourself, "What exactly is in that feedback network and how do I figure out ◈ ?" Usually, the feedback network just needs to produce a loss – it has to scale ◈◈◈◈down to ◈◈◈◈◈◈◈◈◈◈ . The simplest item for the job would be a resistive voltage divider. (It is possible to have complex frequency dependent or non-linear elements in the network as we shall see in the future). An example is presented in Figure 3.4.2 . If you study this diagram for a moment, you will notice that the feedback factor ◈ is really nothing more than the voltage divider loss. ◈◈◈◈is the input to the feedback network and appears across ◈◈+◈◈||◈◈◈ . The output of the network is ◈◈◈◈◈◈◈◈◈◈ , which appears across ◈◈ . Simply put, the ratio is

$$\beta = \frac{R_i||Z_{in}}{R_f + R_i||Z_{in}}$$

If the ◈◈◈ is large enough to ignore, as in most op amps, this simplifies to

$$\beta = \frac{R_i}{R_f + R_i}$$

By substituting this Equation into our approximate gain Equation 3.4.9 , we find

$$A_{sp} = R_f + R_i Ri = \frac{R_f}{R_i} + 1$$

(9.4.10)

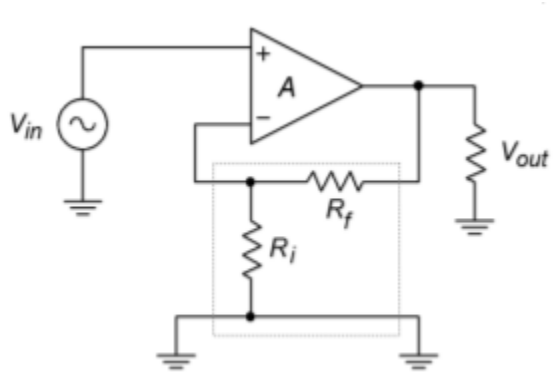


Figure 3.4.2 : Simple voltage divider for ◈ .

Note that the values of ◈◈ and ◈◈ are not really important, rather, their ratio is. We would arrive at the same gain if ◈◈=10◈Ω and ◈◈=1◈Ω , or ◈◈=20◈Ω and ◈◈=2◈Ω . We obviously have quite a bit of latitude when designing circuits for a specific gain, but we do face a few practical limits. If the resistors are too small we will run into problems with op amp output current. On the other hand, if the resistors are too large, excessive noise, offset, drift and loading effects will result. As a guideline for general-purpose circuits, ◈◈+◈◈ is usually in the range of 10 k to 100 k Ω .

COMPUTER SIMULATION

As evidenced earlier, if the open loop gain is very high, its precise value does not matter. We'll examine this effect using Example 3.4.1 , but with much higher open loop gains. The simulation is shown in Figure 3.4.3 using Multisim and the basic op amp model presented in Chapter Two. The circuit uses ◈◈=24◈Ω and ◈◈=1◈Ω for an ideal closed loop voltage gain of 25. The input signal is set to 40 millivolts.

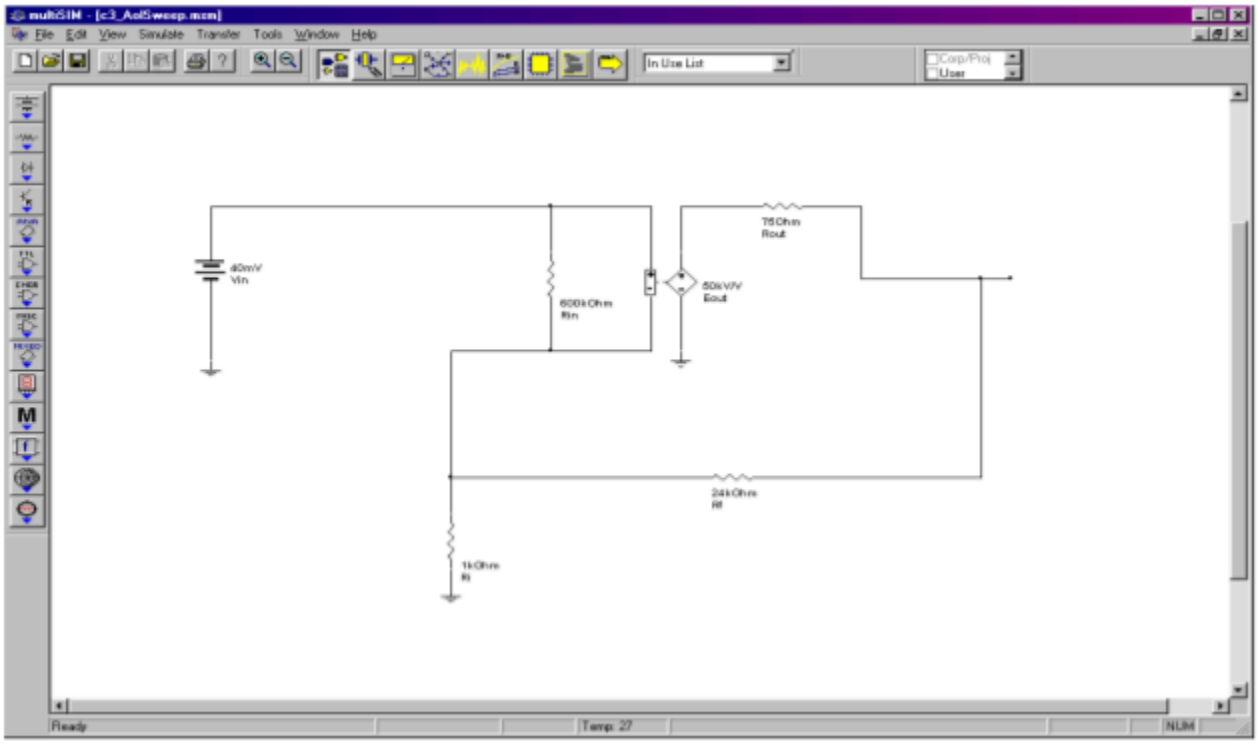


Figure 3.4.3 : Multisim schematic for gain sweep.

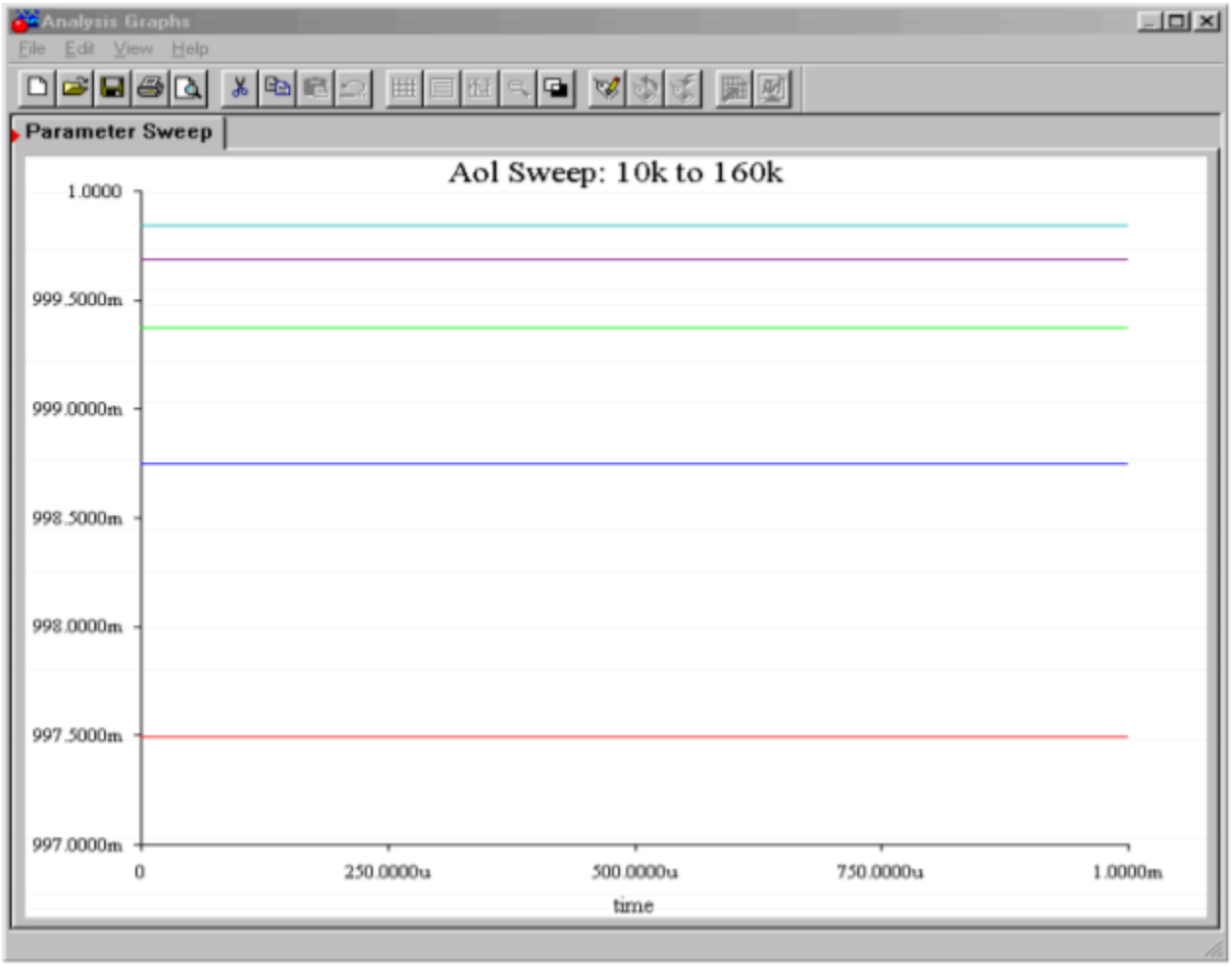


Figure 3.4.3◈: Output variation due to open-loop gain change.

Using the Parameter Sweep option, the open loop gain is initialized at 10,000 and is progressively doubled to a maximum of 160,000. These are reasonable gain values for a production op amp. The output transient analysis shows that in spite of a 16 to 1 variation in open loop gain, all output voltages are approximately 1 V, achieving a closed loop gain of nearly 25 in all cases.

Example 9.4.2

Let's say that the microphone that you use for acoustic instruments produces a signal that is just too weak for you to record without excessive noise. After a little experimentation in lab, you discover that you need about 20 dB of voltage gain before you can successfully capture a softly picked guitar. Using Figure 3.4.4 as a guide, design this amplifier with the following device: ◈◈◈ = 50,000, ◈◈◈–◈◈=600◈Ω .

First, note that our ◈◈◈ and ◈◈◈–◈◈ values are more than sufficient for us to use the approximation formulas. Because our formulas all deal with ordinary gain, we must convert 20 dB.

$$A_{sp} = log^{-1}\frac{A'_{sp}}{20}$$

$$A_{sp} = log^{-1}\frac{20dB}{20}$$

$$A_{sp} = 10$$

By rearranging Equation 3.4.10 ,

$$\frac{R_f}{R_i} = A_{sp} - 1$$

$$\frac{R_f}{R_i} = 9$$

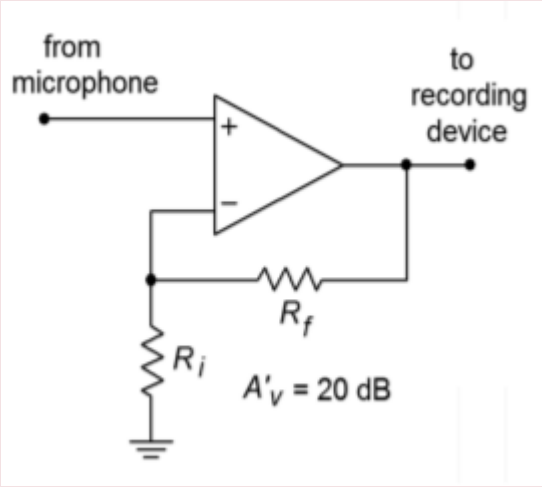


Figure 3.4.4 : Microphone amplifier.

We see that ⬦⬦ must be 9 times larger than ⬦⬦ . There is no single right answer here; there are many possibilities.

COMPUTER SIMULATION

One viable solution for Example 3.4.2 is simulated using Multisim in Figure 3.4.5 . This circuit uses ⬦⬦=9⬦Ω and ⬦⬦=1⬦Ω . The op amp model is the simple dependent source version examined in Chapter Two. The input signal is set to 0.1 V DC for simplicity. Both the output and feedback potentials are presented. The results of this simulation verify our hand calculations. In order to note the sensitivity of the design, you can alter certain parameters of the input file and rerun the simulation. Two of the more interesting areas are the absolute values of the feedback resistors and the open loop gain of the op amp. You will note that as these quantities are lowered, our approximation formulas become less accurate. With the given values, the approximations deviate from the simulation results by less than 1.

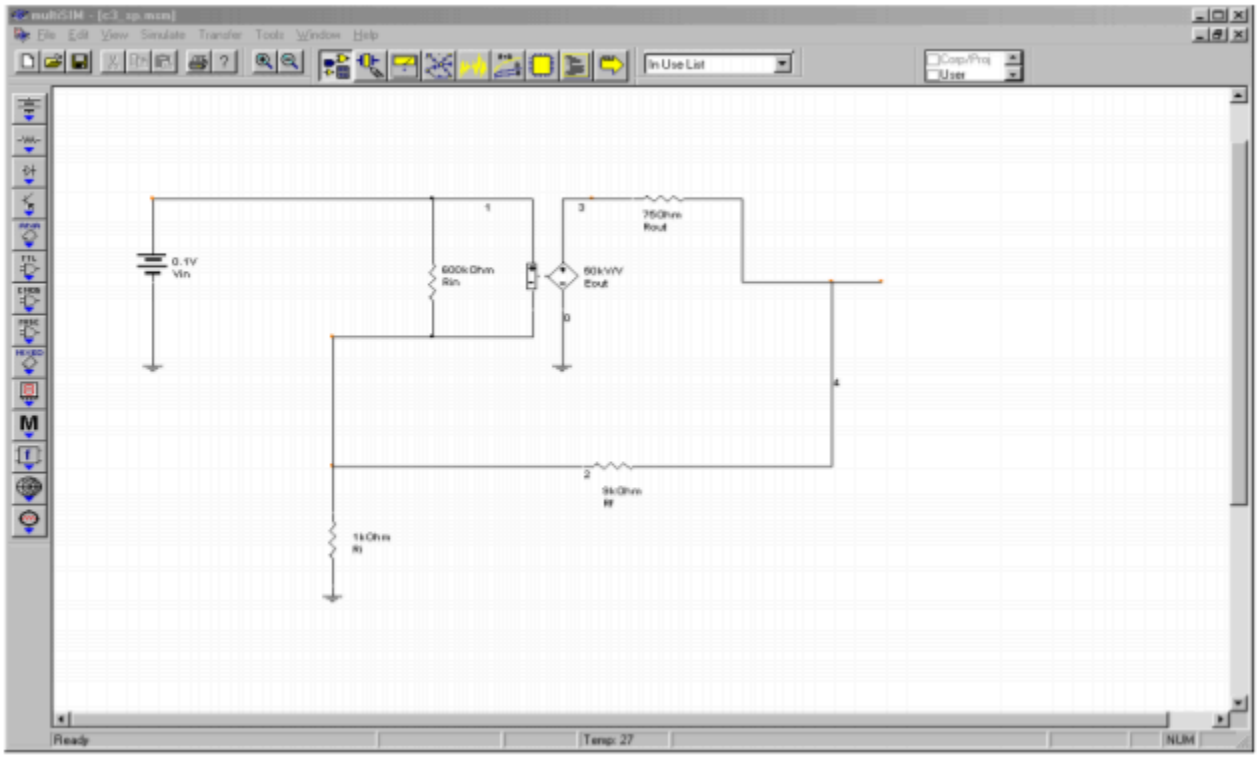


Figure 3.4.5 : Multisim schematic for analysis of the simple op amp model of Example 3.4.2 .

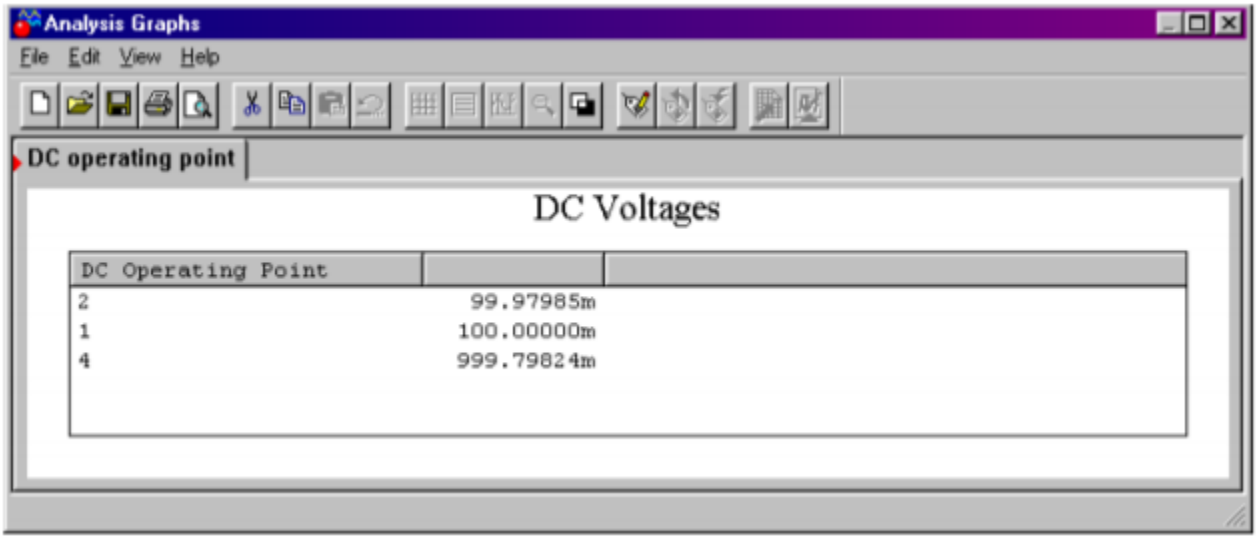


Figure 3.4.5 : Multisim DC outputs.

SP IMPEDANCE EFFECTS

As noted earlier, negative feedback affects the closed-loop input and output impedances of our system. Series connections increase the impedance and parallel connections decrease the impedance.

Let's see exactly how this works in the SP case. First, we must distinguish between the ◇◇◇ of the amplifier itself and the ◇◇◇ of the system with feedback. We shall call them ◇◇◇–◇◇ and ◇◇◇–◇◇ respectively. Figure 3.4.6 shows this difference by using a simple model of the amplifier. By definition,

$$Z_{in-sp} = \frac{V_{in}}{I_{in}}$$

The idea here is to notice that the source only needs to supply enough signal current to develop the ◇◇◇◇◇◇ drop across the amplifier's ◇◇◇–◇◇ . As far as the ◇◇◇signal source is concerned, ◇◇◇◇◇◇◇◇◇ is a voltage source, not a voltage drop. Therefore, ◇◇◇=◇◇◇◇◇◇/◇◇◇–◇◇ .

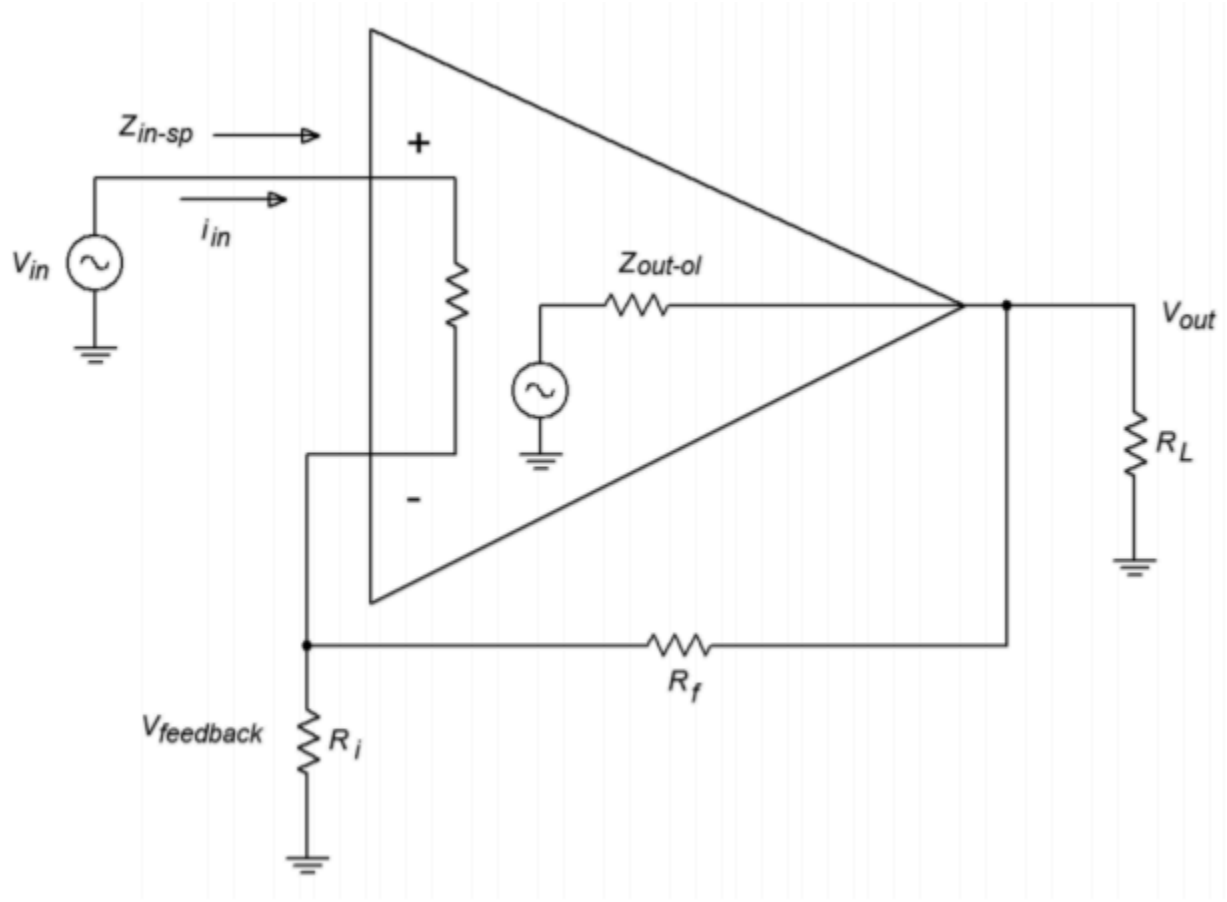


Figure 3.4.6 : Series-parallel input impedance.

We may now say,

$$Z_{in-sp} = \frac{V_{in}}{\left(\frac{V_{error}}{Z_{in-ol}}\right)} = \frac{Z_{in-ol}V_{in}}{V_{error}}$$

Because ◇◇◇◇◇◇ ideally equals ◇◇◇◇/◇◇◇

$$Z_{in-sp} = \frac{Z_{in-ol}A_{ol}V_{in}}{V_{out}} = \frac{Z_{in-ol}A_{ol}}{A_{sp}}$$

Sacrifice factor ◇ , is defined as ◇◇◇/◇◇◇ , so,

$$Z_{in-sp} = Z_{in-ol}S$$

This is our ideal SP input impedance. Obviously, even moderate open-loop ◇◇◇◇ with moderate sacrifice factors can yield high closed-loop ◇◇◇◇ . The upper limit to this will be the impedance seen from each input to ground. In the case of a typical op amp, this is sometimes referred to as

the common-mode input impedance, ◈◈◈–◈◈ , and can be very high (perhaps hundreds of mega ohms). This is the impedance presented to common mode signals. This value effectively appears in parallel with our calculated ◈◈◈–◈◈ , above. An example is shown in Figure 3.4.7 . Note that because ◈◈◈–◈◈ is measured with the inputs of the op amp in parallel, each input has approximately twice the value to ground. In the case of a discrete amplifier, you would be most concerned with the noninverting input's ◈◈◈ . In any case, because ◈ drops as the frequency increases, ◈◈◈–◈◈ decreases as well. At very high frequencies, input and stray capacitances dominate, and the real system input impedance may be a small fraction of the low frequency value. Negative feedback cannot reduce effects that live outside of the loop.

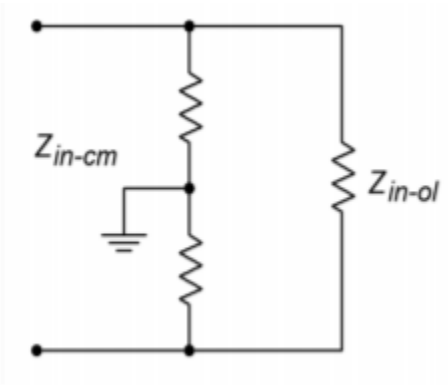


Figure 3.4.7 : Common -mode input impedance.

Now for our ◈◈◈◈–◈◈ . Refer to Figure 3.4.8 . ◈◈◈◈–◈◈ is the Thevenin output impedance. In order to find this, we will drive the amplifier's output with a voltage source and reduce all other independent voltage sources to zero (we have no current sources. Remember, this is a "paper" analysis technique and may not work in lab due to other factors). By figuring out the resulting output current, we can find ◈◈◈◈–◈◈ (by definition, ◈◈◈◈–◈◈=◈◈◈◈/◈◈◈◈).

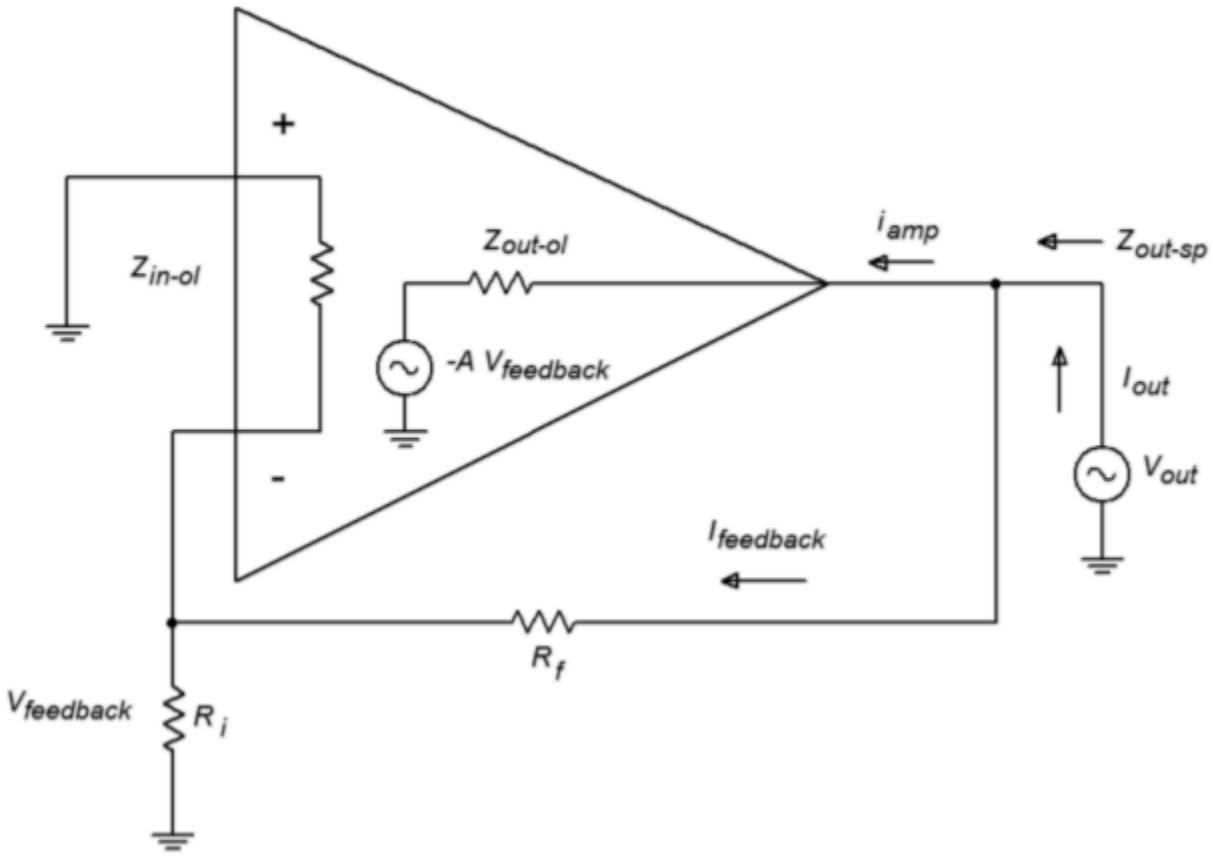


Figure 3.4.8 : Series-parallel output impedance.

First, notice that ⓪⓪⓪⓪ is made of two pieces, ⓪⓪⓪⓪⓪⓪⓪⓪⓪ and ⓪⓪⓪⓪ . If we can find the two impedances associated with these parts, we can simply perform a parallel equivalent in order to determine ⓪⓪⓪⓪–⓪⓪ . The ⓪⓪⓪⓪⓪⓪⓪⓪⓪ portion is very easy to determine. Ignoring the inverting input's loading effects on ⓪⓪ , this impedance is just ⓪⓪+⓪⓪ . Finding the output impedance of the amplifier itself is a little more involved. ⓪⓪⓪⓪ is found by taking the drop across ⓪⓪⓪⓪–⓪⓪ and using Ohm's Law. The voltage across ⓪⓪⓪⓪–⓪⓪ is the difference between ⓪⓪⓪and the signal created by the feedback path to the inverting input. This signal is –⓪⓪⓪⓪⓪⓪⓪⓪⓪⓪⓪⓪ .

$$I_{amp} = \frac{V_{out} - (-A_{ol}V_{feedback})}{Z_{out-ol}}$$

Since ⓪⓪⓪⓪⓪⓪⓪⓪⓪=⓪⓪⓪⓪⓪

$$I_{amp} = \frac{V_{out} + A_{ol}\beta V_{out}}{Z_{out-ol}}$$

$$I_{amp} = \frac{V_{out} + (1 + A_{ol}\beta)}{Z_{out-ol}}$$

By using Equation 3.4.8 , this may be simplified to,

$$I_{amp} = \frac{SV_{out}}{Z_{out-ol}}$$

Because

$$Z_{outamp} = \frac{V_{out}}{I_{amp}}$$

This is the part of ⬚⬚⬚⬚–⬚⬚ contributed by ⬚⬚⬚⬚ . To find ⬚⬚⬚⬚–⬚⬚ , just combine the two pieces in parallel:

$$Z_{out-sp} = \frac{Z_{out-ol}}{S} || (R_f + R_i)$$

With op amps, ⬚⬚+⬚⬚ is much larger than the first part and can be ignored. For example, a typical device may have ⬚⬚⬚⬚–⬚⬚=75Ω . Even a very modest sacrifice factor will yield a value many times smaller than a typical ⬚⬚+⬚⬚ combo (generally over 1⬚Ω). Discrete circuits using common emitter or common base connections will have larger ⬚⬚⬚⬚–⬚⬚ values, and therefore the feedback path may produce a sizable effect. As in the case of input impedance, ⬚⬚⬚⬚–⬚⬚ is a function of frequency. As ⬚ decreases as the frequency rises, ⬚⬚⬚⬚–⬚⬚ will increase.

Example 9.4.3

An op amp has the following open-loop specs: ⬚⬚⬚=300⬚Ω ⬚⬚⬚⬚=100Ω and A = 50,000. What are the low frequency system input and output impedances if the closed loop gain is set to 100?

First we must find ⬚

$$S = \frac{A_{ol}}{A_{sp}}$$

$$S = \frac{50,000}{100}$$

$$S = 500$$

We may now find the approximate solutions.

$$Z_{in-sp} = SZ_{in-ol}$$

$$Z_{in-sp} == 500 \times 300\ k\Omega = 150\ M\Omega$$

$$Z_{out-sp} = \frac{Z_{out-ol}}{S}$$

$$Z_{out-sp} = \frac{100}{500} = 0.2\Omega$$

The effects are quite dramatic. Note that with such high ⬚⬚⬚ values, op amp circuits may be used in place of FETs in some applications. One example would be the front-end amplifier/buffer in an electrometer (electrostatic voltmeter).

DISTORTION EFFECTS

As noted earlier, negative feedback lowers static forms of distortion, such as THD. The question, as always, is "How much?" Something like harmonic distortion is internally generated, so we may model it as a voltage source in series with the output source. An example of this model can be seen in Figure 3.4.9 . ◆◆◆◆◆ is the distortion generator. In the case of a simple input sine wave, ◆◆◆◆◆ will contain harmonics at various amplitudes (sine waves at integer multiples of the input frequency). These amplitudes are directly related to the input signal's amplitude. If we assume that ◆◆◆◆ is small enough to ignore, ◆◆◆◆◆ will appear at the output of the open-loop circuit. Thus, the total output voltage is the desired ◆◆◆◆plus ◆◆◆◆◆ .

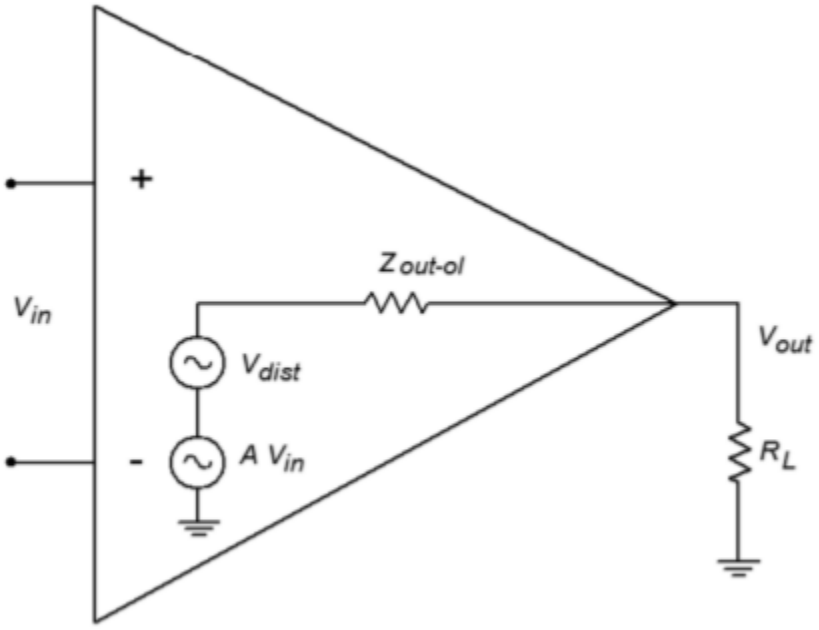


Figure 3.4.9 : Distortion model (open-loop)

When we add feedback, as in Figure 3.4.10 , this distortion signal is fed back to the inverting input, and because it is now out of phase, it partially cancels the internally generated distortion. Thus the SP distortion signal (◆◆◆◆–◆◆) is much smaller.

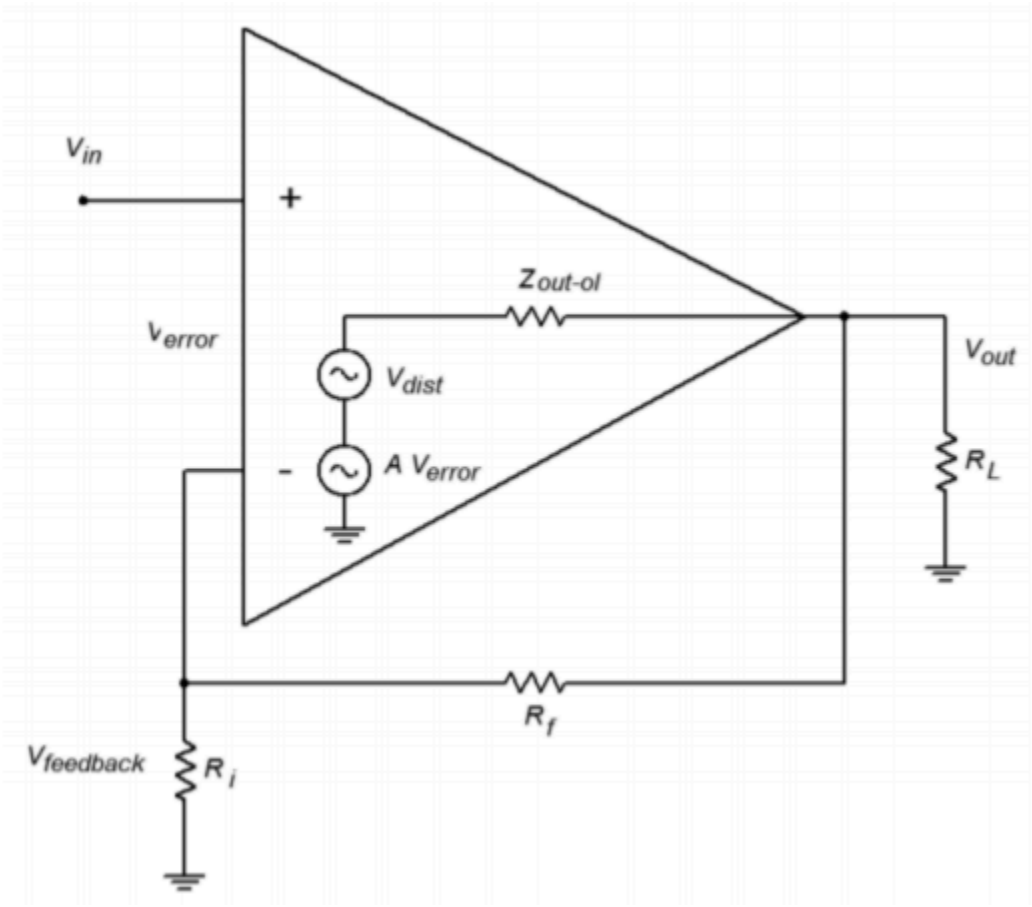


Figure 3.4.10 : Distortion model (closed-loop)

The SP output signal is

$$V_{out} = A_{ol}V_{error} + V_{dist}$$
$$V_{out} = A_{ol}(V_{in}V_{feedback}) + V_{dist}$$
$$V_{out} = A_{ol}(V_{in} - \beta V_{out}) + V_{dist}$$

We now perform some algebra in order to get this into a nicer form and solve for ◈◈◈◈

$$V_{out} = A_{ol}V_{in} - A_{ol}\beta V_{out} + V_{dist}$$
$$V_{out} + A_{ol}\beta V_{out} = A_{ol}V_{in} + V_{dist}$$
$$V_{out}(1 + A_{ol}\beta) = A_{ol}V_{in} + V_{dist}$$

Remember that 1+◈◈◈◈=◈ , so,

$$V_{out}S = A_{ol}V_{in} + V_{dist}$$
$$V_{out} = A_{ol}\frac{V_{in}}{S} + \frac{V_{dist}}{S}$$

Because ◈◈◈/◈ is just ◈◈◈ , this reduces to,

$$V_{out} = A_{sp}V_{in} + \frac{V_{dist}}{S}$$

The internally generated distortion is reduced by the sacrifice factor. As you can see, large sacrifice factors can drastically reduce distortion. An amplifier with 10% THD and a sacrifice factor of 100 produces an effective distortion of only 0.1%. This analysis does assume that the open-loop distortion is not overly grotesque. If the distortion is large, we cannot use this superposition approach

(remember, superposition assumes that the circuit is essentially linear). Also, we are ignoring any additional distortion created by feeding this distortion back into the amplifier. For any reasonably linear amplifier, this extra distortion is a second-order effect, and thus constitutes only a small part of the total output signal.

COMPUTER SIMULATION

An example of the reduction of distortion is simulated using the Distortion Analyzer from Multisim. A basic amplifier is shown in Figure 3.4.11 . The value for ⍰⍰ is changed from 999 k Ω to 99 k Ω to 9 k Ω . ⍰⍰⍰is scaled accordingly so that the output of the amplifier remains at approximately 10 volts. The test frequency was set to 1 kHz and a total of 20 harmonics (up to 20 kHz) were used in the analysis. Using the LF411 op amp, the high gain version shows a THD of 0.09%. Reducing the gain by a factor of 10 (and thus increasing sacrifice factor by 10 fold) yields a THD of 0.011%. Finally, a further 10 fold reduction of gain yields a THD of 0.001%. Although the reduction in distortion is not exactly a factor of 10 each time, the trend can be seen clearly. The precise distortion values will depend on the accuracy of the op amp model used, the test frequency, and the number of harmonics kept in the analysis. 20 kHz was chosen here because that represents the upper limit of human hearing, and thus it would be appropriate for an audio amplifier.

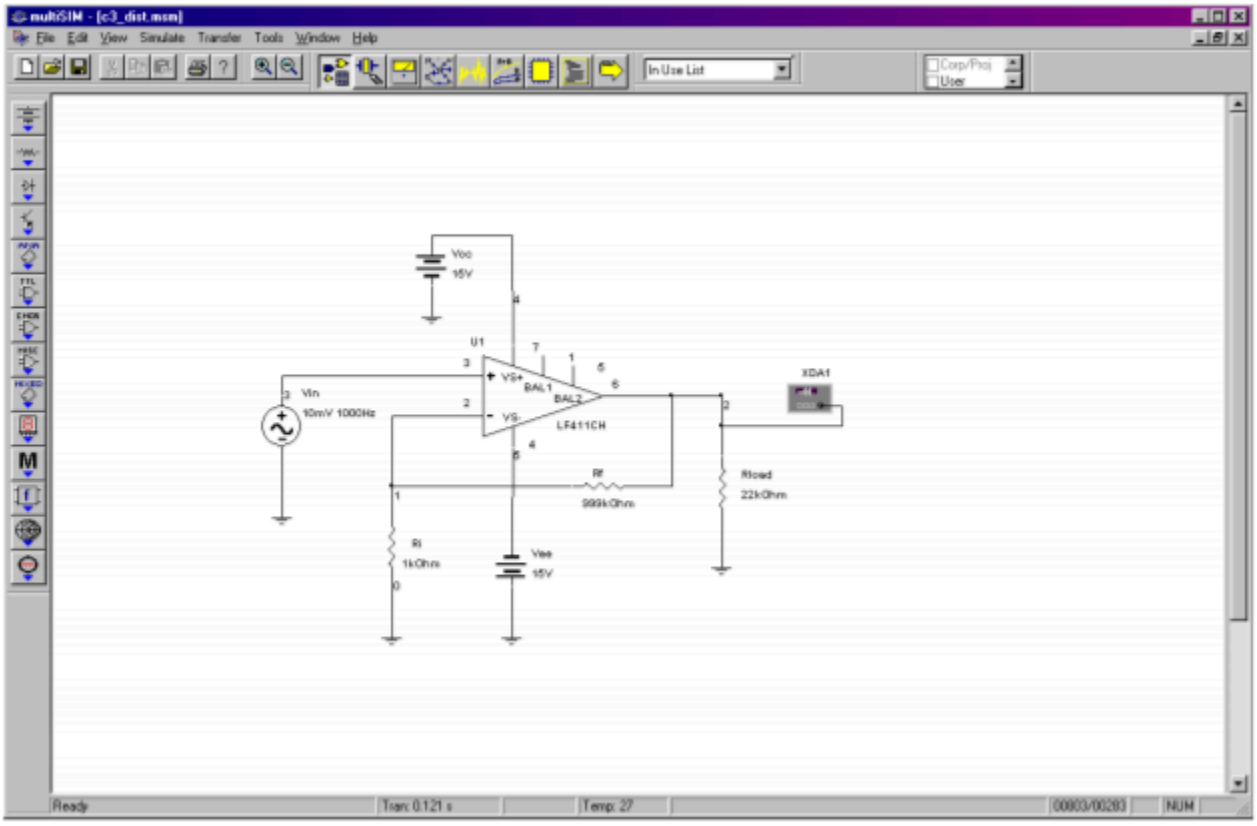


Figure 3.4.11⍰: Multisim distortion analysis

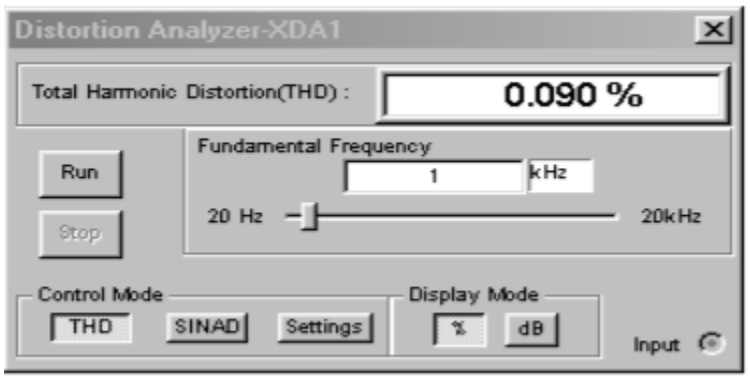


Figure 3.4.11◈: Distortion analyzer results for high-gain version.

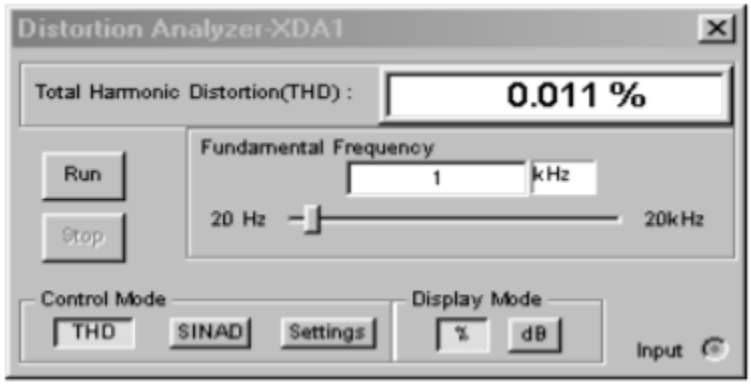


Figure 3.4.11◈: Distortion analyzer results for medium-gain version.

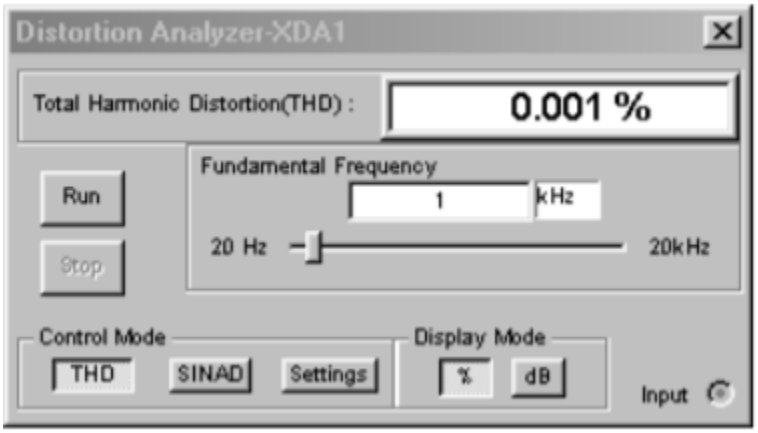


Figure 3.4.11◈: Distortion analyzer results for low-gain version.

As we have seen, the sacrifice factor is a very useful item. Our gain, distortion, and ◈◈◈◈ are all reduced by ◈ , while ◈$_2$ and ◈◈◈ are increased by ◈ .

NOISE

It is possible to model noise effects in much the same way as we just modeled distortion effects. By doing so, you will discover that noise can also be decreased by a large amount. Unfortunately, there is one major flaw. Unlike our distortion generator, a noise generator will produce a signal that is not dependent on the input signal. The net result is that although the noise level does drop by the sacrifice factor, so does the desired output signal. Thus, the signal-to-noise ratio at the output is unchanged.

In contrast, the distortion signal is proportional to the input signal, so that when the desired signal is cut by ◈ , the distortion signal sees a further cut by ◈ (i.e., the distortion drops by ◈ relative to the desired signal). As a matter of fact, it is quite possible that the noise produced in following stages may add up to more noise than the circuit had without feedback. Sad but true, negative feedback doesn't help us when it comes to signal-to-noise ratio.

PARALLEL-SERIES (PS)

The Parallel-Series connection is the opposite of the Series-Parallel form. PS negative feedback is used to make the ideal current amplifier. Its gain is dimensionless, but for convenience, it is normally given the units A/A (amps per amp). It produces a low ◈◈◈ (perfect for sensing ◈◈◈) and a high ◈◈◈◈ (making for an ideal current source). An example of PS is shown in Figure 3.4.12 .

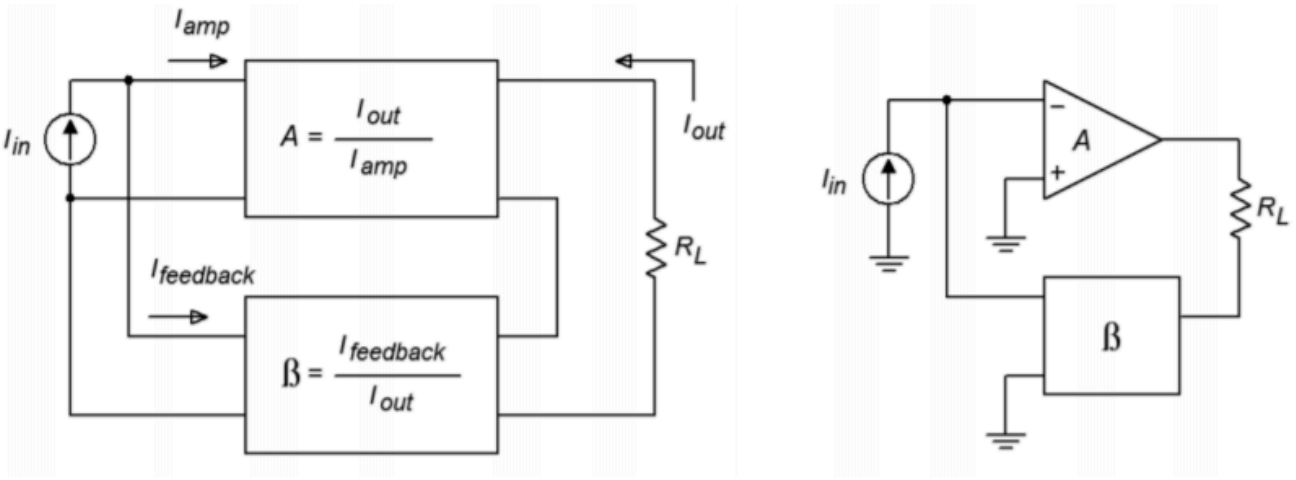


Figure 3.4.12 : Parallel-series connection.

The signal source's current splits in two, with part going into the amplifier, and part going through the feedback network. This is how you know that you have a parallel input connection. The output current on the other hand, passes through the load and then enters the feedback network, indicating a series output connection. Note that this general model is an inverting type, and that the load is floating (i.e., not ground referenced). It is possible though to use ground referenced loads with some additional circuitry. As PS is used for current amplification, let's see how we can find the current gain. Figure 3.4.13 will help us along. PS current gain is defined as:

$$A_{ps} = \frac{I_{out}}{I_{in}}$$

(9.4.11)

The signal source's current splits into two paths, so

$$I_{in} = I_{amp} + I_{feedback}$$

(9.4.12)

Because ◈◈◈◈ times the amplifier's open-loop current gain is ◈◈◈◈ , we can also say,

$$I_{amp} = \frac{I_{out}}{A_{ol}}$$

9.4.13)

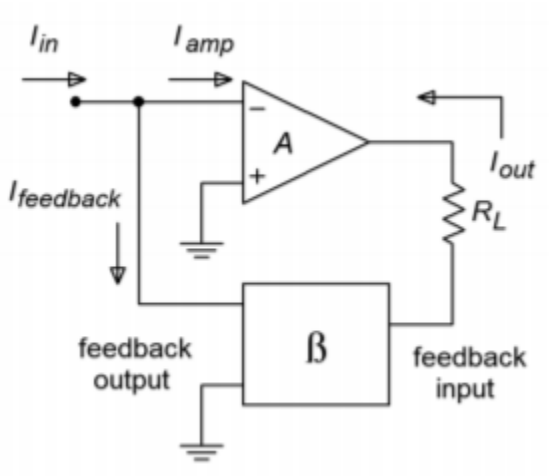


Figure 3.4.13 : Parallel-series analysis.

The feedback network is nothing more than a current divider, where the feedback network's output (◈◈◈◈◈◈◈◈◈) is ◈ times smaller than its input (◈◈◈◈). Don't let the arbitrary current direction fool you – the feedback "flow" is still from right to left, as always. In this case the amplifier is sinking current instead of sourcing it. As ◈ is just a fraction, we may say

$$\beta = \frac{I_{feedback}}{I_{out}} \text{or,}$$

$$I_{feedback} = \beta I_{out}$$

$$I_{feedback} = A_{ol}\beta I_{amp}$$

(9.4.14)

By substituting Equation 3.4.14 into Equation 3.4.12 we see that

$$I_{in} = I_{amp} + A_{ol}\beta I_{amp} \text{ or,}$$

$$I_{in} = I_{amp}(1 + A_{ol}\beta)$$

(9.4.15)

After substituting this into Equation 3.4.11 we find that

$$A_{ps} = \frac{I_{out}}{I_{amp}(1 + A_{ol}\beta)}$$

or, with the help of Equation 3.4.13

$$A_{ps} = \frac{I_{out}}{1 + A_{ol}\beta}$$

(9.4.16)

To make a long story short, the open-loop gain is reduced by the sacrifice factor. (Where have we seen this before?) The one item that you should note is that we have used only current gains in our derivation (compared to voltage gains in the SP case).

It is possible to perform a derivation using the open-loop voltage gain; however, the results are basically the same, as you might have guessed. Once again, our approximation for gain can be expressed as 1/◈ .

Example 3.4.4

A co-worker asks you to measure the output current of a circuit that she's just built. If all is working correctly, this circuit should produce 100 microamps. Unfortunately, your hand-held DMM will only accurately measure down to 1 milliamp. It is 10 times less sensitive than it needs to be. In order to be read accurately, the current will need to be boosted. Does the amplifier of Figure 3.4.14 have the current gain you need?

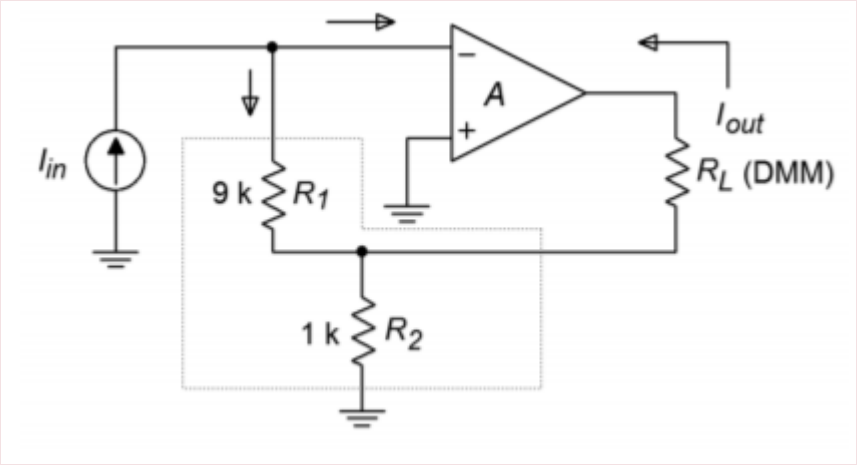


Figure 3.4.14 : Current amplifier for Example 3.4.4 .

If we are using an op amp, we can assume that ◈◈◈=1/◈ . The question then becomes, "what is ◈ ?" ◈ is the current divider ratio. The resistors ◈1 and ◈2 make up the current divider. The output current splits between ◈1 and ◈2 , where the ◈1 path is ◈◈◈◈◈◈◈◈◈ . According to the current divider rule,

$$\beta = \frac{R_2}{R_1 + R_2}$$

Consequently the gain magnitude is

$$A_{ps} = \frac{R_1 + R_2}{R_2} \text{or,}$$

$$A_{ps} = \frac{R_1}{R_2} + 1$$

This Equation looks a lot like the one in the SP example! Solving for ◈◈◈ yields

$$A_{ps} = \frac{9\ k}{1\ k} + 1$$

$$A_{ps} = 10$$

Yes, this circuit has just the gain you need. All you have to do is connect your co-worker's circuit to the input, and replace the load resistor with your hand-held DMM. Note that most DMMs are floating instruments and are not tied to ground (unlike an oscilloscope), so using it as a floating load presents no problems. The accuracy of the gain (and thus your measurement) depends on the accuracy of the resistors

and the relative size of the op amp's input bias current. Therefore, it would be advisable to use a bi-FET type device (i.e., an op amp with an FET diff amp) and precision resistors.

PS IMPEDANCE EFFECTS

As you have seen, the gain derivation for PS is similar to that for SP. The same is true for the impedance equations. First, we'll take a look at input impedance with Figure 3.4.15 .

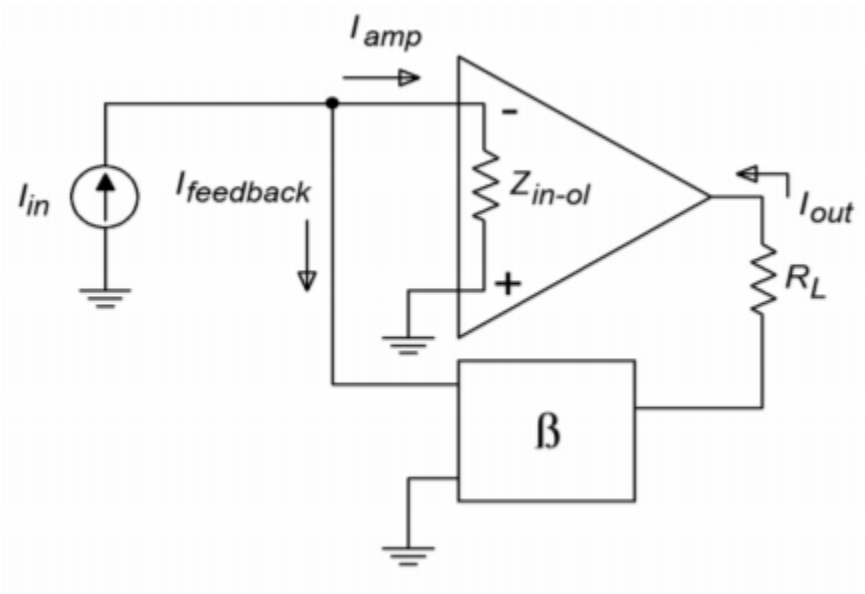


Figure 3.4.15 : Parallel-series input impedance.

As always, we start with our base definition,

$$Z_{in-ps} = \frac{V_{in}}{I_{in}}$$

Recalling Equation 3.4.15 , this can be rewritten as,

$$Z_{in-ps} = \frac{V_{in}}{I_{amp}(1 + A_{ol}\beta)}$$

(9.4.17)

◈◈◈is merely the voltage that appears from the inverting input to ground. By using Ohm's Law, we can say,

$$V_{in} = I_{amp}Z_{in-ol}$$

Where ◈◈◈–◈◈ is the open-loop input impedance. Finally, substituting this into Equation 3.4.17 gives,

$$Z_{in-sp} = \frac{I_{amp}Z_{in-ol}}{I_{amp}(1 + A_{ol}\beta)}$$

$$Z_{in-sp} = \frac{Z_{in-ol}}{(1 + A_{ol}\beta)}$$

The input impedance is lowered by the sacrifice factor. If you're starting to wonder whether everything is altered by the sacrifice factor, the answer is, yes. This is, of course, an approximation. What do you think is going to happen to the output impedance? At this point, it shouldn't be too surprising. For this proof, refer to Figure 3.4.16 . We shall use the same general technique to find ◈◈◈◈ as we did with the SP configuration. In this case we replace the load with a current source, and then determine the resulting output voltage.

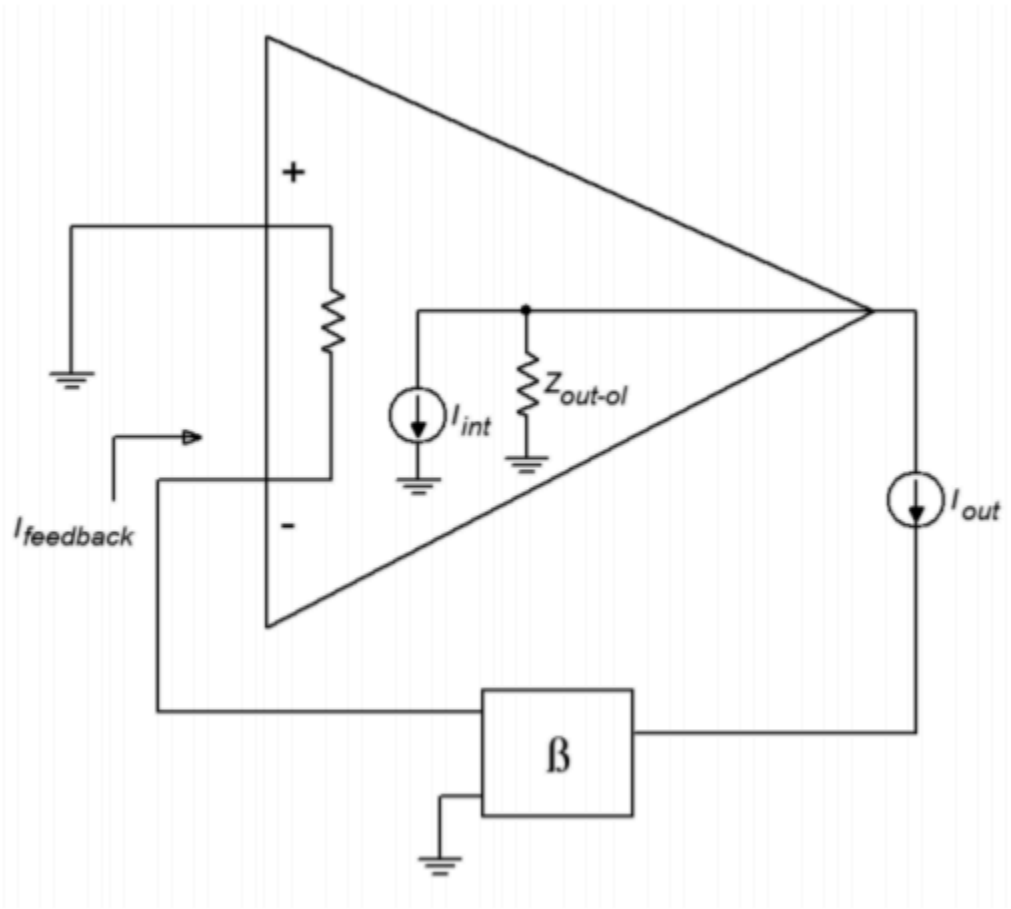


Figure 3.4.16 : Parallel-series output impedance.

Note that the input signal current source is opened. The output current drives the feedback network and produces the feedback current. The feedback current is then multiplied by the amplifier's open-loop current gain. Because the feedback current enters the inverting input, the internal source is sinking current. Because we are driving the circuit from the output,

$$Z_{out-ps} = \frac{V_{out}}{I_{out}}$$

◈◈◈◈is found through Ohm's Law

$$V_{out} = Z_{out-ol}(I_{out} + I_{int})$$

We now expand on our currents.

$$I_{int} = I_{feedback} A_{ol}$$

$$I_{int} = I_{out} \beta A_{ol}$$

$$V_{out} = Z_{out-ol}(I_{out} + I_{out} \beta A_{ol})$$

$$V_{out} = Z_{out-ol} I_{out} (1 + A_{ol}\beta)$$
$$V_{out} = Z_{out-ol} I_{out} S$$

Finally, we see that,

$$Z_{out-ps} = \frac{Z_{out-ol} I_{out} S}{I_{out}}$$

$$Z_{out-ps} = Z_{outol} S$$

As expected, the series output connection increased ⬦⬦⬦⬦ by the sacrifice factor. The remainder of the PS equations are essentially those used in our earlier SP work. Once again, the bandwidth will be increased by ⬦ , and the distortion will be reduced by ⬦ . This is also true for the Parallel-Parallel and Series-Series connections. As a matter of fact, the ⬦⬦⬦ and ⬦⬦⬦⬦ relations are just what you might expect. The proofs are basically the same as those already presented, so we won't go into them. Suffice to say that parallel connections reduce the impedance by ⬦ , and series connections increase it by ⬦ .

PARALLEL-PARALLEL (PP) AND SERIES-SERIES (SS)

Unlike our two earlier examples, the concept and modeling of gain is not quite as straight forward in the Parallel-Parallel and Series-Series cases. These forms do not produce gain, so to speak. They are neither voltage amplifiers, nor current amplifiers. Instead, these connections are used as transducers. Parallel-Parallel turns an input current into an output voltage. It is shown in Figure 3.4.17 .

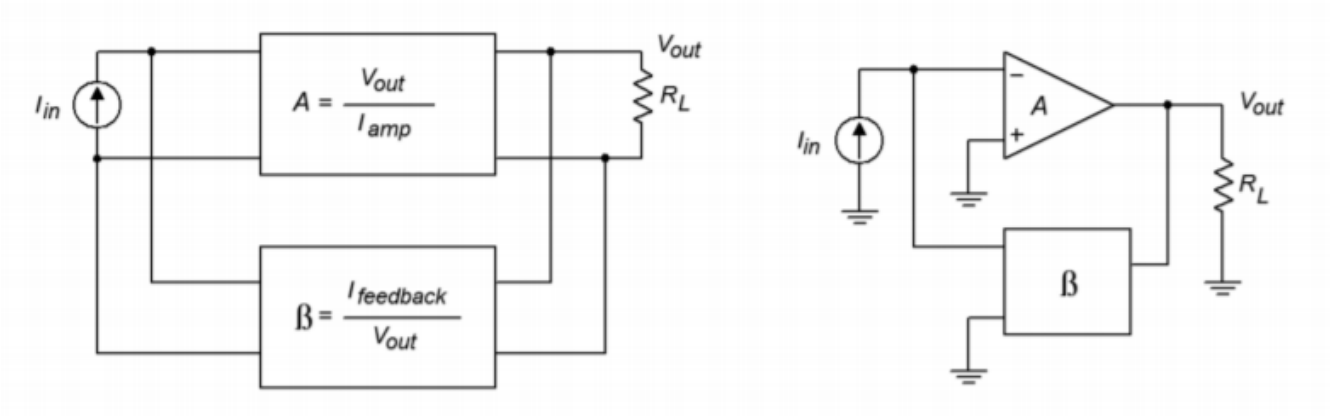


Figure 3.4.17 : Parallel-parallel connection.

Series-Series turns an input voltage into an output current. It is shown in Figure 3.4.18 . Our normal gain and feedback factors now have units associated with them. Because our output versus input quantities are measured in volts-per-amp (PP) or amps-per-volt (SS), the appropriate units for our factors are Ohms and Siemens. To be specific, we refer to our PP gain as a transresistance value, and the SS gain as a transconductance value. Although it is quite possible (and useful) to derive formulas for gain based on these, it is often done only for discrete designs. Quite simply, you cannot find transresitance or transconductance values on typical op amp data sheets.

Fortunately, we can make a few approximations and create some very useful circuits utilizing PP and SS feedback with op amps. These design and analysis shortcuts are presented in the next chapter, along with practical applications.

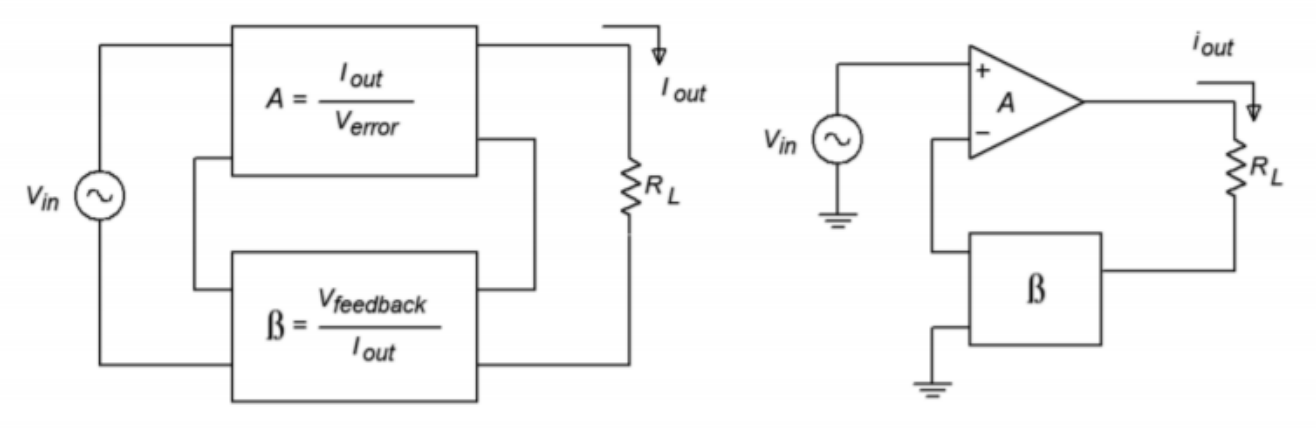


Figure 3.4.18 : Series-series connection.

9.5 LIMITATIONS ON THE USE OF NEGATIVE FEEDBACK

From the fore-going discussion, you may well think that negative feedback can do just about anything, short of curing a rainy day. Such is not the case. Yes, negative feedback can drastically lower distortion and increase bandwidth. Yes, it can have a very profound effect on input and output impedance. And yes, it certainly does stabilize our gains. What then, is the problem? Like all things, negative feedback has specific limitations. The first thing you should note is that ◈ is a function of frequency. This was graphically depicted back in Figure 3.3.2. The amount of change seen in impedances and distortion is a function of ◈ , it follows that these changes must be a function of frequency. Because ◈ drops as the frequency increases, the effects of negative feedback diminish as well. For example, if an SP amplifier has an open-loop ◈◈◈ of 200 k Ω and the low frequency ◈ is 500, the resulting ◈◈◈ with feedback is 100 M Ω . If we increase the input frequency past the open-loop $◈_2$, the open-loop gain drops and thus, ◈ drops. One decade up, ◈ will only be 50, so the ◈◈◈ with feedback will only be 10 M Ω . If this amplifier has a lower break frequency ($◈_1$), ◈ will drop as the frequency is reduced (below $◈_1$). The same sort of thing occurs with distortion, however, the harmonics each see a different ◈ , so the calculation is a bit more involved. Along with the reduction in gain, there is also a change in phase. If the phase around the feedback loop varies from -180 ° , incomplete cancellation takes place, and thus, the effects of feedback are lessened. The bottom line is that the effects of negative feedback weaken as we approach the frequency extremes.

The other item that must be kept in mind is the fact that negative feedback does not change specific fundamental characteristics of the amplifier. Negative feedback cannot get a circuit to do something beyond its operational parameters. For example, feedback has no effect on clipping level (saturation point). Further, feedback has no effect on slew rate (the maximum rate of output signal change, and an item that we will examine in a later chapter). Actually, when an amplifier slews, feedback is effectively blocked. An accurate output signal is no longer sent back to the input, but the amplifier can't correct for the errors any faster than it already is. In a similar manner, even though feedback can be used to lower the output impedance of a system, this does not imply that the system can produce more output current.

One possible "problem" with negative feedback is really the fault of the designer. It can be very tempting to sloppily design an amplifier with poor characteristics and then correct for them with large amounts of feedback. No matter how much feedback you use, the result will never be as good as a system that was designed carefully from the start. An example of this effect can be seen with TIMD (Transient Inter-Modulation Distortion). TIMD is a function of non-linearities in the first stages of an amplifier, and the excessive application of negative feedback will not remove it. On the other hand, if the initial stages of the system are properly designed, TIMD is not likely to be a problem.[1]

1. 1See E.M.Cherry and K.P.Dabke, "Transient Intermodulation Distortion- Part 2: Soft Nonlinearity", Journal of the Audio Engineering Society, Vol.34 No.1/2 (1986): 19–35

9.6 SUMMARY

We have seen that negative feedback can enhance the performance of amplifier circuits. This is done by sampling a portion of the output signal and summing it out of phase with the input signal. In order to maintain stability, the gain of the amplifier must be less than unity by the time its phase reaches -180°. There are four basic variants of negative feedback: Series-Parallel, Parallel-Series, Parallel-Parallel, and Series-Series. In all cases, gain and distortion are lowered by the sacrifice factor ◈ , and the bandwidth is increased by ◈ . Parallel connections reduce impedance by ◈ , whereas series connections increase the impedance by ◈ . At the input, parallel connections are current sensing, and series connections are voltage sensing. At the output, parallel connections produce a voltage-source model, and series connections produce a current-source model. The sacrifice factor is the ratio between the openand closed-loop gains. It is a function of frequency, and therefore, the effects of negative feedback lessen at the frequency extremes.

9.7 PROBLEMS

REVIEW QUESTIONS

1. Give two examples of how negative feedback is used in everyday life.
2. What circuit parameters will negative feedback alter, and to what extent?
3. What is meant by the term “Sacrifice Factor”?
4. What is the usage of Gain and Phase Margin?
5. Name the different negative feedback connections (i.e., variants or forms).
6. How might negative feedback accidentally turn into positive feedback?
7. What circuit parameters won’t negative feedback effect?
8. In practical amplifiers, when does negative feedback “stop working”, and why?

PROBLEMS

Analysis Problems

1. An amplifier’s open-loop gain plot is given in Figure 3.7.1 . If the amplifier is set up for a closed-loop gain of 100, what is the sacrifice factor (⛢) at low frequencies? What is ⛢ at 1 kHz?

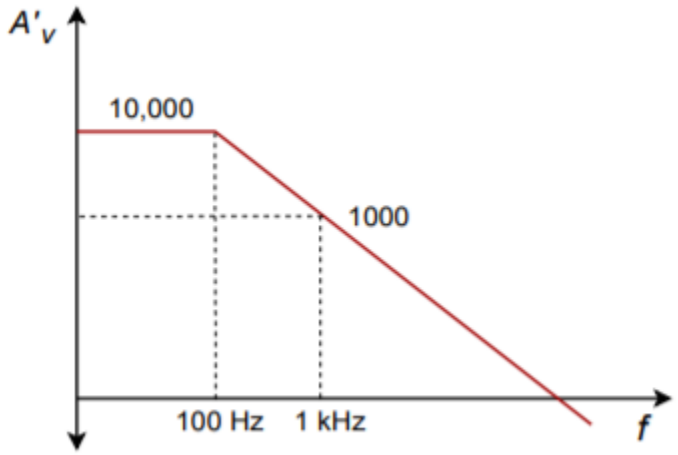


Figure 3.7.1

2. If the amplifier in Problem 1 has an open-loop THD of 5%, what is the closed-loop THD at low frequencies? Assuming that the open-loop THD doesn’t change with frequency, what is the closed-loop THD at 1 kHz?
3. If an amplifier has an open-loop response as given in Figure 3.7.1 , and a feedback factor (⛢)

of 0.05 is used, what is the exact low frequency closed loop gain? What is the approximate low frequency gain? What is the approximate gain at 1 kHz?

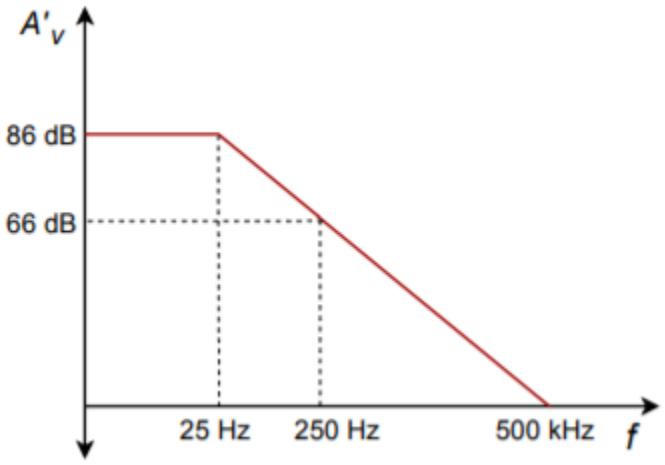


Figure 3.7.1

4. Using the open loop response curve in Figure 3.7.2 , determine exact and approximate values of ◈ for a closed-loop gain of 26 dB.
5. Determine the closed-loop ◈$_2$ for the circuit of Problem 4.
6. What is the maximum allowable phase shift at 500 kHz for stable usage of negative feedback in Figure 3.7.2 ?
7. Determine the gain and phase margins for the amplifier response given in Figure 3.7.3 . Is this amplifier a good candidate for negative feedback?

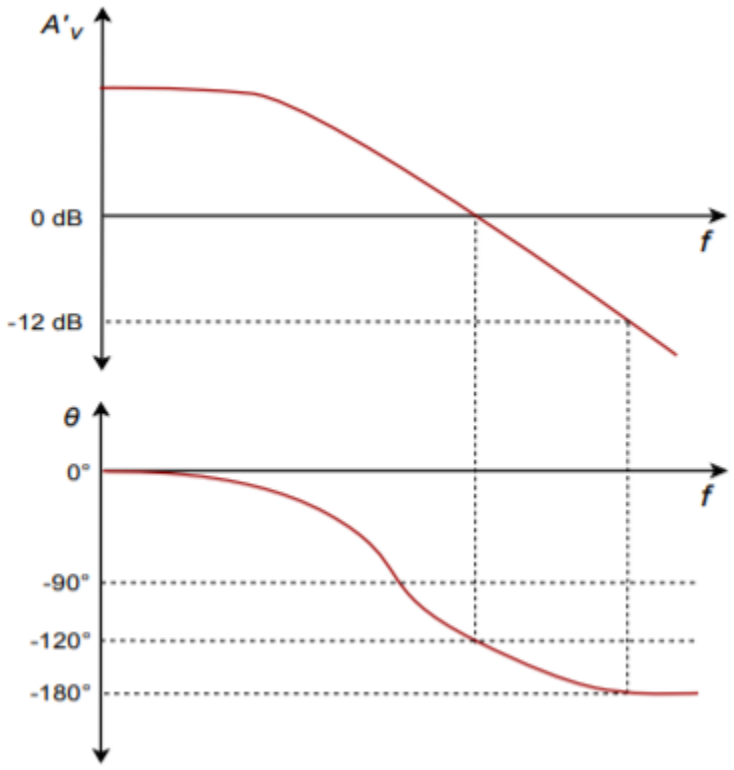


Figure 3.7.3

8. Determine the closed-loop (mid-band) gain in Figure 3.7.4 .
9. What is the closed-loop ??? in Figure 3.7.4 ? What is ???? ?
10. What is the low-frequency sacrifice factor in Figure 3.7.4?

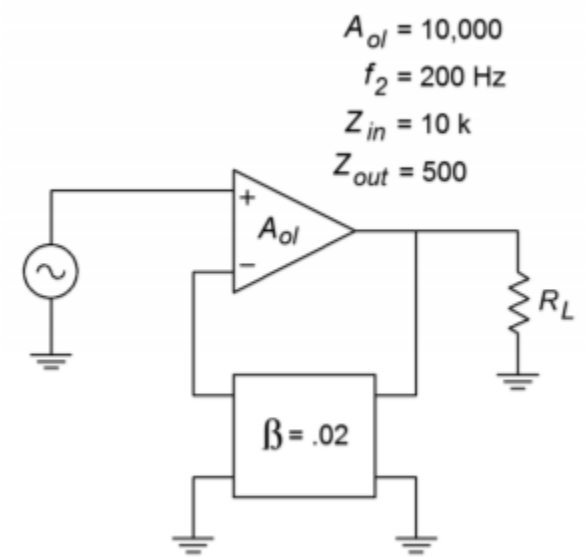


Figure 3.7.4

11. How much distortion reduction can we hope for in Figure 3.7.4 ?

12. How much of a signal/noise improvement can we expect in Figure 3.7.4?
13. Assuming ◇◇◇(◇)=0.1sin2◇500◇ , in Figure 3.7.4 , what is ◇◇◇◇(◇) ?
14. If the circuit in Figure 3.7.4 had an open-loop $◇_1$ of 10 Hz, what would the closed-loop $◇_1$ be?
15. Determine an appropriate pair of resistors to set ◇ to 0.1 in Figure 3.7.4 .

Challenge Problems

16. If the feedback network of Figure 3.7.4 produces a phase shift of -200 ° at 4 kHz, what effect will this have on circuit operation in Problem 10?
17. Consider the circuit of Figure 3.4.4. In general, what effect will the following alterations to the feedback network have on the closed loop system response? A) Placing a capacitor across ◇◇ B) Placing a capacitor across ◇◇ C) Placing a rectifying diode across ◇◇ (both polarities).

Computer Simulation Problems

18. Rerun the simulation of Figure 3.4.5 using the following open loop gains: 1 k, 10 k, and 100 k. What can you conclude from the results?
19. Verify the results of Problem 13 using a circuit simulator. It is possible to extend the basic op amp model with a lag network in order to mimic $◇_2$–◇◇ .
20. Verify the stability of the circuit used in Problem 13 with regard to the open loop gain. Run simulations with the given ◇◇◇ , and with values one decade above and below. Compare the resulting simulations and determine the maximum deviation of ◇◇◇ .

UNIT 10: BASIC OP AMP CIRCUITS

Learning Objectives

After completing this chapter, you should be able to:

- Relate each op amp circuit back to its general feedback form.
- Detail the general op amp circuit analysis idealizations.
- Solve inverting and noninverting voltage amplifier circuits for a variety of parameters,
- including gain and input impedance.
- Solve voltage/current transducer circuits for a variety of parameters.
- Solve current amplifier circuits for a variety of parameters.
- Define the term virtual ground.
- Analyze and design differential amplifiers.
- Analyze and design inverting and noninverting summing amplifiers.
- Discuss how output current capability may be increased.
- Outline the circuit modifications required for operation from a single polarity power supply.

10.1 INTRODUCTION

In this chapter we will be examining some common uses of op amps. Where Chapter Three focused on the theory of negative feedback in a more abstract way, this chapter zeroes in on the practical results of using negative feedback with op amps. Now that you know what negative feedback is and how it works, you can reap the benefits. The main theme is in the design and analysis of simple, small-signal, linear amplifiers. We will also be looking at some convenient approximation methods of analysis that can prove quite efficient. By the end of the chapter, you should be able to design single and multistage voltage amplifiers, voltage followers, and even amplifiers that sense current and/or produce constant current output. The voltage amplifiers offer the possibilities of inverting or not inverting the signal. Simple differential amplifiers are examined too. The chapter wraps up with a section on using op amps with a single polarity power supply and how to increase the available output current.

This is an introductory design section so the details of high frequency response, noise, offsets, and other important criteria are ignored. These items await a detailed analysis in Chapter Five. For the most part then, all calculations and circuit operations are assumed to be in the midband region.

10.2 INVERTING AND NONINVERTING AMPLIFIERS

As noted in our earlier work, negative feedback can be applied in one of four ways. The parallel input form inverts the input signal, and the series input form doesn't. Because these forms were presented as current-sensing and voltage-sensing respectively, you might get the initial impression that all voltage amplifiers must be noninverting. This is not the case. With the simple inclusion of one or two resistors, for example, we can make inverting voltage amplifiers or noninverting current amplifiers. Virtually all topologies are realizable. We will look at the controlledvoltage source forms first (those using SP and PP negative feedback).

For analysis, you can use the classic treatment given in Chapter Three; however, due to some rather nice characteristics of the typical op amp, approximations will be shown. These approximations are only valid in the midband and say nothing of the high frequency performance of the circuit. Therefore, they are not suitable for general-purpose discrete work. The idealizations for the approximations are:

- The input current is virtually zero (i.e., ◆◆◆ is infinite).
- The potential difference between the inverting and noninverting inputs is virtually zero (i.e., loop gain is infinite). This signal is also called the error signal.

Also, note for clarity that the power supply connections are not shown in most of the diagrams.

THE NONINVERTING VOLTAGE AMPLIFIER

The noninverting voltage amplifier is based on SP negative feedback. An example is given in Figure 4.2.1 . Note the similarity to the generic SP circuits of Chapter Three. Recalling the basic action of SP negative feedback, we expect a very high ◆◆◆ , a very low ◆◆◆◆ , and a reduction in voltage gain. Idealization 1 states that ◆◆◆ must be infinite. We already know that op amps have low ◆◆◆◆ , the second item is taken care of. Now let's take a look at voltage gain.

$$A_v = \frac{V_{out}}{V_{in}}$$

Because ideally ◆◆◆◆◆◆=0

$$V_{in} = V_{Ri}$$

Also,

$$V_{out} = V_{Ri} + V_{Rf}$$

$$A_v = \frac{V_{Ri} + V_{Rf}}{V_{Ri}}$$

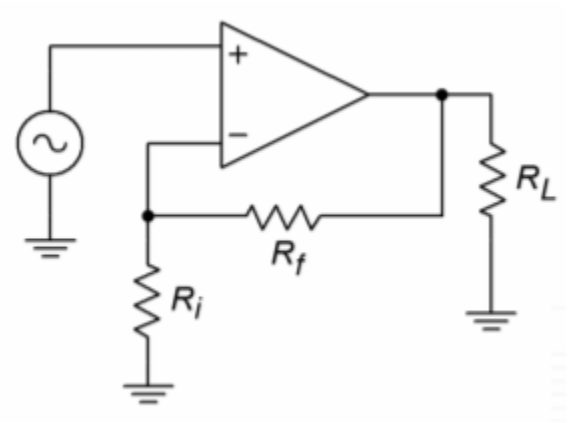


Figure 4.2.1 : Noninverting voltage amplifier.

Expansion gives

$$A_v = \frac{R_i I_{Ri} + R_f I_{Rf}}{RiI_{Ri}}$$

Because ⍰⍰⍰=0 , ⍰⍰⍰=⍰⍰⍰ , and finally we come to

$$A_v = \frac{R_i + R_f}{R_i} \text{or}$$

$$A_v = 1 + \frac{R_f}{R_i}$$

(4.2.1)

Now that's convenient. The gain of this amplifier is set by the ratio of two resistors. The larger ⍰⍰ is relative to ⍰⍰ , the more gain you get. Remember, this is an approximation. The closed loop gain can never exceed the open loop gain, and eventually, ⍰⍰ will fall off as frequency increases. Note that the calculation ignores the effect of the load impedance. Obviously, if ⍰⍰ is too small, the excessive current draw will cause the op amp to clip.

Example 4.2.1

What are the input impedance and gain of the circuit in Figure 4.2.2 ?

First off, ⍰⍰⍰ is ideally infinite. Now for the gain:

$$A_v = 1 + \frac{R_f}{R_i}$$

$$A_v = 1 + \frac{10k}{1k}$$

$$A_v = 11$$

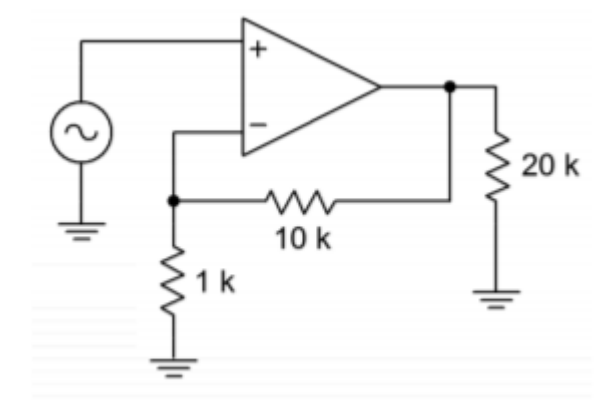


Figure 4.2.2 : Noninverting circuit for Example 4.2.1 .

The opposite process of amplifier design is just as straightforward.

Example 4.2.2

Design an amplifier with a gain of 26 dB and an input impedance of 47 k Ω . For the gain, first turn 26 dB into ordinary form. This is a voltage gain of about 20.

$$A_v = 1 + \frac{R_f}{R_i}$$

$$\frac{R_f}{R_i} = A_v - 1$$

$$\frac{R_f}{R_i} = 19$$

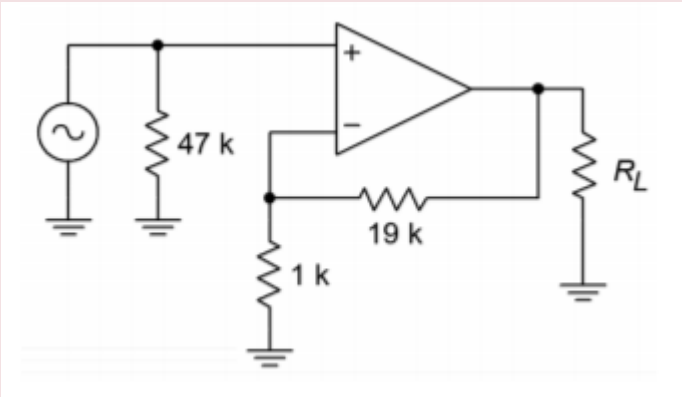


Figure 4.2.3 : Noninverting design for Example 4.2.2 .

At this point, choose a value for one of the resistors and solve for the other one. For example, the following would all be valid:

$R_i = 1k\Omega,\ R_f = 19k\Omega$

$R_i = 2k\Omega,\ R_f = 38k\Omega$

$R_i = 500\Omega,\ R_f = 9.5k\Omega$

Most of these are not standard values, though, and will need slight adjustments for a production circuit (see Appendix B). A reasonable range is 100◈Ω>◈◈+◈◈>10◈Ω . The accuracy of this gain will depend on the accuracy of the resistors. Now for the ◈◈◈ requirement. This is deceptively simple. ◈◈◈ is assumed to be infinite, so all you need to do is place a 47 k Ω in parallel with the input. The resulting circuit is shown in Figure 4.2.3 .

If a specific ◈◈◈ is not specified, a parallel input resistor is not required. There is one exception to this rule. If the driving source is not directly coupled to the op amp input (e.g., it is capacitively coupled), a resistor will be required to establish a DC return path to ground. Without a DC return path, the input section's diff amp stage will not be properly biased. This point is worth remembering, as it can save you a great deal in future headaches. For example, in the lab a circuit like the one in Figure 4.2.2 may work fine with one function generator, but not with another. This would be the case if the second generator used an output coupling capacitor and the first one didn't.

Example 4.2.3

Design a voltage follower (i.e., ideally infinite ◈◈◈ and a voltage gain of 1).

The ◈◈◈ part is straightforward enough. As for the second part, what ratio of ◈◈ to ◈◈ will yield a gain of 1?

$$A_v = 1 + \frac{R_f}{R_i}$$

$$\frac{R_f}{R_i} = A_v - 1$$

$$\frac{R_f}{R_i} = 0$$

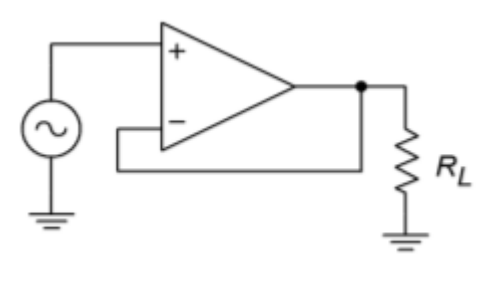


Figure 4.2.4 : Voltage follower for Example 4.2.3 .

This says that ◈◈ must be 0 Ω . Practically speaking, that means that ◈◈ is replaced with a shorting wire. What about ◈◈ ? Theoretically, almost any value will do. As long as there's a choice, consider infinite. Zero divided by infinite is certainly zero. The practical benefit of choosing ◈◈=∞ is that you may delete ◈◈ . The resulting circuit is shown in Figure 4.2.4 . Remember, if the source is not directly coupled, a DC return resistor will be needed. The value of this resistor has to be large enough to avoid loading the source.

As you can see, designing with op amps can be much quicker than its discrete counterpart. As a result, your efficiency as a designer or repair technician can improve greatly. You are now free to concentrate on the system, rather than on the specifics of an individual biasing resistor. In order to make multi-stage amplifiers, just link individual stages together.

Example 4.2.4

What is the input impedance of the circuit in Figure 4.2.5 ? What is ◈◈◈? As in any multi-stage amplifier, the input impedance to the first stage is the system ◈◈◈ . The DC return resistor sets this at 100 k Ω .

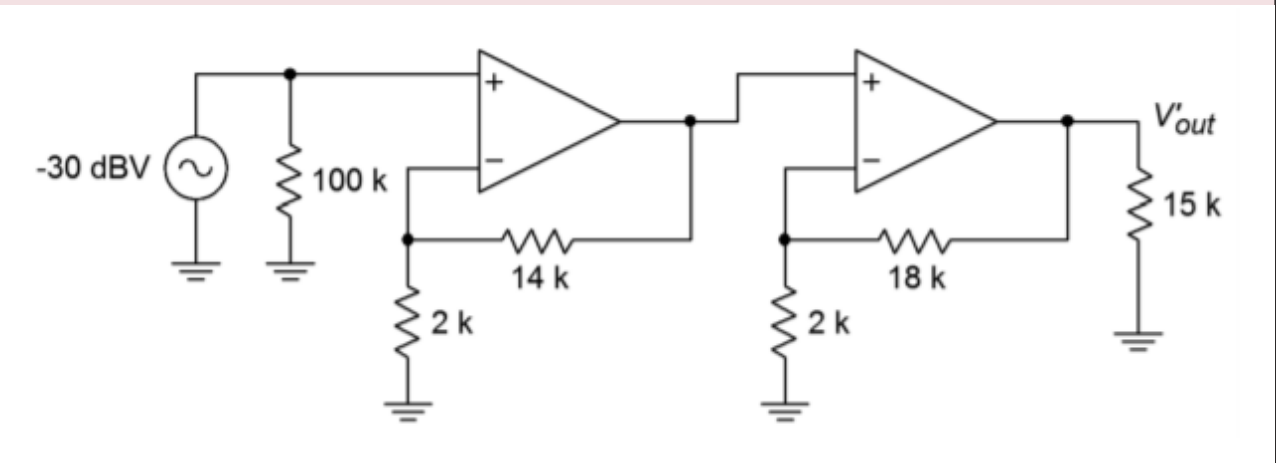


Figure 4.2.5 : Multistage circuit for Example 4.2.4 .

To find ◈◈◈, we need to find the gain (in dB).

$$A_{v1} = 1 + \frac{R_f}{R_i}$$

$$A_{v1} = 1 + \frac{14k}{2k}$$

$$A_{v1} = 8$$

$$A_{v1}^{'} = 18dB$$

$$A_{v2} = 1 + \frac{R_f}{R_i}$$

$$A_{v2} = 1 + \frac{18k}{2k}$$

$$A_{v2} = 10$$

$$A_{v2}^{'} = 20dB$$

$$A_{vt}^{'} = 18dB + 20dB = 38dB$$

$$V_{out}^{'} = A_{vt}^{'} + V_{in}^{'}$$

$$V_{out}^{'} = 38dB + (-30dBV)$$

$$V_{out}^{'} = 8dBV$$

Because 8 dBV translates to about 2.5 V, there is no danger of clipping either.

INVERTING VOLTAGE AMPLIFIER

The inverting amplifier is based on the PP negative feedback model. The base form is shown in Figure 4.2.6 . By itself, this form is current sensing, not voltage sensing. In order to achieve voltage sensing, an input resistor, ◈◈ , is added. See Figure 4.2.7 . Here's how the circuit works: ◈◈◈◈◈◈ is virtually zero, so the inverting input potential must equal the noninverting input potential. This means that the inverting input is at a virtual ground. The signal here is so small that it is negligible. Because of this, we may also say that the impedance seen looking into this point is zero. This last point may cause a bit of confusion. You may ask, "How can the impedance be zero if the current into the op amp is zero?" The answer lies in the fact that all of the entering current will be drawn through ◈◈ , thus bypassing the inverting input.

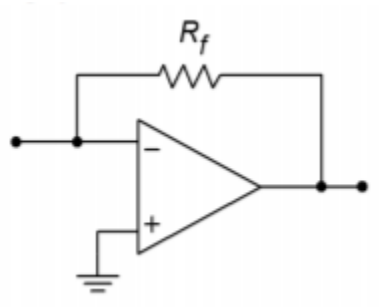


Figure 4.2.6 : A basic parallel-parallel amplifier.

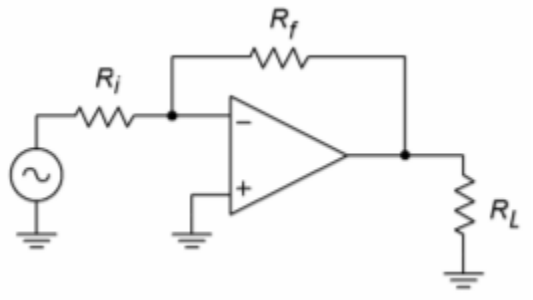


Figure 4.2.7 : Inverting voltage amplifier.

Refer to Figure 4.2.8 for the detailed explanation. The right end of ⓿⓿ is at virtual ground, so all of the input voltage drops across it, creating ⓿⓿⓿ , the input current. This current cannot enter the op amp and instead will pass through ⓿⓿ . Because a positive signal is presented to the inverting input, the op amp will sink output current, thus drawing ⓿⓿⓿ through ⓿⓿ . The resulting voltage drop across ⓿⓿ is the same magnitude as the load voltage. This is true because ⓿⓿ is effectively in parallel with the load. Note that both elements are tied to the op amp's output and to (virtual) ground. There is a change in polarity because we reference the output signal to ground. In short, ⓿⓿⓿⓿is the voltage across ⓿⓿ , inverted.

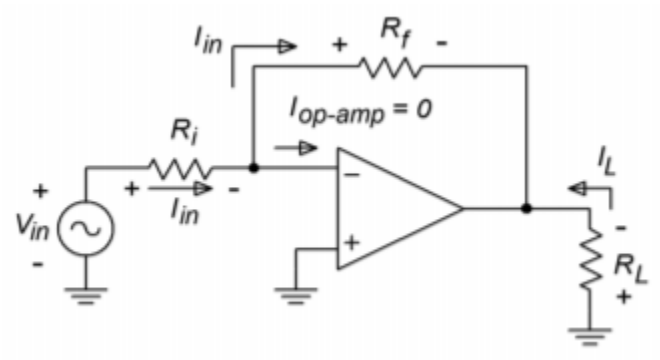


Figure 4.2.8 : Analysis of the inverting amplifier from Figure 4.2.7

$$A_v = \frac{V_{out}}{V_{in}}$$

$$V_{in} = I_{in} R_i$$
$$V_{out} = -V_{R_f}$$

$$V_{Rf} = I_{in} R_f$$

Substitution yields

$$A_v = -\frac{I_{in} R_f}{I_{in} R_i}$$

$$A_v = -\frac{R_f}{R_i}$$

(4.2.2)

Again, we see that the voltage gain is set by resistor ratio. Again, there is an allowable range of values.

The foregoing discussion points up the derivation of input impedance. Because all of the input signal drops across ◈◈ , it follows that all the driving source "sees" is ◈◈ . Quite simply, ◈◈ sets the input impedance. Unlike the noninverting voltage amp, there is a definite interrelation between ◈◈◈(◈◈) and ◈◈(–◈◈/◈◈) . This indicates that it is very hard to achieve both high gain and high ◈◈◈ with this circuit.

Example 4.2.5

Determine the input impedance and output voltage for the circuit in Figure 4.2.9 .

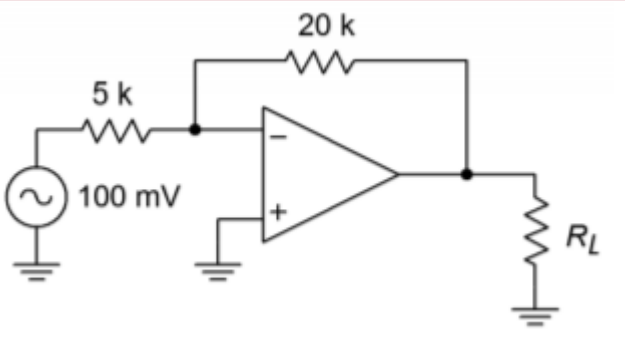


Figure 4.2.9 : Inverting amplifier for Example 4.2.9 .

The input impedance is set by ◈◈ . ◈◈=5◈Ω , therefore ◈◈◈=5◈Ω .

$$V_{out} = V_{in} A_v$$

$$A_v = -\frac{R_f}{R_i}$$

$$A_v = -\frac{20k}{5k}$$

$$A_v = -4$$

$$V_{out} = 100mV \times (-4)$$

$$V_{out} = -400mV, \text{ (i.e., inverted)}$$

Example 4.2.6

Design an inverting amplifier with a gain of 10 and an input impedance of 15 k Ω . The input impedance tells us what ⬖⬖ must be

$$Z_{in} = R_i$$

$$R_i = 15k$$

Knowing ⬖⬖ , solve for ⬖⬖ :

$$A_v = -\frac{R_f}{R_i}$$

$$R_f = R_i(-A_v)$$

$$R_f = 15k \times (-(-10))$$

$$R_f = 150k$$

COMPUTER SIMULATION

A Multisim simulation of the result of Example 4.2.6 is shown in Figure 4.2.10 , along with its schematic. This simulation uses the simple dependent source model presented in Chapter Two. The input is set at 0.1V DC for simplicity. Note that the output potential is negative, indicating the inverting action of the amplifier. Also, note that the virtual ground approximation is borne out quite well, with the inverting input potential measuring in the ⬖ V region.

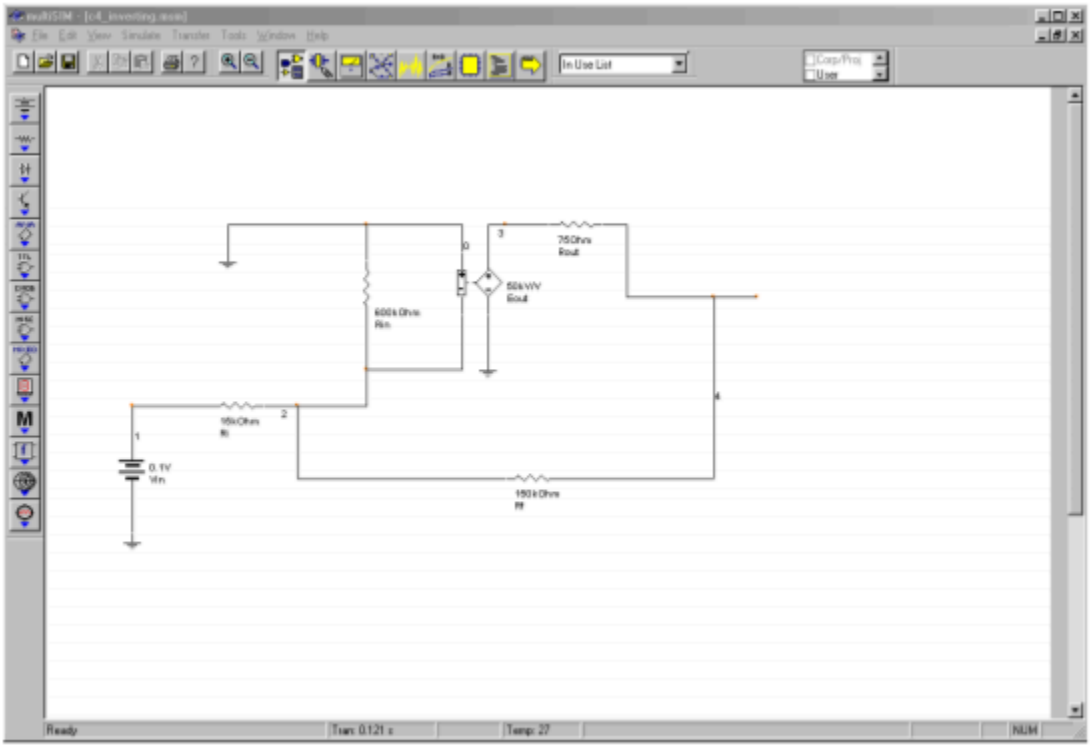


Figure 4.2.10 : Multisim simulation of the simple op amp model for Example 4.2.6 . a. Schematic.

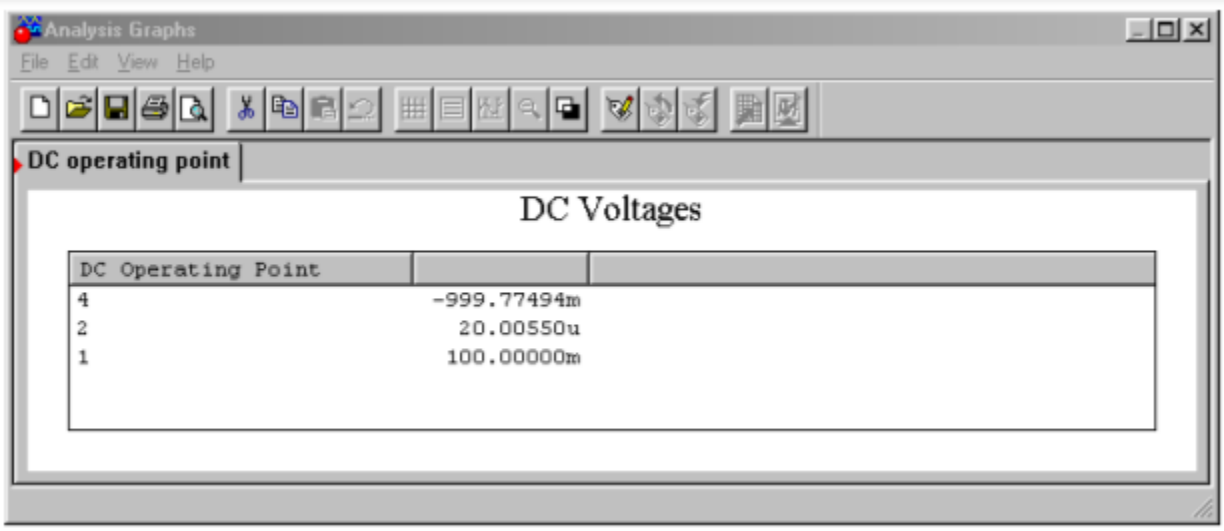


Figure 4.2.10 : Multisim simulation of the simple op amp model for Example 4.2.6 . b. Output listing.

Example 4.2.7

The circuit of Figure 4.2.11 is a pre-amplifier stage for an electronic music keyboard. Like most musicians' pre-amplifiers, this one offers adjustable gain. This is achieved by following the amplifier with a pot. What are the maximum and minimum gain values?

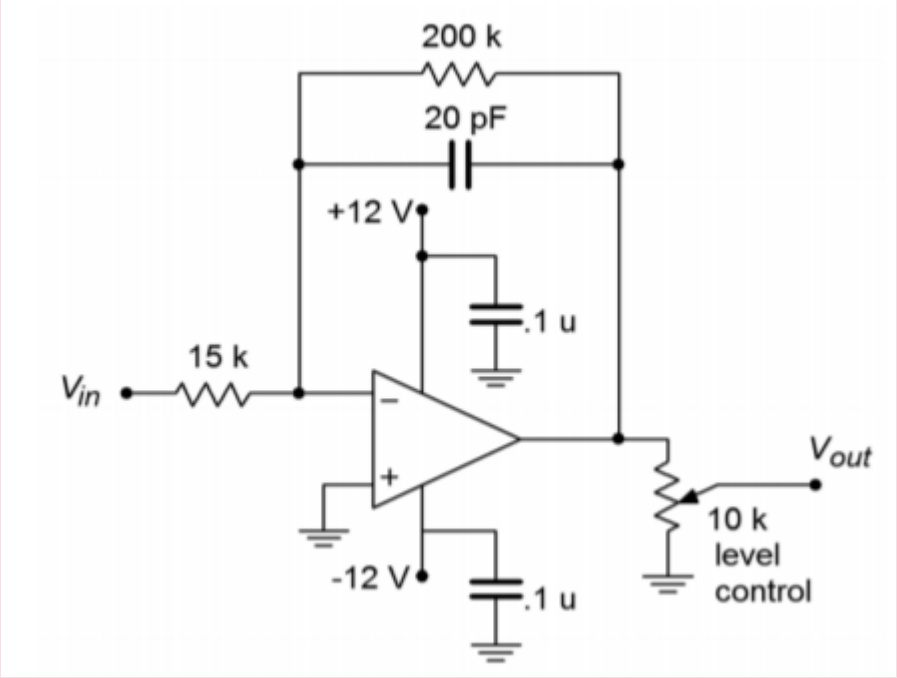


Figure 4.2.11 : Musical instrument preamplifier for Example 4.2.7 .

Note that the gain for the pre-amp is the product of the op amp gain and the voltage divider ratio produced by the pot. For maximum gain, use the pot in its uppermost position. Because the pot acts as a voltage divider, the uppermost position provides no divider action (i.e., its gain is unity). For midband frequencies, the 20 pF may be ignored.

$$A_{v-max} = -\frac{R_f}{R_i}$$

$$A_{v-max} = -\frac{200k}{15k}$$

$$A_{v-max} = -13.33$$

$$A_{v-max}^{'} = 22.5dB$$

For minimum gain, the pot is dialed to ground. At this point, the divider action is infinite, and thus the minimum gain is 0 (resulting in silence).

◈◈◈ for the system is about 15 k Ω . As far as the extra components are concerned, the 20 pF capacitor is used to decrease high frequency gain. The two 0.1 ◈ F bypass capacitors across the power supply lines are very important. Virtually all op amp circuits use bypass capacitors. Due to the high gain nature of op amps, it is essential to have good AC grounds at the power supply pins. At higher frequencies the inductance of power supply wiring may produce a sizable impedance. This impedance may create a positive feedback loop that wouldn't exist otherwise. Without the bypass capacitors, the circuit may oscillate or produce spurious output signals. The precise values for the capacitors are usually not critical, with 0.1 to 1 ◈ F being typical.

INVERTING CURRENT-TO-VOLTAGE TRANSDUCER

As previously mentioned, the inverting voltage amplifier is based on PP negative feedback, with an extra input resistor used to turn the input voltage into a current. What happens if that extra resistor is left out, and a circuit such as Figure 4.2.6 is used? Without the extra resistor, the input is at virtual ground, thus setting ◈◈◈ to 0 Ω . This is ideal for sensing current. This input current will pass through ◈◈ and produce an output voltage as outlined above. The characteristic of transforming a current to a voltage is measured by the parameter transresistance. By definition, the transresistance of this circuit is the value of ◈◈ . To find ◈◈◈◈, multiply the input current by the transresistance. This circuit inverts polarity as well.

$$V_{out} = I_{in} R_f$$

(4.2.3)

Example 4.2.8

Design a circuit based on Figure 4.2.6 if an input current of -50 ◈ A should produce an output of 4 V.

The transresistance of the circuit is ◈◈

$$R_f = -\frac{V_{out}}{I_{in}}$$

$$R_f = -\frac{4V}{50\mu A}$$

$$R_f = 80k$$

The input impedance is assumed to be zero.

At first glance, the circuit applications of the topology presented in the prior example seem very limited. In reality, there are a number of linear integrated circuits that produce their output in current form.[1] In many cases, this signal must be turned into a voltage in order to properly interface with other circuit elements. The current-to-voltage transducer is widely used for this purpose.

NONINVERTING VOLTAGE-TO-CURRENT TRANSDUCER

This circuit topology utilizes SS negative feedback. It senses an input voltage and produces a current. A conceptual comparison can be made to the FET (a voltage controlled current source). Instead of circuit gain, we are interested in transconductance. In other words, how much input voltage is required to produce a given output current? The op amp circuit presented here drives a floating load. That is, the load is not referenced to ground. This can be convenient in some cases, and a real pain in others. With some added circuitry, it is possible to produce a grounded load version, although space precludes us from examining it here.

1. Most notably, operational transconductance amplifiers and digital to analog converters, which we will examine in Chapters Six and Twelve, respectively.

A typical voltage-to-current circuit is shown in Figure 4.2.12 . Because this uses series-input type feedback, we may immediately assume that ⍰⍰⍰ is infinite. The voltage to current ratio is set by feedback resistor ⍰⍰ . As ⍰⍰⍰⍰⍰⍰ is assumed to be zero, all of ⍰⍰⍰drops across ⍰⍰ , creating current ⍰⍰⍰ . The op amp is assumed to have zero input current, so all of ⍰⍰⍰ flows through the load resistor, ⍰⍰ . By adjusting ⍰⍰ , the load current may be varied.

$$I_{load} = I_{R_i}$$

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

$$I_{load} = \frac{V_{in}}{R_i}$$

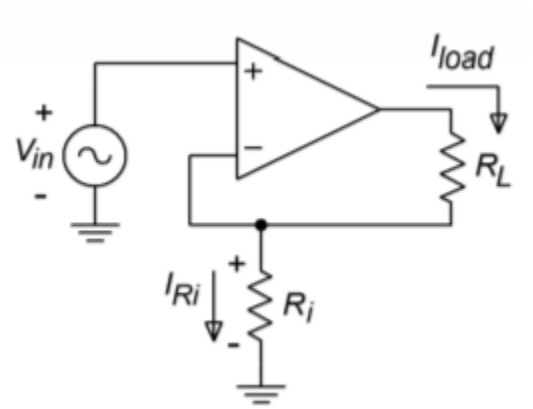


Figure 4.2.12 : Voltage-to-current transducer.

By definition,

$$g_m = \frac{I_{load}}{V_{in}}$$

$$g_m = \frac{1}{R_i}$$

(4.2.4)

So, the transconductance of the circuit is set by the feedback resistor. As usual, there are practical limits to the size of ⍰⍰ . If ⍰⍰ and ⍰⍰ are too small, the possibility exists that the op amp will "run out" of output current and go into saturation. At the other extreme, the product of the two resistors and the ⍰⍰⍰⍰⍰ cannot exceed the power supply rails. As an example, if ⍰⍰ plus ⍰⍰ is 10 k Ω , ⍰⍰⍰⍰⍰ cannot exceed about 1.5 mA if standard ± 15 V supplies are used.

Example 4.2.9

Given an input voltage of 0.4 V in the circuit of Figure 4.2.13 , what is the load current?

$$g_m = \frac{1}{R_i}$$

$$g_m = \frac{1}{20k}$$

$$g_m = 50\mu S$$

$$I_{load} = g_m V_{in}$$

$$I_{load} = 50\mu S \times 0.4V$$

$$I_{load} = 20\mu A$$

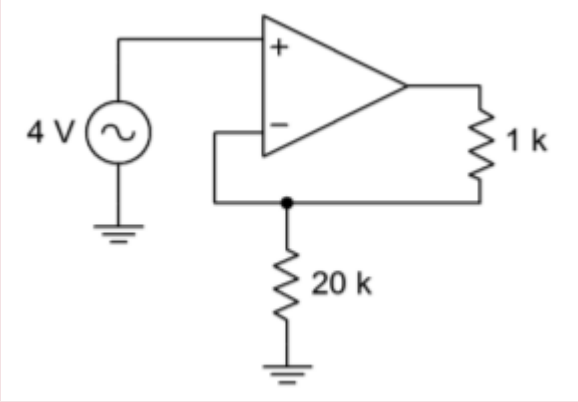


Figure 4.2.13 : Voltage-to-current transducer for Example 4.2.9 .

There is no danger of current overload here as the average op amp can produce about 20 mA, maximum. The output current will be 20 ◈ A regardless of the value of ◈◈ , up to clipping. There is no danger of clipping in this situation either. The voltage seen at the output of the op amp to ground is

$$V_{max} = (R_i + R_l)I_{load}$$

$$V_{max} = (20k + 1k) \times 20\mu A$$

$$V_{max} = 420mV$$

That's well below clipping level.

COMPUTER SIMULATION

A simulation of the circuit of Example 4.2.9 is shown in Figure 4.2.14 . Multisim's ideal op amp model has been chosen to simplify the layout. The load current is exactly as calculated at 20 ◈ A. An interesting trick is used here to plot the load current, as many simulators only offer plotting of node voltages. Using Multisim's Post Processor, the load current is computed by taking the difference between the node voltages on either side of the load resistor and then dividing the result by the load resistance.

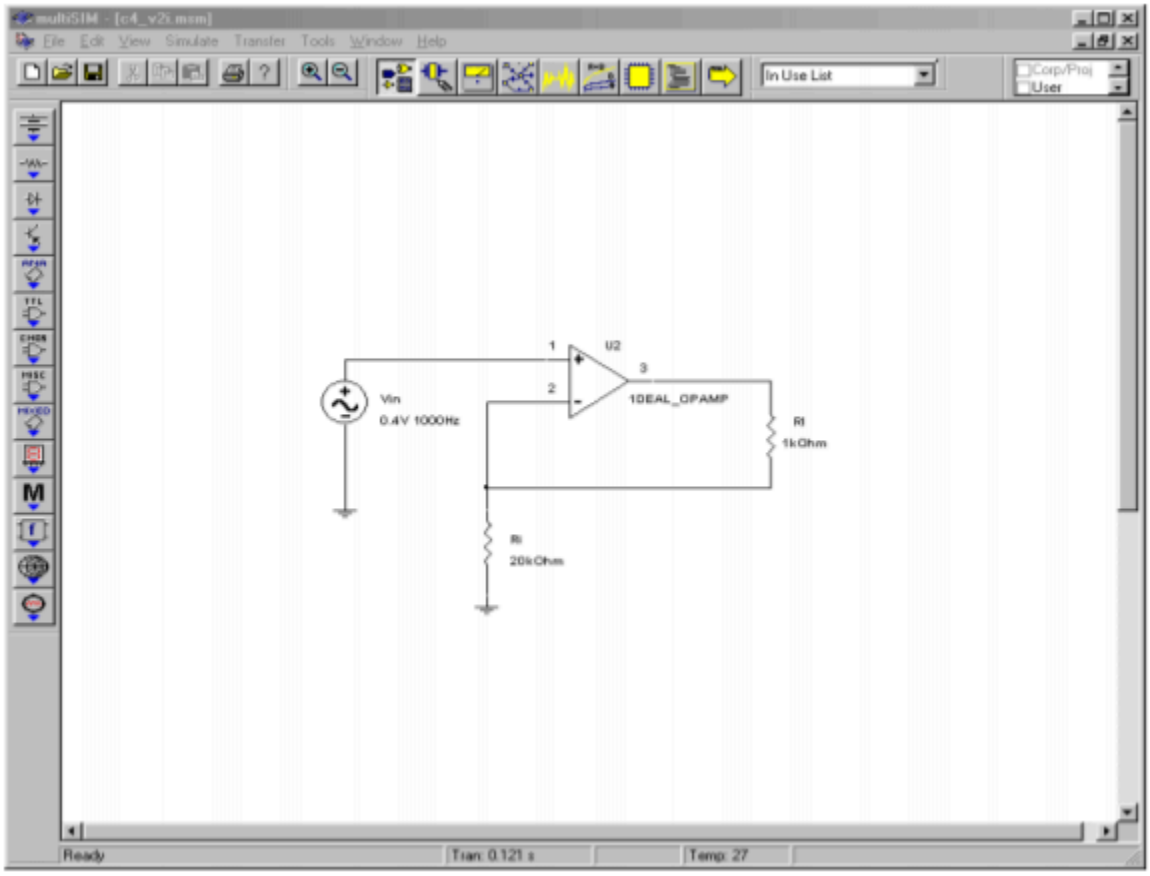


Figure 4.2.14 : Voltage-to-current transducer simulation schematic.

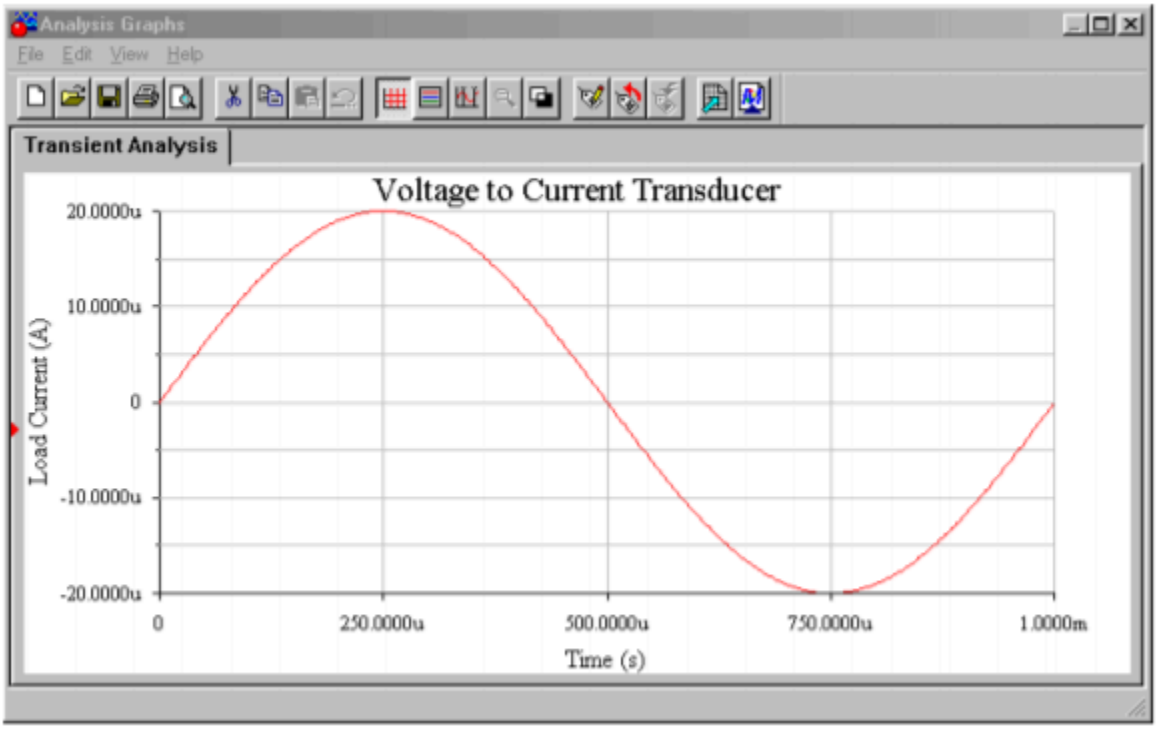


Figure 4.2.14 : Simulation results.

Example 4.2.10

The circuit of Figure 4.2.15 can be used to make a high-input-impedance DC voltmeter. The load in this case is a simple meter movement. This particular meter requires 100 ◈ A for full-scale deflection. If we want to measure voltages up to 10 V, what must ◈◈ be?

First, we must find the transconductance.

$$g_m = \frac{I_{load}}{V_{in}}$$

$$g_m = \frac{100\mu A}{10V}$$

$$g_m = 10\mu S$$

$$R_i = \frac{1}{g_m}$$

$$R_i = \frac{1}{10\mu S}$$

$$R_i = 100k$$

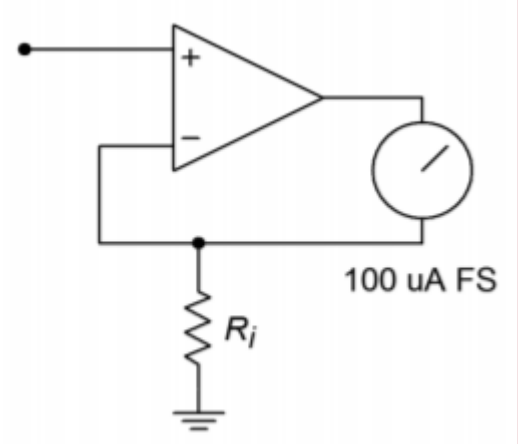


Figure 4.2.15 : DC voltmeter for Example 4.2.10 .

The meter deflection is assumed to be linear. For example, if the input signal is only 5 V, the current produced is halved to 50 ◈ A. 50 ◈ A should produce half-scale deflection. The accuracy of this electronic voltmeter depends on the accuracy of ◈◈ , and the linearity of the meter movement. Note that this little circuit can be quite convenient in a lab, being powered by batteries. In order to change scales, new values of ◈◈ can be swapped in with a rotary switch. For a 1 V scale, ◈◈ equals 10 k Ω . Note that for higher input ranges, some form of input attenuator is needed. This is due to the fact that most op amps may be damaged if input signals larger than the supply rails are used.

INVERTING CURRENT AMPLIFIER

The inverting current amplifier uses PS negative feedback. As in the voltage-to-current transducer, the load is floating. The basic circuit is shown in Figure 4.2.16 . Due to the parallel negative feedback connection at the input, the circuit input impedance is assumed to be zero. This means that the input point is at virtual ground. The current into the op amp is negligible, so all input current flows through ◈◈ to node A. Effectively, ◈◈ and ◈◈ are in parallel (they both share node A and ground; actually virtual ground for ◈◈). Therefore, ◈◈◈ and ◈◈◈ are the same value. This means that a current is flowing through ◈◈ , from ground to node A. These two currents join to form the load current. In this manner, current gain is achieved. The larger ◈◈◈ is relative to ◈◈◈ , the more current gain there is. As the op amp is sinking current, this is an inverting amplifier.

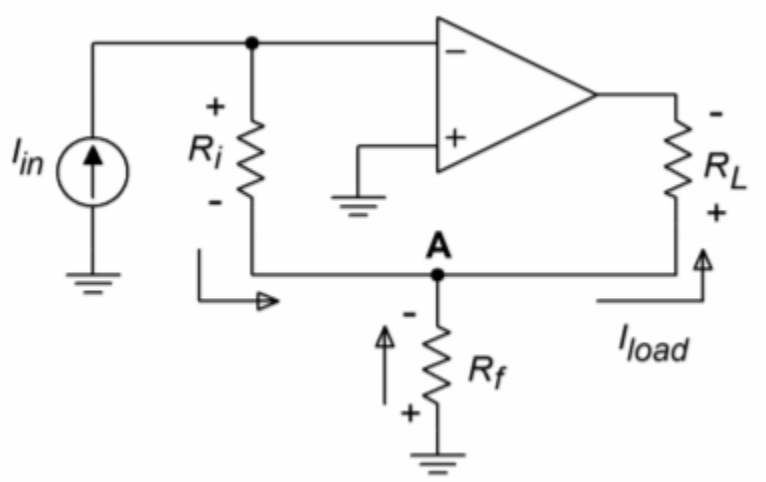


Figure 4.2.16 : Inverting current amplifier

$$A_i = -\frac{I_{out}}{I_{in}}$$

$$I_{out} = I_{Rf} + I_{Ri}$$

(4.2.5)

$$I_{Ri} = I_{in}$$

$$I_{Rf} = \frac{V_{Rf}}{R_f}$$

Because ◈◈◈ is the same value as ◈◈◈ ,

$$I_{Rf} = \frac{V_{Ri}}{R_f}$$

(4.2.6)

$$V_{Ri} = I_{in}R_i$$

(4.2.7)

Substitution of 4.2.7 into 4.2.6 yields

$$I_{Rf} = \frac{I_{in}R_i}{R_f}$$

Substituting into 4.2.5 produces

$$I_{out} = I_{in} + \frac{I_{in} R_i}{R_f}$$

$$I_{out} = I_{in} \left(1 + \frac{R_i}{R_f}\right)$$

$$A_i = -\left(1 + \frac{R_i}{R_f}\right)$$

(4.2.8)

As you might expect, the gain is a function of the two feedback resistors. Note the similarity of this result with that of the noninverting voltage amplifier.

Example 4.2.11

What is the load current in Figure 4.2.17 ?

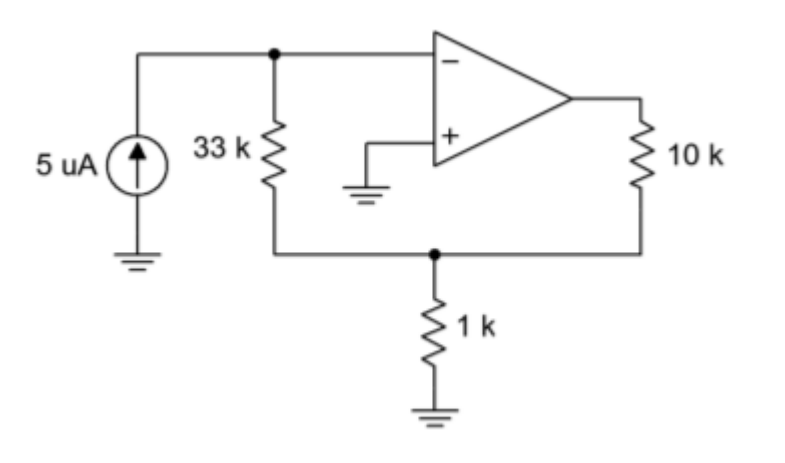


Figure 4.2.17 : Current amplifier for Example 4.2.11 .

$$I_{out} = -A_i I_{in}$$

$$A_i = -(1 + \frac{R_i}{R_f})$$

$$A_i = -(1 + \frac{33k}{1k})$$

$$A_i = -34$$

$$I_{out} = -34 \times 5\mu A$$

$$I_{out} = -170\mu A \text{ (sinking)}$$

We need to check to make sure that this current doesn't cause output clipping. A simple Ohm's Law check is all that's needed.

$$V_{max} = I_{out}R_{load} + I_{in}R_i$$
$$V_{max} = 170\mu A \times 10k + 5\mu A \times 33k$$
$$V_{max} = 1.7V + .165V$$
$$V_{max} = 1.865V \text{ (no problem)}$$

Example 4.2.12

Design an amplifier with a current gain of -50. The load is approximately 200 k Ω . Assuming a typical op amp (◈◈◈◈–◈◈◈ = 20 mA with ± 15 V supplies), what is the maximum load current obtainable?

$$A_i = -\left(1 + \frac{R_i}{R_f}\right)$$

$$\frac{R_i}{R_f} = -Ai - 1$$

$$\frac{R_i}{R_f} = 50 - 1$$

$$\frac{R_i}{R_f} = 49$$

Therefore, ◈◈ must be 49 times larger than ◈◈ . Possible solutions include:

$$R_i = 49k\Omega, R_f = 1k\Omega$$
$$R_i = 98k\Omega, R_f = 2k\Omega$$
$$R_i = 24.5k\Omega, R_f = 500\Omega$$

As far as maximum load current is concerned, it can be no greater than the op amp's maximum output of 20 mA, but it may be less. We need to determine the current at clipping. Due to the large size of the load resistance, virtually all of the output potential will drop across it. Ignoring the extra drop across the feedback resistors will introduce a maximum of 1% error (that's worst case, assuming resistor set number two).

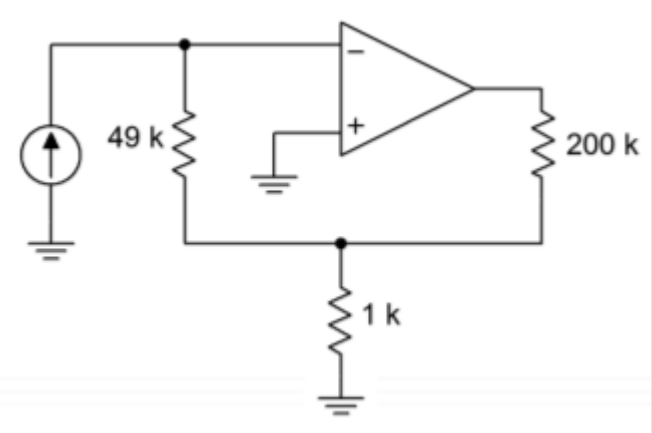


Figure 4.2.18 : Current amplifier design for Example 4.2.12 .

With 15 V rails, a typical op amp will clip at 13.5 V. The resulting current is found through Ohm's Law:

$$I_{max} = \frac{13.5V}{200k}$$

$$I_{max} = 67.5\mu A$$

Another way of looking at this is to say that the maximum allowable input current is 67.5 ◈ A/50, or 1.35 ◈ A. One possible solution is shown in Figure 4.2.18 .

SUMMING AMPLIFIERS

It is very common in circuit design to combine several signals into a single common signal. One good example of this is in the broadcast and recording industries. The typical modern music recording will require the use of perhaps dozens of microphones, yet the final product consists of typically two output signals (stereo left and right). If signals are joined haphazardly, excessive interference, noise, and distortion may result. The ideal summing amplifier would present each input signal with an isolated load not affected by other channels.

The most common form of summing amplifier is really nothing more than an extension of the inverting voltage amplifier. Because the input to the op amp is at virtual ground, it makes an ideal current summing node. Instead of placing a single input resistor at this point, several input resistors may be used. Each input source drives its own resistor, and there is very little affect from neighboring inputs. The virtual ground is the key. A general summing amplifier is shown in Figure 4.2.19 .

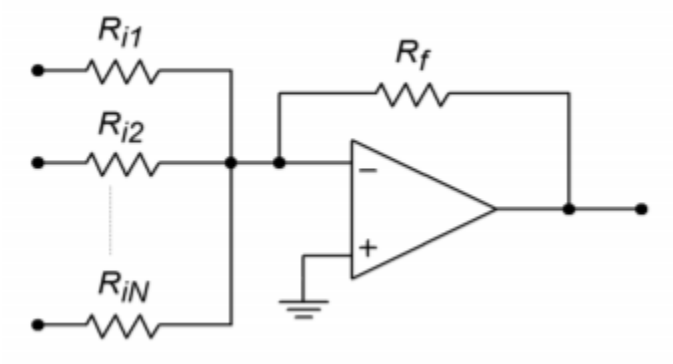


Figure 4.2.19 : Summing amplifier.

The input impedance for the first channel is ◈◈1 , and it's voltage gain is –◈◈/◈◈1 . For channel 2, the input impedance is ◈◈2 , with a gain of –◈◈/◈◈2 . In general, then for channel N we have

$$Z_{inN} = R_{iN}$$

$$A_{vN} = -\frac{R_f}{R_{iN}}$$

The output signal is the sum of all inputs multiplied by their associated gains.

$$V_{out} = V_{in1}A_{v1} + V_{in2}A_{v2} + \cdots + V_{inN}A_{vN}$$

which is written more conveniently as

$$V_{out} = \sum_{i=1}^{n} V_{in_i}A_{v_i}$$

(4.2.9)

A summing amplifier may have equal gain for each input channel. This is referred to as an equal-weighted configuration.

Example 4.2.13

What is the output of the summing amplifier in Figure 4.2.20 , with the given DC input voltages?

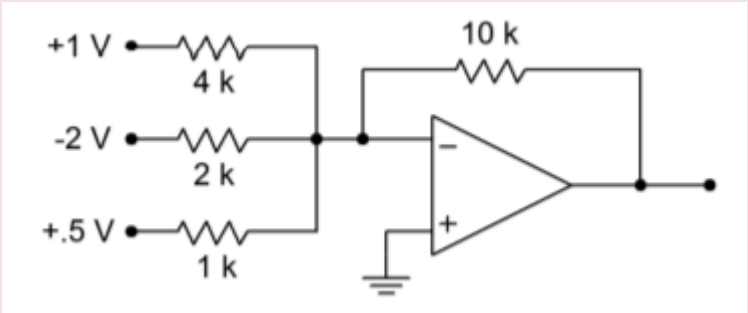


Figure 4.2.20 : Summing amplifier for Example 4.2.13 .

The easy way to approach this is to just treat the circuit as three inverting voltage amplifiers, and then add the results to get the final output.

Channel 1:

$$A_v = -\frac{R_f}{R_i}$$

$$A_v = -\frac{10k}{4k}$$

$$A_v = -2.5$$

$$V_{out} = -2.5 \times 1V$$

$$V_{out} = -2.5V$$

Channel 2:

$$A_v = -\frac{R_f}{R_i}$$

$$A_v = -\frac{10k}{2k}$$

$$A_v = -5$$

$$V_{out} = -5 \times -2V$$

$$V_{out} = 10V$$

Channel 3:

$$A_v = -\frac{R_f}{R_i}$$

$$A_v = -\frac{10k}{1k}$$

$$A_v = -10$$

$$V_{out} = -10 \times .5V$$

$$V_{out} = -5V$$

The final output is found via summation:

$$V_{out} = -2.5V + 10V + (-5V)$$

$$V_{out} = 2.5V$$

If the inputs were AC signals, the summation is not quite so straightforward. Remember, AC signals of differing frequency and phase do not add coherently. You can perform a calculation similar to the preceding to find the peak value, however, an RMS calculation is needed for the effective value (i.e., square root of the sum of the squares).

For use in the broadcast and recording industries, summing amplifiers will also require some form of volume control for each input channel and a master volume control as well. This allows the levels of

various microphones or instruments to be properly balanced. Theoretically, individual-channel gain control may be produced by replacing each input resistor with a potentiometer. By adjusting ⍰⍰ , the gain may be directly varied. In practice, there are a few problems with this arrangement. First of all, it is impossible to turn a channel completely off. The required value for ⍰⍰ would be infinite. Second, because ⍰⍰ sets the input impedance, a variation in gain will produce a ⍰⍰⍰ change. This change may overload or alter the characteristics of the driving source. One possible solution is to keep ⍰⍰ at a fixed value and place a potentiometer before it, as in Figure 4.2.21 . The pot produces a gain from 1 through 0. The ⍰⍰/⍰⍰ combo is then set for maximum gain. As long as ⍰⍰ is several times larger than the pot's value, the channel's input impedance will stay relatively constant. The effective ⍰⍰⍰ for the channel is ⍰⍰⍰⍰ in parallel with ⍰⍰ , at a minimum, up to ⍰⍰⍰⍰ .

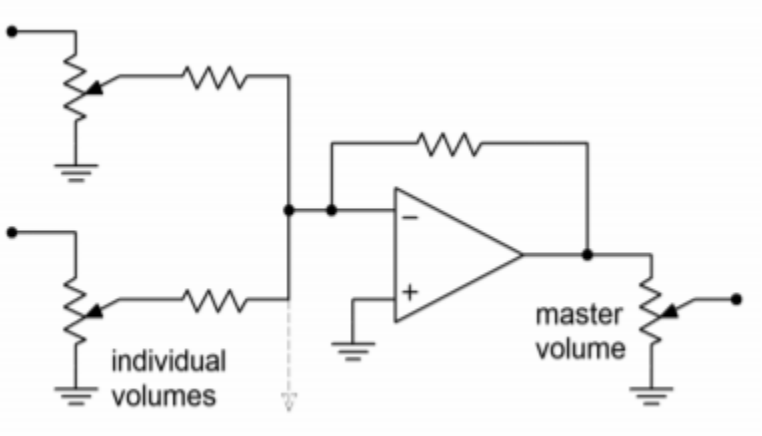


Figure 4.2.21 : Audio mixer.

As far as a master volume control is concerned, it is possible to use a pot for ⍰⍰ . Without a limiting resistor though, a very low master gain runs the risk of over driving the op amp due to the small effective ⍰⍰ value. This technique also causes variations in offset potentials and circuit bandwidth. A technique that achieves higher performance involves using a stage with a fixed ⍰⍰ value, followed with a pot, as in Figure 4.2.21 .

Still another application of the summing amplifier is the level shifter. A level shifter is a two-input summing amplifier. One input is the desired AC signal, and the second input is a DC value. The proper selection of DC value lets you place the AC signal at a desired DC offset. There are many uses for such a circuit. One possible application is the DC offset control available on many signal generators.

NONINVERTING SUMMING AMPLIFIER

Besides the inverting form, summing amplifiers may also be produced in a noninverting form. Noninverting summers generally exhibit superior high frequency performance when compared to the inverting type. One possible circuit is shown in Figure 4.2.22 . In this example, three inputs are shown, although more could be added. Each input has an associated input resistor. Note that it is not possible to simply wire several sources together in hopes of summing their respective signals. This is because each source will try to bring its output to a desired value that will be different from the values created by the other sources. The resulting imbalance may cause excessive (and possibly damaging) source currents. Consequently, each source must be isolated from the others through a resistor.

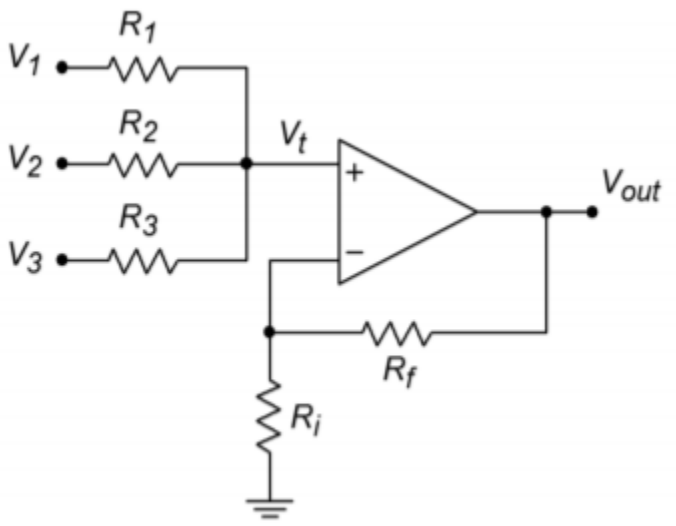


Figure 4.2.22 : Noninverting summing amplifier.

In order to understand the operation of this circuit it is best to break it into two parts: the input source/resistor section and the noninverting amplifier section. The input signals will combine to create a total input voltage, ◈◈ . By inspection, you should see that the output voltage of the circuit will equal ◈◈ times the noninverting gain, or

$$V_{out} = V_t \left(1 + \frac{R_f}{R_i}\right)$$

All that remains is to determine ◈◈ . Each of the input channels contributes to ◈◈ in a similar manner, so the derivation of the contribution from a single channel will be sufficient.

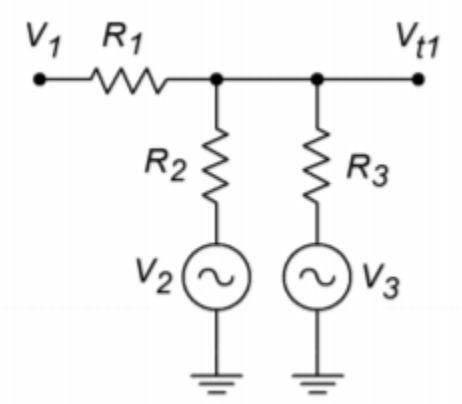


Figure 4.2.23 : Channel 1 input equivalent circuit.

Unlike the inverting summer, the noninverting summer does not take advantage of the virtual-ground summing node. The result is that individual channels will affect each other. The equivalent circuit for channel 1 is redrawn in Figure 4.2.23 . Using superposition, we would first replace the input generators of channels 2 and 3 with short circuits. The result is a simple voltage divider between ◈1 and ◈◈1 .

$$V_{t1} = V_1 \frac{R_2||R_3}{R_1 + R_2||R_3}$$

In a similar manner, we can derive the portions of ?? due to channel 2

$$V_{t2} = V_2 \frac{R_1||R_3}{R_2 + R_1||R_3}$$

and due to channel 3

$$V_{t3} = V_3 \frac{R_1||R_2}{R_3 + R_1||R_2}$$

?? is the summation of these three portions.

$V_t = V_{t1} + V_{t2} + V_{t3}$

Thus, by combining these elements, we find that the output voltage is

$$V_{out} = \left(1 + \frac{R_f}{R_i}\right)\left(V_1 \frac{R_2||R_3}{R_1 + R_2||R_3} + V_2 \frac{R_1||R_3}{R_2 + R_1||R_3} + V_3 \frac{R_1||R_2}{R_3 + R_1||R_2}\right)$$

For convenience and equal weighting, the input resistors are often all set to the same value. This results in a circuit that averages together all of the inputs. Doing so simplifies the Equation to

$$V_{out} = \left(1 + \frac{R_f}{R_i}\right)\frac{V_1 + V_2 + V_3}{3}$$

or in a more general sense,

$$V_{out} = \left(1 + \frac{R_f}{R_i}\right)\frac{\sum_{i=1}^{n} V_n}{n}$$

(4.2.10)

where ? is the number of channels.

One problem still remains with this circuit, and that is inter-channel isolation or crosstalk. This can be eliminated by individually buffering each input, as shown in Figure 4.2.24 .

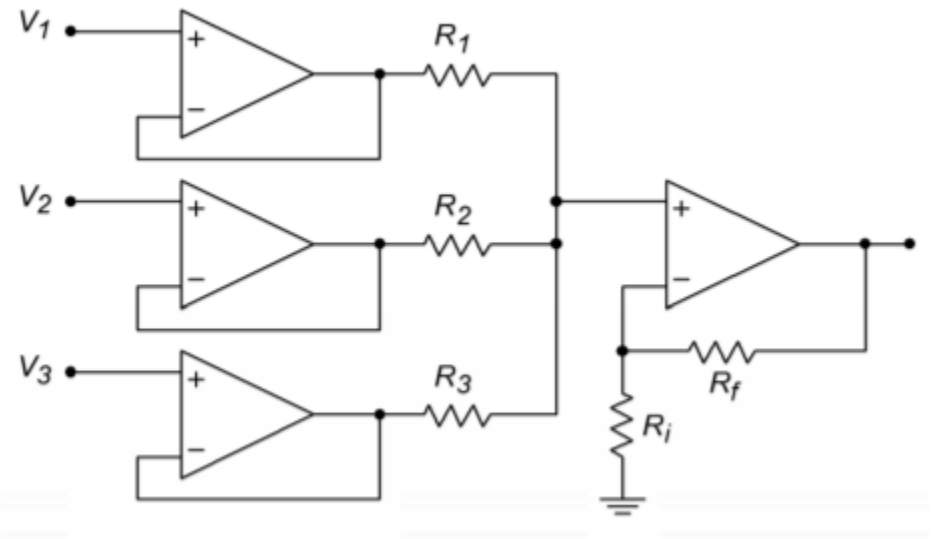


Figure 4.2.24 : Buffered and isolated noninverting summing amplifier.

Example 4.2.14

A noninverting summer such as the one shown in Figure 4.2.22 is used to combine three signals. ⍰1 = 1 V DC, ⍰2 = -0.2 V DC, and ⍰3 is a 2 V peak 100 Hz sine wave. Determine the output voltage if ⍰1=⍰2=⍰3=⍰⍰ = 20 k Ω and ⍰⍰ = 5 k Ω .

Because all of the input resistors are equal, we can use the general form of the summing equation.

$$V_{out} = \left(1 + \frac{R_f}{R_i}\right) \frac{V_1 + V_2 + +V_n}{\text{Number of channels}}$$

$$V_{out} = \left(1 + \frac{20k}{5k}\right) \frac{1VDC + (-0.2VDC) + 2\sin 2\pi 100t}{3}$$

$$V_{out} = 5\frac{0.8VDC + 2\sin 2\pi 100t}{3}$$

$$V_{out} = 1.33VDC + 3.33\sin 2\pi 100t$$

So we see that the output is a 3.33 V peak sine wave riding on a 1.33 V DC offset.

DIFFERENTIAL AMPLIFIER

As long as the op amp is based on a differential input stage, there is nothing preventing you from making a diff amp with it. The applications of an op amp based unit are the same as the discrete version examined in Chapter One. In essence, the differential amplifier configuration is a combination of the inverting and noninverting voltage amplifiers. A candidate is seen in Figure 4.2.25 . The analysis is identical to that of the two base types, and Superposition is used to combine the

results. The obvious problem for this circuit is that there is a large mismatch between the gains if lower values are used. Remember, for the inverting input the gain magnitude is ⓡⓕ/ⓡⓘ , whereas the noninverting input sees ⓡⓕ/ⓡⓘ + 1. For proper operation, the gains of the two halves should be identical. The noninverting input has a slightly higher gain, so a simple voltage divider may be used to compensate. This is shown in Figure 4.2.26 . The ratio should be the same as the ⓡⓕ/ⓡⓘ ratio. The target gain is ⓡⓕ/ⓡⓘ , the present gain is 1 + ⓡⓕ/ⓡⓘ , which may be written as (ⓡⓕ+ⓡⓘ)/ⓡⓘ . To compensate, a gain of ⓡⓕ/(ⓡⓕ+ⓡⓘ) is used.

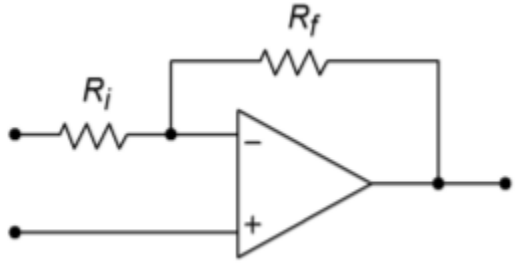


Figure 4.2.25 : Differential amplifier candidate.

$$A_{v+} = \frac{R_f + R_i}{R_i}\frac{R_f}{R_f + R_i}$$

$$A_{v+} = \frac{R_f(R_f + R_i)}{R_i(R_f + R_i)}$$

$$A_{v+} = \frac{R_f}{R_i}$$

(4.2.11)

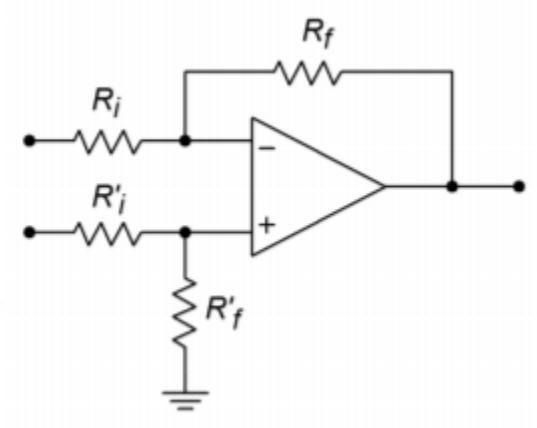


Figure 4.2.26 : Differential amplifier with compensation for mismatched gains.

For a true differential amplifier, ⓡ′ⓘ is set to ⓡⓕ , and ⓡ′ⓕ is set to ⓡⓘ . A small potentiometer is typically placed in series with ⓡ′ⓘ in order to compensate for slight gain imbalances due to component tolerances. This makes it possible for the circuit's common-mode rejection ratio to reach its maximum value. Another option for a simple difference amplifier is to set ⓡ′ⓘ plus ⓡ′ⓕ equal to

. Doing so will maintain roughly equal input impedance between the two halves if two different input sources are used.

Once the divider is added, the output voltage is found by multiplying the differential input signal by /.

Example 4.2.15

Design a simple difference amplifier with an input impedance of 10 k Ω per leg, and a voltage gain of 26 dB.

First of all, converting 26 dB into ordinary form yields 20. Because sets , set = 10 k Ω, from the specifications.

$$A_v = \frac{R_f}{R_i}$$

$$R_f = \frac{A_v}{R_i}$$

$$R_f = 20 \times 10k$$

For equivalent inputs,

$$R_i^{'} + R_f^{'} = R_i$$

$$R_i^{'} + R_f^{'} = 10k$$

Given that = 20,

$$R_f^{'} = 20 \times R_i^{'}$$

Therefore,

$$21 \times R_i^{'} = 10k$$

$$R_i^{'} = 476$$

$$R_f^{'} = 20 \times R_i^{'}$$

$$R_f^{'} = 9.52k$$

The final result is shown in Figure 4.2.27 . As you will see later in Chapter Six, the differential amplifier figures prominently in another useful circuit, the instrumentation amplifier.

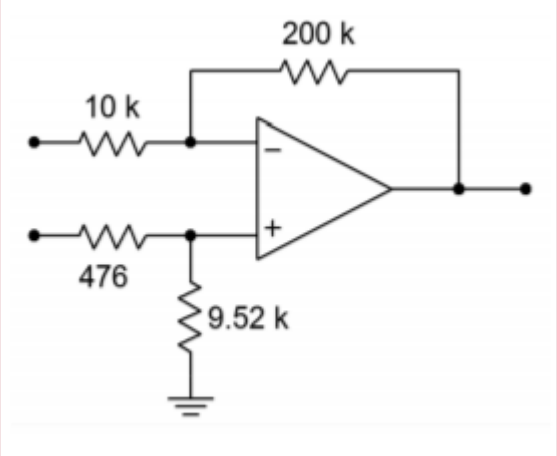


Figure 4.2.27 : Difference amplifier for Example 4.2.15 .

ADDER/SUBTRACTOR

If inverting and noninverting summing amplifiers are combined using the differential amplifier topology, an adder/subtractor results. Normally, all resistors in an adder/subtractor are the same value. A typical adder/subtractor is shown in Figure 4.2.28 .

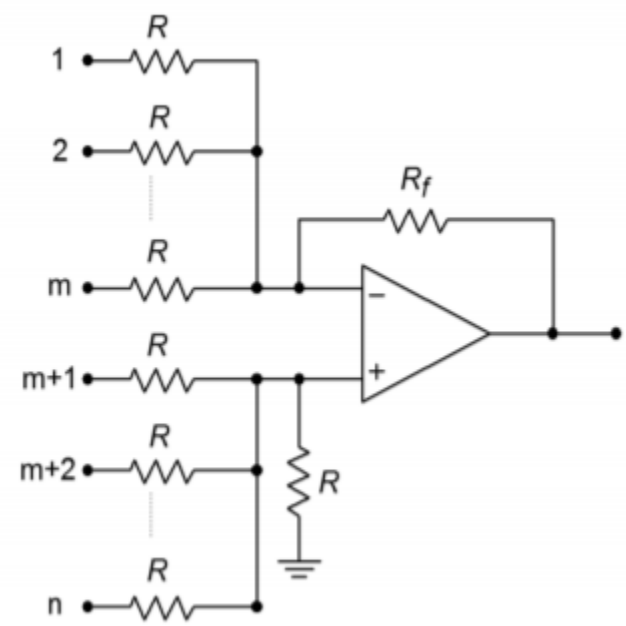


Figure 4.2.28 : Adder-subtractor.

The inverting inputs number from 1 through ◈ , and the noninverting inputs number from ◈+1 through ◈ . The circuit can be analyzed by combining the preceding proofs of Equations 4.2.9 through 4.2.11 via the Superposition Theorem. The details are left as an exercise (Problem 4.45). When all resistors are equal, the input weightings are unity, and the output is found by:

$$V_{out} = \sum_{i=m+1}^{n} V_{in_i} \sum_{j=1}^{m} V_{in_j}$$

(4.2.12)

In essence, you can think of the output voltage in terms of subtracting the inverting input summation from the noninverting input summation.

ADJUSTABLE INVERTER/NONINVERTER

A unique adjustable gain amplifier is shown in Figure 4.2.29 . What makes this circuit interesting is that the gain is continuously variable between an inverting and noninverting maximum. For example, the gain might be set for a maximum of 10. A full turn of the potentiometer would swing the gain from +10 through -10. The exact middle setting would produce a gain of 0. In this way, a single knob controls both the phase and magnitude of the gain.

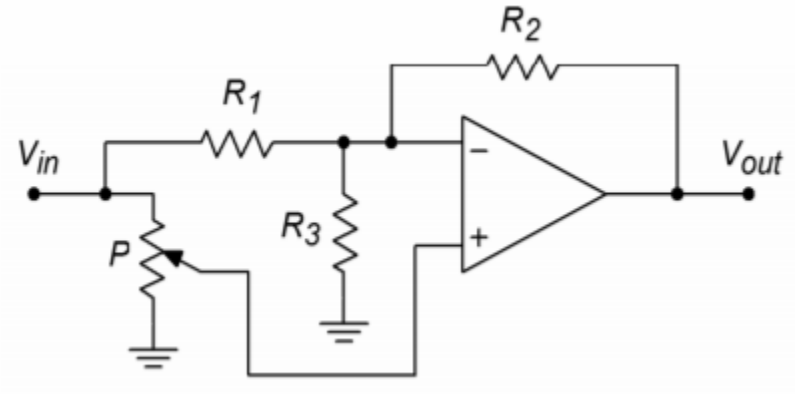


Figure 4.2.29 : Adjustable inverter/ noninverter.

For the analysis of the circuit, refer to Figure 4.2.30 .

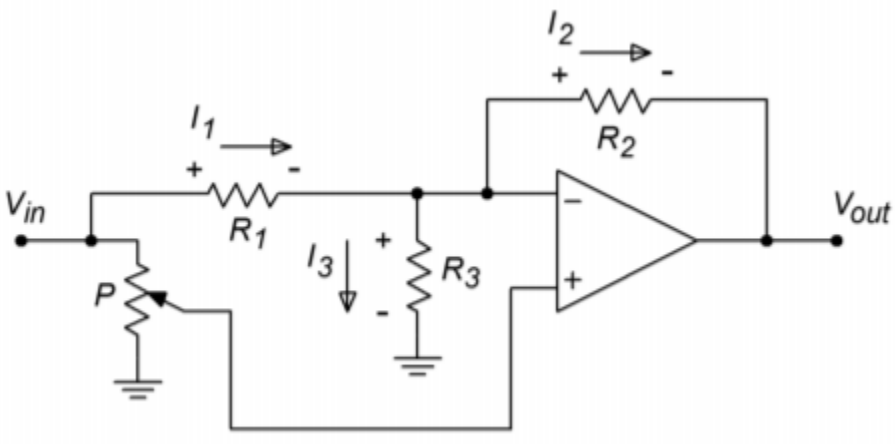


Figure 4.2.30 : Inverter/noninverter analysis.

As you would expect, the gain of the circuit is defined as the ratio of the output to input voltages. It is important to note that unlike a normal inverting amplifier, the magnitude of the output voltage is not necessarily equal to the voltage across ⟐2 . This is because the inverting terminal of the op amp

is not normally a virtual ground. Instead, the voltage across ◈3 must also be considered. Because the two inputs of the op amp must be at approximately the same potential (i.e., ◈◈◈◈◈◈ must be 0), the voltage at the inverting terminal must be the same as the voltage tapped off of the potentiometer. Representing the potentiometer's voltage divider factor as ◈ , we find:

$$V_{out} = kV_{in} - V_{R2}$$

(4.2.13)

The drop across ◈2 is simply ◈2◈2 . ◈2 is found by Kirchhoff's Current Law and the appropriate voltage-resistor substitutions:

$$I_2 = I_1 - I_3$$

$$I_2 = \frac{V_{in} kV_{in}}{R_1} - \frac{kV_{in}}{R_3}$$

$$I_2 = V_{in}\left(\frac{1-k}{R_1} - \frac{k}{R_3}\right)$$

Thus, V_{R2} is found to be

$$V_{R2} = V_{in}\left((1-k)\frac{R_2}{R_1} - k\frac{R_2}{R_3}\right)$$

(4.2.14)

Combining Equations 4.2.13 and ??? and then solving for gain, we find

$$A_v = k - \left((1-k)\frac{R_2}{R_1} - k\frac{R_2}{R_3}\right)$$

$$A_v = k - \left(\frac{R_2}{R_1} - k\frac{R_2}{R_1} - k\frac{R_2}{R_3}\right)$$

$$A_v = -\frac{R_2}{R_1} + k\left(1 + \frac{R_2}{R_1} + k\frac{R_2}{R_3}\right)$$

The value of ◈3 is chosen so that ◈1=◈2||◈3 . This means that

$$R_3 = \frac{1}{\frac{1}{R_1} - \frac{1}{R_2}}$$

Substituting this into our gain Equation and simplifying yields

$$A_v = \frac{R_2}{R_1}(2k1)$$

In essence, the resistors ◈1 and ◈2 set the maximum gain. The potentiometer sets ◈ from 0 through 1. If ◈=1 , then ◈◈=◈2/◈1 , or maximum noninverting gain. When ◈=0 , then ◈◈=–◈2/◈1 , or maximum inverting gain. Finally, when the potentiometer is set to the mid-point, ◈=0.5 and ◈◈=0 .

10.3 SINGLE SUPPLY BIASING

Up to this point, all of the example circuits have used a bipolar power supply, usually ± 15 V. Sometimes this is not practical. For example, a small amount of analog circuitry may be used along with a predominantly digital circuit that runs off a unipolar supply. It may not be economical to create an entire negative supply just to run one or two op amps. Although it is possible to buy op amps that have been specially designed to work with unipolar supplies1, the addition of simple bias circuitry will allow almost any op amp to run from a unipolar supply. This supply can be up to twice as large as the bipolar counterpart. In other words, a circuit that normally runs off a ± 15 V supply can be configured to run off of a +30 V unipolar supply, producing similar performance. We will look at examples using both the noninverting and inverting voltage amplifiers

The idea is to bias the input at one-half of the total supply potential. This can be done with a simple voltage divider. A coupling capacitor may be used to isolate this DC potential from the driving stage. For proper operation, the op amp's output should also be sitting at one-half of the supply. This fact implies that the circuit gain must be unity. This may appear to be a very limiting factor, but in reality, it isn't. The thing to remember is that the gain need only be unity for DC. The AC gain can be just about any gain you'd like.

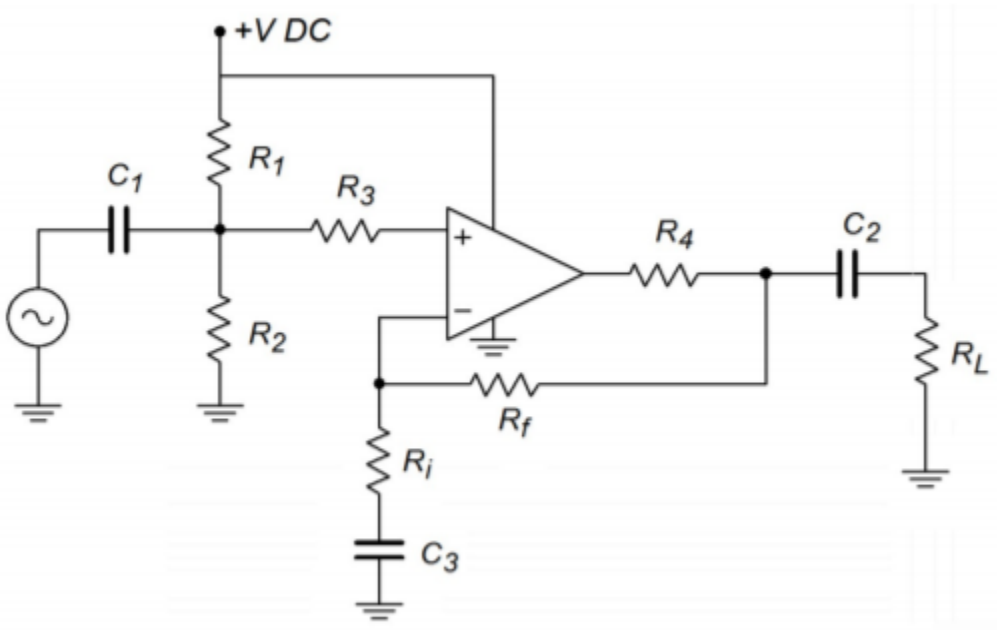


Figure 4.3.1 : Single-supply bias in a noninverting amplifier.

An example using the noninverting voltage amplifier is shown in Figure 4.3.1 . In order to set DC gain to unity without affecting the AC gain, capacitor ⍰3 is placed in series with ⍰⍰ . ⍰1 and ⍰2 establish the 50% bias point. Their parallel combination sets the input impedance too. Resistors ⍰3 and ⍰4 are

used to prevent destructive discharge of the coupling capacitors ⍰1 and ⍰2 into the op amp. They may not be required, but if present, typically run around 1 k Ω and 100 Ω respectively.

The inclusion of the capacitors produces three lead networks. A standard frequency analysis and circuit simplification shows that the approximate critical frequencies are

$$f_{in} = \frac{1}{2\pi C_1 R_1 || R_2}$$

$$f_{out} = \frac{1}{2\pi C_2 R_{load}}$$

$$f_{fdbk} = \frac{1}{2\pi C_3 R_i}$$

The input bias network can be improved by using the circuit of Figure 4.3.2 . This reduces the hum and noise transmitted from the power supply into the op amp's input. It does so by creating a low impedance at node A. This, of course, does not affect the DC potential. ⍰5 now sets the input impedance of the circuit.

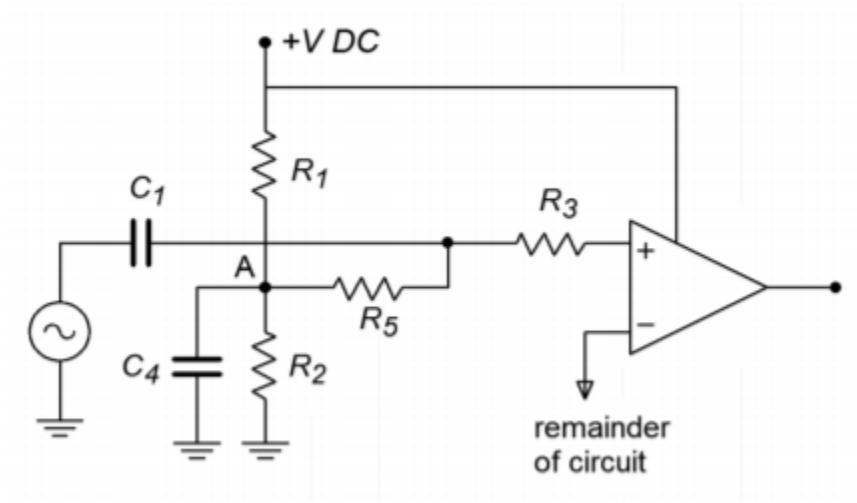


Figure 4.3.2 : Improved bias for the circuit in Figure 4.3.1 .

The important points to remember here are that voltage gain is still 1+⍰⍰/⍰⍰ in the midband, ⍰⍰⍰ is now set by the biasing resistors ⍰1 and ⍰2 , or ⍰5 (if used), and that frequency response is no longer flat down to zero Hertz.

A single-supply version of the inverting voltage amplifier is shown in Figure 4.3.3 . It uses the same basic techniques as the noninverting form. The bias setup uses the optimized low-noise form. Note that there is no change in input impedance, it is still set by ⍰⍰ . The approximate lead network critical frequencies are found through

$$f_{in} = \frac{1}{2\pi C_1 R_i}$$

$$f_{out} = \frac{1}{2\pi C_2 R_{load}}$$

$$f_{bias} = \frac{1}{2\pi C_3 R_1 || R_2}$$

Note the general similarity between the circuits of Figures 4.3.3 and 4.3.1 . A simple redirection of the input signal creates one form from the other.

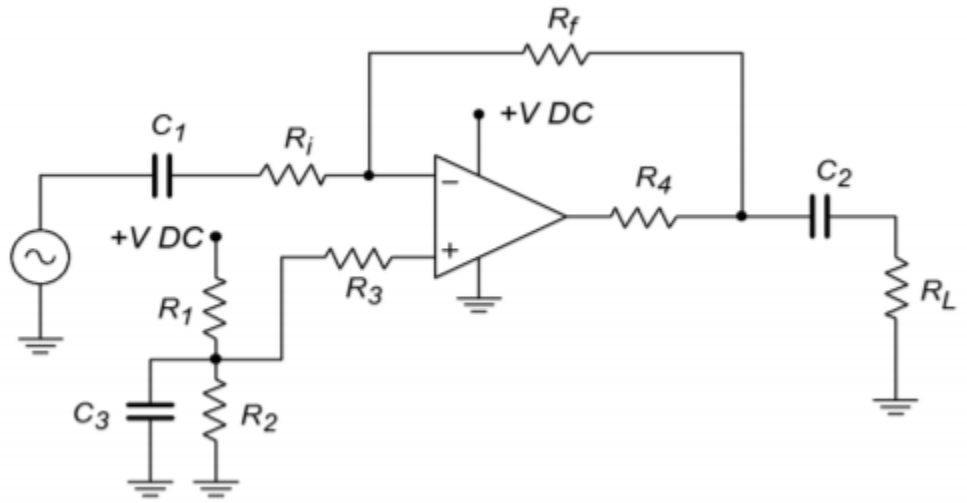


Figure 4.3.3 : Single-supply inverting amplifier.

10.4 CURRENT BOOSTING

As previously noted, general-purpose op amps produce a maximum output current of around 20 mA. This is sufficient for a wide variety of uses. If the load is less than about 1 k Ω , the op amp will start to clip on the higher output signals. The average op amp cannot drive low impedance loads. A few examples of applications whose loads are inappropriate include distribution amplifiers, small audio power amplifiers such as a headphone amplifier, and small motors. This is most unfortunate, as we have already seen how useful these devices can be. There is a way out, though. It is possible to include a current gain stage right after the op amp. All that is needed is a simple class B or class AB push-pull follower. This follower will be able to produce the higher current required by low impedance loads. The op amp only needs to drive the follower stage. In order to increase system linearity and lower distortion, the follower can be placed inside of the op amp's feedback loop. Because the follower is noninverting, there is no problem with maintaining correct feedback (this assumes that the power devices used have a wider bandwidth than the op amp). An example of this is shown in Figure 4.4.1 .

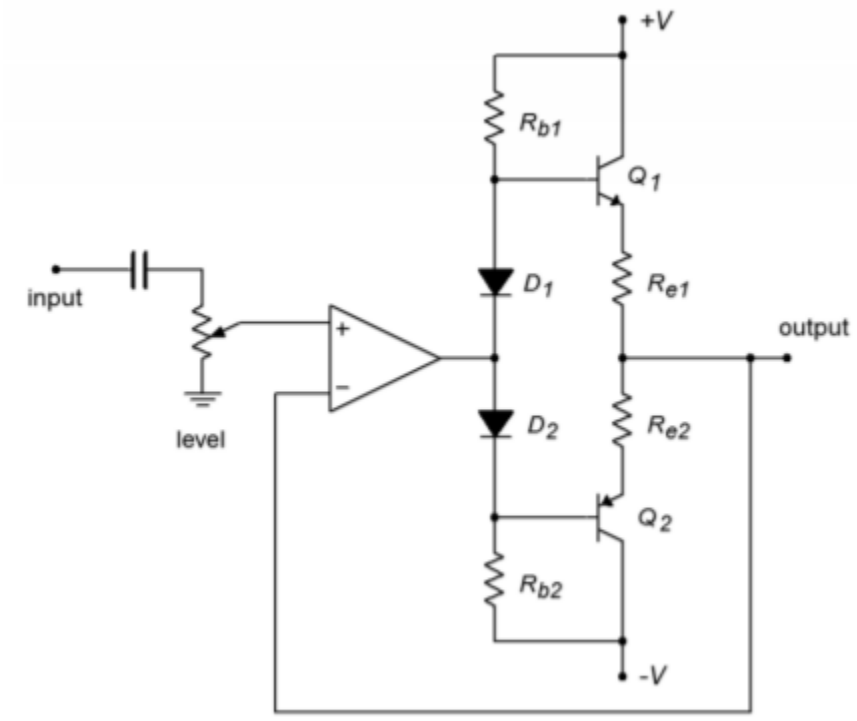


Figure 4.4.1 : Current boosting.

This circuit is typical of an electronic crossover or distribution amplifier (an electronic crossover design example is given in Chapter Eleven). This output circuit needs to drive relatively low impedances through long cable runs of perhaps several hundred feet. The excessive capacitance resulting from long cable runs increases the current demand above that of a purely resistive load.

Circuits like the one in Figure 4.4.1 can produce currents of several hundred milliamps or more. Many times, small resistors are placed in the emitter or collector as a means of limiting maximum

current or reducing distortion. The maximum output current limitation is a function of the class B devices. Some manufacturers offer current-boosting ICs to further simplify the design. The current-booster is a drop-in replacement for the class B follower. For very high current demands, Darlingtons or multi-stage designs may be required. It is even possible to provide voltage gain stages. Indeed, several consumer audio power amplifiers have been designed in exactly this way. In essence, the designers produce a discrete power amplifier and then "wrap it" within an op amp feedback loop.

COMPUTER SIMULATION

A basic current booster is simulated using Multisim in Figure 4.4.2 . In order to see secondary effects, the LF411 model has been chosen instead of the ideal device model.

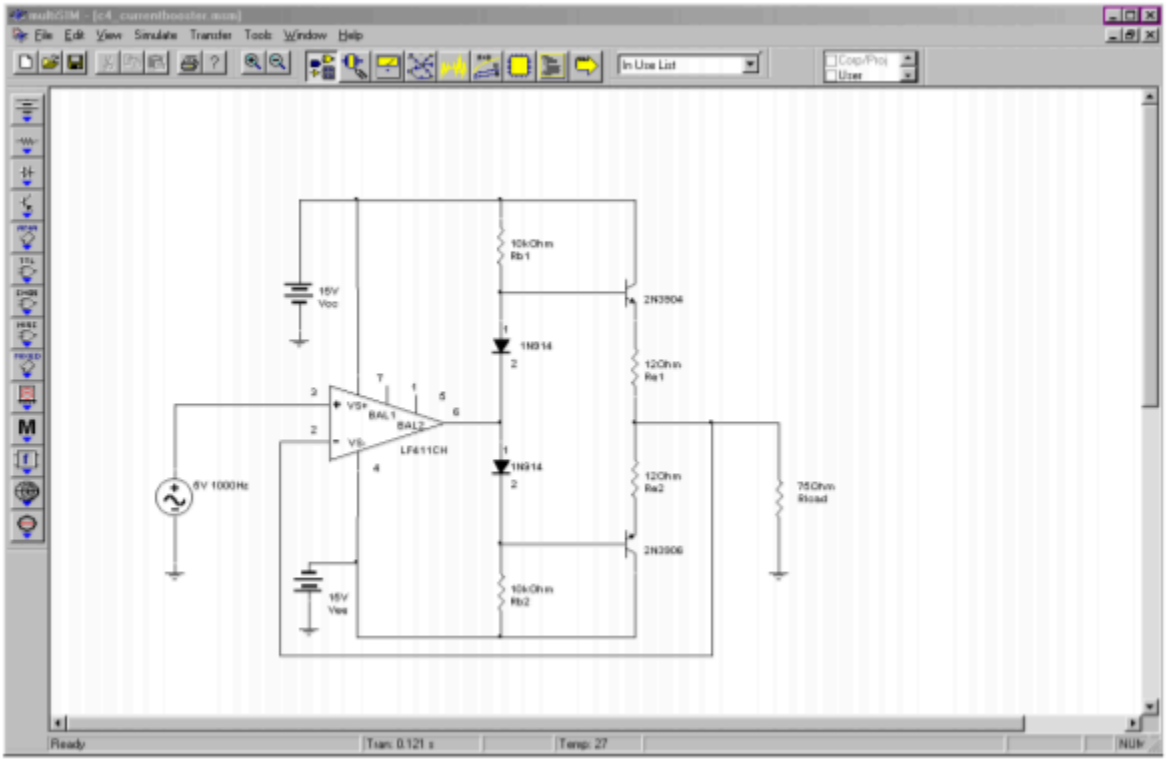


Figure 4.4.2 : Current booster simulation schematic.

The circuit is configured for a voltage gain of unity, thus the 5 volt input signal should yield a 5 volt output. According to its data sheet, the short-circuit current of the LF411 is approximately 25 mA at room temperature. It is not capable of driving a 75 Ω load to 5 volts by itself. The Transient Analysis shows a full 5 volt output signal, indicating the effectiveness of the current boosting stage. Also, a close inspection of the output waveform shows no obvious forms of distortion such as the cross-over distortion typical of simple class B stages. This shows that keeping the class B stage within the feedback loop does indeed minimize distortion.

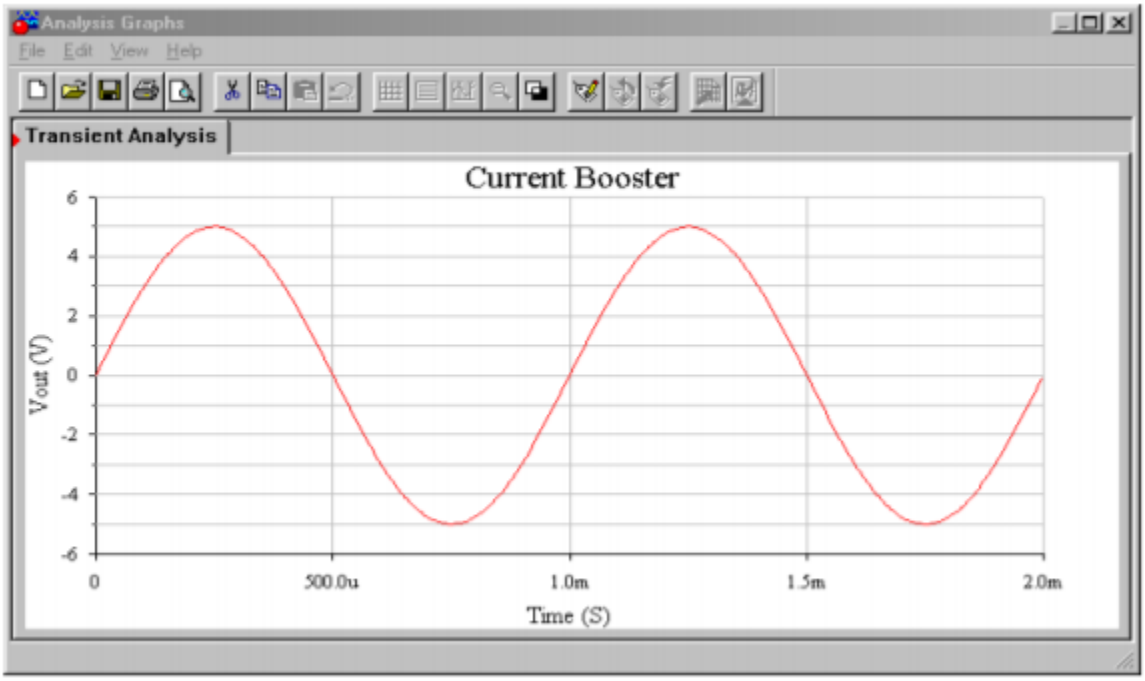


Figure 4.4.2 : Simulation of output waveform.

10.5 SUMMARY

In this chapter we have explored a variety of basic op amp circuits and learned a few analysis shortcuts. The basic assumptions are that the error voltage (differential input voltage) is zero and that the op amp's input current is zero. In all circuits, the gain or transfer parameter is a function of just one or two resistors. Circuits can be made that produce voltage gain, current gain, voltage-to-current conversion, or current-to-voltage conversion. The most popular op amp circuits are the noninverting voltage amplifier and the inverting voltage amplifier. These are based on SP and PP negative feedback, respectively. The noninverting type shows an ideally infinite input impedance, whereas the inverting type has its input impedance set by one of the feedback resistors. A variation of the inverting voltage amplifier is the summing amplifier. This adds its several input channels together in order to arrive at its single output signal. The input node is at virtual ground. The differential amplifier is basically the simultaneous use of both the inverting and noninverting voltage amplifier forms.

The voltage-to-current transducer is based on SS feedback. Its transconductance is set by a single feedback resistor. In a similar manner, the current-to-voltage transducer is based on PP feedback and has a single feedback resistor to set its transresistance. The current amplifier is based on PS feedback.

Although op amps are designed to run off bipolar power supplies, they can be used with unipolar supplies. Extra circuitry is needed for the proper bias. There is no restriction on AC gain; however, DC gain must be set to unity. Because lead networks are introduced, the system gain cannot be flat down to zero Hertz.

Finally, if higher output current requirements need to be met, it is possible to boost the op amp's capabilities with a discrete output stage. This stage is typically a class B or class AB push-pull follower. In order to lower system distortion, the follower is kept within the op amp's feedback loop.

10.6 PROBLEMS

REVIEW QUESTIONS

1. What forms of feedback are used for the inverting and noninverting voltage amplifiers?
2. What forms of feedback are used for the current-to-voltage and voltage-to-current transducers?
3. What form of feedback is used for the inverting current amplifier?
4. What are the op amp analysis idealizations?
5. What is virtual ground?
6. What is a summing amplifier?
7. How can output current by increased?
8. What circuit changes are needed in order to bias an op amp with a unipolar supply?
9. What operational parameters change when a circuit is set up for single supply biasing?
10. How might a circuit's gain be controlled externally?
11. What is meant by the term "floating load"?

PROBLEMS

Analysis Problems

1. What is the voltage gain in Figure 4.6.1 ? What is the input impedance?

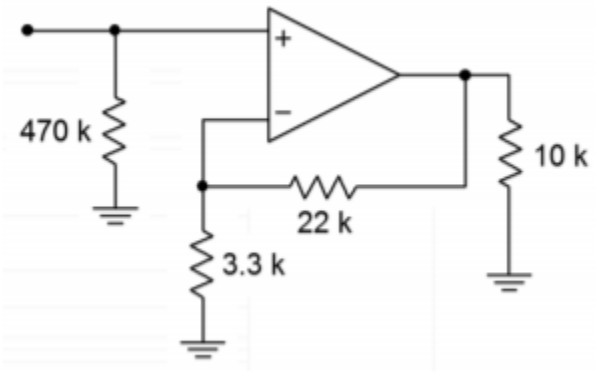


Figure 4.6.1

2. What is the voltage gain for the first stage of Figure 4.6.2 ? What is the input impedance?

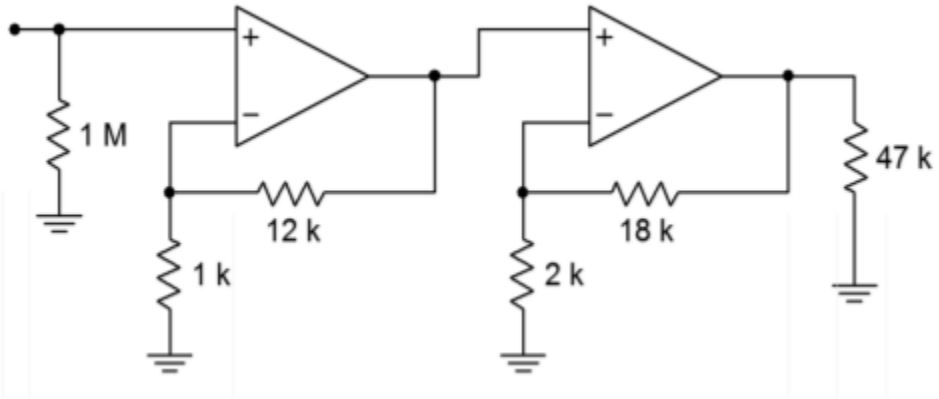


Figure 4.6.2

3. What is the voltage gain for the second stage of Figure 4.6.2 ? What is the input impedance?
4. What is the system voltage gain in Figure 4.6.24.6.2? What is the input impedance?
5. If the input to Figure 4.6.2 is -52 dBV, what is ?'??? ?
6. What is the voltage gain in Figure 4.6.3? What is the input impedance?
7. If the input voltage to the circuit of Figure 4.6.3 is 100 mV, what is ????????

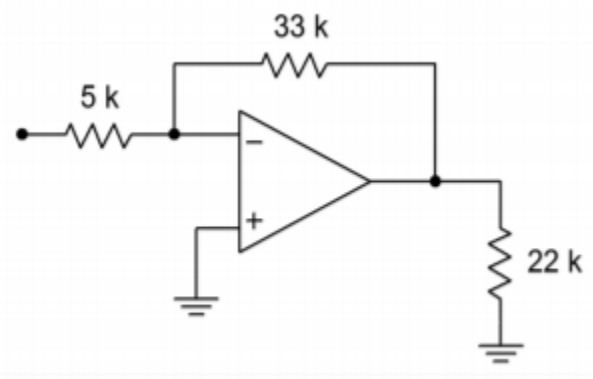


Figure 4.6.3

8. What is the system input impedance in Figure 4.6.4 ? What is the system gain?

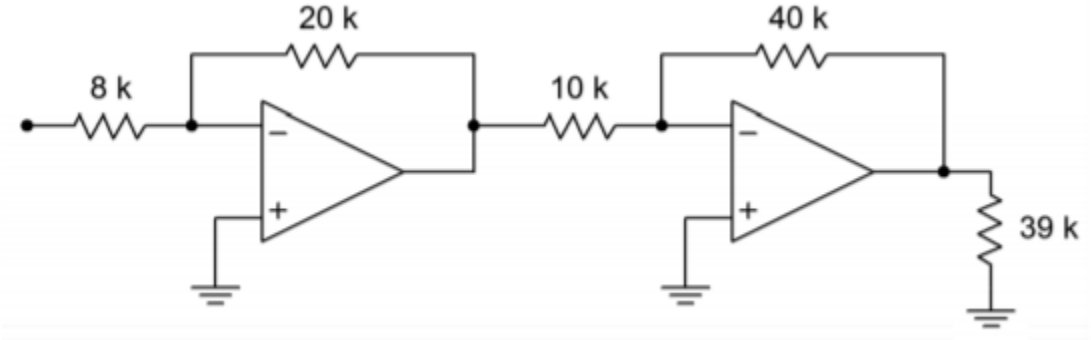


Figure 4.6.4

9. Redesign Figure 4.6.4 for an input impedance of 20 k Ω .

10. Given an input current of 2 ⋄ A, what is the output voltage in Figure 4.6.5 ?

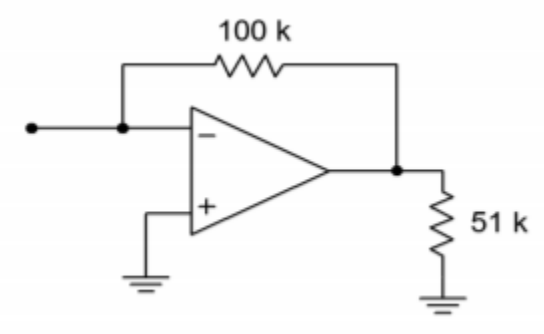


Figure 4.6.5

11. What is the meter deflection in Figure 4.6.6 if the input voltage is 1 V?

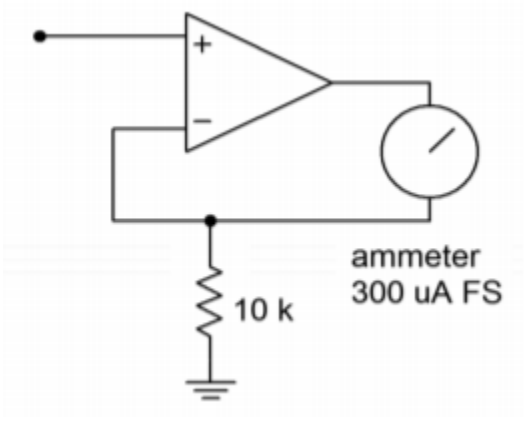


Figure 4.6.6

12. What input voltage will produce full-scale deflection in figure 4.6.6 ?
13. Determine a new value for the 10 k Ω resistor in Figure 4.6.6 such that a 0.1 V input will produce full-scale deflection.
14. What is the current gain in Figure 4.6.7 ?

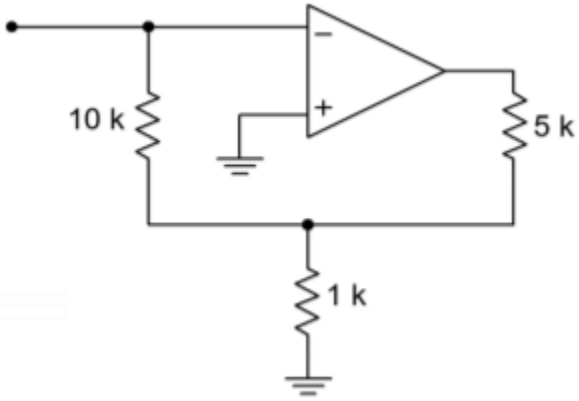


Figure 4.6.7

15. What is the maximum input current in Figure 4.6.7 , assuming the circuit is running off of ±

15 V supplies, and the op amp has a maximum output current of 25 mA?

16. If the differential input signal is 300 mV in Figure 4.6.8 , what is ?

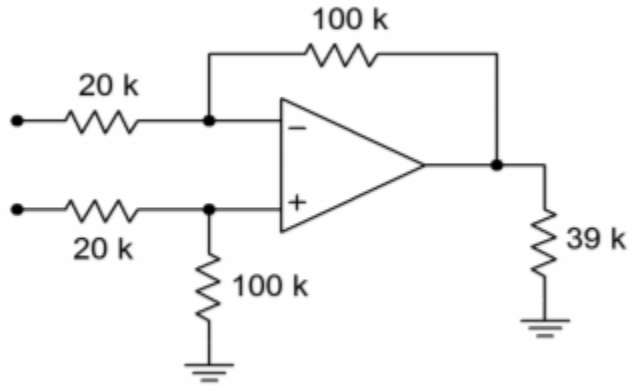


Figure 4.6.8

17. Determine new values for the voltage divider resistors in Figure 4.6.8 , such that the resulting input impedance is balanced.

Design Problems

18. Design a noninverting amplifier with a voltage gain of 32 dB and an input impedance of 200 k Ω.
19. Design a voltage follower with a gain of 0 dB.
20. Design an inverting amplifier with a voltage gain of 14 dB and an input impedance of 15 k Ω.
21. Design a current-to-voltage transducer such that a 20 A input current will produce a -1 V output.
22. Design a voltage-to-current transducer such that a 100 mV input will produce a 1 mA output.
23. Design a current amplifier with a gain of -20.
24. Design a differential amplifier with a gain of 18 dB and a balanced input impedance of 25 k Ω per input.
25. Design a voltage to current transducer with a transconductance of 1 mS. If is 200 mV, what is ?
26. Design a current to voltage transducer with a transresistance of 10 k Ω . If the input current is 500 A, what is ?
27. Redesign the circuit of Figure 4.6.1 for single supply operation (don't bother calculating capacitor values).
28. Redesign the circuit of Figure 4.6.3 for single supply operation (don't bother calculating capacitor values).
29. Design a summing amplifier such that channel 1 has a gain of 10, channel 2 has a gain of 15, and channel 3 has a gain of 5. The minimum channel input impedance should be 1 k Ω .

30. Determine capacitor values for Problem 27 if the lower break frequency ◈1 , is set to 20 Hz.
31. Determine capacitor values for Problem 28 if the lower break frequency f1, is set to 10 Hz.

Challenge Problems

32. Design a three channel summing amplifier such that: channel 1 ◈◈◈≥10◈Ω , ◈◈ = 6 dB; channel 2 ◈◈◈≥22◈Ω , ◈◈ = 10 dB; and channel 3 ◈◈◈≥5◈Ω , ◈◈ = 16 dB
33. Assuming 10% resistor values, determine the production gain range for Figure 4.6.1 .
34. Assuming 5% resistor values, determine the highest gain produced in Figure 4.6.2 .
35. Design an inverting amplification circuit with a gain of at least 40 dB, and an input impedance of at least 100 k Ω . No resistor used may be greater than 500 k Ω . Multiple stages are allowed.
36. Redesign the circuit of Figure 4.6.3 as a voltmeter with 500 mV, 2 V, 5 V, 20 V, and 50 V ranges.
37. Assuming 1% precision resistors and a meter accuracy of 5%, what range of input values may produce a full-scale reading of 2 V, for the circuit of Problem 36?
38. Design an amplifier with a gain range from -10 dB to +20 dB, with an input impedance of at least 10 k Ω .
39. Design an amplifier with a gain range from 0 to 20. The input impedance should be at least 5 k Ω .
40. What is the input impedance in Figure 4.6.9 ? What is ◈◈ ?

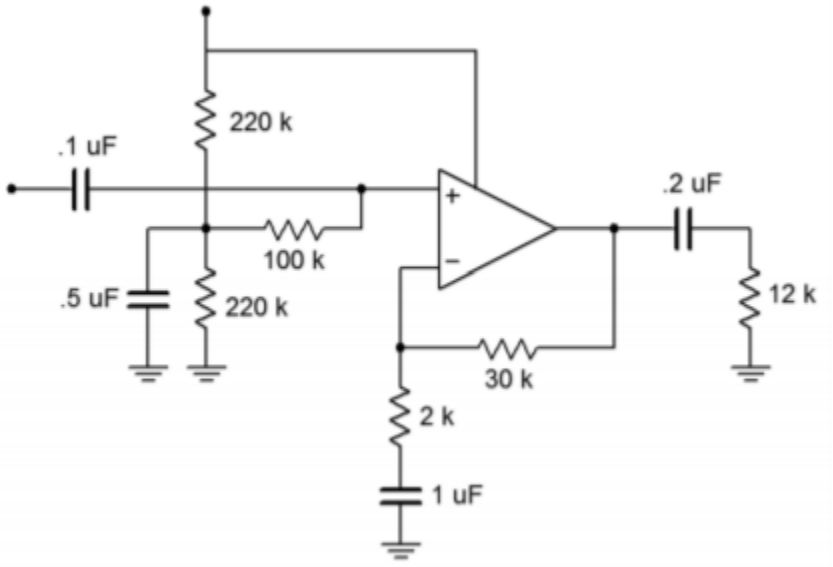


Figure 4.6.9

41. How much power supply ripple attenuation does the input biasing network of Figure 4.6.9 produce (assume ◈◈◈◈◈◈◈ = 120 Hz)?
42. Assume that the circuit of Figure 4.6.10 utilizes a standard 20 mA output op amp. If the output devices are rated for a maximum collector current of 5 amps and a Beta of 50, what is

the maximum load current obtainable?

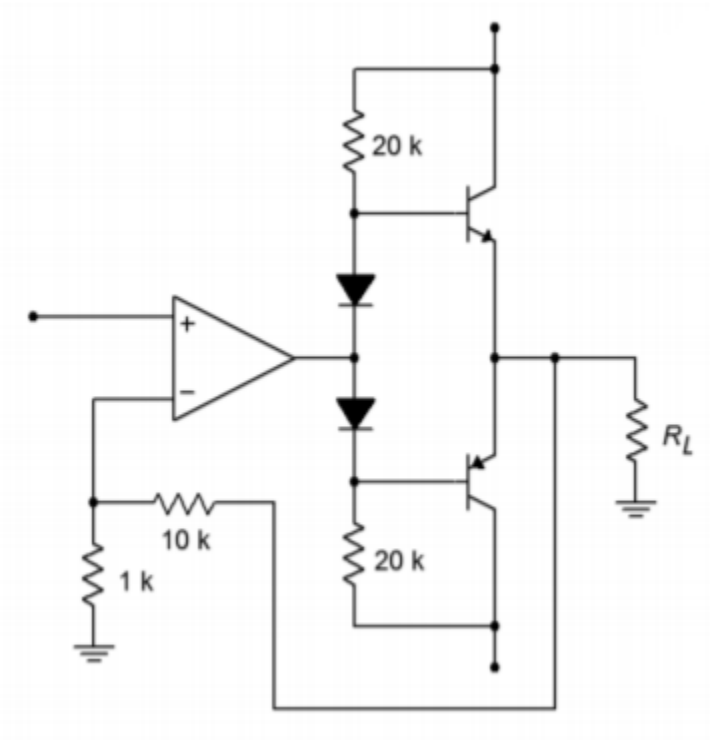


Figure 4.6.10

43. What are the voltage gain and input impedance in Figure 4.6.10 ?
44. Given a summer based on Figure 4.2.22, sketch the output waveform if ◈1=◈2= 10 k Ω , ◈3=◈◈= 30 k Ω , ◈◈ = 15 k Ω , ◈1 = 0.3 V DC, ◈2=0.1sin2◈50◈ and ◈3=–0.2sin2◈200◈ .
45. Prove Equation 4.2.15 for the case when all resistors are of equal value.

Computer Simulation Problems

46. Simulate the operation of the circuit in Figure 4.2.9. Verify the output voltage and the virtual ground at the inverting input.
47. Use a simulator to verify the maximum and minimum gains of the circuit in Figure 4.2.11.
48. Use a simulator to verify the load current and the voltage of the circuit in Figure 4.2.17.
49. Verify the output potential of the circuit in Figure 4.2.20.
50. Simulate the output voltage of the circuit of Figure 4.6.4 for the following inputs:
    A. ◈◈◈=0.1 VDC,
    B. ◈◈◈(◈)=1sin2◈10◈,
    C. ◈◈◈=5 VDC.

    Also, note the potential at the output of the first stage. How might your op amp model affect the results?
51. Simulate the circuit in Figure 4.2.27. Determine the output potential for the following inputs:

A. ⬦⬦⬦+(⬦)=0.1sin2⬦10⬦ , ⬦⬦⬦–(⬦)=0.1sin2⬦10⬦,

B. ⬦⬦⬦+(⬦)=0.1sin2⬦10⬦ , ⬦⬦⬦–(⬦)=–0.1sin2⬦10⬦.

52. Simulate the circuit of Figure 4.6.10 , and determine the output of the circuit and op amp for inputs of 0.1 V DC and 1 V DC.

Use the space below creatively. Draw a cartoon. Compose a song. Write a limerick. Whatever works for you. If you're stuck, try starting with this: ♫

UNIT 11: PRACTICAL LIMITATIONS OF OP AMP CIRCUITS

Learning Objectives

After completing this chapter, you should be able to:

- Define gain-bandwidth product and describe its use in circuit design and analysis.
- Determine upper and lower break frequencies in a multi-stage circuit.
- Define slew rate and power bandwidth, and calculate their effect on circuit performance.
- Understand the difference between power bandwidth and small-signal bandwidth.
- Detail the differences between compensated, noncompensated, and decompensated op amps.
- Calculate the DC offset of an op amp circuit and understand how to minimize it.
- Calculate the DC drift of an op amp circuit and understand how to minimize it.
- Discuss which factors affect the noise performance of an op amp circuit.
- Calculate the noise voltage of an op amp circuit.
- Analyze the CMRR, PSRR, and S/N performance of an op amp circuit.

11.1 INTRODUCTION

Up to now, the op amp has been treated as an ideal device. Although these idealizations are very useful in their place, closer examination must follow. Without this knowledge, it will be impossible to accurately predict a circuit's performance for very high or low frequencies, to judge its noise characteristics, or to determine its stability with temperature or power supply variations. With this information, you will be able to optimize circuit performance for given applications. A major part of this is determining the most desirable op amp for the job. The function of this chapter, then, is to delve deeper into the specifics of individual op amps and to present methods for determining system parameters such as frequency response, noise level, offsets, and drift. The primary interest is in investigating the popular inverting and noninverting voltage amplifier topologies.

11.2 FREQUENCY RESPONSE

In Chapter Four, a number of equations were presented for the various amplifier topologies. These enabled you to find the circuit gain, among other things. These equations are, of course, only valid in the midband region of the amplifier. They say nothing of the amplifier response at the frequency extremes. Chapter One showed that all amplifiers eventually roll off their gain as the input frequency increases. Some amplifiers exhibit a rolloff as the input frequency is decreased as well. Op amp circuits are no exception. There are two things we can say about the average op amp circuit's frequency response: (1) if there are no coupling or other lead network capacitors, the circuit gain will be flat from midband down to DC; and (2) there will eventually be a well-controlled high-frequency rolloff that is usually very easy to find. Item one should come as no great surprise, but you may well wonder about the second. For general-purpose op amps, the high frequency response may be determined with a parameter called the gain-bandwidth product, often abbreviated GBW.

11.3 GAIN-BANDWIDTH PRODUCT

The open loop frequency response of a general-purpose op amp is shown in Figure 5.3.1 . Although the exact frequency and gain values will differ from model to model, all devices will exhibit this same general shape and 20 dB per decade rolloff slope. This is because the lag break frequency ( ) is determined by a single capacitor called the compensation capacitor. This capacitor is usually in the Miller position (i.e., straddling input and output) of an intermediate stage, such as in Figure 5.3.1 . Although this capacitor is rather small, the Miller effect drastically increases its apparent value. The resulting critical frequency is very low, often in the range of 10 to 100 Hz. The other circuit lag networks caused by stray or load capacitances are much higher, usually over 1 MHz. As a result, a constant 20 dB per decade rolloff is maintained from up to very high frequencies. The remaining lag networks will not affect the open-loop response until the gain has already dropped below zero dB.

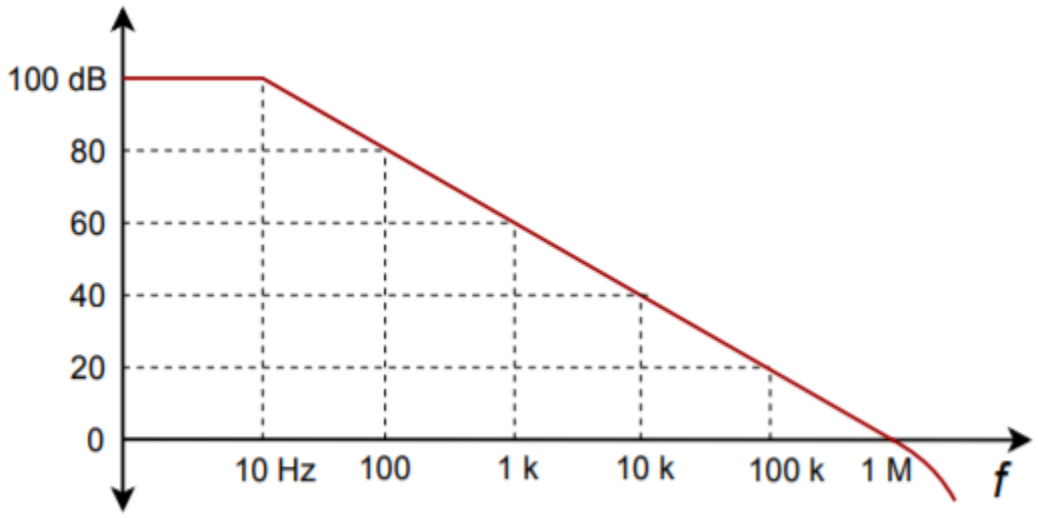


Figure 5.3.1 : Open-loop frequency response.

This type of frequency response curve has two benefits: (1) The most important benefit is that it allows you to set almost any gain you desire with stability. Only a single network is active, thus satisfactory gain and phase margins will be maintained. Therefore, your negative feedback never turns into positive feedback (as noted in Chapter Three). 2) The product of any break frequency and its corresponding gain is a constant. In other words, the gain decreases at the same rate at which the frequency increases. In Figure 5.3.1 , the product is 1 MHz. As you might have guessed, this parameter is the gain-bandwidth product of the op amp (GBW). GBW is also referred to as (the frequency at which the open loop gain equals one). You will find both terms used on manufacturer's spec sheets.

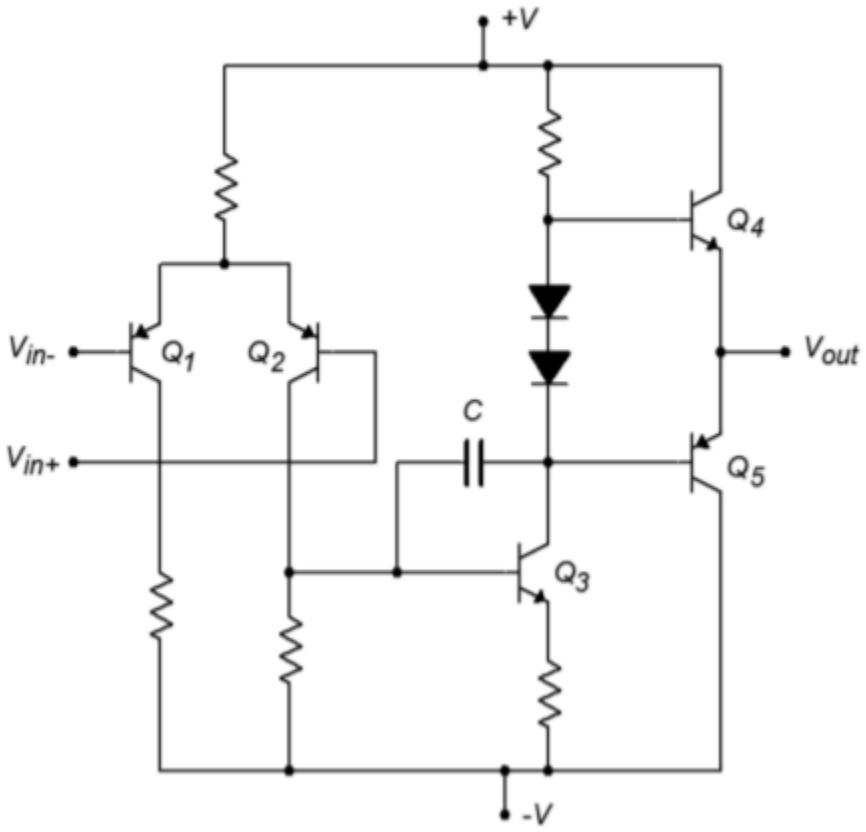


Figure 5.3.1 : Miller compensation capacitor.

As you already know, operating an op amp with negative feedback lowers the midband gain. To a first approximation, this gain will continue until it reaches the open loop response. At this point, the closed loop response will follow the open-loop rolloff. Remember, this is due to the reduction in loop gain, as seen in Chapter Three. This effect is shown in Figure 5.3.2 .

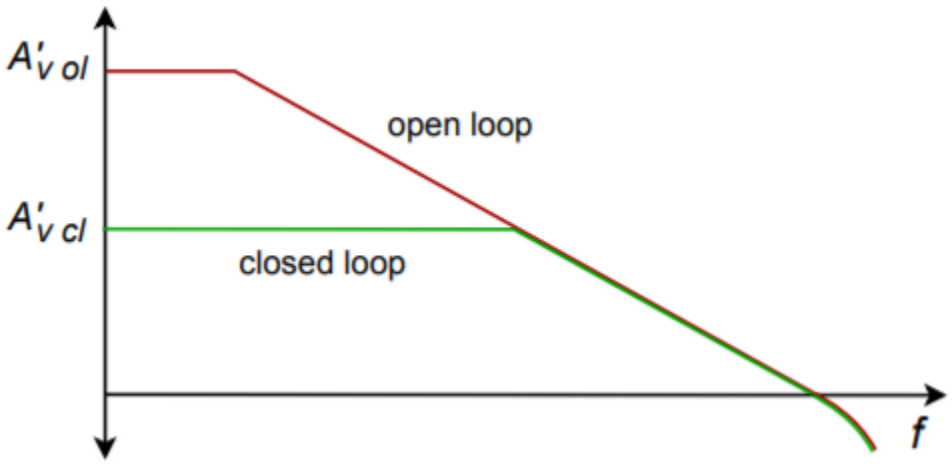


Figure 5.3.2 : Comparison of open-loop and closed-loop responses.

By knowing GBW and the gain, the associated break frequency can be quickly determined. For the inverting and noninverting voltage amplifiers,

$$f_2 = \frac{GBW}{A_{noise}}$$

(5.3.1)

The use of noise gain versus ordinary voltage gain simplifies things and actually makes the results a bit more accurate. Noise gain is the same for both the inverting and noninverting voltage amplifiers. The use of noise gain helps us to take into account the true (non-ideal) feedback effects and circuit imperfections. An example of these limitations is that the open loop gain of an op amp is never infinite. To find the noise gain for any circuit, short all voltage sources and open all current sources. The only item remaining for each source should be its internal resistance. At this point, simplify the circuit as required, and find the gain from the noninverting input to the output of the op amp. This gain is the noise gain. For the standard inverting and noninverting voltage amplifiers, we find

$$A_{noise} = 1 + \frac{R_f}{R_i}$$

Noise gain is the same as ordinary voltage gain for the noninverting voltage amplifier, but is one unit larger than the inverting amplifier's ordinary gain (◈◈/◈◈) . The deviation is only noticeable at lower gains. This does imply though, that for the same gain, noninverting amplifiers will exhibit a higher break frequency than inverting types. Thus, for maximum bandwidth with low gain circuits, the noninverting form is generally preferred. The worst case occurs with an ordinary voltage gain of 1. For the noninverting configuration, the noise gain will also equal 1, and the closed loop bandwidth will equal ◈◈◈◈◈◈ . On the other hand, an inverting amplifier with a voltage gain of 1 will produce a noise gain of 2 and will exhibit a small-signal bandwidth of ◈◈◈◈◈◈/2 . Never use the gain in dB form for this calculation!

Example 5.3.1

Using a 741 op amp, what is the upper break frequency for a noninverting amplifier with a gain of 20 dB?

A 741 data sheet shows a typical GBW of 1 MHz. The noise gain for a noninverting amplifier is the same as its ordinary gain. Converting 20 dB into ordinary form yields a gain of 10.

$$f_2 = \frac{GBW}{A_{noise}}$$

$$f_2 = \frac{1MHz}{10}$$

$$f_2 = 100kHz$$

So, the gain is constant at 10 up to 100 kHz. Above this frequency the gain rolls off at 20 dB per decade.

Example 5.3.2

Sketch the frequency response of the circuit in Figure 5.3.3

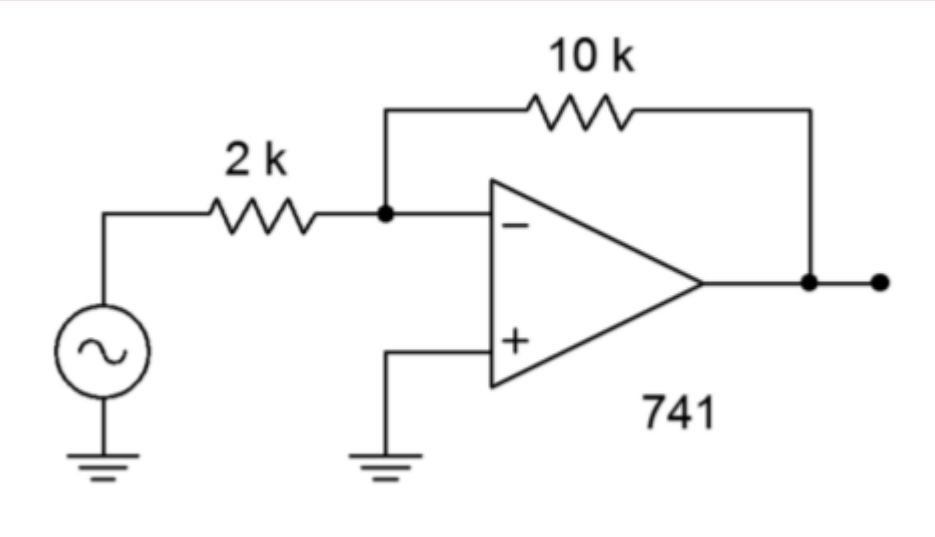


Figure 5.3.3 : Circuit for Example 5.3.2 .

This is an inverting voltage amplifier. The gain is

$$A_v = -\frac{R_f}{R_i}$$

$$A_v = -\frac{10k}{2k}$$

$$A_v = -5$$

For noise gain

$$A_{noise} = 1 + \frac{R_f}{R_i}$$

$$A_{noise} = 1 + \frac{10k}{2k}$$

$$A_{noise} = 6$$

From a data sheet, GBW for a 741 is found to be 1 MHz.

$$f_2 = \frac{GBW}{A_{noise}}$$

$$f_2 = \frac{1MHz}{167kHz}$$

$$f_2 = 167kHz$$

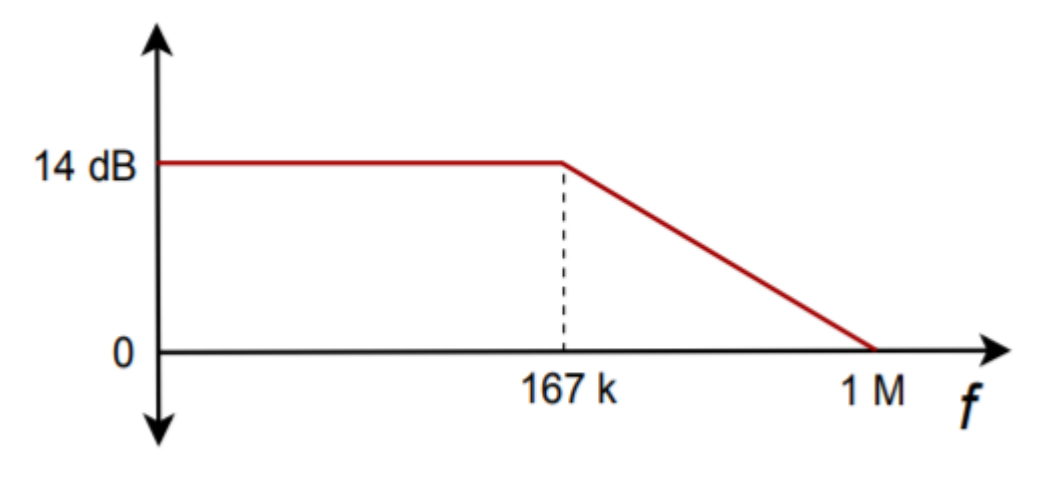


Figure 5.3.4 : Bode plot for the circuit.

The resulting gain Bode plot is shown in Figure 5.3.4 . Note that if a "faster" op amp is used (i.e., one with a higher GBW, such as the LF411), the response will extend further. As you might guess, faster op amps are more expensive.

COMPUTER SIMULATION

The simulation results for Example 5.3.2 are also shown in Figure 5.3.4 . The low frequency gain agrees with the hand calculation of approximately 14 dB. The 3 dB down point ($_2$) also agrees with the calculated break of approximately 167 kHz.

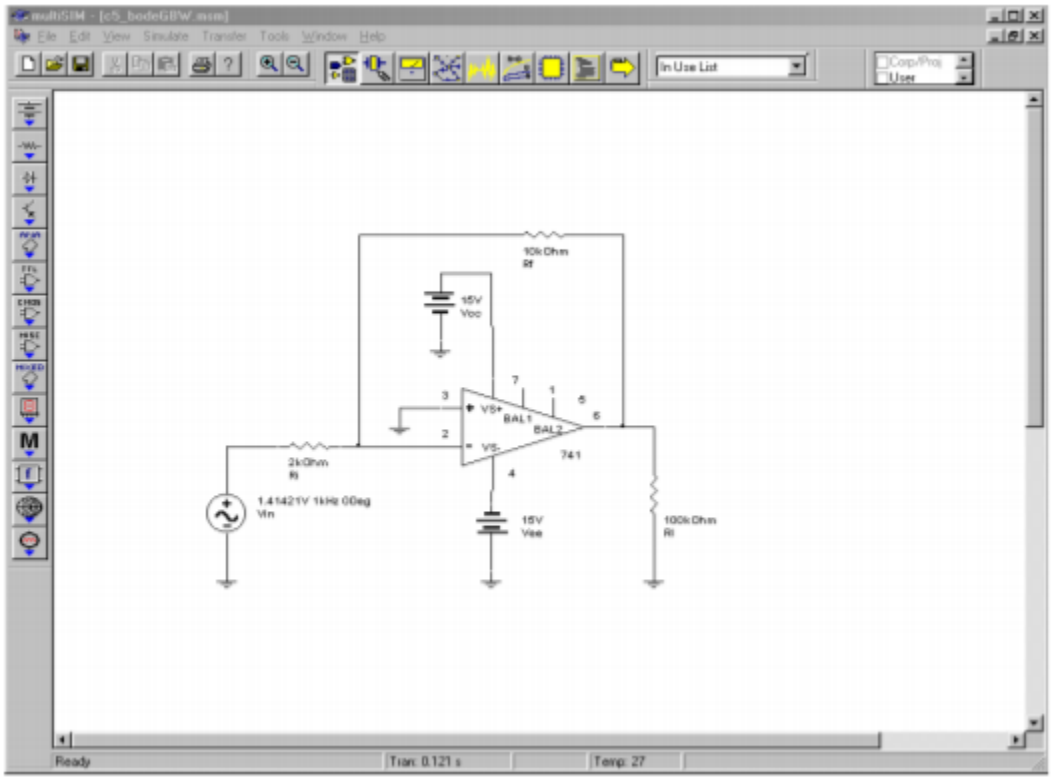


Figure 5.3.4 : Multisim schematic for Bode plot simulation.

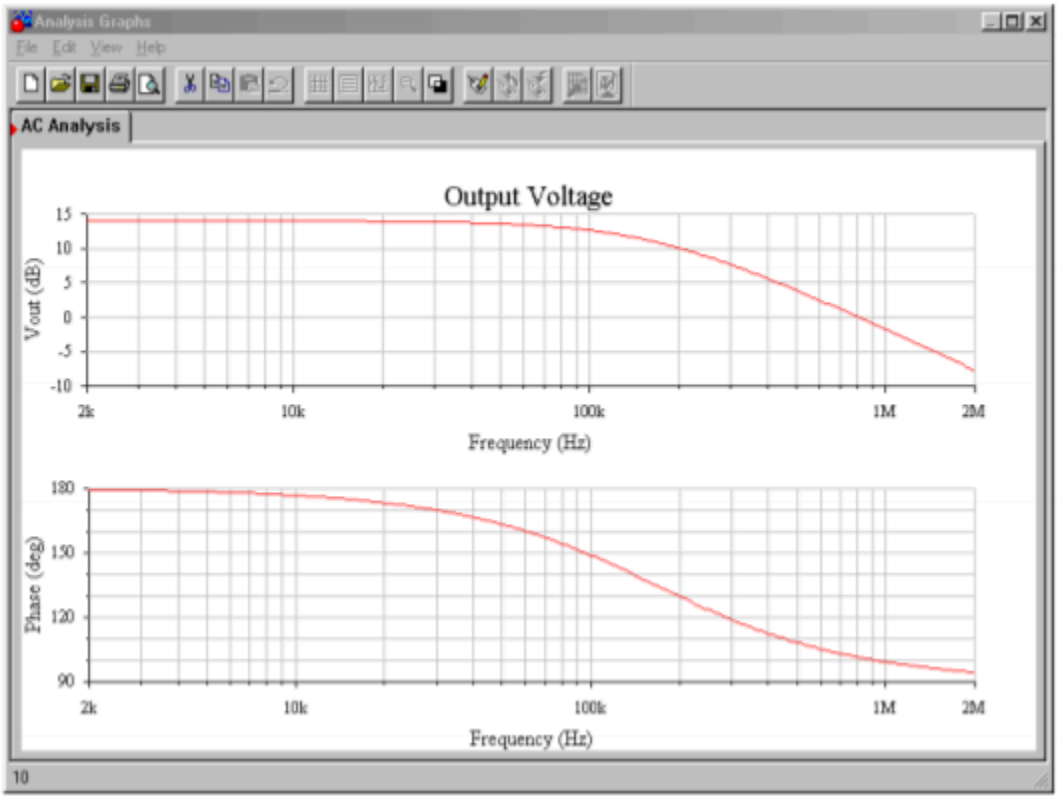


Figure 5.3.4 : Simulation of Bode plot.

It is important to note that the simple dependent source model presented in Chapter Two cannot be used because it does not have the proper frequency response. Instead, manufacturers offer accurate

models for their op amps. There are several variations on the theme. The model presented here is typical. It is fairly complex and is quite accurate. The op amp model is comprised of two basic parts, a differential amplifier input portion and a dependent source output section. The input portion utilizes a pair of NPN transistors with simple resistors for the loads (◈◈1 and ◈◈2). Resistors ◈◈1 and ◈◈2 serve as swamping or emitter-degeneration resistors. The tail-current source is set by the independent source ◈◈◈ . The non-ideal internal impedance and frequency limitations of this current source are taken into account by ◈◈ and ◈◈ , whereas ◈1 helps to model the high frequency loading of the diff amp's output. The output portion revolves around a series of voltage-controlled current sources. GCM models common mode gain, GA models the ordinary gain, and ◈2 serves as the combined internal impedance of these sources. ◈2 is the system compensation capacitor and has a value of 30 pF. ◈◈1 and ◈◈2 serve to model the output impedance of the op amp. Diodes ◈1 through ◈4 and voltage sources ◈◈ , ◈◈ , and ◈◈ model the limits of the op amp's class AB output stage.

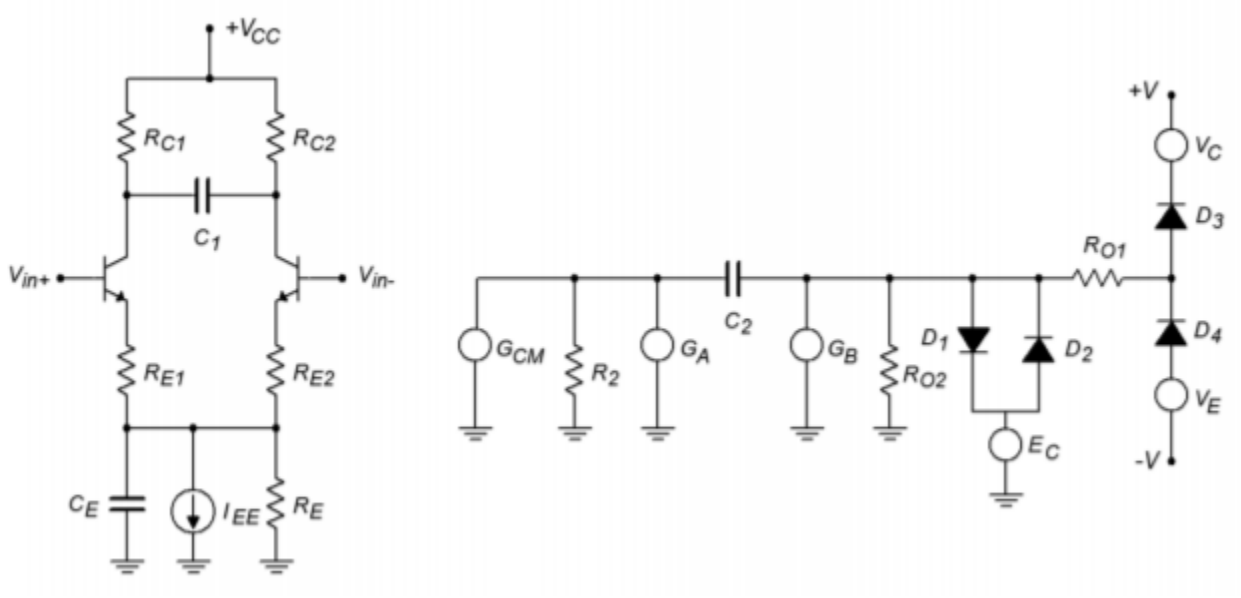


Figure 5.3.4◈: Typical op amp model.

In spite of their accuracy, models such as this are time consuming and tedious to recreate. Fortunately, many manufacturers offer simulation models for their components in library form. To use these models, all you need to do is reference the appropriate part number from the library. The op amp models normally use typical rather than worst-case values.

Example 5.3.3

Determine the minimum acceptable ◈◈◈◈◈◈ for the circuit of Figure 5.3.5 if response should extend to at least 50 kHz.

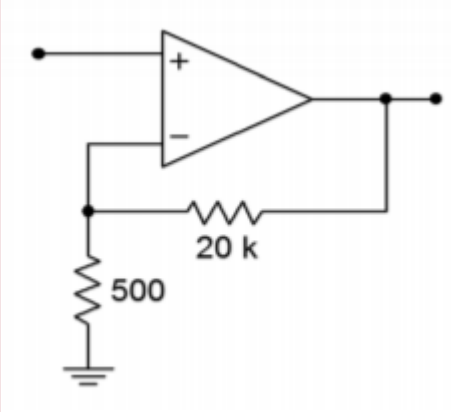


Figure 5.3.5 : Circuit for Example 5.3.3 .

$$A_{noise} = 1 + \frac{R_f}{R_i}$$

$$A_{noise} = 1 + \frac{20k}{500}$$

$$A_{noise} = 41$$

$$f_{unity} = A_{noise} f_2$$

$$f_{unity} = 41 \times 50kHz$$

$$f_{unity} = 2.05MHz$$

For this application, a stock 741 would not be fast enough; however, a 411 would be fine. From the foregoing, it is apparent that there is a direct trade off between circuit gain and high frequency performance for a given device. For an application requiring both high gain and wide bandwidth, a multistage approach should be considered.

MULTI-STAGE CONSIDERATIONS

By combining two or more wide-bandwidth, low-gain stages, a single high-gain, wide-bandwidth system may be produced. Although the overall system gain will simply be the combination of the individual stage gains, the upper break frequency calculation can be a little tricky. Chances are, in a multi-stage op amp design all stages will not exhibit the same upper break frequency. In this case the system's upper break is approximately equal to the lowest of the stage [illegible] . In other words, the system is treated as though it was a discrete stage with multiple lag networks. On the other hand, if the break frequencies are close, this approximation can lead to a sizable error. This is best illustrated with a quick example. Imagine two stages exhibiting a 100 kHz break. If each stage produces a 3 dB loss at 100 kHz, it is obvious that the cascaded system must be producing a 6 dB loss at 100 kHz. Therefore, the system's critical frequency (i.e., -3 dB point) must be somewhat lower than 100 kHz (to be exact, it is the frequency at which each stage produces a 1.5 dB loss). Taking this a step further, if we cascade three identical stages, the total loss at 100 kHz will be 9 dB. The system break will be the point at which each of the three stages produces a 1 dB loss. The more identical stages that are

added, the lower the effective break becomes. If we make a few assumptions about the exact shape of the rolloff curve, we can reduce this to a simple equation. In Chapter One, we derived the general Equation describing the amplitude response of a lead network (1.3.3). In a similar vein, the response for a lag network may be determined to be

$$A_v = \frac{1}{\sqrt{1 + \frac{f^2}{f_c^2}}}$$

(5.3.2)

Where f is the frequency of interest and ◈꜀ is the critical frequency.

It is more convenient to write this Equation in terms of a normalized frequency of interest. Instead of being expressed in Hertz, the frequency of interest is represented as a factor relative to ◈꜀ . If we call this normalized frequency ◈◈ , we may rewrite the amplitude response equation.

$$A_v = \frac{1}{\sqrt{1 + k_n^2}}$$

5.3.3)

We now solve for k

$$\frac{1}{A_v} = \sqrt{1 + k_n^2}$$

$$k_n^2 + 1 = \frac{1}{A_v^2}$$

$$k_n^2 = \frac{1}{A_v^2} - 1$$

$$k_n = \sqrt{\frac{1}{A_v^2} - 1}$$

(5.3.4)

We will now find the gain contribution of each stage. If all stages are critical at the same frequency, each stage must produce the same gain as the other stages at any other frequency. Because the combined gain of all stages must, by definition, be -3 dB or 0.707 at the system's break frequency, we may find the gain of each stage at this new frequency.

$$A_v^n = 0.707$$

(5.3.5)

Where n is the number of stages involved.

We may rewrite this as

$$A_v = 0.707^{\frac{1}{n}}$$

(5.3.6)

Combining 5.3.6 with 5.3.4 yields

$$k_n = \sqrt{\frac{1}{(0.707^{\frac{1}{n}})^2} - 1}$$

$$k_n = \sqrt{2^{\frac{1}{n}} - 1}$$

(5.3.7)

As ◈◈ is nothing more than a factor, this may be rewritten into a final convenient form.

$$f_{2-system} = f_2 k_n$$

$$f_{2-system} = f_2 \sqrt{2^{\frac{1}{n}} - 1}$$

(5.3.8)

where ◈ is the number of identical stages.

Example 5.3.4

Assuming that all stages in Figure 5.3.6 use 741's, what is the system gain and upper break frequency?

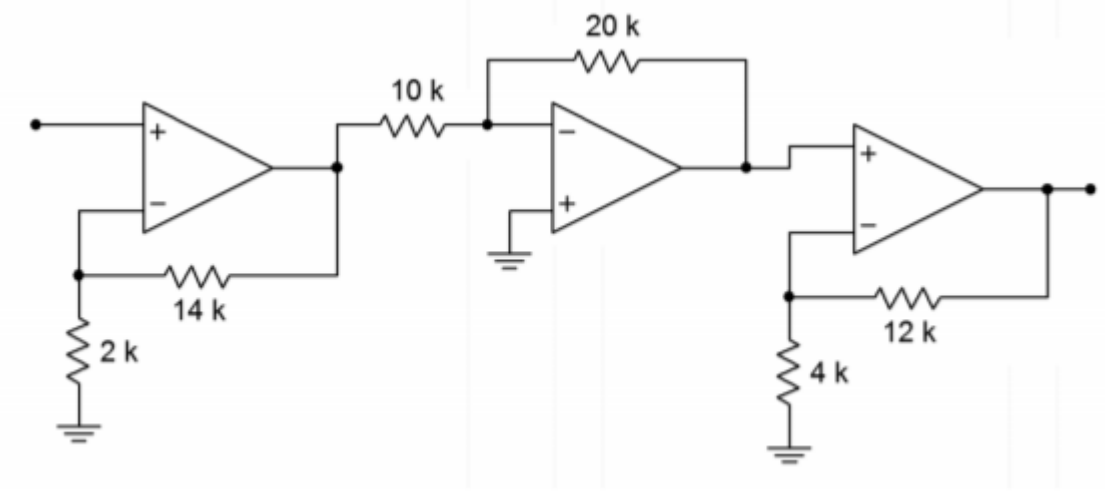


Figure 5.3.6 : Multistage circuit for Example 5.3.4 .

Stage 1:

$$A_v = 1 + \frac{R_f}{R_i}$$

$$A_v = 1 + \frac{14k}{2k}$$

$$A_v = 8$$

$$A_{noise} = 1 + \frac{R_f}{R_i}$$

$$A_{noise} = 8$$

$$f_2 = \frac{GBW}{A_{noise}}$$

$$f_2 = \frac{1MHz}{8}$$

$$f_2 = 125kHz$$

Stage 2:

$$A_v = -\frac{R_f}{R_i}$$

$$A_v = \frac{-20k}{10k}$$

$$A_v = -2$$

$$A_{noise} = 1 + \frac{R_f}{R_i}$$

$$A_{noise} = 3$$

$$f_2 = \frac{GBW}{A_{noise}}$$

$$f_2 = \frac{1MHz}{3}$$

$$f_2 = 333kHz$$

Stage 3:

$$A_v = -\frac{R_f}{R_i}$$

$$A_v = -\frac{12k}{4k}$$

$$A_v = 4$$

$$A_{noise} = -\frac{R_f}{R_i}$$

$$A_{noise} = 4$$

$$f_2 = \frac{GBW}{A_{noise}}$$

$$f_2 = \frac{1MHz}{4}$$

$$f_2 = 250kHz$$

System:

$$A_v = 8 \times (-2) \times 4$$

$$A_v = -64$$

◈2 = dominant stage. The dominant break here is 125 kHz (stage 1).

The system has a gain of 64 and an upper break of 125 kHz. If this level of performance is to be achieved with a single op amp, it would need a gainbandwidth product of 125 kHz times 64, or 8 MHz.

Example 5.3.5

A three-stage amplifier uses identical noninverting voltage stages with gains of 10 each. If the op amps used have an ◈◈◈◈◈◈ of 4 MHz, what is the system gain and upper break?

Because these are noninverting amplifiers, the noise gain equals the signal gain. The break frequency for each stage is:

$$f_2 = \frac{f_{unity}}{A_{noise}}$$

$$f_2 = \frac{4MHz}{10}$$

$$f_2 = 400kHz$$

Because the three stages are identical, the system will roll off before 400 kHz.

$$f_2 = \frac{f_{unity}}{A_{noise}}$$

$$f_2 = \frac{4MHz}{10}$$

$$f_2 = 400kHz$$

Note that the system response in this case is reduced about an octave from the single-stage response.

Example 5.3.6

Using only LF411 op amps, design a circuit with an upper break frequency of 500 kHz and a gain of 26 dB.

A gain of 26 dB translates to an ordinary gain of 20. Assuming a noninverting voltage stage, and noting that ◇◇◇◇◇◇=◇◇ for noninverting form, a single op amp would require an ◇◇◇◇◇◇ of:

$$f_{unity} = f_2 A_v \; f_{unity} = 500kHz \times 20$$

$$f_{unity} = 10MHz$$

Because the 411 has a typical ◇◇◇◇◇◇ of 4 MHz, at least two stages are required.

There are many possibilities. One option is to set one stage as the dominant stage and set its gain to produce the desired $◇_2$. The second stage will then be used to make up the difference in gain to the desired system gain.

Stage 1:

$$f_2 = \frac{f_{unity}}{A_v}$$

$$A_v = \frac{f_{unity}}{f_2}$$

$$A_v = \frac{4MHz}{500kHz}$$

$$A_v = 8$$

Stage 2: To achieve a final gain of 20, stage two requires a gain of 2.5. Its $◇_2$ is:

$$f_2 = \frac{f_{unity}}{A_v}$$

$$f_2 = \frac{4MHz}{2.5}$$

$$f_2 = 1.6MHz$$

Note that if this frequency worked out to less than 500 kHz, three or more stages would be needed. To set the resistor values, the rules of thumb presented in Chapter Four may be used. One possible solution is shown in Figure 5.3.7 .

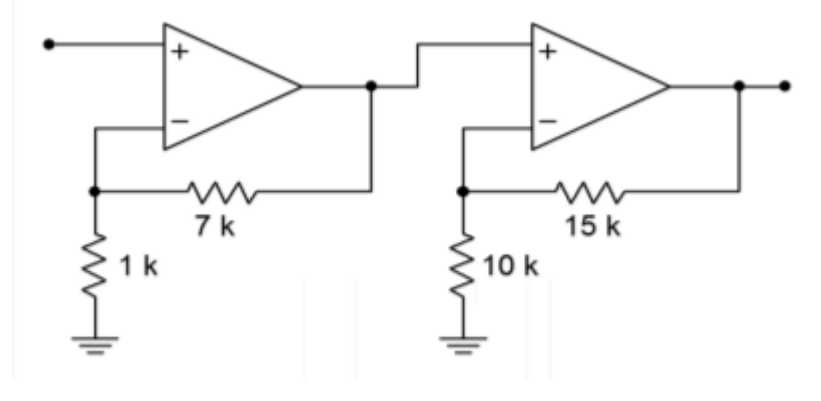


Figure 5.3.7 : Completed design for Example 5.3.6 .

LOW FREQUENCY LIMITATIONS

As mentioned earlier, standard op amps are direct-coupled. That is, their gain response extends down to 0 Hz. Consequently, many op amp circuits have no lower frequency limit. They will amplify DC signals just as easily as AC signals. Sometimes it is desirable to introduce a low frequency rolloff. Two cases of this are single-supply biasing (Chapter Four) and interference rejection (the removal of undesired signals, such as low frequency rumble). In both cases, the circuit designer produces a low frequency rolloff (lead network) by introducing coupling capacitors. For single-supply circuits, these capacitors are a necessary evil. Without them, stages would quickly overload from the large DC input. Also, signal sources and loads may be very intolerant of the DC bias potential. The result could be gross distortion or component failure. Even if a circuit uses a normal bipolar supply, a lead network may be used to reduce interference signals. For example, a well-chosen coupling capacitor can reduce 60 Hz hum interference while hardly affecting the quality of a voice transmission. Generally, these coupling capacitors can be simplified into the straightforward lead networks discussed back in Chapter One. (Remember, for lead networks, the highest critical frequency is the dominant one.) Also, if multiple networks are dominant, the resulting critical frequency will be higher than the individual break frequency. The relationship is the mirror image of Equation 5.3.8 . The proof for the following Equation is very similar to that of Equation 5.3.8 , and is left as an exercise.

$$f_{1-system} = \frac{f_1}{\sqrt{2^{\frac{1}{n}} - 1}}$$

(5.3.9)

If you decide to add coupling capacitors in order to reduce interference, remember that the op amp will need a DC return resistor. An example is shown in Figure 5.3.8 . The 100 k Ω resistor is needed so that the inverting input's half of the diff amp stage is properly biased. Note that for a typical op amp, this 100 k Ω also ends up setting the input impedance. Assuming a relatively low source impedance, the lead network simplification boils down to the 0.1 ◈ F capacitor along with the 100 k Ω . The critical frequency is:

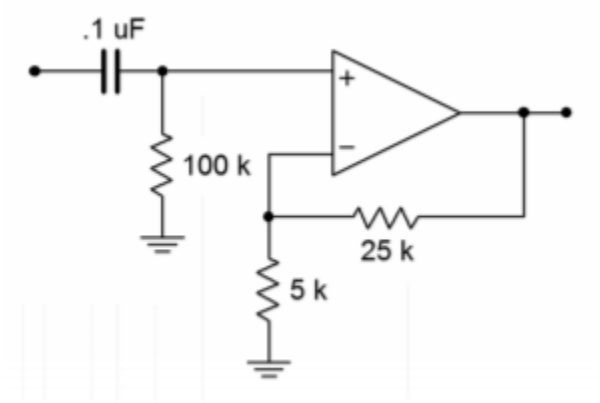


Figure 5.3.8 : DC return resistor (100 k).

$$f_c = \frac{1}{2\pi RC}$$

$$f_c = \frac{1}{2\pi \times 100k \times 0.1\mu F}$$

$$f_c = 15.9Hz$$

To sum up, then, when using general-purpose op amps, if no signal-coupling capacitors are being used, the gain response extends back to 0 Hz. If coupling capacitors are used, general lead network analysis techniques can be used to find the critical frequencies.

11.4 SLEW RATE AND POWER BANDWIDTH

As noted in the previous section, general-purpose op amps contain a compensation capacitor that is used to control the open loop frequency response. The signal developed across this capacitor will be amplified in order to create the final output signal. In essence, this capacitor serves as the load for the preceding stage inside of the op amp. Like all stages, this one has a finite current output capability. Due to this, the compensation capacitor can be charged no faster than a rate determined by the standard capacitor charge equation:

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

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

The rate of change of voltage versus time is ◈◈/◈◈ . By definition, this parameter is called slew rate (SR). The base unit for slew rate is volts per second, however, given the speed of typical devices, slew rate is normally specified in volts per microsecond. Slew rate is very important in that it helps determine whether or not a circuit can accurately amplify high-frequency or pulse-type waveforms. In order to create a fast op amp, either the charging current ◈ must be large, or the compensation capacitor ◈ must be very small. Because ◈ also plays a role in determining the gain bandwidth product, there is a lower limit to its size. A typical op amp might use a 30 pF compensation capacitor, and the driving stage may effectively produce a charging current of 100 ◈ A. The resulting slew rate would be:

$$SR = \frac{dv}{dt} = \frac{i}{C}$$

$$SR = \frac{100\mu A}{30pF}$$

$$SR = 3.33 \text{ Megavolts/second}$$

$$SR = 3.33V/\mu s$$

This means that the output of the op amp can change no faster than 3.33 V over the course of one microsecond. It would take this op amp about 3 microseconds for its output signal to change a total of 10 V. It can go no faster than this. The ideal op amp would have an infinite slew rate. Although this is a practical impossibility, it is possible to find special high-speed devices that exhibit slew rates in the range of several thousand volts per microsecond. Comparative slew rates for a few selected devices are found in Table 11.4.1 .

Device	Slew Rate
uA741	0.5 V/ ◈ s
LF411	15 V/ ◈ s
OPA134	20 V/ ◈ s
LM318	70 V/ ◈ s
LM6364	300 V/ ◈ s
LT1363	1000 V/ ◈ s

Table 11.4.1

Slew rate is always output-referred. This way, the circuit gain need not be taken into account. Slew rate is normally the same regardless of whether the signal is positive or negative going. There are a few devices that exhibit an asymmetrical slew rate. One example is the 3900. It has a slew rate of 0.5 V/ ◈ s for positive swings, but shows 20 V/ ◈ s for negative swings.

THE EFFECT OF SLEW RATE ON PULSE SIGNALS

An ideal pulse waveform will shift from one level to the other instantaneously, as shown in Figure 5.4.1 .

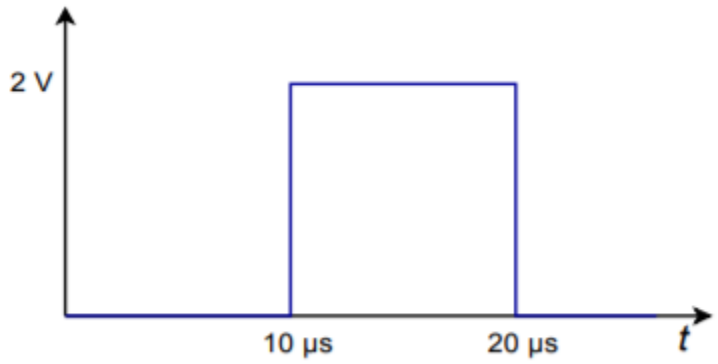


Figure 5.4.1 : Ideal output.

In reality, the rising and falling edges are limited by the slew rate. If this signal is fed into a 741 op amp, the output pulse would be decidedly trapezoidal, as shown in Figure 5.4.2 . The 741 has a slew rate of 0.5 V/ ◈ s. Because the voltage change is 2 V, it takes the 741 4 ◈ s to traverse from low to high, or from high to low. The resulting waveform is still recognizable as a pulse, however.

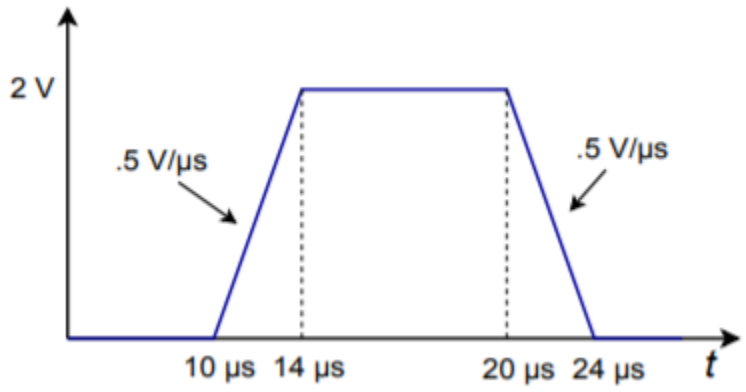


Figure 5.4.2 : Slewed output of the 741.

Gross distortion of the pulse occurs if the pulse width is decreased, as in Figure 5.4.3 . Here, the pulse width is only 3 microseconds, so the 741 doesn't even have enough time to reach the high level. In 3 microseconds, the 741 can only change 1.5 V. By the time the 741 gets to 1.5 V, the input signal is already swinging low, so the 741 attempts to track it. The result is a triangular waveform of reduced amplitude as shown in Figure 5.4.4 . This same effect can occur if the amplitude of the pulse is increased. Obviously, then, pulses that are both fast and large require high slew rate devices. Note that a 411 op amp would produce a nice output in this example. Because its slew rate is 15 V/ ◈ s, it requires only 0.134 microseconds for the 2 V output swing. Its output waveform is shown in Figure 5.4.5 .

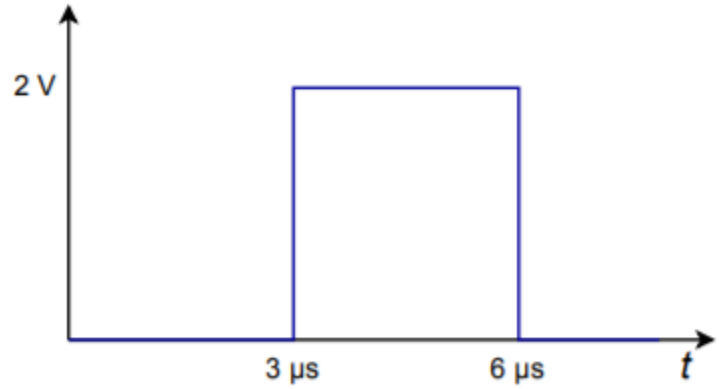


Figure 5.4.3 : Ideal output.

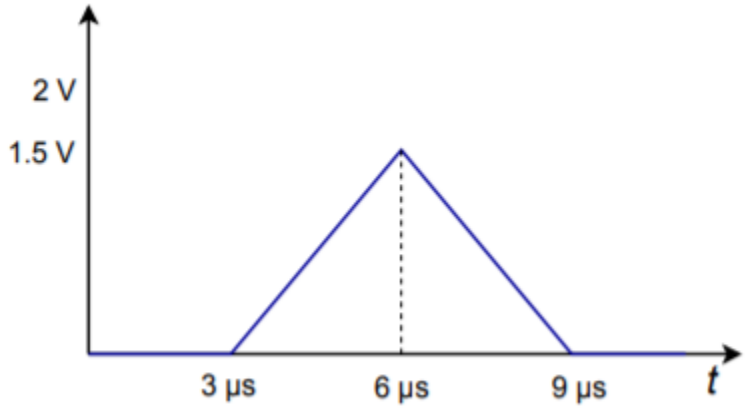


Figure 5.4.4 : Slewed output of the 741.

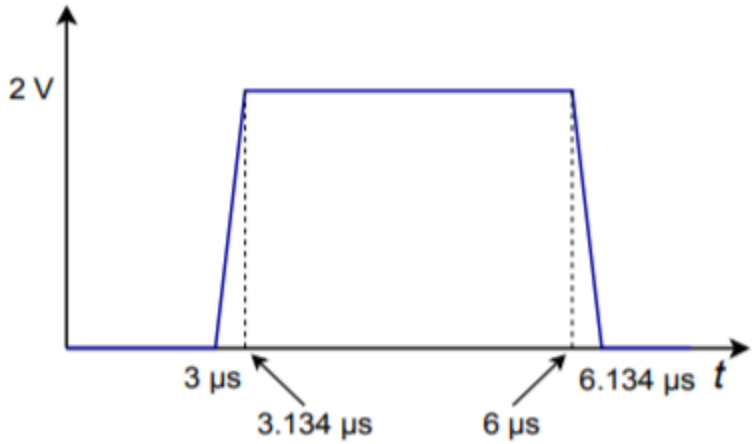


Figure 5.4.5 : Slewed output of the 411.

THE EFFECT OF SLEW RATE ON SINUSOIDAL SIGNALS AND POWER BANDWIDTH

Slew rate limiting produces an obvious effect on pulse signals. Slew rate limiting can also affect sinusoidal signals. All that is required for slewing to take place is that the signal change faster than the device's slew rate. If the rate of change of the signal is never greater than the slew rate, slewing will never occur. To find out just how fast a given sine wave does change, we need to find the first derivative with respect to time. Assume that the input sine wave has a frequency , and a peak amplitude .

$$v(t) = K \sin 2\pi f t$$

$$\frac{dv}{dt} = 2\pi f K \cos 2\pi f t$$

The rate of change of the signal with respect to time is . The maximum rate of change will occur when the sine wave passes through zero (i.e., at t = 0). To find this maximum value, substitute 0 in for t, and solve the equation.

$$\frac{dv}{dt} = 2\pi f K$$

(5.4.1)

So, the rate of change of the signal is directly proportional to the signal's frequency (◈), and its amplitude (◈). From this, it is apparent that high-amplitude, high-frequency signals require high slew rate op amps in order to prevent slewing. We can rewrite our Equation in a more convenient form:

$$\text{Slew Rate required } = 2\pi V_p f_{max}$$

(5.4.2)

where ◈◈ is the peak voltage swing required and ◈◈◈◈ is the highest frequency sine wave reproduced. Often, it is desirable to know just how "fast" a given op amp is. A further rearranging yields

$$f_{max} = \frac{\text{Slew Rate}}{2\pi V_p}$$

In this case, ◈◈◈◈ represents the highest frequency sine wave that the op amp can reproduce without producing Slewing Induced Distortion (SID). This frequency is commonly referred to as the power bandwidth. To be on the conservative side, set ◈◈ to the op amp's clipping level. Note that slew rate calculations are not dependent on either the circuit gain or small-signal bandwidth. Power bandwidth and small signal bandwidth ($◈_2$) are not the same thing. This is a very important point!

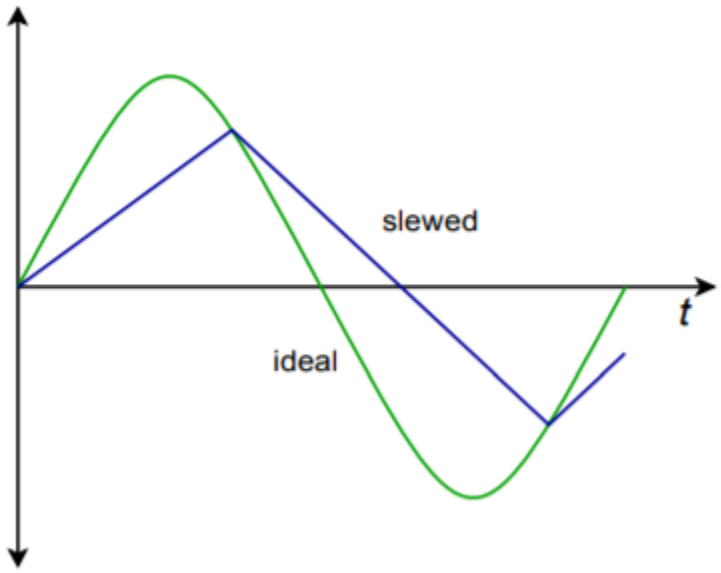


Figure 5.4.6 : Sine wave distorted by heavy slewing.

The effects of slewing can be either subtle or dramatic. Small amounts of SID are very difficult to see directly on an oscilloscope and require the use of a distortion analyzer or a spectrum analyzer for verification. Heavy slewing will turn a sine wave into a triangular wave. An example of this is shown in Figure 5.4.6 .

Example 5.4.1

A 741 is used as part of a motor control system. If the highest reproducible frequency is 3 kHz and the maximum output level is 12 V peak, does slewing ever occur?

Another way of stating the problem is to ask "Is the 741's power bandwidth at least 3 kHz?"

$$f_{max} = \frac{SlewRate}{2\pi V_p}$$

$$f_{max} = \frac{0.5V/\mu s}{2\pi \times 12V}$$

$$f_{max} = \frac{0.5MV/s}{\pi \times 24V}$$

$$f_{max} = 6631Hz$$

For this application, the 741 is twice as fast as it needs to be. Note in the calculation how the slew rate is transferred from V/ ◈ s into V/s, and how the volts units cancel between denominator and numerator. This leaves units of "1/seconds", which is another way of saying "Hertz". If the calculation produced a smaller value, say 2 kHz, then slewing is a possibility for certain signals.

Example 5.4.2

An audio pre-amplifier needs to reproduce signals as high as 20 kHz. The maximum output swing is 10 V peak. What is the minimum acceptable slew rate for the op amp used?

$$\text{Slew Rate} = 2\pi V_p f_{max}$$

$$\text{Slew Rate} = 2\pi \times 10V \times 20kHz$$

$$\text{Slew Rate} = 1.257MV/s$$

$$\text{Slew Rate} = 1.257V/\mu s$$

For this design, a 741 would not be fast enough. The aforementioned 411 through 318 would certainly be satisfactory, whereas the 1363 would probably be overkill.

COMPUTER SIMULATION

To verify the results of Example 5.4.2 , a simple noninverting voltage amplifier may be used with differing op amp models. The simulation is shown Figure 5.4.7 using Multisim.

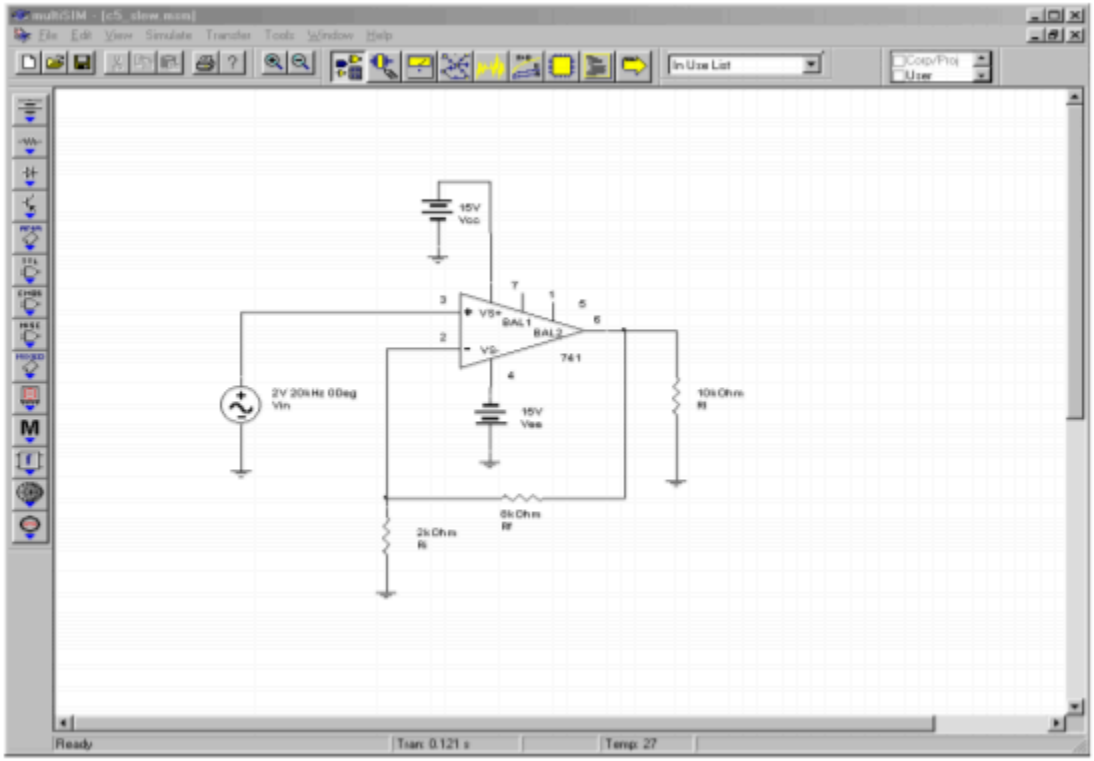


Figure 5.4.7 : Multisim schematic for slew simulation.

The circuit is configured with a 2 volt input at 20 kHz and a gain of 5. This will yield the worst case output of 10 volts at 20 kHz. For the first Transient Analysis, a 741 is used. Note how the output waveform is essentially triangular. It is also below the expected peak output level. Clearly, this waveform is severely slewed, and results in undesired distortion and a reduction in audio quality.

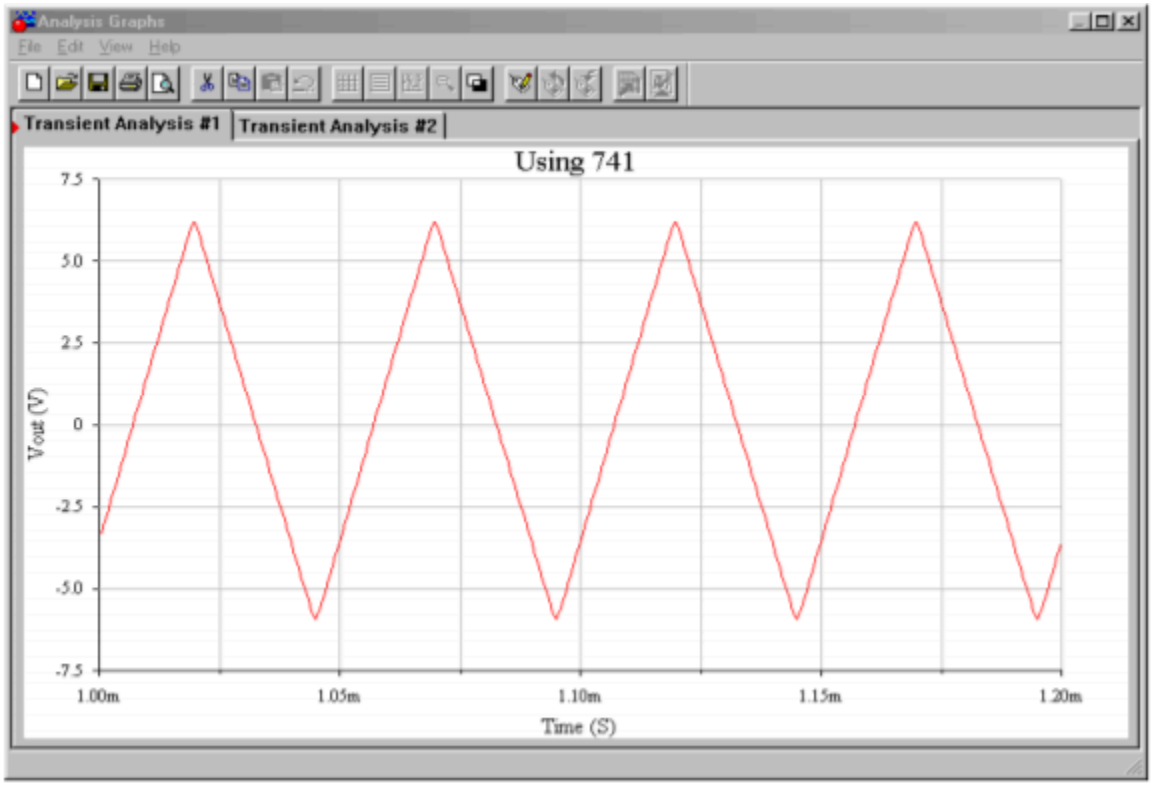


Figure 5.4.7 : Output with 741.

The second simulation is performed using the faster LF411. In this case, the simulation shows a full 10 volt peak output with no discernable distortion. The LF411 would certainly meet the circuit requirements.

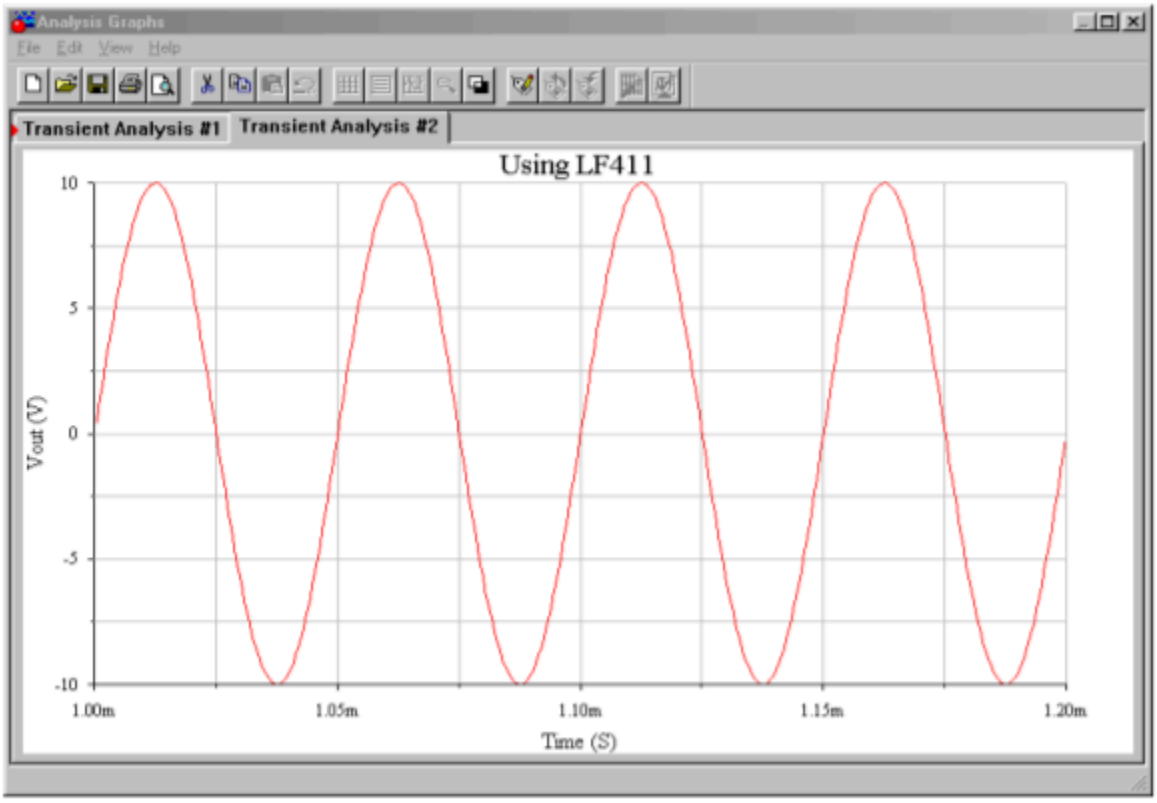


Figure 5.4.7 : Output with LF411.

DESIGN HINT

There is a convenient way of graphically determining whether the output of an op amp will be distorted. It involves graphing output levels versus frequency. The two major distortion causes are clipping and slewing. We start with a grid measuring frequency on the horizontal axis, and output voltage on the vertical axis. The first step involves plotting the output level limit imposed by clipping. Clipping is dependent on the circuit's power supply and is independent of frequency. Therefore, a horizontal line is drawn across the graph at the clipping level (see Figure 5.4.8). If we assume a standard ± 15 V power supply, this level will be around ± 13 V. The output level cannot swing above this line because clipping will be the result. Everything below this line represents unclipped signals. The second step is to plot the slewing line. To do this, a point needs to be calculated for . In Example 5.4.1 a 741 was used and 12 V produced an fmax of 6631 Hz. Plot this point on the graph. Now, as the slew rate is directly proportional to and , it follows that doubling while halving results in the same slew rate. This new point lets you graphically determine the slope of the slew limiting line. Plot this new point and connect the two points with a straight line (see Figure 5.4.9). Everything above this line represents slewed signals, and everything below this line represents non-slewed signals. As long as the desired output signal falls within the lower intersection area of the two lines, the signal will not experience either slewing or clipping. A quick glance at the graph allows you to tell what forms of distortion may affect a given signal.

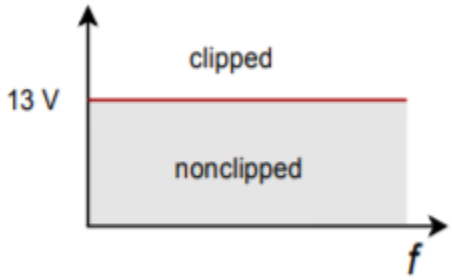


Figure 5.4.8 : Limit due to clipping.

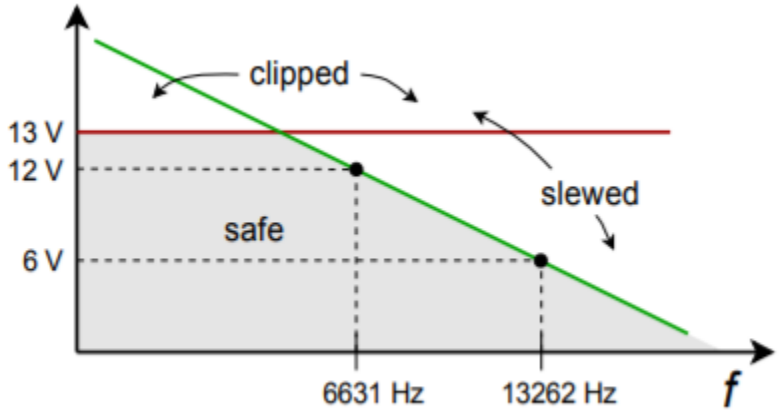


Figure 5.4.9 : Limits due to clipping and slewing.

SLEW RATE AND MULTIPLE STAGES

Consider the three-stage circuit shown in Figure 5.4.10 . The slew rates for each device are found in Table 5.4.1 . What is the effective slew rate of the system? You might think that it is set by the slowest device (741 at 0.5 V/ ⍰ s), or perhaps by the final device (318 at 70 V/ ⍰ s). The fact is that the system slew rate could be set by any of the devices, and it depends on the gains of the stages.

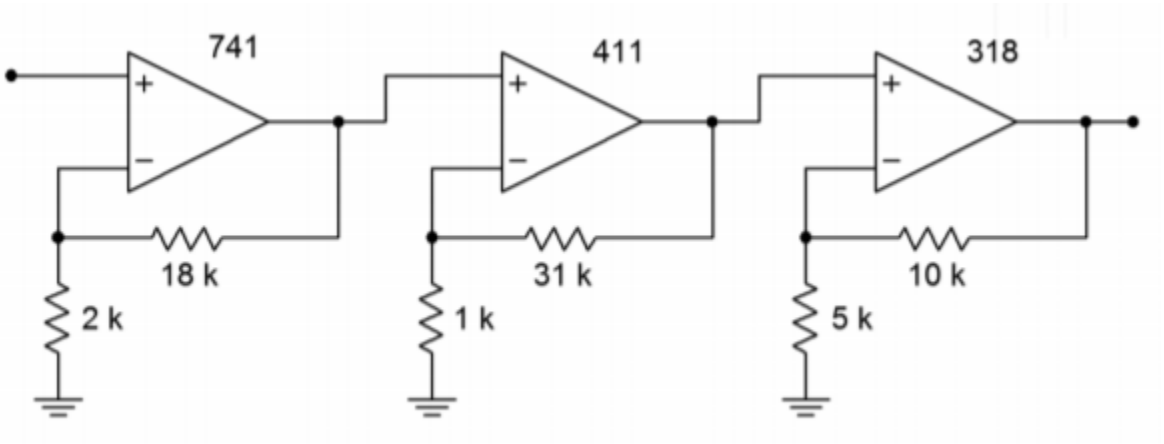


Figure 5.4.10 : Multistage circuit incorporating the 741, 411 and 318.

The one thing that you can say immediately is that the system slew rate will never be faster than the

final device. In this example, the slew rate cannot be greater than 70 V/ ◇ s. It may be less than this, however. The trick to finding the effective system slew rate is to start at the output of the first stage, and then determine the maximum rate of change for the following stages in sequence. Looking at stage 1, its maximum output rate is 0.5 V/ ◇ s. This is the maximum rate of change going into stage 2. Because stage 2 has a gain of 32, it will attempt to increase this rate to 16 V/ ◇ s. This cannot happen, however, because the 411 has a slew rate of only 15 V/ ◇ s. Therefore the 411 is the limiting factor at this point. The maximum rate of change out of stage 2 is 15 V/ ◇ s. This signal is then applied to stage 3, which has a gain of 3. So, the 318 triples its input signal to 45 V/ ◇ s. Because the 318 is capable of changing as fast as 70 V/ ◇ s, 45 V/ ◇ s becomes the limiting output factor. The system slew rate is 45 V/ ◇ s. This is the value used to calculate the system power bandwidth, if needed. The first op amp to slew in this circuit is the 411, even though it is about 30 times faster than the 741 used in stage 1. The reason for this is that it must handle signals 32 times as large. Note that if the final stage had a larger gain, say 5, the 318 would become the limiting factor. The important thing to remember is that the front end stages of a system don't need to be as fast as the final stages, as they handle smaller signals.

NONCOMPENSATED DEVICES

As discussed in Chapter Three, all op amps need some form of frequency compensation in order to ensure that their closed loop response is stable. The most straightforward way to do this is to add a compensation capacitor, which forces a 20 dB/decade rolloff to ◇◇◇◇◇◇ . In this way, no matter what gain you choose, the circuit will be stable. Although this is very convenient, it is not the most efficient form of compensation for every circuit. High-gain circuits need less compensation capacitance than a low-gain circuit does. The advantage of using a smaller compensation capacitor is that slew rate is increased. Also, available loop gain at higher frequencies is increased. This allows the resulting circuit to have a wider small-signal bandwidth (the effect is as if ◇◇◇◇◇◇ increased). Therefore, if you are designing a high gain circuit, you are not producing the maximum slew rate and small-signal bandwidth that you might. The compensation capacitor is large enough to achieve unity gain stability, but your circuit is a high-gain design. The bottom line is that you are "paying" for unity gain stability with slew rate and bandwidth.

To get around this, manufacturers offer noncompensated op amps. No internal capacitor is used. Instead, connections are brought out to the IC package so that you may add your own capacitor. This way, the op amp may be tailored to your application. There is no set way of determining the values for the external compensation circuit (it may be more complex than a single capacitor). Compensation details are given on manufacturers data sheets. One example of a noncompensated op amp is the 301. You can think of a 301 as a 741 without a compensation capacitor. If a 33 pF compensation capacitor is used, the 301 will be unity gain stable and produce an ◇◇◇◇◇◇ of 1 MHz and a slew rate of 0.5 V/ ◇ s. For higher gains, a smaller capacitor may be used. A 10 pF unit will produce an effective ◇◇◇◇◇◇ of 3 MHz, and a slew rate of 1.5 V/ ◇ s. Comparative Bode gain plots are shown in Figure 5.4.11 for the 301.

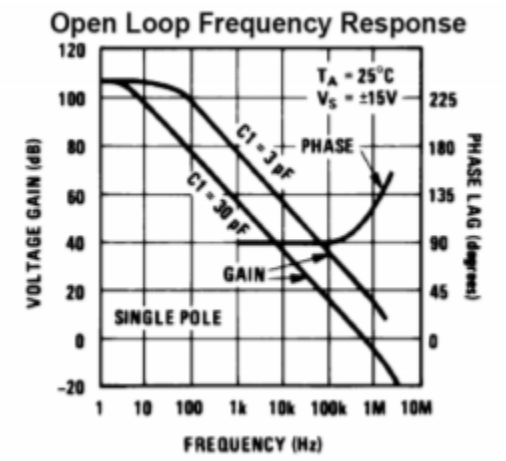


Figure 5.4.11 : Bode gain plots for the 301. Reprinted courtesy of Texas Instrutments

FEEDFORWARD COMPENSATION

Besides the ordinary form of compensation, noncompensated devices like the 301 may utilize other methods that can make them even faster, such as feedforward compensation. The concept of feedforward is in direct contrast to feedback. As its name suggests, feedforward involves adding a portion of the input signal to the output, thus bypassing certain sections of the system. Compared to ordinary feedback, feedforward is seldom used as part of the design of an amplifier.[1]

Ordinarily, a compensation capacitor must be large enough to maintain sufficient gain and phase margin for the slowest stage inside of an op amp. Quite often, the slowest stage in an op amp is one of the first stages, such as a level shifter. Here is where feedforward comes into play. If the high frequency content of the input signal can be shunted around this slow stage, the effective bandwidth and speed of the op amp may be increased. This is the key behind the technique of feedforward compensation. Normally, manufacturers will provide details of specific feedforward realizations for their op amps. (It would be very difficult to create successful feedforward designs without detailed knowledge of the internal design of the specific op amp being used.) Not all op amps lend themselves to feedforward techniques.

In summary, noncompensated op amps require a bit more work to configure than fully compensated devices, but offer higher performance. This means faster slew rates and higher-upper break frequencies.

DECOMPENSATED DEVICES

Straddling the worlds of compensated and uncompensated op amps is the decompensated device. Decompensated devices are also known as partially compensated devices. They include some compensation capacitance, but not enough to make them unity gain stable. Usually these devices are stable for gains above 3 to 5. Because the majority of applications require gains in this area or above,

1. It can offer similar advantages though, such as a reduction in distortion. Also, feedforward and feedback techniques may be combined in order to achieve complementary increases in performance. For an example, see M.J.Hawksford, "Reduction of Transistor Slope Impedance Dependent Distortion in Large Signal Amplifiers",Journal of the Audio Engineering Society, Vol.36 No.4 (1988): 213–222

decompensated devices offer the ease of use of compensated devices and the increased performance of customized noncompensated units. If required, extra capacitance can usually be added to make the circuit unity gain stable. One example of this type is the 5534. Its gain Bode plot is shown in Figure 5.4.12 . Note how the addition of an extra 22 pF reduces the open loop gain. This 22 pF is enough to make the device unity gain stable. It also has a dramatic effect on the slew rate, as seen in the 5534's spec sheet. Without the 22 pF capacitor, the slew rate is 13 V/ ◈ s, but with it, the slew rate drops to 6 V/ ◈ s. The performance that you give up in order to achieve unity gain stability is obvious here.

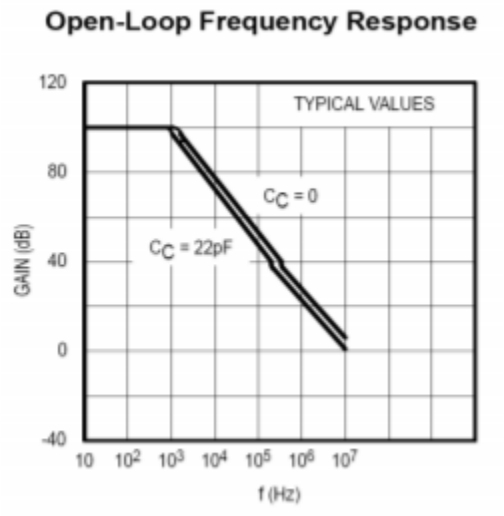


Figure 5.4.12 : Bode gain plot for the 5534. Reprinted courtesy of Philips Semiconductors

There is one more interesting item to note about the 5534. A close look at the Bode gain curve shows a hump in the rolloff region. Some other devices, such as the 318, exhibit this hump too. This is a nice extra. In essence, the manufacturer has been able to achieve a slightly higher open loop gain than a normal device would allow. This means that your circuits will have higher loop gains, and therefore, the nice effects of negative feedback will remain active to higher frequencies. This also means that very high gain circuits will be able to achieve a higher ◈$_2$ than is predicted by the gain-bandwidth calculation.

11.5 OFFSETS

Offsets are undesirable DC levels appearing at the output of a circuit. If op amps were perfect, there would be no such thing as an offset. Even though part matching is very close when ICs are made, the parts will not be identical. One possible example is the fact that the transistors used for the differential amplifier stage will not have identical characteristics. Because of this, their DC bias points are slightly different. This difference, or unbalance, is amplified by the remaining stages and will eventually produce a DC voltage at the output. Because all op amps are slightly different, you never know what the exact output offset will be. For measurement applications, this offset creates uncertainty in readings. For example, if the circuit output measures 100 mV, the signal might be 99 mV with 1 mV of offset. It might also be 101 mV with -1mV offset. In other applications, offsets can harm following stages or loads. Dynamic loudspeakers and headphones are two loads that should not be fed DC signals. This will reduce their maximum volume and increase their distortion. In short, offsets are not desired. Let's see what the causes are and how we can reduce or eliminate their effect.

OFFSET SOURCES AND COMPENSATION

For bipolar input sections the major cause of input current mismatch is the variation of beta. Base-emitter junction voltage variation is the major cause of input voltage deviation. For field effect devices, current variation is much less of a problem as the magnitude of input current is very low to begin with. Unfortunately, FETs do suffer from larger input voltage variations due to transconductance curve mismatches.

As mentioned in Chapter Two, the input current into the bases (or gates, in the case of an FET) of the first stage is called ◇◇ , the input bias current. In reality, this is an average of the two input currents, ◇◇+ and ◇◇– . The magnitude of their difference is called the input offset current, ◇◇◇ (some manufacturers use the symbol ◇◇◇). Note that the actual direction of ◇◇ is normally not specified, but can usually be determined from the manufacturer's circuit diagram. ◇◇ flows into the op amp if the input devices are NPN, and out of the op amp if the input devices are PNP.

$$I_B = \frac{I_{B+} + I_{B-}}{2}$$

$$I_{OS} = |I_{B+} - I_{B-}|$$

(5.5.1)

The voltage difference for the input stage is referred to as the input offset voltage, ◇◇◇ (some manufacturers use the symbol ◇◇◇). This is the potential required between the two inputs to null the output, that is, to re-align the output to 0 V DC. Both ◇◇◇ and ◇◇◇ are available on data sheets. The absolute magnitude of these offsets generally gets worse at temperature extremes. Table 5.5.1 shows some typical values. Note the low ◇◇ and ◇◇◇ values for the FET input 411.

Device	
◇◇	
◇◇◇	
◇◇◇	
5534	
800 nA	
10 nA	
0.5 mV	
411	
50 pA	
25 pA	
0.8 mV	
318	
150 nA	
30 nA	
4 mV	
741	
80 nA	
20 nA	
1 mV	

Table 11.5.1

Remember, these numbers are absolutes, so when ◇◇◇ is specified as 10 nA, it means that the actual ◇◇◇ can be anywhere between -10 nA and +10 nA. ◇◇ , ◇◇◇ and ◇◇◇ combine with other circuit elements to produce an output offset voltage. As this is a linear circuit, superposition may be used to separately calculate their effects. The model in Figure 5.5.1 will be used. ◇◇ and ◇◇ are the standard feedback components, and ◇◇◇◇ is called the offset compensation resistor (in some cases it may be zero). Because the input signal is grounded, this model is valid for both inverting and noninverting amplifiers.

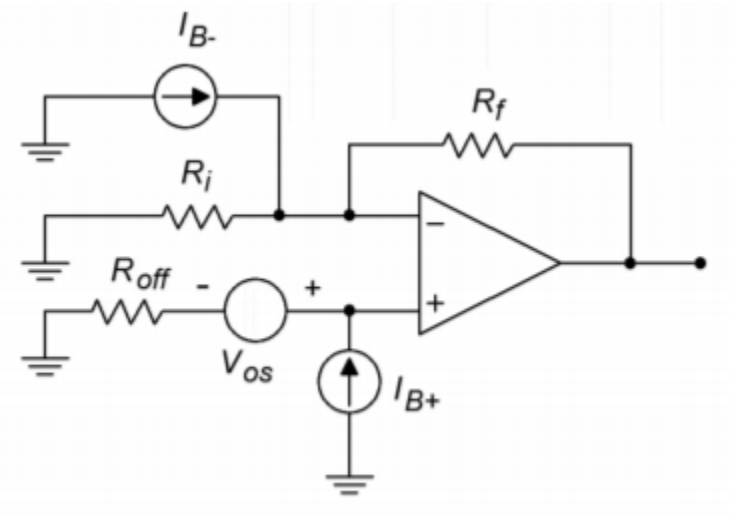


Figure 5.5.1 : Offset model.

◈◈◈ is seen as a small input voltage and is multiplied by the circuit's noise gain in order to find its contribution to the output offset. Offsets are by nature DC, so it is important to use the DC noise gain. Consequently, any capacitors found within the feedback loop should be mathematically "opened" for this calculation (for example, when working with the filter circuits presented in Chapter Eleven)

$$V_{outoffset1} = A_{noise} V_{OS}$$

(5.5.2)

where

$$A_{noise} = 1 + \frac{R_f}{R_i} = \frac{R_i + R_f}{R_i}$$

◈◈ and ◈◈◈ pass through input and feedback resistors to produce their output contributions. First, consider the effect of ◈◈+ . This creates a voltage across ◈◈◈◈ . This voltage is then multiplied by the circuit noise gain to yield its portion of the output offset.

$$V_{out-offset2} = I_{B+} R_{off} A_{noise}$$

(5.5.3)

For ◈◈– , recall that the inverting input is at virtual ground. This implies that the voltage across ◈◈ must be zero, and therefore, the current through ◈◈ must be zero. Consequently, all of ◈◈– flows through ◈◈ . This creates a relative negative potential at the output.

$$V_{out-offset3} = I_{B-} R_f$$

(5.5.4)

So the combination of the input bias current effects is:

$$V_{out-offset(I_B)} = |I_{B+} R_{off} A_{noise} - I_{B-} R_f|$$

(5.5.5)

Expanding this produces

$$V_{out-offset(I_B)} = I_{B+} R_{off} \frac{R_i + R_f}{R_i} - I_{B-} R_f$$

$$V_{out-offset(I_B)} = \left(I_{B+} R_{off} \frac{R_i + R_f}{R_i R_f} I_{B-} \right) R_f$$

By noting the product-sum rule for resistor combination ◆◆ , ◆◆ , this can be further simplified to

$$V_{out-offset(I_B)} = \left(\frac{I_{B+} R_{off}}{R_i || R_f} I_{B-} \right) R_f$$

(5.5.6)

If ◆◆◆◆ is set to equal ◆◆||◆◆ , this reduces to:

$$V_{out-offset(I_B)} = (I_{B+} - I_{B-}) R_f$$

By definition,

$$I_{OS} = |I_{B+} - I_{B-}|$$

so we finally come to

$$V_{out-offset(I_B)} = I_{OS} R_f$$

(5.5.7)

If it is possible, ◆◆◆◆ should be set to ◆◆||◆◆ . This drastically reduces the effect of the input bias current. Note that the value of ◆◆◆◆ includes the driving source internal resistance. If ◆◆||◆◆=2◆Ω and the driving source resistance is 100 Ω , the required resistance value would be 1.9 k Ω . If setting ◆◆◆◆ to the optimum value is not possible, you can at least reduce the effect of ◆◆ by using a partial value. Also, note that it is possible to determine the polarity of the offset caused by ◆◆+ and ◆◆– (Equation 5.5.6) if actual currents and the type of device used in the diff amp stage are known. The circuit of Figure 5.5.1 assumes that NPN devices are being used, hence the currents are drawn as entering the op amp. PNP input devices would produce the opposite polarity. Typically, though, we don't have precise values for ◆◆+ and ◆◆– , and thus can only compute the worst-case magnitude.

For a final result, we may combine our components:

$$V_{out-offset} = V_{OS} A_{noise} + I_{OS} R_f$$

(5.5.8)

if ◆◆◆◆=◆◆||◆◆ , and

$$V_{out-offset} = V_{OS} A_{noise} + |I_{B+} R_{off} A_{noise} - I_B R_f|$$

(5.5.9)

if ◆◆◆◆≠◆◆||◆◆ .

There is one special case involving the selection of Roff, and that deals with a voltage follower. Normally for a follower, ◆◆=0Ω . What if the driving source resistance is perhaps 0 Ω ? The calculation would require an ◆◆◆◆ of 0 Ω , and thus a -50 Ω resistor to compensate for the source resistance. This is of course, impossible! To compensate for the 50 Ω source, use 50 Ω for ◆◆ . The circuit gain will still be unity, but ◆◆ will now be compensated for. This is shown in Figure 5.5.2 .

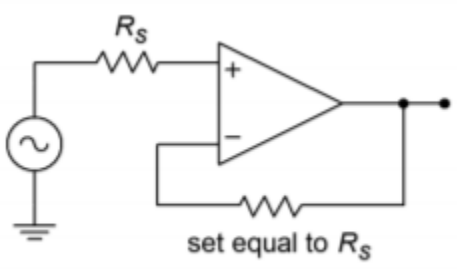


Figure 5.5.2 : Offset compensation for a follower.

Example 5.5.1

Determine the typical output offset voltage for the circuit of Figure 5.5.3 if ⓡⓞⓕⓕ is 0 Ω . Then determine an optimum size for ⓡⓞⓕⓕ and calculate the new offset.

From the data sheet for the 5534, we find ⓥⓞⓢ=0.5ⓜⓥ , ⓘⓑⓘ=10ⓝⓐ , and ⓐⓥ=800ⓥⓜ . Because this is an approximation, assume ⓘⓑ+=ⓘⓑ−=ⓘⓑ .

$$A_{noise} = 1 + \frac{R_f}{R_i}$$

$$A_{noise} = 1 + \frac{10k}{1k}$$

$$A_{noise} = 11$$

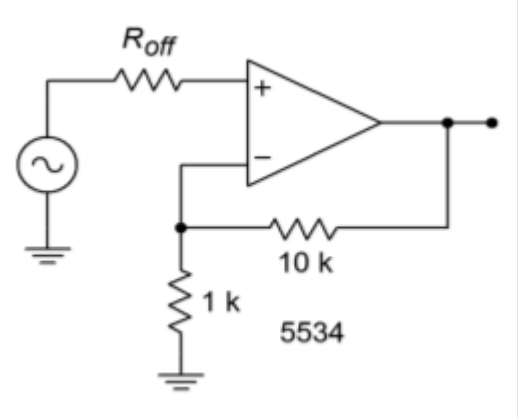


Figure 5.5.3 : Circuit for Example 5.5.1 .

$$V_{out-offset} = V_{OS}A_{noise} + |I_{B+}R_{off}A_{noise} - I_{B-}R_f|$$

$$V_{out-offset} = 0.5mV \times 11 + |800nA \times 0 \times 11 - 800nA \times 10k|$$

$$V_{out-offset} = 5.5mV + 8mV$$

$$V_{out-offset} = 13.5mV$$

Remember, this is the magnitude of the offset, it could be anywhere within ± 13.5 mV. It might be worse if this is not a typical device. Now we find the optimum offset compensating resistor:

$$R_{off} = R_i||R_f$$

$$R_{off} = 1k||10k$$

$$R_{off} = 909\Omega$$

For this case, the offset Equation reduces to

$$V_{out-offset} = V_{OS}A_{noise} + I_{OS}R_f$$

$$V_{out-offset} = 0.5mV \times 11 + 10nA \times 10k$$

$$V_{out-offset} = 5.5mV + 100\mu V$$

$$V_{out-offset} = 5.6mV$$

By adding ◈◈◈◈ , the output offset voltage is more than halved. This may lead you to think that it is always wise to add ◈◈◈◈ . Such is not the case. There are two times when you may prefer to leave it out: 1) to optimize noise characteristics, as we will see shortly, and 2) when using FET input op amps. FET input devices have very small input bias and offset currents to begin with, so their effect is negligible when using typical resistor values.

Example 5.5.2

The circuit of Figure 5.5.4 is used as part of a measurement system. Assuming that the DC input signal is 3 mV, how much uncertainty is there in the output voltage typically?

The desired output from the amplifier is

$$V_{out} = A_v V_{in}$$

$$V_{out} = -\frac{R_f}{R_i}V_{in}$$

$$V_{out} = -\frac{20k}{5k}3mV$$

$$V_{out} = -12mV$$

The typical specs for the 411 are ◈◈◈=0.8◈◈ , ◈◈◈=25◈◈ , ◈◈=50◈◈

$$A_{noise} = 1 + \frac{R_f}{R_i}$$

$$A_{noise} = 1 + \frac{20k}{5k}$$

$$A_{noise} = 5$$

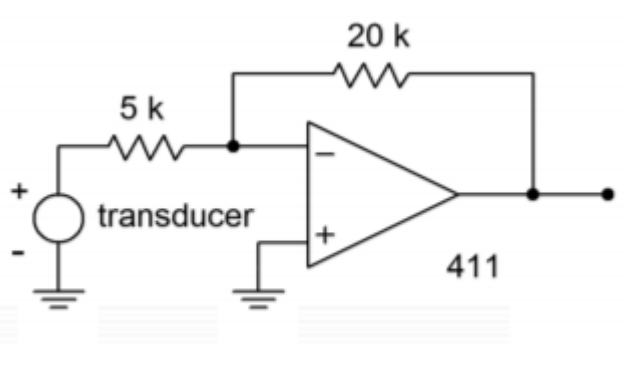


Figure 5.5.4 : Circuit for Example 5.5.2 .

$$V_{out-offset} = V_{OS}A_{noise} + |I_{B+}R_{off}A_{noise} - I_{B-}R_f|$$

$$V_{out-offset} = 0.8mV \times 5 + |50pA \times 0 \times 11 - 50pA \times 20k|$$

$$V_{out-offset} = 4mV + 1\mu V$$

$$V_{out-offset} = 4mV$$

The output can vary by as much as ± 4 mV. As this a DC measurement system, the results are devastating. The output can be anywhere from -12 mV – 4 mV = -16 mV, to -12 mV + 4 mV = -8 mV. That's a 2:1 spread, and it's caused solely by the op amp. Note that the addition of ◈◈◈◈ would have little effect here. Because the 411 uses a FET input, its ◈◈ contribution is only 1 ◈ V.

So, then, how do you keep output offsets to a minimum? First and foremost, make sure that the op amp chosen has low ◈◈◈ and ◈◈◈ ratings. Second, use the offset compensation resistor, ◈◈◈◈ . Third, keep the circuit resistances as low as possible. Finally, if the output offset is still too large, it can be reduced by manually nulling the circuit.

Nulling involves summing in a small signal that is of opposite polarity to the existing offset. By doing this, the new signal will completely cancel the offset and the output will show 0 V DC. This is much easier than it sounds. Most op amps have connections for null circuits. These are specified by the manufacturer and usually consist of a single potentiometer and perhaps one or two resistors. An example nulling connection is shown in Figure 5.5.5 . Usually the potentiometer is a multi-turn trim type to allow for fine adjustment. To null the circuit, the technician monitors the output with a very sensitive DC voltmeter. The input is grounded (or perhaps tied to ground through a resistor equal to the driving source resistance if it's large). The potentiometer is then adjusted until the meter reads zero.

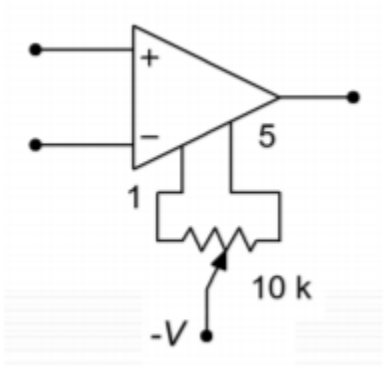


Figure 5.5.5 : A typical nulling connection.

The drawback to this procedure is that it requires someone (or perhaps some thing) to perform the nulling. Also, the unit will require periodic adjustment to compensate for aging and environmental effects.

COMPUTER SIMULATION

As stated, it is very important to match the input resistors in order to keep offsets low. This can be seen clearly in the simulation shown in Figure 5.5.6 . A noninverting voltage amplifier is modeled here using the 741. In order to focus on the offset current effect, the contribution of ⍰⍰⍰ is kept small by keeping the voltage gain low. The simulation is run twice for DC Operating Point. In the case where the resistances are matched, the DC output voltage is less than 1 mV. For the unmatched case, ⍰⍰⍰⍰ is set to a fraction of an Ohm. The resulting DC output voltage is much larger at approximately 17 mV. It is worthy to note that this is approximately equal to ⍰⍰⍰⍰⍰ times ⍰⍰ (80 nA times 200 k Ω). This simulation also points out the poor performance caused by excessively large resistor values. If the simulation is re-run with the resistors scaled down in size, the offsets will be lessened.

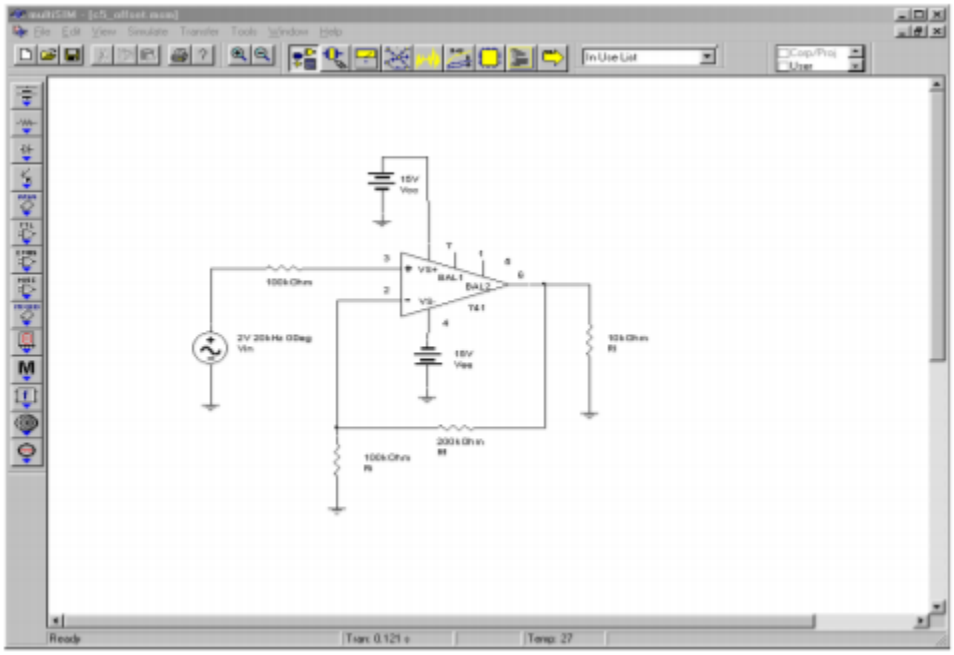


Figure 5.5.6⍰ : Multisim schematic for offset simulation.

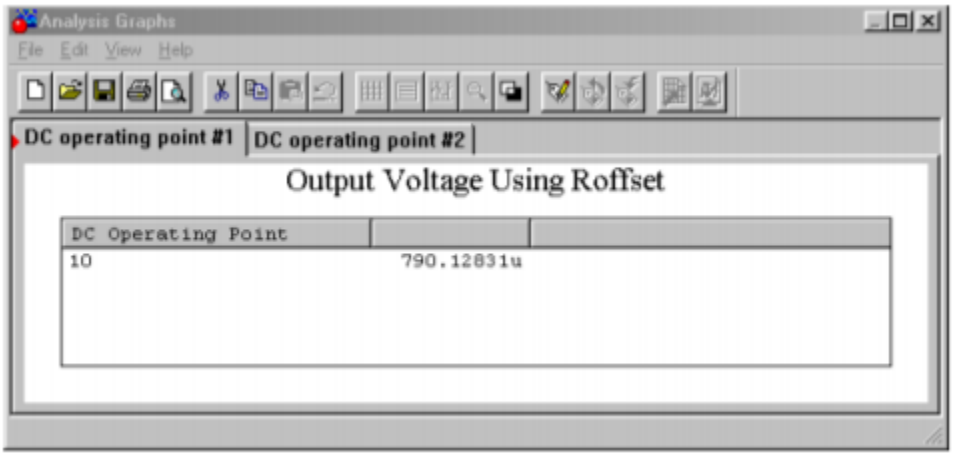


Figure 5.5.6 : Results with offset resistor.

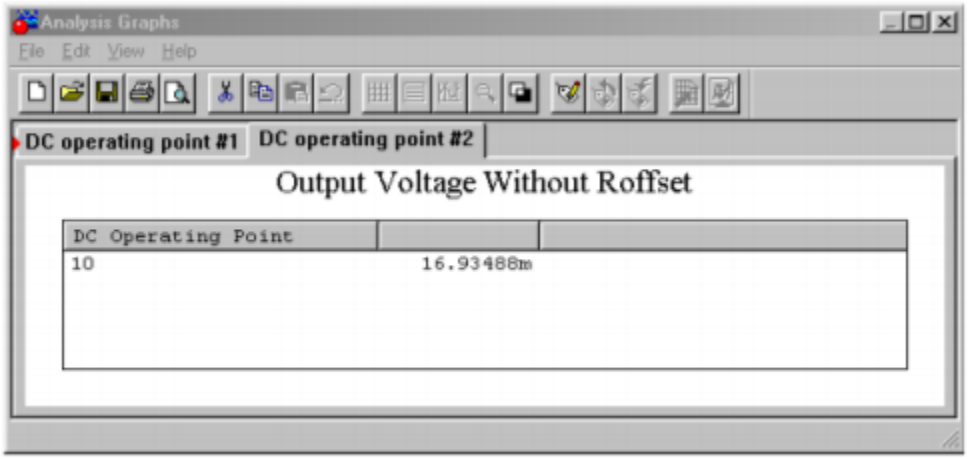


Figure 5.5.6 : Results without offset resistor.

11.6 DRIFT

Drift is a variation in the output offset voltage. Often, it is temperature induced. Even if a circuit has been manually nulled, an output offset can be produced if the temperature changes. This is because ⍰⍰⍰ and ⍰⍰⍰ are temperature sensitive. The only way around this is to keep the circuit in a constant temperature environment. This can be very costly. If the drift can be kept within an acceptable range, the added cost of cooling and heating equipment may be removed. As you might expect, the magnitude of the drift depends on the size of the temperature change. It also depends on the ⍰⍰⍰ and ⍰⍰⍰ sensitivities. These items are Δ⍰⍰⍰/Δ⍰ , the change in ⍰⍰⍰ with respect to temperature, and Δ⍰⍰⍰/Δ⍰ , the change in ⍰⍰⍰ with respect to temperature. Drift rates are specified in terms of change per centigrade degree. These parameters are usually specified as worst-case values and can produce either a positive or negative potential.

The development of the drift Equation pretty much follows that of the Equation for offsets. The only difference is that offset parameters are replaced by their temperature coefficients and the temperature change. The products of the coefficients and the change in temperature produce an input offset voltage and current.

$$V_{drift} = \frac{\Delta V_{OS}}{\Delta T}\Delta T A_{noise} + \frac{\Delta I_{OS}}{\Delta T}\Delta T R_f$$

(5.21)

As with the offset calculation, the drift result may be either positive or negative. Also, because ⍰⍰⍰ is so small for FET input devices, Δ⍰⍰⍰/Δ⍰ is often not listed, as it is almost always small enough to ignore. For lowest drift, it is assumed that the op amp uses the offset compensation resistor ⍰⍰⍰⍰ , and that the circuit has been nulled.

Example 5.6.1

Determine the output drift for the circuit of Figure 5.5.3 for a target temperature of 80 ° C. Assume that ⍰⍰⍰⍰=909Ω and that the circuit has been nulled at 25 ° C.

The parameters for the 5534 are Δ⍰⍰⍰/Δ⍰=5⍰⍰/⍰ ° , Δ⍰⍰⍰/Δ⍰=200⍰⍰/⍰ ° .

The noise gain was already determined to be 11 in Example 5.5.1. The total temperature change is from 25 ° C to 80 ° C, or 55C ° .

$$V_{drift} = \frac{\Delta V_{OS}}{\Delta T}\Delta T A_{noise} + \frac{\Delta I_{OS}}{\Delta T}\Delta T R_f$$

$$V_{drift} = 5\mu V/C^{\circ} \times 55C^{\circ} \times 11 + 200pA/C^{\circ} \times 55C^{\circ} \times 10k$$

$$V_{drift} = 3.025mV + .11mV$$

$$V_{drift} = 3.135mV$$

Note that for this circuit the ◈◈◈ drift is the major source of error. At 80 ° C, the output of the circuit may have up to ± 3.135 mV of DC error.

As with offsets, drift is partially a function of the circuit gain. Therefore high gain circuits often appear to have excessive drift. In order to compare different amplifiers, input referred drift is often used. To find input referred drift, just divide the output drift by the signal gain of the amplifier. Don't use the noise gain! In this way, both inverting and noninverting amplifiers can be compared on an equal footing. For the circuit just examined, the input referred drift is

$$V_{drift(input)} = \frac{V_{drift}}{A_v}$$

$$V_{drift(input)} = \frac{3.135mV}{11}$$

$$V_{drift(input)} = 285\mu V$$

For many applications, particularly those primarily concerned with AC performance, drift specification is not very important. A communications amplifier, for example, might use an output coupling capacitor to block any drift or offset from reaching the output if it had to. Drift is usually important for applications involving DC or very low frequencies where coupling capacitors are not practical.

11.7 CMRR AND PSRR

CMRR stands for Common Mode Rejection Ratio. It is a measure of how well the two halves of the input differential amplifier stage are matched. A common-mode signal is a signal that is present on both inputs of the diff amp. Ideally, a differential amplifier completely suppresses or rejects common-mode signals. Common-mode signals should not appear at the circuit output. Due to the non-perfect matching of transistors, some portion of the common-mode signal will make it to the output. Exactly how much this signal is reduced relative to desired signals is measured by the CMRR.

Ideally, CMRR is infinite. A typical value for CMRR would be 100 dB. In other words, if an op amp had both desired (i.e., differential) and common-mode signals at its input that were the same size, the common-mode signal would be 100 dB smaller than the desired signal at the output.

CMRR is particularly important when using the op amp in differential mode (Chapter Four) or when making an instrumentation amplifier (Chapter Six). There are two broad uses for these circuits. First, the amplifier may be receiving a low level, balanced signal over a considerable distance.[1] Good examples of this are a microphone cable in a recording studio and an instrumentation cable on a factory floor. Interference signals tend to be induced into the cable in-phase (i.e., common-mode). Because the desired signal is presented out-of-phase (i.e., differential), a high CMRR will effectively remove the interference signal. Second, the op amp may be used as part of a bridge-type measurement system. Here, the desired signal is seen as a small variation between two DC potentials. The op amp must amplify the difference signal, but suppress the DC outputs of the bridge circuit.

Example 5.7.1

An amplifier has a closed loop voltage gain of 20 dB and a CMRR of 90 dB. If a common-mode signal is applied to the input at -60 dBV, what is the output?

If the input signal were differential instead of common-mode, the output would be:

$$V_{out}^{'} = A_v^{'} + V_{in}^{'}$$

$$V_{out}^{'} = 20dB + (-60dBV)$$

$$V_{out}^{'} = -40dBV$$

Because this is a common-mode signal, it is reduced by the CMRR

$$V_{out}^{'} = -40dBV - 90dB$$

$$V_{out}^{'} = -130dBV$$

1. A balanced system uses two signal-carrying conductors and a shield (ground). The two signals are the reference, or in-phase signal, and the inverted, or out-of-phase signal.

This signal is so small that it is probably overshadowed by the circuit noise.

One final note concerning CMRR is that it is specified for DC. In truth, CMRR is frequency dependent. The shape of its curve is reminiscent of the open loop gain curve. The stated CMRR may remain at its DC level up to perhaps 100 or 1000 Hz, and then fall off as frequency increases. For example, the 741 data sheet found in the Appendix states a typical CMRR of 90 dB. By looking at the CMRR graph, though, you can see that it starts to roll off noticeably around 1 kHz. By the time it hits 1 MHz, only 20 dB of rejection remains. A more gentle rolloff is exhibited by the 411. At 1 MHz, almost 60 dB of rejection remains. What this means is that the op amp cannot suppress high frequency interference signals as well as it suppresses low frequency interference.

Similar to CMRR is PSRR, or Power Supply Rejection Ratio. Ideally, all ripple, hum, and noise from the power supply will be prevented from reaching the output of the op amp. PSRR is a measure of exactly how well the op amp reaches this ideal. Typical values for PSRR are in the 100 dB range. Like CMRR, PSRR is frequency-dependent and shows a rolloff as frequency increases. If an op amp is powered by a 60 Hz source, the ripple frequency from a standard full-wave rectifier will be 120 Hz. At the output, this ripple will be reduced by the PSRR. Higher frequency noise components on the power supply line are not reduced as much because PSRR rolls off. Normally, PSRR is consistent between power rails. Sometimes, there is a marked performance difference between the positive and negative PSRR. One good example of this is the 411. The positive rail exhibits about a 30 dB improvement over the negative rail. Note that PSRR is only a few decibels for the negative rail by the time it reaches 1 MHz.

Example 5.7.2

An op amp is operated off a power supply that has a peak to peak ripple voltage of 0.5 V. If the op amp's PSRR is 86 dB, how much of this ripple is seen at the circuit output?

First, determine the PSRR as an ordinary value.

$$PSRR = log_{10}^{-1}\frac{PSRR}{20}$$

$$PSRR = log_{10}^{-1}\frac{86dB}{20}$$

$$PSRR = 20,000$$

Divide the ripple voltage by the PSRR to find the amount that is seen at the output.

$$V_{out-ripple} = \frac{V_{ripple}}{PSRR}$$

$$V_{out-ripple} = \frac{0.5Vpp}{20,000}$$

$$V_{out-ripple} = 25\mu Vpp$$

11.8 NOISE

Generally speaking, noise refers to undesired output signals. The background hiss found on audio tape is a good example of noise. If noise levels get too high, the desired signals are lost. We will narrow our definition down a bit by only considering noise signals that are created by the op amp circuit. Noise comes from a variety of places. First of all, all resistors have thermal, or Johnson noise. This is due to the random effects that thermal energy produces on the electrons. Thermal noise is also called white noise, because it is equally distributed across the frequency spectrum. Semiconductors exhibit other forms of noise. Shot noise is caused by the fact the charges move as discrete particles (electrons). It is also white. Popcorn noise is dominant at lower frequencies and is caused by manufacturing imperfections. Finally, flicker noise has a $1/f$ spectral density. This means that it increases as the frequency drops. It is sometimes referred to as $1/f$ noise.

Any op amp circuit exhibits noise from all of these sources. Because we are not designing the op amps, we don't really need to distinguish the exact sources of the noise, rather, we'd just like to find out how much total noise arrives at the output. This will allow us to determine the signal-to-noise ratio (S/N) of the circuit, or how quiet the amplifier is.

If noise performance for a particular design is not paramount, it can be quickly estimated from manufacturer's data sheets. Some manufacturers will specify RMS noise voltages for specified signal bandwidths and source impedances. A typical plot is found in Figure 5.8.1 . To use this, simply find the source impedance of your circuit on the horizontal axis, and by using the appropriate signal bandwidth curve, find the noise voltage on the vertical axis. This noise voltage is input-referred. In order to find the output noise voltage, multiply this number by the noise gain of the circuit. The result will not be exact, but it will put you in the ballpark.

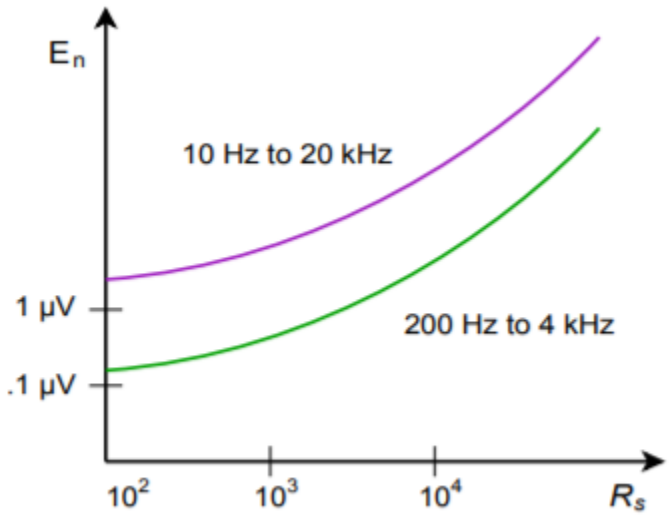


Figure 5.8.1 : Noise voltage for given bandwidth versus source resistance.

A more exact approach involves the use of two op amp parameters, input noise voltage density,

◇◇◇◇ , and input noise current density, ◇◇◇◇ . Nanovolts per root Hertz are used to specify ◇◇◇◇ . Picoamps per root Hertz are used to specify ◇◇◇◇ . Refer to the 5534 data sheet for example specifications. (Some manufacturer's square these values and give units of volts squared per Hertz and amps squared per Hertz. To translate to the more common form, just take the square root of the values given. The data sheet for the 741 is typical of this form). These two parameters take into account noise from all internal sources. As a result, these parameters are frequency dependent. Due to the flicker noise component, the curves tend to be rather flat at higher frequencies and then suddenly start to increase at lower frequencies. The point at which the graphs start to rise is called the noise corner frequency. Generally, the lower this frequency, the better.

In order to find the output noise we will combine the noise from three sources:

1. ◇◇◇◇
2. ◇◇◇◇
3. The thermal noise of the input and feedback resistors

Before we start, there are a few points to note. First, the strength of the noise is dependent on the noise bandwidth of the circuit. For general-purpose circuits with 20 dB/decade rolloffs, the noise-bandwidth, ◇◇◇◇◇◇◇ , is 1.57 times larger than the small-signal bandwidth. It is larger because some noise still exists in the rolloff region. The noise-bandwidth would only equal the small-signal bandwidth if the rolloff rate were infinitely fast. Second, because the noise-bandwidth factor is common to all three sources, instead of calculating its effect three times, we will combine the partial results of the three sources first, and then apply the noise bandwidth effect. This will make the calculation faster, as we have effectively factored out ◇◇◇◇◇◇◇ . Each of the three sources will be noise voltage densities, all having units of volts per root Hertz. Because ◇◇◇◇ is already in this form, we only need to calculate the thermal and ◇◇◇◇ effects before doing the summation. Finally, because noise signals are random, they do not add coherently. In order to find the effective sum we must perform an RMS summation, that is, a square root of the sum of the squares. This will result in the input noise voltage. To find the output noise voltage we will then multiply by ◇◇◇◇◇◇ .

The first thing that must be done is to determine the input noise resistance. ◇◇◇◇◇◇ is the combination of the resistance seen from the inverting input to ground and from the noninverting input to ground. To do this, short the voltage source and ground the output. You will end up with circuit like Figure 5.8.2 . Note that ◇◇ and ◇◇ are effectively in parallel. Therefore,

$$R_{noise} = R_s + R_i||R_f$$

(5.8.1)

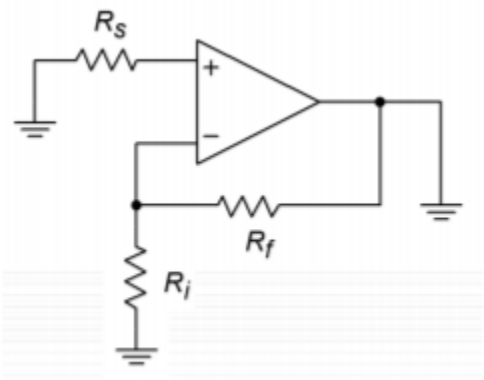


Figure 5.8.2 : Equivalent noise analysis circuit.

You might note that ⟐⟐ is in the same position as ⟐⟐⟐⟐ for the offset calculations. For absolute minimum noise, the offset compensating resistor is not used. ⟐⟐⟐⟐⟐⟐ is used to find the thermal noise and the contribution of ⟐⟐⟐⟐ . As always, Ohm's Law still applies, so as you might expect, ⟐⟐⟐⟐ ⟐⟐⟐⟐⟐⟐ produces a noise voltage density with our desired units of volts per root Hertz. The only source left is the thermal noise.

The general Equation for thermal noise is:

$$e_{th} = \sqrt{4KTBW_{noise}R_{noise}}$$

(5.8.2)

Where

⟐⟐*h* is the thermal noise.

⟐ is Boltzmann's constant, 1.38·10–23 Joules/Kelvin degree.

⟐ is the temperature in Kelvin degrees (Celsius + 273).

⟐⟐⟐⟐⟐⟐⟐ is the effective noise bandwidth.

⟐⟐⟐⟐⟐⟐ is the equivalent noise resistance.

Because we are interested in finding the noise density, we can pull out the ⟐⟐⟐⟐⟐⟐⟐ factor. Also, when we perform the RMS summation, this quantity will need to be squared. Instead of taking the square root and then squaring it again, we can just leave it as ⟐2⟐*h* . These two considerations leave us with the mean squared thermal noise voltage density, or

$$e_{th}^2 = 4KTR_{noise}$$

(5.8.3)

Now that we have the components, we may perform the summation.

$$e_{total} = \sqrt{v_{ind}^2 + (i_{ind} \times R_{noise})^2 + e_{th}^2}$$

(5.8.4)

The total input noise voltage density is ⟐⟐⟐⟐⟐⟐ . Its units are in volts per root Hertz. At this point we may now include the noise bandwidth effect. In order to find ⟐⟐⟐⟐⟐⟐⟐ , multiply the small-signal bandwidth by 1.57. If the amplifier is DC coupled, the small-signal bandwidth is equal to ⟐2 , otherwise it is equal to ⟐2–⟐1 . For most applications setting the bandwidth to ⟐2 is sufficient.

$$BW_{noise} = 1.57f_2$$

(5.8.5)

For the final step, ⟐⟐⟐⟐⟐⟐ is multiplied by the square root of ⟐⟐⟐⟐⟐⟐⟐ . Note that the units for ⟐⟐⟐⟐⟐⟐ are volts per root Hertz. Consequently, we need a root Hertz bandwidth. We are

performing a mathematical shortcut here. You could square ⍰⍰⍰⍰⍰⍰ in order to get units of volts squared per Hertz, multiply by ⍰⍰⍰⍰⍰⍰⍰ , and then take the square root of the result to get back to units of volts, but the first way is quicker. Anyway, we end up with the input referred RMS noise voltage.

$$e_n = e_{total}\sqrt{BW_{noise}} \tag{5.8.6}$$

Example 5.14

Determine the output noise voltage for the circuit of Figure 5.8.3 . For a nominal output of 1 V RMS, what is the signal-to-noise ratio? Assume T = 300 ° K (room temperature).

The 5534 shows the following specs, ⍰⍰⍰⍰=4⍰⍰/⍰⍰—√ , ⍰⍰⍰⍰=0.6⍰⍰/⍰⍰—√ . These values do rise at lower frequencies, but we will ignore this effect for now. Also, ⍰⍰⍰⍰⍰⍰ is 10 MHz.

$$A_v = 1 + \frac{R_f}{R_i}$$

$$A_v = 1 + \frac{99k}{1k}$$

$$A_v = 100$$

$$A_v^{'} = 40dB$$

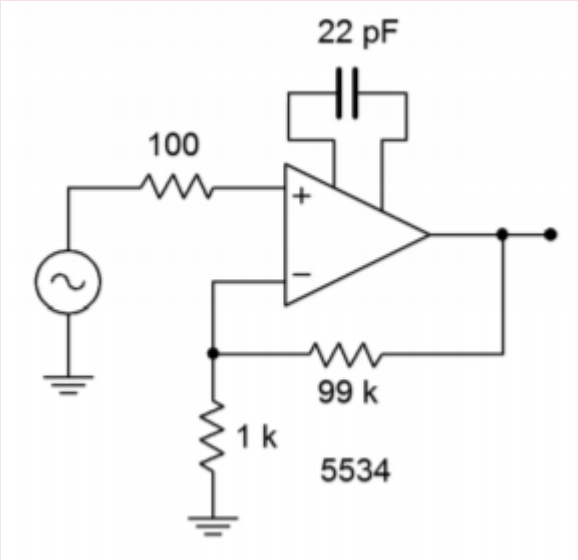


Figure 5.8.3 : Circuit for Example 5.8.1 .

For the noninverting amplifier, ⍰⍰=⍰⍰⍰⍰⍰⍰ , so

$$A_{noise} = 100 = 40dB$$

$$R_{noise} = R_s + R_f||R_i$$

$$R_{noise} = 100 + 99k||1k$$

$$R_{noise} = 1090\Omega$$

$$e_{th}^2 = 4KTR_{noise}$$

$$e_{th}^2 = 4 \times 1.38 \times 10^{-23} \times 300 \times 1090$$

$$e_{th}^2 = 1.805 \times 10^{-17} \text{ Volts squared per Hertz}$$

$$e_{total} = \sqrt{v_{ind}^2 + (i_{ind}R_{noise})^2 + e_{th}^2}$$

$$e_{total} = \sqrt{(4nV/\sqrt{Hz})^2 + (.6pA/\sqrt{Hz} \times 1090\Omega)^2 + 1.805 \times 10^{17}V^2/Hz}$$

$$e_{total} = \sqrt{1.6 \times 10^{-17} + 4.277 \times 10^{-19} + 1.805 \times 10^{-17}}$$

$$e_{total} = 5.87nV/\sqrt{Hz}$$

Note that the major noise contributors are ◈◈◈◈ and ◈◈h . Now to find ◈◈◈◈◈◈◈ .

$$f_2 = \frac{f_{unity}}{A_{noise}}$$

$$f_2 = \frac{10MHz}{100}$$

$$f_2 = 100kHz$$

$$BW_{noise} = f_2 1.57$$

$$BW_{noise} = 100kHz \times 1.57$$

$$BW_{noise} = 157kHz$$

$$e_n = e_{total}\sqrt{BW_{noise}}$$

$$e_n = 5.87nV/\sqrt{Hz}\sqrt{157kHz}$$

$$e_n = 2.33\mu VRMS$$

To find the output noise, multiply by the noise gain.

$$e_{n-out} = e_n A_{noise}$$

$$e_{n-out} = 2.33\mu V \times 100$$

$$e_{n-out} = 233\mu VRMS$$

For a nominal output signal of 1 V RMS, the signal-to-noise ratio is

$$S/N = \frac{Signal}{Noise}$$

$$S/N = \frac{1V}{233\mu V}$$

$$S/N = 4290$$

Normally S/N is given in dB

$$S/N^{'} = 20\log_{10} S/N$$

$$S/N^{'} = 20\log_{10} 4290$$

$$S/N^{'} = 72.6dB$$

As was noted earlier, the noise curves increase at lower frequencies. How do you take care of this effect? First of all, when using a wide-band design with a noise corner frequency that is relatively low (as in Example 5.8.1), it can be safely ignored. If the low frequency portion takes up a sizable chunk of the signal frequency range, it is possible to split the calculation into two or more segments. One segment would be for the constant part of the curves. Other segments can be made for the lower frequency portions. In these regions, an averaged value would be used for ⍰⍰⍰⍰ and ⍰⍰⍰⍰ . This is a rather advanced treatment, and we will not pursue it here.

Finally, it is common to use the parameter input referred noise voltage. Input referred noise voltage is the output noise voltage divided by the circuit signal gain.

$$e_{in-ref} = \frac{e_{n-out}}{A_v}$$

(5.8.7)

This value is the same as ⍰⍰ for noninverting amplifiers, but varies a bit for inverting amplifiers because ⍰⍰⍰⍰⍰⍰ does not equal ⍰⍰ for inverting amplifiers.

11.8 SUMMARY

In this chapter we have taken a closer look at op amp characteristics. First of all, we find that the upper frequency limit is a function of the op amp parameter ◇◇◇◇◇◇ , also known as the gain-bandwidth product, and the circuit's noise gain. The higher the gain is, the lower the upper break frequency will be. Op amps are capable of flat response down to DC. If coupling capacitors are used, the lower break frequency may be found by using standard lead network analysis. When stages are cascaded, the results echo those of cascaded discrete stages. The lowest $◇_2$ is dominant and becomes the system $◇_2$. The highest $◇_1$ is dominant and sets the system $◇_1$. If more than one stage exhibits the dominant critical frequency, the actual critical frequency will be somewhat lower for $◇_2$ and somewhat higher for $◇_1$.

In order to make the op amp unconditionally stable, a compensation capacitor is used to tailor the open loop frequency response. Besides setting ◇◇◇◇◇◇ , this capacitor also sets the slew rate. Slew rate is the maximum rate of change of output voltage with respect to time. Slewing slows down the edges of pulse signals and distorts sinusoidal signals. The highest frequency that an amplifier can produce without slewing is called the power bandwidth. In order to optimize ◇◇◇◇◇◇ and slew rate, some amplifiers are available without the compensation capacitor. The designer then adds just enough capacitance to make the design stable.

Due to slight imperfections between the input transistors, op amps may produce small DC output voltages called offsets. Offsets may be reduced through proper resistor selection. Simple nulling circuits may be used to completely remove the offset. A variable offset due to temperature variation is called drift. The larger the temperature variation, the larger the drift will be. The transistor mismatch also means that common-mode signals will not be completely suppressed. Just how well common mode signals are suppressed is measured by the common-mode rejection ratio, CMRR. Similar to CMRR is PSRR, the power-supply rejection ratio. PSRR measures how well power-supply noise and ripple are suppressed by the op amp. Both PSRR and CMRR are frequency-dependent. Their maximum values are found at DC and then they decrease as frequency increases.

Finally, noise is characterized as an undesired random output signal. The noise in op amp circuits may be characterized by three components: the thermal noise of the input and feedback resistors, the op amp's input noise voltage density, ◇◇◇◇ , and its input noise current density, ◇◇◇◇ . The combination of these elements requires an RMS summation. In order to find the output noise voltage, the input noise voltage is multiplied by the circuit's noise gain. The ratio of the desired output signal and the noise voltage is called, appropriately enough, the signal-to-noise ratio. Normally, signal-to-noise ratio is specified in decibels.

11.10 PROBLEMS

REVIEW QUESTIONS

1. Define gain-bandwidth product. What is its use?
2. How do you determine f_2 and f_1 for a multi-stage circuit?
3. What happens if two or more stages share the same break frequency?
4. What is slew rate?
5. How is power bandwidth determined?
6. How do power bandwidth and small-signal bandwidth differ?
7. What are the advantages and disadvantages of noncompensated op amps?
8. What are decompensated op amps?
9. What causes DC offset voltage?
10. What causes DC drift voltage?
11. What is CMRR?
12. What is PSRR?
13. What parameters describe an op amps noise performance?
14. What is S/N?

PROBLEMS

Analysis Problems

1. Determine f_2 for the circuit in Figure 5.3.3 if a 411 op amp is used.
2. Determine f_2 for the circuit of Figure 5.3.5 if a 318 op amp is used. Note: f_{GBW} = 15 MHz for a 318.
3. What is the minimum acceptable f_{GBW} for the op amp in Figure 5.3.3 if the desired f_2 is 250 kHz?
4. What is the minimum acceptable f_{GBW} for the op amp in Figure 5.3.5 if the desired f_2 is 20 kHz?
5. Determine the power bandwidth for Problem 5.1. Assume V_{PP} = 10 V.
6. 6. Determine the power bandwidth for Problem 5.3. Assume V_{PP} = 12 V.
7. What is the minimum acceptable slew rate for the circuit of Figure 5.3.3 if the desired power bandwidth is 20 kHz with a V_{PP} of 10 V?

8. What is the minimum acceptable slew rate for the circuit of Figure 5.3.5 if the desired power bandwidth is 40 kHz with a ◇◇ of 5 V?
9. A circuit has the following specifications: ± 15 V power supply, voltage gain equals 26 dB, desired power bandwidth equals 80 kHz at clipping. Determine the minimum acceptable slew rate for the op amp.
10. Determine the system $◇_2$ in Figure 5.3.6 if all three devices are 318s.
11. Determine the system slew rate for Figure 5.3.7. The first device is a 741 and the second unit is a 411.
12. Find the output offset voltage for Figure 5.3.3.
13. If ◇◇=◇◇=100Ω in Figure 5.5.2, find the output offset voltage using a 318 op amp.
14. Assume that the circuit of Figure 5.5.3 is nulled at 25 ° C and that an optimum value for ◇◇◇◇ is used. Determine the drift at 75 ° C.
15. Assuming that the 120 Hz power supply ripple in Figure 5.10.1 is 50 mV, how large is its contribution to the output?

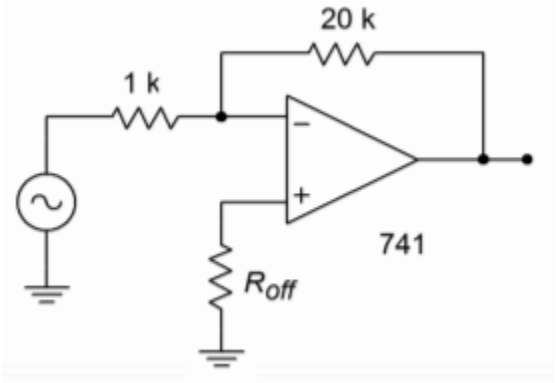


Figure 5.10.1

16. Utilizing a 5534 op amp, what is the approximate input noise voltage for a source resistance of 1 k Ω and a bandwidth from 10 Hz to 20 kHz?
17. Assuming that the op amp of Problem 5.16 is connected like Figure 5.10.1 , what is the approximate output noise voltage? What is the approximate input referred noise voltage?
18. Assume that ◇◇=0Ω , ◇◇=500Ω , and ◇◇=10◇Ω in Figure 5.8.2. Find the input noise voltage if the op amp is a 411.
19. For a nominal output voltage of 2 V RMS, determine the signal-to-noise ratio for Problem 5.18. Assume that the input to Figure 5.5.2 is a 5 V peak 50 kHz square wave. Draw the output waveform if a 741 is used.
20. Repeat Problem 5.20 using a 10 V peak, 100 kHz sine wave.

Design Problems

21. Determine an optimum value for ◇◇◇◇ in Figure 5.3.3 and determine the resulting offset voltage.

22. Determine the optimum value for ◇◇◇◇ in Figure 5.10.1 . Assuming that the circuit has been nulled at 25 ° C, find the drift at 60 ° C.
23. Determine a new value for the 100 k Ω resistor in Figure 5.3.8 in order to minimize the output offset.
24. Using the optimum resistor found in Problem 5.24, determine a new value for the input capacitor that will maintain the original $\diamond_1$.
25. Design a circuit with a gain of 32 dB and an $\diamond_2$ of at least 100 kHz. You may use any of the following: 741, 411, 318.
26. Design a circuit with a gain of 50 and an ◇◇◇◇ of at least 50 kHz, given a maximum output swing of 10 V peak. You may use any of the following: 741, 411, 318.
27. Design a circuit with a gain of 12 dB, a small-signal bandwidth of at least 100 kHz, and an ◇◇◇◇ of at least 100 kHz for a peak output swing of 12 V.
28. Utilizing two or more stages, design a circuit with a gain of 150 and a small signal bandwidth of at least 600 kHz.

Challenge Problems

29. Determine the system $\diamond_2$ in Figure 5.4.10.
30. Determine the input noise voltage for the circuit of Figure 5.10.1 . Assume ◇◇◇◇ = 950 Ω .
31. Determine the output noise voltage and the input-referred noise voltage for Problem 5.31.
32. Assuming that driving source resistance in Figure 5.3.7 is 0 Ω , how much offset voltage is produced at the output of the circuit? Assume that both devices are 411's.
33. Assume that you have one each of: 411, 741, 318. Determine the combination that will yield the highest system slew rate in Figure 5.10.2 .

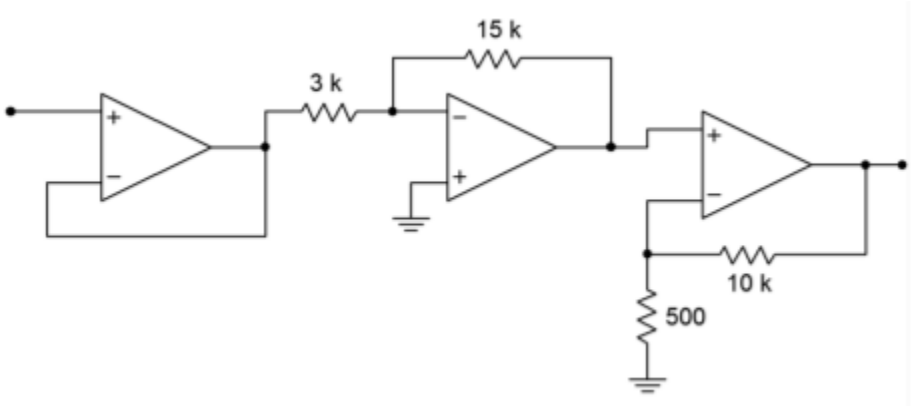


Figure 5.10.2

34. Repeat Problem 5.34 in order to produce the highest system $\diamond_2$.
35. Assuming that the driving source resistance is 0 Ω in Figure 5.3.7. Determine the output noise voltage if both devices are 411's.
36. Derive Equation 5.3.9 from the text.

Computer Simulation Problems

37. Use a simulator to create a Bode plot for Problem 5.10. If a macro model for the LM318 is not available, use the 741 instead.
38. Create a time domain representation of the output voltage of Problem 5.20 using a simulator.
39. Create a time domain representation of the output voltage of Problem 5.21 using a simulator. Repeat the simulation using an LM318 op amp in place of the 741. What do the results indicate?
40. Simulate the circuit designed in Problem 5.26. Verify $\diamondsuit_2$ and $\diamondsuit\diamondsuit$ through a Bode plot.
41. Generate a Bode plot using a simulator and verify $\diamondsuit_2$ and $\diamondsuit\diamondsuit$ for the circuit designed in Problem 5.29.
42. Generate a Bode plot using a simulator and verify $\diamondsuit_2$ for the circuit designed in Problem 5.35.

UNIT 12: VOLTAGE REGULATION

Learning Objectives

After completing this chapter, you should be able to:

- Understand the need and usefulness of voltage regulators.
- Detail the difference between load and line regulation.
- Detail the need for a stable reference voltage.
- Describe the operation of a basic linear regulator.
- Describe the operation of a basic switching regulator.
- Outline the differences between linear and switching regulators, giving the advantages and disadvantages of each.
- Detail the function of a pass transistor in both linear and switching regulators.
- Utilize many of the popular linear IC regulators, such as the 78XX and 723.
- Analyze system heat sink requirements.

12.1 INTRODUCTION

In this chapter, the need for voltage regulation is examined. The different schemes for achieving voltage regulation and the typical ICs used are presented. By the end of this chapter, you should be able to use standard voltage regulator devices in your work and understand the advantages and disadvantages of the various types.

Generally, voltage regulators are used to keep power supply potentials constant in spite of changes in load current or source voltage. Without regulation, some circuits may be damaged by the fluctuations present in the power supply voltage. Even if devices are not damaged, the fluctuations may degrade circuit performance.

Two general forms are presented, the linear regulator and the switching regulator. Both forms have distinct advantages and drawbacks. The trick is to figure out which form works best in a given situation. Due to their wide use in modern power supply design, many regulators have achieved high levels of internal sophistication and robust performance. Often, the inclusion of voltage regulator ICs is a very straightforward – almost cookbook – affair.

The chapter concludes with a discussion of heat sink theory and application. As regulators are called upon to dissipate a bit of power, attention to thermal considerations is an important part of the design process.

12.2 THE NEED FOR REGULATION

Modern electronic circuits require stable power supply voltages. Without stable potentials, circuit performance may degrade, or if variations are large enough, the circuit may cease to function all together and various components may be destroyed. There are many reasons why a power supply may fluctuate. No matter what the cause, it is the job of the regulator to absorb the fluctuations, and thereby maintain constant operating potentials.

A basic power supply circuit is shown in Figure 8.2.1 . First, a transformer is used to isolate the circuit from the AC power source. It is also used to step down (or step up) the AC source potential to a more manageable level. A rectifier then converts the scaled AC signal to pulsating DC, in the form of either half-wave or, more typically, full-wave rectification. The variations in the pulsating DC signal are then filtered out in order to produce a (hopefully) stable DC potential, which feeds the load. The filter may take the form of a simple capacitor or, possibly, more complex networks comprising both capacitors and inductors.

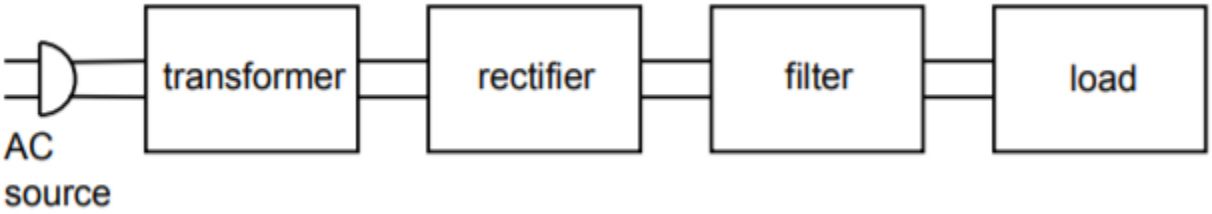


Figure 8.2.1 : Basic power supply.

There are two main causes of power supply output variation. First, if the AC source signal changes, a proportional change will be seen at the output. If, for example, a brown out1 occurs, the nominal 120 V AC source used in the US, may drop to, say, 100 V AC. This represents a decrease of one sixth, or about 16.7%. This same change will be reflected by the transformer, so the output of the transformer will be about 16.7% low as well. This reduction carries right through the rectifier and filter to the load. In some circuits, this will not present a major problem. For example, this may mean that an op amp will be running off a 12 V supply instead of a 15 V supply. On the other hand, a TTL logic circuit may not operate correctly if only 4 V instead of 5 V are applied. Of course, if an over-voltage occurred, both the op amp and the logic circuit could be damaged.

How well a circuit handles variations in the AC line signal is denoted by the parameter, line regulation.

$$Line\ Regulation = \frac{V_{max} - V_{min}}{V_{min}} \times 100$$

(8.2.1)

where ◈◈◈◈ is the load voltage produced at the maximum AC line potential and ◈◈◈◈ is the load voltage for the lowest AC line potential. Normally, this value is expressed as a percentage.

Ideally, a power supply will always produce the same output potential, and therefore, the perfect line regulation figure would be 0%.

The other major source of supply variation is variation of load current. Indeed, load current variations can be far greater than the usual variations seen in the AC line. This normally has the effect of increasing or decreasing the amount of AC ripple seen in the output voltage. Ideally, there would be no ripple in the output. Ripple is caused by the fact that heavy load currents effectively reduce the discharge time constant of the filter. The result is that the filter gives up its stored energy faster and cannot successfully "fill in the gaps" of the pulsating DC signal fed to it. The ripple signal is effectively an AC signal that rides on the DC output. This rapid variation of supply potential can find its way into an audio or signal processing path and create a great deal of interference. Along with this variation, the effective DC value of the supply may drop as the load current is increased. The figure of merit for stability in spite of load changes is called load regulation.

$$Load\ Regulation = \frac{V_{max} - V_{min}}{V_{min}} \times 100$$

(8.2.2)

where ◈◈◈◈ is the largest load voltage produced, and ◈◈◈◈ is the minimum load voltage produced. These points usually occur at the minimum and maximum load currents, respectively. Again, this value is normally expressed as a percentage and would ideally be 0%.

In order to maintain a constant output voltage, the power supply regulator needs to sense its output and then compensate for any irregularities. This implies a number of things. First, some form of reference is needed for stable comparison to the output signal. Second, some form of comparator or amplifier is required in order to make use of this comparison. Finally, some form of control element is needed to absorb the difference between the input to the regulator circuit and the desired output. This control element may either appear in series or in parallel with the load, as seen in Figure 8.2.2 . The first form is shown in Figure 8.2.2◈ and is normally referred to as a series mode regulator. The control element allows current to pass through to the load, but drops a specific amount voltage. The voltage that appears across the control element is the difference between the filter's output and the desired load voltage.

Figure 8.2.2◈ shows a shunt mode regulator. Here, the control element is in parallel with the load and draws enough current to keep the output level constant. For many applications, shunt mode regulators are not as efficient as series mode regulators are. One example of a simple shunt mode regulator is the resistor/Zener diode arrangement shown in Figure 8.2.3 . Note that under no-load conditions (i.e., when the load impedance is infinite) considerable current flows in the regulating circuit.

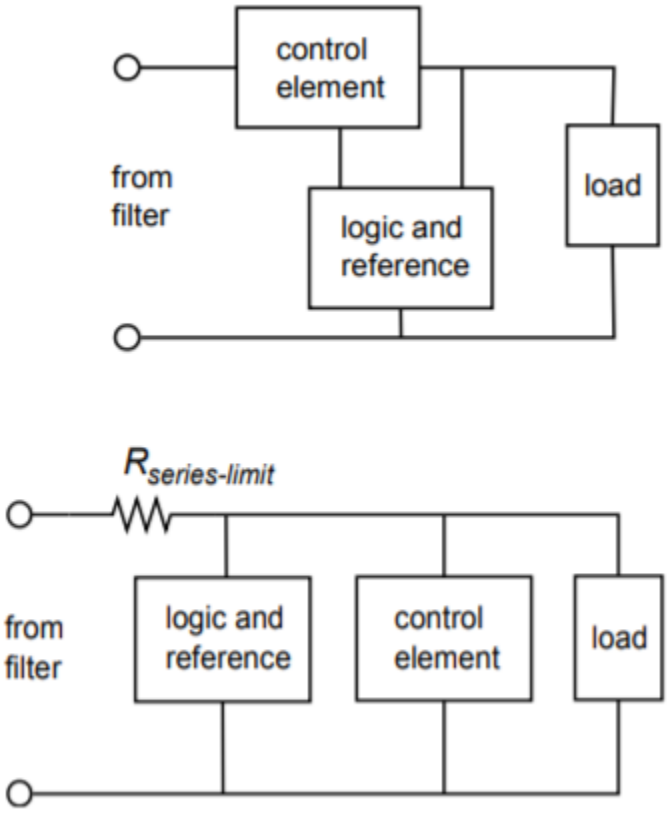


Figure 8.2.2 : Two forms of regulator control elements. a. Series regulator (top). b. Parallel regulator (bottom).

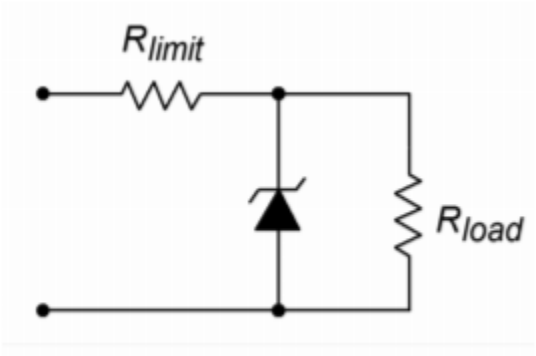


Figure 8.2.3 : Simple zener shunt regulator.

If a regulator's control element operates in its linear region, the regulator is said to be a linear regulator. Most linear regulator ICs are series mode types. The main advantages of linear regulators are their ease of use and accuracy of control. Their main disadvantage is low efficiency.

A basic linear regulator is shown in Figure 8.3.1 . The series control element is the transistor ◈1 . This component is most often referred to as a pass transistor, because it allows current to pass through to the load. The pass transistor is driven by the op amp. The pass transistor's job is to amplify the output current of the op amp. The op amp constantly monitors its two inputs. The noninverting input sees the Zener voltage of device ◈1 . ◈1 is used to properly bias ◈1 . The op amp's inverting input sees the voltage drop across ◈3 . Note that ◈2 and ◈3 make up a voltage divider with ◈◈◈◈◈ as the divider's source. Remembering the basics of negative feedback, recall that the op amp will produce enough current to keep its two inputs at approximately the same level, thus keeping ◈◈◈◈◈◈ at zero. In other words, the voltage across ◈3 should be equal to the Zener potential. As long as the op amp has a high enough output current capability, this equality will be maintained. Note that the◈◈◈drop produced by ◈1 is compensated for, as it is within the feedback loop. The difference between the filtered input signal and the final ◈◈◈◈◈ is dropped across the collector/emitter of ◈1 .

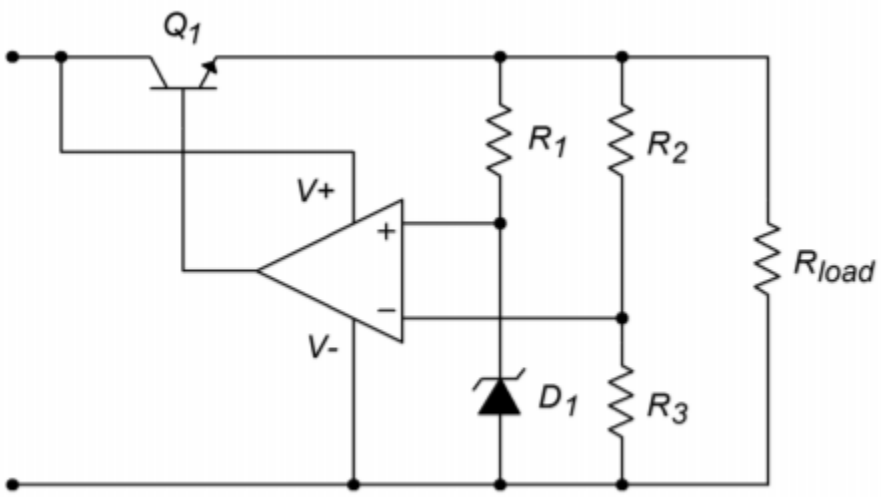


Figure 8.3.1 : Basic op amp linear regulator.

In essence, the circuit of Figure 8.3.1 is a noninverting amplifier. ◈◈◈◈◈◈ is the input potential, and ◈2 and ◈3 take the place of resistors ◈◈ and ◈◈ , respectively. ◈◈◈◈◈ is found by using a variation of the basic gain formula:

$$V_{load} = V_{zener}\frac{R_2 + R_3}{R_3}$$

(8.3.1)

By choosing the appropriate ratios of resistors along with a suitable Zener, a wide range of output potentials may be achieved. Note that the op amp is fed from the nonregulated input. Any ripple that exists on this line will hardly affect the op amp's function. The ripple will be reduced by the op amp's PSRR (power supply rejection ratio). The Zener bias resistor, ◈1 , is chosen so that it sets up a current that will guarantee Zener conduction. This is usually in the low milliamp range.

Example 8.3.1

Determine the output of Figure 8.3.1 if ◈1=5◈Ω , ◈2=20◈Ω , ◈3=10◈Ω and ◈◈◈◈◈◈=3.9◈ . Assume that the filtered input is 20 V DC with no more than 3 V peak-to-peak ripple.

First, note that the input signal varies between a minimum of 18.5 V to a maximum of 21.5 V (20 V ± 1.5V peak). As long as the circuit doesn't try to produce more than the minimum input voltage and as long as all current and power dissipation limits are obeyed, everything should function correctly.

$$V_{load} = V_{zener}\frac{R_2 + R_3}{R_3}$$

$$V_{load} = 3.9V \times \frac{20k + 10k}{10k}$$

$$V_{load} = 11.7V$$

Due to the◈◈◈drop, the op amp needs to produce about 0.7 V more than this. Also, note that the Zener current may now be found.

$$I_{zener} = \frac{V_{load} - V_{zener}}{R_1}$$

$$I_{zener} = \frac{11.7V - 3.9V}{5k}$$

$$I_{zener} = 1.56mA$$

This is a reasonable value.

Example 8.3.2

Determine the power dissipation for ◈1 and the circuit efficiency for Example 8.3.1 if the effective load resistance is 20 Ω .

First, ◈◈◈◈◈ must be determined.

$$I_{load} = \frac{V_{load}}{R_{load}}$$

$$I_{load} = \frac{11.7V}{20}$$

$$I_{load} = 0.585A$$

The dissipation of ◈1 is the product of the current passing through it and the voltage across it. The current through ◈1 is the load current. The voltage across ◈1 is the difference between the filter's output and the load voltage. Because the filter's output contains a relatively small AC signal riding on DC, the average will equal the DC value, in this case, 20 V.

$$V_{CE} = V_C - V_E$$
$$V_{CE} = 20V11.7V$$
$$V_{CE} = 8.3V$$
$$P_D = I_C V_{CE}$$
$$P_D = 0.585A \times 8.3V$$
$$P_D = 4.86W$$

So, the pass transistor must dissipate 4.86 W and tolerate a current of 0.585 amps. The maximum collector-emitter voltage will occur at the peak input of 21.5 V. Therefore, the maximum differential is 9.8 V.

Finally, note that a 0.585 amp output would require a minimum ◈ of

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

$$\beta = \frac{585mA}{20mA}$$

$$\beta = 29.25$$

This assumes that the op amp can produce 20 mA and also ignores the small Zener and voltage divider currents at the output.

As far as efficiency is concerned, the input power and the load power need to be calculated. For load power,

$$P_{load} = I_{load} V_{load}$$
$$P_{load} = .585A \times 11.7V$$
$$P_{load} = 6.844W$$

Ignoring the current requirements of the op amp, Zener and ◈2/◈3 divider, the supplied current equals 0.585 A. The average input voltage is 20 V.

$$P_{in} = I_{in} V_{in}$$
$$P_{in} = 0.585A \times 20V$$
$$P_{in} = 11.7W$$

Efficiency, ◈ , is defined as the ratio of useful output to required input, so,

$$\eta = \frac{P_{load}}{P_{in}}$$

$$\eta = \frac{6.844W}{11.7W}$$

$$\eta = 0.585 \; or \; 58.5\%$$

Therefore, 41.5% of the input power is wasted. In order to minimize the waste and maximize the efficiency, the input/output differential voltage needs to be as small as possible. For proper operation of the op amp and pass transistor, this usually means a differential lower than 2 to 3 V is impossible. Consequently, this form of regulation is very inefficient when the power supply must produce low output voltage levels.

COMPUTER SIMULATION

A Multisim simulation of an op amp-based linear regulator, such as the one in Figure 8.3.1 , is shown in Figure 8.3.2 . The input signal is 20 V average, with a 120 Hz 2 V peak sine wave riding on it representing ripple. A 741 op amp is used as the comparison element along with a generic transistor for the pass device. The reference is a 5.2 V Zener diode. Given the circuit values, a manual calculation shows

$$V_{out} = V_{zener}(1 + \frac{R_f}{R_i})$$

$$V_{out} = 5.2V \times (1 + \frac{14k}{11k})$$

$$V_{out} = 11.8V$$

The output plot indicates a constant output at approximately 12.6 V. The discrepancy can be attributed to the non-ideal nature of the Zener diode (e.g., its internal resistance). Indeed, the initial solution indicates that the Zener potential (node 5) is approximately 5.5 V. A more accurate specification of the Zener diode (in particular, the BV and IBV parameters) would result in a much closer prediction. Another way of looking at this is to say that the 5.2 V Zener specification is too far below the actual Zener bias point for utmost accuracy (i.e., the manual calculation is somewhat sloppy and is, therefore, the one in error). This minor discrepancy aside, the time-domain plot does show the stability of the output signal in spite of the sizable input variations.

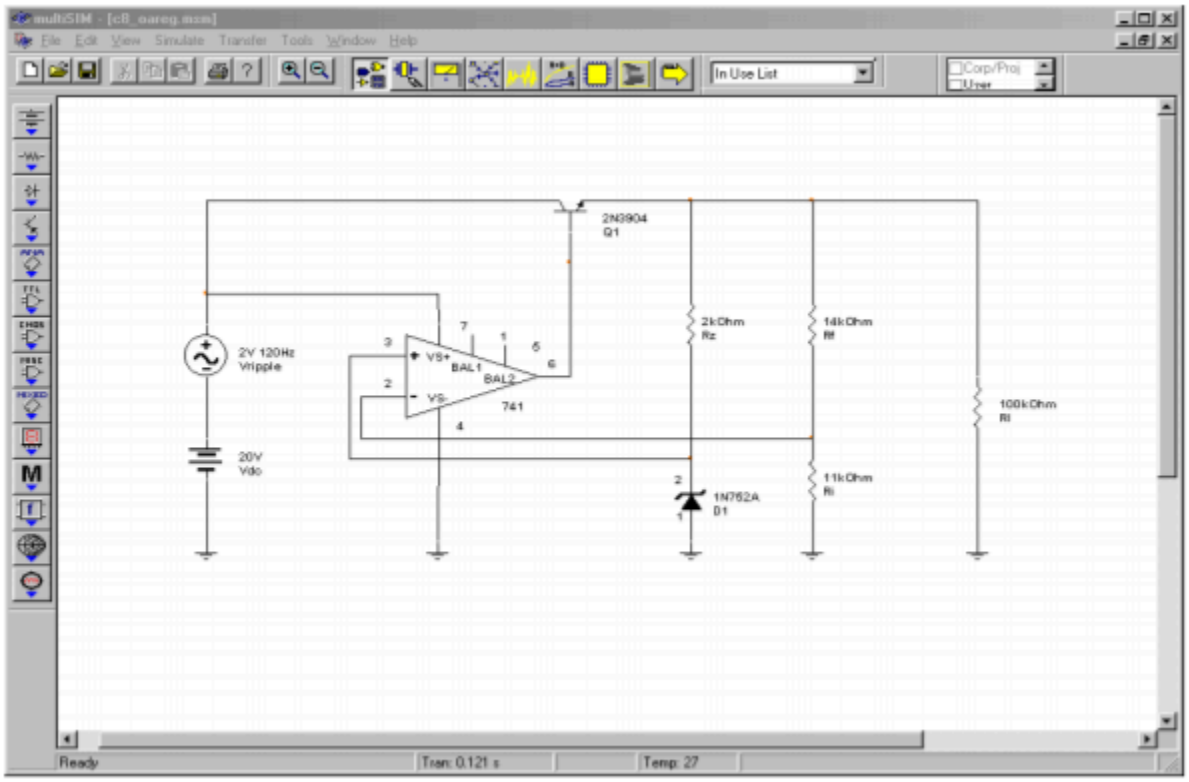


Figure 8.3.2 : Op amp regulator in Multisim.

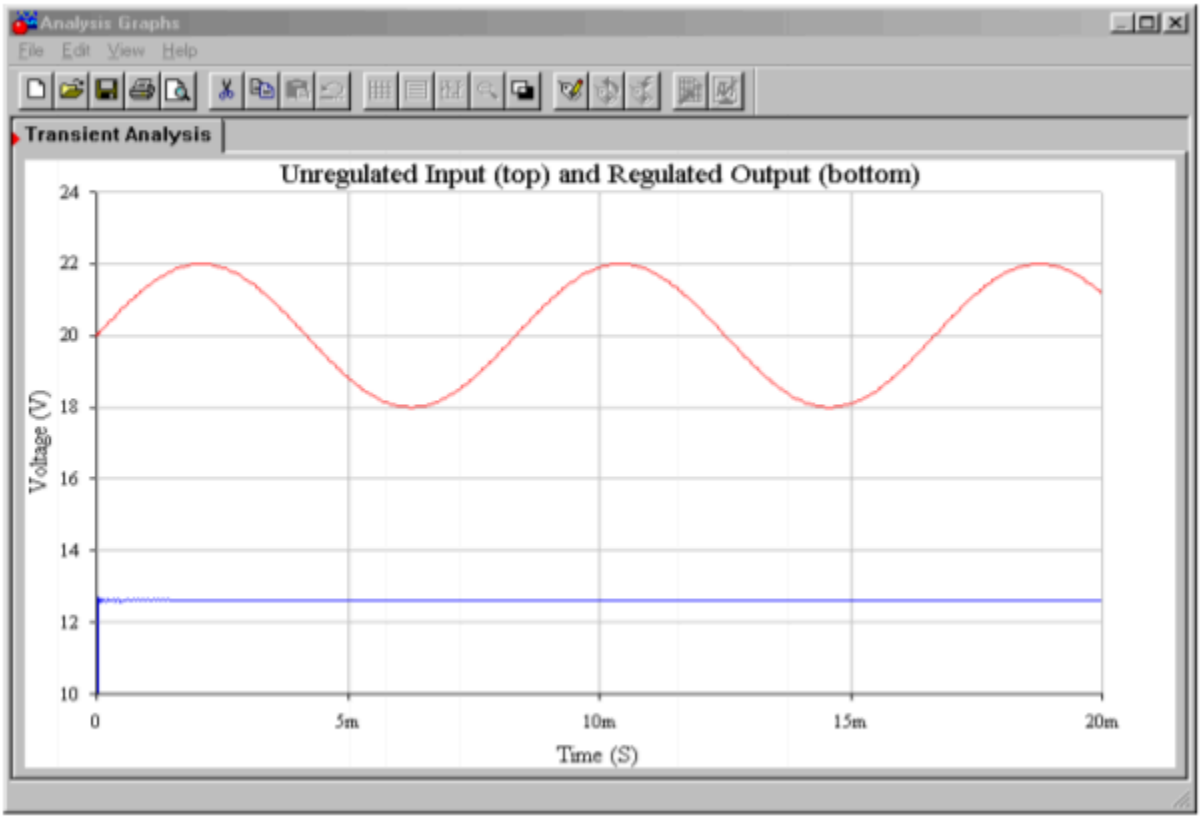


Figure 8.3.2 : Input and output waveforms from Multisim.

THREE TERMINAL DEVICES

In an effort to make the designer's job ever easier, manufacturers provide circuits such as the one shown in Figure 8.3.1 in a single package. If you notice, the circuit really needs only three pins with which to connect to the outside world: input from the filter, ground, and output. These devices are commonly known as three pin regulators – hardly a creative name tag, but descriptive at least. Several different devices are available for varying current and power dissipation demands.

A typical "3 pin" device family is the LM340-XX/LM78XX and LM360- XX/LM79XX. The LM340-XX/LM78XX series is for positive outputs, while the LM360-XX/LM79XX series produces negative outputs. The "XX" indicates the rated load voltage. For example, the LM340-05 is a +5 V regulator, while the LM7912 is a -12 V unit. The most popular sizes are the 5, 12, and 15 V units. For simplicity, the series will be referred to as the LM78XX from here on.

A data sheet for the LM78XX series is shown in Figure 8.3.3 . This regulator comes in several variants including TO-3, TO-220, and surface-mount versions. The TO-3 case version offers somewhat higher power dissipation capability. Output currents in excess of 1 A are available. As a side note, for lighter loads with lower current demands, the LM78LXX may be used. This regulator offers a 100 mA output and comes in a variety of packages.

LM340/LM7800C
Electrical Characteristics (Note 4)

$0°C \le T_J \le +125°C$ unless otherwise specified

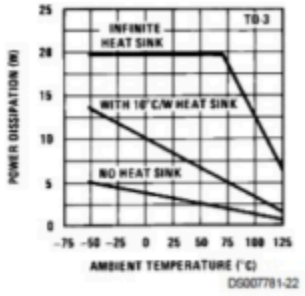
Symbol	Output Voltage			5V			12V			15V			Units
	Input Voltage (unless otherwise noted)			10V			19V			23V			
	Parameter	Conditions		Min	Typ	Max	Min	Typ	Max	Min	Typ	Max	
V_O	Output Voltage	$T_J = 25°C$, $5\ mA \le I_O \le 1A$		4.8	5	5.2	11.5	12	12.5	14.4	15	15.6	V
		$P_D \le 15W$, $5\ mA \le I_O \le 1A$		4.75		5.25	11.4		12.6	14.25		15.75	V
		$V_{MIN} \le V_{IN} \le V_{MAX}$		$(7.5 \le V_{IN} \le 20)$			$(14.5 \le V_{IN} \le 27)$			$(17.5 \le V_{IN} \le 30)$			V
ΔV_O	Line Regulation	$I_O = 500\ mA$	$T_J = 25°C$		3	50		4	120		4	150	mV
			ΔV_{IN}	$(7 \le V_{IN} \le 25)$			$(14.5 \le V_{IN} \le 30)$			$(17.5 \le V_{IN} \le 30)$			V
			$0°C \le T_J \le +125°C$			50			120			150	mV
			ΔV_{IN}	$(8 \le V_{IN} \le 20)$			$(15 \le V_{IN} \le 27)$			$(18.5 \le V_{IN} \le 30)$			V
		$I_O \le 1A$	$T_J = 25°C$			50			120			150	mV
			ΔV_{IN}	$(7.5 \le V_{IN} \le 20)$			$(14.6 \le V_{IN} \le 27)$			$(17.7 \le V_{IN} \le 30)$			V
			$0°C \le T_J \le +125°C$			25			60			75	mV
			ΔV_{IN}	$(8 \le V_{IN} \le 12)$			$(16 \le V_{IN} \le 22)$			$(20 \le V_{IN} \le 26)$			V
ΔV_O	Load Regulation	$T_J = 25°C$	$5\ mA \le I_O \le 1.5A$		10	50		12	120		12	150	mV
			$250\ mA \le I_O \le 750\ mA$			25			60			75	mV
		$5\ mA \le I_O \le 1A$, $0°C \le T_J \le +125°C$				50			120			150	mV
I_Q	Quiescent Current	$I_O \le 1A$	$T_J = 25°C$			8			8			8	mA
			$0°C \le T_J \le +125°C$			8.5			8.5			8.5	mA
ΔI_Q	Quiescent Current Change	$5\ mA \le I_O \le 1A$				0.5			0.5			0.5	mA
		$T_J = 25°C$, $I_O \le 1A$				1.0			1.0			1.0	mA
		$V_{MIN} \le V_{IN} \le V_{MAX}$		$(7.5 \le V_{IN} \le 20)$			$(14.8 \le V_{IN} \le 27)$			$(17.9 \le V_{IN} \le 30)$			V
		$I_O \le 500\ mA$, $0°C \le T_J \le +125°C$				1.0			1.0			1.0	mA
		$V_{MIN} \le V_{IN} \le V_{MAX}$		$(7 \le V_{IN} \le 25)$			$(14.5 \le V_{IN} \le 30)$			$(17.5 \le V_{IN} \le 30)$			V
V_N	Output Noise Voltage	$T_A = 25°C$, $10\ Hz \le f \le 100\ kHz$			40			75			90		μV
$\frac{\Delta V_{IN}}{\Delta V_{OUT}}$	Ripple Rejection	$f = 120\ Hz$	$I_O \le 1A$, $T_J = 25°C$	62	80		55	72		54	70		dB
			or $I_O \le 500\ mA$, $0°C \le T_J \le +125°C$	62			55			54			dB
		$V_{MIN} \le V_{IN} \le V_{MAX}$		$(8 \le V_{IN} \le 18)$			$(15 \le V_{IN} \le 25)$			$(18.5 \le V_{IN} \le 28.5)$			V
R_O	Dropout Voltage	$T_J = 25°C$, $I_O = 1A$			2.0			2.0			2.0		V
	Output Resistance	$f = 1\ kHz$			8			18			19		mΩ
	Short-Circuit Current	$T_J = 25°C$			2.1			1.5			1.2		A
	Peak Output Current	$T_J = 25°C$			2.4			2.4			2.4		A
	Average TC of V_{OUT}	$0°C \le T_J \le +125°C$, $I_O = 5\ mA$			−0.6			−1.5			−1.8		mV/°C
V_{IN}	Input Voltage Required to Maintain Line Regulation	$T_J = 25°C$, $I_O \le 1A$		7.5			14.6			17.7			V

Note 1: Absolute Maximum Ratings are limits beyond which damage to the device may occur. Operating Conditions are conditions under which the device functions but the specifications might not be guaranteed. For guaranteed specifications and test conditions see the Electrical Characteristics.

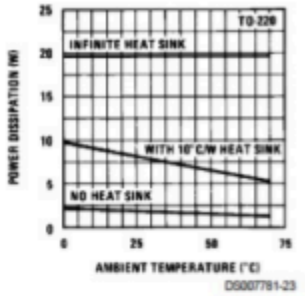
Figure 8.3.3 : LM78XX data sheet.

Typical Performance Characteristics

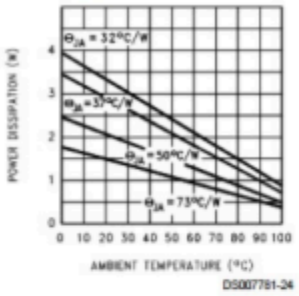
Maximum Average Power Dissipation



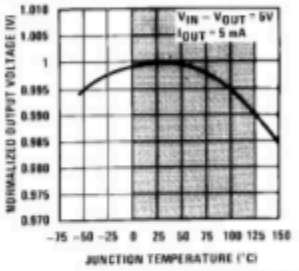
Maximum Average Power Dissipation



Maximum Power Dissipation (TO-263) (See Note 2)

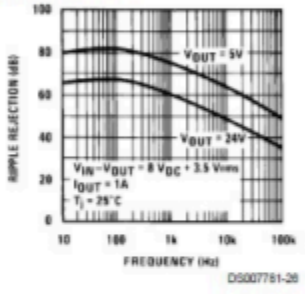


Output Voltage (Normalized to 1V at T_j = 25°C)

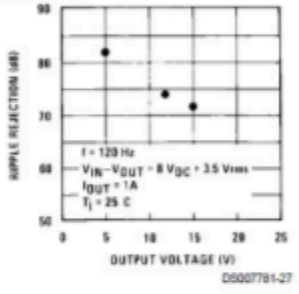


Note: Shaded area refers to LM340A/LM340, LM7805C, LM7812C and LM7815C.

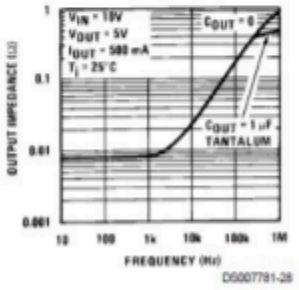
Ripple Rejection



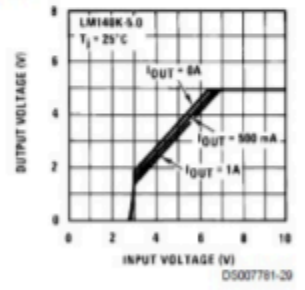
Ripple Rejection



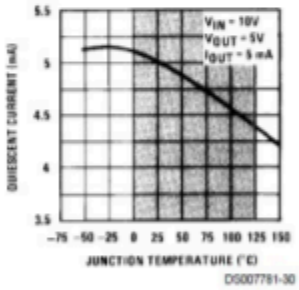
Output Impedance



Dropout Characteristics

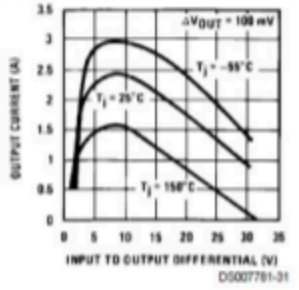


Quiescent Current

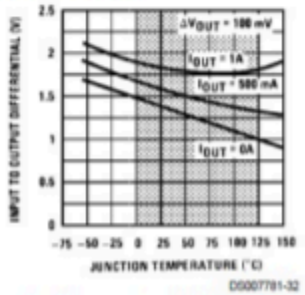


Note: Shaded area refers to LM340A/LM340, LM7805C, LM7812C and LM7815C.

Peak Output Current



Dropout Voltage



Note: Shaded area refers to LM340A/LM340, LM7805C, LM7812C and LM7815C.

Quiescent Current

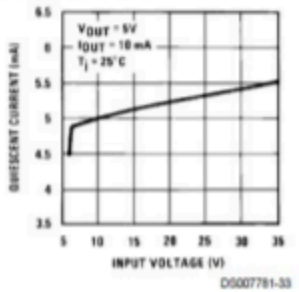


Figure 8.3.3◈: LM78XX data sheet (continued). Reprinted from of Texas Instruments

The practical power dissipation limit of the LM78XX series, like any power device, is highly dependent on the type of heat sink used. The device graphs show that it can be as high as 20 W with the TO-3 case. The typical output voltage is within approximately 5 percent of the nominal value. This indicates the inherent accuracy of the IC, and is not a measure of its regulation abilities. Figures for both load and line regulation are given in the data sheet. For a variance in the line voltage of over 2:1, we can see that the output voltage varies by no more than 1 percent of the nominal output. Worst-case load regulation is just as good, showing no more than 1 percent deviation for a load current change from 5 mA up to 1.5 A. The regulator can also be seen to draw very little standby current, only 8.5 mA over a wide temperature range.

The output of the LM78XX series is fairly clean. The LM7815 shows an output noise voltage of 90 μV, typically, and an average ripple rejection of 70 dB. This means that ripple presented to the input of the regulator is reduced by 70 dB at the output. Because 70 dB translates to a factor of over 3000, this means that an input ripple signal of 1 V will be reduced to less than one-third of a millivolt at the output. It is worthwhile to note that this figure is frequency-dependent, as shown by the Ripple Rejection graph. Fortunately, the maximum value occurs in the desired range of 60 to 120 Hz.

The last major item in the data sheet is an extremely important one: Input Voltage Required To Maintain Line Regulation. This value is approximately 2.5 to 2.7 V greater than the nominal rating of the regulator. Without this headroom, the regulator will cease to function properly. Although it is desirable to keep the input/output differential voltage low in order to decrease device power dissipation and maximize efficiency, too low of a value will cripple the regulator. As seen in the Peak Output Current graph, the highest load currents are achieved when the differential is in the 5 to 10 V range.

In all cases, the devices offer thermal shutdown and output short-circuit protection. The maximum input voltage is limited to 35 V. Besides the LM78XX series, special high-current types with outputs in the 3 A and higher range are available.

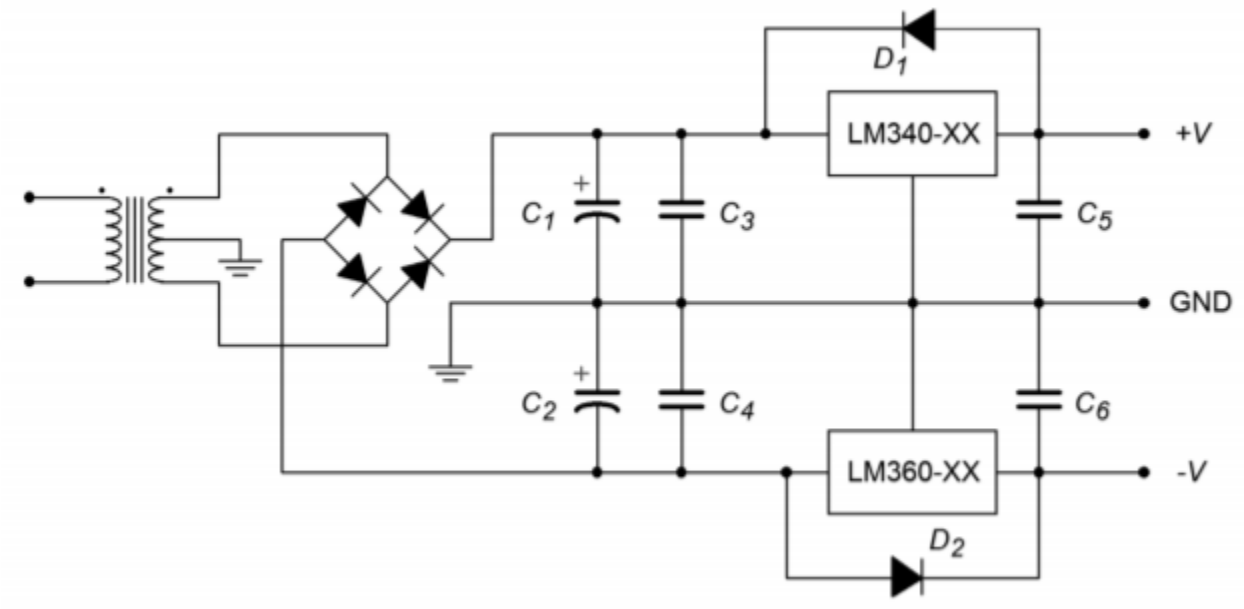


Figure 8.3.4 : Dual supply for op amp circuits.

As an example, a bipolar power supply is shown in Figure 8.3.4 . A center-tapped transformer is used

to generate the required positive/negative polarities. ◈1 and ◈2 serve as the filter capacitors. The positive peak across each capacitor should not exceed 35 V, or the regulators may be damaged. The minimum voltage across the capacitors should be at least 3 V greater than the desired output level. The exact size of the capacitor depends on the amount of load current expected. In a similar manner, the magnitude of the load current will determine which case style will be used. Capacitors ◈3 and ◈4 are 220 nF units and are only required if the regulators are located several inches or more from the filter capacitors. If their use is required, ◈3 and ◈4 must be located very close the regulator. ◈5 and ◈6 are used to improve the transient response of the regulator. They are not used for filtering or for circuit stabilization. Through common misuse, ◈5 and ◈6 are often called "stability capacitors", although this is not their function. Finally, diodes ◈1 and ◈2 protect the regulators from output over-voltage conditions, such as those experienced with inductive loads.

As you can see, designing moderate fixed voltage regulated power supplies with this template can be a very straightforward exercise. Once the basic power supply is configured, all that needs to be added are the regulators and a few capacitors and diodes. Quite simply, the regulator is "tacked on" to a basic capacitor-filtered unregulated supply. For a simple single polarity supply, the additional components can be reduced to a single regulator IC, the extra capacitors and diode being ignored.

CURRENT BOOSTING

For very high output currents, it is possible to utilize external pass transistors to augment the basic linear regulator ICs. An example is shown in Figure 8.3.5 . When the input current rises to a certain level, the voltage drop across ◈1 will be large enough to turn on transistor ◈1 . When ◈1 starts to conduct, a current will flow through ◈3 . As this current rises, the potential across ◈3 will increase to the point where the power transistor ◈2 will turn on. ◈2 will handle any further increases in load current. Normally, ◈1 is set so that the regulator is running at better than one-half of its maximum rating before switch over occurs. A typical value would be 22 Ω . For example, if switch over occurs at 1 A, the regulator will supply all of the load's demand up to 1 A. If the load requires more than 1 A, the regulator will supply 1 A, and the power transistor will supply the rest. Circuits of this type often need a minimum load current to function correctly. This "bleed-off" can be achieved by adding a single resistor in parallel with the load.

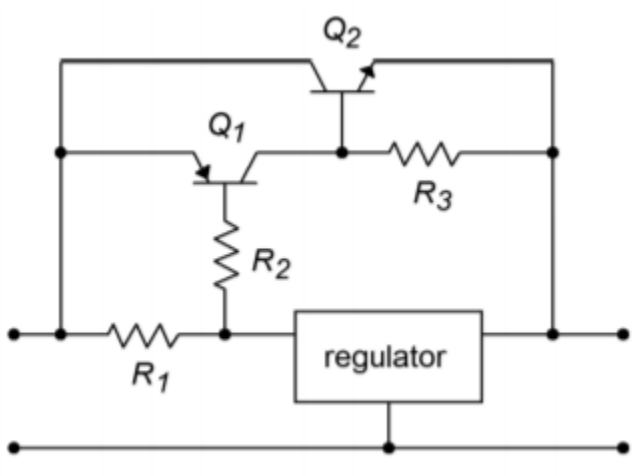


Figure 8.3.5 : Pass transistors for increased load current.

Another possibility for increased output current is by paralleling devices and adding small ballast resistors. An example of this is shown in Figure 8.3.6 . The ballast resistors are used to create local feedback. This reduces current hogging and forces the regulators to equally share the load current. (This is the same technique that is commonly used in high-power amplifiers so that multiple output transistors may be connected in parallel.) The size of the ballast resistors is rather low, usually 0.5 Ω or less.

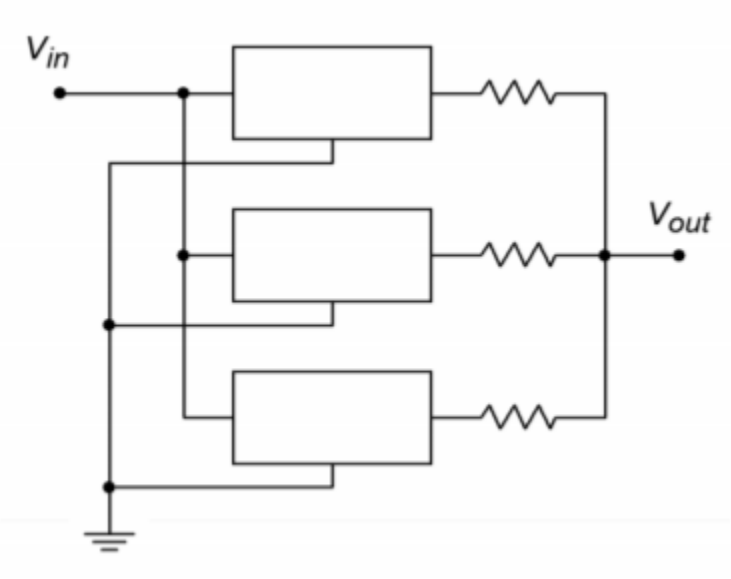


Figure 8.3.6 : Parallel devices for increased load current

LOW DROPOUT REGULATORS

Low dropout regulators (usually shortened to LDO) are a special subclass of ordinary linear regulators. Generally, they operate the same way, but with one major exception. Unlike normal linear regulators, LDOs do not require a large input-output differential voltage of several volts. Instead, LDOs will regulate with a differential as little as a few tenths of a volt. This minimum input-output differential is known as the dropout voltage. With a lowered headroom requirement, the LDO is much more efficient than the standard regulator, especially with low voltage outputs. An example is the LM2940. Like the LM78XX, this regulator is available at many popular output potentials, including 5, 12, and 15 V. It is rated for a 1 A output current. At maximum current, the dropout voltage is typically 0.5 V. At an output of 100 mA, the dropout voltage is typically 110 mV.

PROGRAMMABLE AND TRACKING REGULATORS

Along with the simple three-pin fixed regulators, a number of adjustable or programmable devices are available. Some devices also include features such as programmable current limiting. Also, it is possible to configure multiple regulators so that they track, or follow, each other.

One popular adjustable regulator is the LM317. This device is functionally similar to the 340 series discussed in the previous section. In essence, it has an internal reference of 1.25 V. By using an external voltage divider, a wide range of output potentials are available. The LM317 will produce a maximum current of 1.5 A and a maximum voltage of 37 V. A basic connection diagram is shown in Figure 8.3.7 . The ◈1/◈2 divider sets the output voltage according to the formula

$$V_{out} = 1.25V(1 + \frac{R_2}{R_1}) + I_{adj}R_2$$

◈◈◈◈ is the current flowing through the bottom adjust pin. ◈◈◈◈ is about 100 μA, and being this small, may be ignored to a first approximation. ◈1 is set to 240 Ω . D1 serves as a protection device for output over-voltages as seen in the preceding section. The 10 μF capacitor is used to increase the ripple rejection of the device. The addition of this capacitor will increase ripple rejection by at least 10 dB. Diode ◈2 is used to prevent possible destructive discharges from the 10 μF capacitor. In practice, ◈2 is set to a fixed size resistor for static supplies, or is a potentiometer for user-adjustable supplies.

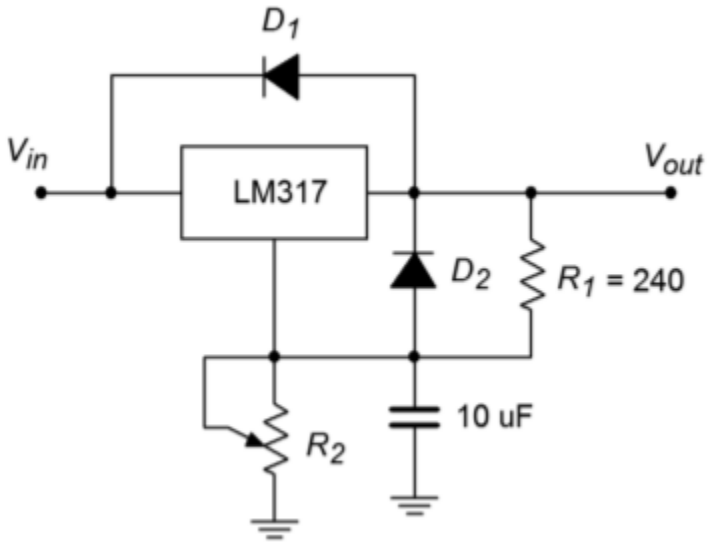


Figure 8.3.7 : Basic LM317 regulator.

Example 8.3.3

Determine a value for ◈2 so that the output is adjustable from a minimum of 1.25 V to a maximum of 15 V.

For the minimum value, ◈2 should be 0 Ω . Ignoring the effect of ◈◈◈◈ , the maximum value is found by

$$V_{out} = 1.25V(1 + \frac{R_2}{R_1})$$

$$R_2 = R_1 \times (\frac{V_{out}}{1.25V} - 1)$$

$$R_2 = 240 \times (\frac{15V}{1.25V} - 1)$$

$$R_2 = 2.64k$$

Normally, such a value is not available for stock potentiometers. A 2.5 k Ω unit may be readily available, so ◈1 may be reduced a bit in order to compensate.

COMPUTER SIMULATION

Figure 8.3.8 shows a simulation of the regulator designed in Example 8.3.3 . The LM117 model used is very similar to the LM317. The maximum potentiometer value of 2.5 k Ω is used here in order to see just how far off the design is from the 15 volt target. As in the earlier chapter simulation, a sine wave riding on a DC offset is used as the input to mimic the presence of ripple on the unregulated power source. The Transient Analysis of Multisim is used to plot the input and output waveforms. The circuit produces a very stable DC output as desired. Also, the level is only a few tenths of a volt shy of the desired 15 volt maximum, as expected. The simulation verifies the manual calculations quite well.

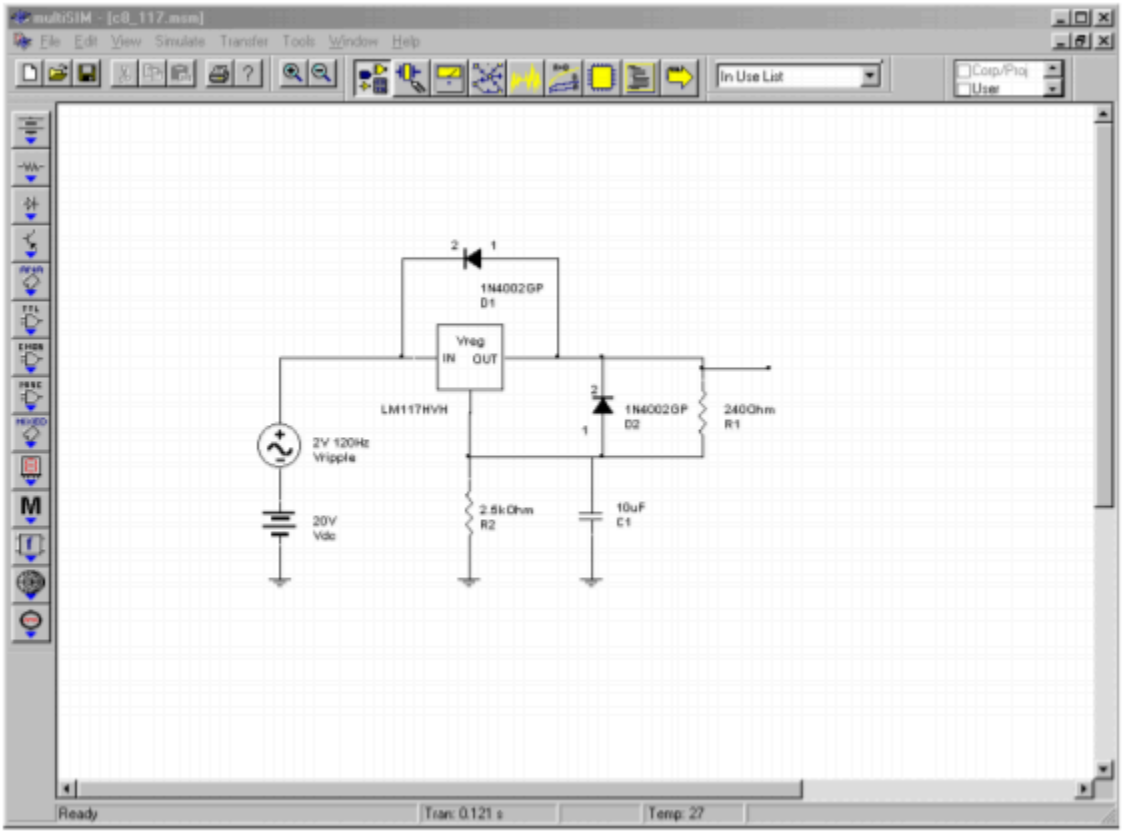


Figure 8.3.8 ◈ : LM317/LM117 simulation.

When using the LM317, if a minimum value greater than 1.25 V is needed, the pot may be placed in series with a fixed resistor. Alternately, precise preset values may be obtained through the use of a rotary switch and a bank of fixed resistors, as shown in Figure 8.3.9 . As a matter of fact, these techniques may also be applied to the fixed three-pin devices presented in the previous section. For example, a LM7805 may be used as a 15 V regulator just by adding an external divider network as shown in Figure 8.3.10 . You can think of the LM7805 as an LM317 with a 5 V reference. The exact values are not critical; what is important is that the ratio of the two resistors is 2:1.

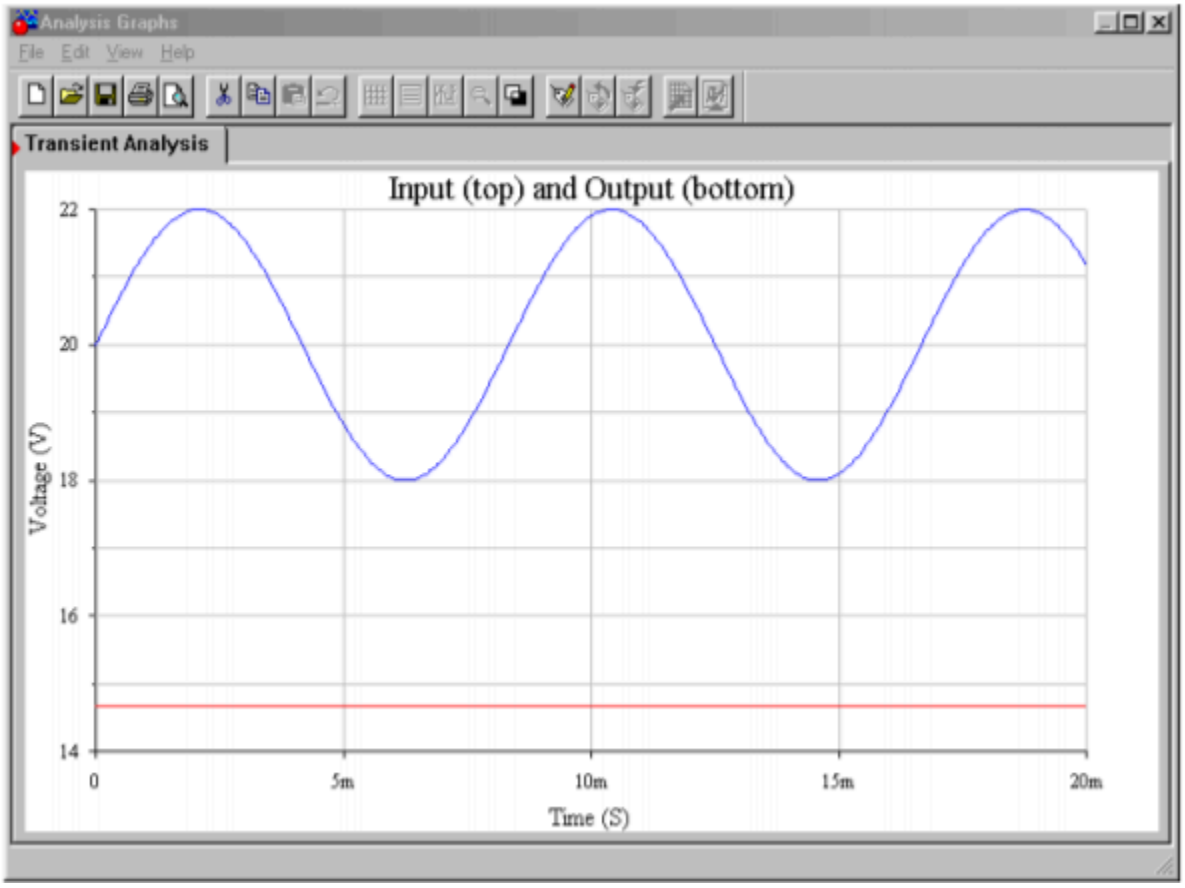


Figure 8.3.8 : Waveforms from Multisim.

One of the major problems with adjustable linear regulators is a "bottom end" limitation. It is easier for a regulator of this type to generate a high load current at a high load voltage than it is to generate a high load current at a low load voltage. The reason for this is that at low output voltages, the internal pass transistor sees a very high differential voltage. This results in very high power dissipation. If the device gets too hot, the thermal protection circuits may activate. From a user's standpoint, this means that the power supply may be able to produce a 15 V, 1 A output, but be unable to produce a 5 V, 1 A output.

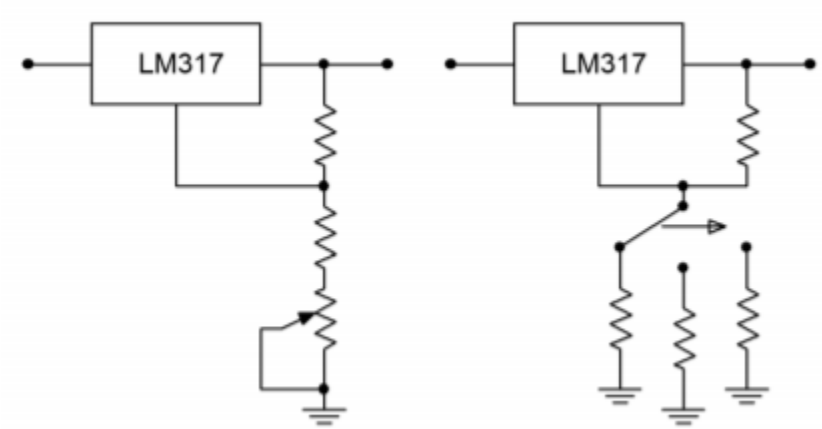


Figure 8.3.9 : Adjustable regulation.

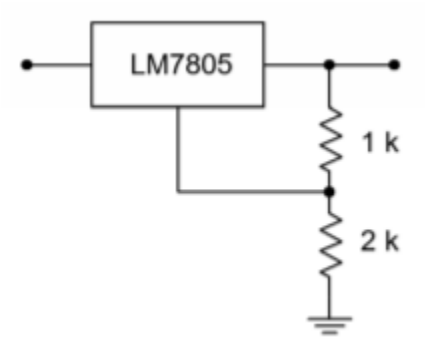


Figure 8.3.10 : Varying output voltage with a fixed regulator.

Another popular regulator IC is the LM723. This device is of a more modular design and allows for preset current limiting. The equivalent circuit of 723 is shown in Figure 8.3.11 . By itself, the LM723 is only capable of producing 150 mA. With the use of external pass transistors, an LM723-based regulator may produce several amps of load current. The internal reference is approximately 7.15 V. Because of this, two basic configurations exist: one for outputs below 7 V, and one for outputs above 7 V. These configurations and their output formulas are shown in Figure 8.3.12 .

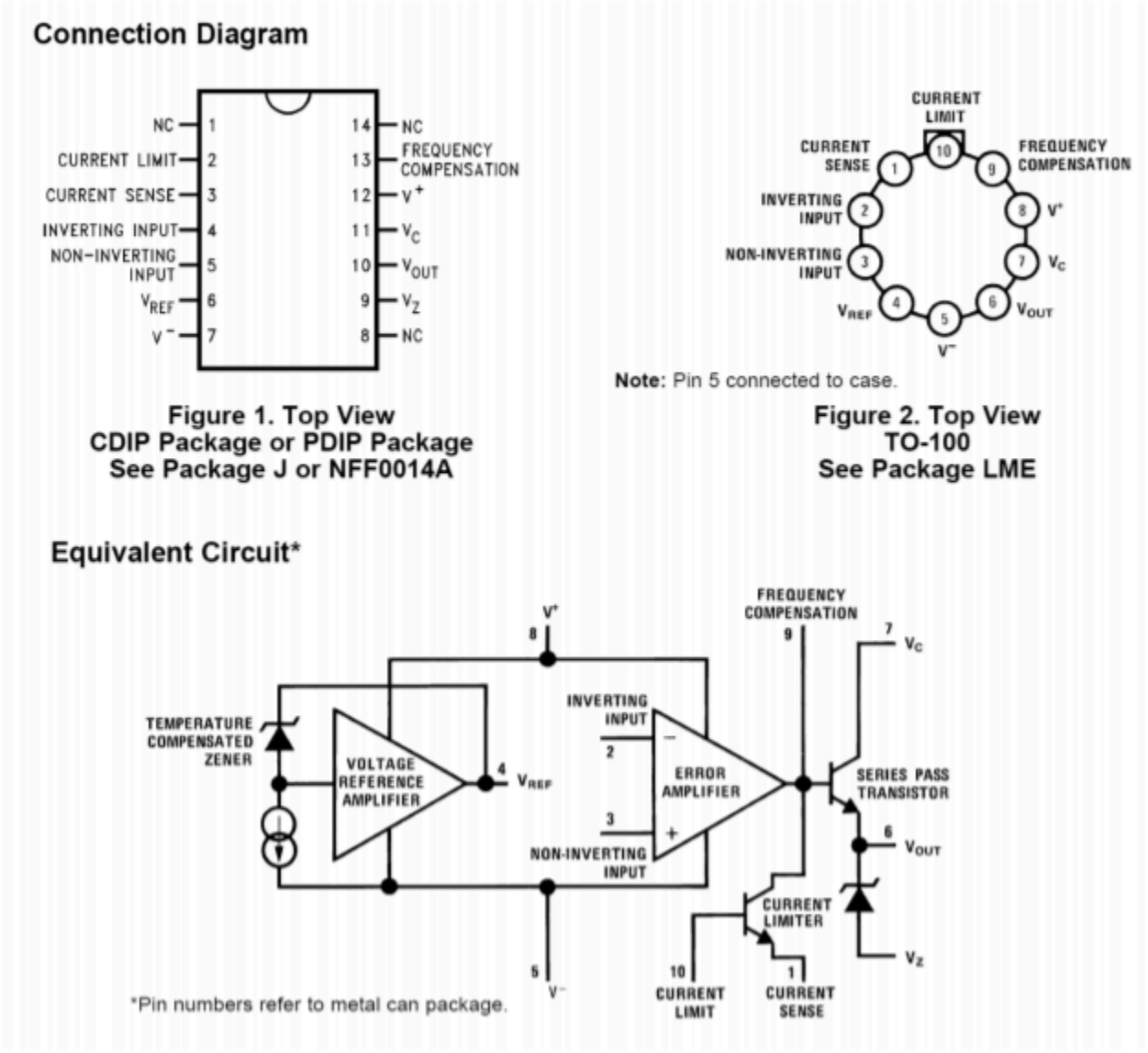


Figure 8.3.11 : LM723 equivalent circuit. Reprinted courtesy of Texas Instruments

By combining Figure 8.3.10 with Figure 8.3.11 and redrawing the results in Figures 8.3.13 and 8.3.13 you can see a strong resemblance to the basic op amp regulator encountered earlier. Figure 8.3.13 is used for outputs greater than the 7.15 V reference. As such, resistors 1 and 2 are used to set the gain of the internal amplifier (i.e., the amount by which the reference will be multiplied). Resistor 3 simply serves as DC bias compensation for 1 and 2 . For outputs less than the reference, 1 and 2 serve as a voltage divider, which effectively reduces the reference, whereas the amplifier operates with a gain of unity. Again, R3 serves as bias compensation. In short, the output voltage is once again determined by the reference voltage in conjunction with a pair of resistors. In both cases, the output current limit is set by

$$I_{limit} = \frac{V_{sense}}{R_{sc}}$$

Where is approximately 0.65 V at room temperature (for other temperatures, a graph is given in the LM723 data sheet). This Equation may be found on the manufacturer's data sheet, but may be readily derived by inspecting Figure 8.3.13 . Resistor is

placed in series with the load, and thus, the current through it is the load current (ignoring the small current required by the adjustment resistors). The current limit transistor is connected so that the voltage across ⍰⍰⍰ is applied to this transistor's base-emitter junction. The collector of the current limit transistor is connected to the base of the output pass transistor. If the output current rises to the point where the voltage across ⍰⍰⍰ exceeds approximately 0.65 V, the current limit transistor will turn on, thus shunting output drive current away from the output pass transistor. As you can see, the formula for ⍰⍰⍰⍰⍰⍰ is little more than the direct application of Ohm's Law. This scheme is similar to the method examined in Chapter Two to safely limit an op amp's output current.

Table 1. Resistor Values (kΩ) for Standard Output Voltage

Positive Output Voltage	Applicable Figures	Fixed Output ±5%		Output Adjustable ±10%(1)			Negative Output Voltage	Applicable Figures	Fixed Output ±5%		5% Output Adjustable ±10%		
	See(2)	R1	R2	R1	P1	R2			R1	R2	R1	P1	R2
+3.0	Figure 4, Figure 19, Figure 21, Figure 24, Figure 27 (Figure 19)	4.12	3.01	1.8	0.5	1.2	+100	Figure 22	3.57	102	2.2	10	91
+3.6	Figure 4, Figure 19, Figure 21, Figure 24, Figure 27 (Figure 19)	3.57	3.65	1.5	0.5	1.5	+250	Figure 22	3.57	255	2.2	10	240
+5.0	Figure 4, Figure 19, Figure 21, Figure 24, Figure 27 (Figure 19)	2.15	4.99	0.75	0.5	2.2	−6(3)	Figure 18, (Figure 25)	3.57	2.43	1.2	0.5	0.75
+6.0	Figure 4, Figure 19, Figure 21, Figure 24, Figure 27 (Figure 19)	1.15	6.04	0.5	0.5	2.7	−9	Figure 18, Figure 25	3.48	5.36	1.2	0.5	2.0
+9.0	Figure 17, Figure 19, (Figure 19, Figure 21, Figure 24, Figure 27)	1.87	7.15	0.75	1.0	2.7	−12	Figure 18, Figure 25	3.57	8.45	1.2	0.5	3.3
+12	Figure 17, Figure 19, (Figure 19, Figure 21, Figure 24, Figure 27)	4.87	7.15	2.0	1.0	3.0	−15	Figure 18, Figure 25	3.65	11.5	1.2	0.5	4.3
+15	Figure 17, Figure 19, (Figure 19, Figure 21, Figure 24, Figure 27)	7.87	7.15	3.3	1.0	3.0	−28	Figure 18, Figure 25	3.57	24.3	1.2	0.5	10
+28	Figure 17, Figure 19, (Figure 19, Figure 21, Figure 24, Figure 27)	21.0	7.15	5.6	1.0	2.0	−45	Figure 23	3.57	41.2	2.2	10	33
+45	Figure 22	3.57	48.7	2.2	10	39	−100	Figure 23	3.57	97.6	2.2	10	91
+75	Figure 22	3.57	78.7	2.2	10	68	−250	Figure 23	3.57	249	2.2	10	240

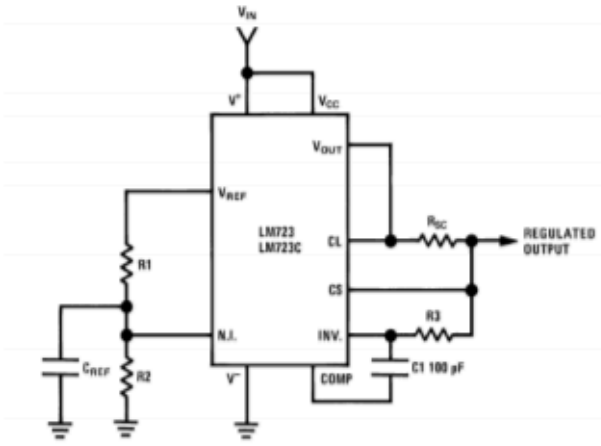
(1) Replace R1/R2 in figures with divider shown in Figure 28.
(2) Figures in parentheses may be used if R1/R2 divider is placed on opposite input of error amp.
(3) V⁻ and V_{CC} must be connected to a +3V or greater supply.

Table 2. Formulae for Intermediate Output Voltages

Outputs from +2 to +7 volts	Outputs from +4 to +250 volts	Current Limiting
(Figure 4 Figure 19 Figure 20 Figure 21 Figure 24 Figure 27	(Figure 22)	
$V_{OUT} = \left(V_{REF} \times \frac{R2}{R1 + R2}\right)$	$V_{OUT} = \left(\frac{V_{REF}}{2} \times \frac{R2 - R1}{R1}\right)$; R3 = R4	$I_{LIMIT} = \frac{V_{SENSE}}{R_{SC}}$
Outputs from +7 to +37 volts	**Outputs from −6 to −250 volts**	**Foldback Current Limiting**
(Figure 17 Figure 19 Figure 20 Figure 21 Figure 24 Figure 27)	(Figure 18 Figure 23 Figure 25)	$I_{KNEE} = \left(\frac{V_{OUT}\,R3}{R_{SC}\,R4} + \frac{V_{SENSE}\,(R3 + R4)}{R_{SC}\,R4}\right)$
$V_{OUT} = \left(V_{REF} \times \frac{R1 + R2}{R2}\right)$	$V_{OUT} = \left(\frac{V_{REF}}{2} \times \frac{R1 + R2}{R1}\right)$; R3 = R4	$I_{SHORT\ CKT} = \left(\frac{V_{SENSE}}{R_{SC}} \times \frac{R3 + R4}{R4}\right)$

Figure 8.3.12 : LM723 "hook ups". Reprinted courtesy of Texas Instrutments

Typical Application

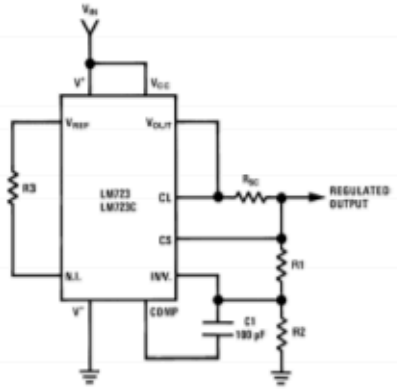


Note: R3 = $\frac{R1\,R3}{R1 + R2}$
for minimum temperature drift.

Typical Performance

Regulated Output Voltage	5V
Line Regulation (ΔV_{IN} = 3V)	0.5mV
Load Regulation (ΔI_L = 50 mA)	1.5mV

Figure 4. Basic Low Voltage Regulator (V_{OUT} = 2 to 7 Volts)

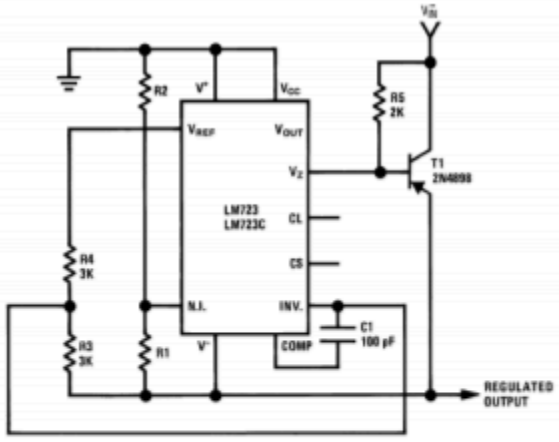


Note: R3 = $\frac{R1\,R2}{R1 + R2}$
for minimum temperature drift.
R3 may be eliminated for minimum component count.

Typical Performance

Regulated Output Voltage	15V
Line Regulation (ΔV_{IN} = 3V)	1.5 mV
Load Regulation (ΔI_L = 50 mA)	4.5 mV

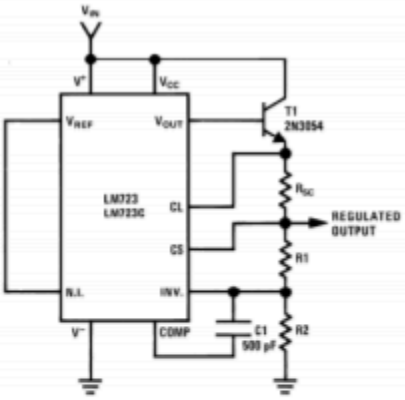
Figure 17. Basic High Voltage Regulator (V_{OUT} = 7 to 37 Volts)



Typical Performance

Regulated Output Voltage	−15V
Line Regulation (ΔV_{IN} = 3V)	1 mV
Load Regulation (ΔI_L = 100 mA)	2 mV

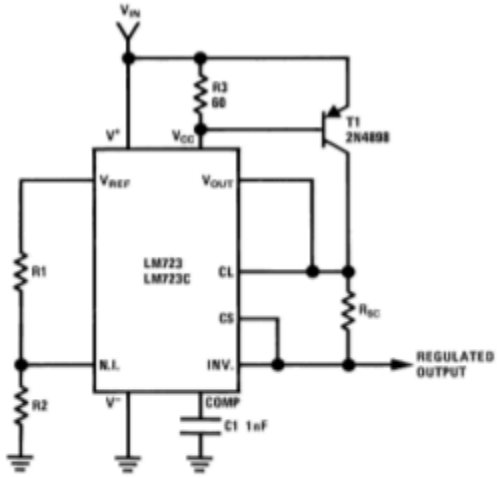
Figure 18. Negative Voltage Regulator



Typical Performance

Regulated Output Voltage	+15V
Line Regulation (ΔV_{IN} = 3V)	1.5 mV
Load Regulation (ΔI_L = 1A)	15 mV

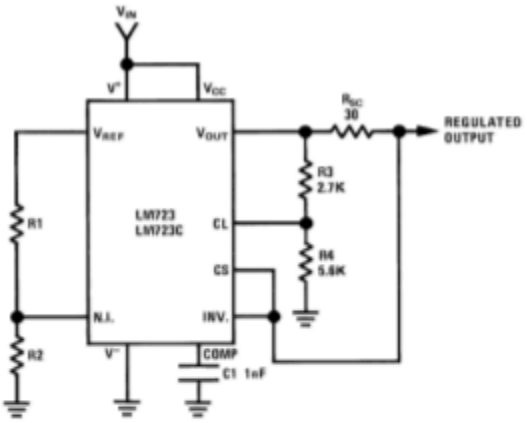
Figure 19. Positive Voltage Regulator (External NPN Pass Transistor)



Typical Performance

Regulated Output Voltage	+5V
Line Regulation (ΔV_{IN} = 3V)	0.5 mV
Load Regulation (ΔI_L = 1A)	5 mV

Figure 20. Positive Voltage Regulator (External PNP Pass Transistor)

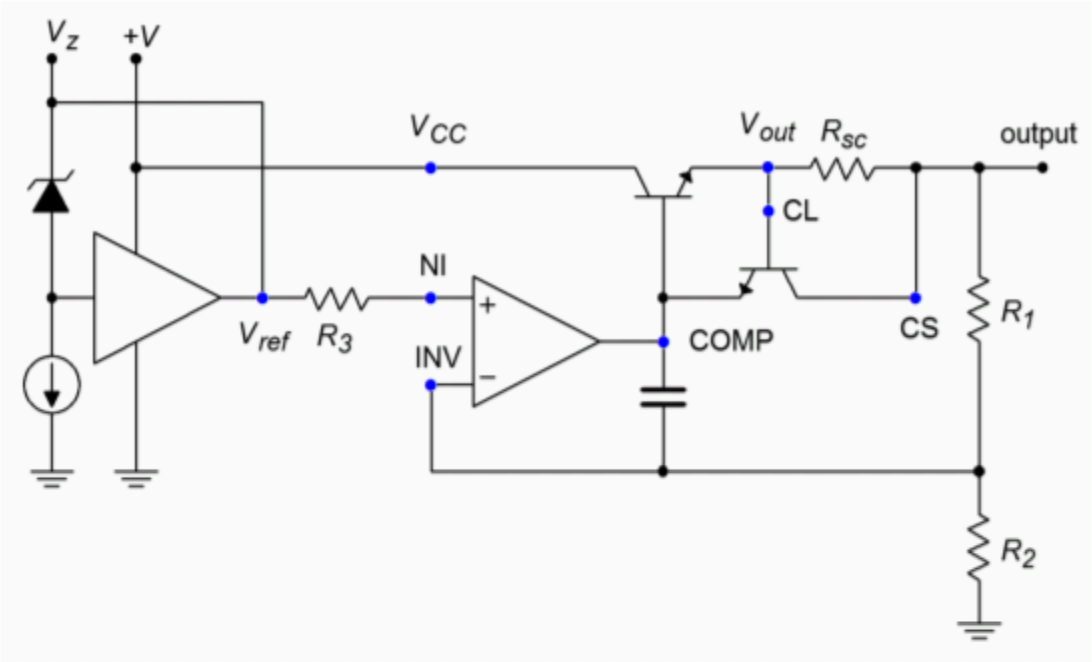


Typical Performance

Regulated Output Voltage	+5V
Line Regulation (ΔV_{IN} = 3V)	0.5 mV
Load Regulation (ΔI_L = 10 mA)	1 mV
Short Circuit Current	20 mA

Figure 21. Foldback Current Limiting

Figure 8.3.12 (continued): LM723 "hook ups". Reprinted courtesy of Texas Instrutments



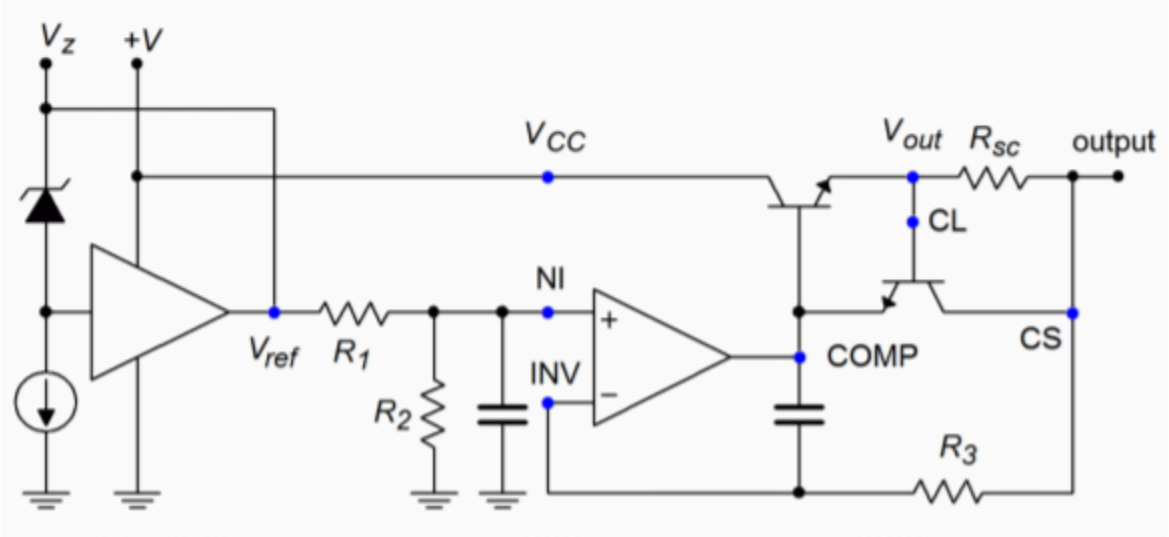


Figure 8.3.13 : Two basic configurations of an LM723 regualtor. a. Output > 7.15 volts (top). b. Output < 7.15 volts (bottom).

Example 8.3.4

Design a +12 V regulator using the LM723, with a current limit of 50 mA.

The basic form for this is the version shown in 8.3.12.2 . The appropriate output Equation is

$$V_{out} = V_{ref}\frac{R_1 + R_2}{R_2}$$

Choosing an arbitrary value for ◈2 of 10 k Ω , and then solving for ◈1 ,

$$R_1 = \frac{V_{out}}{V_{ref}} R_2 - R_2$$

$$R_1 = \frac{12V}{7.15V} 10k - 10k$$

$$R_1 = 6.78k$$

For the current sense resistor,

$$I_{limit} = \frac{V_{sense}}{R_{sc}}$$

$$R_{sc} = \frac{V_{sense}}{I_{limit}}$$

$$R_{sc} = \frac{0.65V}{50mA}$$

$$R_{sc} = 13\Omega$$

Finally, for minimum temperature drift, ◈3 is included, and set to ◈1||◈2

$$R_3 = R_1||R_2$$

$$R_3 = 10k||6.78k$$

$$R_3 = 4.04k$$

The completed circuit is shown in Figure 8.3.14 .

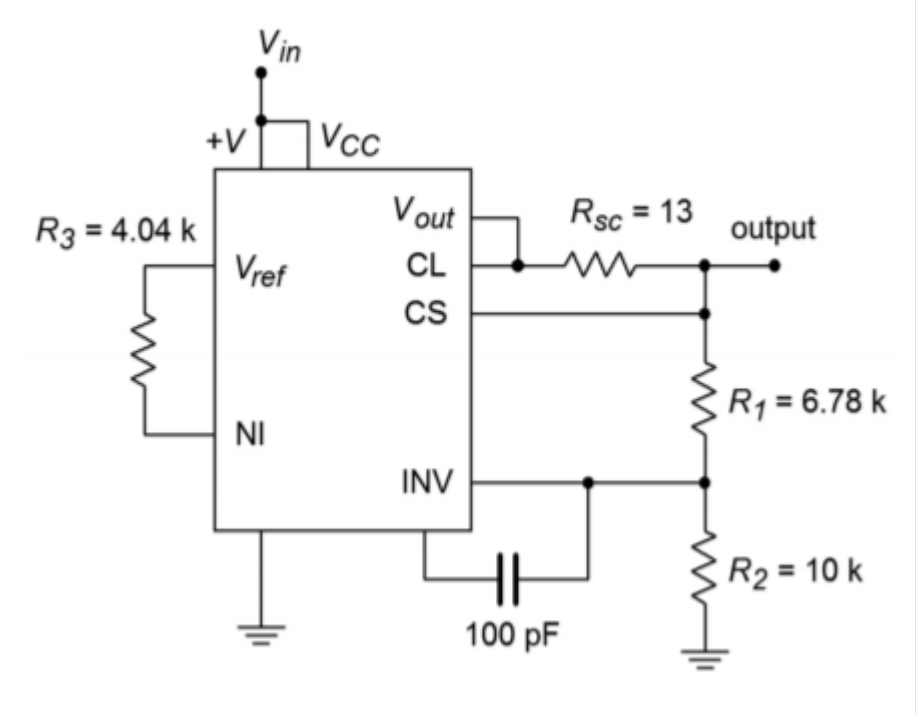


Figure 8.3.14 : Completed 12 volt circuit for Example 8.3.4 .

Example 8.3.5

Using the LM723, design a continuously adjustable 2V to 5V supply, with a current limit of 1.0 A.

The basic form for this is the version shown in Figure 8.3.12.1 . The appropriate output Equation is

$$V_{out} = V_{ref} \frac{R_2}{R_1 + R_2}$$

We need to make a few modifications to the basic form in order to accommodate the high output current and the output voltage adjustment. One possibility is shown in Figure 8.3.15 .

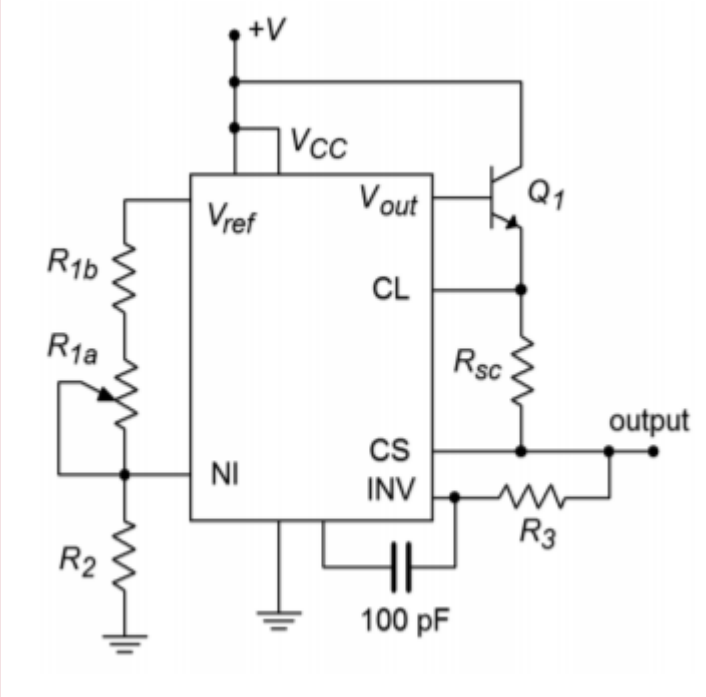


Figure 8.3.15 : Circuit for Example 8.3.5 : 2 V to 5 V regulator, 1 amp.

In order to produce a 1 A load current, an external pass transistor will be used. To obtain the desired voltage adjustment, resistor ◈1 is replaced with a series potentiometer/resistor combination (◈1◈,◈1◈). In this fashion, the minimum value for R1 will be R1b, and the maximum value will be ◈1◈+◈1◈ . There are several ways in which we can approach the calculation of these three resistors. Perhaps the easiest is to pick a value for ◈2 and then determine values for ◈1◈ and ◈1◈ . Though this is fairly straightforward, it is not very practical because you will most likely wind up with an odd size for the potentiometer. A better, though admittedly more involved, approach revolves around the selection of a reasonable pot value, such as 10 k Ω . Given the desired output potentials, the other two resistors may be found.

First of all, note that these three resistors are nothing more than a voltage divider. The output voltage Equation may be rewritten as

$$\frac{R_1 + R_2}{R_2} = \frac{V_{ref}}{V_{out}}$$

For the maximum case, we have

$$\frac{R_{1b} + R_2}{R_2} = \frac{V_{ref}}{V_{out}}$$

$$\frac{R_{1b} + R_2}{R_2} = \frac{7.15V}{5V}$$

$$\frac{R_{1b} + R_2}{R_2} = 1.43$$

If we consider ◈2 to be unity, we may say that the ratio of the two resistors to ◈2 is 1.43:1, or, that the

ratio of ◈1◈ to ◈2 is 0.43:1.
For the minimum case, we have

$$\frac{R_{1a} + R_{1b} + R_2}{R_2} = \frac{V_{ref}}{V_{out}}$$

$$\frac{R_{1a} + R_{1b} + R_2}{R_2} = \frac{7.15V}{2V}$$

$$\frac{R_{1a} + R_{1b} + R_2}{R_2} = 3.575$$

We may say that the ratio of the three resistors to ◈2 is 3.575:1, or that the ratio of ◈1◈+◈1◈ to ◈2 is 2.575:1. Because we already know that the ratio of ◈1◈ to ◈2 is 0.43:1, the ratio of ◈1◈ to ◈2 must be the difference, or 2.145:1. Because we chose 10 k Ω for ◈1◈ ,

$$R_2 = \frac{R_{1a}}{2.145}$$

$$R_2 = \frac{10k}{2.145}$$

$$R_2 = 4.66k$$

Similarly,

$$R_{1b} = R_2 \times 0.43$$
$$R_{1b} = 4.66k \times 0.43$$
$$R_{1b} = 2k$$

For the current sense resistor,

$$I_{limit} = \frac{V_{sense}}{R_{sc}}$$

$$R_{sc} = \frac{V_{sense}}{I_{limit}}$$

$$R_{sc} = \frac{0.65V}{1.0A}$$

$$R_{sc} = 0.65\Omega$$

For minimum temperature drift, ◈3 is included, and set to ◈1||◈2 . Because ◈1 is adjustable, a midpoint value will be used.

$$R_3 = R_1||R_2$$

$$R_3 = 6k||4.66k$$

$$R_3 = 2.62k$$

The final calculation involves the external pass transistor. This design uses a 1 A output, so this device must be able to handle this current continuously. Also, a minimum ⟐ specification is required. As the LM723 will be driving the pass transistor, the LM723 only needs to produce base drive current. With a maximum output of 150 mA, this translates to a minimum ⟐ of

$$\beta_{min} = \frac{I_c}{I_b}$$

$$\beta_{min} = \frac{1A}{150mA}$$

$$\beta_{min} = 6.67$$

This value should pose no problem for a power transistor.

Besides the applications we have just examined, the LM723 can also be used to make negative regulators, switching regulators, and other types as well. One useful variation on the basic theme is the use of foldback current limiting, as seen in Figure 8.3.12.6 . Unlike the ordinary form of current limiting, foldback limiting actually produces a decrease in output current once the limit point is reached. Figure 8.3.16⟐ shows the effect of ordinary limiting. Once the limit point is reached, further demands by the load will be ignored. The problem with this arrangement is that under short circuit conditions, the pass transistor will be under heavy stress. Because the load is shorted, so ⟐⟐⟐⟐⟐ = 0, and therefore, a large potential will drop across the pass transistor. This device is already handling the full current draw, thus the resulting power dissipation can be very high. Foldback limiting gets around this problem by lowering the output current as the pass transistor's voltage increases. Note how in Figure 8.3.16⟐ , the current limit curve does not drop straight down to the limit point, but rather, as the load demand increases, the current drops back to ⟐⟐⟐ . By limiting the current in this fashion, a much lower power dissipation is achieved.

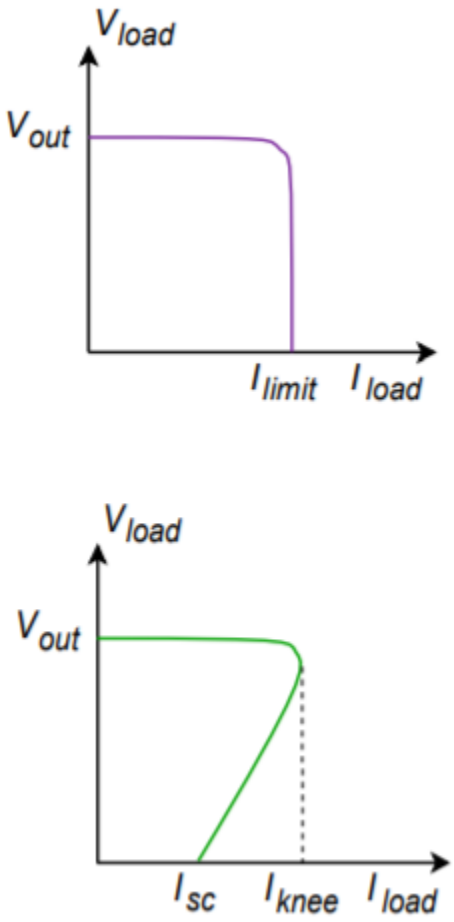


Figure 8.3.16 : Current limiting. a. Ordinary (top). b. Fold back (bottom).

Our last item of interest in this section is the LT3032 dual regulator series. These devices are of particular interest to the op amp technician and designer. The LT3032 is available as a fixed bipolar regulator at outputs of ± 3.3, ± 5, ± 12 or ± 15 volts. Another variant offers adjustable output from ± 1.22 volts up to ± 20 volts. Output current capability is up to 150 mA. These devices are ideal for powering general purpose op amp circuits.

The circuit in Figure 8.3.17 is a minimum parts-configuration dual regulator requiring only two resistors and a capacitor per side.. All that is needed ahead of this circuit is a standard transformer, rectifier and filter capacitor arrangement, such as that found in Figure 8.3.4 . The maximum output (unloaded) of the filter capacitors should be no more than 20 V in order to prevent damage to the LT3032. This circuit is certainly simpler than the LM317-based regulator of Figure 8.3.7 , and not much different from an LM340-based design. In fact, note the similarity of the design equations presented for the LM317 and the LT3032. Again, we see that the output potential is found essentially by multiplying a reference voltage by a series-parallel voltage gain. The downside is that its power dissipation and maximum currents are considerably lower than its LM340 or LM317 counterparts. Still, a 150 mA output capability is sufficient to drive a large number of op amps and other small signal devices.

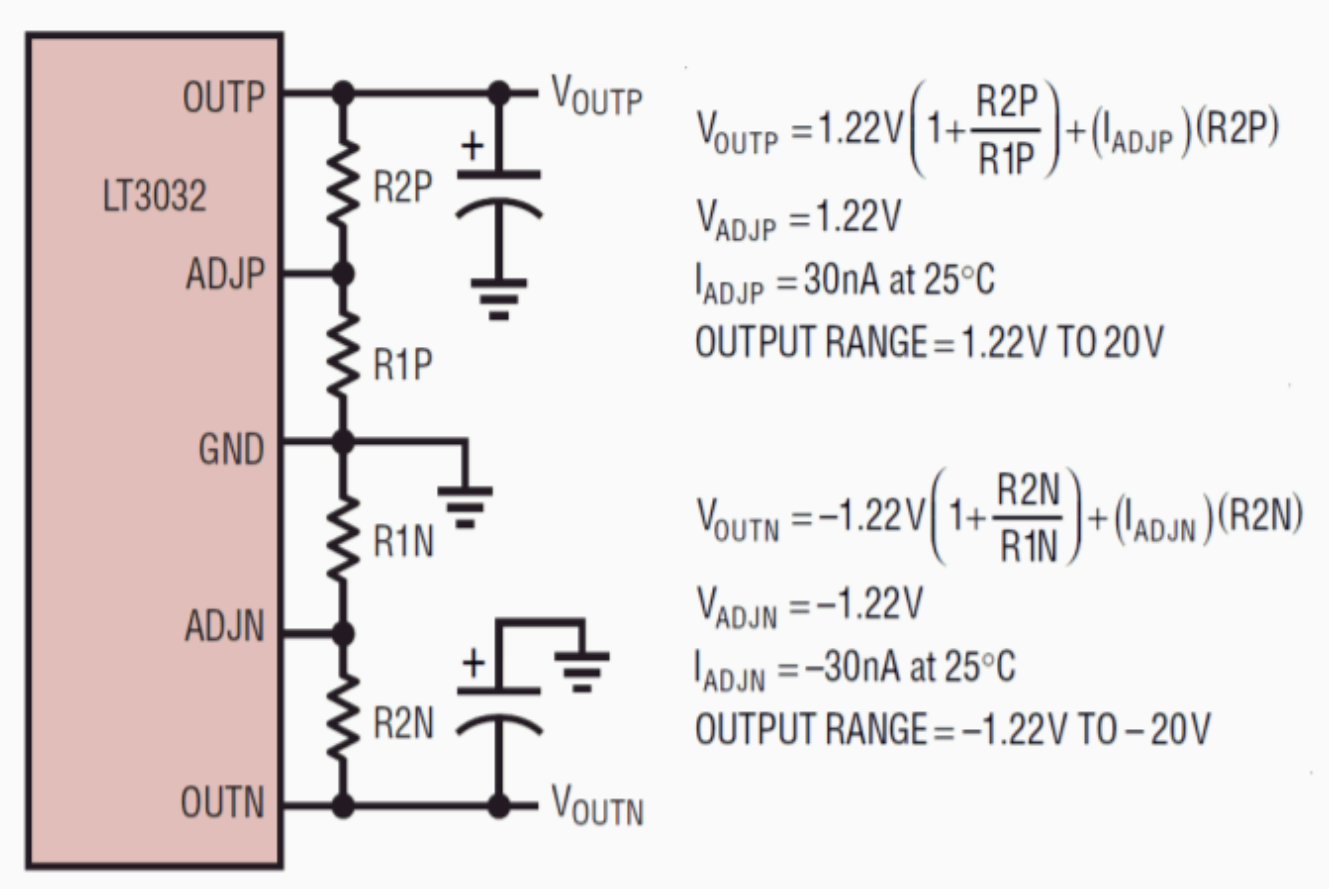


Figure 8.3.17 : LT3032 pinout and formulas. Reprinted with permission of Linear Technology

Current limiting and thermal shutdown are built-in. Other design features include the fact that the LT3032 is a low drop-out regulator requiring only a 300 millivolt differential per channel. ESD (electrostatic discharge) protection is included and reverse output polarity protection diodes are not required. Further, the addition of a small 10 nF capacitor at each output will reduce regulator noise voltage down to the 20 to 30 microvolt RMS range.

There are many other linear regulator ICs available to the designer than have been presented here. Many of these devices are rather specialized, and all units seem to have their own special set of operational formulas and graphs. There are, however, a few common threads among them all. First, due to the relative internal complexity, manufacturers often give very specific application guidelines for their particular ICs. The resulting design sequence is rather like following a cookbook and makes the designer's life much easier. Second, as mentioned at the outset, all linear regulators tend to be rather inefficient. This inefficiency is inherent in the design and implementation of the linear regulation circuits, and cannot be avoided. At best, the inefficiency can be minimized for a given application. In order to achieve high efficiency, a different topology must be considered. One alternative is the switching regulator.

12.4 SWITCHING REGULATORS

The major cause of the inefficiency of a linear regulator is that its pass transistor operates in the linear region. This means that it constantly sees both a high current and a high (or at least moderate) voltage. The result is a sizable power dissipation. In contrast, switching regulators rely on the efficiency of the transistor switch. When the transistor is off, no current flows, and thus, no power is dissipated. When the transistor is on (in saturation), a high current flows, but the voltage across the transistor (◈◈◈(◈◈◈)) is very small. This results in modest power dissipation. The only time that both the current and voltage are relatively large at the same time is during the switching interval. This time period is relatively short compared to cycle time, so again, the power dissipation is rather small.

Switching the pass transistor on and off can be very efficient, but unfortunately, the resulting pulses of current are not appropriate for most loads. Some way of smoothing out the pulses into a constant DC level is needed. One way to do this is through an inductor/capacitor arrangement. This concept is essential to the switching regulator.

Although the switching regulator does offer an increased efficiency over the linear regulator, it is not without its detractions. First of all, switching regulators tend to be more complex than linear types. This complexity may outstrip the efficiency advantage, particularly for low-power designs. Some of the newer "single chip" switching regulators make switching power supply design almost as straightforward as linear design, so this problem is not quite as great as it once was. The other major problem is that the switching process can create a great deal of radiated electrical noise. Without proper shielding and similar precautions, the induced noise signals can produce grave interference in nearby analog or digital circuits. In spite of these obstacles, switching power supplies have seen great acceptance in a variety of areas, including powering personal computers.

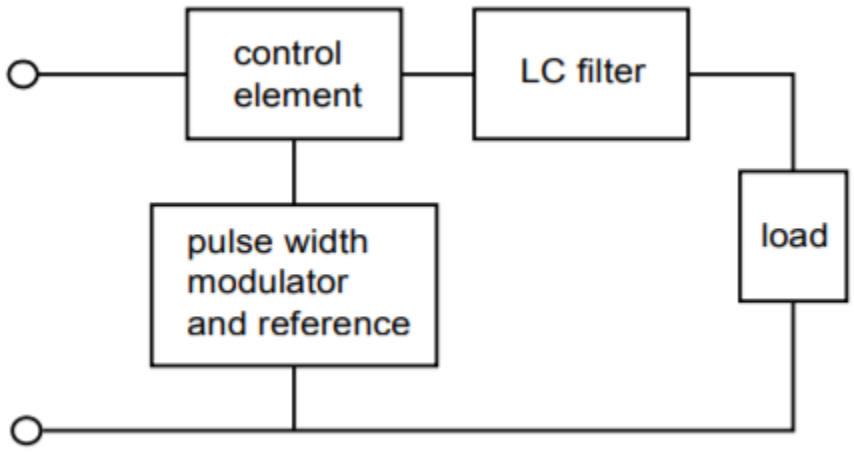


Figure 8.4.1 : Basic switching regulator.

A basic switching regulator outline is shown in Figure 8.4.1 . Once again, a control device is used to adjust the incoming signal. Unlike the linear regulator, the signal out of the control element is a pulse

train. In order to produce a smooth DC output, the pulse train is passed through a capacitor/inductor network. The heart of the circuit is the pulse-width modulator. This circuit creates a variable duty cycle pulse waveform, which alternately turns the control element on and off. The duty cycle of the pulse is proportional to the load current demand. Low load impedances require large currents and, thus, will create pulse trains with long "on" times. The peak value of the pulses will tend to be much greater than the average load current demand, however, if this current pulse is averaged over one cycle, the result will equal the load current demand. Effectively, the ⧫⧫ network serves to integrate the current pulses. This is shown graphically in Figure 8.4.2 . Because the lower load impedance is receiving a proportionally larger current, the load voltage remains constant. This is shown graphically in Figure 8.4.3 .

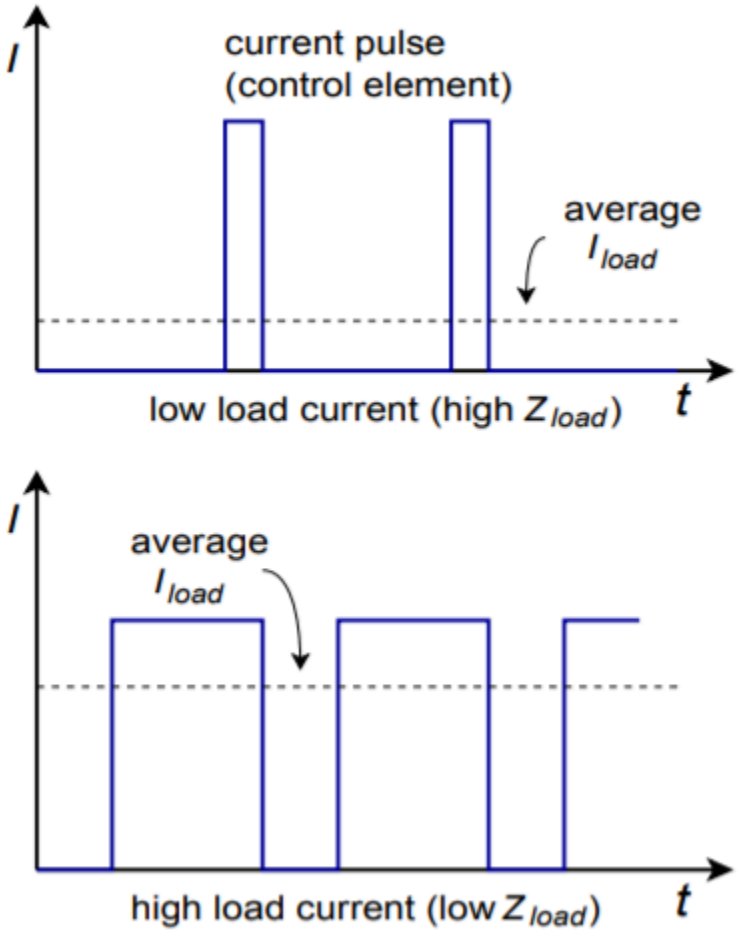


Figure 8.4.2 : Current waveforms.

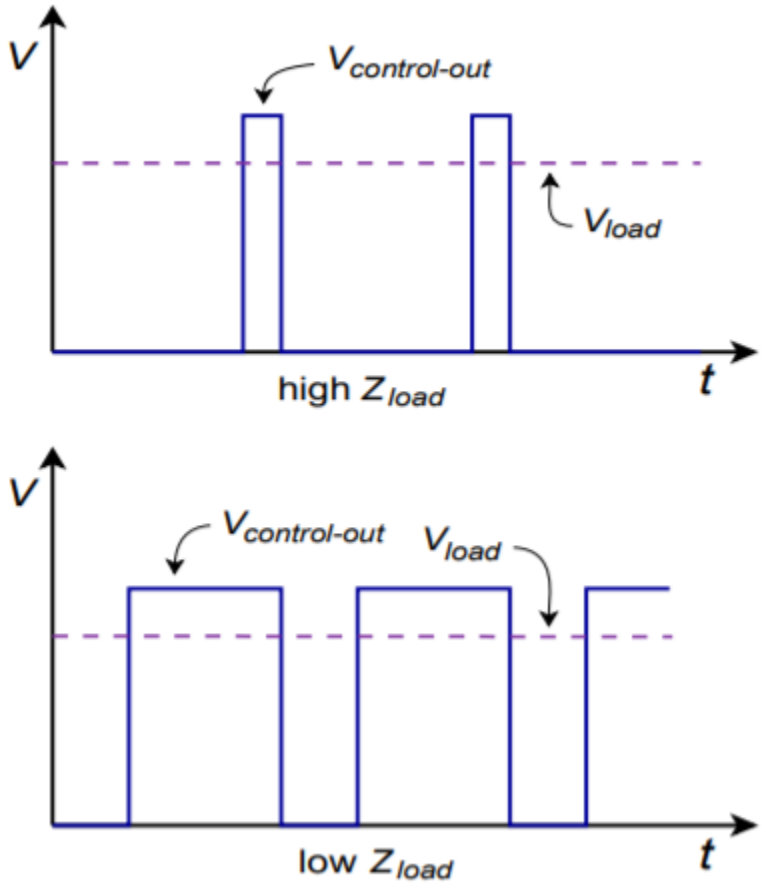


Figure 8.4.3 : Voltage waveforms.

Due to the constant charging and discharging of the ⍰ and ⍰ components, switching regulators tend to respond to load changes somewhat slower than linear regulators do. Also, if the frequency of the switching pulse is high, the required sizes for ⍰ and ⍰ may be reduced, resulting in a smaller circuit and possible cost reductions. Many switching regulators run in the 20 kHz to 100 kHz region. The practical upper limit for switching speed is determined by the speed of the control element. Devices with fast switching times will prove to be more efficient. Power FETs are very attractive for this application because of their inherent speed and low drive requirements. Their negative temperature coefficient of transconductance also helps reduce thermal runaway problems. For the highest powers, bipolar devices are often the only choice. Finally, it is possible to configure many switchers for step down, step up, or inverter operation. As an example, we will investigate the step down, or buck, configuration.

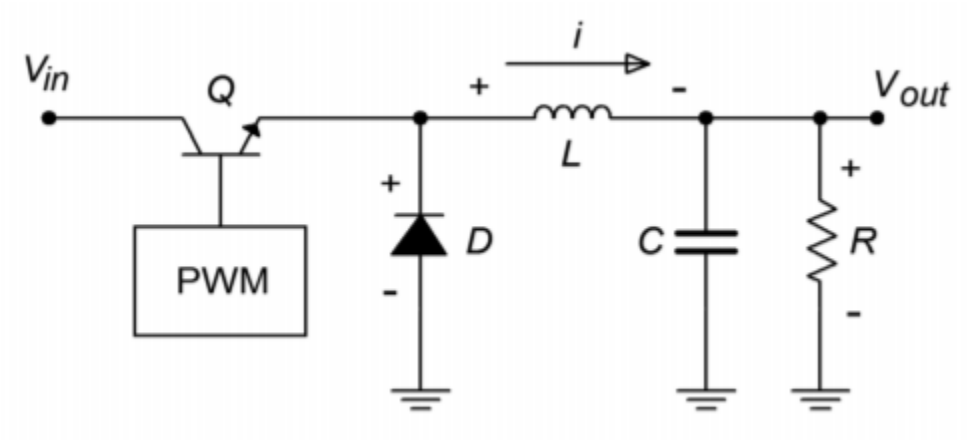


Figure 8.4.4 : Basic stepdown switching regulator. a. on: Source supplies current and charges inductor.

A block diagram of a step down switching regulator is shown in Figure 8.4.4 . Here is how it works: When the transistor, , is on, current flows through , , and the load. The inductor current rises at a rate equal to the inductor voltage divided by the inductance. The inductor voltage is equal to the input voltage minus the load voltage and the transistor's saturation potential. While is charging (i.e.,), provides load current. When turns off, the inductor's magnetic field starts to collapse causing an effective polarity reversal. In other words, the inductor is now acting as a source for load current. At this point, diode is forward biased, effectively removing the left portion of the circuit. also supplies current to capacitor during this time period. Eventually, the inductor current will drop below the value required by the load, and will start to discharge, making up the difference. Before completely discharges, the transistor switch will turn back on, repeating the cycle. Because the inductor is never fully discharged, this is called continuous operation. Discontinuous operation is also possible, although we will not pursue it here. One point worth noting is that for proper continuation of this cycle, some load current draw must always be present. This minimum level can be achieved through the use of a bleed resistor in parallel with the load.

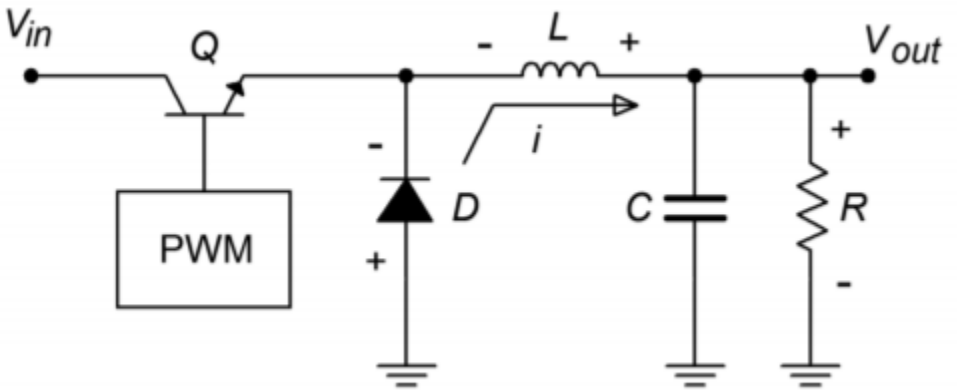


Figure 8.4.4 : Basic stepdown switching regulator (continued). b. off: Inductor supplies current

The values for and are dependent on the input and output voltages, desired output current, the switching frequency, and the particulars of the switching circuit used. Manufacturers generally give look-up tables and charts for appropriate values and/or formulas. Earlier switching ICs contained the

necessary base components and required only a moderate amount of external circuitry. The newest ICs may be configured with no more than three or four external passive parts.

One example of a switch mode regulator is the LM3578A. A functional diagram of the inner workings of this IC is shown in Figure 8.4.5 . The AND gates, comparator, oscillator, and associated latches combine to form the pulse-width modulator. Notice that this circuit includes an internal reference and mediumpower switching transistor. Thermal shutdown and current limiting are available. The circuit will operate at inputs up to 40 V and can produce output currents up to 750 mA. The maximum switching frequency is 100 kHz. Step up, step down, and inverter configurations are all possible with this device. To make the design sequence is fast as possible, the manufacturer has included a design chart. This is shown in Figure 8.4.6 . As with virtually all highly specialized ICs, specific design equations only apply to particular devices, and probably cannot be used with ICs that produce similar functions. Consequently, it is recommended that you do not memorize these formulas!

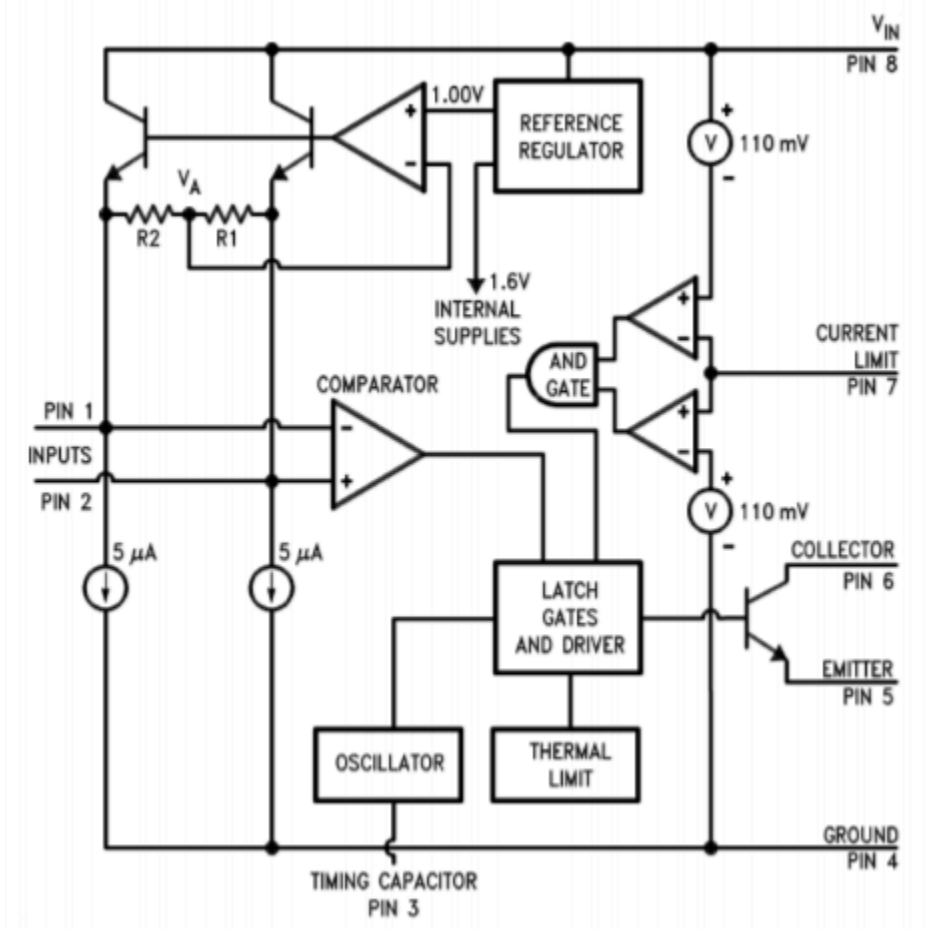


Figure 8.4.5 : LM3578 equivalent circuit. Reprinted from of Texas Instruments

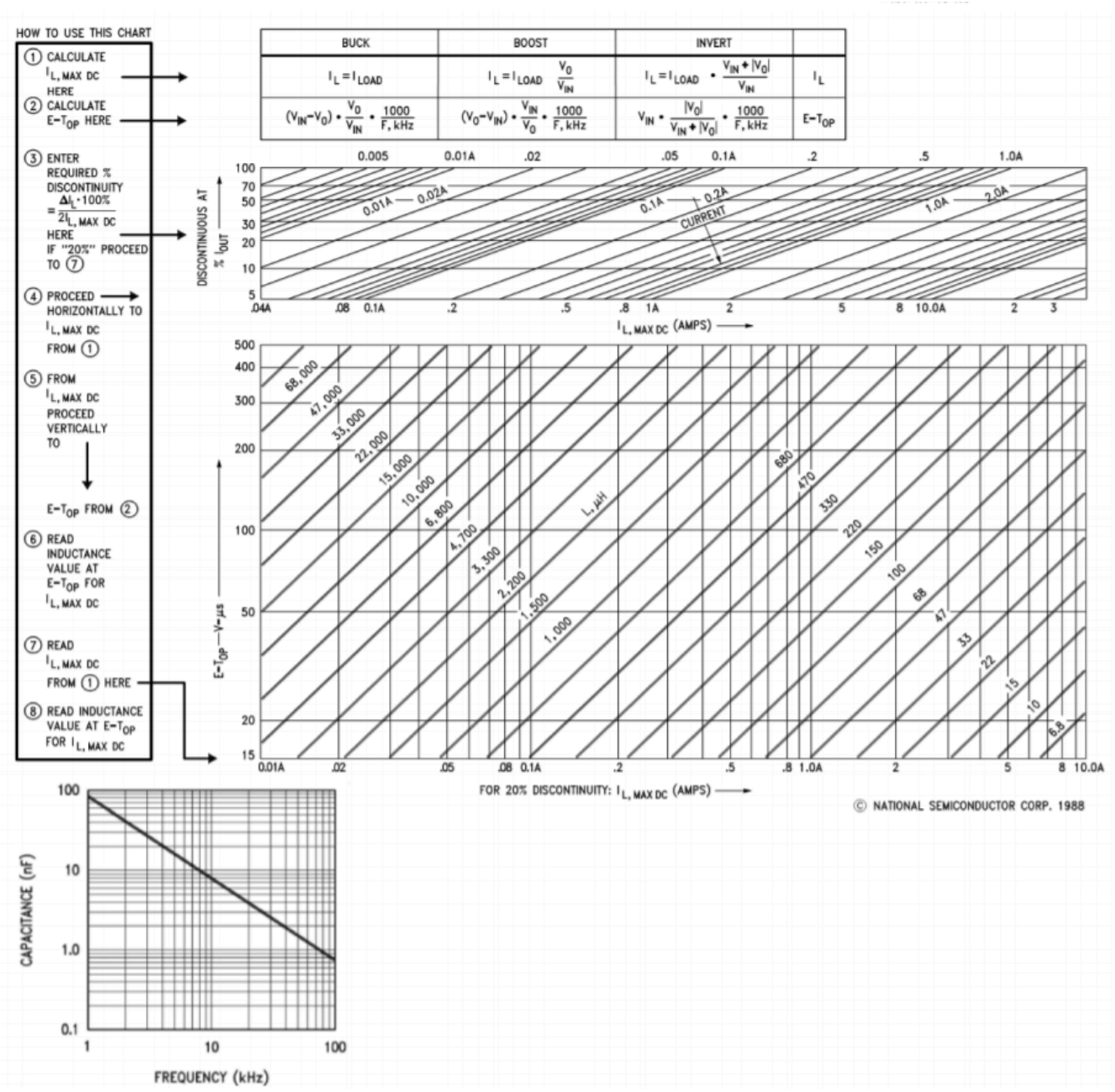


Figure 8.4.6 : LM3578 design chart. Reprinted from of Texas Instruments

A step down regulator using the LM3578A is shown in Figure 8.4.7 . ◈1 is the frequency-selection capacitor and can be found from the manufacturer's chart. ◈3 is necessary for continuous operation and is generally in the vicinity of 10 to 30 pF. ◈1 should be a Schottky-type rectifier. ◈1 and ◈2 set the step down ratio (they are functionally the same as ◈◈ and ◈◈ in the previous work). ◈3 sets the current limit. Finally, ◈1 and ◈2 are used for the final output filtering.

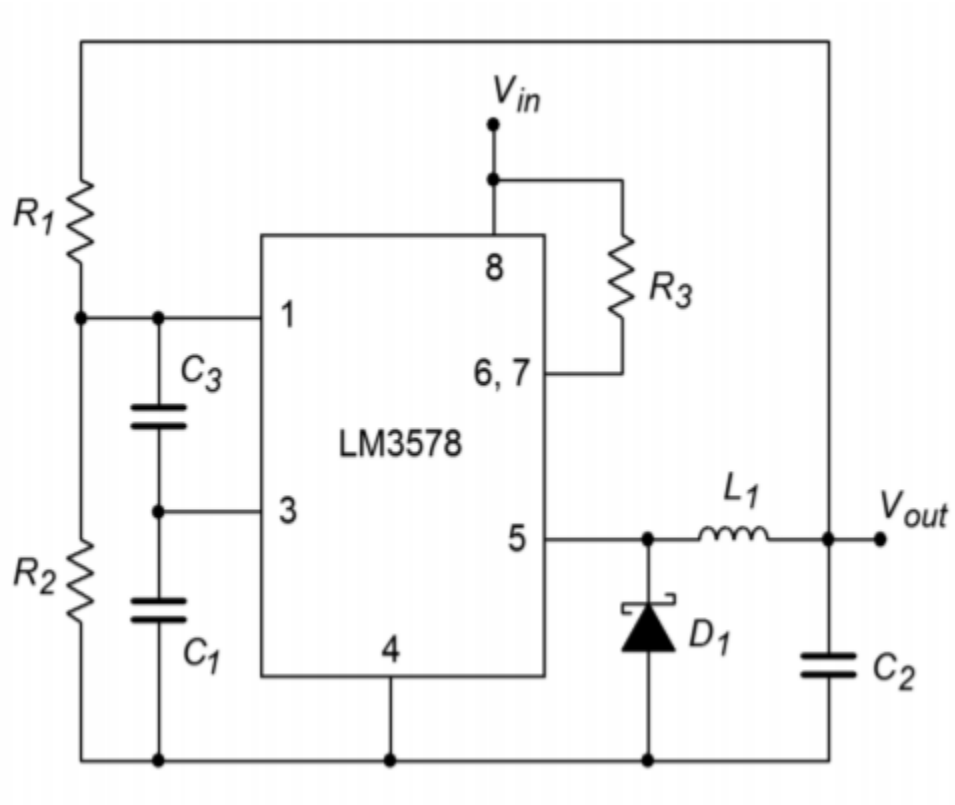


Figure 8.4.7 : Step down regulator using an LM3578.

The relevant equations from the manufacturer's data sheets are:

$$V_{out} = \frac{R_1}{R_2} + 1 \ (in \ volts)$$

$$R_3 = \frac{0.11V}{I_{sw(max)}}$$

$$C_2 \geq V_{out} \frac{V_{in} - V_{out}}{8f^2 V_{in} V_{ripple} L_1}$$

$$L_1 = V_{out} \frac{V_{in} - V_{out}}{\Delta I_{out} V_{in} f}$$

$$\Delta I_{out} = 2I_{out} \ Discontinuity \ Factor \ (\text{typically } 0.2)$$

Where ◈◈◈(◈◈◈) is the maximum current through the switching element, and ◈◈◈◈◈◈◈ is in peak-to-peak volts. Some values may be found using the look-up chart method shown in Figure 8.4.6 .

Example 8.4.1

Using the LM3578A, design a step down regulator that delivers 12 V from a 20 V source, with 200 mA of

load current. Use an oscillator frequency of 50 kHz, and a discontinuity factor of 0.2 (20%). Ripple should be no more than 40 mV.

First determine the ◈1 and ◈2 values. ◈2 is arbitrarily chosen at 10 k Ω .

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

$$R_1 = R_2(V_{out} - 1V)$$

$$R_1 = 10k(12V - 1V)$$

$$R_1 = 110k$$

$$R_3 = \frac{0.11V}{I_{sw(max)}}$$

$$R_3 = \frac{0.11V}{0.75A}$$

$$R_3 = 0.15\Omega$$

Note that ◈3 will always be 0.15 Ω for this circuit form

From the oscillator graph, ◈1 is estimated at 1700 pF, and ◈3 is set to 20 pF, as suggested by the manufacturer.

$$\Delta I_{out} = 2I_{out}\ Discontinuity\ Factor$$

$$\Delta I_{out} = 2 \times 200mA \times 0.2$$

$$\Delta I_{out} = 80mA$$

$$L_1 = V_{out}\frac{V_{in} - V_{out}}{\Delta I_{out} V_{in} f}$$

$$L_1 = 12V \times \frac{20V - 12V}{80mA \times 20V \times 50kHz}$$

$$L_1 = 1.2mH$$

$$C_2 \geq V_{out} \times \frac{V_{in} - V_{out}}{8f^2 V_{in} V_{ripple} L_1}$$

$$C_2 \geq 12V \times \frac{20V - 12V}{8 \times 50kHz^2 \times 20V \times 40mV \times 1.2mH}$$

$$C_2 \geq 5\mu F$$

To be on the conservative side, ◈2 is usually set a bit higher, so a standard 33 ◈ F or 47 ◈ F might be used.

A basic inverting switcher is shown in Figure 8.4.8 . This circuit produces a negative output potential from a positive input. It works as follows: When the transistor switch is closed, the inductor L is charged. Diode ◈ is in reverse bias, as its anode is negative. When the switch opens, the inductor's collapsing field causes it to appear as a source of opposite polarity. The diode is forward-biased because its cathode is now forced to be lower than its anode. The inductor is now free to deliver current to the load. Eventually, the inductor will discharge to the point where the transistor switch will turn back on. While the inductor is charging, capacitor ◈ will supply the load current. The process repeats, thus maintaining a constant output potential.

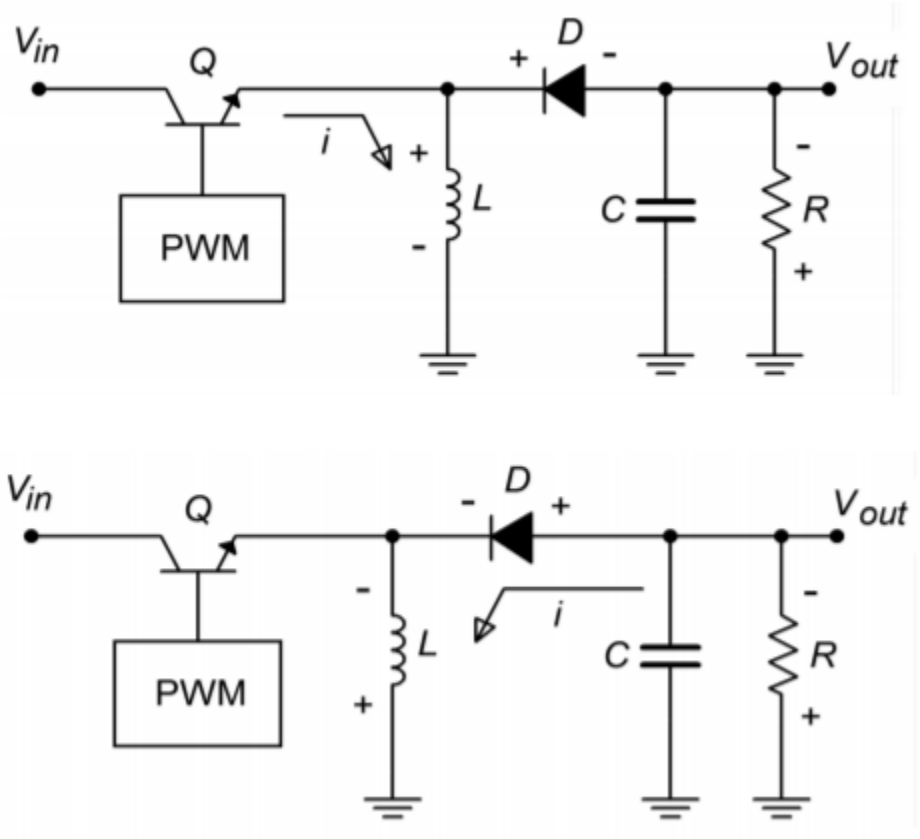


Figure 8.4.8 : Basic inverting switcher. a. ◈ on: Inductor is charged by the source (top). b. ◈ off: Inductor supplies current (bottom).

A basic step up switcher is shown in Figure 8.4.9 . This variation is used when a potential greater than the input is desired, such as deriving a 15 V supply from an existing 5 V source. Here is how the circuit works: When the transistor switch is closed, the inductor ◈ charges. During this time period, the capacitor ◈ is supplying load current. Because the output potential will be much higher than the saturation voltage of the transistor switch, the diode ◈ will be in reverse bias. When the transistor turns off, the magnetic field of the inductor collapses, causing the inductor to appear as a source. This potential is added to the driving source potential, as these two elements are in series. This combined voltage is what the load sees; hence, the load voltage is greater than the driving source. Eventually, the inductor will discharge to the point where the transistor switches back on, thus reverse-biasing the diode and recharging ◈ . The capacitor will continue to supply load current during this time. This process will continue in this fashion, producing the desired output voltage.

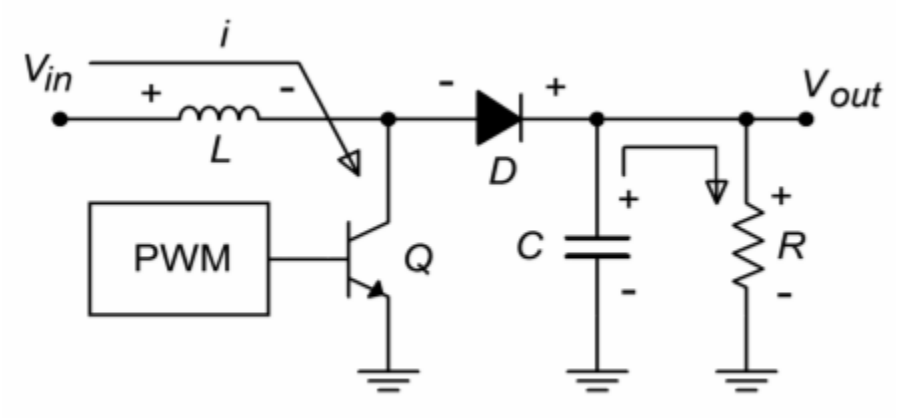


Figure 8.4.9 : Basic step-up switching regulator. on: Source charges inductor.

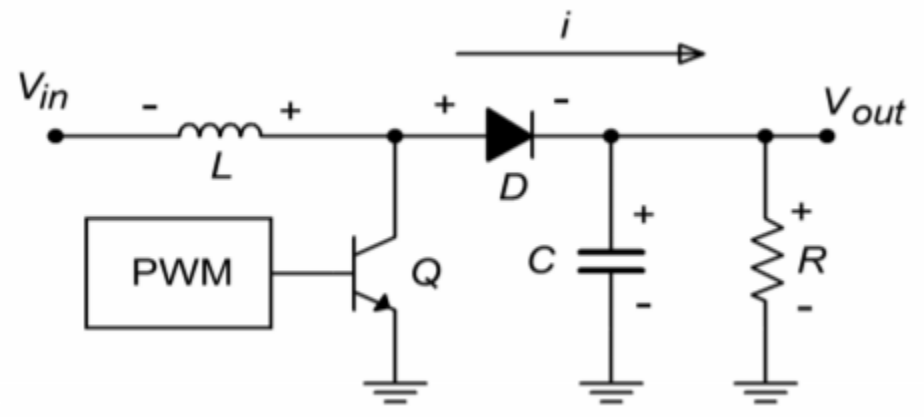


Figure 8.4.9 : Basic step-up switching regulator. off: Inductor supplies current.

For the step up and inverter forms, other sets of equations are used for the LM3578A. Although the design sequence is certainly not quite as straightforward as in the linear regulator circuits, it is definitely not a major undertaking, either. If 750 mA is not sufficient, an external pass transistor may be added to the LM3578A. Other switching regulator ICs are available from different manufacturers. Each unit operates on the same basic principle, but the realization of the design may take considerably different routes. Specific device data sheets must be consulted for each model.

For specific applications, some manufacturers offer switching regulators that are almost drop-in replacements for basic 3-pin linear regulators. These devices are not nearly as flexible as the generic switching regulator ICs, though. A good example is the LM2576 regulator. This is a 3 A output, step down regulator (i.e., buck-mode only). It is available at a variety of output potentials ranging from 3.3 V to 15 V. An adjustable version is also available. A block diagram and typical circuit are shown in Figure 8.4.10 . As you can see, a typical fixed output design requires a minimum of external components: 2 capacitors, an inductor, and a diode. Other members of this family include the LM2574 0.5 A step down regulator, and the LM2577 step up regulator.

LM2576/LM2576HV Series SIMPLE SWITCHER® 3A Step-Down Voltage Regulator

FEATURES

- 3.3V, 5V, 12V, 15V, and Adjustable Output Versions
- Adjustable Version Output Voltage Range,1.23V to 37V (57V for HV Version) ±4% Max Over Line and Load Conditions
- Specified 3A Output Current
- Wide Input Voltage Range, 40V Up to 60V for HV Version
- Requires Only 4 External Components
- 52 kHz Fixed Frequency Internal Oscillator
- TTL Shutdown Capability, Low Power Standby Mode
- High Efficiency
- Uses Readily Available Standard Inductors
- Thermal Shutdown and Current Limit Protection
- P+ Product Enhancement Tested

APPLICATIONS

- Simple High-Efficiency Step-Down (Buck) Regulator
- Efficient Pre-Regulator for Linear Regulators
- On-Card Switching Regulators
- Positive to Negative Converter (Buck-Boost)

DESCRIPTION

The LM2576 series of regulators are monolithic integrated circuits that provide all the active functions for a step-down (buck) switching regulator, capable of driving 3A load with excellent line and load regulation. These devices are available in fixed output voltages of 3.3V, 5V, 12V, 15V, and an adjustable output version.

Requiring a minimum number of external components, these regulators are simple to use and include internal frequency compensation and a fixed-frequency oscillator.

The LM2576 series offers a high-efficiency replacement for popular three-terminal linear regulators. It substantially reduces the size of the heat sink, and in some cases no heat sink is required.

A standard series of inductors optimized for use with the LM2576 are available from several different manufacturers. This feature greatly simplifies the design of switch-mode power supplies.

Other features include a specified ±4% tolerance on output voltage within specified input voltages and output load conditions, and ±10% on the oscillator frequency. External shutdown is included, featuring 50 µA (typical) standby current. The output switch includes cycle-by-cycle current limiting, as well as thermal shutdown for full protection under fault conditions.

TYPICAL APPLICATION

(Fixed Output Voltage Versions)

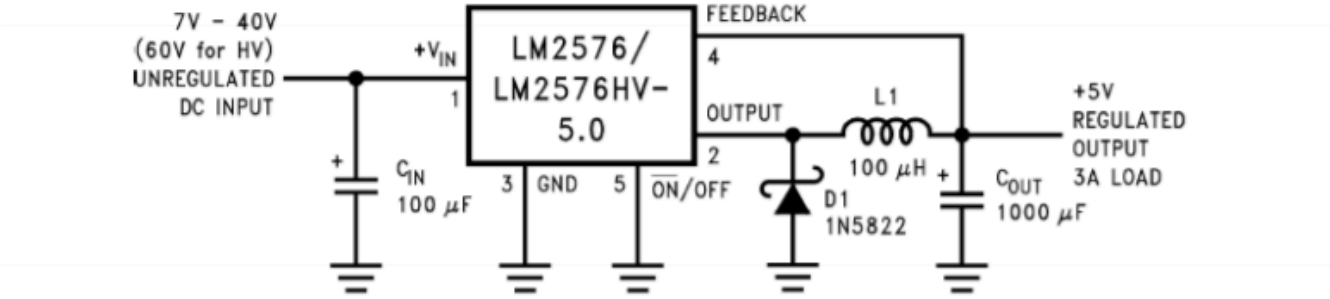


Figure 1.

Block Diagram

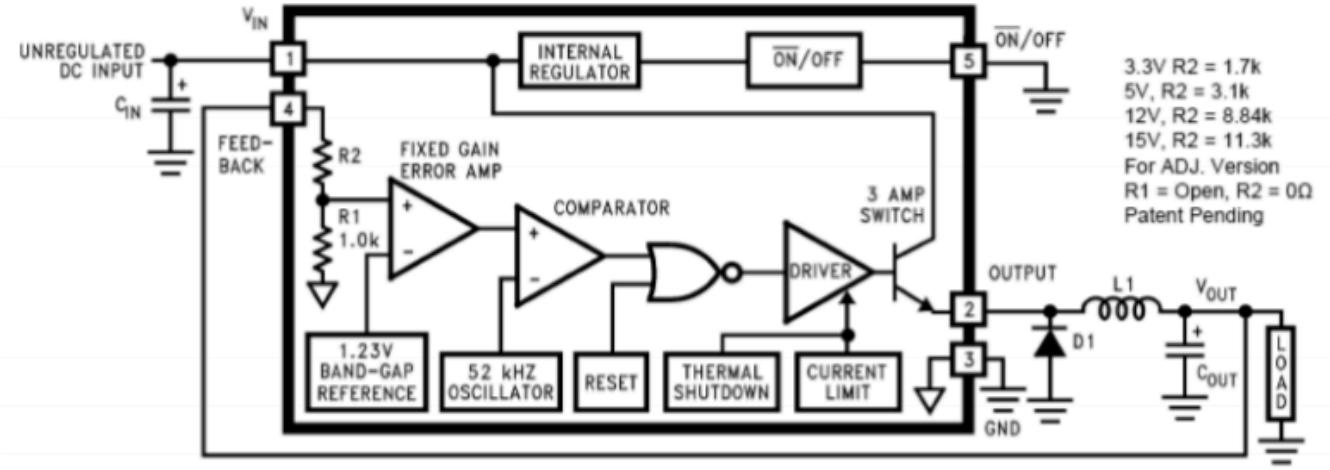


Figure 8.4.10 : LM2576 switching regulator. Reprinted courtesy of Texas Instruments

12.5 HEAT SINK USAGE

Whenever appreciable amounts of power are dissipated by semiconductor devices, some form of cooling element needs to be considered. Power supply regulation circuits are no exception. The pass transistors used in both linear and switching regulators can be forced to dissipate large amounts of power. The result of this is the production of heat. Generally, the life-span of semiconductors drops as the operation temperature rises. Most silicon-based devices exhibit maximum allowable junction temperatures in the 150 ° C range. Although power transistors utilize heavier metal cases, they are generally not suitable for high dissipation applications by themselves.

In order to increase the thermal efficiency of the device, an external heat sink is used. Heat sinks are normally made of aluminum and appear as a series of fins. The fins produce a large surface, which enhances the process of heat convection. In other words, the heat sink can transfer heat to the surrounding atmosphere faster than the power transistor can. By bolting the transistor to the heat sink, the device will be able to dissipate more power at a given operating temperature. Some typical heat sinks are shown in Figure 8.5.1 .

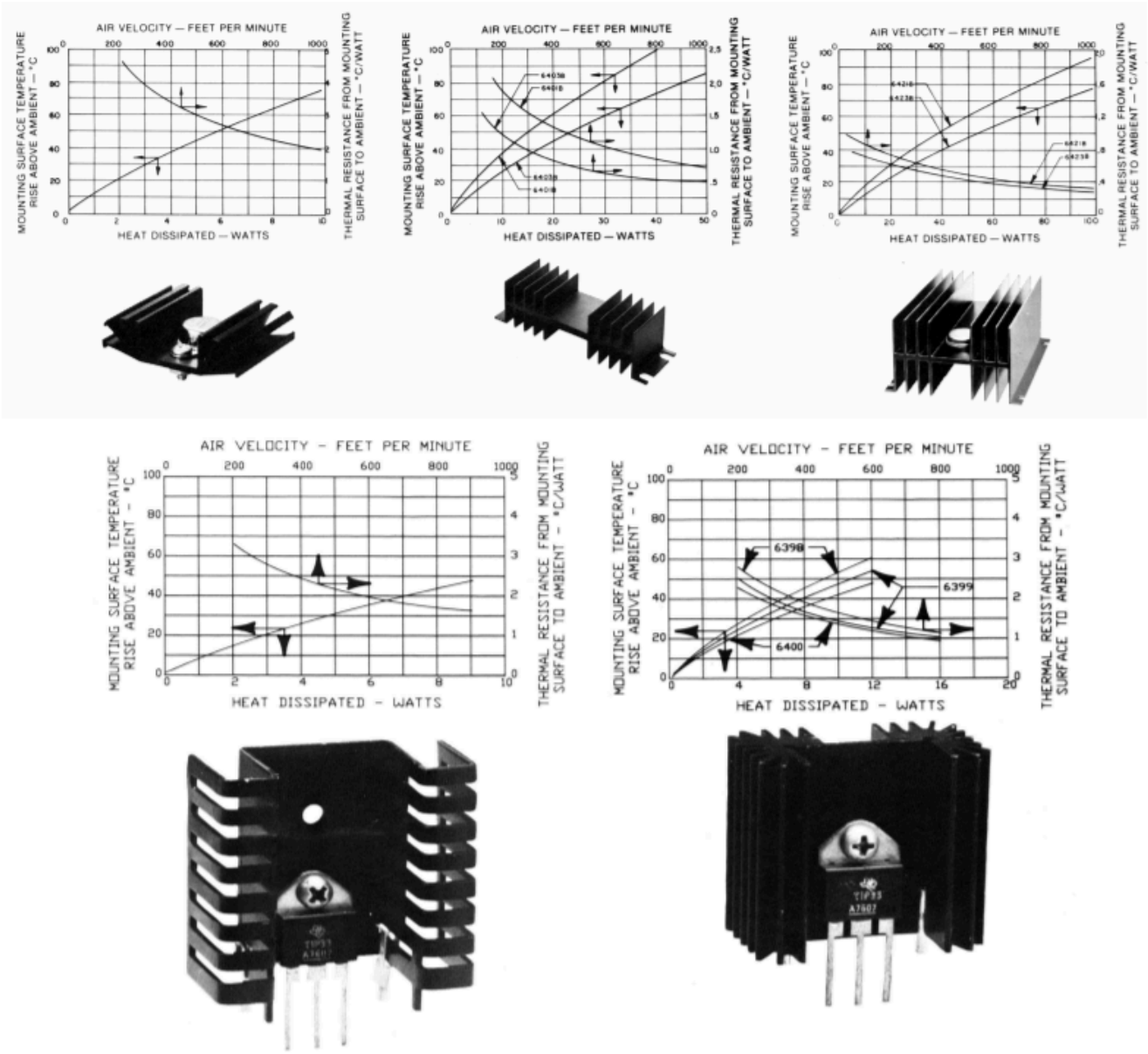


Figure 8.5.1 : Typical heatsinks. Reprinted courtesy of Thermalloy, Inc.

PHYSICAL REQUIREMENTS

Heat sinks are designed to work with specific device-case styles. The most common case styles for regulators are the TO-220 “power tab”, and the TO-3 “can”. Heat sinks are available for these specific styles, including the requisite mounting hardware and insulation spacers. Some of the lower power regulators utilize TO-5 “mini can” or DIP-type cases. Heat sinks are available for these package types, too, but are not quite as common.

There are a couple of 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. Note that excessive quantities of heat sink grease will actually 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 (e.g., fans).

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 ◇ , and the units are Centigrade degrees per watt. In this model, temperature is analogous to voltage, and thermal power dissipation is analogous to current. A useful Equation is,

$$P_D = \frac{\Delta T}{\theta_{total}}$$

(8.5.1)

Where ◇◇ is the power dissipated by the semiconductor device in watts, Δ◇ is the temperature differential, and ◇◇◇◇◇◇ 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 8.5.2 . ◇◇ is the semiconductor junction temperature. This heat energy source heats the device case to ◇◇ . The thermal resistance between the two entities is ◇◇◇ . The case, in turn, heats the heat sink via the interconnection. This thermal resistance is ◇◇◇ , and the resulting temperature is ◇◇ .

Finally, the heat sink passes the thermal energy to the surrounding air, which is sitting at ◇◇ . The thermal resistance of the heat sink is ◇◇◇ . The equivalent thermal model is shown in Figure 8.5.3 . (Although this does not have perfect correspondence with normal circuit analysis, it does illustrate the main points.)

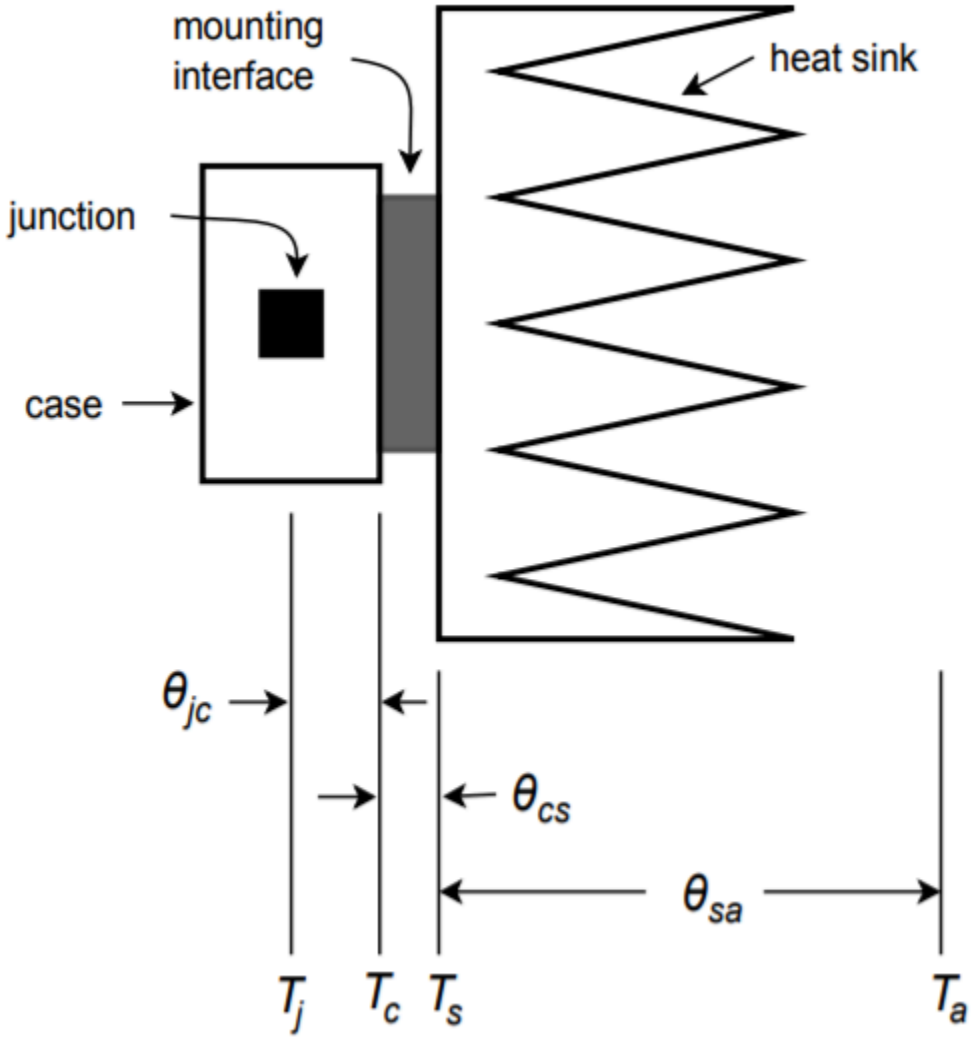


Figure 8.5.2 : Device and heatsink.

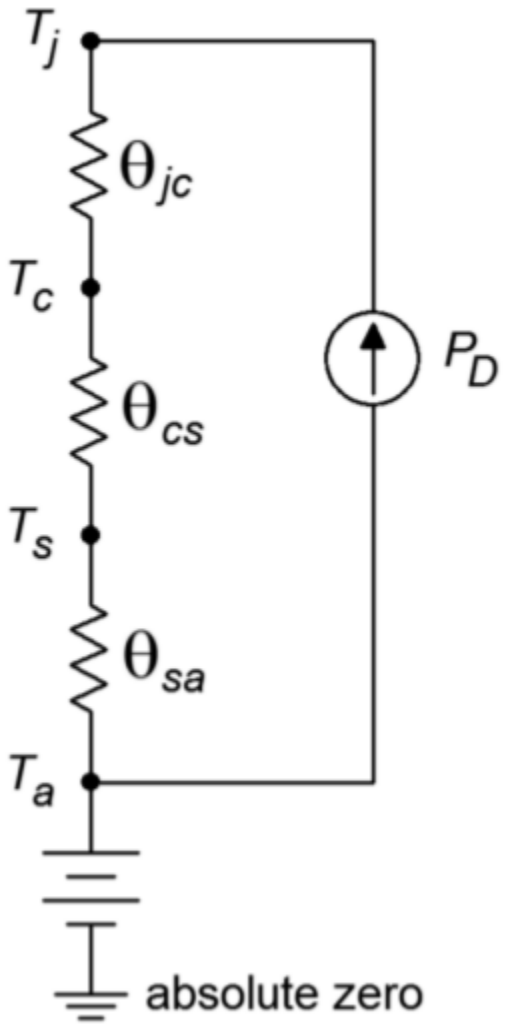


Figure 8.5.3 : Equivalent thermal model of Figure 8.5.2 .

In this model, ground represents a temperature of absolute zero. The circuit is sitting at an ambient temperature ◈◈ , thus a voltage source of ◈◈ 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 ◈◈ , higher power dissipations require lower thermal resistances. As ◈◈◈ is set by the device manufacturer, you have no control over that element. However, ◈◈◈ is a function of the case style and the insulation material used, so you do have some control (but not a lot) over that. On the other hand, as the person who specifies the heat sink, you have a great deal of control over ◈◈◈ . Values for ◈◈◈ are given by heat sink manufacturers. A useful variation of Equation 8.5.1 is

$$P_D = \frac{T_j - T_a}{\theta_{jc} + \theta_{cs} + \theta_{sa}}$$

(8.5.2)

Normally, power dissipation, junction and ambient temperatures, ◈◈◈ and ◈◈◈ are known. The

idea is to determine an appropriate heat sink. Both ⍰⍰ and ⍰⍰⍰ are given by the semiconductor device manufacturer. The ambient temperature, ⍰⍰ , 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 8.5.4 , may be used to determine ⍰⍰ .

Example 8.5.1

Determine the appropriate heat sink rating for a power device rated as follows: ⍰⍰(⍰⍰⍰) = 150 ° C, TO-220 case style, ⍰⍰⍰ = 3.0 C ° /W. The device will be dissipating a maximum of 6 W in an ambient temperature of 40 ° C. Assume that the heat sink will be mounted with heat sink grease and a 0.002 mica insulator.

First, find ⍰⍰⍰ from the TO-220 graph. Curve 3 is used. The approximate (conservative) value is 1.6 C ° /W.

$$P_D = \frac{T_j - T_a}{\theta_{jc} + \theta_{cs} + \theta_{sa}}$$

$$\theta_{sa} = \frac{T_j - T_a}{P_D} - \theta_{jc} - \theta_{cs}$$

$$\theta_{sa} = \frac{150°C - 40°C}{6W} - 3.0C°/W - 1.6C°/W$$

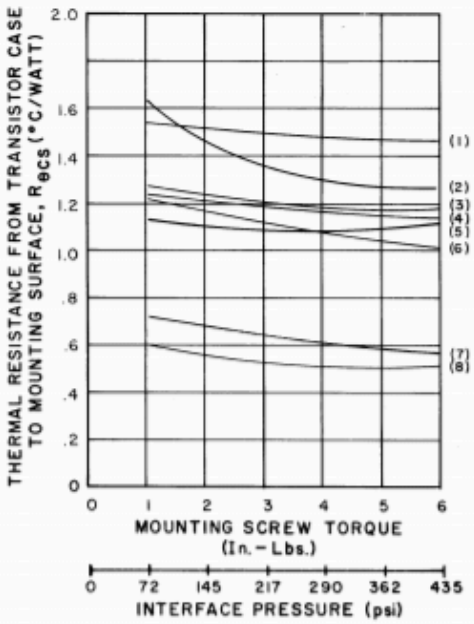
$$\theta_{sa} = 13.73C°/W$$

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 2 C ° /W or so. Also, note the generally lower values of ⍰⍰⍰ 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 ⍰⍰⍰ .

WITHOUT THERMAL GREASE

JEDEC TO-3

Interface Thermal Resistance versus Mounting Screw Torque for a TO-3 Semiconductor Device using Various Insulating Materials. No Thermal Joint Compound Used in the Interface Area.

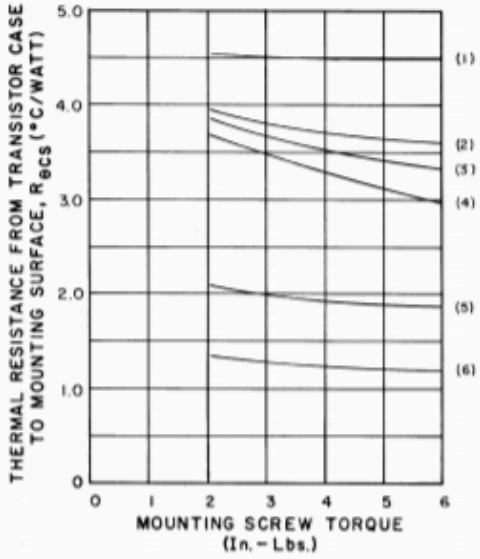


LEGEND:

(1) THERMALFILM, .002(.05) THK.
(2) MICA, .003(.08) THK.
(3) MICA, .002(.05) THK.
(4) HARD ANODIZED, .020(.51) THK.
(5) THERMALSIL, .008(.20) THK.
(6) ALUMINUM OXIDE, .062(1.57) THK.
(7) BERYLLIUM OXIDE, .062(1.57) THK.
(8) BARE JOINT – NO FINISH

JEDEC TO-220

Interface Thermal Resistance versus Mounting Screw Torque for a TO-220 Semiconductor Device using Various Insulating Materials. No Thermal Joint Compound Used in the Interface Area.



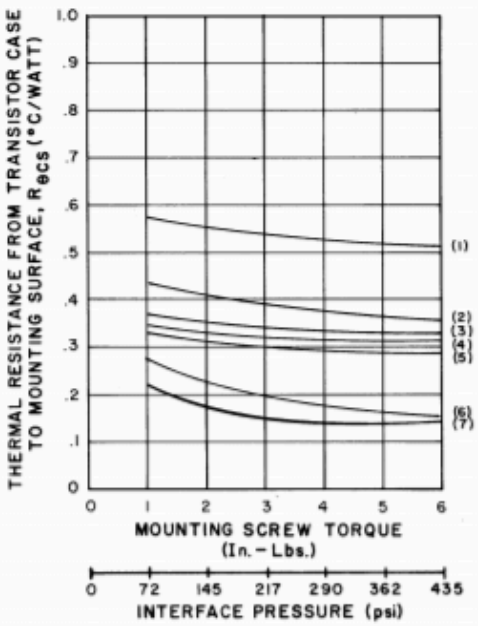
LEGEND:

(1) THERMALFILM, .002(.05) THK.
(2) MICA, .003(.08) THK.
(3) MICA, .002(.05) THK.
(4) HARD ANODIZED, .020(.51) THK.
(5) THERMALSIL, .008(.20) THK.
(6) BARE JOINT – NO FINISH

WITH THERMAL GREASE

JEDEC TO-3

Interface Thermal Resistance versus Mounting Screw Torque for a TO-3 Semiconductor Device Using Various Insulating Materials. Thermalcote Thermal Joint Compound Used in the Interface Area.

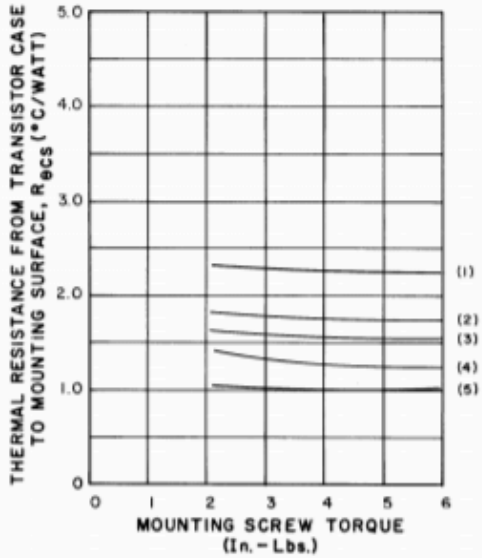


LEGEND:

(1) THERMALFILM, .002(.05) THK.
(2) MICA, .003(.08) THK.
(3) MICA, .002(.05) THK.
(4) ALUMINUM OXIDE, .062(1.57) THK.
(5) HARD ANODIZED, .020(.51) THK.
(6) BERYLLIUM OXIDE, .062(1.57) THK.
(7) BARE JOINT – NO FINISH

JEDEC TO-220

Interface Thermal Resistance versus Mounting Screw Torque for a TO-220 Semiconductor Device using Various Insulating Materials. Thermalcote Thermal Joint Compound Used in the Interface Area.



LEGEND:

(1) THERMALFILM, .002(.05) THK.
(2) MICA, .003(.08) THK.
(3) MICA, .002(.05) THK.
(4) HARD ANODIZED, .020(.51) THK.
(5) BARE JOINT – NO FINISH

Figure 8.5.4 : ◈◈◈ for TO-3 and TO-220. Reprinted courtesy of Thermalloy, Inc.

12.6 EXTENDED TOPIC - PRIMARY SWITCHER

The switching regulators examined earlier are referred to as secondary switchers because the switching elements are found on the secondary side of the power transformer. In contrast to this is the primary or forward switcher. The switching circuitry in these designs is placed prior to the primary of the power transformer. This positioning offers a distinct advantage over the secondary switcher. The power transformer and secondary rectifier will be handling much higher frequencies, thus they can be made much smaller. The result is a physically smaller and lighter design.

One possible configuration of a primary switcher is shown in Figure 8.6.1 . This is known as a push-pull configuration. The two power transistors are alternately pulsed on and off. That is, when one device is conducting, the other is off. By doing this, opposite polarity pulses are fed into the primary of the transformer, creating a high frequency alternating current (as is the case with secondary switchers, primary switchers operate at frequencies well above the nominal 60 Hz power line). This high frequency waveform is then stepped up or down to the secondary, where it is again rectified and then filtered, producing a DC output signal.

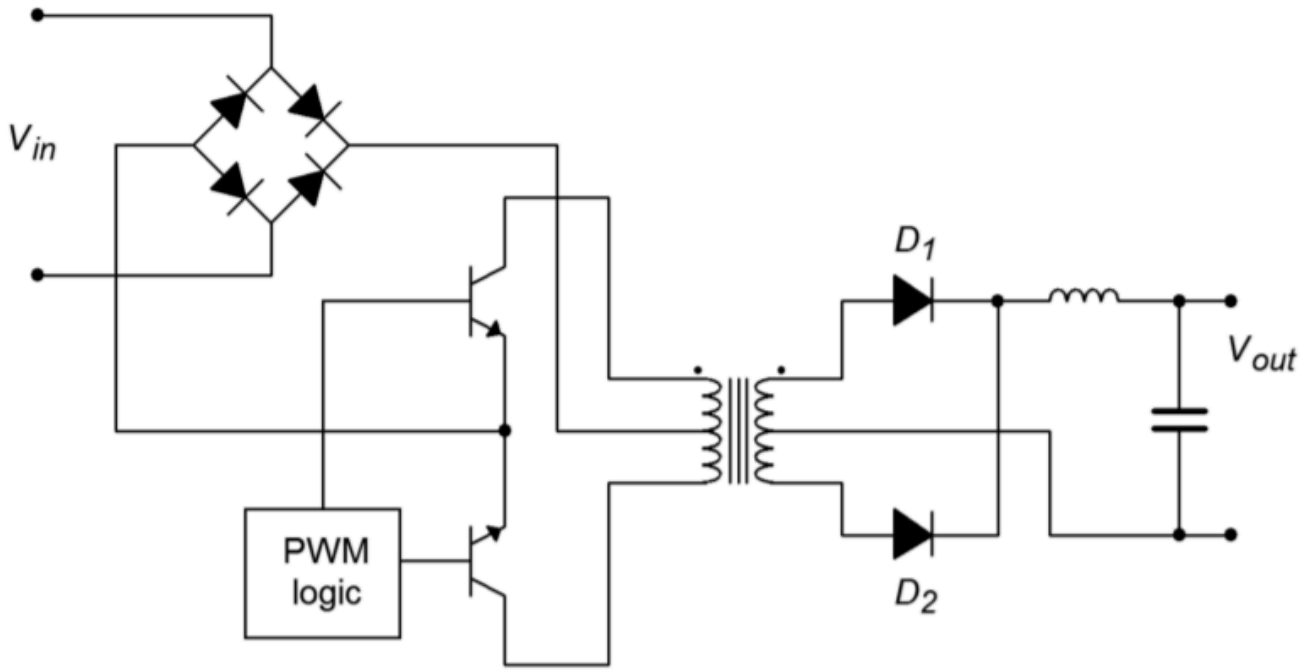


Figure 8.6.1 : Push-pull primary switching regulator.

Although the transformer and secondary rectifier/filter may be reduced in size, it is important to note that the main input rectifier and switching transistors are not isolated from the AC power source, as is the case in other power supply designs. These devices must handle high input potentials. For an ordinary 120 V AC line, that translates to over 170 V peak. Also, the power transistors will see an off-

state potential of twice ⍰⍰⍰, or over 340 V in this case. Some form of in-rush current limiting will also need to be added.

A somewhat more sophisticated approach is taken in Figure 8.6.2 . This circuit is known as a full bridge switcher. In this configuration, diagonal pairs of devices are simultaneously conducting (i.e., ⍰1/⍰4 and ⍰2/⍰3). By eliminating the center-tapped primary, each device sees a maximum potential of ⍰⍰⍰, or one-half that of the pushpull switcher. The obvious disadvantage is the need for four power devices instead of just two.

Primary switchers do offer a size and weight advantage over secondary switchers and linear regulators. They also maintain the high efficiency characteristics of the secondary switcher. They do tend to be somewhat more complex, though, and their application is therefore best suited to cases where circuit size, weight, and efficiency are paramount.

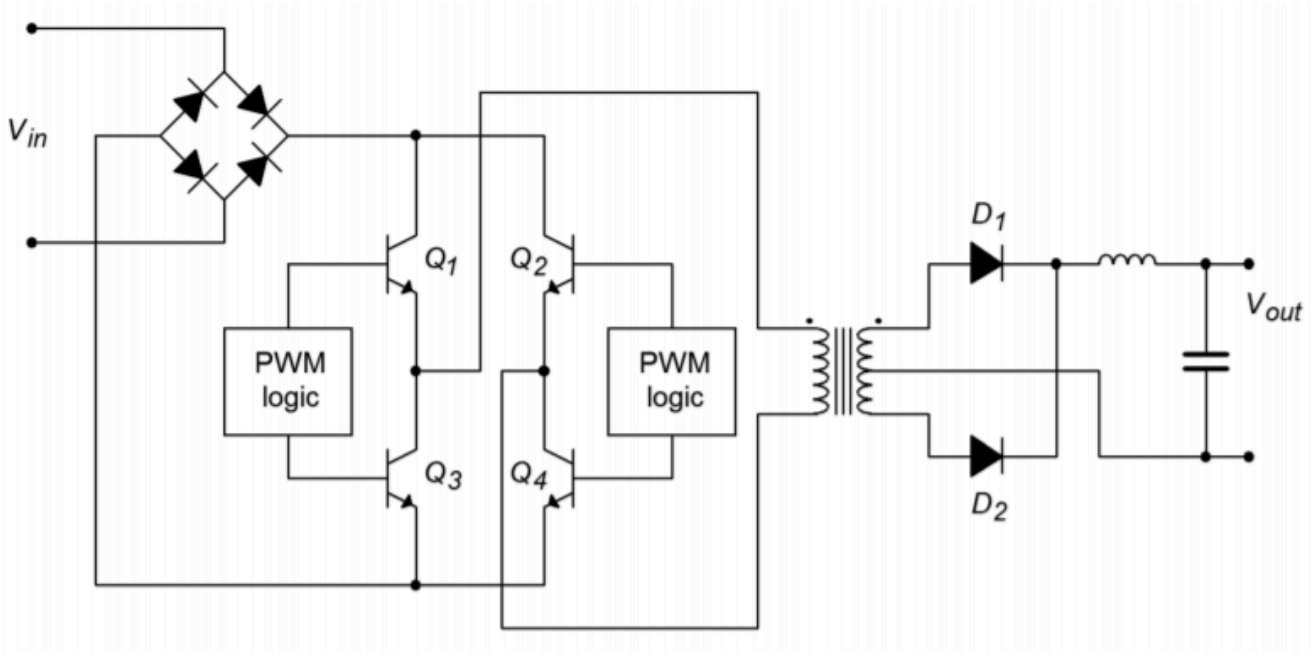


Figure 8.6.2 : Full-bridge primary switching regulator.

12.7 SUMMARY

In this chapter you have examined the basic operation of voltage regulators. Their purpose is simple: to provide constant, non-varying output voltages despite changes in either the AC source, or in the load current demand. Voltage regulation circuits are an integral part of just about every piece of modern electronic equipment. Due to their wide use, a number of specialized voltage regulator ICs are available from a variety of manufacturers.

Voltage regulation may be achieved through two main methods. These methods are linear regulation and switching regulation. In both cases, a portion of the output voltage is compared to a stable internal reference. The result of this comparison is used to drive a control element, usually a power transistor. If the output voltage is too low, the control element allows more current to flow to the load from the rectified AC source. Conversely, if the output is too high, the control element constricts the current flow. In the case of the linear regulator, the control element is always in the active, or linear, state. Because of this, the linear regulator tends to dissipate quite a bit of power and, as a result, is rather inefficient. On the plus side, the linear regulator is able to quickly react to load variations, and thus exhibits good transient response.

In contrast to the linear regulator, the control device in the switching regulator is either fully on or fully off. As a result, its power dissipation tends to be reduced. For best performance, fast control devices are needed. The control device is driven by a pulse-width modulator. The output of this modulator is a rectangular pulse whose duty cycle is proportional to the load-current demand. As the control device produces current pulses instead of a constant current, some means of smoothing the pulses is necessary. This function is performed by an ◈◈ filter. The main advantage of the switching regulator is its high efficiency. On the down side, switching regulators are somewhat more difficult to design, do not respond as fast to transient load conditions, and tend to radiate high frequency interference.

No matter what type of regulator is used, power dissipation can be rather large in the control device, so heat sinking is generally advisable. Heat sinks allow for a more efficient transferal of heat energy to the surrounding atmosphere than the control device exhibits on its own.

12.8 PROBLEMS

REVIEW QUESTIONS

1. What is the function of a voltage regulator?
2. What is the difference between load regulation and line regulation?
3. Why do regulators need a reference voltage?
4. What is the functional difference between a linear regulator and a switching regulator?
5. What are the main advantages of using linear regulators versus switching regulators?
6. What are the main advantages of using switching regulators versus linear regulators?
7. What is the function of a pass transistor?
8. Describe two ways in which to increase the output current of an IC-based regulator.
9. How can fixed "three-pin" regulators be used to regulate at other than their rated voltage?
10. What is the purpose of the output inductor and capacitor in the switching regulator?
11. Explain the correlation between the output current demand and the pulsewidth modulator used in switching regulators.
12. What is the purpose of a heat sink?
13. What is meant by the term thermal resistance?
14. What are the thermal resistance elements that control heat flow in a typical power-device/heat-sink connection?
15. What are the general rules that should be considered when using heat sinks?

PROBLEMS

Analysis Problems

1. If the average input voltage to the circuit of Problem 9 is 22 V, determine the maximum device dissipation for a 900 mA output.
2. If the average input voltage is 25 V for the circuit of Problem 11, determine the maximum output current for each output voltage. Use the TO-220 case style (◇◇ = 15W, ◇◇◇◇◇◇ = 1.5A).
3. Draw a block diagram of a complete ± 12 V regulated power supply using LM78XX and LM79XX series parts.
4. Determine the maximum allowable thermal resistance for a heat sink given the following: Ambient temperature = 50 ° C, maximum operating temperature = 150 ° C, TO-3 case style

with thermal grease and Thermalfilm isolator, power dissipation is 30 W, and the device's thermal resistance is 1.1 C° /W, junction to case.

5. A pass transistor has the following specifications: maximum junction temperature = 125° C, TO-220 case, junction to case thermal resistance = 1.5 C° /W. Determine the maximum power dissipation allowed if this device is connected to a 20 C° /W heat sink with thermal grease, using a 0.003 mica insulator. The ambient temperature is 35° C.
6. The thermal resistance of the LM723 is 25 C° /W, junction to case. Its maximum operating temperature is 150° C. For a maximum dissipation of 500 mW and an ambient temperature of 30° C, determine the maximum allowable thermal resistance for the heat-sink/insulator-interface combination.

Design Problems

7. Using Figure 8.3.1, design a 15 V regulator using a 3.3 V Zener. The Zener bias current should be 2 mA, the output should be capable of 500 mA.
8. Using Figure 8.3.1 as a guide, design a variable power supply regulator with a 5 to 15 V output range using a 3.9 V Zener. ◇◇◇◇◇◇ = 3 mA.
9. Design a +12 V regulator using the LM317. The output current capability should be at least 900 mA.
10. Design a +3 to +15 V regulator using the LM317. The output should be continuously variable.
11. Using the LM317, configure a regulator to produce either +5V, +12V, or +15V.
12. Design a +12 V regulator using the LM7805.
13. Design a +9 V regulator using the LM723. Use a current limit of 100 mA.
14. Design a +5 V regulator with 100 mA current limiting using the LM723.
15. Configure a ± 12 V regulator with 70 mA current limiting. Use the LM326
16. Reconfigure the circuit of Problem 15 for ± 15 V.
17. Using the LM3578A, design a 5 V, 400 mA, regulator. The input voltage is 15 V. Use a discontinuity factor of 0.2, and an oscillator frequency of 75 kHz. No more than 10 mV of ripple is allowed.
18. Repeat Problem 17 for a 9 V output.

Challenge Problems

19. Based on the LM723 adjustable regulator example, design a regulator that will produce a continuously variable output from 5 V to 12 V.
20. The LM317 has a maximum operating temperature of 125° C. The TO-220 case version shows a thermal resistance of 4 C° /W, junction to case. It also shows 50 C° /W, junction to ambient (no heat sink used). Assuming an ambient temperature of 50° C. What is the maximum allowable power dissipation for each setup? Assume that the first version uses a 15 C° /W heat sink with a 2 C° /W case to heat sink interconnection.

21. Forced air cooling of a heat-sink/power-device can significantly aid in removing heat energy. As a rule of thumb, forced air cooling at a velocity of 1000 feet per minute will effectively increase the efficiency of a heat sink by a factor of 5. Assuming such a system is applied to the circuit of Problem 17, calculate the new power dissipation.
22. An LM317 (TO-3) is used for a 5 V, 1 A power supply. The average voltage into the regulator is 12 V. Assume a maximum operating temperature of 125 ° C, and an ambient temperature of 25 ° C. First, determine whether or not a heat sink is required. If it is, determine the maximum acceptable thermal resistance for the heat-sink/insulator combination. For the LM317, thermal resistance = 2.3 C ° /W, junction to case, and 35 C ° /W, junction to ambient.

Computer Simulation Problems

23. Using a simulator, plot the time-domain response of the circuit of Figure 8.3.10, assuming an input of 22 V with 3 V peak ripple. How does the simulation change if the ripple is increased to 8 V peak?
24. Verify the output waveform for the circuit of Figure 8.3.14 using a simulator. Use various loads in order to test the current limit operation. The source is 18 V DC, with 2 V peak ripple.
25. Verify the adjustment range for the regulator designed in Example 8.3.5 in the text using a simulator. Use a load of 200 Ω , and a source equal to 10 V, with 1 V peak ripple.
26. Use several different loads with a simulator in order to test the current limit portion of the regulator designed in Example 8.3.5 in the text.

UNIT 13: OSCILLATORS AND FREQUENCY GENERATORS

Learning Objectives

After completing this chapter, you should be able to:

- Explain the differences between positive and negative feedback.
- Define the Barkhausen criterion, and relate it to individual circuits.
- Detail the need for level-detecting circuitry in a practical oscillator.
- Analyze the operation of Wien bridge and phase shift op amp oscillators.
- Analyze the operation of comparator-based op amp oscillators.
- Detail which factors contribute to the accuracy of a given oscillator.
- Explain the operation of a VCO.
- Explain the operation of a PLL, and define the terms capture range and lock range.
- Explain the operation of basic timers and waveform generators.

13.1 INTRODUCTION

Oscillators are signal sources. Many times, it is necessary to generate waveforms with a known wave shape, frequency, and amplitude. A laboratory signal generator perhaps first comes to mind, but there are many other applications. Signal sources are needed to create and receive radio and television signals, to time events, and to create electronic music, among other uses. Oscillators may produce very low frequencies (a fraction of a cycle per second) to very high frequencies (microwaves, > 1 GHz). Oscillators employing op amps are generally used in the area below 1 MHz. Specialized linear circuits may be used at much higher frequencies. The output wave shape may be sinusoidal, triangular, pulse, or some other shape. Oscillators can generally be broken into two broad categories: fixed frequency or variable. For many fixed frequency oscillators, absolute accuracy and freedom from drift are of prime importance. For variable oscillators, ease of tuning and repeatability are usually important. Also, a variable frequency oscillator might not be directly controlled by human hands, rather, the oscillator may be tuned by another circuit. A VCO, or voltage-controlled oscillator, is one example of this. Depending on the application, other factors such as total harmonic distortion or rise time may be important.

No matter what the application or how the oscillator design is realized, the oscillator circuit will normally employ positive feedback. Unlike negative feedback, positive feedback is regenerative – it reinforces change. Generally speaking, without some form of positive feedback, oscillators could not be built. In this chapter we are going to look at positive feedback and the requirements for oscillation. A variety of small oscillators based on op amps will be examined. Finally, more powerful integrated circuits will be discussed.

13.2 OP AMP OSCILLATORS

POSITIVE FEEDBACK AND THE BARKHAUSEN CRITERION

In earlier work, we examined the concept of negative feedback. Here, a portion of the output signal is sent back to the input and summed out of phase with the input signal. The difference between the two signals then, is what is amplified. The result is stability in the circuit response because the large open-loop gain effectively forces the difference signal to be very small. Something quite different happens if the feedback signal is summed in phase with the input signal, as shown in Figure 9.2.1 . In this case, the combined signal looks just like the output signal. As long as the open-loop gain of the amplifier is larger than the feedback factor, the signal can be constantly regenerated. This means that the signal source can be removed. In effect, the output of the circuit is used to create its own input. As long as power is maintained to the circuit, the output signal will continue practically forever. This self-perpetuating state is called oscillation. Oscillation will cease if the product of open loop gain and feedback factor falls below unity or if the feedback signal is not returned perfectly in phase (0 ° or some integer multiple of 360 °). This combination of factors is called the Barkhausen Oscillation Criterion. We may state this as follows:

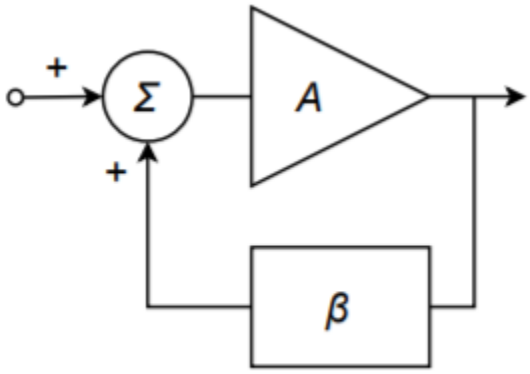


Figure 9.2.1 : Positive feedback.

In order to maintain self-oscillation, the closed-loop gain must be unity or greater, and the loop phase must be ◈ 360 ° , where ◈=0,1,2,3…

Note that when we examined linear amplifiers, we looked at this from the opposite end. Normally, you don't want amplifiers to oscillate, and thus you try to guarantee that the Barkhausen Criterion is never met by setting appropriate gain and phase margins.

A good example of positive feedback is the "squeal" sometimes heard from improperly adjusted public address systems. Basically, the microphone is constantly picking up the ambient room noise, which is then amplified and fed to the loudspeakers. If the amplifier gain is high enough or if the acoustic loss is low enough (i.e., the loudspeaker is physically close to the microphone), the signal that the microphone picks up from the loudspeaker can be greater than the ambient noise. The result is

that the signal constantly grows in proper phase to maintain oscillation. The result is the familiar squealing sound. In order to stop the squeal, either the gain or the phase must be disrupted. Moving the microphone may change the relative phase, but it is usually easier to just reduce the volume a little. It is particularly interesting to listen to a system that is on the verge of oscillation. Either the gain or the phase is just not quite perfect, and the result is a rather irritating ringing sound, as the oscillation dies out after each word or phrase.

There are a couple of practical considerations to be aware of when designing oscillators. First of all, it is not necessary to provide a "start-up" signal source as seen in Figure 9.2.1 . Normally, there is enough energy in either the input noise level or possibly in a turn-on transient to get the oscillator started. Both the turn-on transient and the noise signal are broad spectrum signals, so the desired oscillation frequency is contained within either of them. The oscillation signal will start to increase as time progresses due to the closed-loop gain being greater than one. Eventually, the signal will reach a point where further level increases are impossible due to amplifier clipping. For a more controlled, low-distortion oscillator, it is desirable to have the gain start to roll off before clipping occurs. In other words, the closed-loop gain should fall back to exactly one. Finally, in order to minimize frequency drift over time, the feedback network should be selective. Frequencies either above or below the target frequency should see greater attenuation than the target frequency. Generally, the more selective (i.e., higher ◈) this network is, the more stable and accurate the oscillation frequency will be. One simple solution is to use an ◈◈◈ tank circuit in the feedback network. Another possibility is to use a piezoelectric crystal. A block diagram of a practical oscillator is shown in Figure 9.2.2 .

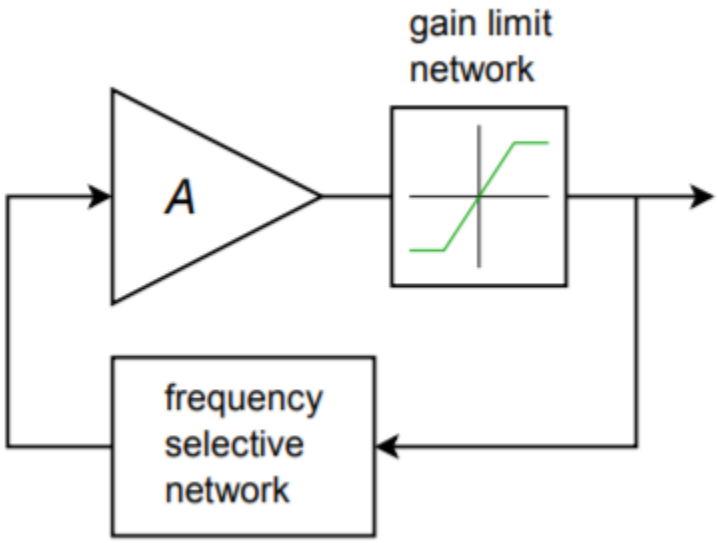


Figure 9.2.2 : Practical oscillator

A BASIC OSCILLATOR

A real-world circuit that embodies all of the elements is shown in Figure 9.2.3 . This circuit is not particularly efficient or cost-effective, but it does illustrate the important points. Remember, in order to maintain oscillation the closed-loop gain of the oscillator circuit must be greater than 1, and the loop phase must be a multiple of 360 ° .

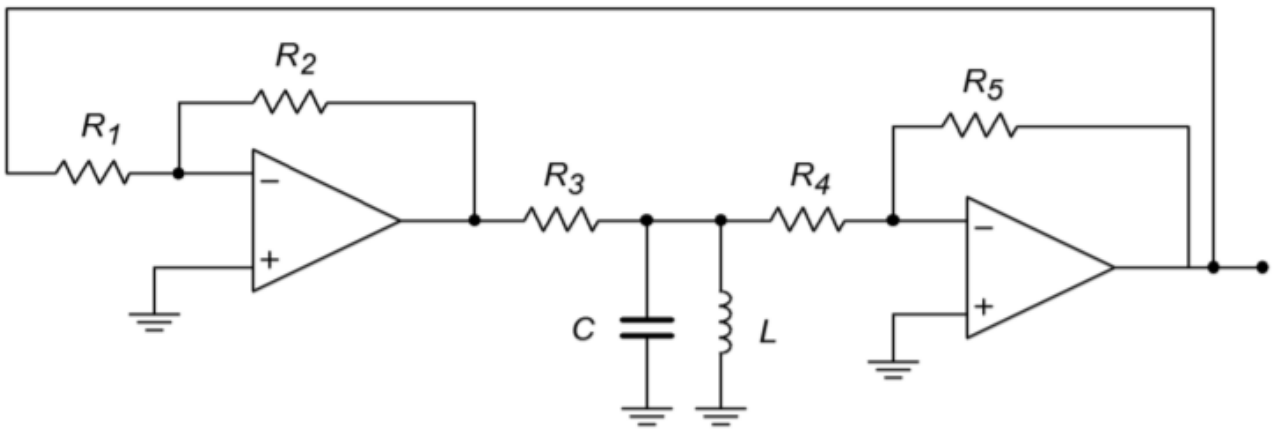


Figure 9.2.3 : A basic oscillator.

To provide gain, a pair of inverting amplifiers is used. Note op amp 2 serves to buffer the output signal. As each stage produces a 180 ° shift, the shift for the pair is 360 ° . The product of the gains has to be larger than the loss produced by the frequency selecting network. This network is made up of ⍰3 , ⍰ , and ⍰ . Because the ⍰⍰ combination produces an impedance peak at the resonant frequency, ⍰⍰ , a minimum loss will occur there. Also, at resonance, the circuit is basically resistive, so no phase change occurs. Consequently, this circuit should oscillate at the fo set by ⍰ and ⍰ . This circuit can be easily tested in lab. For example, if you drop the gain of one of the op amp stages, there will not be enough system gain to over come the tank circuit's loss, and thus, oscillation will cease. You can also verify the phase requirement by replacing one of the inverting amplifiers with a noninverting amplifier of equal gain. The resulting loop phase of 180 ° will halt oscillation. This circuit does not include any form of automatic gain adjustment, so the output signal may be clipped. If properly chosen, the slew rate of the op amp may be used as the limit factor. (A 741 will work acceptably for ⍰⍰ in the low kHz range). Although this circuit does work and points out the specifics, it is certainly not a top choice for an oscillator design based on op amps.

WIEN BRIDGE OSCILLATOR

A relatively straightforward design useful for general-purpose work is the Wien bridge oscillator. This oscillator is far simpler than the generalized design shown in Figure 9.2.3 , and offers very good performance. The frequency selecting network is a simple lead/lag circuit, such as that shown in Figure 9.2.4 . This circuit is a frequency-sensitive voltage divider. It combines the response of both the simple lead and lag networks. Normally, both resistors are set to the same value. The same may be said for the two capacitors. At very low frequencies, the capacitive reactance is essentially infinite, and thus, the upper series capacitor appears as an open. Because of this, the output voltage is zero. Likewise, at very high frequencies the capacitive reactance approaches zero, and the lower shunt capacitor effectively shorts the output to ground. Again, the output voltage is zero. At some middle frequency the output voltage will be at a peak. This will be the preferred, or selected, frequency and will become the oscillation frequency so long as the proper phase relation is held. We need to determine the phase change at this point as well as the voltage divider ratio. These items are needed in order to guarantee that the Barkhausen conditions are met.

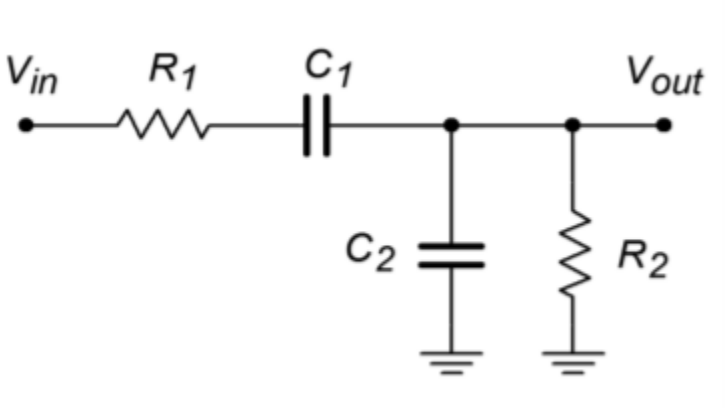


Figure 9.2.4 : Lead/lag network.

First, note that

$$\beta = \frac{Z_2}{Z_1 + Z_2}$$

where ◈1=◈1–◈◈◈1 , and ◈2=◈2||–◈◈◈2 .

$$Z_2 = \frac{-jX_{C_2}R_2}{-jX_{C_2} + R_2}$$

$$Z_2 = \frac{-jX_{C_2}R_2}{-jX_{C_2}(1 + \frac{R_2}{-jX_{C_2}})}$$

$$Z_2 = \frac{R_2}{1 + \frac{R_2}{-jX_{C_2}}}$$

Recalling that ◈◈=1/◈◈ , we find

$$Z_1 = R_1 - \frac{j}{\omega C_1}$$

$$Z_2 = \frac{R_2}{1 + j\omega R_2 C_2}$$

So,

$$\beta = \frac{\frac{R_2}{1+j\omega R_2 C_2}}{\frac{R_2}{1+j\omega R_2 C_2} + R_1 - \frac{j}{\omega C_1}}$$

$$\beta = \frac{R_2}{R_2 + R_1 - \frac{j}{\omega C_1} + j\omega R_1 R_2 C_2 + \frac{R_2 C_2}{C_1}}$$

$$\beta = \frac{R_2}{R_2\left(1+\frac{C_2}{C_1}\right)+R_1+j\left(\omega R_1 R_2 C_2 - \frac{1}{\omega C_1}\right)}$$

We can determine the desired frequency from the imaginary portion of Equation 9.2.1 .

$$\omega R_1 R_2 C_2 = \frac{1}{\omega C_1}$$

$$\omega^2 = \frac{1}{R_1 R_2 C_1 C_2}$$

$$\omega = \frac{1}{\sqrt{R_1 R_2 C_1 C_2}}$$

(9.2.2)

Normally, ◈1=◈2 and ◈1=◈2 , so Equation 9.2.2 reduces to

$$\omega = \frac{1}{RC} \text{ or,}$$

$$f_o = \frac{1}{2\pi RC}$$

(9.2.3)

To find the magnitude of the feedback factor, and thus the required forward gain of the op amp, we need to examine the real portion of Equation 9.2.1 .

$$\beta = \frac{R_2}{R_2\left(1+\frac{C_2}{C_1}\right)+R1}$$

Assuming that equal components are used, this reduces to

$$\beta = \frac{R}{3R} \text{ or simply,}$$

$$\beta = \frac{1}{3}$$

The end result is that the forward gain of the amplifier must have a gain of slightly over 3 and a phase of 0° in order to maintain oscillation. It also means that the frequency of oscillation is fairly easy to set and can even be adjusted if potentiometers are used to replace the two resistors. The final circuit is shown in Figure 9.2.5 .

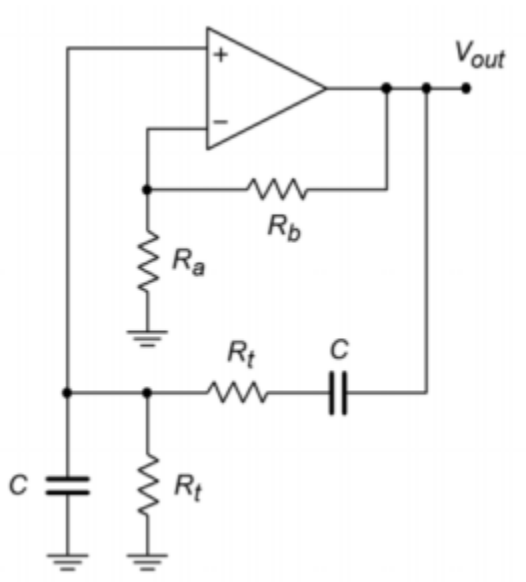


Figure 9.2.5 : Wien bridge oscillator.

This circuit uses a combination of negative feedback and positive feedback to achieve oscillation. The positive feedback loop utilizes ⍰⍰ and ⍰ . The negative feedback loop utilizes ⍰⍰ and ⍰⍰ . ⍰⍰ must be approximately twice the size of ⍰⍰ . If it is smaller, the ⍰⍰ product will be less than unity and oscillation cannot be maintained. If the gain is significantly larger, excessive distortion may result. Indeed, some form of gain reduction at higher output voltages is desired for this circuit. One possibility is to replace ⍰⍰ with a lamp. As the signal amplitude increases across the lamp, its resistance increases, thus decreasing gain. At a certain point the lamp's resistance will be just enough to produce an ⍰⍰ product of exactly 1. Another technique is shown in Figure 9.2.6 . Here an opposite approach is taken. Resistor ⍰⍰ is first broken into two parts, the smaller part, ⍰⍰2 , is shunted by a pair of signal diodes. For lower amplitudes, the diodes are off and do not affect the circuit operation. At higher amplitudes, the diodes start to turn on, and thus start to short ⍰⍰2 . If correctly implemented, this action is not instantaneous and does not produce clipping. It simply serves to reduce the gain at higher amplitudes.

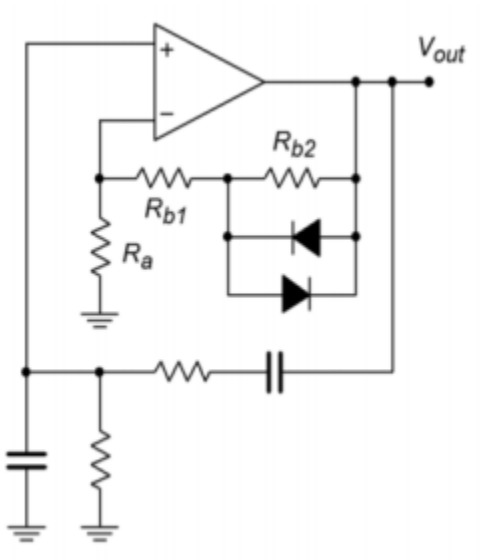


Figure 9.2.6 : Wien bridge oscillator with gain adjustment.

Another way of drawing the Wien bridge oscillator is shown in Figure 9.2.7 . This form clearly shows the Wien bridge configuration. Note that the output of the bridge is the differential input voltage (i.e., error voltage). In operation, the bridge is balanced, and thus, the error voltage is zero.

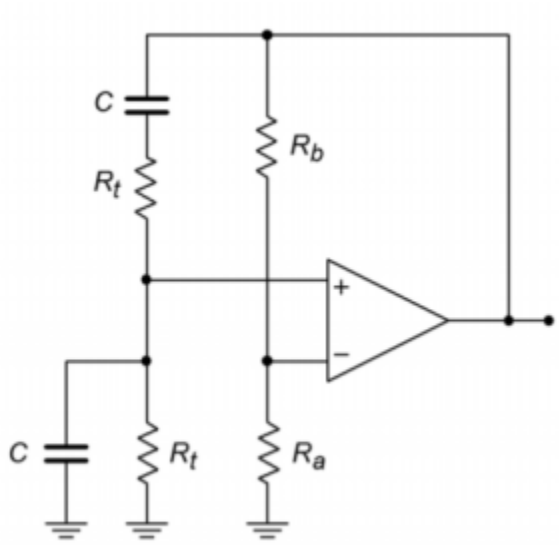


Figure 9.2.7 : Wien bridge oscillator redrawn.

Example 9.2.1

Determine the frequency of oscillation for the circuit of Figure 9.2.8 .

$$f_o = \frac{1}{2\pi RC}$$

$$f_o = \frac{1}{2\pi \times 50k \times .01\mu F}$$

$$f_o = 318Hz$$

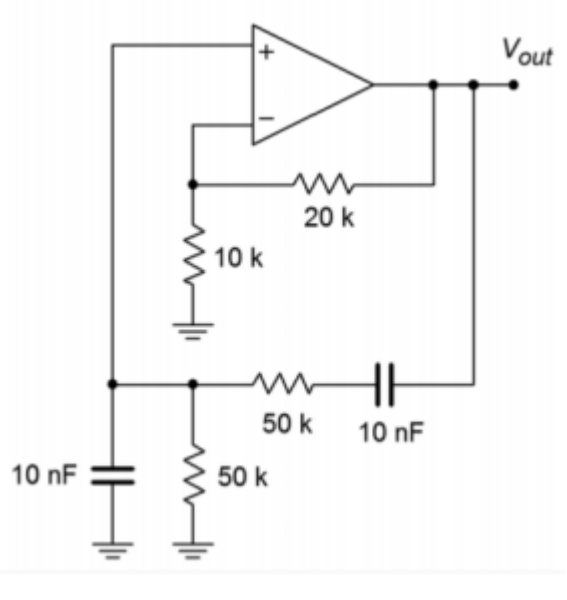


Figure 9.2.8 : Oscillator for Example 9.2.1 .

For other frequencies, either ⍰ or ⍰ may be altered as required. Also, note that the forward gain works out to exactly 3, thus perfectly compensating the positive feedback factor of 1/3. In reality, component tolerances make this circuit impractical. To overcome this difficulty, a small resistor/diode combination may be placed in series with the 20 k Ω , as shown in Figure 9.2.6 . A typical resistor value would be about one-fourth to one-half the value of ⍰⍰ , or about 5 k Ω to 10 k Ω in this example. ⍰⍰ would be decreased slightly as well (or, ⍰⍰ might be increased).

The ultimate accuracy of ⍰⍰ depends on the tolerances of ⍰ and ⍰ . If 10% parts are used in production, a variance of about 20% is possible. Also, at higher frequencies, the op amp circuit will produce a moderate phase shift of its own. Thus, the assumption of a perfect noninverting amplifier is no longer valid, and some error in the output frequency will result. With extreme values in the positive feedback network, it is also possible that some shift of the output frequency may occur due to the capacitive and resistive loading effects of the op amp. Normally, this type of loading is not a problem, as the op amp's input resistance is very high, and its input capacitance is quite low.

Example 9.2.2

Figure 9.2.9 shows an adjustable oscillator. Three sets of capacitors are used to change the frequency range, whereas a dual-gang potentiometer is used to adjust the frequency within a given range. Determine the maximum and minimum frequency of oscillation within each range.

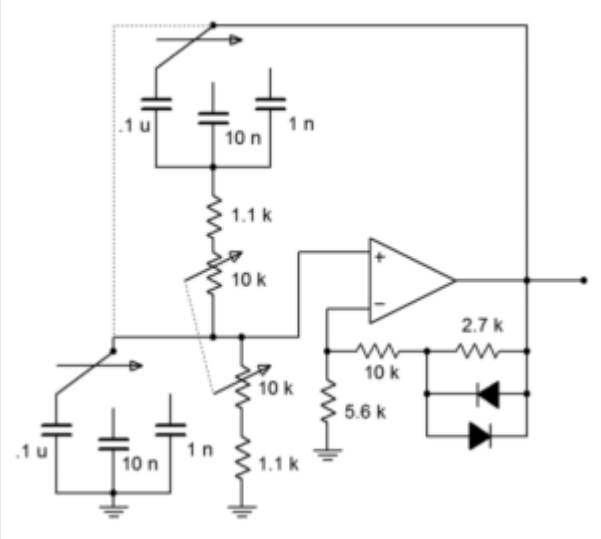


Figure 9.2.9 : Adjustable oscillator.

First, note that the capacitors are spaced by decades. This means that the resulting frequency ranges will also change by factors of 10. The 0.1 ◈ F capacitor will produce the lowest range, the 10 nF will produce a range 10 times higher, and the 1 nF range will be 10 times higher still. Thus, we only need to calculate the range produced by the 0.1 ◈ F.

The maximum frequency of oscillation within a given range will occur with the lowest possible resistance. The minimum resistance is seen when the 10 k Ω pot is fully shorted, the result being 1.1 k Ω . Conversely, the minimum frequency will occur with the largest resistance. When the pot is fully in the circuit, the resulting sum is 11.1 k Ω . Note that a dual-gang potentiometer means that both units are connected to a common shaft; thus, both pots track in tandem.

For ◈◈◈◈◈◈◈◈ with 0.1 ◈ F:

$$f_o = \frac{1}{2\pi RC}$$

$$f_o = \frac{1}{2\pi \times 11.1k \times 0.1\mu F}$$

$$f_o = 143.4Hz$$

For ◈◈◈◈◈◈◈◈ with 0.1 ◈ F:

$$f_o = \frac{1}{2\pi RC}$$

$$f_o = \frac{1}{2\pi \times 1.1k \times 0.1\mu F}$$

$$f_o = 1.447kHz$$

For the 0.01 ◈ F, the ranges would be 1.434 kHz to 14.47 kHz, and for the 0.001 ◈ F, the ranges would be 14.34 kHz to 144.7 kHz. Note that each range picks up where the previous one left off. In this way, there are no "gaps", or unobtainable frequencies. For stable oscillation, this circuit must have a gain of 3. For low-level outputs, the diodes will not be active, and the forward gain will be

$$A_v = 1 + \frac{R_f}{R_i}$$

$$A_v = 1 + \frac{10k + 2.7k}{5.6k}$$

$$A_v = 3.27$$

As the signal rises, the diodes begin to turn on, thus shunting the 2.2 k Ω resistor and dropping the gain back to exactly 3.

PHASE SHIFT OSCILLATOR

Considering the Barkhausen Criterion, it should be possible to create an oscillator by using a simple phase shift network in the feedback path. For example, if the circuit uses an inverting amplifier (-180 ° shift), a feedback network with an additional 180 ° shift should create oscillation.

The only other requirement is that the gain of the inverting amplifier be greater than the loss produced by the feedback network. This is illustrated in Figure 9.2.10 . The feedback network can be as simple as three cascaded lead networks. The lead networks will produce a combined phase shift of 180 ° at only one frequency. This will become the frequency of oscillation. In general, the feedback network will look something like the circuit of Figure 9.2.11 . This form of ◈◈ layout is usually referred to as a ladder network.

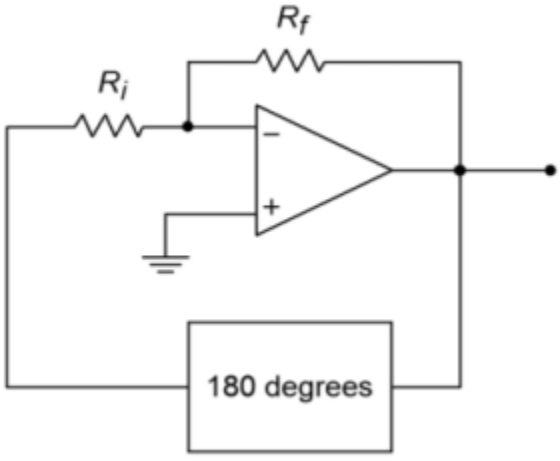


Figure 9.2.10 : Block diagram of phase shift oscillator.

There are many ways in which ◈ and ◈ may be set in order to create the desired 180 ° shift. Perhaps the most obvious scheme is to set each stage for a 60 ° shift. The components are determined by

finding a combination that produces a 60 ° shift at the desired frequency. In order to avoid loading effects, each stage must be set to higher and higher impedances. For example, ◈2 might be set to 10 times the value of ◈1 , and ◈3 set to 10 times the value of ◈2 . The capacitors would see a corresponding decrease. Because the tangent of the phase shift yields the ratio of ◈◈ to ◈ , at our desired 60 ° we find

$$\tan 60 = \frac{X_C}{R}$$

$$1.732 = \frac{X_C}{R}$$

$$X_C = 1.732R$$

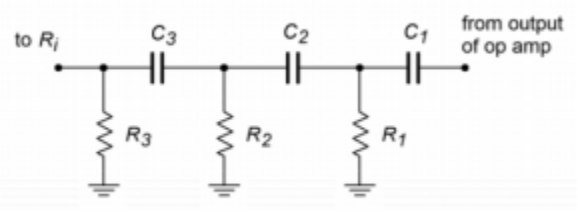


Figure 9.2.11 : Phase shift network

Using this in the general reactance formula produces

$$f_o = \frac{1}{2\pi 1.732RC}$$

(9.2.4)

Likewise, a look at the magnitude shows the approximate loss per stage of

$$\beta = \frac{R}{\sqrt{R^2 + X_C^2}}$$

$$\beta = \frac{R}{\sqrt{R^2 + (1.732R)^2}}$$

$$\beta = .5$$

Because there are three stages, the total loss for the feedback network would be 0.125. Therefore, the inverting amplifier needs a gain of 8 in order to set the ◈◈ product to unity. Remember, these results are approximate and depend on minimum interstage loading. A more exacting analysis will follow shortly.

Example 9.2.3

Determine the frequency of oscillation in Figure 9.2.12 .

$$f_o = \frac{1}{2\pi 1.732 RC}$$

$$f_o = \frac{1}{2\pi \times 1.732 \times 1k \times 100nF}$$

$$f_o = 919Hz$$

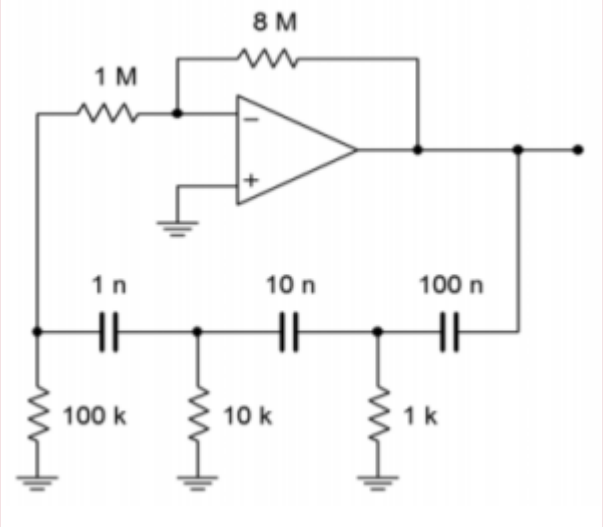


Figure 9.2.12 : Phase shift oscillator (minimum loading form).

Figure 9.2.12 graphically points up the major problem with the "60 ° per stage" concept. In order to prevent loading, the final resistors must be very high. In this case a feedback resistor of 8 M Ω is required. It is possible to simplify the circuit somewhat by omitting the 1 M Ω and connecting the 100 k Ω directly to the op amp as shown in Figure 9.2.13 . This saves one part and does allow the feedback resistor to be dropped in value, but the resulting component spread is still not ideal.

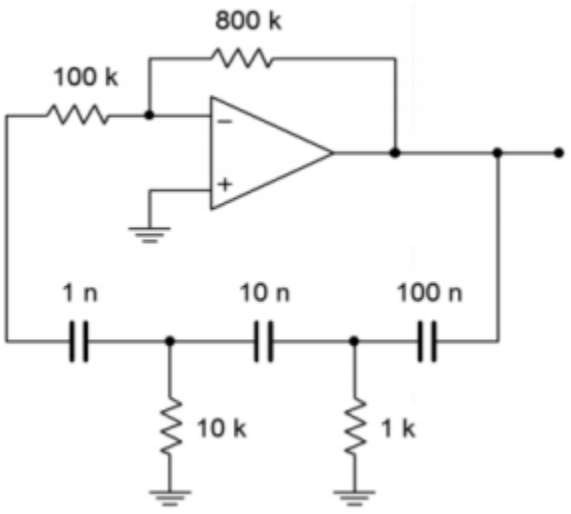


Figure 9.2.13 : Improved phase shift oscillator.

If we accept the loading effect, we can simplify things a bit by making each resistor and capacitor the same size, as shown in Figure 9.2.14 . With these values, we can be assured that the resulting frequency will no longer be 919 Hz, as Equation 9.2.4 is no longer valid. Also, it is quite likely that the loss produced by the network will no longer be 0.125.

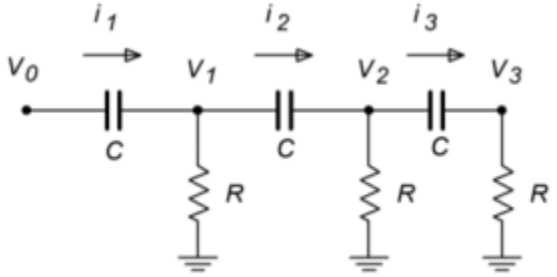


Figure 9.2.14 : Phase shift analysis.

We need to determine the general input/output (◈0/◈3) ratio of the ladder network, and from this, find the gain and frequency relations for a net phase shift of -180 ° . One technique involves the use of simultaneous loop equations. Because all resistors and capacitors are equal in this variation, we will be able to simplify our equations readily. By inspection, the three loop equations are (from left to right):

$$V_0 = (R + X_C)I_1 - RI_2$$

(9.2.5)

$$0 = -RI_1 + (2R + X_C)I_2 - RI_3$$

(9.2.6)

$$0 = -RI_2 + (2R + X_C)I_3$$

(9.2.7)

Also, note that

$$V_3 = I_3R$$

(9.2.8)

We now have expressions for ◈0 and ◈3 , however, ◈0 is in terms of ◈1 and ◈2 , and ◈3 is in terms of ◈3 . Write ◈1 and ◈2 in terms of ◈3 so that we can substitute these back into Equation 9.2.5 . Rewriting Equation 9.2.7 yields an expression for ◈2

$$I_2 = I_3\left(2 + \frac{X_C}{R}\right)$$

(9.2.9)

For ◈1 , rewrite Equation 9.2.6

$$I_1 = \left(2 + \frac{X_C}{R}\right)I_2 - I_3$$

(9.2.10)

Substituting Equation 9.2.9 into Equation 9.2.10 yields

$$I_1 = I_3 \left(\left(2 + \frac{X_C}{R} \right)^2 - 1 \right)$$

(9.2.11)

Thus, ◈0 may be rewritten as

$$V_0 = I_3(R + X_C) \left(\left(2 + \frac{X_C}{R} \right)^2 - 1 \right) - I_3 R \left(2 + \frac{X_C}{R} \right)$$

(9.2.12)

Equation 9.2.12 may be simplified to

$$V_0 = I_3 \left(R \left(2 + \frac{X_C}{R} \right)^2 + X_C \left(2 + \frac{X_C}{R} \right)^2 - 3R - 2X_C \right)$$

(9.2.13)

The input/output expression is simplified as follows.

$$\frac{V_0}{V_3} = \left(R \left(2 + \frac{X_C}{R} \right)^2 + X_C \left(2 + \frac{X_C}{R} \right)^2 - 3R - 2X_C \right) \left(\frac{1}{R} \right)$$

(9.2.14)

$$\frac{V_0}{V_3} = \left(2 + \frac{X_C}{R} \right)^2 + \frac{X_C}{R} (2 + XCR)^2 - 3 - 2\frac{X_C}{R}$$

$$\frac{V_0}{V_3} = 1 + \frac{6X_C}{R} + 5\frac{X_C^2}{R^2} + \frac{X_C^3}{R^3}$$

(9.2.15)

At this point, we are nearly done with the general equation. All that is left is to substitute 1/◈◈◈ in place of ◈◈ . Remember, ◈2=–1

$$\frac{V_0}{V_3} = 1 + \frac{6}{j\omega CR} - \frac{5}{\omega^2 C^2 R^2} - \frac{1}{j\omega^3 C^3 R^3}$$

(9.2.16)

This Equation contains both real and imaginary terms. For this Equation to be satisfied, the imaginary components (6/◈◈◈◈ and 1/◈◈3◈3◈3) must sum to zero, and the real components must similarly sum to zero (as there are only two terms, their magnitudes must be equal). We can use these facts to find both the gain and the frequency.

$$\frac{6}{j\omega CR} = \frac{1}{j\omega^3 C^3 R^3}$$

$$\frac{1}{j\omega CR} = \frac{1}{6j\omega^3 C^3 R^3}$$

$$1 = \frac{1}{6j\omega^2 C^2 R^2}$$

$$\omega^2 = \frac{1}{6C^2R^2}$$

(9.2.17)

$$\omega = \frac{1}{\sqrt{6}CR} \; or,$$

(9.2.18)

$$f_o = \frac{1}{2\pi\sqrt{6}CR}$$

(9.2.19)

For the gain, we solve Equation 9.2.16 in terms of the voltage ratio and zero the imaginary terms, because the result must be real.

$$\frac{V_0}{V_3} = 1 - \frac{5}{\omega^2 C^2 R^2}$$

(9.2.20)

Substituting Equation 9.2.17 into Equation 9.2.20 yields

$$\frac{V_0}{V_3} = 1 - \frac{5}{\frac{1}{6C^2R^2}C^2R^2}$$

(9.2.21)

$$\frac{V_0}{V_3} = 1 - 5 \times 6$$

$$\frac{V_0}{V_3} = -29$$

(9.2.22)

The gain of the ladder network is ◈3/◈0 , or the reciprocal of Equation 9.2.22 , or

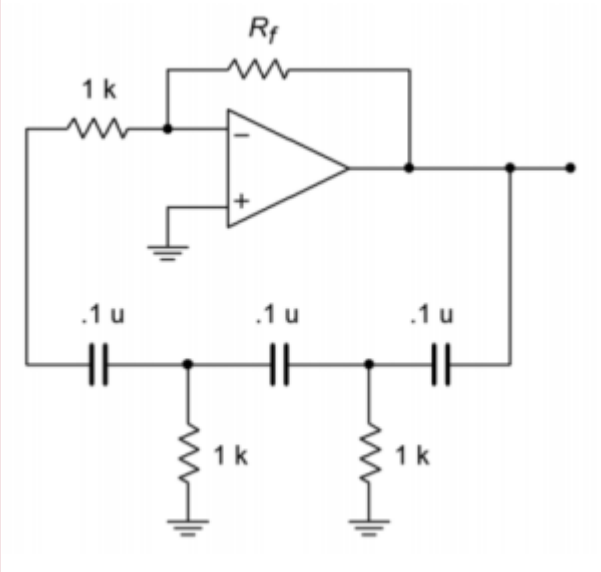
$$\beta = \frac{1}{-29}$$

(9.2.23)

The loss produced will be 1/29. This has the disadvantage of requiring a forward gain of 29 instead of 8 (as in the previous form). This disadvantage is minor compared to the advantage of reasonable component values.

Example 9.2.4

Determine a value for ◈◈ in Figure 9.2.15 in order to maintain oscillation. Also determine the oscillation frequency.



Equation 9.2.23 shows that the inverting amplifier must have a gain of 29.

$$A_v = -\frac{R_f}{R_i}$$

$$R_f = -\frac{R_i}{A_v}$$

$$R_f = -1k \times -29$$

$$R_f = 29k$$

Of course, the higher standard value will be used. Also, in order to control the gain at higher levels, a diode/resistor combination (as used in the Wien bridge circuits) should be placed in series with ◈◈ . Without a gain limiting circuit, excessive distortion may occur.

$$f_o = \frac{1}{2\pi\sqrt{6}RC}$$

$$f_o = \frac{1}{2\pi\sqrt{6} \times 1k \times 0.1\mu F}$$

$$f_o = 650Hz$$

COMPUTER SIMULATION

In order to verify Equations 9.2.3 and 9.2.19 , the gain and phase responses of the feedback networks of Figures 9.2.13 and 9.2.15 are found in Figure 9.2.16 . These graphs were obtained via the standard

Multisim route. Note that the phase for both circuits hits -180 ° very close to the predicted frequencies. The equal value network prediction is very accurate, whereas the staggered network prediction is off by only a few percent. Once the phase exceeds -180 ° , the Multisim grapher wraps it back to +180 ° , so it is very easy to see this frequency. Likewise, the gain response agrees with the derivations for attenuation at the oscillation frequency. It can be very instructive to analyze these circuits for the gain and phase response at each stage as well.

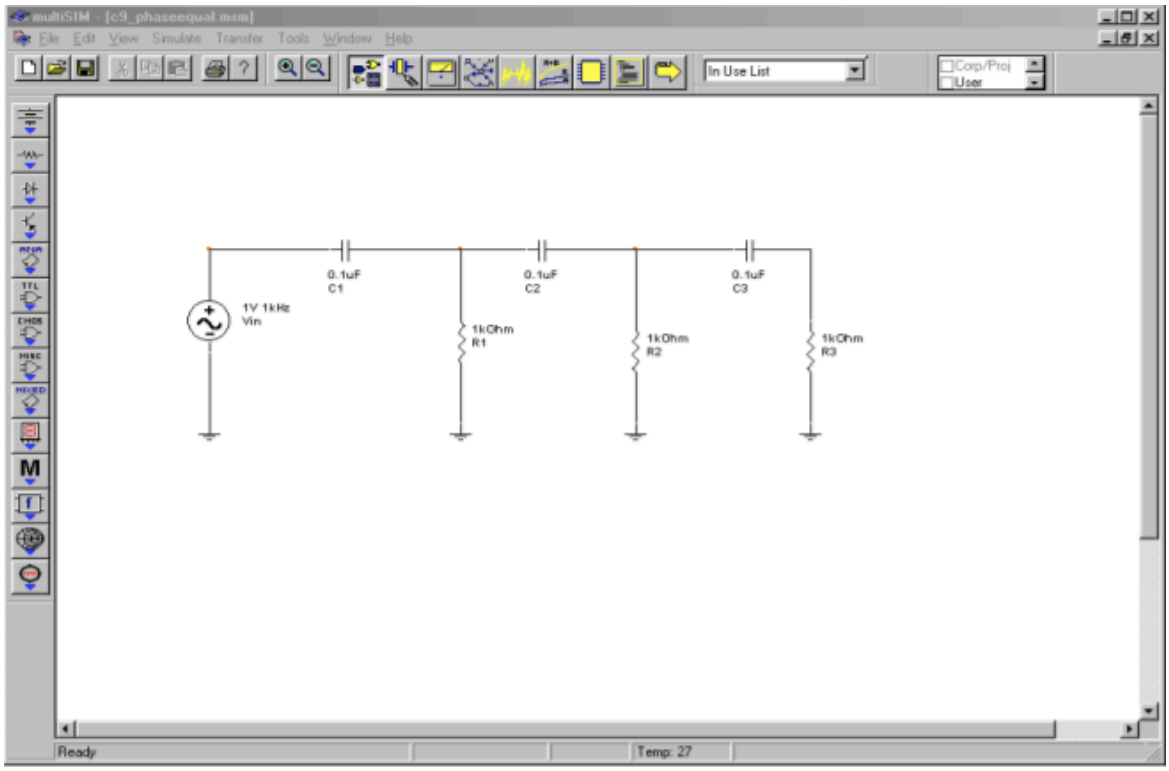


Figure 9.2.16 : Equal value network in Multisim.

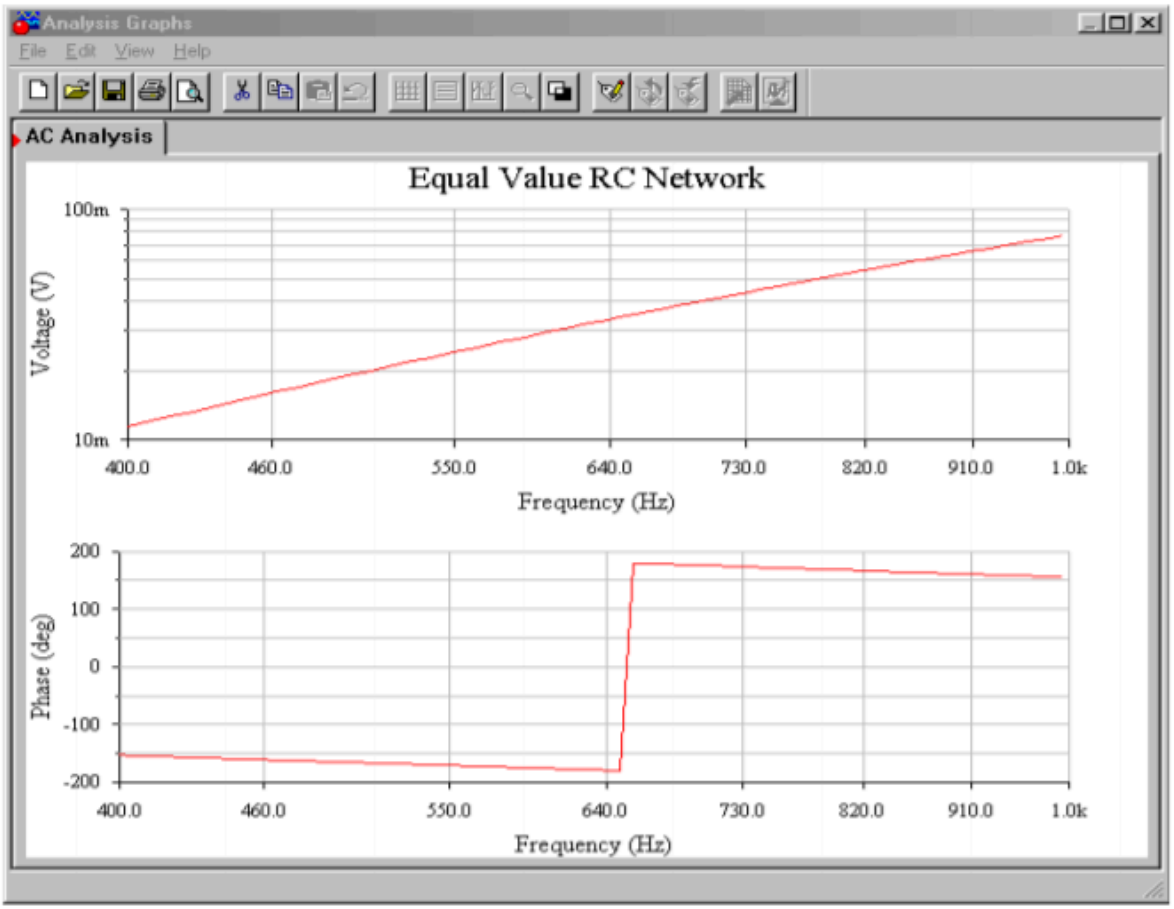


Figure 9.2.16 : Response of equal value network.

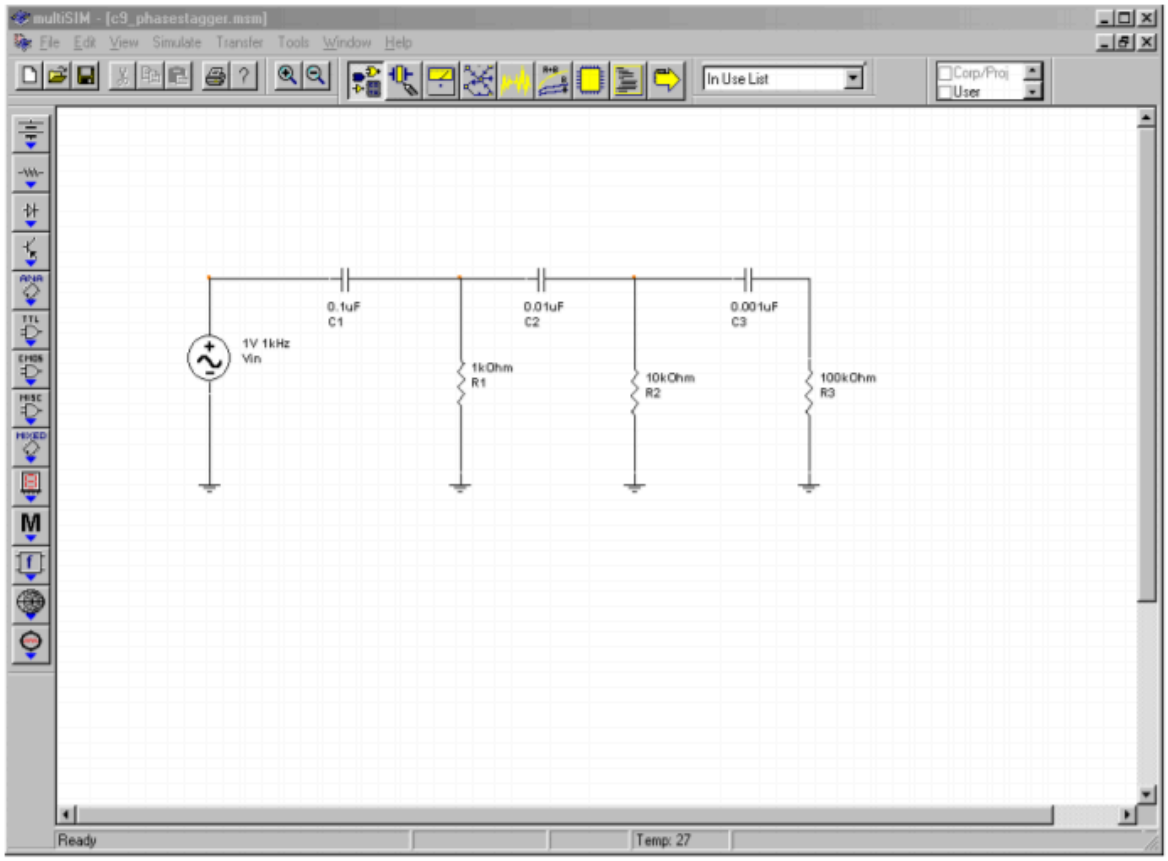


Figure 9.2.16 : Staggered value network in Multisim.

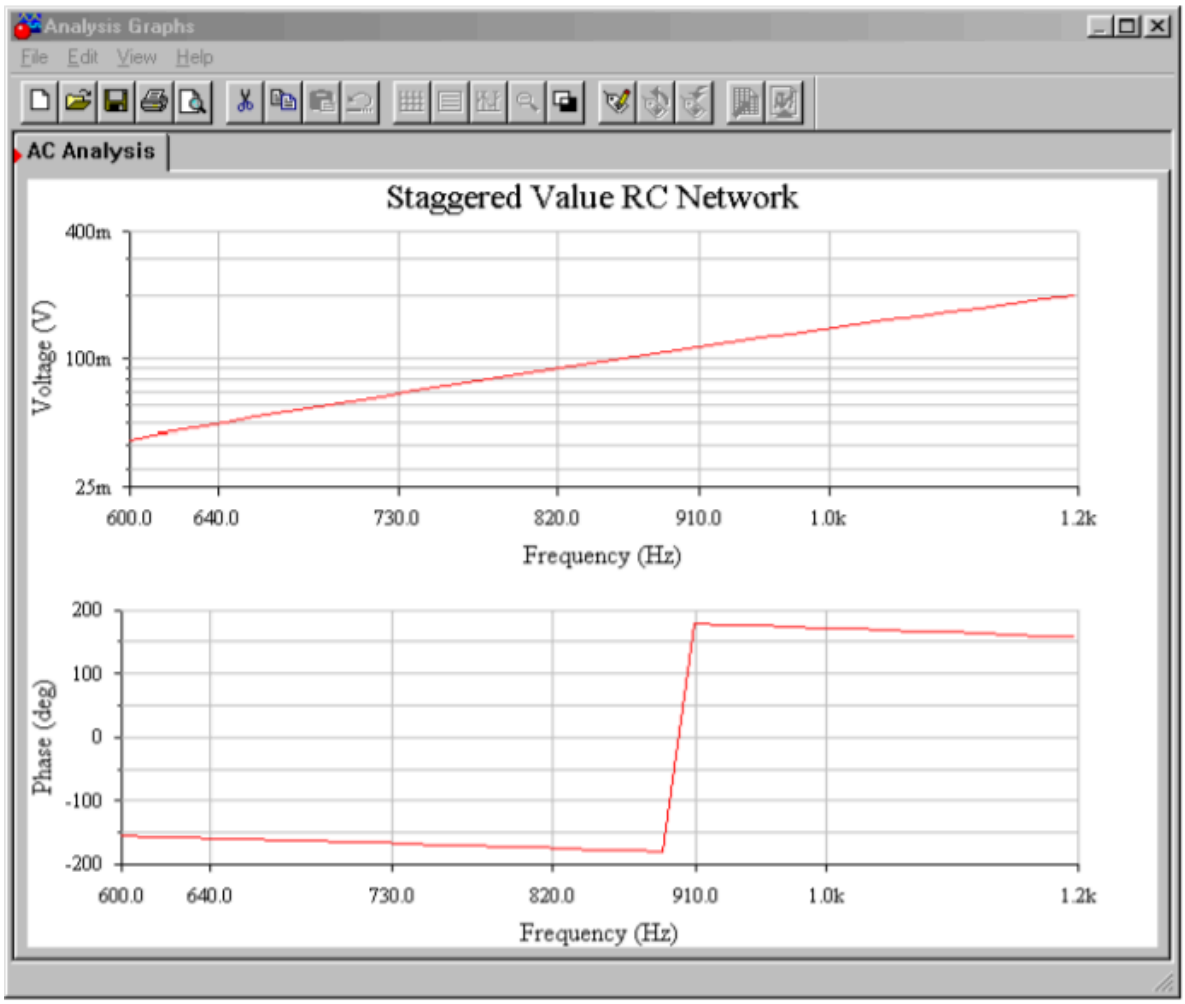


Figure 9.2.16 : Response of staggered value network.

SQUARE/TRIANGLE FUNCTION GENERATOR

Besides generating sine waves, op amp circuits may be employed to generate other wave shapes such as ramps, triangle waves, or pulses. Generally speaking, squareand pulse-type waveforms may be derived from other sources through the use of a comparator. For example, a square wave may be derived from a sine wave by passing it through a comparator, such as those seen in Chapter Seven. Linear waveforms such as triangles and ramps may be derived from the charge/discharge action of a capacitor. As you may recall from basic circuit theory, the voltage across a capacitor will rise linearly if it is driven by a constant current source. One way of achieving this linear rise is with the circuit of Figure 9.2.17 .

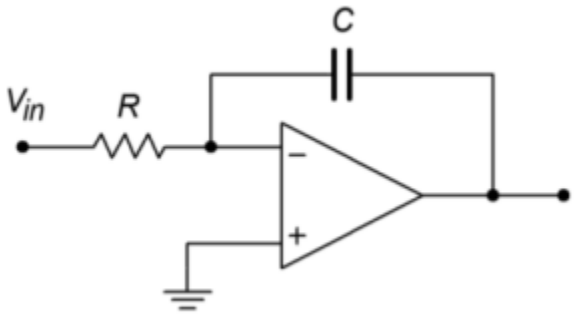


Figure 9.2.17 : Ramp generator.

In essence, this circuit is an inverting amplifier with a capacitor taking the place of ◈◈ . The input resistor, ◈ , turns the applied input voltage into a current. Because the current into the op amp itself is negligible, this current flows directly into capacitor ◈ . As in a normal inverting amplifier, the output voltage is equal to the voltage across the feedback element, though inverted. The relationship between the capacitor current and voltage is

$$\frac{d_v}{d_t} = \frac{i}{C}$$

(9.2.24)

$$V(t) = \frac{1}{C}\int idt$$

$$V_{out} = -\frac{1}{C}\int idt$$

(9.2.25)

As expected, a fast rise can be created by either a small capacitor or a large current. (As a side note, this circuit is called an integrator and will be examined in greater detail in the next chapter.)

By choosing appropriate values for ◈ and ◈ , the ◈◈◈◈ramp may be set at a desired rate. The polarity of the ramp's slope is determined by the direction of the input current; a positive source will produce a negative going ramp and vice versa. If the polarity of the input changes at a certain rate, the output ramp will change direction in tandem. The net effect is a triangle wave. A simple way to generate the alternating input polarity is to drive ◈ with a square wave. As the square wave changes from plus to minus, the ramp changes direction. This is shown in Figure 9.2.18 .

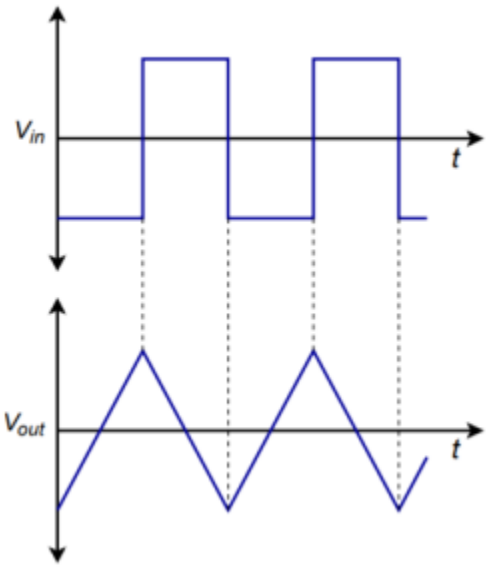


Figure 9.2.18 : Ramp generator waveforms.

So, we are now able to generate a triangle wave. The only problem is that a square wave source is needed. How do we produce the square source? As mentioned earlier, a square wave may be derived by passing an AC signal through a comparator. Logically then, we should be able to pass the output triangle wave into a comparator in order to create the needed square wave. The resulting circuit is shown in Figure 9.2.19 . A comparator with hysteresis is used to turn the triangle into a square wave. The square then drives the ramp circuit. The circuit produces two simultaneous outputs: a square wave that swings to ± saturation and a triangle wave that swings to the upper and lower comparator thresholds. This is shown in Figure 9.2.20 . The thresholds may be determined from the equations presented in Chapter Seven. In order to determine the output frequency, the V/s rate of the ramp is determined from Equation 9.2.24 . Knowing the peak-to-peak swing of the triangle is ◈upper thres–◈lower thres , the period of the wave may be found. The output frequency is the reciprocal of the period.

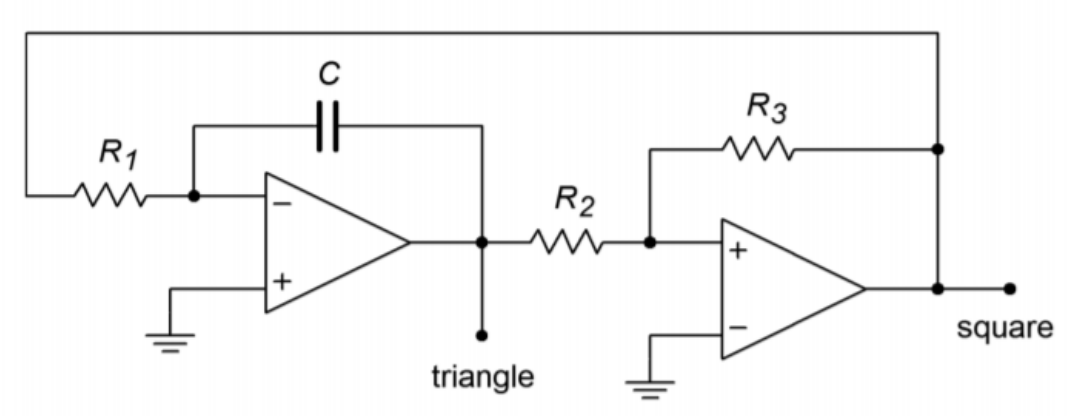


Figure 9.2.19 : Triangle/square generator.

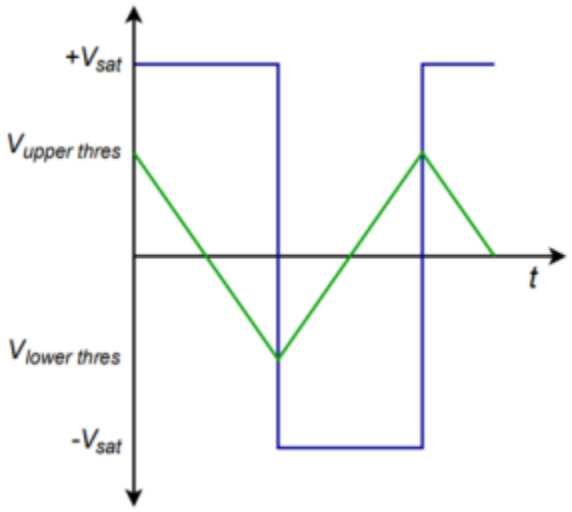


Figure 9.2.20 : Output waveforms of triangle/square generator.

Example 9.2.5

Determine the output frequency and amplitudes for the circuit of Figure 9.2.21 . Use ⍰⍰⍰⍰=±13⍰ .

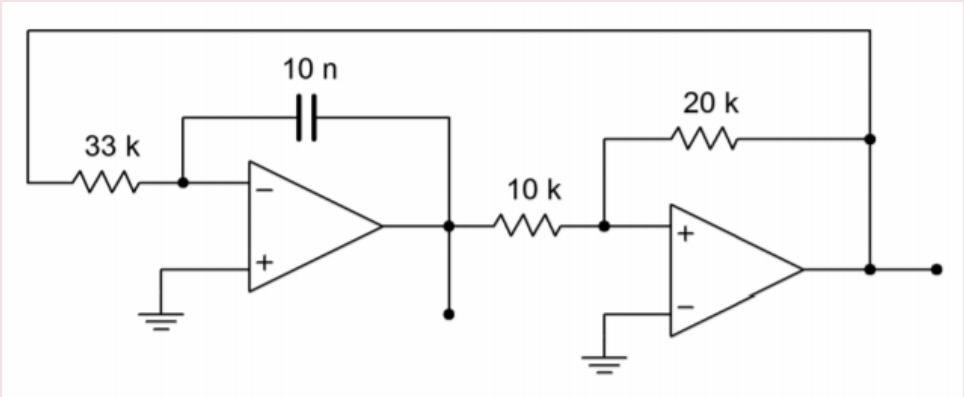


Figure 9.2.21 : Signal generator for Example 9.2.5 .

First, note that the comparator always swings between +⍰⍰⍰⍰ and –⍰⍰⍰⍰ . Now, determine the upper and lower thresholds for the comparator.

$$V_{\text{upper thres}} = V_{sat}\frac{R_2}{R_3}$$

$$V_{\text{upper thres}} = 13V\frac{10k}{20k}$$

$$V_{\text{upper thres}} = 6.5V$$

The lower threshold will be -6.5 V. We now know that the triangle wave output will be 13 V peak-to-peak. From this we may determine the output period.

Because the ramp generator is driven by a square wave with an amplitude of ◈◈◈◈ , Equation 9.2.24 may be rewritten as

$$\frac{d_v}{d_t} = \frac{V_{sat}}{RC}$$

$$\frac{d_v}{d_t} = \frac{13V}{33k \times 0.01\mu F}$$

$$\frac{d_v}{d_t} = 39,394V/s$$

The time required to produce the 13 V peak-to-peak swing is

$$T = \frac{13V}{39,394V/s}$$

$$T = 330\mu s$$

This represents one half-cycle of the output wave. To go from +6.5 V to -6.5 V and back will require 660 ◈ s. Therefore, the output frequency is

$$f = \frac{1}{T}$$

$$f = \frac{1}{660\mu s}$$

$$f = 1.52kHz$$

The resulting frequency of Example 9.2.5 may be adjusted by changing either the 33 k Ω resistor or the 10 nF capacitor. Changing the comparator's resistors can alter the thresholds, and thus alter the frequency, but this is generally not recommended, as a change in output amplitude will occur as well. By combining steps, the above process may be reduced to a single equation:

$$f = \frac{1}{\frac{2V_{pp}}{V_{sat}}RC}$$

(9.2.26)

where ◈◈◈ is the difference between ◈◈◈◈◈◈◈*h*◈◈◈ and ◈◈◈◈◈◈◈*h*◈◈◈ . Note that if ◈3 is 4 times larger than ◈2 in the comparator, Equation 9.2.26 reduces to

$$f = \frac{1}{RC}$$

and the peak triangle wave amplitude is one-fourth of ◈◈◈◈ .

Generally, circuits such as this are used for lower frequency work. For clean square waves, very fast

op amps are required. Finally, for lower impedance loads, the outputs should be buffered with voltage followers.

COMPUTER SIMULATION

The Multisim simulation for the signal generator of Example 9.2.5 is shown in Figure 9.2.22 . The square and triangle outputs are plotted together so that the switching action can be seen. Note how each wave is derived form the other. The output plot is delayed 5 milliseconds in order to guarantee a plot of the steady state output. Failure to delay the plotting times will result in a graph of the initial turn-on transients. It may take many milliseconds before the waveforms finally stabilize, depending on the desired frequency of oscillation and the initial circuit conditions. Finally, note the sharp rising and falling edges of the square wave. This is due to the moderately fast slew rate of the LF411 op amp chosen. Had a slower device such as the 741 been used, the quality of the output waveforms would have suffered.

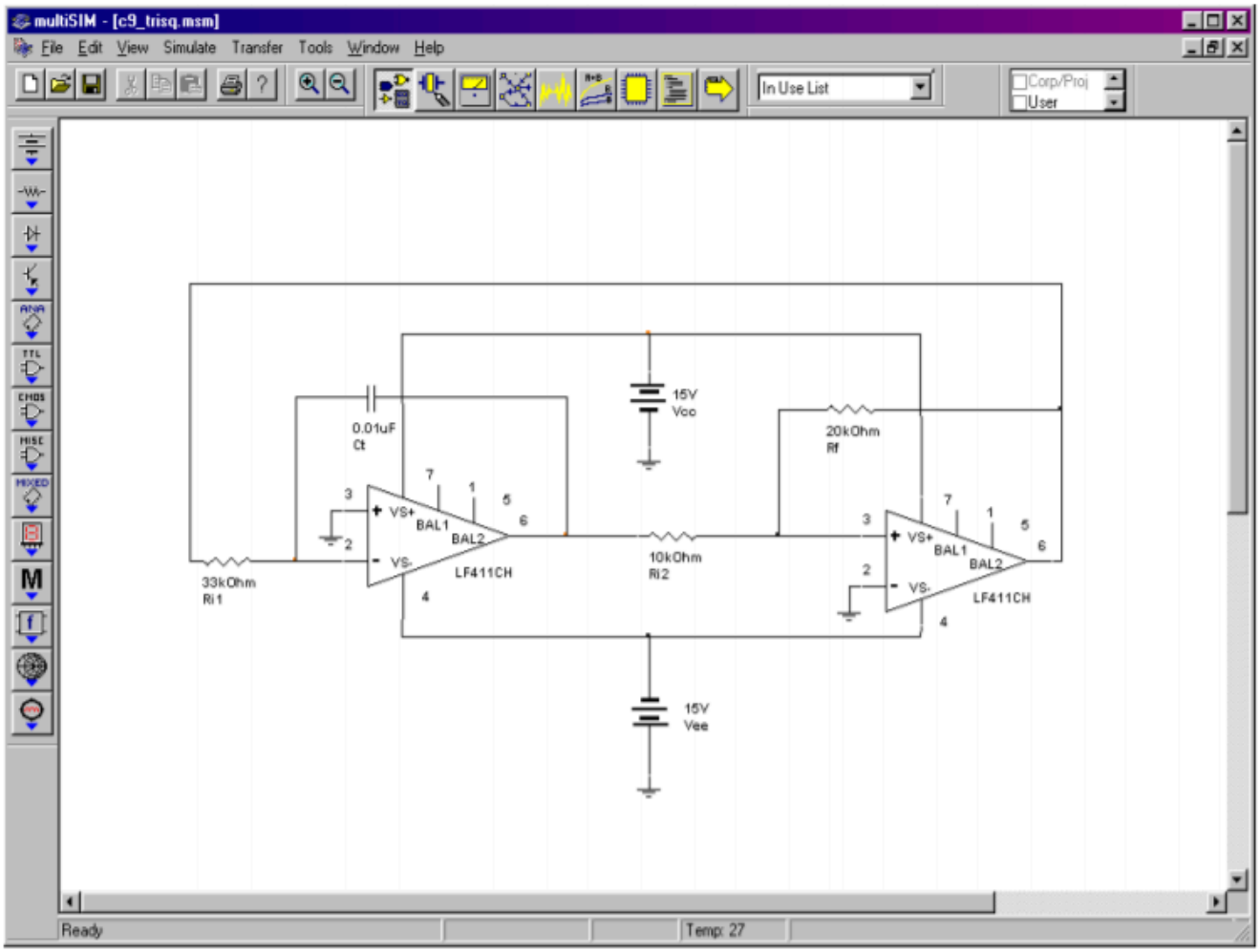


Figure 9.2.22 : Triangle/square generator in Multisim.

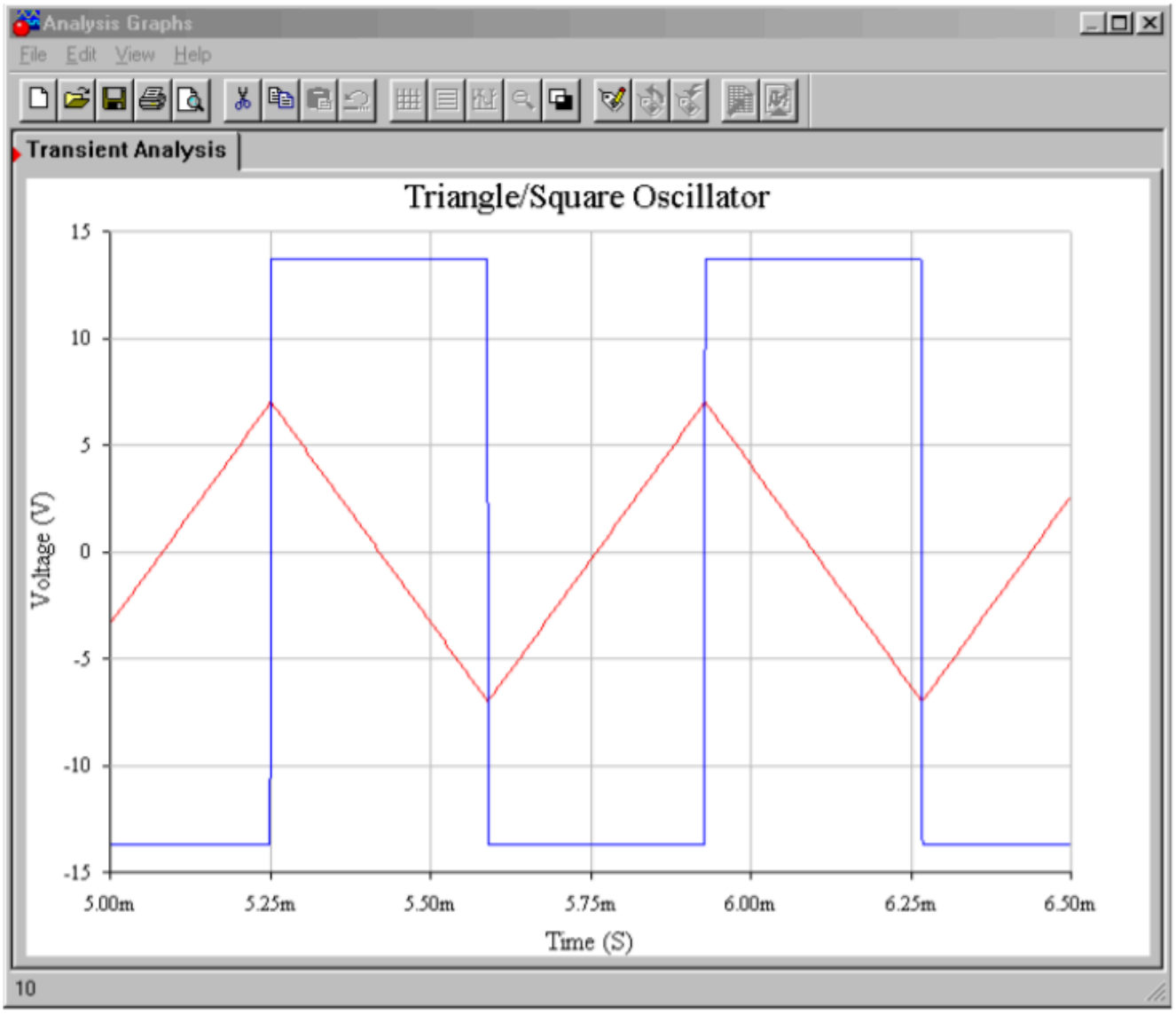


Figure 9.2.22 : Output waveforms from simulator.

If an accurate triangle wave is not needed, and only a square wave is required, the circuit of Figure 9.2.19 may be reduced to a single op amp stage. This is shown in Figure 9.2.23 . This circuit is, in essence, a comparator. Resistors 1 and 2 form the positive feedback portion and set the effective comparator trip point, or threshold. The measurement signal is the voltage across the capacitor. The potentials of interest are shown in Figure 9.2.24 . If the output is at positive saturation, the noninverting input will see a percentage of this, depending on the voltage divider produced by 1 and 2 . This potential is h . Because the output is at positive saturation, the capacitor , will be charging towards it. Because it is charging through resistor , the waveform is an exponential type. Once the capacitor voltage reaches h , the noninverting input will no longer be greater than the inverting input, and the device will change to the negative state. At this point, will reverse its course and move towards negative saturation. At the lower threshold, the op amp will again change state, and the process repeats. In order to determine the frequency of oscillation, we need to find how long it takes the capacitor to charge between the two

threshold points. Normally the circuit will be powered from equal magnitude supplies, and therefore +◻◻◻◻=–◻◻◻◻ and ◻◻◻◻◻◻◻h◻◻◻=◻◻◻◻◻◻◻h◻◻◻ . By inspection,

$$V_{thres} = V_{sat}\frac{R_1}{R_1 + R_2}$$

(9.2.27)

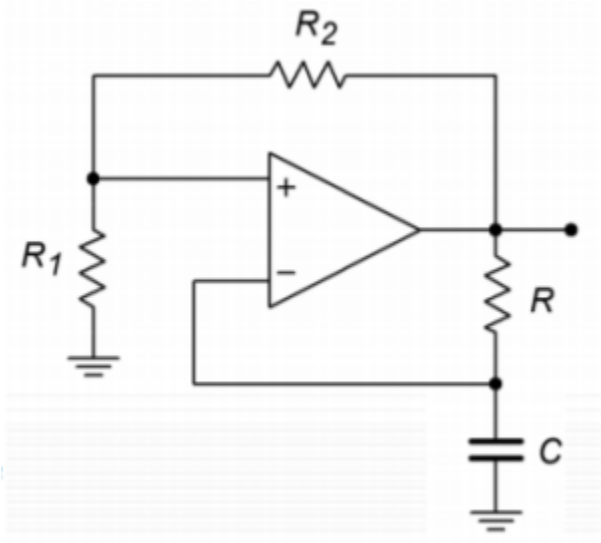


Figure 9.2.23 : Simple square wave generator.

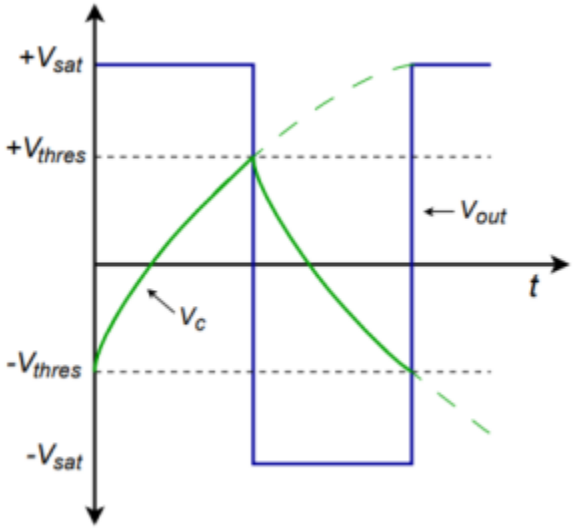


Figure 9.2.24 : Waveforms of a simple square wave generator.

The capacitor voltage is

$$V_C(t) = V_k(1 - \epsilon^{\frac{-t}{RC}})$$

(9.2.28)

where ◻◻ is the total potential applied to the capacitor. Because the capacitor will start at one threshold and attempt to charge to the opposite saturation limit, this is

$$V_k = V_{sat} + V_{thres}$$

(9.2.29)

Combining Equation 9.2.27 , Equation 9.2.28 , and Equation 9.2.29 yields

$$V_C(t) = (V_{sat} + V_{thres}) \left(1\epsilon^{\frac{-t}{RC}}\right)$$

(9.2.30)

At the point where the comparator changes state,

$$V_C = 2V_{thres}$$

(9.2.31)

Combining Equation 9.2.30 and Equation 9.2.31 produces

$$2V_{thres} = (V_{sat} + V_{thres}) \left(1\epsilon^{tRC}\right)$$

[1–\epsilon^{\frac{–t}{RC}} = \frac{2 V_{thres}}{V_{sat} + V_{thres}} \]

[1–\epsilon^{\frac{–t}{RC}} = \frac{2 V_{sat} \frac{R_1}{R_1+ R_2}}{V_{sat}\left(1+ \frac{R_1}{R_1+ R_2} \right)} \]

[1–\epsilon^{\frac{–t}{RC}} = \frac{2 R_1}{2 R_1+R_2} \]

[\epsilon^{\frac{–t}{RC}} = \frac{R_2}{2 R_1+R_2} \]

[\frac{–t}{RC} = \ln \left(\frac{R_2}{2 R_1+R_2} \right) \]

[t = RC \ln \left(\frac{2 R_1+ R_2}{R_2} \right) \]

This represents the charge time of the capacitor. One period requires two such traverses, so we may say

$$T = 2RC \ \ln\left(\frac{2R_1 + R_2}{R_2}\right) \ \ or,$$

$$f_o = \frac{1}{2RC \ \ln\left(\frac{2R_1+R_2}{R_2}\right)}$$

(9.2.32)

We can transform Equation 9.2.32 into “nicer” forms by choosing values for ◈1 and ◈2 such that the log term turns into a convenient number, such as 1 or 0.5. For example, if we set ◈1=0.859◈2 , the log term is unity, and consequently Equation 9.2.32 becomes ◈◈=1/2◈◈

Example 9.2.6

Design a 2 kHz square wave generator using the circuit of Figure 9.2.23 . For convenience, set ◈1=0.859◈2 . If ◈1 is arbitrarily set to 10 k Ω , then

$$R_1 = 0.859R_2$$

$$R_2 = \frac{R_1}{0.859}$$

$$R_2 = \frac{10k}{0.859}$$

$$R_2 = 11.64k$$

In order to set the oscillation frequency, ◈ is arbitrarily set to 10 k Ω , and ◈ is then determined.

$$f_o = \frac{1}{2RC}$$

$$C = \frac{1}{2Rf_o}$$

$$C = \frac{1}{2 \times 10k \times 2kHz}$$

$$C = 25nF$$

COMPUTER SIMULATION

A simulation of the square wave generator of Example 9.2.6 is shown in Figure 9.2.25 . In order to graphically illustrate the importance of the op amp having sufficient bandwidth and slew rate, the simulation is run twice, once using the moderately fast LF411 and a second time using the much slower 741.

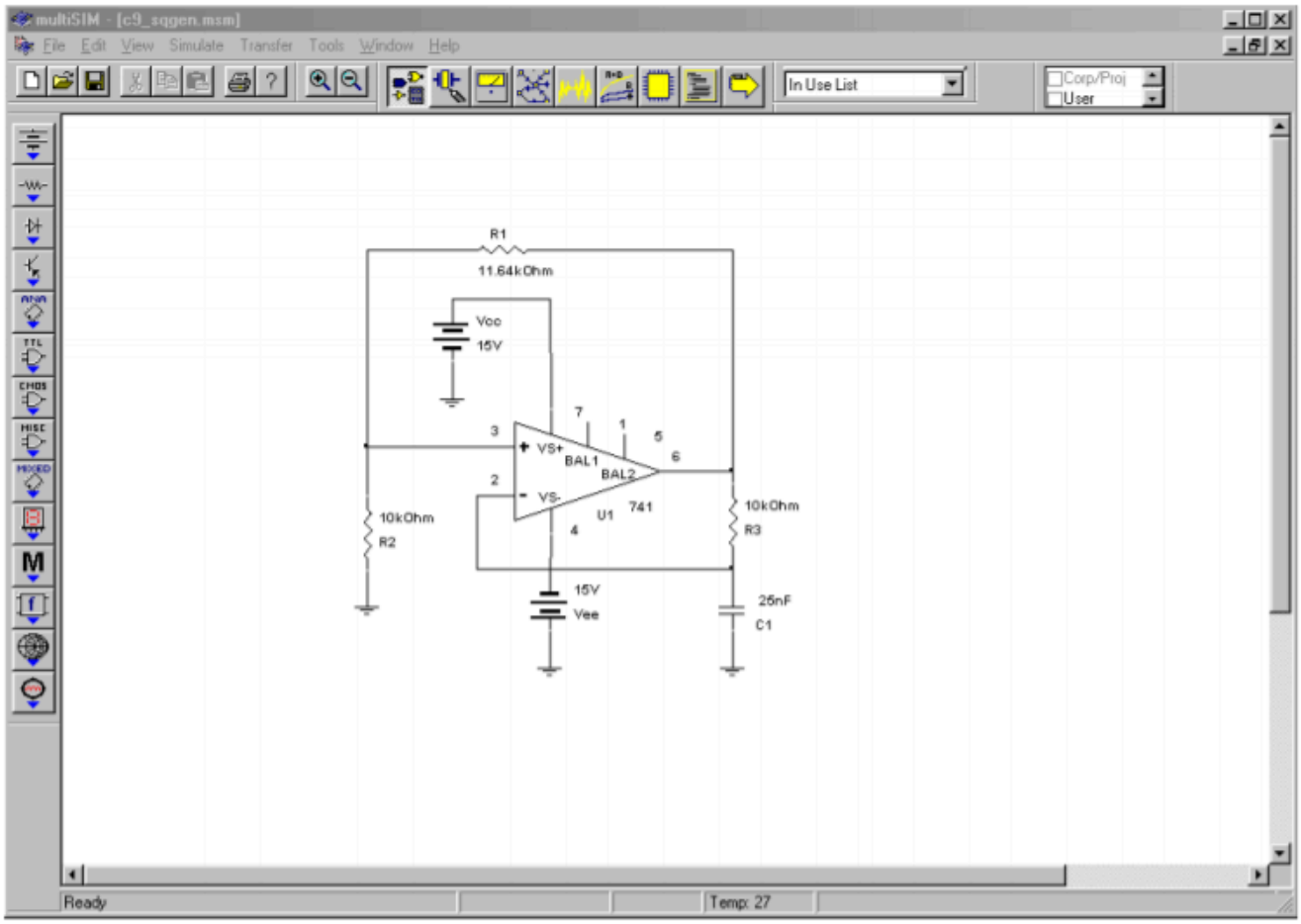


Figure 9.2.25 : Square wave generator in Multisim.

Both the output and capacitor voltages are plotted from the Transient Analysis. Using the LF411, the output waveform is very crisp with sharp rising and falling edges. The capacitor voltage appears exactly as it should. The resulting frequency is just a little lower than the 2 kHz target. In contrast, the 741 plots show some problems. First, the square wave has noticeable slew rate limiting on the transitions. Second, due to the slewing problems, the capacitor voltage waveshape appears distorted (note the excessive rounding of the peaks). These effects combine to produce a frequency about 15 percent lower than the target, or about 1.7 kHz. The end result is a lackluster output waveform.

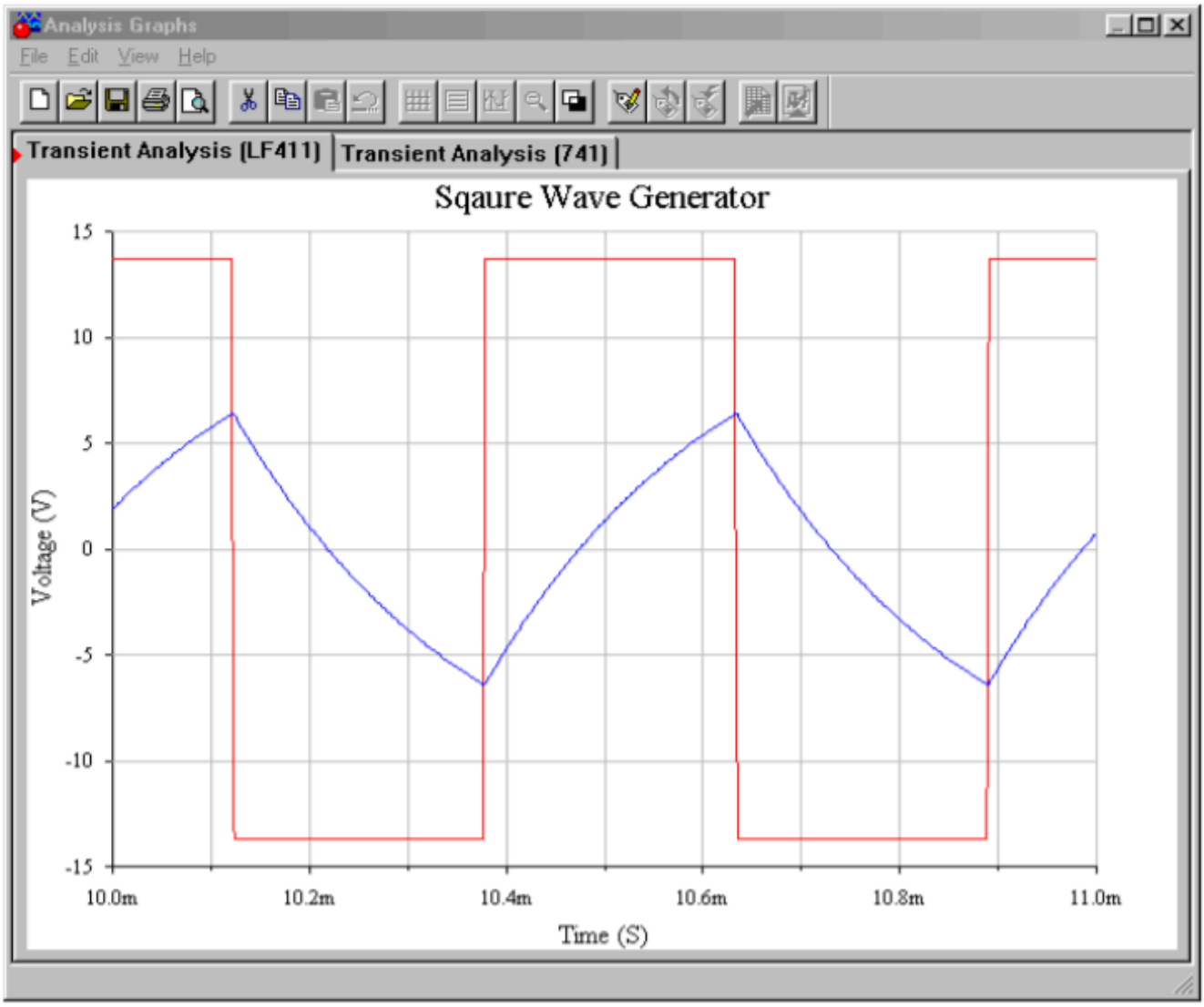


Figure 9.2.25 : Waveforms using LF411.

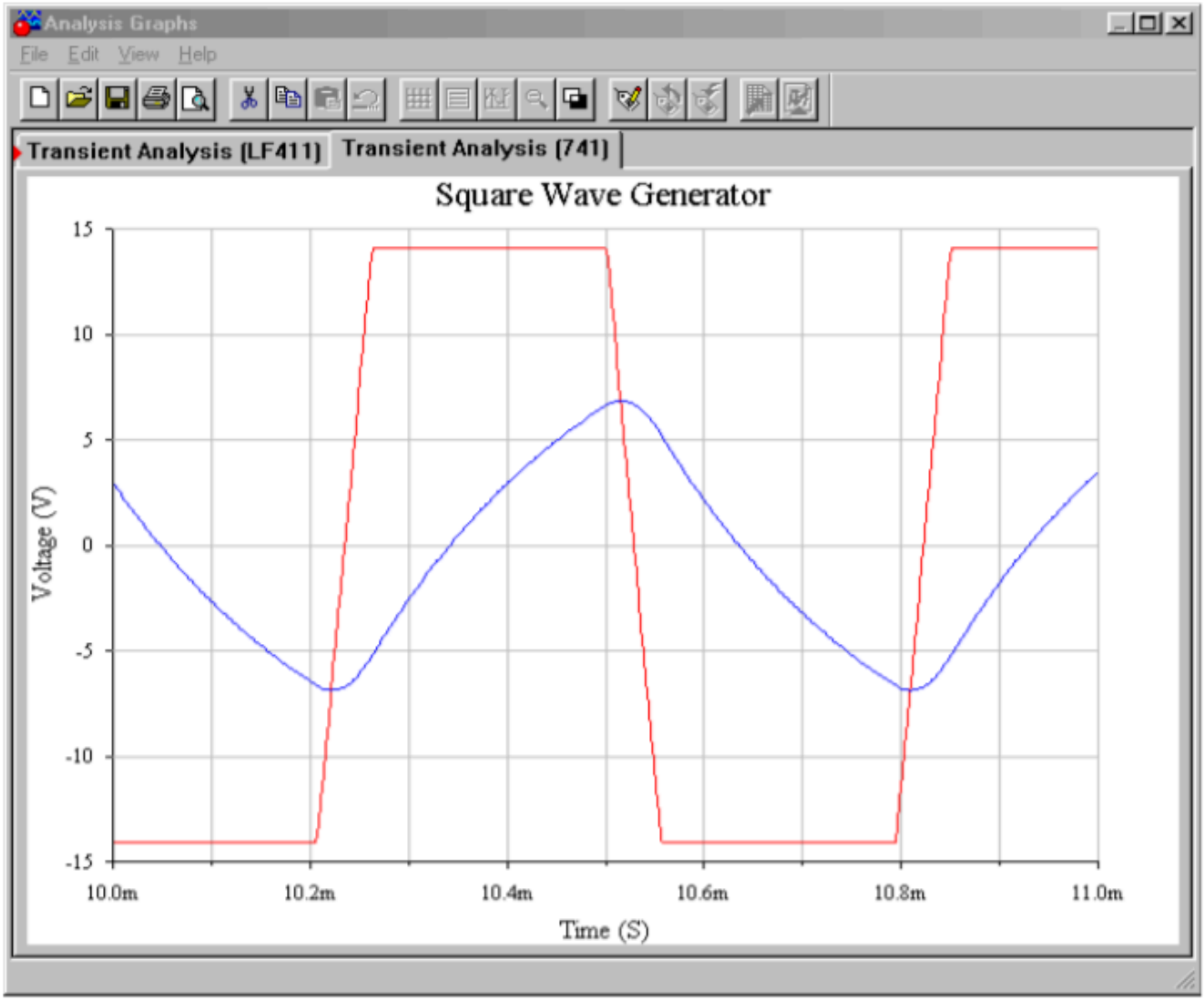


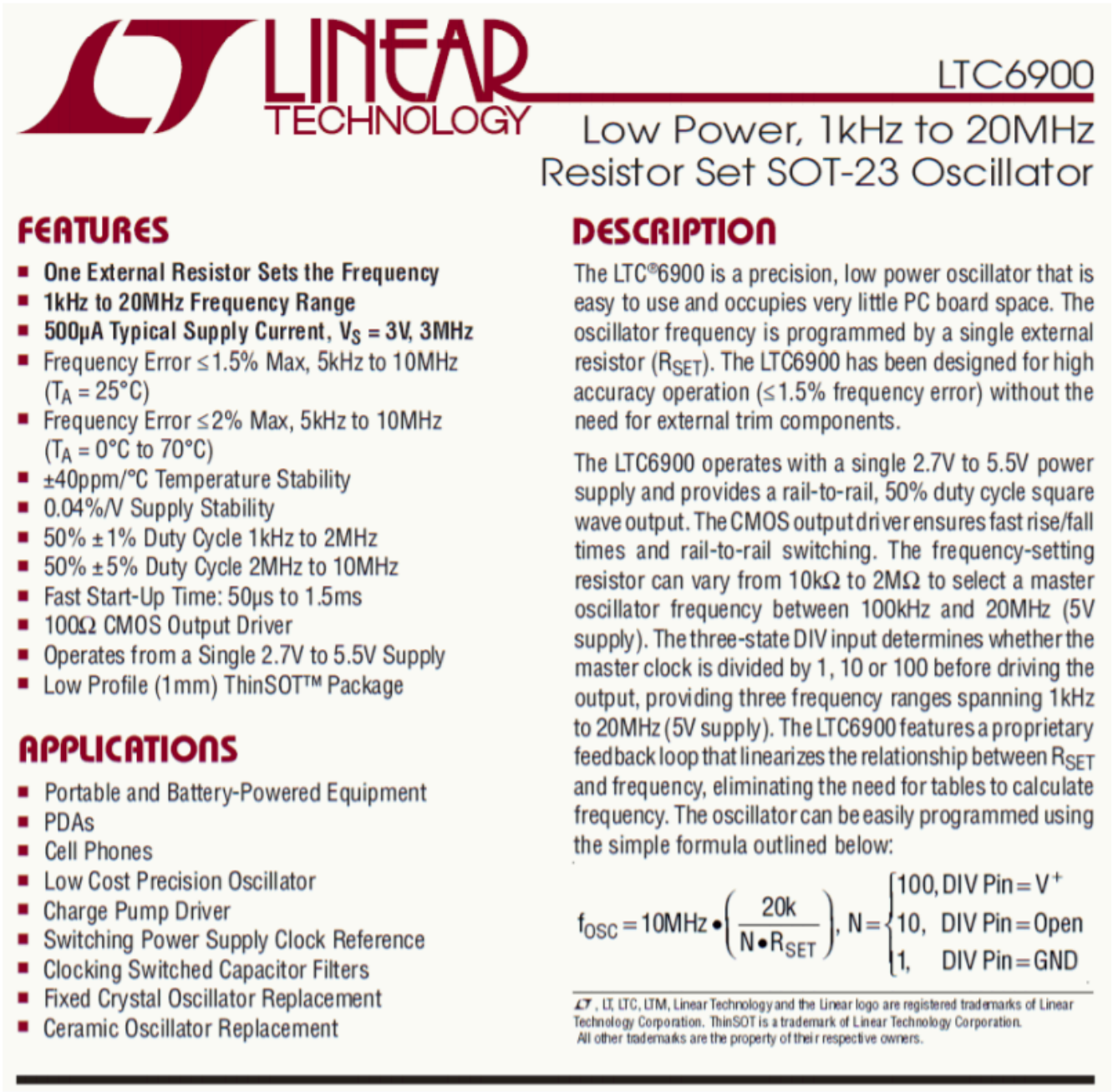
Figure 9.2.25 : Waveforms using 741.

13.3 SINGLE CHIP OSCILLATORS AND FREQUENCY GENERATORS

The generation of signals is a basic requirement for a wide variety of applications, thus a number of manufacturers produce a selection of single IC oscillators and frequency generators. Some of these tend to work in the range below 1 MHz and usually require some form of external resistor/capacitor network to set the operating frequency. Other, highly specialized circuits for targeted applications are also available. In this section we shall examine a few of the ICs that are generically referred to as clock generators, voltage-controlled oscillators, phase-locked loops and timers.

SQUARE WAVE/CLOCK GENERATOR

The need for stable, low cost, easy-to-use integrated circuits to generate square waves for clocking needs is widespread. Several companies manufacture such devices. One example is the LTC6900 from Linear Technology. A description sheet with a basic programming formula is shown in Figure 9.3.1 .



LINEAR TECHNOLOGY

LTC6900

Low Power, 1kHz to 20MHz Resistor Set SOT-23 Oscillator

FEATURES

- **One External Resistor Sets the Frequency**
- **1kHz to 20MHz Frequency Range**
- **500µA Typical Supply Current, V_S = 3V, 3MHz**
- Frequency Error ≤1.5% Max, 5kHz to 10MHz (T_A = 25°C)
- Frequency Error ≤2% Max, 5kHz to 10MHz (T_A = 0°C to 70°C)
- ±40ppm/°C Temperature Stability
- 0.04%/V Supply Stability
- 50% ±1% Duty Cycle 1kHz to 2MHz
- 50% ±5% Duty Cycle 2MHz to 10MHz
- Fast Start-Up Time: 50µs to 1.5ms
- 100Ω CMOS Output Driver
- Operates from a Single 2.7V to 5.5V Supply
- Low Profile (1mm) ThinSOT™ Package

APPLICATIONS

- Portable and Battery-Powered Equipment
- PDAs
- Cell Phones
- Low Cost Precision Oscillator
- Charge Pump Driver
- Switching Power Supply Clock Reference
- Clocking Switched Capacitor Filters
- Fixed Crystal Oscillator Replacement
- Ceramic Oscillator Replacement

DESCRIPTION

The LTC®6900 is a precision, low power oscillator that is easy to use and occupies very little PC board space. The oscillator frequency is programmed by a single external resistor (R_{SET}). The LTC6900 has been designed for high accuracy operation (≤1.5% frequency error) without the need for external trim components.

The LTC6900 operates with a single 2.7V to 5.5V power supply and provides a rail-to-rail, 50% duty cycle square wave output. The CMOS output driver ensures fast rise/fall times and rail-to-rail switching. The frequency-setting resistor can vary from 10kΩ to 2MΩ to select a master oscillator frequency between 100kHz and 20MHz (5V supply). The three-state DIV input determines whether the master clock is divided by 1, 10 or 100 before driving the output, providing three frequency ranges spanning 1kHz to 20MHz (5V supply). The LTC6900 features a proprietary feedback loop that linearizes the relationship between R_{SET} and frequency, eliminating the need for tables to calculate frequency. The oscillator can be easily programmed using the simple formula outlined below:

$$f_{OSC} = 10MHz \bullet \left(\frac{20k}{N \bullet R_{SET}}\right),\ N = \begin{cases} 100, & \text{DIV Pin} = V^+ \\ 10, & \text{DIV Pin} = \text{Open} \\ 1, & \text{DIV Pin} = \text{GND} \end{cases}$$

LT, LTC, LTM, Linear Technology and the Linear logo are registered trademarks of Linear Technology Corporation. ThinSOT is a trademark of Linear Technology Corporation. All other trademarks are the property of their respective owners.

Figure 9.3.1 : LTC6900 description. Reprinted courtesy of Linear Technology

The LTC6900 is a 5 volt low power circuit available in an SOT-23 (5 pin) package. It operates from 1 kHz to 20 MHz. The output frequency is programmable via a single resistor and the connection to its divider pin (labeled DIV). The frequency of the master oscillator is given by the equation

$$f_o = 10MHz\frac{20k}{R_{set}} \tag{9.3.1}$$

◇◇◇◇ is connected from the power supply pin to the SET pin. Acceptable values range between 10 k Ω and 2 M Ω . If the DIV pin is grounded, the output frequency will be as calculated. If the DIV pin is left unconnected, the output frequency will be divided by 10 and if the DIV pin is connected to

+5 volts, the output frequency will be reduced by a factor of 100. This is summarized graphically in Figure 9.3.2 .

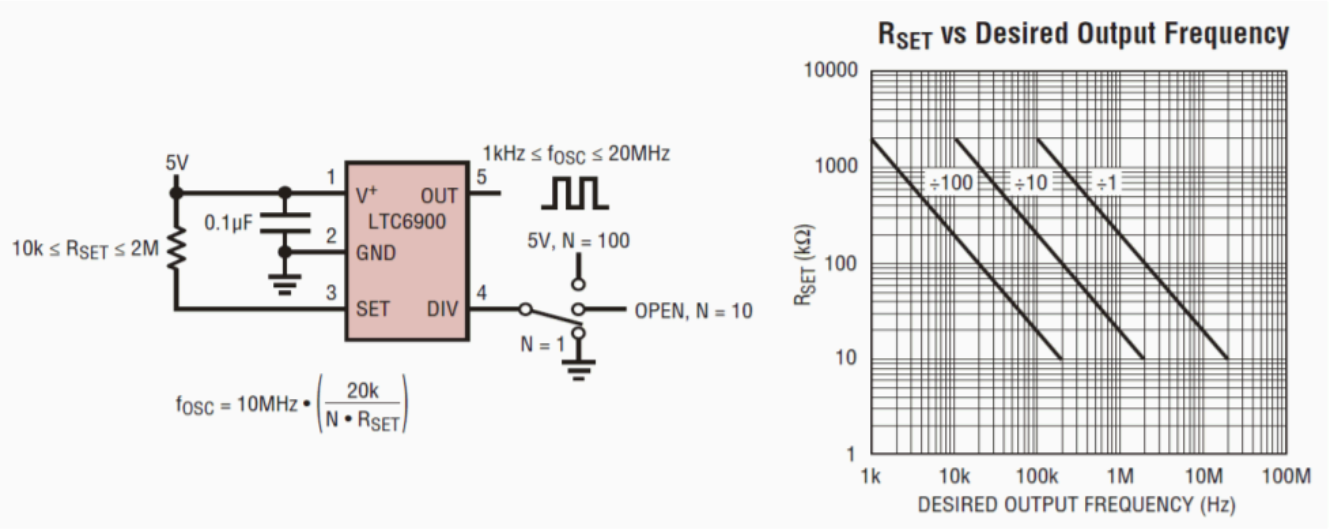


Figure 9.3.2 : LTC6900 oscillator. operation Reprinted courtesy of Linear Technology

Example 9.3.1

Using the LTC6900, design a 10 kHz square wave oscillator.

10 kHz is well within the range of this IC. To achieve this comfortably, we will need a divide-by-100 setting based on the graph of Figure 9.3.2 . This will require us to tie the DIV pin to +5 volts. The value of ⍰⍰⍰⍰ can be approximated from the graph or computed directly.

$$f_{osc} = 10MHz\frac{20k}{N \times R_{set}}$$

$$R_{set} = 10MHz\frac{20k}{N \times f_{osc}}$$

$$R_{set} = 10MHz\frac{20k}{100 \times 100kHz}$$

$$R_{set} = 200k$$

VOLTAGE-CONTROLLED OSCILLATOR

A voltage-controlled oscillator (usually abbreviated as VCO) does not produce a fixed output frequency. As its name suggests, the output frequency of a VCO is dependent on a control voltage. There is a fixed relationship between the control voltage and the output frequency. Theoretically, just about any oscillator can be turned into a VCO. For example, if a resistor is used as part of the tuning circuit, it could be replaced with some form of voltage-controlled resistor, such as an FET or a

light-dependent resistor/lamp combination. By doing this, an external potential can be used to set the frequency of oscillation. This is very useful if the frequency needs to be changed quickly or accurately swept through some range.

A classic example of the usefulness of a VCO is shown in Figure 9.3.3 . This is a simplified schematic of an analog monophonic musical keyboard synthesizer. The keys on the synthesizer are little more than switches. These switches tap potentials off a voltage divider. As the musician plays up the keyboard, the switches engage higher and higher potentials. These levels are used to control a very accurate VCO. The higher the control voltage, the higher the output frequency or pitch will be.

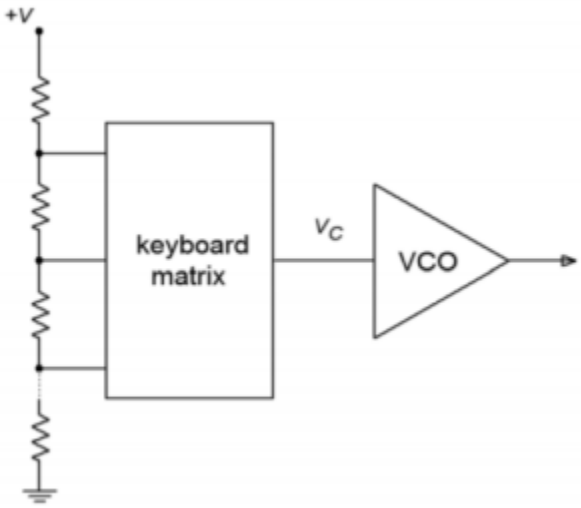


Figure 9.3.3 : Simplified music synthesizer using VCO.

VCOs can be used for a number of other applications, including swept frequency spectrum analyzers, frequency modulation and demodulation, and control systems. It is also an integral part of the phase-locked loop, as we will see later in this chapter.

An example of a VCO is the LTC6990. It is part of Linear Technology's TimerBlox series of timer/counter/clock ICs. The series includes clock sources that operate in excess of 100 MHz and timers that switch at multi-hour rates. The LTC6990 operates in the range of just under 500 Hz to 2 MHz. While it can be used for fixed frequency applications, it also makes for a flexible VCO. An overview is shown in Figure 9.3.4 .

LINEAR TECHNOLOGY

LTC6990

TimerBlox: Voltage Controlled Silicon Oscillator

FEATURES

- **Fixed-Frequency or Voltage-Controlled Operation**
  - **Fixed: Single Resistor Programs Frequency with <1.5% Max Error**
  - **VCO: Two Resistors Set VCO Center Frequency and Tuning Range**
- **Frequency Range: 488Hz to 2MHz**
- 2.25V to 5.5V Single Supply Operation
- 72µA Supply Current at 100kHz
- 500µs Start-Up Time
- VCO Bandwidth >300kHz at 1MHz
- CMOS Logic Output Sources/Sinks 20mA
- 50% Duty Cycle Square Wave Output
- Output Enable (Selectable Low or Hi-Z When Disabled)
- –55°C to 125°C Operating Temperature Range
- Available in Low Profile (1mm) SOT-23 (ThinSOT™) and 2mm × 3mm DFN Package

APPLICATIONS

- Low Cost Precision Programmable Oscillator
- Voltage-Controlled Oscillator
- High Vibration, High Acceleration Environments
- Replacement for Fixed Crystal and Ceramic Oscillators
- Portable and Battery-Powered Equipment

DESCRIPTION

The LTC®6990 is a precision silicon oscillator with a programmable frequency range of 488Hz to 2MHz. It can be used as a fixed-frequency or voltage-controlled oscillator (VCO). The LTC6990 is part of the TimerBlox® family of versatile silicon timing devices.

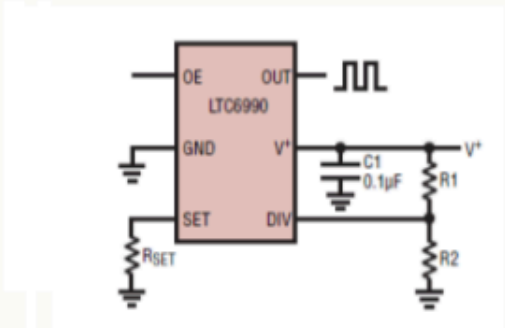
A single resistor, R_{SET}, programs the LTC6990's internal master oscillator frequency. The output frequency is determined by this master oscillator and an internal frequency divider, N_{DIV}, programmable to eight settings from 1 to 128.

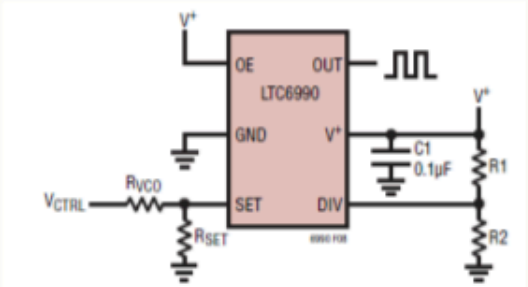
$$f_{OUT} = \frac{1MHz}{N_{DIV}} \bullet \frac{50k\Omega}{R_{SET}}, N_{DIV} = 1, 2, 4 \ldots 128$$

Optionally, a second resistor at the SET input provides linear voltage control of the output frequency and can be used for frequency modulation. A narrow or wide VCO tuning range can be configured by the appropriate selection of the two resistors.

The LTC6990 includes an enable function that is synchronized with the master oscillator to ensure clean, glitch-free output pulses. The disabled output can be configured to be high impedance or forced low.

For easy configuration of the LTC6990, download the TimerBlox Designer tool at www.linear.com/timerblox.





$$f_{OUT} = \frac{1MHz \bullet 50k}{N_{DIV} \bullet R_{VCO}} \bullet \left(1 + \frac{R_{VCO}}{R_{SET}} - \frac{V_{CTRL}}{V_{SET}}\right)$$

Figure 9.3.4 : LTC6990 VCO. Reprinted courtesy of Linear Technology

Like the LTC6900, the LTC6990 is programmed with as little as one resistor and has a frequency divider option. Unlike its brother, the divider capabilities are much more broad, spanning eight

power-of-two settings versus just three decade settings. A basic fixed frequency oscillator is shown at the bottom left of Figure 9.3.4 where the master oscillation frequency is controlled by ◈◈◈◈ . Standard VCO operation is shown at the bottom right. The LTC6990 also has the option of a high impedance output state, making a total of 16 divider/output possibilities. This set up is programmed typically through the use of two external resistors. The programming table is reproduced in Figure 9.3.5 .

DIVCODE Programming

DIVCODE	Hi-Z	N_{DIV}	Recommended f_{OUT}	R1 (k)	R2 (k)	V_{DIV}/V^+
0	0	1	62.5kHz to 1MHz	Open	Short	≤ 0.03125 ±0.015
1	0	2	31.25kHz to 500kHz	976	102	0.09375 ±0.015
2	0	4	15.63kHz to 250kHz	976	182	0.15625 ±0.015
3	0	8	7.813kHz to 125kHz	1000	280	0.21875 ±0.015
4	0	16	3.906kHz to 62.5kHz	1000	392	0.28125 ±0.015
5	0	32	1.953kHz to 31.25kHz	1000	523	0.34375 ±0.015
6	0	64	976.6Hz to 15.63kHz	1000	681	0.40625 ±0.015
7	0	128	488.3Hz to 7.813kHz	1000	887	0.46875 ±0.015
8	1	128	488.3Hz to 7.813kHz	887	1000	0.53125 ±0.015
9	1	64	976.6Hz to 15.63kHz	681	1000	0.59375 ±0.015
10	1	32	1.953kHz to 31.25kHz	523	1000	0.65625 ±0.015
11	1	16	3.906kHz to 62.5kHz	392	1000	0.71875 ±0.015
12	1	8	7.813kHz to 125kHz	280	1000	0.78125 ±0.015
13	1	4	15.63kHz to 250kHz	182	976	0.84375 ±0.015
14	1	2	31.25kHz to 500kHz	102	976	0.90625 ±0.015
15	1	1	62.5kHz to 1MHz	Short	Open	≥ 0.96875 ±0.015

Figure 9.3.5 : LTC6990 oscillator. programming Reprinted courtesy of Linear Technology

The output state depends on the combination of the Output Enable pin (OE) and the Hi-Z logic. When OE is high, the output will be active. If OE is low and Hi-Z is low, then the output will be held low. Finally, if OE is low and Hi-Z is high, then the output will go to a high impedance state.

The voltage present at the DIVCODE pin sets the frequency divider and the impedance mode. This voltage is interpreted by an internal 4 bit analog-to-digital converter (AD converters are the topic of Chapter Twelve). While it is possible to feed this pin with some external source, the more practical method is to simply create a voltage divider with a pair of 1% tolerance resistors; one tied from the power supply to the DIVCODE pin, and the second connected from DIVCODE to ground.

The master oscillator of the LTC6990 is controlled by the current at the SET pin. Internally, the voltage at this pin is maintained at 1 volt, therefore the frequency can be set by a single resistor, ◈◈◈◈ , connected from this pin to ground. It can then be divided down to lower frequency. This is essentially the same situation we found with the LTC6900. A 20 ◈ A current (i.e., 50 k Ω) will produce the top rate of 1 MHz. Lower currents (higher resistances) will produce proportionately lower frequencies.

The DIVCODE value will divide this base frequency down further by powers of two. We can express this relation with the following formula,

$$f_{osc} = 1MHz\frac{50k}{N_{DIV}R_{set}}$$

(9.3.2)

where ◈◈◈◈ is found from the table in Figure 9.3.5

The design process starts by identifying the appropriate frequency range. It is best if the desired oscillation frequency is not located at the extremes of any given range. Once a range is determined, the corresponding value for ◈◈◈◈ is found, and along with it, the required divider resistor values, ◈1 and ◈2 . From there it is a simple matter to solve Equation 9.3.2 in terms of ◈◈◈◈ .

Example 9.3.2

An LTC6990 is connected as shown in Figure 9.3.6 . It is being used as a light-to-frequency converter. That is, the output frequency will be controlled by the amount of light hitting a sensor. The sensor is a CdS (Cadmium Sulfide) cell that is connected in the position of ◈◈◈◈ . Under low light conditions the cell will produce a high resistance and as the light level increases the resistance drops. Assuming that the cell varies from 500 k down to 60 k, determine the range of output frequencies. First, determine the DIVCODE value. This can be found by computing the voltage divider ratio of ◈1 and ◈2 , but in this circuit recommended values have been used from the DIVCODE table. By observation, ◈◈◈◈=16 .

Next, we calculate the limit frequencies.

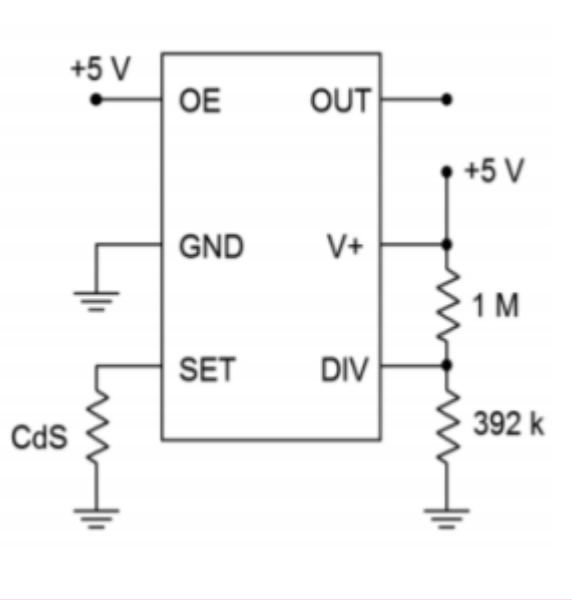


Figure 9.3.6 : Light-to-frequency converter circuit for Example 9.3.2 .

$$f_{osc} = 1MHz\frac{50k}{N_{DIV}R_{set}}$$

$$f_{osc} = 1MHz\frac{50k}{16 \times 500k}$$

$$f_{osc} = 6.25kHz$$

$$f_{osc} = 1MHz\frac{50k}{N_{DIV}R_{set}}$$

$$f_{osc} = 1MHz\frac{50k}{16x60k}$$

$$f_{osc} = 52.08kHz$$

Note that as light levels increase, frequency increases in proportion.

When used in VCO mode, the most important element to remember is that the master oscillator frequency is set by the current coming out of the SET pin, ◈◈◈◈ , as expressed by the following formula

$$f_o = 1MHz \times 50k\frac{I_{set}}{V_{set}}$$

◈◈◈◈ is kept to 1 volt internally so this reduces to

$$f_o = 1MHz \times 50k \times I_{set}$$

(9.3.3)

◈◈◈◈ is then divided down by ◈◈◈◈ . The final oscillation frequency may be expressed as

$$f_{osc} = 1MHz \times 50k\frac{I_{set}}{N_{DIV}}$$

(9.3.4)

Note that ◈◈◈◈ is exiting the chip. Further, note the frequency of oscillation and ◈◈◈◈ are directly proportional. Also, keep in mind that the ◈◈◈◈ variation, and hence, the frequency variation, should be kept to 16:1 for best performance, where the maximum value of ◈◈◈◈ is 20 ◈ A. A simple method for obtaining voltage control is shown in Figure 9.3.4 . There are some potential issues here. First, the range of voltage from the control circuit may not be able to achieve the desired frequency with that circuit. Second, note that higher control voltages will produce lower output frequencies, that is, an inverse relation. This can be an issue in some applications. Consequently, we shall examine a more generic method of controlling the circuit through the use of an external op amp for scaling and offsetting.

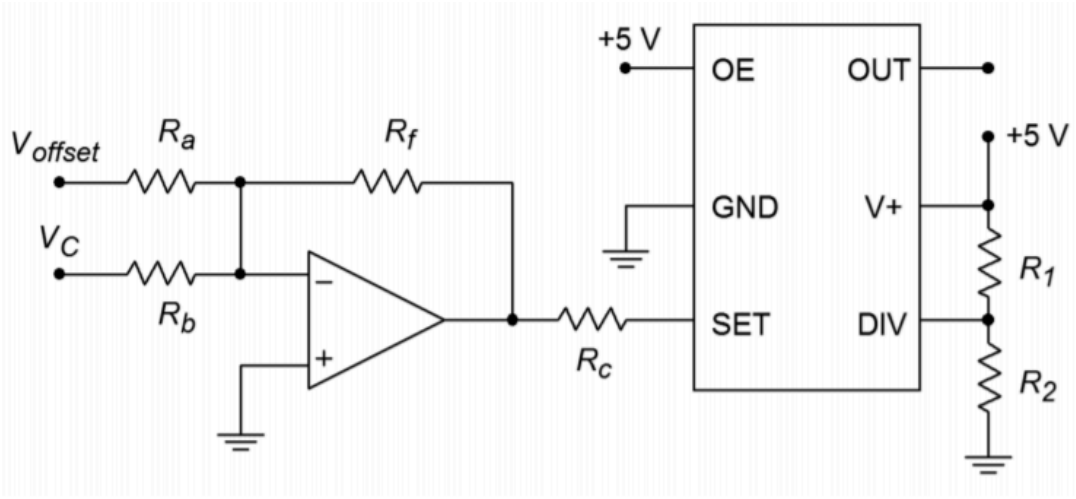


Figure 9.3.7 : A method of mapping the control voltage.

The circuit of Figure 9.3.7 presents a method of mapping the existing control voltage onto the LTC6990 or any similar VCO. In this circuit a simple summing amplifier is used for scaling and offsetting. The control voltage, ◈◈ , is scaled by one channel of the weighted summer. This signal is offset by a DC voltage, ◈◈◈◈◈◈◈ , fed through another channel. The voltage at the output of the op amp is used to sink current from the VCO via the control resistor, ◈◈ . Recall that the SET pin of the IC produces 1 volt internally and it is the exiting current, ◈◈◈◈ , that sets the master oscillator frequency, as expressed in Equation 9.3.3 . Obviously, the op amp output voltage must be less than 1 volt in order for the op amp to sink current (i.e., in order for ◈◈◈◈ to be exiting the LTC6990). The voltage difference between the op amp's output and the 1 volt at the SET pin drops across ◈◈ and this is what creates ◈◈◈◈ . Note that as the control voltage grows more positive at the op amp's input, its output, and hence ◈◈◈◈ , also increases. Thus, frequency increases as control voltage increases.

Example 9.3.3

Using Figure 9.3.7 as a guide, design a VCO circuit that will produce output frequencies from 20 kHz through 50 kHz when driven by control voltages from 6 to 8 volts (i.e., 6 volts will produce 20 kHz, 7 volts will produce 35 kHz, 8 volts will produce 50 kHz, etc.)

First, note that the frequency range is 2.5:1. As the LTC6990 can always cover any 8:1 span (as high as 16:1) and the maximum frequency of 50 kHz is well below the LTC6990's maximum, we know this IC is a good candidate for the design. We now need to determine the divider resistor values. Consulting Figure 9.3.5 shows that we can achieve this range using an ◈◈◈◈ of 4, 8 or 16. Choosing the middle value, and assuming we don't care about Hi-Z state, we arrive at DIVCODE=3 with ◈1 =1 M Ω and ◈2 =280 k Ω .

Our frequency range is 2.5:1 which means that our ◈◈◈◈ range must also be 2.5:1. For convenience, choose the op amp's output to be 0 volts for the minimum frequency. This will yield 1 volt across ◈◈ and occurs when the control voltage into the op amp is at its 6 volt minimum. When the control voltage is at its maximum of 8 volts we'll need 2.5 volts across ◈◈ (i.e., 2.5 times the prior ◈◈◈◈). This means that the op amp's output must go to -1.5 volts. Note that a 2 volt change in the input control voltage will produce a 1.5 volt change at the op amp's output. Thus, the gain of this channel is -0.75. If we choose ◈◈ =100 k Ω then ◈◈ =75 k Ω .

At this point we need to add an offset. With only the gain scaling, the 6 volt ⍰⍰ produces -0.75 times 6, or -4.5 volts, and a ⍰⍰ of 8 volts similarly produces -6 volts. Consequently, we need to add a +4.5 volt offset to the output. If we tie ⍰⍰⍰⍰⍰⍰⍰ to the op amp's -15 volt power rail then we will need a gain of 4.5/(-15), or -0.3. With an ⍰⍰ of 100 k Ω , ⍰⍰ must be 333.3 k Ω (the nearest 1% standard value is 332 k Ω).

Finally, to determine ⍰⍰ , refer to Equation 9.3.4 and solve for ⍰⍰⍰⍰

$$I_{set} = \frac{f_{osc} \times N_{DIV}}{1MHz \times 50k}$$

Using the minimum fosc of 20 kHz yields

$$I_{set} = \frac{20kHz \times 8}{1MHz \times 50k}$$

$$I_{set} = 3.2\mu amps$$

This occurs with 1 volt across ⍰⍰ . Therefore ⍰⍰ = 312.5 k Ω . Crosschecking, when ⍰⍰ = 8 V, we see 2.5 volts across ⍰⍰ for a current of 8 ⍰ A. Inserting this into Equation 9.3.4 yields 50 kHz, our desired maximum frequency.

In closing, note that the way in which the frequency sweeps depends on the wave shape of ⍰⍰ . If a sinusoid is used, the output frequency will vary smoothly between the stated limits. On the other hand, if the wave shape for ⍰⍰ is a ramp, the output frequency will start at one extreme and then move smoothly to the other limit as the ⍰⍰ ramp continues. When the ramp resets itself, the output frequency will jump back to its starting point. An example of this is shown in Figure 9.3.8 . Finally, if the control wave shape is a square, the output frequency will abruptly jump from the minimum to the maximum frequency and back. This effect is shown in Figure 9.3.9 , and can be used to generate FSK (frequency shift key) signals. FSK is used in the communications industry to transmit binary information.

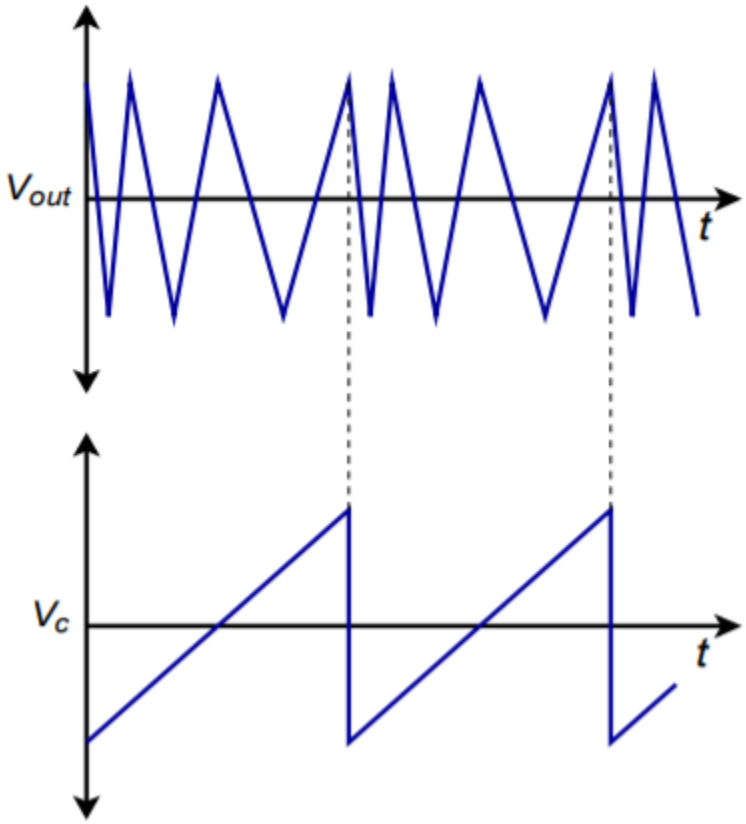


Figure 9.3.8 : VCO frequency sweep using a ramp.

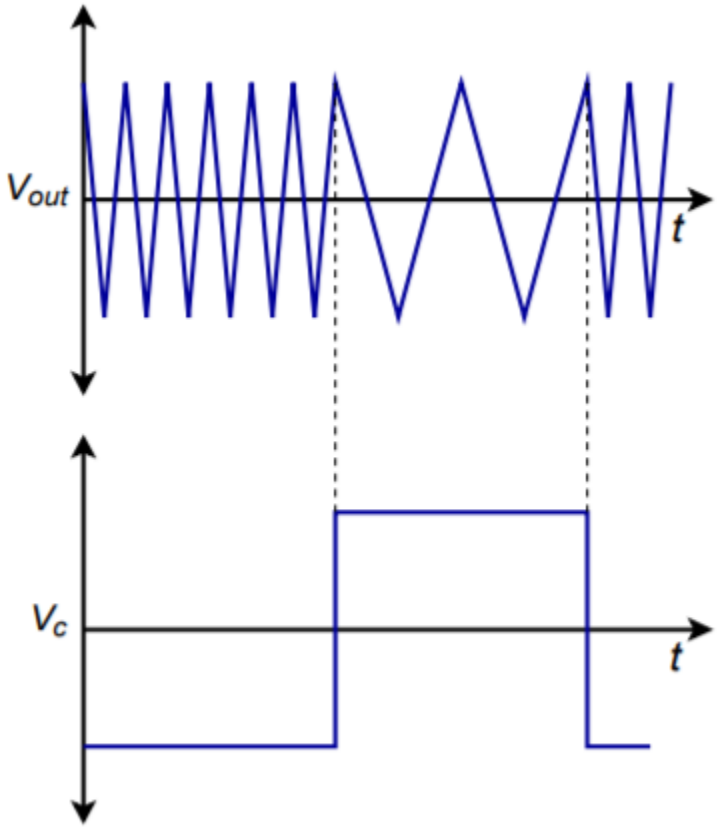


Figure 9.3.9 : VCO two tone output using a square wave.

PHASE-LOCKED LOOP

One step up from the VCO is the Phase-Locked Loop, or PLL. The PLL is a selfcorrecting circuit; it can lock onto an input frequency and adjust to track changes in the input. PLLs are used in modems, for FSK systems, frequency synthesis, tone decoders, FM signal demodulation, and other applications. A block diagram of a basic PLL is shown in Figure 9.3.10 .

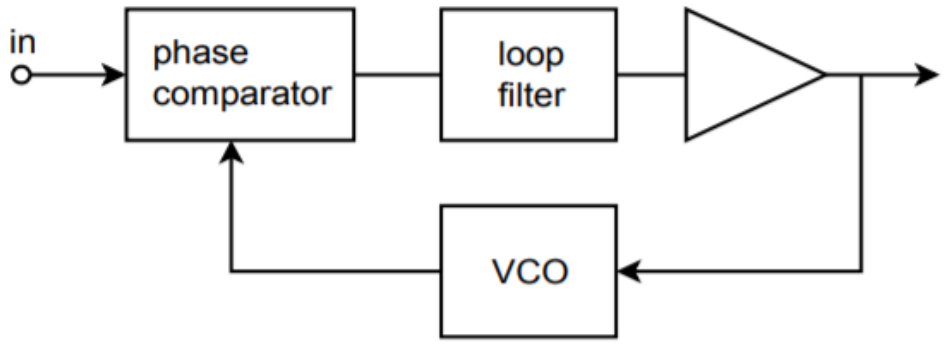


Figure 9.3.10 : Phase-locked loop.

In essence, the PLL uses feedback in order to lock an oscillator to the phase and frequency of an incoming signal. It consists of three major parts; a phase comparator, a loop filter (typically, a lag network of some form), and a VCO. An amplifier may also exist within the loop. The phase

comparator is driven by the input signal and the output of the VCO. It produces an error signal that is proportional to the phase difference between its inputs. This error signal is then filtered in order to remove spurious high-frequency signals and noise. The resulting error signal is used as the control voltage for the VCO, and as such, sets the VCO's output frequency. As long as the error signal is not too great, the loop will be selfstabilizing. In other words, the error signal will eventually drive the VCO to be in perfect frequency and phase synchronization with the input signal. When this happens, the PLL is said to be in lock with the input. The range of frequencies over which the PLL can stay in lock as the input signal changes is called the lock range. Normally, the lock range is symmetrical about the VCO's free-running, or center, frequency. The deviation from the center frequency out to the edge of the lock range is called the tracking range, and is therefore one-half of the lock range. This is illustrated in Figure 9.3.11 .

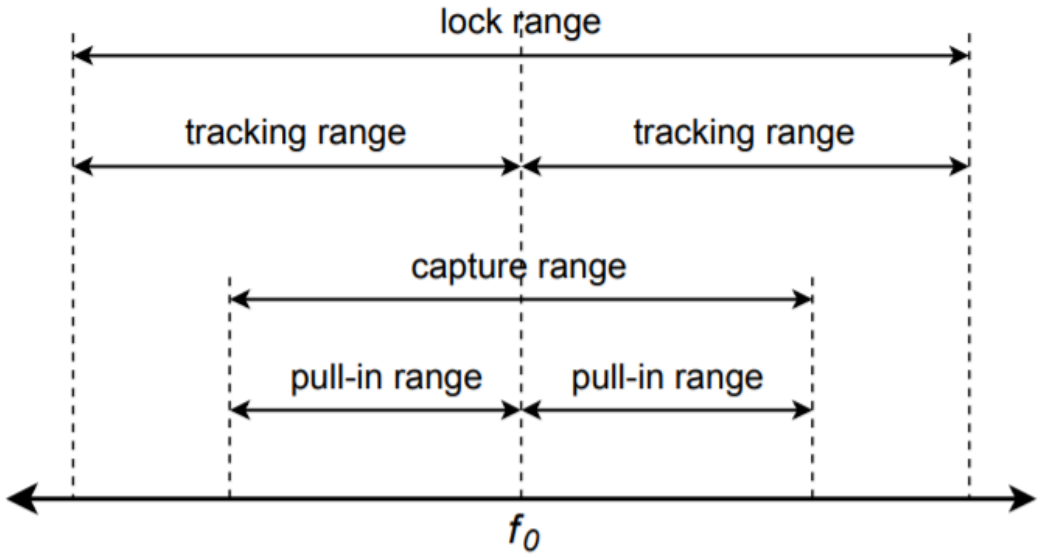


Figure 9.3.11 : Operating ranges for phase-locked loop.

Although a PLL may be able to track changes throughout the lock range, it may not be able to initially acquire sync with frequencies at the range limits. A somewhat narrower band of frequencies, called the capture range, indicates frequencies that the PLL will always be able to lock onto. Again, the capture range is usually symmetrical about ◈◈ . The deviation on either side of ◈◈ is referred to as the pull-in range. For a PLL to function properly, the input frequency must first be within the capture range. Once the PLL has locked onto the signal, the input frequency may vary throughout the larger lock range. The VCO center frequency is usually set by an external resistor or capacitor. The loop filter may also require external components. Depending on the application, the desired output signal from the PLL may be either the VCO's output, or the control voltage for the VCO.

One way to transmit binary signals is via FSK. This may be used to allow two computers to exchange data over telephone lines. Due to limited bandwidth, it is not practical to directly transmit the digital information in its normal pulse-type form. Instead, logic high and low can be represented by distinct frequencies. A square wave, for example, would be represented as an alternating set of two tones. FSK is very easy to generate. All you need to do is drive a VCO with the desired logic signal. To recover the data, the reception circuit needs to create a high or low level, depending on which tone is received. A PLL may be used for this purpose. The output signal will be the error signal that drives the VCO. The logic behind the circuit operation is deceptively simple. If the PLL is in lock, the

output frequency of its VCO must be the same as the input signal. Remembering that the incoming FSK signal is itself derived from a VCO, for the VCOs to be in lock, they must be driven with identical control signals. Therefore, the control signal that drives the PLL's internal VCO must be the same as the control signal that originally generated the FSK signal. The PLL control signal can then be fed to a comparator in order to properly match the signal to the following logic circuitry.

Along the same lines as the FSK demodulator is the standard FM signal demodulator. Again, the operational logic is the same. In order for the PLL to remain in lock, its VCO control signal must be the same as the original modulating signal. In the case of typical radio broadcasts, the modulating signal is either voice or music. The output signal will need to be AC coupled and amplified further. The PLL serves as the intermediate frequency amplifier, limiter, and demodulator. The result is a very cost-effective system.

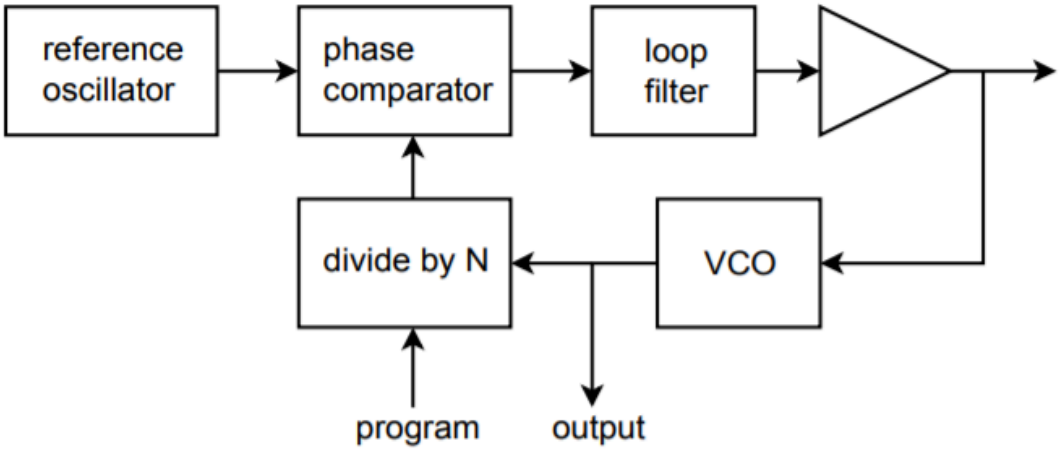


Figure 9.3.12 : PLL frequency synthesizer.

Another usage for the PLL is in frequency synthesis. From a single, accurate signal reference, a PLL may be used to derive a number of new frequencies. A block diagram is shown in Figure 9.3.12 . The major change is in the addition of a programmable divider between the VCO and the phase comparator. The PLL can only remain in lock with the reference oscillator by producing the same frequency out of the divider. This means that the VCO must generate a frequency ⋄ times higher than the reference oscillator. We can use the VCO output as desired. In order to change the output frequency, all that needs to be changed is the divider ratio. Normally, a highly accurate and stable reference, such as a quartz crystal oscillator, is used. In this way, the newly synthesized frequencies will also be very stable and accurate.

One example of an advanced digital PLL is the LTC6950. This device operates at up to 1.4 GHz and has five outputs. Each of the outputs has an independently programmable divider and VCO clock cycle delay. The input reference frequency is set between 2 MHz and 250 MHz. Due to the multiple outputs and syncing capabilities, this device can be used for large distributed systems that require precisely controlled multiple clocks. Indeed, one device can be used to control several other LTC6950s for very large systems. An example of this would be a system making use of several high-speed high-resolution analog-to-digital or digitalto-analog converters. The accuracy of these devices depends greatly on very accurate and stable clock sources. We will examine analog-to-digital-to-analog conversion in Chapter Twelve.

555 TIMER

The 555 timer is a versatile integrated circuit first introduced in the early 1970's by Signetics. It has remained a popular building block in a variety of applications ranging from simple square wave oscillators to burglar alarms to pulse-width modulators and beyond. In its most basic forms, the one-shot, or monostable, and the astable oscillator, the 555 requires only a handful of external components. Usually, only two capacitors and two resistors are needed for basic functions. The 555 is made by different manufacturers and in a few forms. The 556, for example, is a dual 555. The 555 can produce frequencies up to approximately 500 kHz. The output current is specified as 200 mA, although this entails fairly high internal voltage drops. A more reasonable expectation would be below 50 mA. The circuit may be powered from supplies as low as 5 volts and as high as 18 volts. This makes the 555 suitable for both TTL digital logic and typical op amp systems. Rise and fall times for the output square wave are typically 100 ns.

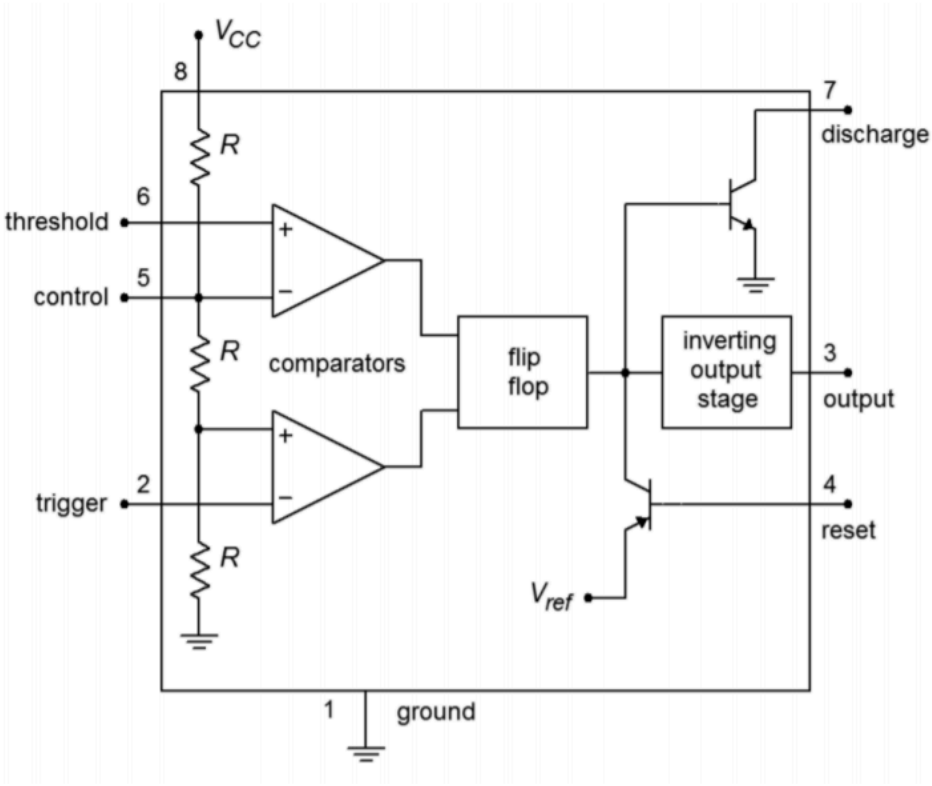


Figure 9.3.13 : Block diagram of 555 timer.

A block diagram of the 555 is shown in Figure 9.3.13 . It is comprised of a pair of comparators tied to a string of three equal-valued resistors. Note that the upper, or Threshold, comparator sees approximately 2/3 of ◈◈◈ at its inverting input, assuming no external circuitry is tied to the Control pin. (If the Control pin is unused, a 10 nF capacitor should be placed between the pin and ground.) The lower, or Trigger, comparator sees approximately 1/3 of ◈◈◈ at its noninverting input. These two comparators feed a flip-flop, which in turn feeds the output circuitry and Discharge and Reset transistors. If the flip-flop output is low, the Discharge transistor will be off. Note that the output stage is inverting, so that when the flip-flop output is low, the circuit output is high. In contrast, if the input to the Reset transistor is low, this will inhibit the output signal. If Reset capabilities are not needed, the Reset pin should be tied to ◈◈◈ .

Returning to the comparators, if the noninverting input of the Threshold comparator were to rise above 2/3 ◈◈◈ , the comparator's output would change state, triggering the flip-flop and producing a low out of the 555. Similarly, if the input to the inverting input of the Trigger comparator were to drop below 1/3 ◈◈◈ , the comparator's output would change, and ultimately, the 555 output would go high.

555 MONOSTABLE OPERATION

The basic monostable circuit is shown in Figure 9.39. In this form, the 555 will produce a single pulse of predetermined width when a negative going pulse is applied to the trigger input. Note that the three input components, ◈◈◈ , ◈◈◈ , and ◈ serve to limit and differentiate the applied pulse. In this way, a very narrow pulse will result which reduces the possibility of false triggers. To see how the circuit works, refer to the waveforms presented in Figure 9.40.

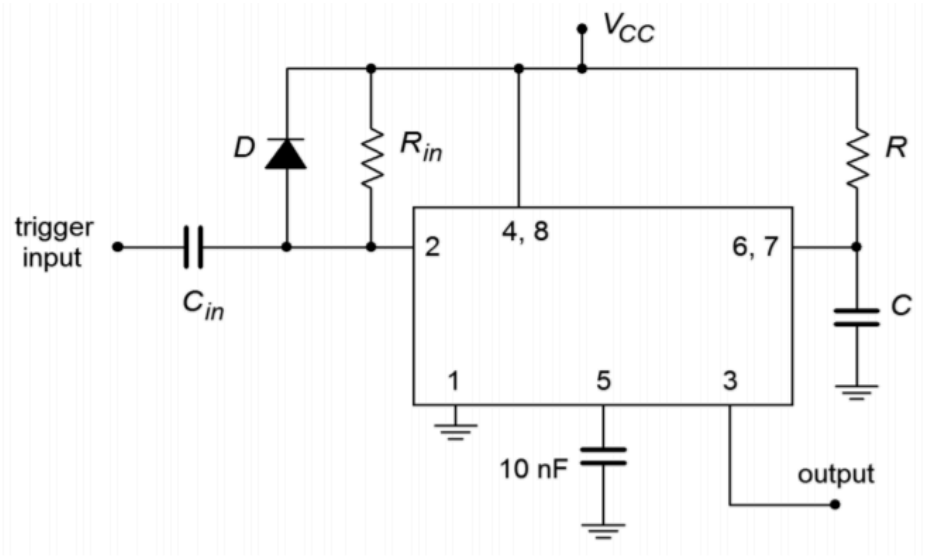


Figure 9.3.14 : 555 monostable connection.

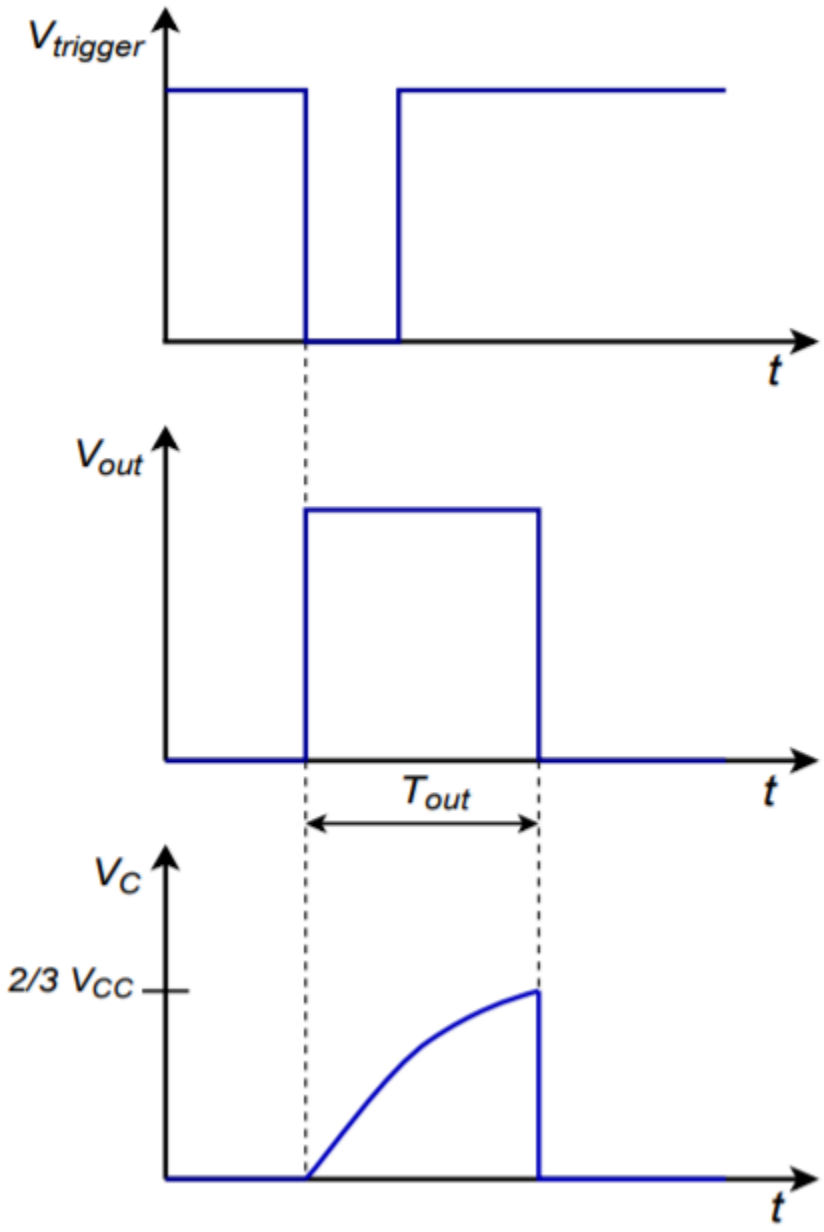


Figure 9.3.15 : 555 monostable waveforms.

Assume that the output of the 555 is initially low. This implies that the Discharge transistor is on, shorting the timing capacitor ⍰ . A narrow low pulse is applied to the input of the circuit. This will cause the Trigger comparator to change state, firing the flip-flop, which in turn will cause the output to go high and also turn off the Discharge transistor. At this point, ⍰ begins to charge toward ⍰⍰⍰ through ⍰ . When the capacitor voltage reaches 2/3 ⍰⍰⍰ , the Threshold comparator fires, setting the output low and turning on the Discharge transistor. This drains the timing capacitor, and the circuit is ready for the application of a new input pulse. Note that without the input waveshaping network, the trigger pulse must be narrower than the desired output pulse. The Equation for the output pulse width is

$$T_{out} = 1.1RC$$

An interesting item to note is that the value of ◈◈◈ does not enter into the equation. This is because the comparators are always comparing the input signals to specific percentages of ◈◈◈ rather than to specific voltages.

Example 9.3.4

Determine values for the timing resistor and capacitor to produce a 100 ◈ s output pulse from the 555.

A reasonable choice for ◈ would be 10 k Ω .

$$T_{out} = 1.1RC$$

$$C = \frac{T_{out}}{1.1R}$$

$$C = \frac{100\mu s}{1.1 \times 10k}$$

$$C = 9.09nF$$

The nearest standard value would be 10 nF, so a better choice for ◈ might be 9.1 k Ω (also a standard value). This pair would yield the desired pulse width quite accurately.

555 ASTABLE OPERATION

Figure 9.3.16 shows the basic astable, or free-running form, for a square wave generator. Note the similarities to the monostable circuit. The obvious difference is
that the former trigger input is now tied into the resistor-capacitor timing network. In effect, the circuit will trigger itself continually. To see how the circuit works,
refer to Figure 9.3.17 for the waveforms of interest.

Assume initially that the 555 output is in the high state. At this point, the Discharge transistor is off and capacitor C is charging toward Vcc through RA and RB. Eventually, the capacitor voltage will exceed 2/3 Vcc causing the Threshold comparator to trigger the flip-flop. This will turn on the Discharge transistor and make the 555 output go low. The Discharge transistor effectively places the upper end of RB at ground, removing RA and Vcc from consideration. C now discharges through RB toward 0. Eventually, the capacitor voltage will drop below 1/3 Vcc. This will fire the Trigger comparator, which will in turn place the circuit back to its initial state, and the cycle will repeat.

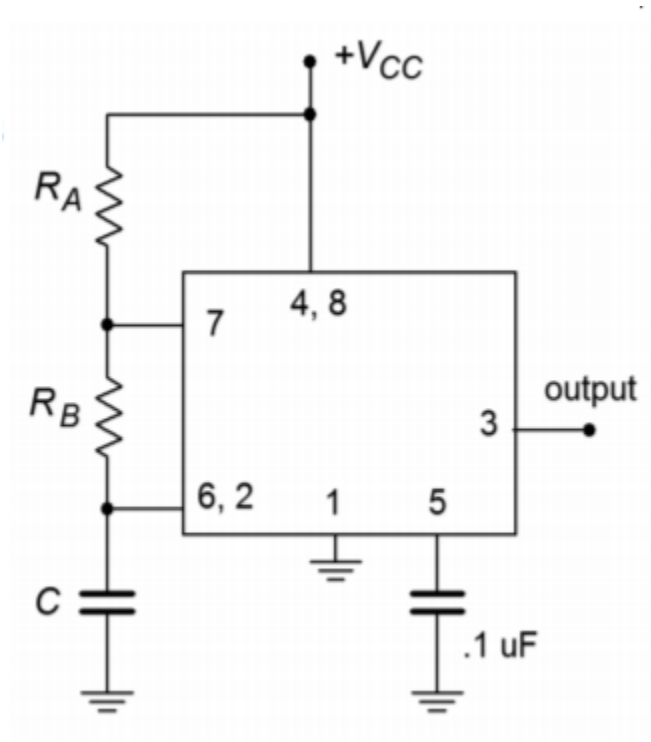


Figure 9.3.16 : 555 astable connection.

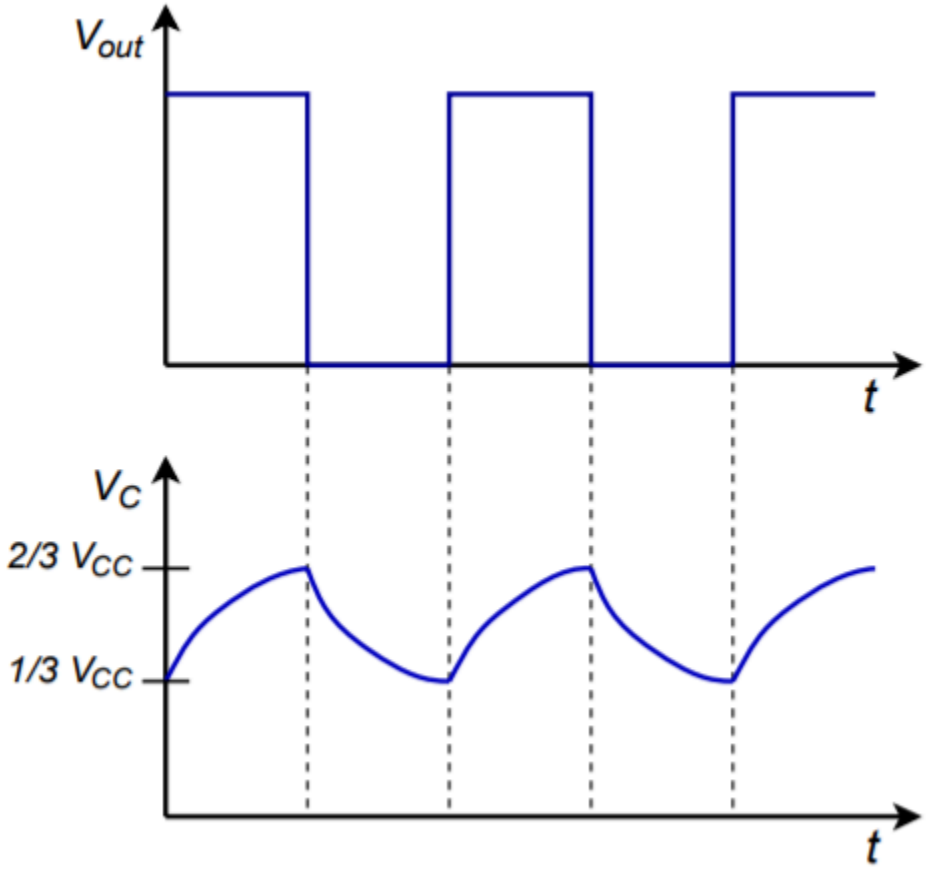


Figure 9.3.17 : 555 astable waveforms.

The frequency of oscillation clearly depends only on C, RA, and RB. The time periods are

$$T_{high} = 0.69(R_A + R_B)C$$

$$T_{low} = 0.69R_BC$$

This results in a frequency of

$$f = \frac{1.44}{R_A + 2R_B}$$

The duty cycle is normally defined as the high time divided by the period. The 555 documentation often reverses this definition, but we will stick with the industry norm.

$$DutyCycle = \frac{R_A + R_B}{R_A + 2R_B}$$

A quick examination of the duty cycle Equation shows that there is no reasonable combination of resistors that will yield 50% duty cycle, let alone anything smaller. There is a simple trick to solve this problem, though. All you need to do is place a diode in parallel with RB as illustrated in Figure 9.3.18 . The diode will be forward biased during the high time period and will effectively short out RB. During the low time period the diode will be reverse-biased, and RB will still be available for the discharge phase. If RA and RB are set to the same value, the end result will be 50% duty cycle. Of course, due to the non-ideal nature of the diode, this will not be perfect, so some adjustment of the resistor values may be in order. Further, note that if ⟐⟐ is also replaced with a potentiometer (and perhaps a series limiting resistor) a tunable square wave generator will result.

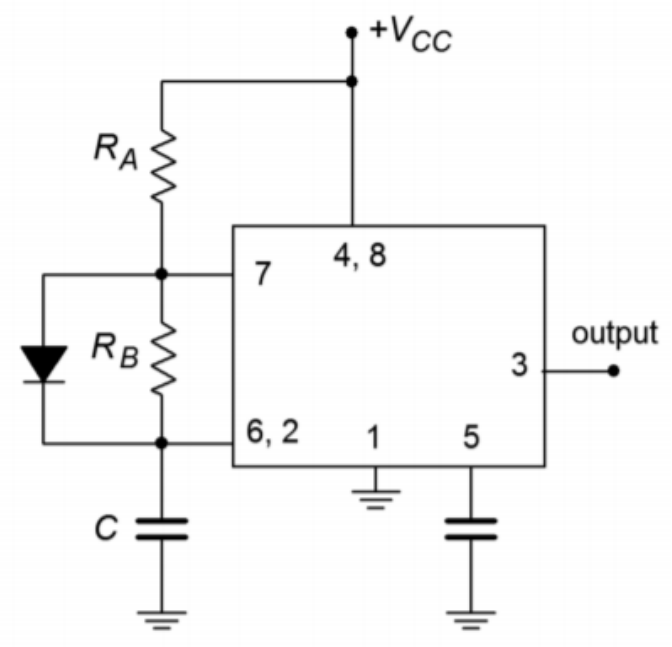


Figure 9.3.18 : 555 with shunting diode for duty cycles ≤ 50%.

Example 9.3.5

Determine component values for a 2 kHz square wave generator with an 80% duty cycle. First, note the

period is the reciprocal of the desired frequency, or 500 ◈ s.
For an 80% duty cycle, that yields

$$T_{high} = \text{ Duty Cycle } \times T$$

$$T_{high} = 0.8 \times 500\mu s$$

$$T_{high} = 400\mu s$$

$$T_{low} = T - T_{high}$$

$$T_{low} = 500\mu s - 400\mu s$$

$$T_{low} = 100\mu s$$

Choosing ◈◈=10◈Ω,

$$T_{low} = 0.69 R_B C$$

$$C = \frac{T_{low}}{0.69 R_B}$$

$$C = \frac{100\mu s}{0.69 \times 10k\Omega}$$

$$C = 14.5nF$$

$$T_{high} = 0.69(R_A + R_B)C$$

$$R_A = \frac{T_{high}}{0.69C} - R_B$$

$$R_A = \frac{400\mu s}{0.69 \times 14.5nF} - 10k\Omega$$

$$R_A = 30k\Omega$$

13.4 SUMMARY

Oscillators and frequency generators find use in a wide variety of applications. They may be realized from simple single op amp topologies or use more elaborate special purpose integrated circuits. Basic op amp oscillators are usually constrained to the frequency range below 1 MHz. Two sine wave oscillators that are based on op amps are the Wien bridge and phase shift types. Both oscillators rely on positive feedback in order to create their outputs. To maintain oscillation, the amplifier/feedback loop must conform to the Barkhausen criterion. This states that in order to maintain oscillation, the loop phase must be 0°, or an integer multiple of 360°. Also, the product of the positive feedback loss and the forward gain must be greater than unity to start oscillations and revert to unity to maintain oscillation. In order to make the gain fall back to unity, some form of gain limiting device, such as a diode or lamp, is included in the amplifier's negative feedback loop. The oscillation frequency is usually set by a simple resistor/capacitor network. As such, the circuits are relatively easy to tune. Their ultimate accuracy will depend on the tolerance of the tuning components and, to a lesser degree, on the characteristics of the op amp used.

Besides sine waves, other shapes such as triangles and squares may be produced. A simultaneous square/triangle generator may be formed from a ramp generator/comparator combination. Only two op amps are required to realize this design.

The voltage-controlled oscillator, or VCO, produces an output frequency that is dependent on an external control voltage. The free-running, or center, frequency is normally set via a resistor/capacitor combination. The control voltage may then be used to increase or decrease the frequency about the center point. The VCO is an integral part of the phase-locked loop, or PLL. The PLL has the ability to lock onto an incoming frequency. That is, its internal VCO frequency will match the incoming frequency. Should the incoming frequency change, the internal VCO frequency will change along with it. Two important parameters of the PLL are the capture range and the lock range. Capture range is the range of frequencies over which the PLL can acquire lock. Once lock is achieved, the PLL can maintain lock over a somewhat wider range of frequencies called the lock range. The PLL is in wide use in the electronics industry and is found in such applications as FM demodulation, FSK based communication systems, and frequency synthesis.

Timers can be used to generate rectangular waves of various duty cycles as well as single-shot pulses.

13.5 PROBLEMS

REVIEW QUESTIONS

1. How does positive feedback differ from negative feedback?
2. Define the Barkhausen criterion.
3. Explain the operation of the Wien bridge op amp oscillator.
4. Detail the operation of the phase shift op amp oscillator.
5. How might a square wave be generated from a sinusoidal or triangular source?
6. Give two ways to make the output frequency of a Wien bridge oscillator user-adjustable.
7. What factors contribute to the accuracy of a Wien bridge oscillator's output frequency?
8. What is a VCO, and how does it differ from a fixed-frequency oscillator?
9. Draw a block diagram of a PLL and explain its basic operation.
10. What is the difference between capture range and lock range for a PLL?
11. Give at least two applications for a fixed-frequency oscillator or VCO.
12. Give at least two applications for the PLL.
13. Explain the difference between astable and monostable operation of a timer.

PROBLEMS

Analysis Problems

Unless otherwise specified, all circuits use ± 15 V power supplies.

1. Given the circuit of Figure 9.5.1 , determine the frequency of oscillation if ⍰1=1.5⍰Ω , ⍰2=⍰3=⍰4=3⍰Ω , and ⍰1=⍰2=22⍰⍰ .

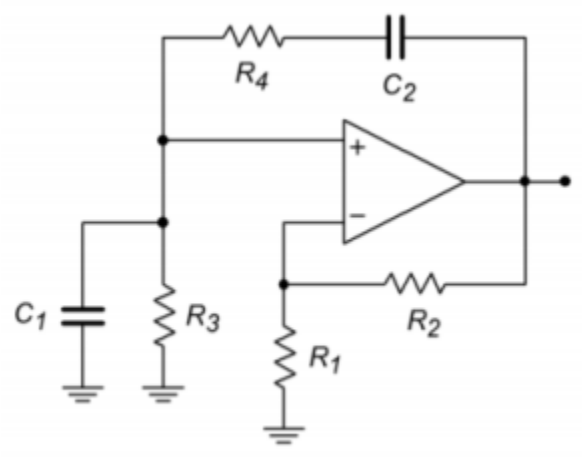


Figure 9.5.1

2. Given the circuit of Figure 9.5.1 , determine the frequency of oscillation if ◈2=22◈Ω , ◈1=◈3=◈4=11◈Ω , and ◈1=◈2=33◈◈ .
3. Given the circuit of Figure 9.5.2 , determine the maximum and minimum fo if ◈1=5.6◈Ω , ◈2=12◈Ω , ◈3=◈4=1◈Ω , ◈1=◈2=10◈Ω , ◈1=◈2=39◈◈ .

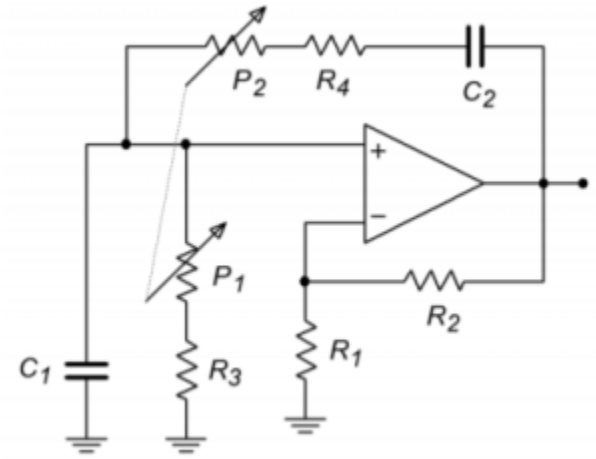


Figure 9.5.2

4. Given the circuit of Figure 9.5.3 , determine fo if ◈4=2◈Ω , ◈3=20◈Ω , ◈2=200◈Ω , ◈1=1.6◈Ω , ◈1=30◈◈ , ◈2=3◈◈ , ◈3=300◈◈ .

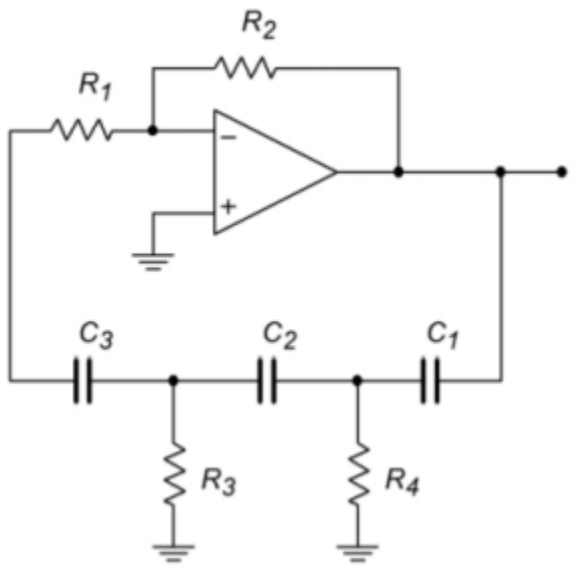


Figure 9.5.3

5. Given the circuit of Figure 9.5.3 , determine ◈◈ if ◈1=◈3=◈4=3.3◈Ω , ◈2=100◈Ω , ◈1=◈2=◈3=86◈◈ .
6. Given the circuit of Figure 9.5.4 , determine ◈◈ if ◈1=◈2=22◈Ω , ◈3=33◈Ω , ◈=3.3◈◈ .

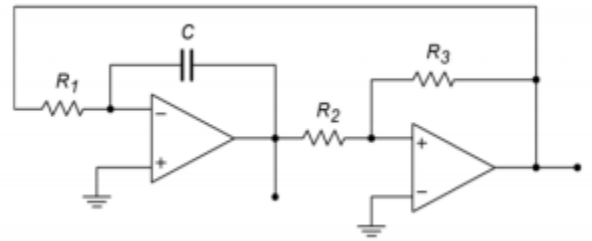


Figure 9.5.4

7. Using the circuit of Figure 9.5.5 , determine the output voltage if ◈◈◈◈=100◈ .

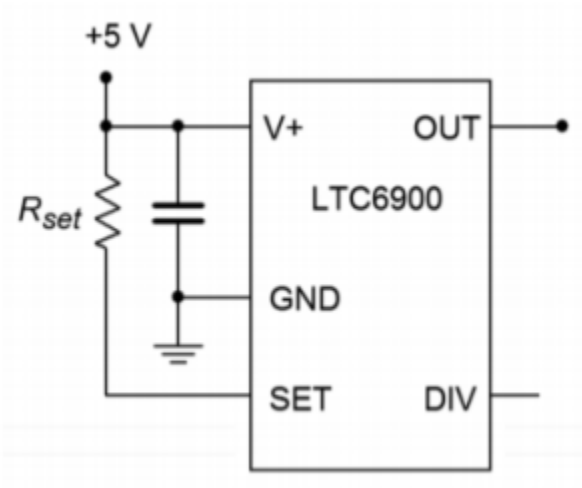


Figure 9.5.5

8. Using the circuit of Figure 9.5.6 , determine the output voltage if ⍰⍰⍰⍰=50⍰ .

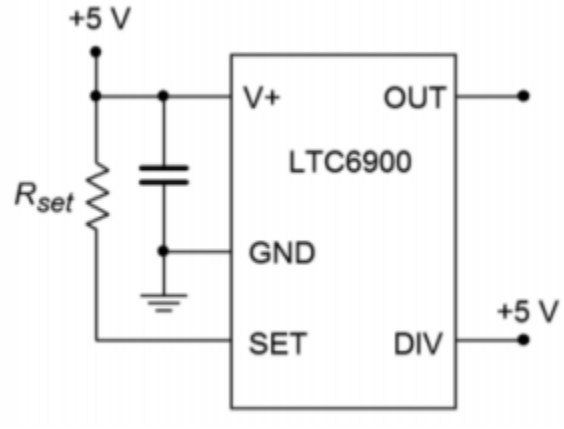


Figure 9.5.6

9. A temperature dependent resistor, or thermistor, is used in Figure 9.5.7 . If the resistance varies between 20 k and 200 k through the temperature range of interest, determine the range of output frequencies.

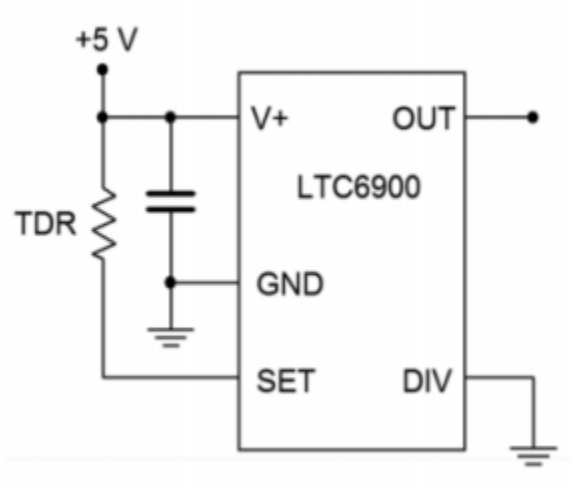


Figure 9.5.7

10. For the circuit of 9.5.8 , determine the range of output frequencies if ⍰⍰ varies between 0 V and -1 V. ⍰⍰=100⍰ , ⍰1=1⍰ , ⍰2=681⍰ .
11. For the circuit of Problem 9.10, determine the output frequencies if ⍰⍰ is a 100 Hz square wave at 0.5 volts peak.
12. Sketch the output waveform for Problem 9.11.

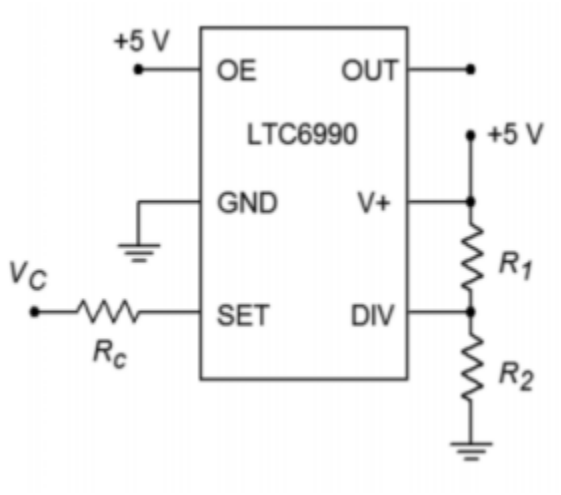


Figure 9.5.8

13. If a control voltage of 0.4 sin 2π60t is used for the circuit of Problem 9.10, find the resulting maximum and minimum output frequencies.
14. Given the circuit of Figure 9.5.9 , determine the output frequency if ◈◈=2◈ , ◈◈◈◈=◈◈◈◈=100◈ , ◈1=976◈ and ◈2=102◈ .

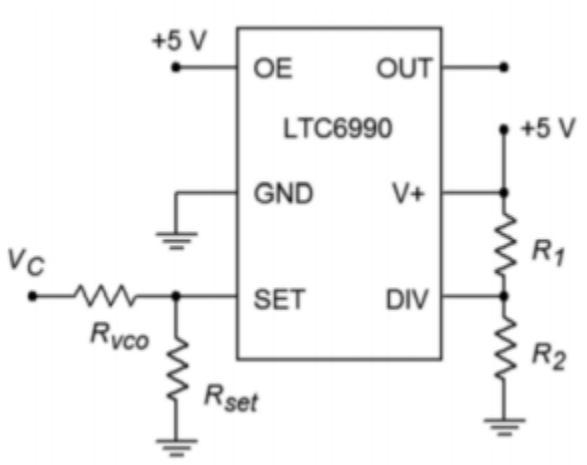


Figure 9.5.9

15. Given the circuit of Figure 9.5.4 , determine fo if ◈1=◈2=22◈Ω , ◈3=33◈Ω , and ◈=3.3◈◈ .
16. Determine the output frequency range in Figure 9.5.10 if ◈◈ varies from 0 to 2 volts, ◈◈=200◈ , ◈◈=100◈ , ◈◈=500◈ , ◈1=976◈ and ◈2=182◈ .

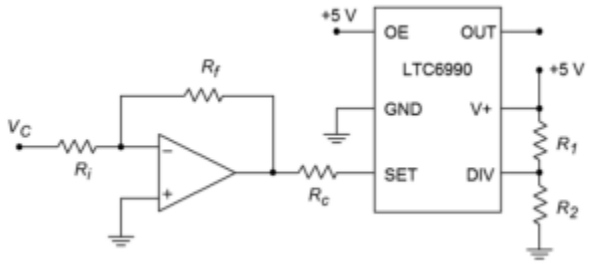


Figure 9.5.10

17. Determine the output frequency range in Figure 9.5.11 if ◈◈ varies from 0 to 2 volts, ◈◈=1◈◈◈◈ , ◈◈=200◈ , ◈◈=100◈ , ◈◈=200◈ , ◈◈=390◈ , ◈1=182◈ and ◈2=976◈ .

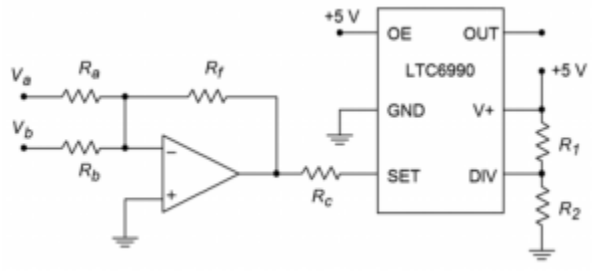


Figure 9.5.11

Design Problems

18. For the circuit of Figure 9.5.1 , determine values for ◈1 and ◈2 if ◈1=6.8◈Ω , ◈2=◈3=◈4=15◈Ω , and ◈◈=30◈◈◈ .
19. For the circuit of Figure 9.5.1 , determine values for ◈3 and ◈4 if ◈1=2.2◈Ω , ◈2=4.7◈Ω , ◈1=◈2=47◈◈ , and ◈◈=400◈◈ .
20. For the circuit of Figure 9.5.1 , determine values for ◈2,◈1 and ◈2 if ◈1=7.2◈Ω,◈3=◈4=3.9◈Ω , and ◈◈=19◈◈◈ .
21. Determine the values required for ◈3,◈4,◈1 , and ◈2 in Figure 9.5.2 if ◈1=◈2=98◈◈ , ◈1=5.6◈Ω,◈2=12◈Ω , ◈◈.◈◈◈=2◈◈◈ , and ◈◈.◈◈◈=20◈◈◈ .
22. Repeat Problem 21 for ◈◈.◈◈◈=10◈◈◈ and ◈◈.◈◈◈=30◈◈◈ .
23. Redesign the circuit of Problem 1 so that exact gain resistors are not needed. Use Figure 9.6 as a model.
24. Redesign the circuit of Problem 3 so that clipping does not occur. Use Figure 9.2.6 as a model.
25. Determine new values for the capacitors of Problem 4 if fo is changed to 10 kHz.
26. Given the circuit of Figure 9.5.3 , determine values for the capacitors if ◈1=◈3=◈4=3.3◈Ω , ◈2=100◈Ω , and ◈◈=7.6◈◈◈ .
27. Given the circuit of Figure 9.5.3 , determine values for the resistors if the capacitors all equal 1100 pF and fo = 15 kHz.
28. Determine the capacitor and resistor values for the circuit of Figure 9.5.3 if ◈2=56◈Ω and

◇◇=1◇◇◇ .

29. Find C in Figure 9.5.4 if ◇◇=5◇◇◇,◇1=◇3=39◇Ω,◇2=18◇Ω .
30. Determine the resistor values in Figure 9.5.4 if ◇◇=20◇◇◇ and ◇=22◇◇ . Set ◇1=◇2 and ◇2=◇3/2 . Sketch the output waveforms as well.
31. Determine the required ratio for ◇2/◇3 to set the triangle wave output to 5 V peak in Figure 9.5.4 .
32. Using the circuit of Figure 9.5.5 , find ◇◇◇◇ for an output of 100 kHz.
33. Determine the value for ◇◇◇◇ in Figure 9.5.6 to set the frequency to 50 kHz.
34. Design a square wave generator that is adjustable from 5 kHz to 20 kHz.
35. For the circuit of Figure 9.5.9 , determine the component values such that a frequency of 250 kHz is produced when ◇◇=0 volts and 125 kHz when ◇◇ is 1 volt.
36. Design a ◇◇◇ circuit and determine the component values such that a frequency of 250 kHz is produced when ◇◇=1 volt and 125 kHz when ◇◇ is 0 volts.

CHALLENGE PROBLEMS

37. Using Figure 9.2.9 as a guide, design a sine wave oscillator that will operate from 2 Hz to 20 kHz, in decade ranges
38. Design a 10 kHz TTL-compatable square wave oscillator using a Wien bridge oscillator and a 311 comparator.
39. Using a triangle or sine wave oscillator and a comparator, design a variable duty cycle pulse generator. Hint: Consider varying the comparator reference.
40. Using a function synthesizer (Chapter Seven) and the oscillator of Figure 9.5.4 , outline a simple laboratory frequency generator with sine, triangle, and square wave outputs.
41. For the preceding problem, outline how amplitude and DC offset controls could be implemented as well.
42. Generate a square wave that smoothly increases from 50 kHz to 300 kHz and back at a rate 100 times each second.
43. Assume that the output of the left-most op amp of Figure 9.5.4 drives Vb of Figure 9.5.11 . Further, assume that a positive DC voltage equal to the peak value of ◇◇ is used to drive ◇◇ . Also, ◇◇=◇◇=◇◇ . Assuming the frequency produced by Figure 9.5.4 is considerably lower than that of Figure 9.5.11 , describe the output waveform of Figure 9.5.11 .

Computer Simulation Problems

44. Perform a simulation of the circuit of Problem 1. Perform a frequency domain analysis of the positive feedback loop's gain and phase, and verify that the Barkhausen Criterion is met.
45. Perform a simulation for the circuit of Problem 4. Perform a frequency domain analysis of the

positive feedback loop's gain and phase, and verify that the Barkhausen Criterion is met.

46. Perform a time-domain simulation analysis for the circuit of Problem 6. Make sure that you check both outputs (a simultaneous plot would be best).

UNIT 14: INTEGRATORS AND DIFFERENTIATORS

Learning Objectives

After completing this chapter, you should be able to:

- Describe the fundamental usefulness and operation of an integrator.
- Describe the fundamental usefulness and operation of a differentiator.
- Detail the modifications required in order to make a practical op amp integrator or differentiator.
- Plot the useful frequency range of a given integrator or differentiator.
- Analyze the operation of integrator circuits using both time-continuous and time-discrete methods.
- Analyze the operation of differentiator circuits using both time-continuous and time-discrete methods.
- Explain the operation of an analog computer.

14.1 INTRODUCTION

Up to this point, we have examined a number of different op amp circuits and applications. Viewed from a purely mathematical perspective, the circuits perform basic functions. Amplifiers multiply an input quantity by a constant. Voltage-controlled or transconductance amplifiers can be used to multiply a quantity by a variable. Absolute value and logarithmic functions can also be produced. There is no reason to stop with just these tools; higher mathematical functions may be enlisted. In this chapter we shall examine circuits that perform integration and differentiation. Although these circuits may appear to be somewhat esoteric at first glance, they can prove to be quite useful. Integrators perform the function of summation over time. They may be used whenever an integration function is required, for example, to solve differential equations. The differentiator is the mirror of the integrator and may be used to find rates of change. One possible application is finding acceleration if the input voltage represents a velocity.

Integrators and differentiators may be combined with summing amplifiers and simple gain blocks to form analog computers, that can be used to model physical systems. This can be a valuable aid in the initial design and testing of such things as mechanical suspension systems or loudspeakers. Unlike their digital counterparts, analog computers do not require the use of a programming language, per se. Also, they can respond in real-time
to an input stimulus.

Besides the obvious use as a direct mathematical tool, integrators and differentiators can be used for wave shaping purposes. As an example, the square/triangle generator of Chapter Nine used an integrator. With the exception of simple sinusoids, both the integrator and differentiator will change the fundamental shape of a waveform. This might be contrasted with a simple amplifier that only makes a wave larger and does not alter the basic shape. For practical work, certain alterations will be required to the ideal circuits. The effects of these changes, both good and bad, will also be studied.

14.2 INTEGRATORS

The basic operation of an integrator is shown in Figure 10.2.1 . The output voltage is the result of the definite integral of ⍰⍰⍰from time = 0 to some arbitrary time ⍰ . Added to this will be a constant that represents the output of the network at ⍰=0 . Remember, integration is basically the process of summation. You can also think of this as finding the "area under the curve". The output of this circuit always represents the sum total of the input values up to that precise instant in time. Consequently, if a static value (non-zero) is used as an input, the output will continually grow over time. If this growth continues unchecked, output saturation will occur. If the input quantity changes polarity, the output may also change in polarity. Having established the basic operation of the circuit, we are left with the design realization and limitations to be worked out.

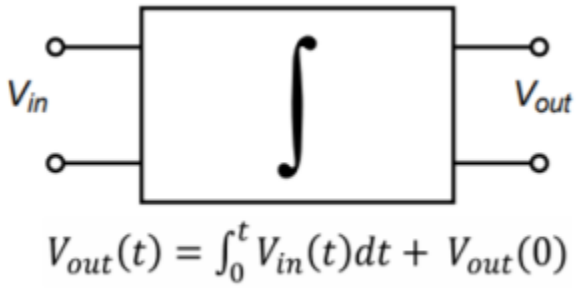


Figure 10.2.1 : A basic integrator.

As noted in earlier work, the response of an op amp circuit with feedback will reflect the characteristics of the feedback elements. If linear elements are used, the resulting response will be linear. If a logarithmic device is used in the feedback loop, the resulting response will have a log or anti-log character. In order to achieve integration, then, the feedback network requires the use of an element that exhibits this characteristic. In other words, the current through the device must be proportional to either the integral or differential of the voltage across it. Inductors and capacitors answer these requirements, respectively. It should be possible then, to create an integrator with either an inductor or a capacitor. Capacitors tend to behave in a more ideal fashion than do inductors. Capacitors exhibit virtually no problems with stray magnetic field interference, do not have the saturation limitations of inductors, are relatively inexpensive to manufacture, and operate reliably over a wide frequency range. Because of these factors, the vast majority of integrators are made using capacitors rather than inductors. The characteristic Equation for a capacitor is

$$i(t) = C\frac{dv(t)}{dt}$$

(10.2.1)

This tells us that the current charging the capacitor is proportional to the differential of the input voltage. By integrating Equation 10.2.1 , it can be seen that the integral of the capacitor current is proportional to the capacitor voltage.

$$v(t) = \frac{1}{C}\int_0^t i(t)dt$$

(10.2.2)

Assume for a moment that the capacitor voltage is the desired output voltage, as in Figure 10.2.2 . If the capacitor current can be derived from the input voltage, the output voltage will be proportional to the integral of the input. The circuit of Figure 10.2.3 will satisfy our requirements. Note that it is based on the parallel-parallel inverting amplifier studied earlier. The derivation of its characteristic Equation follows.

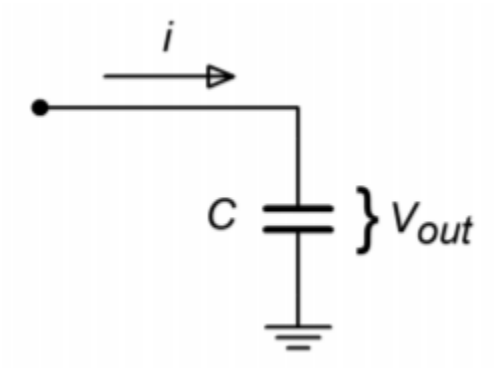


Figure 10.2.2 : Capacitor as integration element.

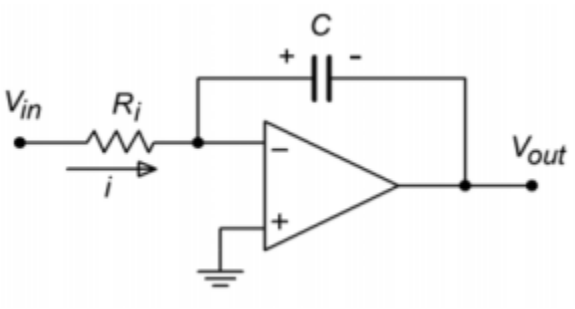


Figure 10.2.3 : A simple op amp integrator.

First, note that the voltage across the capacitor is equal to the output voltage. This is due to the virtual ground at the inverting op amp terminal. Given the polarities marked, the capacitor voltage is of opposite polarity.

$$V_{out}(t) = -V_c(t)$$

(10.2.3)

Due to the virtual ground at the op amp's inverting input, all of ◈◈◈drops across ◈◈ . This creates the input current, ◈ .

$$i(t) = \frac{V_{in}(t)}{R_i}$$

(10.2.4)

Because the current drawn by the op amp's inputs is negligible, all of the input current flows into the capacitor, ◈ . Combining Equations 10.2.2 and 10.2.4 yields

$$V_c(t) = \frac{1}{C} \int \frac{V_{in}(t)}{R_i} dt$$

[V_c (t)= \frac{1}{R_i C} \int V_{in} (t)dt \] (10.2.5)

Combining Equation 10.2.5 with Equation 10.2.3 produces the characteristic Equation for the circuit.

$$V_{out}(t) = -\frac{1}{R_i C} \int V_{in}(t) dt$$

(10.2.6)

Note that this Equation assumes that the circuit is initially relaxed (i.e., there is no charge on the capacitor at time = 0). The only difference between Equation 10.2.6 and the general Equation as presented in Figure 10.2.1 is the preceding constant –1/◈◈◈ . If the inversion or magnitude of this constant creates design problems, it can usually be corrected by gain/attenuation networks and/or inverting buffers. Finally, it is worth noting that due to the virtual ground at the inverting input, the input impedance of this circuit is approximately equal to ◈◈ , as is the case with the ordinary parallel-parallel inverting amplifier.

ACCURACY AND USEFULNESS OF INTEGRATION

As long as the circuit operation follows the model presented above, the accuracy of Equation 10.2.6 is very high. Small errors occur due to the approximations made. One possible source is the op amp's input bias current. If the input signal current is relatively low, the idealization that all of the input current will bypass the op amp and flow directly into the capacitor is no longer realistic. One possible solution is to use an FET input op amp. The limited frequency response and noise characteristics of the op amp will also play a role in the ultimate circuit accuracy. Generally, slew rate performance is not paramount, as integrators tend to "smooth out" signal variations. As long as the input signal stays within the op amp's frequency limits, well above the noise floor, and the output does not saturate, accurate integration results.

Circuits of this type can be used to model any number of physical processes. For example, the heat stress on a given transducer might be found by integrating the signal across it. The resulting signal could then be compared to a predetermined maximum. If the temperature proved to be too high, the circuit could be powered down as a safety measure. Although the initial reaction might be to directly measure the temperature with some form of thermal sensor, this is not always practical. In such a case, simulation of the physical process is the only reasonable route.

OPTIMIZING THE INTEGRATOR

For best performance, high-quality parts are a must. Low offset and drift op amps are needed, as the small DC values that they produce at the input will eventually force the circuit into saturation. Although they are not shown in Figure 10.2.3 , bias and offset compensation additions are usually required.

Very low offset devices such as chopper-stabilized (or commutating auto-zero) op amps may be used for lower frequency applications. For maximum accuracy, a small input bias current is desired, and thus, FET input op amps should be considered. Also, the capacitor should be a stable, low-leakage type, such as polypropylene film. Electrolytic capacitors are generally a poor choice for these circuits.

Even with high-quality parts, the basic integrator can still prove susceptible to errors caused by small DC offsets. The input offset will cause the output to gradually move towards either positive or negative saturation. When this happens, the output signal is distorted, and therefore, useless. To prevent this it is possible to place a shorting switch across the capacitor. This can be used to re-initialize the circuit from time to time. Although this does work, it is hardly convenient. Before we can come up with a more effective cure, we must first examine the cause of the problem.

We can consider the circuit of Figure 10.2.3 to be a simple inverting amplifier, where ⍰⍰ is replaced by the reactance of the capacitor, ⍰⍰ . In this case, we can see that the magnitude of the voltage gain is

$$A_v = -\frac{X_C}{R_i} = -\frac{1}{2\pi f C R_i}$$

Because ⍰⍰ is frequency-dependent, it follows that the voltage gain must be frequency-dependent. Because ⍰⍰ increases as the frequency drops, the voltage gain must increase as the frequency decreases. To be more specific, the gain curve will follow a –6 dB per octave slope. Eventually, the amplifier's response reaches the open-loop response. This is shown in Figure 10.2.4 .

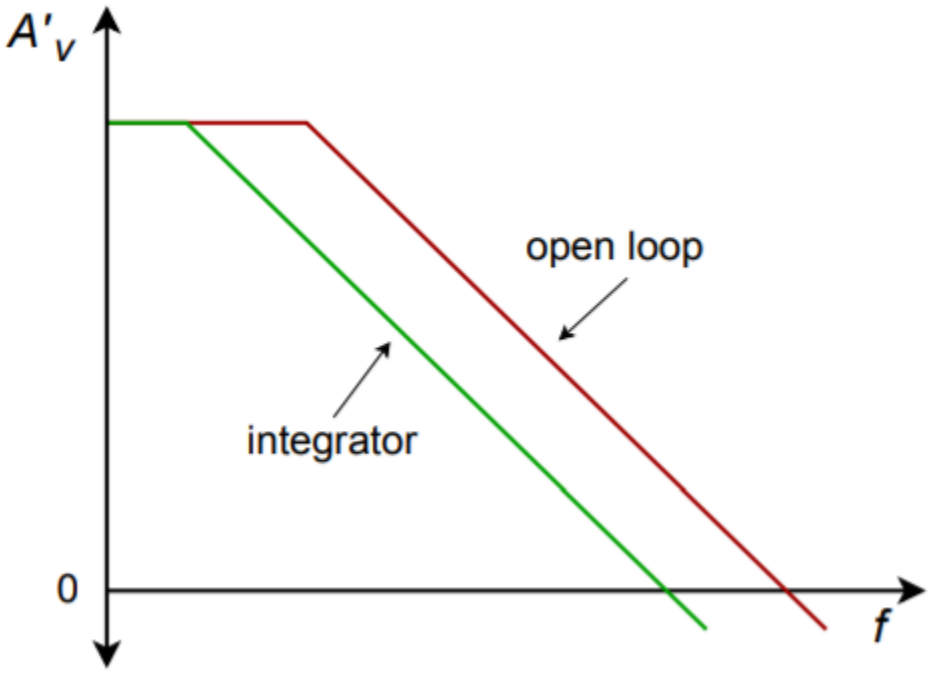


Figure 10.2.4 : Response of a simple integrator.

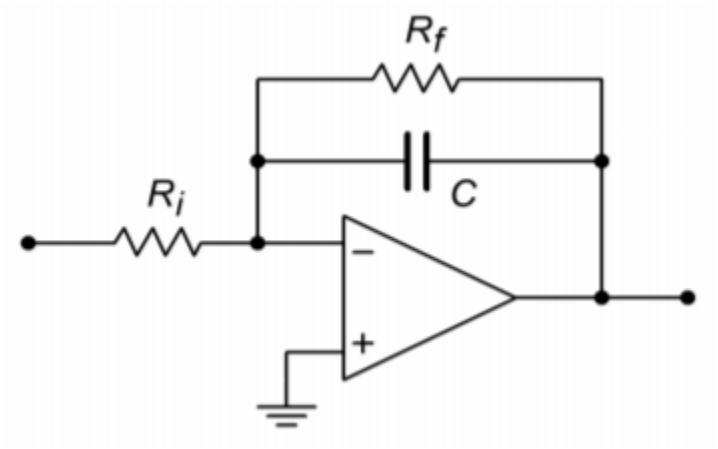


Figure 10.2.5 : A practical integrator.

The DC gain of the system is at the open loop-level, thus it is obvious that even very small DC inputs can play havoc with the circuit response. If the low-frequency gain is limited to a more modest level, saturation problems can be reduced, if not completely eliminated. Gain limiting may be produced by shunting the integration capacitor with a resistor, as shown in Figure 10.2.5 . This resistor sets the upper limit voltage gain to

$$A_{max} = -\frac{R_f}{R_i}$$

As always, there will be a trade-off with modification. By definition, integration occurs where the amplitude response rolls off at –6 dB per octave, as this is the response of our idealized capacitor model. Any alteration to the response curve will impact the ultimate accuracy of the integration. By limiting the low frequency gain, the Bode amplitude response will no longer maintain a constant –6 dB per octave rolloff. Instead, the response will flatten out below the critical frequency set by ⍰ and ⍰⍰ , as shown in Figure 10.2.6 . This critical frequency, ⍰⍰⍰⍰ , is found in the standard manner,

$$f_{low} = \frac{1}{2\pi R_f C}$$

(10.2.7)

This is a very important frequency. It tells us where the useful integration range starts. If the input frequency is less than ⍰⍰⍰⍰ , the circuit acts like a simple inverting amplifier and no integration results. The input frequency must be higher than ⍰⍰⍰⍰ for useful integration to occur. The question then, is how much higher? A few general rules of thumb are useful. First, if the input frequency equals ⍰⍰⍰⍰ , the resulting calculation will only be about 50% accurate.[1] This means that significant signal amplitude and phase changes will exist relative to the ideal. Second, at about 10 times ⍰⍰⍰⍰ , the accuracy is about 99%. At this point, the agreement between the calculated and actual measured values will be quite solid.

1. The term 50% accurate is chosen for convenience. The actual values are 50% of the calculated for power and phase angle and 70.7% of calculated for voltage

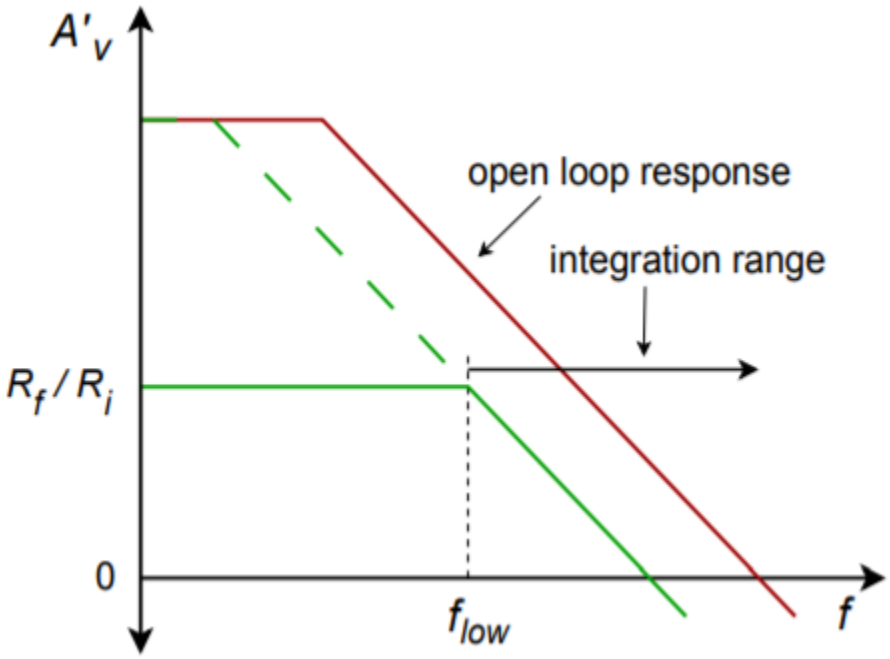


Figure 10.2.6 : Response of a practical integrator.

Example 10.2.1

Determine the Equation for ⧄, and the lower frequency limit of integration for the circuit of Figure 10.2.7 .

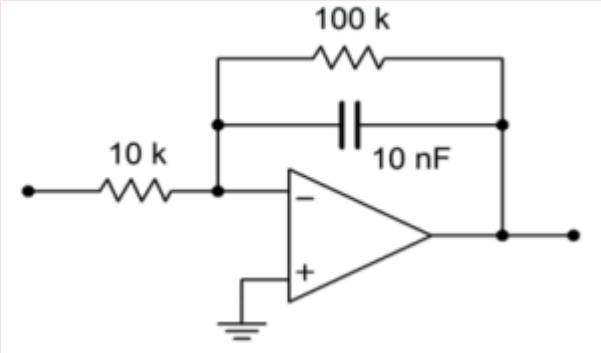


Figure 10.2.7 : Integrator for Example 10.2.1 .

The general form of the output Equation is given by Equation 10.2.6 .

$$V_{out}(t) = \frac{1}{R_i C} \int V_{in}(t) dt$$

[V_{out}(t)=– \frac{1}{10 \text{ k}\times 10 \text{ nF}}\int V_{in} (t) dt \]

[V_{out}(t)=–10^4\int V_{in} (t) dt \]

The lower limit of integration is set by ◈◈◈◈ .

$$f_{low} = \frac{1}{2\pi R_f C}$$

[f_{low} = \frac{1}{2 \pi 100 \text{ k}\times 10 \text{ nF}} \]

[f_{low} =159 \text{ Hz} \]

This represents our 50% accuracy point. For 99% accuracy, the input frequency should be at least one decade above ◈◈◈◈ , or 1.59 kHz. Accurate integration will continue to higher and higher frequencies.

The upper limit to useful integration is set by two factors: the frequency response of the op amp, and the signal amplitude versus noise. Obviously, at much higher frequencies, the basic assumptions of the circuit's operation will no longer be valid. The use of wide-bandwidth op amps reduces this limitation but cannot eliminate it. All amplifiers will eventually reach an upper frequency limit. Perhaps not so obvious is the limitation caused by signal strength. As noted earlier, the response of the integrator progresses at –6 dB per octave, which is equivalent to –20 dB per decade. In other words, a tenfold increase in input frequency results in a tenfold reduction in output amplitude. For higher frequencies, the net attenuation can be very great. At this point, the output signal runs the risk of being lost in the output noise. The only way around this is to ensure that the op amp is a low-noise type, particularly if a wide range of input frequencies will be used. For integration over a narrow band of frequencies, the RC integration constant may be optimized to prevent excessive loss.

Now that we have examined the basic structure of the practical integrator, it is time to analyze its response to different input waveforms. There are two basic ways of calculating the output: time-continuous or time-discrete. The time-continuous method involves the use of the indefinite integral, and is well suited for simple sinusoidal inputs. Waveforms that require a more complex time domain representation, such as a square wave, may be analyzed with the time-discrete method. This corresponds to the definite integral.

ANALYZING INTEGRATORS WITH THE TIME-CONTINUOUS METHOD

The time-continuous analysis approach involves finding a time-domain representation of the input waveform and then inserting it into Equation 10.2.6 . The indefinite integral is taken, and the result is the time-domain representation of the output waveform. The constant term produced by the indefinite integral is ignored. As long as the input waveform can be written in a time-domain form, this method may be used. Although any waveform may be expressed in this way, it is not always practical or the most expedient route. Indeed, relatively common waveforms such as square waves and triangle waves require an infinite series time-domain representation. We shall not deal with these waveforms in this manner. The time-continuous integration of these functions is left as an exercise in the Challenge Problems at the end of this chapter.

Example 10.2.2

Using the circuit of Figure 10.2.7 , determine the output if the input is a 1 V peak sine wave at 5 kHz. First, write the input signal as a function time.

$$V_{in}(t) = 1 \sin 2\pi 5000t$$

Substitute this input into Equation 10.2.6 :

$$V_{out}(t) = -\frac{1}{R_i C} \int V_{in}(t) dt$$

[V_{out}(t) = – \frac{1}{10 \text{ k}\times 10 \text{ nF}} \int 1 \sin 2 \pi 5000t dt \]

[V_{out}(t) = –10^4 \int 1 \sin 2 \pi 5000t dt \]

[V_{out}(t) = -10^{4} \frac{1}{2 \pi 5000t} \int 2 \pi 5000t \sin 2 \pi 5000t dt \]

[V_{out}(t) = -0.318 (- \cos 2 \pi 5000 t) \]

[V_{out}(t) = 0.318 \cos 2 \pi 5000 t \nonumber

Thus, the output is a sinusoidal wave that is leading the input by 90 ° and is 0.318 V peak. Again, note that the constant produced by the integration is ignored. In Example 1 it was noted that ⦁⦁⦁⦁ for this circuit is 159 Hz. Because the input frequency is over 10 times ⦁⦁⦁⦁ , the accuracy of the result should be better than 99%. Note that if the input frequency is changed, both the output frequency and amplitude change. For example, if the input were raised by a factor of 10 to 50 kHz, the output would also be at 50 kHz, and the output amplitude would be reduced by a factor of 10, to 31.8 mV. The output would still be a cosine, though.

COMPUTER SIMULATION

Figure 10.2.8 shows the Multisim simulation of the circuit of Example 10.2.2 . The steady-state response is graphed in Figure 10.2.8⦁ ..

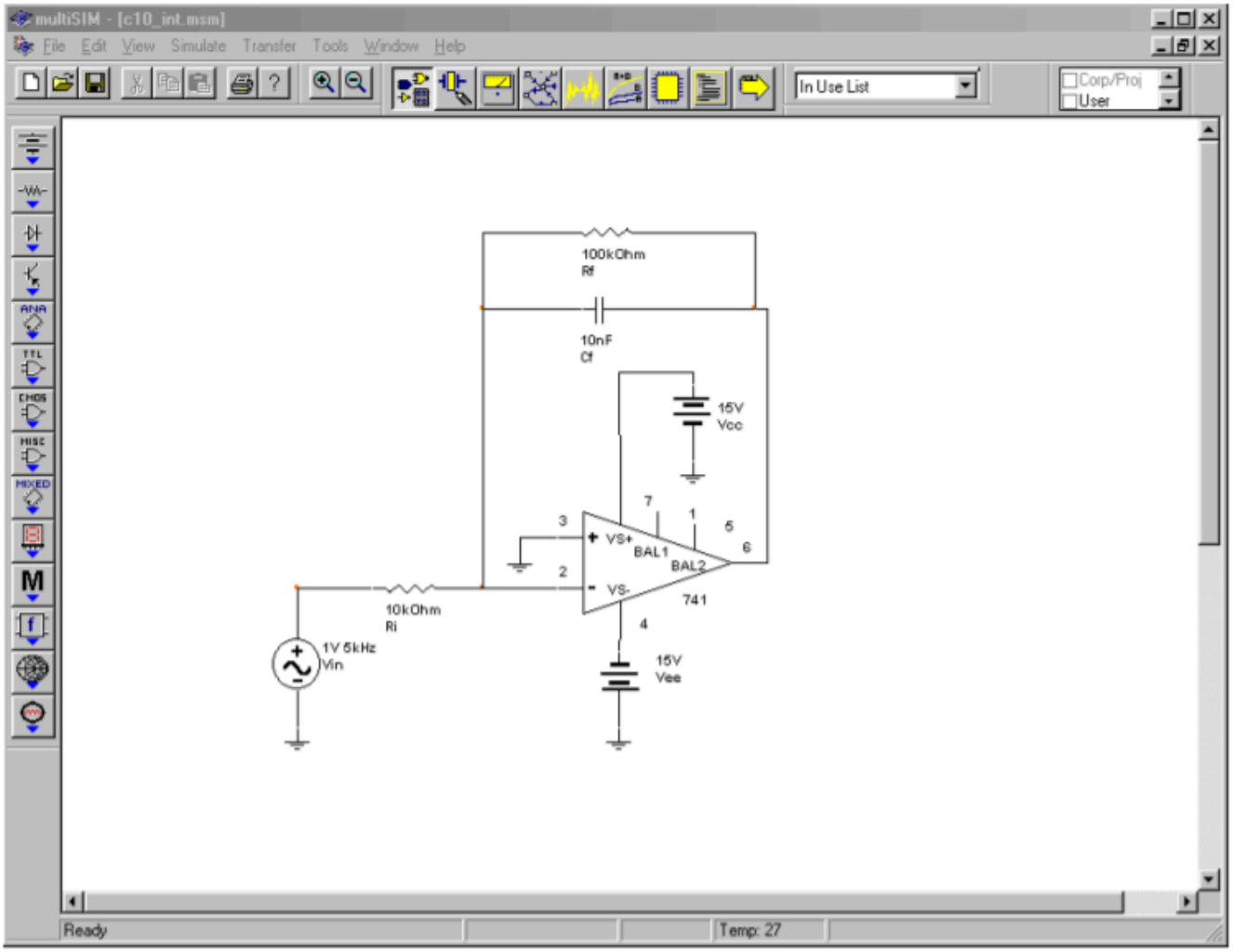


Figure 10.2.8 : Integrator in Multisim.

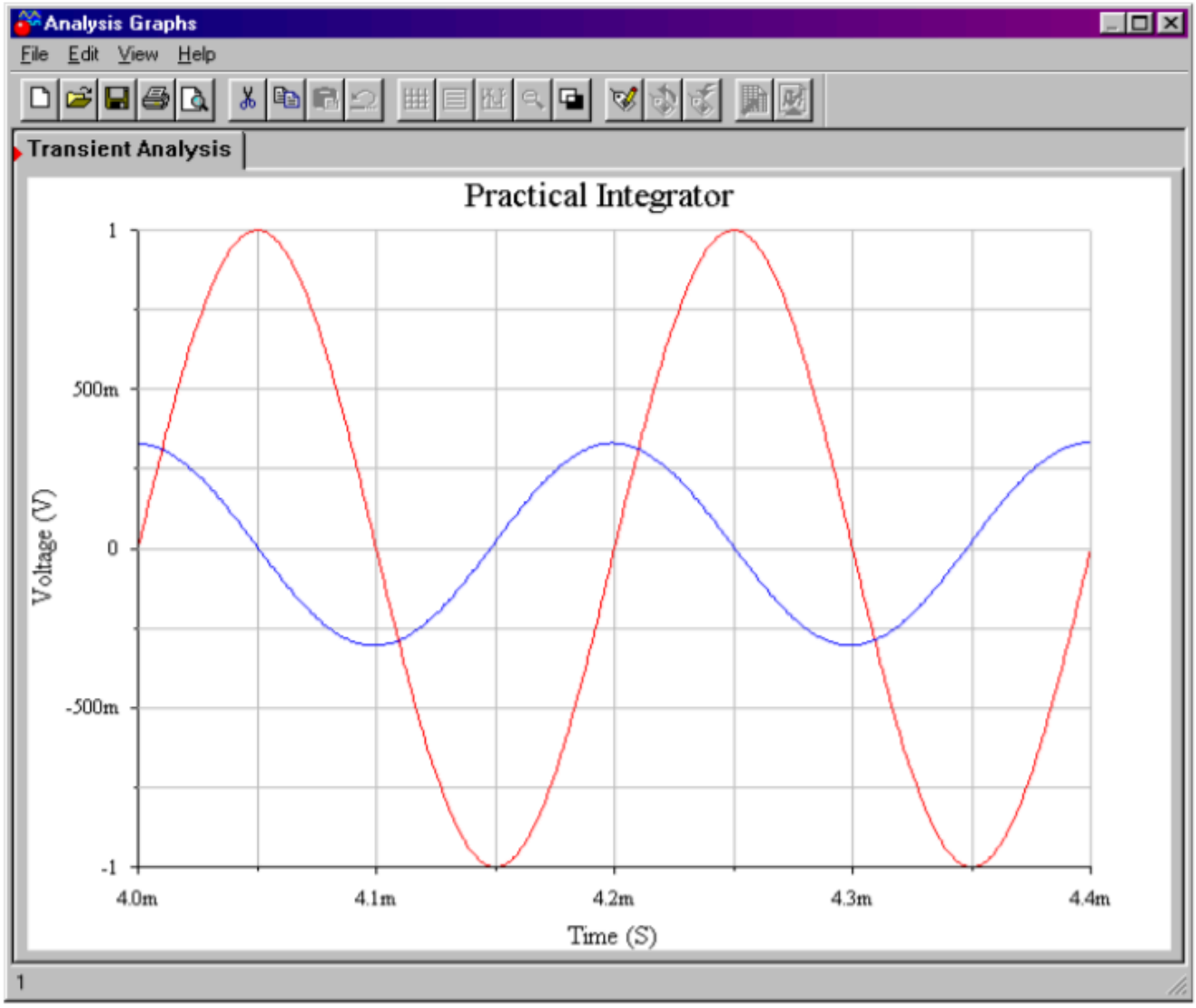


Figure 10.2.8 : Integrator input and output waveforms.

Note that in order to achieve steady-state results, the output is plotted only after several cycles of the input have passed through. Even after some 20 cycles, the output is not perfectly symmetrical. In spite of this, the accuracy of both the amplitude and phase is quite good when compared to the calculated results. This further reinforces the fact that the circuit is operating within its useful range. It can be instructive to rerun the simulation to investigate the initial response as well.

ANALYZING INTEGRATORS WITH THE TIME-DISCRETE METHOD

Unlike the time-continuous approach, this method uses the definite integral and is used to find an output level at specific instances in time. This is useful if the continuous time-domain representation is somewhat complex, and yet the wave shape is relatively simple, as with a square wave. Often, a little logic may be used to ascertain the shape of the resulting wave. The integral is then used to determine the exact amplitude. The time-discrete method also proves useful for more complex waves when modeled within a computer program. Here, several calculations may be performed per cycle, with the results joined together graphically to form the output signal.

The basic technique revolves around finding simple time-domain representations of the input for

specific periods of time. A given wave might be modeled as two or more sections. The definite integral is then applied over each section, and the results joined.

Example 10.2.3

Sketch the output of the circuit shown in Figure 10.2.7 if the input signal is a 10 kHz, 2 V peak square wave with no DC component.

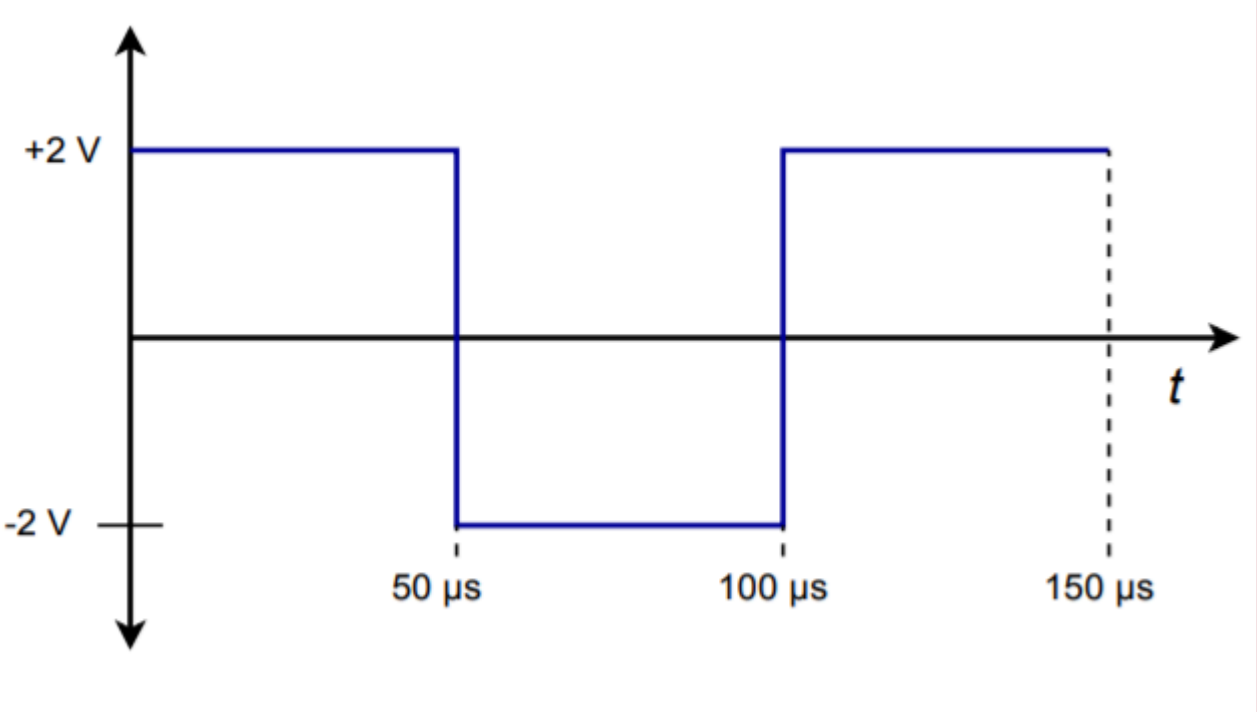


Figure 10.2.9 : Input waveform for Example 10.2.3 .

The first step is to break down the input waveform into simple-to-integrate components and ascertain the basic shape of the result. The input is sketched in Figure 10.2.9 . The input may be broken into two parts: a non-changing 2 V potential from 0 to 50 ◈ s, and a non-changing –2 V potential from 50 ◈ s to 100 ◈ s. The sequence repeats after this.

$$V_{in}(t) = 2 \text{ from } t = 0, \text{ to } t = 50\mu s$$

[V_{in}(t)= – 2 \text{ from } t=50 \mu s, \text{ to } t=100 \mu s \]

In essence, we are saying that the input may be treated as either ± 2 V DC for short periods of time. Because integration is a summing operation, as time progresses, the area swept out from underneath the input signal increases, and then decreases due to the polarity changes. It follows that the output will grow (or shrink) in a linear fashion, as the input is constant. In other words, straight ramps will be produced during these time periods. The only difference between them will be the polarity of the slope. From this we find that the expected output waveform is a triangle wave. All we need to do now is determine the peak value of the output. To do this, perform the definite integral using the first portion of the input wave.

$$V_{out}(t) = -\frac{1}{R_i C}\int V_{in}(t)dt$$

$$V_{out}(t) = - \frac{1}{10 \text{ k} \times 10 \text{ nF}} \int_{0}^{50\mu s} 2 \, dt$$

$$V_{out}(t) = -10^4 \times 2 \times \left. t\right|_{t=0}^{t=50 \mu s}$$

$$V_{out}=-20000 \times 50 \mu s$$

$$V_{out}=-1V$$

This represents the total change over the 50 ◈ s half-cycle interval. This is a peak-to-peak change, saying that the output is 1 V negative relative to its value at the end of the preceding half-cycle. Integration for the second half cycle is similar, and produces a positive change. The resulting waveform is shown in Figure 10.2.10 .

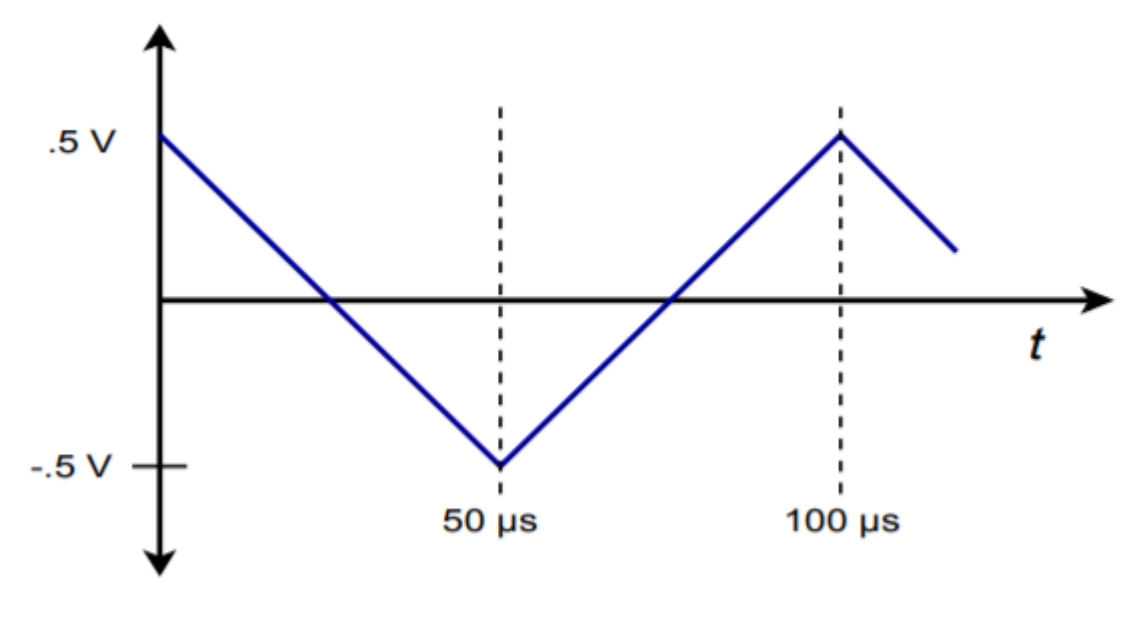


Figure 10.2.10 : Output waveform for Example 10.2.3 .

Note that the wave is effectively inverted. For positive inputs, the output slope is negative. Because the input frequency is much higher than ◈◈◈◈ , we can once again expect high accuracy. Although the solution seems to imply that the output voltage should swing between 0 and –1 V, a real world integrator does indeed produce the indicated ± 0.5 V swing. This is due to the fact that the integration capacitor in this circuit will not be able to maintain the required DC offset indefinitely.

Example 10.2.4

Figure 10.2.11◈ shows an integrator connected to an accelerometer. This device produces a voltage that is proportional to the acceleration it experiences.[2] Accelerometers may be fastened to a variety of physical devices in order to determine how the devices respond to various mechanical inputs. This would be useful, for example, in experimentally determining the mechanical resonance characteristics of a surface. Another

2. A mechanical accelerometer consists of a small mass, associated restoring springs, and some form of transducer that is capable of reading the mass's motion. More recent designs may be fashioned using micro-machined chips, where the deflection of a mass-supporting beam is measured indirectly.

possibility is the determination of the lateral acceleration of an automobile as it travels through a corner. By integrating this signal, it is possible to determine velocity, and a further integration will produce position. If the accelerometer produces the voltage shown in Figure 10.2.11 , determine the shape of the velocity curve (assume a non-repetitive input wave).

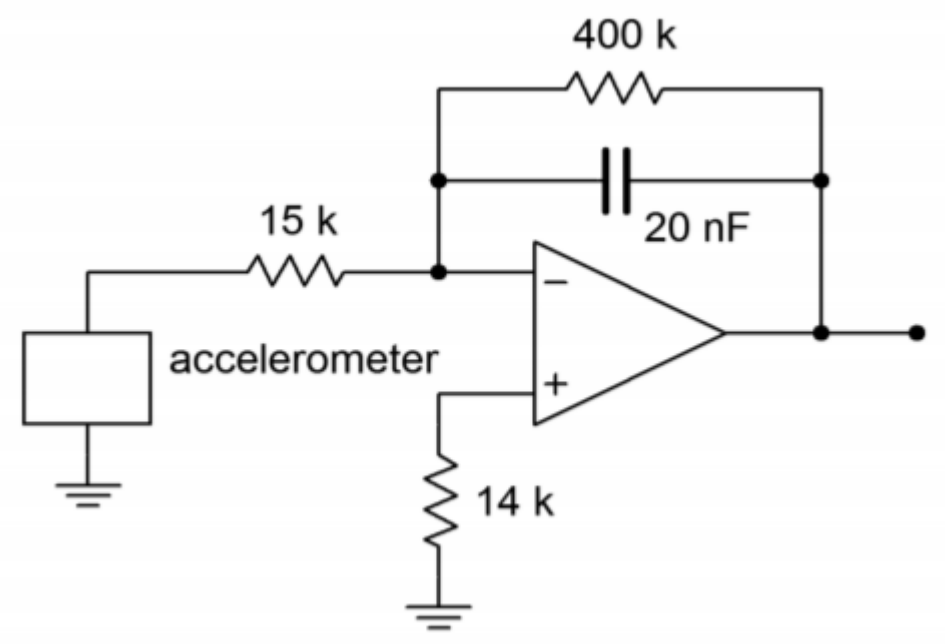


Figure 10.2.11 : Accelerometer with integrator circuit.

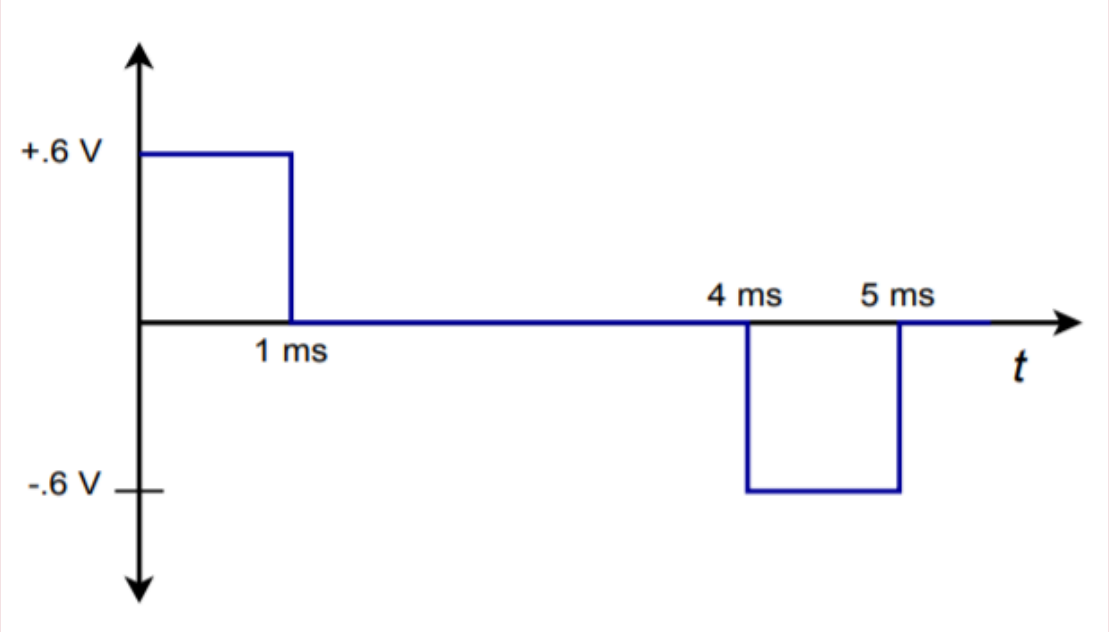


Figure 10.2.11 : Signal produced by accelerometer.

First, check the frequency limit of the integrator in order to see if high accuracy may be maintained. If the input waveform was repetitive, it would be approximately 200 Hz (1/ 5 ms).

$$f_{low} = \frac{1}{2\pi R_f C}$$

[f_{low} = \frac{1}{2 \pi 400 \text{ k} \times 20 \text{ nF}} \]

[f_{low} = 19.9 \text{ Hz}\]

The input signal is well above the lower limit. Note that the input waveshape may be analyzed in piecewise fashion, as if it were a square wave. The positive and negative portions are both 1 ms in duration, and 0.6 V peak. The only difference is the polarity. As these are square pulses, we expect ramp sections for the output. For the positive pulse (◈=0 to ◈=1 ms),

$$V_{out}(t) = -\frac{1}{R_i C}\int V_{in}(t)dt$$

[V_{out}(t) = – \frac{1}{15 \text{ k} \times 20 \text{ nF}} \int_{0}^{10^{-3}} .6 dt \]

[V_{out}(t) = -3333 \times 0.6 \times \left. t\right|_{t=0}^{t=10^{-3}} \]

[V_{out} = -2000 \times 10^{–3} \\ V_{out} = -2 V \]

This tells us that we will see a negative going ramp with a –2 V change. For the time period between 1 ms and 4 ms, the input signal is zero, and thus the change in output potential will be zero. This means that the output will remain at –2 V until ◈=4 ms.

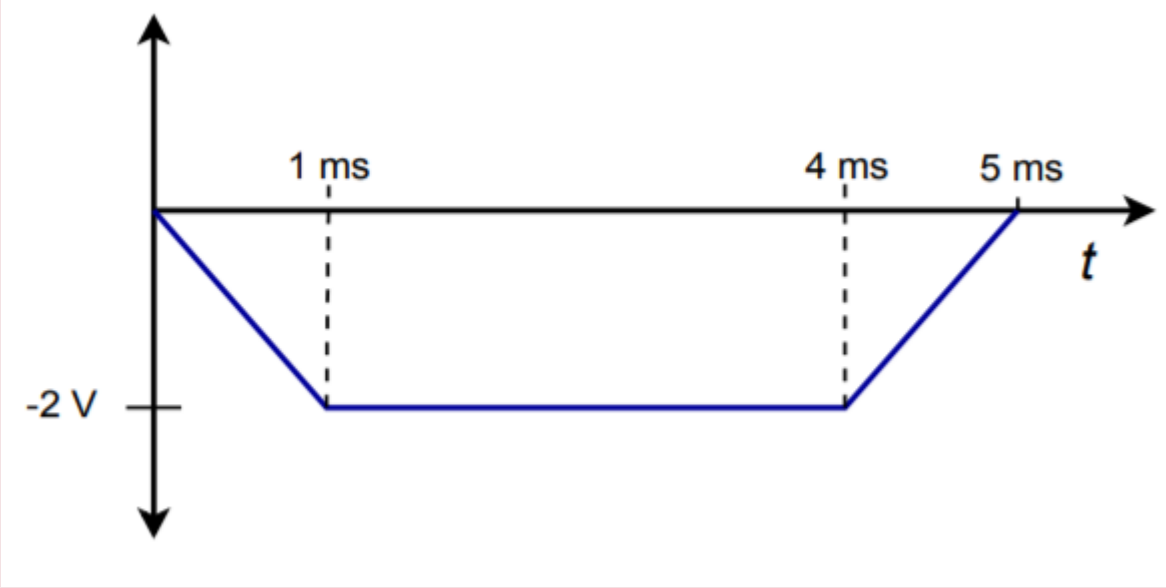


Figure 10.2.12 : Integrator output.

For the period between 4 ms and 5 ms, a positive going ramp will be produced. Because only the polarity is changed relative to the first pulse, we can quickly find that the change is +2 V. The input waveform is nonrepetitive, so the output waveform appears as shown in Figure 10.2.12 . (A repetitive input would naturally cause the integrator to “settle” around ground over time, producing the same basic waveshape, but shifted positively.) This waveform tells us that the velocity of the device under test increases linearly up to a point. After this point, the velocity remains constant until a linear decrease in velocity occurs. (Remember, the integrator inverts the signal, so the output curve is effectively upside down. If the input wave indicates an initial positive acceleration, the actual output is an initial positive velocity.) On a different

time scale, these are the sort of waveforms that might be produced by an accelerometer mounted on an automobile that starts at rest, smoothly climbs to a fixed speed, and then smoothly brakes to a stop.

14.3 DIFFERENTIATORS

Differentiators perform the complementary function to the integrator. The base form of the differentiator is shown in Figure 10.3.1 . The output voltage is the differential of the input voltage. This is very useful for finding the rate at which a signal varies over time. For example, it is possible to find velocity given distance and acceleration given velocity. This can be very useful in process control work.

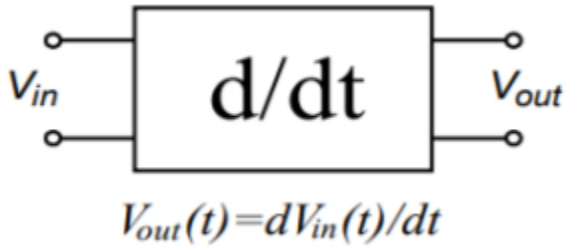


Figure 10.3.1 : A basic differentiator

Essentially, the differentiator tends to reinforce fast signal transitions. If the input waveform is non-changing, (i.e., DC), the slope is zero, and thus the output of the differentiator is zero. On the other hand, an abrupt signal change such as the rising edge of a square wave produces a very large slope, and thus the output of the differentiator will be large. In order to create the differentiation, an appropriate device needs to be associated with the op amp circuit. This was the approach taken with the integrator, and it remains valid here. In fact, we are left with the same two options: using either an inductor, or a capacitor. Again, capacitors tend to be somewhat easier to work with than inductors and are preferred. The only difference between the integrator and the differentiator is the position of the capacitor. Instead of placing it in the ⍰⍰ position, the capacitor will be placed in the ⍰⍰ position. The resulting circuit is shown in Figure 10.3.2 .

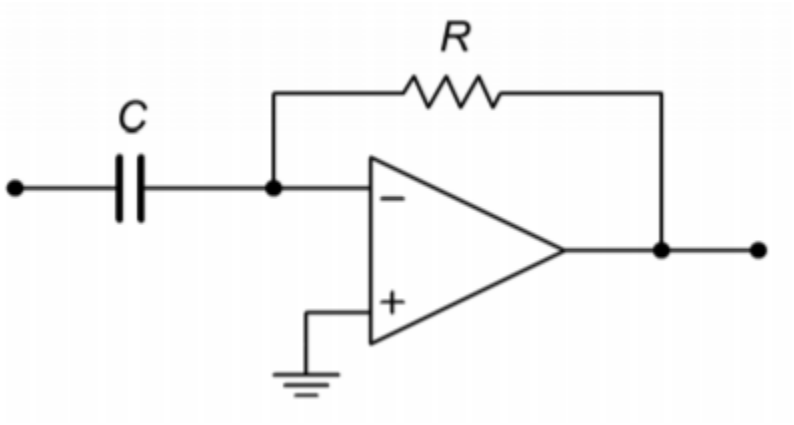


Figure 10.3.2 : A simple op amp differentiator.

The analysis starts with the basic capacitor Equation (Equation 10.2.1):

$$i(t) = C\frac{dv(t)}{dt}$$

We already know from previous work that the output voltage appears across ⍰⍰ , though inverted.

$$V_{out} = -V_{Rf}$$

Also, by Ohm's law,

$$V_{Rf} = iR_f$$

By using the approximation that all input current flows through ⍰⍰ (since the op amp's input current is zero), and then substituting Equation 10.2.1 for the current, we find

$$V_{out}(t) = -R_f C\frac{dv(t)}{dt}$$

A quick inspection of the circuit shows that all of the input voltage drops across the capacitor, because the op amp's inverting input is a virtual ground. Bearing this in mind, we arrive at the final output voltage equation,

$$V_{out}(t) = -R_f C\frac{dV_{in}(t)}{dt}$$

(10.3.1)

As with the integrator, a leading constant is added to the fundamental form. Again, it is possible to scale the output as required through the use of gain or attenuation networks.

ACCURACY AND USEFULNESS OF DIFFERENTIATION

Equation 10.3.1 is an accurate reflection of the circuit response so long as the base assumptions remain valid. As with the integrator, practical considerations tend to force limits on the circuit's operating range. If the circuit is analyzed at discrete points in the frequency domain, it can be modeled as an inverting amplifier with the following gain equation:

$$A_v = -\frac{R_f}{X_C} = 2\pi f C R_f$$

Note that as the frequency decreases, X_C grows, thus reducing the gain. Conversely, as the input frequency is raised, X_C falls in value, causing the gain to rise. This rise will continue until it intersects the open-loop response of the op amp. The resulting amplitude response is shown in Figure 10.3.3 .

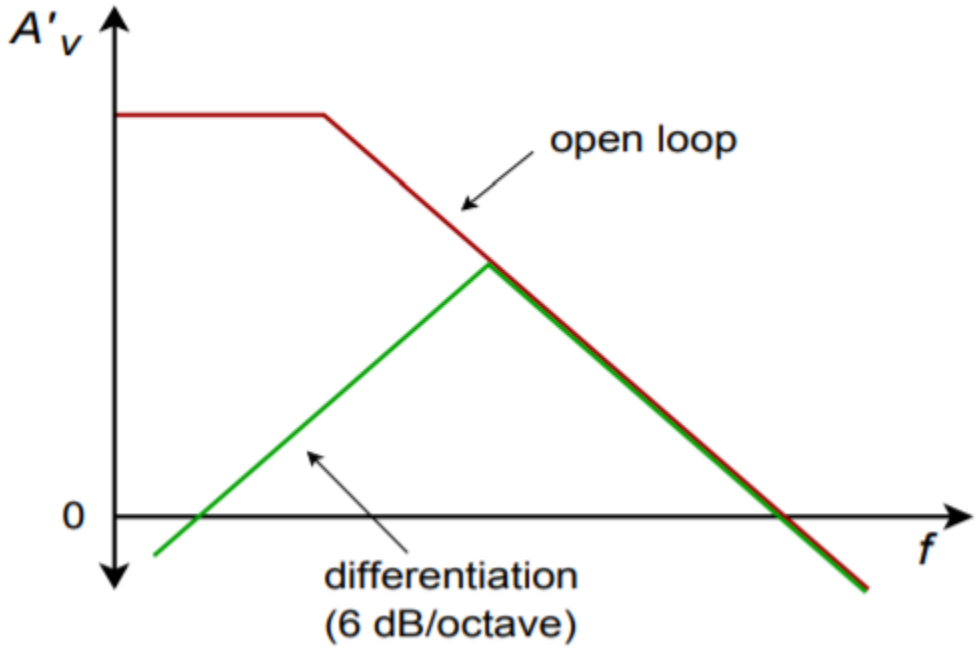


Figure 10.3.3 : Response of a simple differentiator.

This response is the mirror image of the basic integrator response and exhibits a slope of 6 dB per octave. Note that DC gain is zero, and therefore the problems created by input bias and offset currents are not nearly as troublesome as in the integrator. Because of this, there is no limit as to how low the input frequency may be, excluding the effects of signal-to-noise ratio. Things are considerably different on the high end though. Once the circuit response breaks away from the ideal 6 dB per octave slope, differentiation no longer takes place.

OPTIMIZING THE DIFFERENTIATOR

There are a couple of problems with the general differentiator of Figure 10.3.2 . First of all, it is quite possible that the circuit may become unstable at higher frequencies. Also, the basic shape of the amplitude response suggests that high frequencies are accentuated, thus increasing the relative noise level. Both of these problems may be reduced by providing an artificial upper-limit frequency, ◇*h*◇◇*h* . This tailoring may be achieved by shunting ◇◇ with a small capacitor. This reduces the high frequency gain, and thus reduces the noise. The resulting response is shown in Figure 10.3.4 .

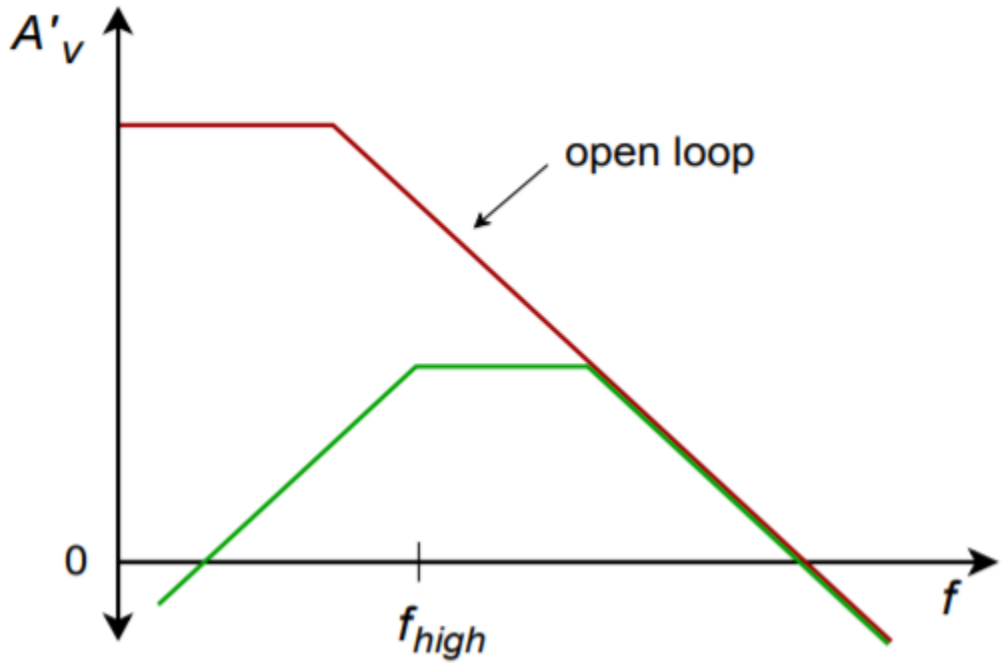


Figure 10.3.4 : Response of a partially optimized differentiator.

We find ◈h◈◈h in the standard manner:

$$f_{high(fdbk)} = \frac{1}{2\pi R_f C_f}$$

(10.3.2)

◈h◈◈h represents the highest frequency for differentiation. It is the 50% accuracy point. For higher accuracy, the input frequency must be kept well below ◈h◈◈h . At about 0.1 ◈h◈◈h , the accuracy of Equation 10.3.1 is about 99%. Generally, you have to be somewhat more conservative in the estimation of accuracy than with the integrator. This is because complex waves contain harmonics that are higher than the fundamental. Even though the fundamental may be well within the high accuracy range, the upper harmonics may not be.

The other major problem of the basic circuit is that the input impedance is inversely proportional to the input frequency. This is because ◈◈ is the sole input impedance factor. This may present a problem at higher frequencies because the impedance will approach zero. To circumvent this problem, a resistor may be placed in series with the input capacitor in order to establish a minimum impedance value. Unfortunately, this will also create an upper break frequency, ◈h◈◈h .

$$f_{high(in)} = \frac{1}{2\pi R_f C}$$

(10.3.3)

The resulting response is shown in Figure 10.3.5 . The effective ◈h◈◈h for the system will be the lower of Equations 10.3.2 and 10.3.3 .

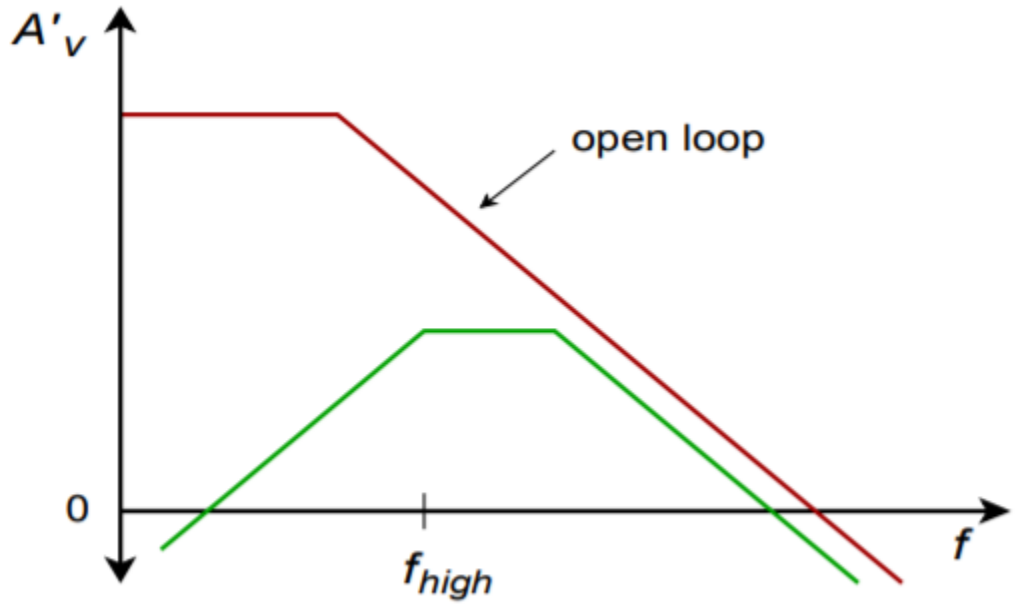


Figure 10.3.7 : Response of a practical differentiator.

The completed, practical differentiator is shown in Figure 10.3.8 . Note that a bias compensation resistor may be required at the noninverting input, although it is not shown.

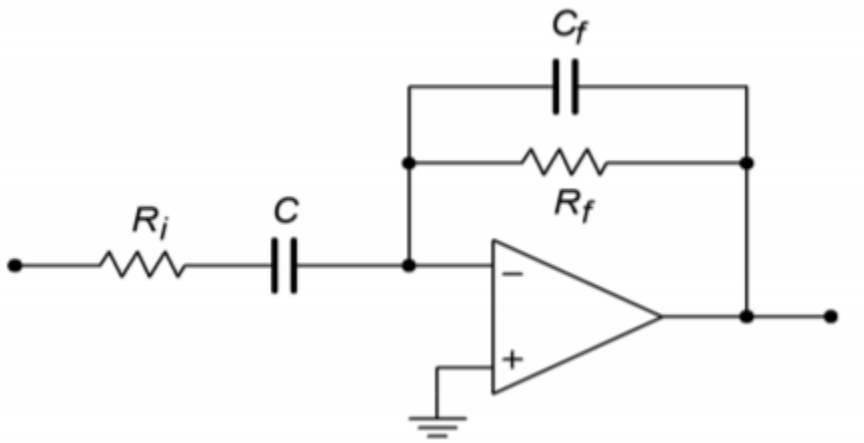


Figure 10.3.8 : A practical differentiator.

ANALYZING DIFFERENTIATORS WITH THE TIME-CONTINUOUS METHOD

The time-continuous method will be used when the input signal may be easily written in the time-domain (e.g., sine waves). For more complex waveforms, such as a triangle wave, a discrete time method will be used. The continuous method will lead directly to a time-domain representation of the output waveform. Specific voltage/time coordinates will not be evaluated.

Example 10.3.1

Determine the useful range for differentiation in the circuit of Figure 10.3.9 . Also determine the output voltage if the input signal is a 2 V peak sine wave at 3 kHz.

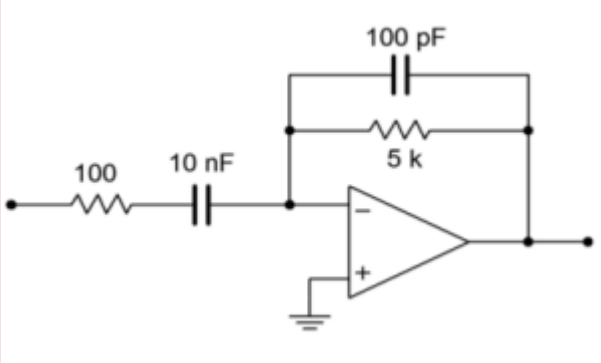


Figure 10.3.9 : Differentiator for Example 10.3.1 .

The upper limit of the useful frequency range will be determined by the lower of the two ◈◈ networks.

$$f_{high(fdbk)} = \frac{1}{2\pi R_f C_f}$$

[f_{high (fdbk)} = \frac{1}{2 \pi \times 5 k \times 100 pF} \]

[f_{high (fdbk)} = 318.3 kHz \\ f_{high (in)} = \frac{1}{2 \pi R_f C} \]

[f_{high (in)} = \frac{1}{2 \pi \times 100 \times 10 nF} \]

[f_{high (in)} = 159.2 kHz \]

Therefore, the upper limit is 159.2 kHz. Remember, the accuracy at this limit is relatively low, and normal operation will typically be several octaves lower than this limit. Note that the input frequency is 3 kHz, so high accuracy should result. First, write ◈◈◈as a time-domain expression:

$$V_{in}(t) = 2 \sin 2\pi 3000t$$

[V_{out}(t) = –R_f C \frac{dV_{in} (t)}{dt} \]

[V_{out}(t) = -5 k \times 10 nF \frac{d 2 \sin 2 \pi 3000t}{dt} \]

[V_{out}(t) = -10^{–4} \frac{d \sin 2 \pi 3000t}{dt} \]

[V_{out}(t) = -1.885 \cos 2 \pi 3000t \]

This tells us that the output waveform is also sinusoidal, but it lags the input by 90 ° . Note that the input frequency has not changed, but the amplitude has. The differentiator operates with a 6 dB per octave slope, thus it can be seen that the output amplitude is directly proportional to the input frequency. If this example is rerun with a frequency of 6 kHz, the output amplitude will be double the present value.

COMPUTER SIMULATION

The Multisim simulation for the circuit of Example 10.3.1 is shown in Figure 10.3.10 . Note the excellent correlation for both the phase and amplitude of the output. As was the case with the integrator simulation, the Transient Analysis output plot is started after the initial conditions have settled.

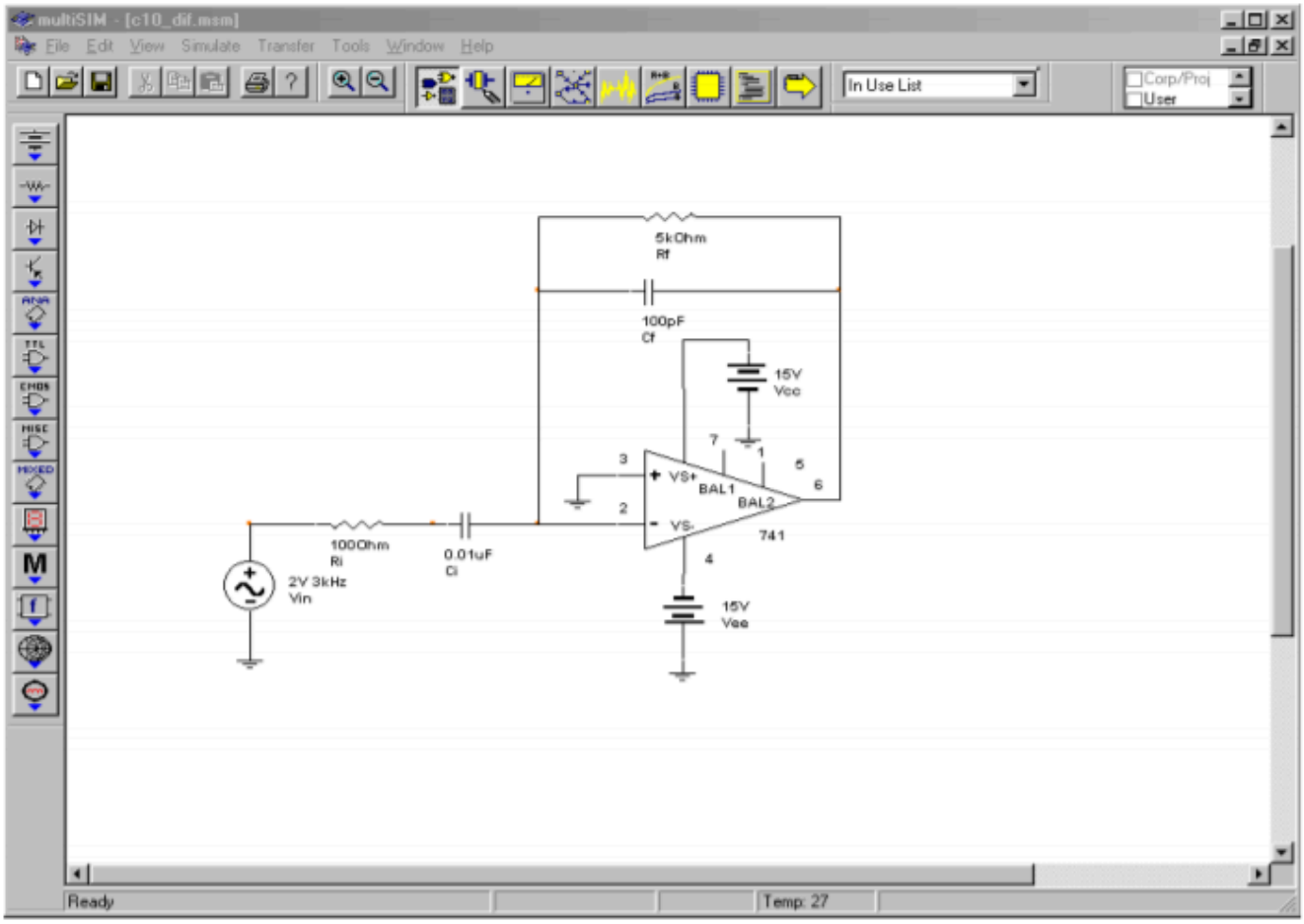


Figure 10.3.10 : Differentiator in Multisim.

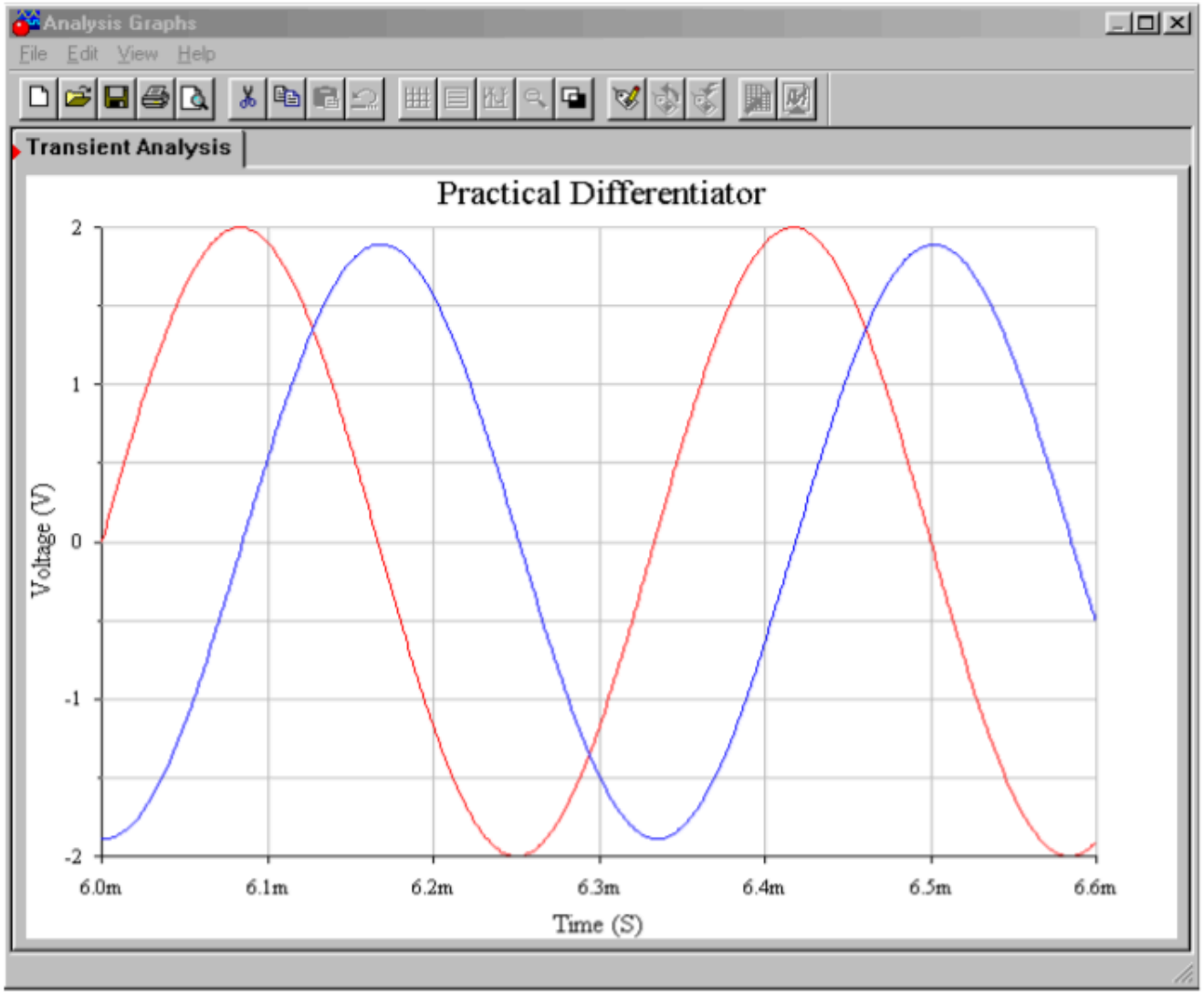


Figure 10.3.10 : Differentiator input and output waveforms.

ANALYZING DIFFERENTIATORS WITH THE TIME-DISCRETE METHOD

For more complex waveforms, it is sometimes expedient to break the waveform into discrete chunks, differentiate each portion, and then combine the results. The idea is to break the waveform into equivalent straight-line segments. Differentiation of a straight-line segment will result in a constant (i.e., the slope, which does not change over that time). The process is repeated until one cycle of the input waveform is completed. The resultant levels are then joined together graphically to produce the output waveform.

Often, waveforms are symmetrical, and only part of the calculation need be performed: a sign change is all that will be needed for the mirror image portions. As an example, a triangle wave may be broken into a positive-going line segment and a negative going-line segment. The slopes should be equal, only the direction (i.e., sign) has changed. A square wave may be broken into four parts: a positive-going edge, a static positive value, a negative-going edge, and a static negative value. The "flat" portions have a slope of zero, so only one calculation must be performed, and that's the positive-going edge. We shall take a look at both of these waveforms in the next two examples.

Example 10.3.2

Sketch the output waveform for the circuit of Figure 10.3.9 if the input is a 3 volt peak triangle wave at 4 kHz.

First, note that the input frequency is well within the useful range of this circuit, as calculated in Example 10.3.1 . (Note that the highest harmonics will still be out of range, but the error introduced will be minor.)

The triangle wave may be broken into a positive-going portion and a negative-going portion. In either case, the total voltage change will be 6 V in one half-cycle. The period of the waveform is

$$T = \frac{1}{4kHz}$$

$$T = 250\mu s$$

Therefore, for the positive going portion, a 6 volt change will be seen in 125 ◈ s (–6 V in 125 ◈ s for the negative going portion). The slope is

$$Slope = \frac{6V}{125\mu s}$$

$$Slope = 48000V/s$$

Which, as a time-domain expression is

$$V_{in}(t) = 48000t$$

Substituting this Equation into Equation 10.3.1 yields

$$V_{out}(t) = -R_f C \frac{dV_{in}(t)}{dt}$$

[V_{out}(t) = -5 k \times 10 nF \frac{d 48000 t}{dt} \]

[V_{out}(t) = -2.4 V \]

During ◈=0 through ◈=125◈ s, the output is –2.4 V. Differentiation of the second half of the wave is similar, but produces a positive output, +2.4 V. The result is a 4 kHz square wave that is 2.4 V peak. The resulting waveform is shown in Figure 10.3.11 .

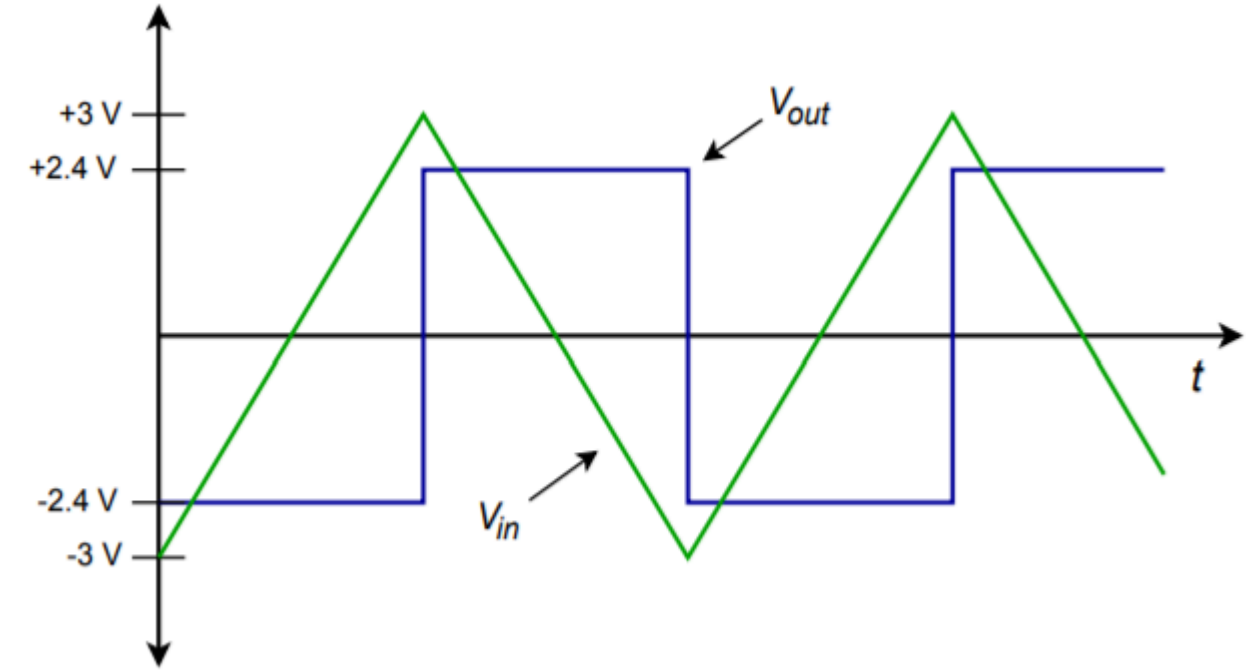


Figure 10.3.11 : Input/output waveforms.

COMPUTER SIMULATION

Some circuits are more sensitive to the choice of op amp than others are, and the effects of an improper choice may not always be immediately apparent without first building or simulating the circuit. A good example of this is shown in Figure 10.3.12 . Multisim was used to create the Transient Analysis for the circuit of Example 10.3.2 with two different op amps. Accurate differentiation requires excellent high frequency response from the op amp. In the first simulation, a 741 op amp is used. This device is not particularly fast. As a result, the output waveform suffers from excessive overshoot and ringing. Also, slew rate limiting is fairly obvious, slowing the transitions of the output waveform. In contrast, using an LF411 in the same circuit yields far superior response. Some overshoot still exists, but its magnitude has been restrained, as has the ringing. Also, slew rate limiting is reduced by a wide margin. Clearly, the second result is much closer to the ideal calculation than the first run.

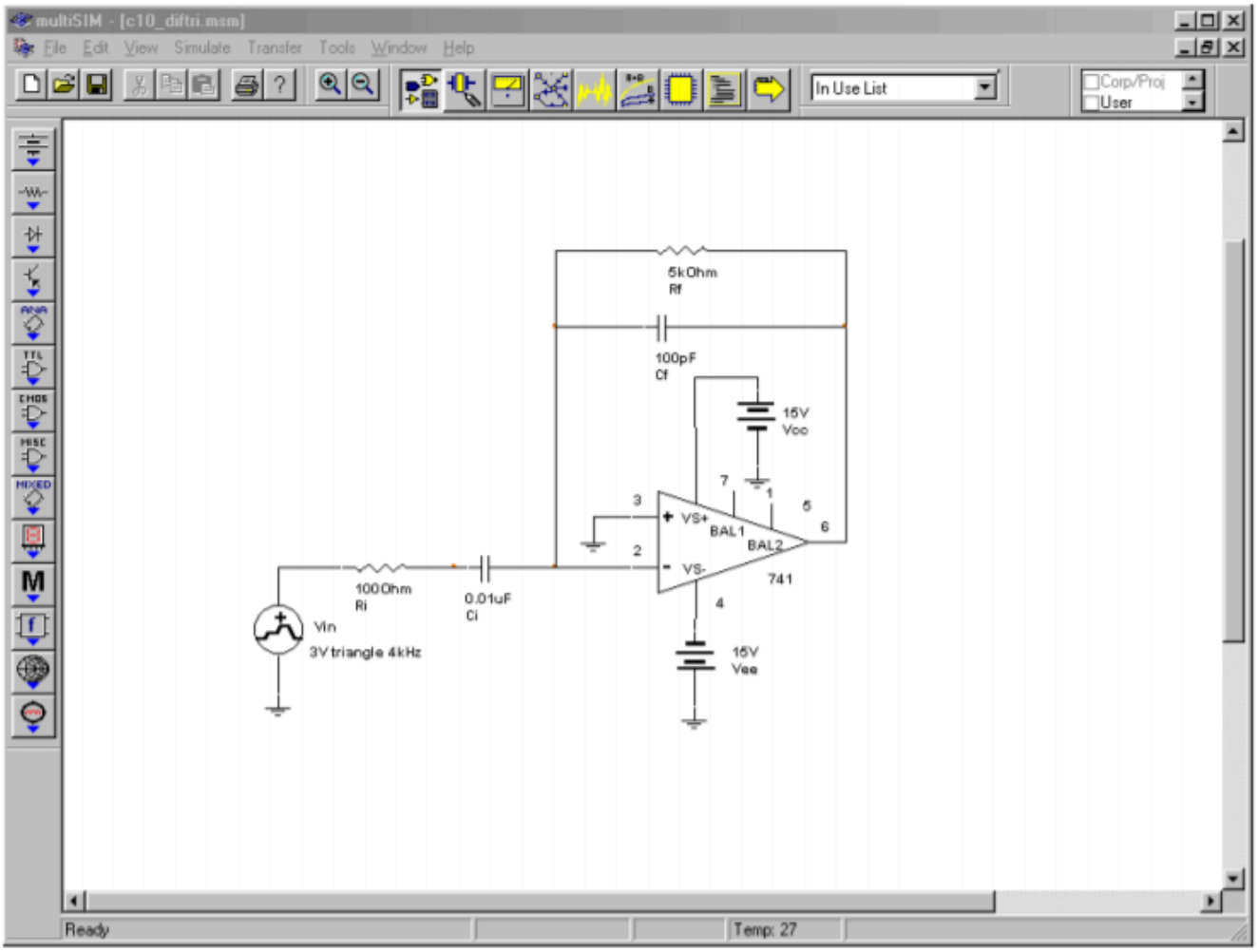


Figure 10.3.12 : Multisim schematic of differentiator.

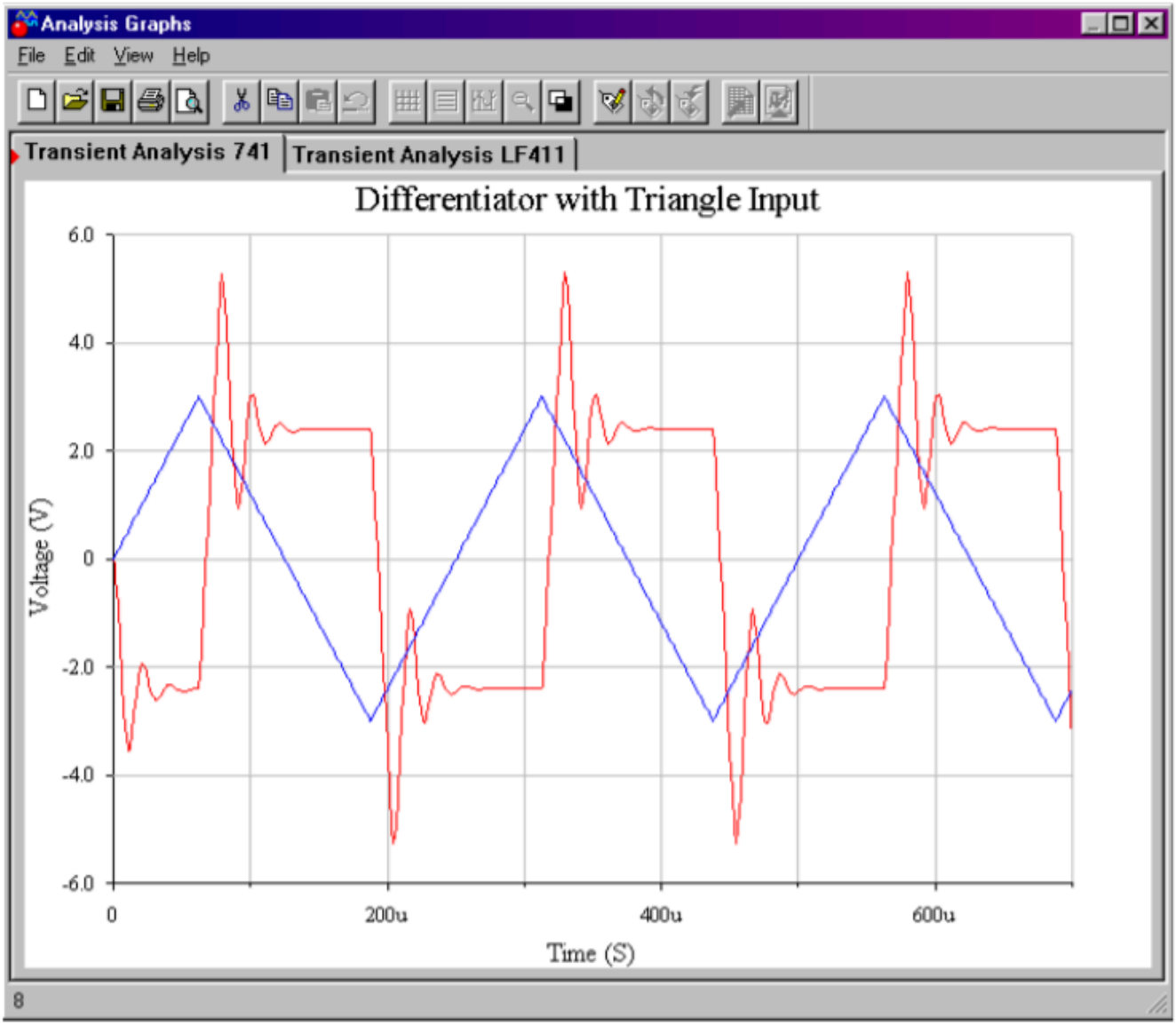


Figure 10.3.12◈: Input and output waveforms using 741.

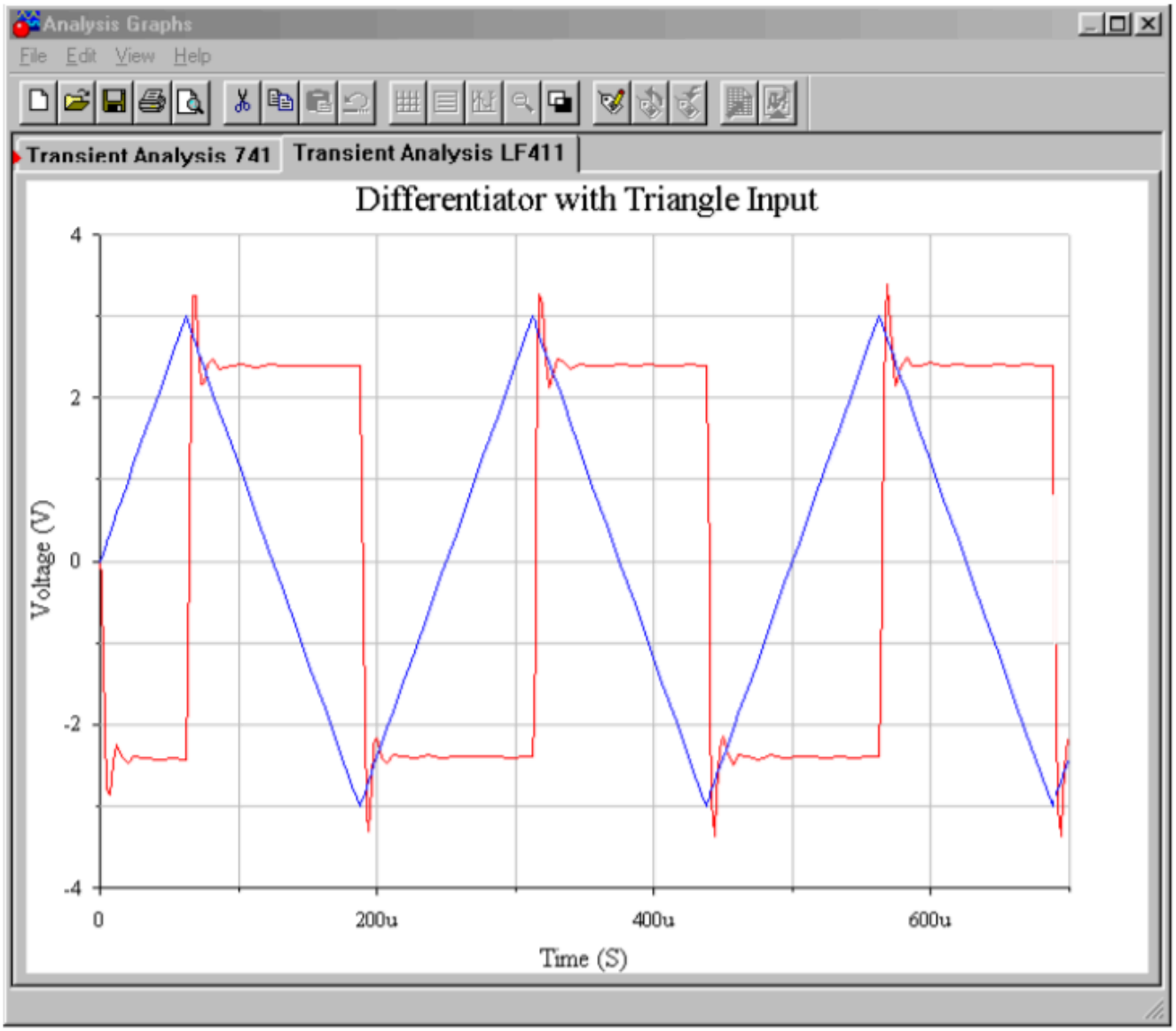


Figure 10.3.12 : Input and output waveforms using LF411.

Example 10.3.3

Repeat Example 10.3.2 with a 3 V peak, 4 kHz square as the input. Assume that the rising and falling edges of the square wave have been slew limited to 5 V/ s.

During the time periods that the input is at ±3 V, the output will be zero. This is because the input slope is, by definition, zero when the signal is "flat". An output is only noted during the transitions between the ±3 V levels. Therefore, we need to find the slope of the transitions. It was stated that due to slew rate limiting (perhaps from some previous amplifier stage) the transitions run at 5 V/ s. As a time-domain expression, this is

$$V_{in}(t) = 5 \times 10^6 t$$

Substituting this Equation into Equation 10.3.1 yields,

$$V_{out}(t) = -R_f C \frac{dV_{in}(t)}{dt}$$

$$V_{out}(t) = -5k10nF \frac{d5 \times 10^6 t}{dt}$$

$$V_{out}(t) = -250V$$

Obviously, when using a standard op amp and ±15 V power supply, clipping will occur in the vicinity of –13.5 V. For the negative going edge, a similar result will be seen (+250 V calculated, with clipping at +13.5 V). The resulting waveform is shown in Figure 10.3.13 . Note that the output waveform spikes will also be limited by the slew rate of the differentiator's op amp.

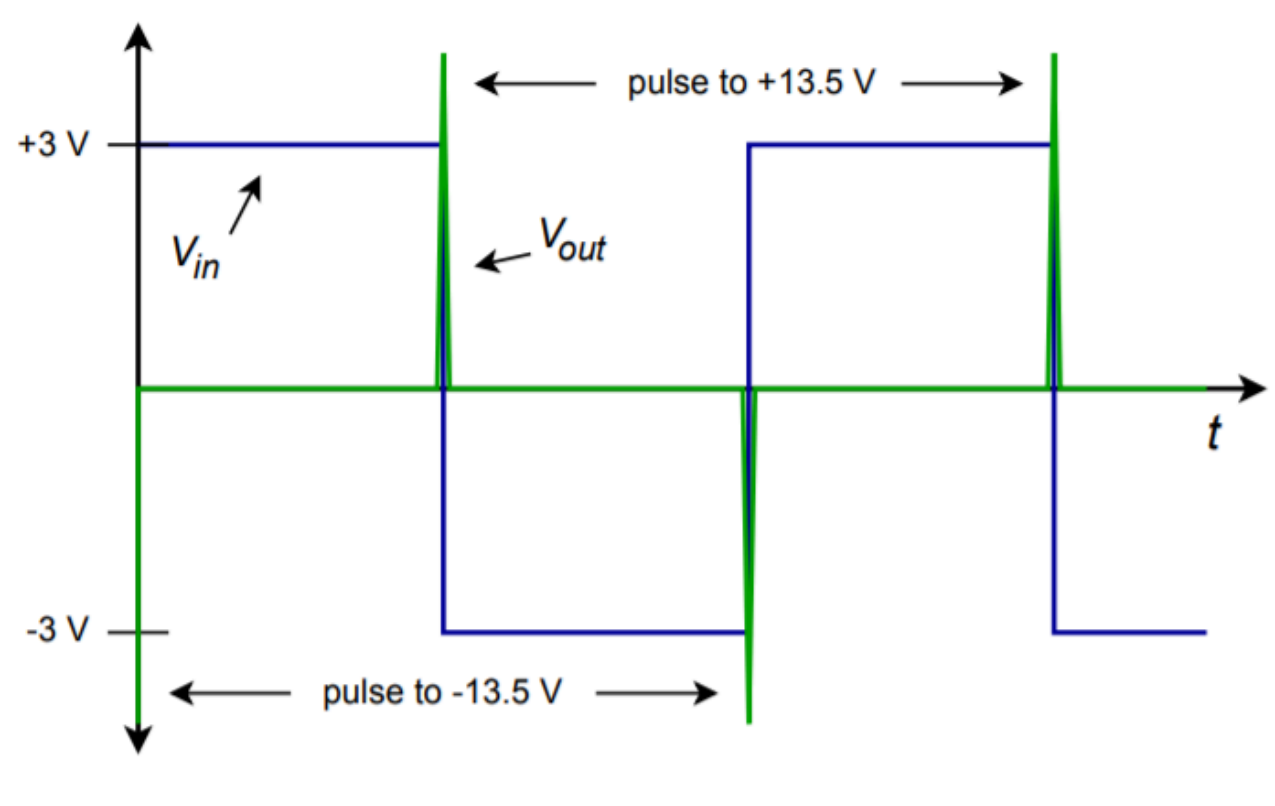


Figure 10.3.13 : Differentiated square wave (note output clipping).

Example 10.3.4

Figure 10.3.14 shows a differentiator receiving a signal from an LVDT, or linear variable differential transformer.[1] An LVDT can be used to accurately measure the position of objects with displacements of

1. An LVDT is a transformer with dual secondary windings and a movable core. The core is connected to a shaft, that is in turn actuated by some external object. The movement of the core alters the mutual inductance between primary and

less than one-thousandth of an inch. This could be useful in a computer-aided manufacturing system. By differentiating this position signal, a velocity signal may be derived. A second differentiation will produce acceleration. If the LVDT produces the wave shown in Figure 10.3.14 , determine the velocity/time curve for the object being tracked.

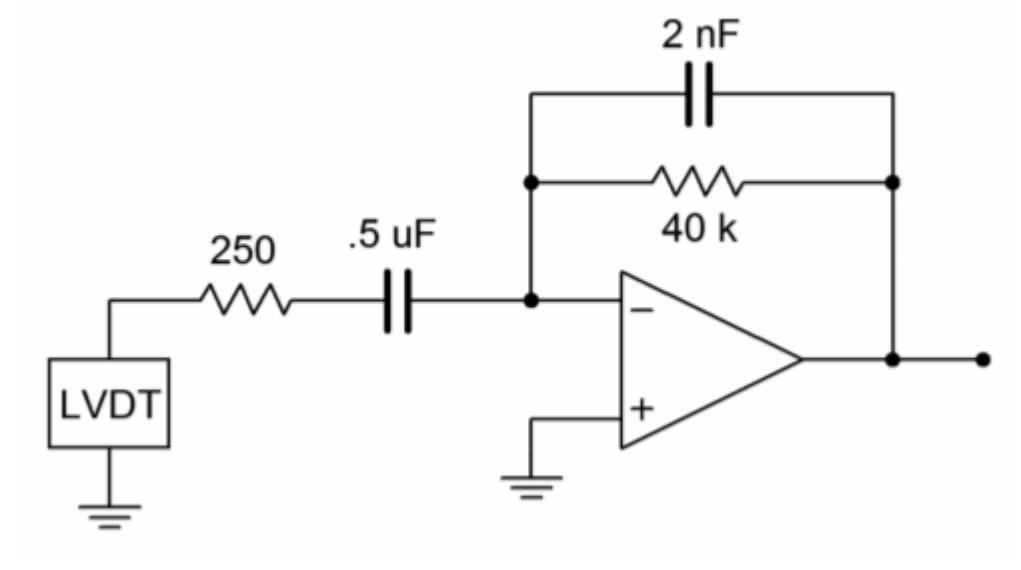


Figure 10.3.14 : Differentiator circuit with LVDT.

First, check the upper frequency limit for the circuit.

$$f_{high(fdbk)} = \frac{1}{2\pi R_f C_f}$$

$$f_{high(fdbk)} = \frac{1}{2\pi \times 40k \times 2nF}$$

$$f_{high(fdbk)} = 1.99kHz$$

$$f_{high(in)} = \frac{1}{2\pi R_i C}$$

$$f_{high(in)} = \frac{1}{2\pi \times 250 \times 500nF}$$

$$f_{high(in)} = 1.273kHz$$

The limit will be the lower of the two, or 1.273 kHz. This is well above the slowly changing input signal, and therefore, high accuracy should be possible.

secondary. A carrier signal is fed into the primary, and the changing mutual inductance alters the strength of the signal induced into the secondaries. This signal change is turned into a simple DC voltage by a demodulator. The resulting DC potential is proportional to the position of the core, and thus, proportional to the position of the object under measurement.

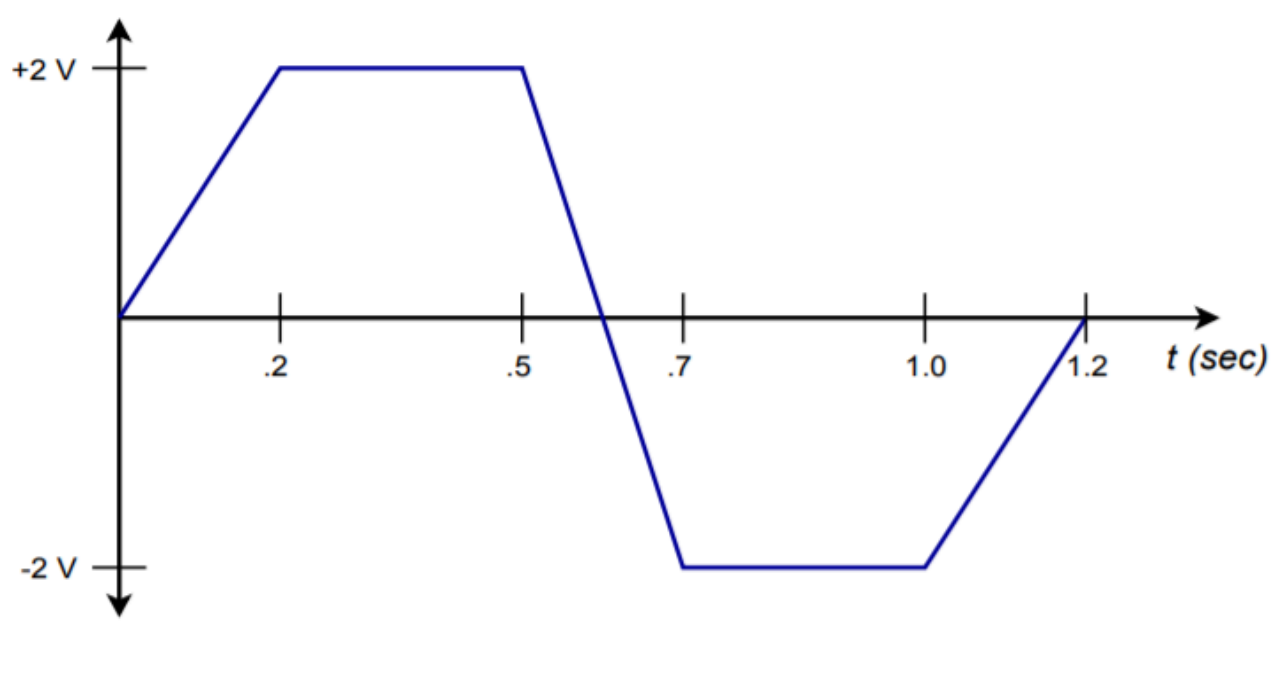


Figure 10.3.14 : Signal produced by LVDT.

This wave can be analyzed in piece-wise fashion. The ramp portions will produce constant output levels and the flat portions will produce an output of 0 V (i.e., the rate of change is zero).

For the first section,

$$Slope = \frac{2V}{0.2s}$$

$$Slope = 10V/s$$

Which, as a time-domain expression, is

$$V_{in}(t) = 10t$$

$$V_{out}(t) = -R_f C \frac{dV_{in}(t)}{dt}$$

$$V_{out}(t) = -40k \times 500nF \frac{d10t}{dt}$$

$$V_{out}(t) = -0.2V$$

So, during =0 through =0.2 s, the output is –0.2 V. The time period between 0.2 s and 0.5 s will produce an output of 0 V. For the negative-going portion,

$$Slope = \frac{4V}{0.2s}$$

$$Slope = 20V/s \quad \text{thus,}$$

$$V_{in}(t) = 20t$$

$$V_{out}(t) = R_f C \frac{dV_{in}(t)}{dt}$$

$$V_{out}(t) = 40k500nF \frac{d20t}{dt}$$

$$V_{out}(t) = 0.4V$$

The output is 0.4 V between 0.5 s and 0.7 s. The third section has the same slope as the first section, and will also produce a –0.2 V level. The output waveform is drawn in Figure 10.3.15 .

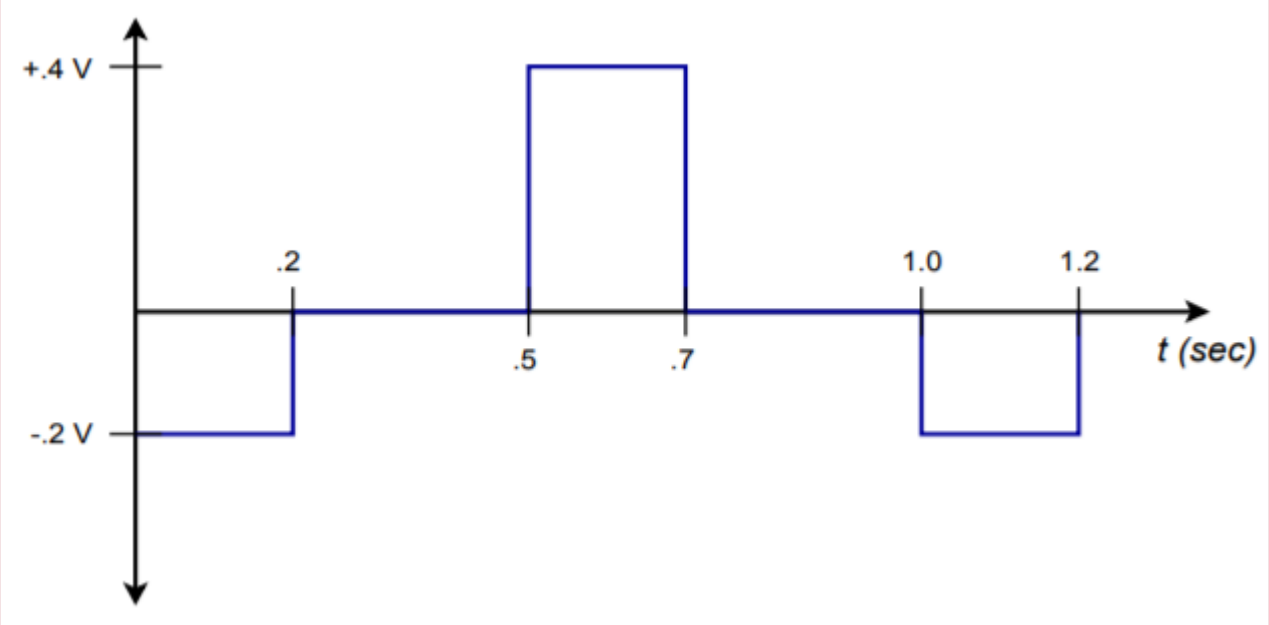


Figure 10.3.15 : Differentiator output.

14.4 ANALOG COMPUTER

Analog computers are used to simulate physical systems. These systems may be electrical, mechanical, acoustical, or what have you. An analog computer is basically a collection of integrators, differentiators, summers, and amplifiers. Due to their relative stability, integrators are favored over differentiators. It is not uncommon for analog computers to be made without any differentiators. Because physical systems may be described in terms of differential equations, analog computers may be used to solve these equations, thus producing as output some system parameter.

The basic advantage of simulation is that several variations of a given system may be examined in real time without actually constructing the system. For a large project this is particularly cost-efficient. The process starts by writing a differential Equation (first-, second-, or third-order) that describes the system in question. The Equation is then solved for its highest-order element, and the result used to create a circuit.

Example 10.4.1

Let's investigate the system shown in Figure 10.4.1 . This is a simple mechanical system that might represent (to a rough approximation) a variety of physical entities, including the suspension of an automobile.

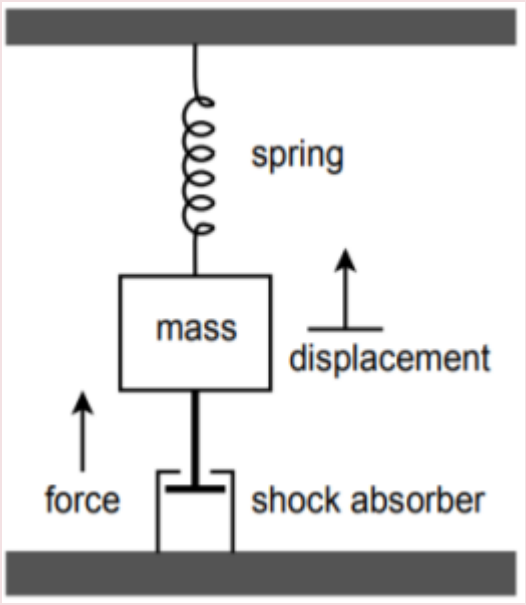


Figure 10.4.1 : Mechanical system.

This system is comprised of a body with mass ⍰ , that is suspended from a spring. The spring has a spring constant, ⍰ . The mass is also connected to a shock absorber that produces damping, ⍰ . If an external force, ⍰ , excites the mass, it will move, producing some displacement, ⍰ . This displacement

depends on the mass, force, spring constant, and damping. Essentially, the spring and shock absorber will create reactionary forces. From basic physics, ◈=◈◈ , where ◈ is the acceleration of the body. If ◈ is the position of the body, then ◈◈/◈◈ is its velocity, and ◈2◈/◈◈2 is its acceleration. Therefore, we can say

$$F = M\frac{d^2X}{dt^2}$$

In this system, the total force is comprised of the excitation force ◈ , and the forces produced by the spring and shock absorber.

$$F - F_{spring} - F_{shock} = M\frac{d^2X}{dt^2}$$

The spring's force is equal to the displacement times the spring constant:

$$F_{spring} = KX$$

The shock absorber's force is equal to the damping constant times the velocity of the body:

$$F_{shock} = R\frac{dX}{dt}$$

By substituting and rearranging the above elements we find that

$$F = M\frac{d^2X}{dt^2} + R\frac{dX}{dt} + KX$$

Here ◈ is seen as the input signal, and ◈ as the output signal. A somewhat less busy notation form is the dot convention. A single dot represents the first derivative with respect to time, two dots represent the second derivative, and so on. The above Equation may be rewritten as

$$F = M\ddot{X} + R\dot{X} + KX$$

This is the final differential equation. Note how it contains only derivatives and no integrals. The last step is to solve the Equation for the highest-order differential. By setting it up in this form, the simulation circuit may be realized without using differentiators. This will indicate how many integrators will be required.

$$\ddot{X} = \frac{F}{M} - \frac{R}{M}\dot{X} - \frac{K}{M}X$$

This says that the second differential of ◈ is the sum of three components. To realize the circuit, start with a summing amplifier with the three desired signals as inputs. This is shown in block form in Figure 10.4.2 . Note that two of the inputs use ◈ and the first derivative of ◈ .

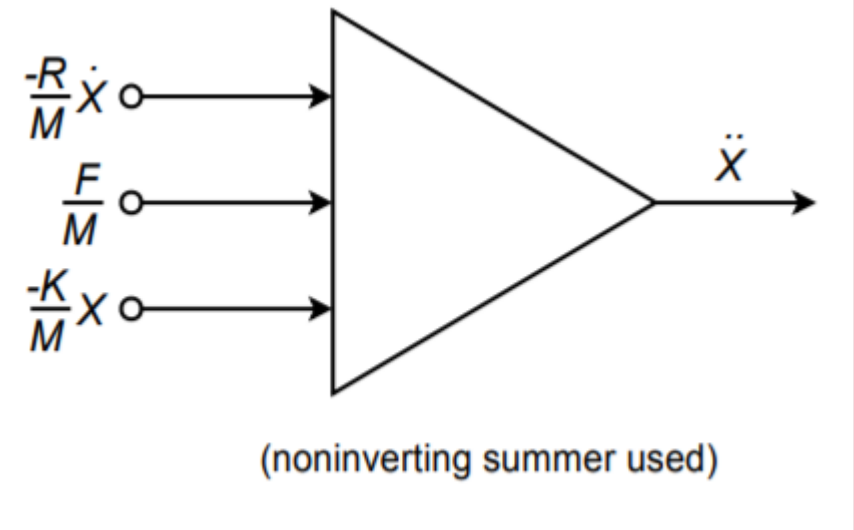


Figure 10.4.2 : Circuit realization (block form).

These elements may be produced by integrating the output of Figure 10.4.2 . Appropriate constants may be used to achieve the desired signal levels. This is shown in Figure 10.4.3 . Certain elements may be combined; for example, a weighted summing amplifier may be used to eliminate unneeded amplifiers.

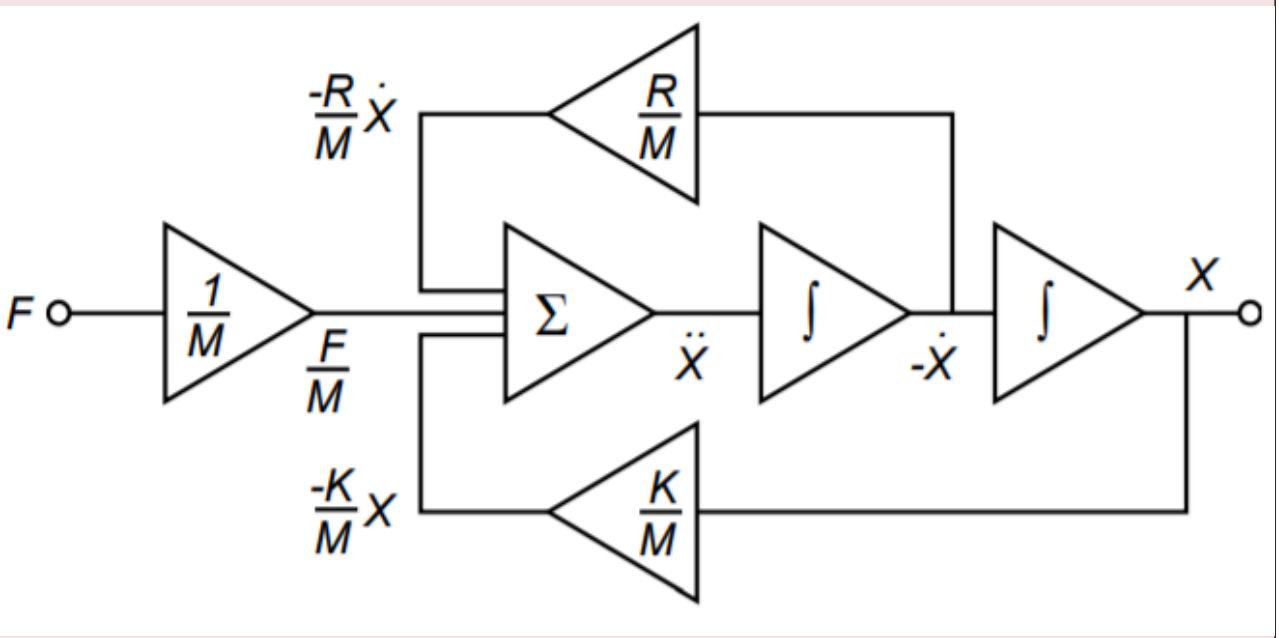


Figure 10.4.3 : Circuit realization using function blocks.

In use, the constants ◈ , ◈ , and ◈ are set by potentiometers (they are essentially nothing more than scaled gain factors). A voltage representing the excitation force is applied to the circuit, and the desired output quantity is recorded. Note that the output of interest could be the acceleration, velocity, or displacement of the body. In order to test the system with a new spring or damping constant, all that is needed is to adjust the appropriate potentiometer. In this manner, a large number of combinations may be tried quickly. The most successful combinations may then be built and tested for the final design. An analog computer such as this would be very useful in testing such items as the suspension of an automobile or a loudspeaker system. To ease the design of the simulation circuit, commercial analog computers are

available. Construction (or programming) of the circuit involves wiring integrator, amplifier, and summer blocks, together with the appropriate potentiometers. In this way, the details of designing and optimizing individual integrators or amplifiers is bypassed.

14.5 ALTERNATIVES TO INTEGRATORS AND DIFFERENTIATORS

There are alternatives to using op amp based integrators and differentiators. As long as systems can be described by a reasonable set of equations, simulation using digital computers is possible. The primary advantages of a digital computer-based simulation scheme is that it is very flexible and potentially very accurate. Before the rise of affordable and powerful desktop digital computers, the speed of response of the digital technique was a severe limitation. Analog computers can be configured as real-time or faster-than-real-time devices: they respond at the same speed or faster than the system that is being simulated. The typically heavy computation load of the digital computer requires prodigious calculation speed to keep up with reasonably fast processes. Fortunately, this is no longer the problem it once was. Another advantage of the analog computer is its immediacy. It is by nature, interactive. This means that an operator can change simulation parameters and immediately see the result. Finally, there are some applications for which the added cost of the computing hardware cannot be justified. If these are not issues, then a digital-based simulation will most likely be preferred. The general idea is to calculate the response for several closely spaced points in time. The results may be used in a variety of ways. For example, graphs may be created that can then be studied at leisure or the resulting digital signal may be used elsewhere in the system. An extreme (although generally impractical) possibility is to design an op amp based simulator and then simulate its response by using a digital circuit simulator! As you might guess, this double simulation is not particularly efficient, especially with large systems.

14.6 EXTENDED TOPIC: OTHER INTEGRATOR AND DIFFERENTIATOR CIRCUITS

The basic integrator and differentiator circuits examined earlier may be extended into other forms. Perhaps the most obvious extension is to add multiple inputs, as in an ordinary summing amplifier. In complex systems, this concept may save the use of several op amps. A summing integrator is shown in Figure 10.6.1 . Note its similarity to a normal summing amplifier. In this circuit, the input currents are summed at the inverting input of the op amp.

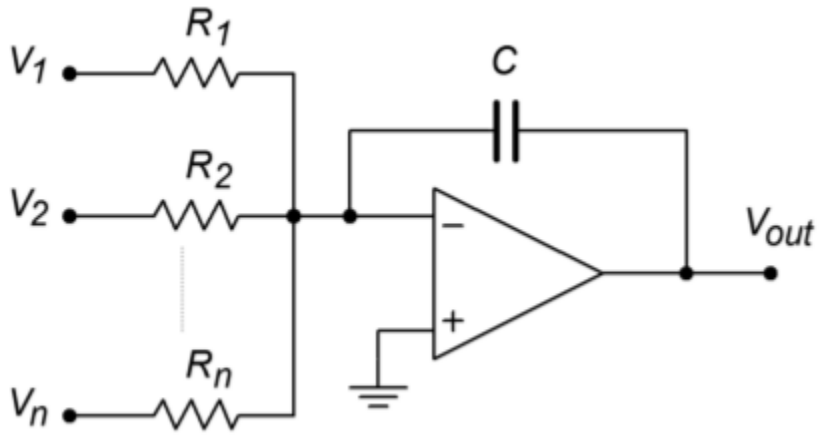


Figure 10.6.1 : Summing integrator.

If the resistors are all set to the same value, we can quickly derive the output Equation by following the original derivation. The result is

$$V_{out} = -\frac{1}{RC}\int V_1 + V_2 + \cdots + V_n dt$$

The output is the negative integral of the sum of the inputs.

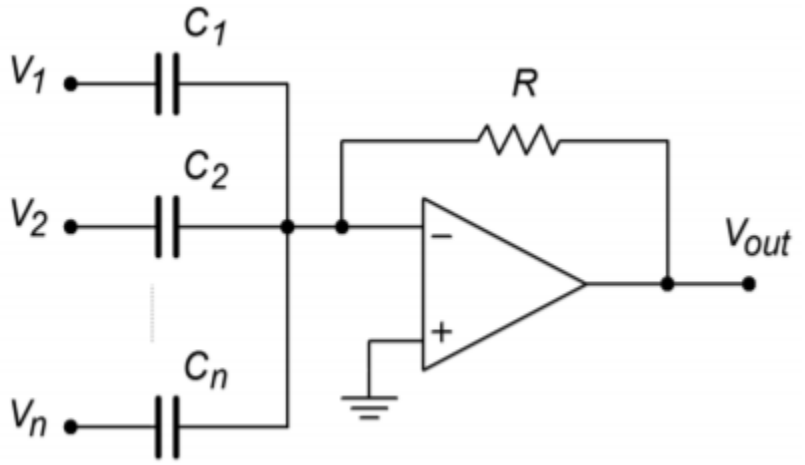


Figure 10.6.2 : Summing differentiator.

In a similar vein, a summing differentiator may be formed. This is shown in Figure 10.6.2 . Again, the proof of its output Equation follows the original differentiator derivation.

$$V_{out} = -RC_1\frac{dV_1}{dt} - RC_2\frac{dV_2}{dt} - \cdots - RC_n\frac{dV_n}{dt}$$

Another interesting adaptation of the integrator is the augmenting integrator. This circuit adds a constant gain portion to the output equation. An augmenting integrator is shown in Figure 10.6.3 . The addition of the feedback resistor ⍰⍰ provides the augmenting action. As you might surmise, the gain portion is directly related to ⍰⍰ and ⍰⍰ .

$$V_{out} = -V_{in}\frac{R_f}{R_i} - \frac{1}{RC}\int V_{in}dt$$

The augmenting integrator can also be turned into a summing/augmenting integrator by adding extra input resistors as in Figure 10.6.1 . Note that the gain portion will be the same for all inputs if the input summing resistors are of equal value.

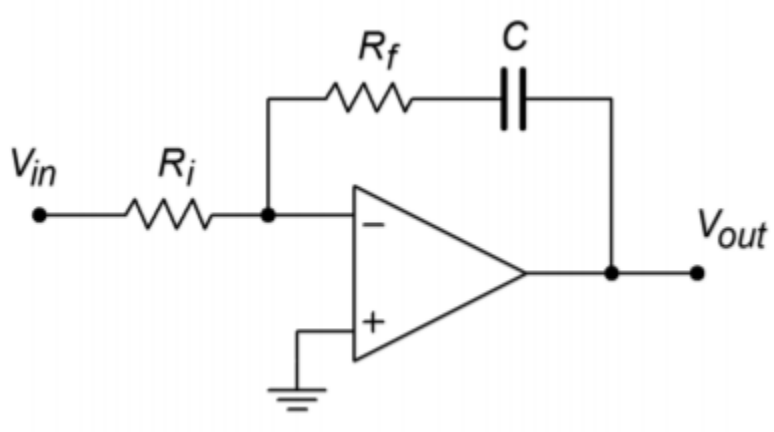


Figure 10.6.3 : Augmenting integrator.

The final variant that we shall note is the double integrator. This design requires two reactive portions in order to achieve double integration. One possibility is shown in Figure 10.6.4 . In this circuit, a pair of ⍰⍰ “Tee” networks are used. The output Equation is

$$V_{out} = -\frac{4}{(RC)^2} \int \int V_{in} dt$$

When properly used, the double integrator can cut down the parts requirement of larger circuit designs.

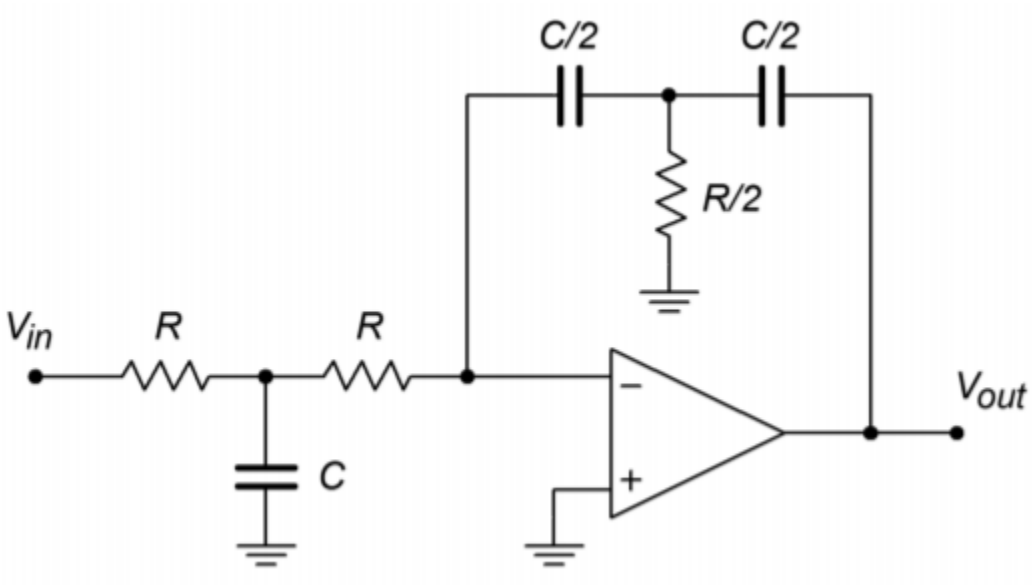


Figure 10.6.4 : Double integrator.

14.7 SUMMARY

In this chapter we have examined the structure and use of integrator and differentiator circuits. Integrators produce a summing action whereas differentiators find the slope of the input. Both types are based on the general parallel-parallel inverting voltage feedback model. In order to achieve integration and differentiation, a capacitor is used in the feedback network in place of the standard resistor. Because the capacitor current is proportional to the rate of change of the capacitor's voltage, a differential or integral response is possible. For the integrator, the capacitor is placed in the normal ⟡⟡ position, and for the differentiator, the capacitor is placed in the normal ⟡⟡ position. As a result, the integrator exhibits a –6 dB per octave slope through the useful integration range. The differentiator exhibits the mirror image, or +6 dB per octave slope, throughout its useful range.

In both circuits, practical limitations require the use of additional components. In the case of the integrator, small DC offsets at the input can force the output into saturation. To avoid this, a resistor is placed in parallel with the integration capacitor in order to limit the low frequency gain. This has the unfortunate side effect of limiting the useful integration range to higher frequencies. In the case of the differentiator, noise, stability, and input impedance limits can pose problems. In order to minimize noise and aid in stability, a small capacitor may be placed in parallel with ⟡⟡ . This reduces the high frequency gain. In order to place a lower limit on the input impedance, a resistor may be placed in series with the differentiation capacitor. The addition of either component will limit the upper range of differentiation.

We examined two general techniques for determining the output signal. The first form is referred to as the time-continuous method and although it may be used with virtually any waveform, our use was with simple sine waves only. In the case of the integrator, it corresponds to the indefinite integral – a time-domain Equation for the output is the result. The second form is the time-discrete method and is useful for waves that may be easily broken into segments. Each segment is analyzed, and the results joined graphically. This has the advantage of an immediate graphical result, whereas the shape of the time-continuous result may not be immediately apparent. The time-discrete method is also useful for producing output tables and graphs with a digital computer. In the case of the integrator, the time-discrete method corresponds to the definite integral.

Integrators and differentiators may be used in combination with summers and amplifiers to form analog computers. Analog computers may be used to model a variety of physical systems in real time. Unlike their digital counterparts, programming an analog computer only requires the proper interconnection of the various building blocks and appropriate settings for the required physical constants.

Finally, it is noted that integrators and differentiators may be used as wave shaping circuits. Integrators may be used to turn square waves into triangle waves. Differentiators can be used to turn triangle waves into square waves. Indeed, as differentiators tend to "seek" rapid changes in the input signal, they can be quite useful as edge detectors.

14.8 PROBLEMS

REVIEW QUESTIONS

1. What is the basic function of an integrator?
2. What is the basic function of a differentiator?
3. What is the function of the capacitor in the basic integrator and differentiator?
4. Why are capacitors used in favor of inductors?
5. What practical modifications need to be done to the basic integrator, and why?
6. What practical modifications need to be done to the basic differentiator, and why?
7. What are the negative side effects of the practical versus basic integrator and differentiator?
8. What is an analog computer, and what is it used for?
9. What are some of the advantages and disadvantages of the analog computer versus the digital computer?
10. How might integrators and differentiators be used as wave-shaping circuits?

PROBLEMS

Analysis Problems

1. Sketch the output waveform for the circuit of Figure 10.8.1 if the input is a 4 V peak square wave at 1 kHz.

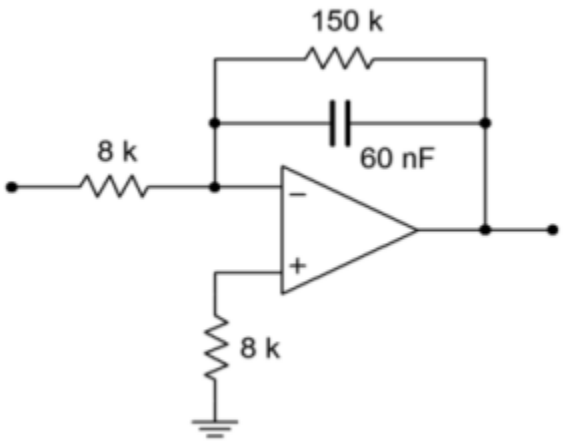


Figure 10.8.1

2. Repeat Problem 1 if ◇◇◇(◇)=5sin2◇318◇ .
3. Repeat Problem 1 if ◇◇◇(◇)=20cos2◇10000◇ .

4. Using the circuit of Problem 1, if the input is a ramp with a slope of 10 V/s, find the output after 1 ms, 10 ms, and 4 s. Sketch the resulting wave.
5. Determine the low frequency gain for the circuit of Figure 10.8.2 .

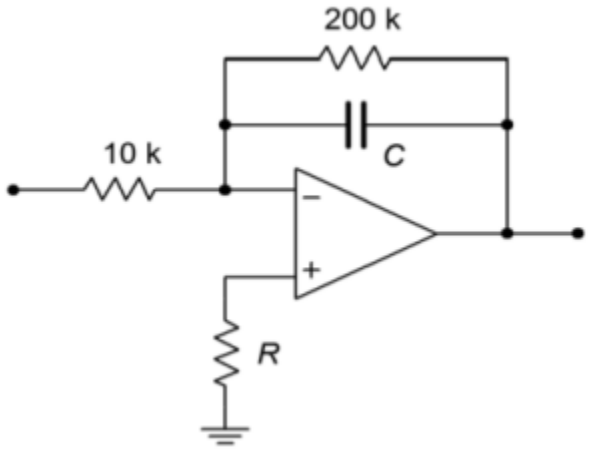


Figure 10.8.2

6. If ⍰=33 nF in Figure 10.8.2 , determine ⍰⍰⍰if ⍰⍰is a 200 mV peak sine wave at 50 kHz.
7. Repeat Problem 6 using a 200 mV peak square wave at 50 kHz.
8. Sketch the output of the circuit of Figure 10.2.7 with the input signal given in Figure 10.8.3 .
9. Assume that the input to the circuit of Figure 10.2.7 is 100 mV DC. Sketch the output waveform, and determine if it goes to saturation.

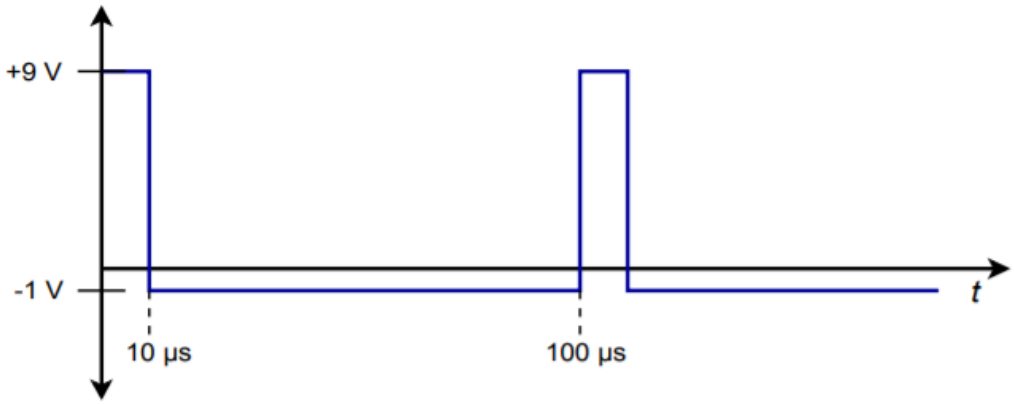


Figure 10.8.3

10. Sketch the output waveform for Figure 10.2.7, given an input of: ⍰⍰⍰(⍰)=0.5cos2⍰9000⍰ .
11. Given the circuit of Figure 10.8.4 , sketch the output waveform if the input is a 100 Hz, 1 V peak triangle wave.

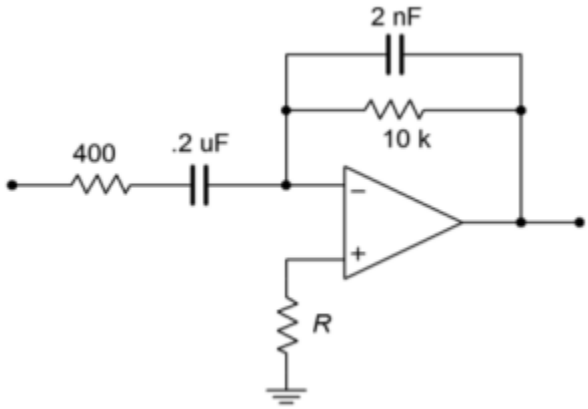


Figure 10.8.4

12. Repeat Problem 11 if the input is a 2 V peak square wave at 500 Hz. Assume that the rise and fall times are 1 ◈ s and are linear (vs. exponential).
13. Repeat Problem 11 for the following input: ◈◈◈(◈)=3cos2◈60◈ .
14. Repeat Problem 11 for the following input: ◈◈◈(◈)=0.5sin2◈1000◈ .
15. Repeat Problem 11 for the following input: ◈◈◈(◈)=10◈2 .

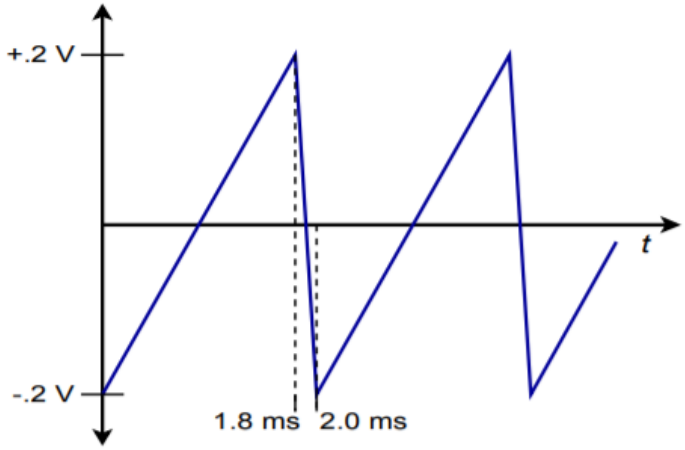


Figure 10.8.5

16. Given the input shown in Figure 10.8.5 , sketch the output of the circuit of Figure 10.19.
17. Given the input shown in Figure 10.8.5 , sketch the output of the circuit of Figure 10.19.

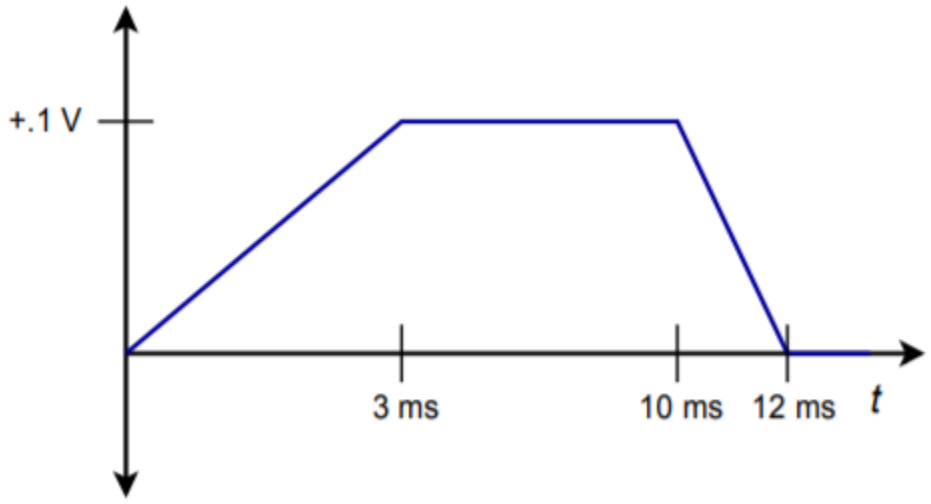


Figure 10.8.6

18. Sketch the output waveform for Figure 10.3.9, given an input of ◈◈◈(◈)=0.5cos2◈4000◈ .
19. Sketch the output waveform for Figure 10.3.9, if ◈◈◈is a 300 mV peak triangle wave at 2500 Hz.

Design Problems

20. Given the circuit of Figure 10.8.1 ,
    A. Determine the value of ◈ required in order to yield an integration constant of –2000.
    B. Determine the required of value of ◈ for minimum integration offset error.
    C. Determine ◈◈◈◈ .
    D. Determine the 99% accuracy point.
21. Given the differentiator of Figure 10.8.7 , determine the value of ◈ that will set the differentiation constant to –103 .

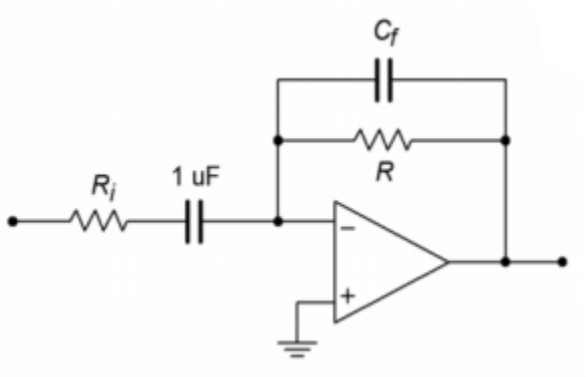


Figure 10.8.7

22. Determine the value of ◈◈ such that the maximum gain is 20 for the circuit of Figure 10.8.7 . (Use the values of Problem 21.)
23. Determine the value of ◈◈ in Problem 22 such that noise above 5 kHz is attenuated.

24. Design an integrator to meet the following specifications: integration constant of –4500, ◈◈◈◈ no greater than 300 Hz, ◈◈◈ at least 6 k Ω , and DC gain no more than 32 dB.

25. Design a differentiator to meet the following specifications: differentiation constant of –1.2◈–4 , ◈*h*◈◈*h* at least 100 kHz, and a minimum ◈◈◈ of 50 Ω .

Challenge Problems

26. Sketch the output waveform for the circuit of Figure 10.2.7 if the input is the following damped sinusoid: ◈◈◈(◈)=3◈–200◈sin2◈1000◈ .

27. A 1.57 V peak, 500 Hz square wave may be written as the following infinite series:
◈◈◈(◈)=2∑∞◈=112◈–1sin2◈500(2◈–1)◈
Using this as the input signal, determine the infinite series output Equation for the circuit of Figure 10.2.7.

28. Sketch the output for the preceding problem. Do this by graphically adding the first few terms of the output.

29. Remembering that the voltage across an inductor is equal to the inductance times the rate of change of current, determine the output Equation for the circuit of Figure 10.8.8 .

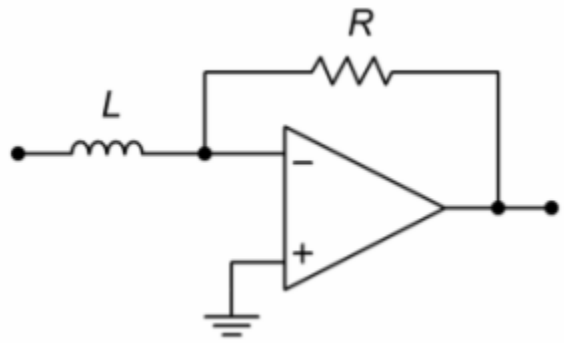


Figure 10.8.8

30. Repeat the preceding Problem for the circuit of Figure 10.8.9 .

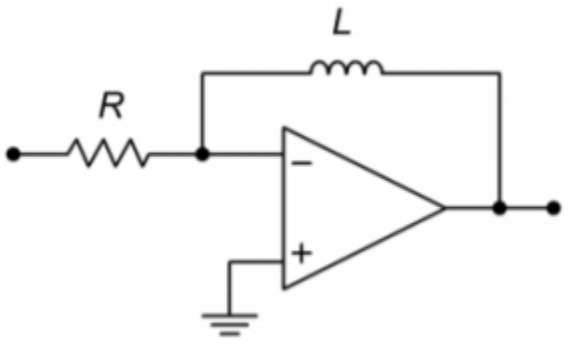


Figure 10.8.9

31. A 2.47 V peak, 500 Hz triangle wave may be written as the following infinite series:
◈◈◈(◈)=2∑∞◈=11(2◈–1)2cos2◈500(2◈–1)◈

Using this as the input signal, determine the infinite series output Equation for the circuit of Figure 10.3.9.

32. Sketch the output for the preceding problem. Do this by graphically adding the first few terms of the output.

Computer Simulation Problems

33. Simulate Problem 1 and determine the steady-state response.
34. Model Problem 8 using a simulator. Determine both the steady-state and initial outputs.
35. Simulate Problem 14. Determine the steady-state output.
36. Model Problem 22 using a simulator and determine the output signal.
37. Compare the resulting waveform produced by the circuit of Problem 8 using both the 741 op amp and the medium-speed LF411. Does the choice of op amp make a discernible difference in this application?

UNIT 15: ACTIVE FILTERS

Learning Objectives

After completing this chapter, you should be able to:

- Describe the four main types of filters.
- Detail the advantages and disadvantages of active versus passive filters.
- Describe the importance of filter order on the shape of a filter's response.
- Describe the importance of filter alignment on the shape of a filter's response.
- Compare the general gain and phase response characteristics of the popular filter alignments.
- Analyze high- and low-pass Sallen and Key filters.
- Analyze both low and high Q band-pass filters.
- Analyze state-variable filters.
- Analyze notch filters.
- Analyze cascaded filters of high order.
- Detail the operation of basic bass and treble audio equalizers.
- Explain the operation of switched capacitor filters, and list their relative advantages and disadvantages compared to ordinary op amp filters.

15.1 INTRODUCTION TO ACTIVE FILTERS

Generally speaking, a filter is a circuit that inhibits the transfer of a specific range of frequencies. Conversely, you can think of a filter as a circuit that allows only certain frequencies to pass through. Filters are used to remove undesirable frequency components from a complex input signal. The uses for this operation are many, including the suppression of power-line hum, reduction of very low or high frequency interference and noise, and specialized spectral shaping. One very common use for filters is bandwidth limiting, which, as you'll see in Chapter Twelve, is an integral part of any analog-to-digital conversion system.

There are numerous variations on the design and implementation of filters. Indeed, an in-depth discussion of filters could easily fill more than one textbook. Our discussion must by necessity be of a limited and introductory nature. This chapter deals with the implementation of a number of popular op amp filter types. Due to the finite space available, every mathematical proof for the design sequences will not be detailed here, but may be found in the references listed at the end of the chapter.

Filter implementations may be classified into two very broad, yet distinct, camps: digital filters and analog filters. Digital filters work entirely in the digital domain, using numeric data as the input signal. The design of digital filters is an advanced topic and will not be examined here.

The second category, analog filters, utilizes standard linear circuit techniques for their construction. Analog filter implementations can be broken into two subcategories: passive and active. Passive filters utilize only resistors, inductors, and capacitors, whereas active filters make use of active devices (i.e., discrete transistors or op amps) as well. Although we will be examining only one subcategory in the world of filters, it is important to note that many of the circuits that we will design can be realized through passive analog filters or digital filters.

15.2 FILTER TYPES

No matter how a filter is realized, it will usually conform to one of four basic response types. These types are: high-pass, low-pass, band-pass, and band-reject (also known as a notch type). A high-pass filter allows only frequencies above a certain break point to pass through. In other words, it attenuates low frequency components. Its general amplitude frequency response is shown in Figure 11.2.1 . A low-pass filter is the logical mirror of the high-pass, that is, it allows only low frequency signals to pass through, while suppressing high-frequency components. Its response is shown in Figure 11.2.2 . The band-pass filter can be thought of as a combination of high- and low-pass filters. It only allows frequencies within a specified range to pass through. Its logical inverse is the band-reject filter that allows everything to pass through, with the exception of a specific range of frequencies. The amplitude frequency response plots for these last two types are shown in Figures 11.2.3 and 11.2.4 , respectively. In complex systems, it is possible to combine several different types together in order to achieve a desired overall response characteristic.

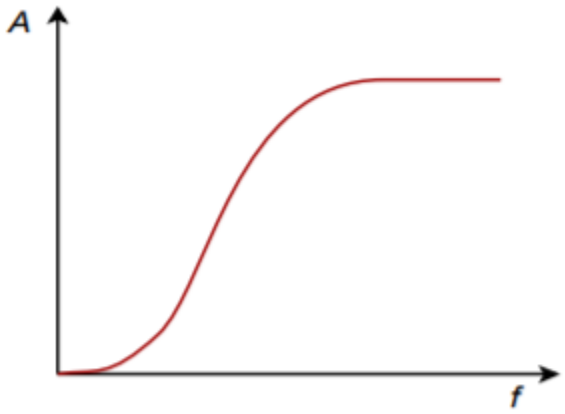


Figure 11.2.1 : High-pass response.

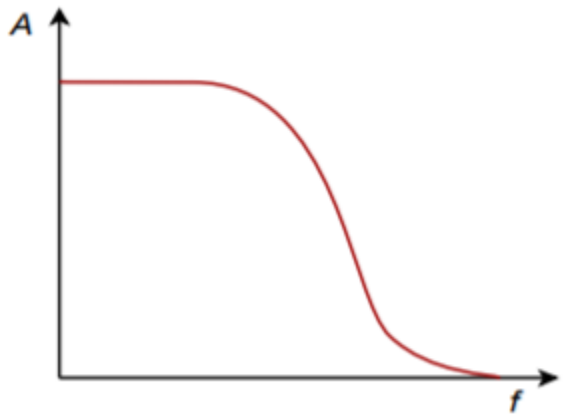


Figure 11.2.2 : Low-pass response.

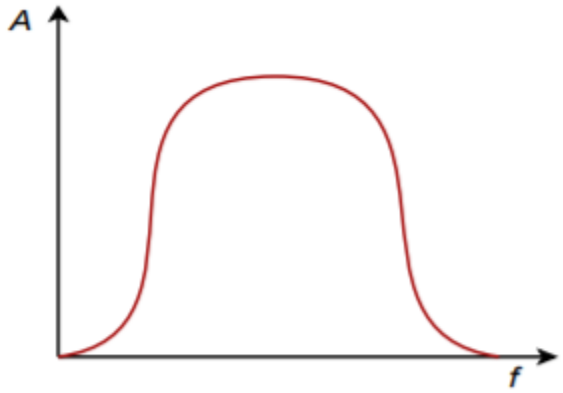


Figure 11.2.3 : Band-pass response.

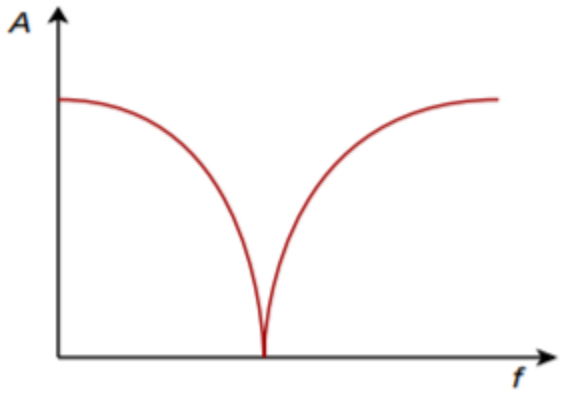


Figure 11.2.4 : Notch response.

In each plot, there are three basic regions. The flat area where the input signal is allowed to pass through is known as the pass band. The edge of the pass band is denoted by the break frequency. The break frequency is usually defined as the point at which the response has fallen 3 dB from its pass-band value. Consequently, it is also referred to as the dB down frequency. It is important to note that the break frequency is not} always equal to the natural critical frequency, ⍰⍰ . The area where the input signal is fully suppressed is called the stop band. The section between the pass band and the stop band is referred to as the transition band. This is shown for the low-pass filter in Figure 11.2.5 . Note that plots 11.2.1 through 11.2.4 only define the general response of the filters. The actual shape of the transition band for a given design will be refined in following sections.

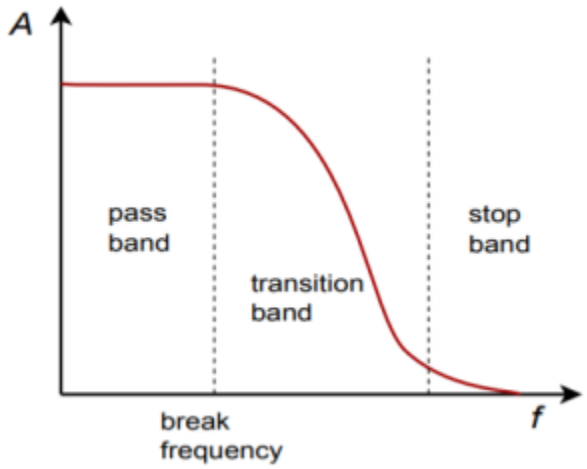


Figure 11.2.5 : Filter regions.

These response plots may be achieved through the use of simple RLC circuits. For example, the simple lag network shown in Figure 11.2.6 may be classified as a low-pass filter. You know from earlier work that lag networks attenuate high frequencies. In a similar vein, a band-pass filter may be constructed as shown in Figure 11.2.7 . Note that at resonance, the tank circuit exhibits an impedance peak, whereas the series combination exhibits its minimum impedance. Therefore, a large portion of ⍰⍰⍰will appear at ⍰⍰⍰⍰. At very high or low frequencies the situation is reversed (the parallel combination shows a low impedance, and the series combination shows a high impedance), and very little of ⍰⍰⍰appears at the output. Note that no active components are used in this circuit. This is a passive filter.

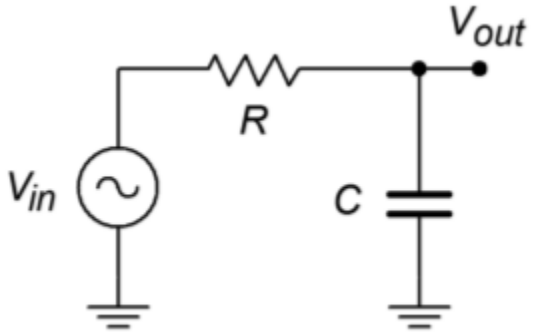


Figure 11.2.6 : Passive low-pass filter.

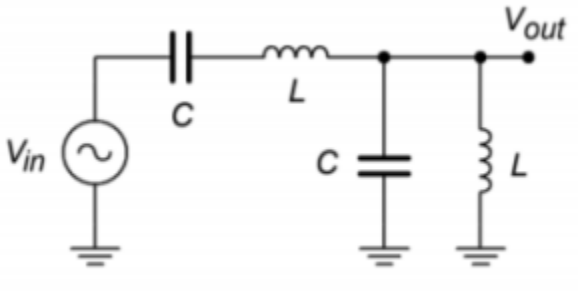


Figure 11.2.7 : Passive band-pass filter.

15.3 THE USE AND ADVANTAGES OF ACTIVE FILTERS

If filters may be made with only resistors, capacitors, and inductors, you might ask why anyone would want to design a variation that required the use of an op amp. This is a good question. Obviously, there must be certain shortcomings or difficulties associated with passive filter designs, or active filters would not exist. Active filters offer many advantages over passive implementations. First of all, active filters do not exhibit insertion loss. This means that the pass-band gain will equal 0 dB. Passive filters always show some signal loss in the pass band. Active filters may be made with pass-band gain, if desired. Active filters also allow for interstage isolation and control of input and output impedances. This alleviates problems with interstage loading and simplifies complex designs. It also produces modest component sizes (e.g., capacitors tend to be smaller for a given response). Another advantage of the active approach is that complex filters may be realized without using inductors. This is desirable, as practical inductors tend to be far less ideal than typical resistors and capacitors and are generally more expensive. The bottom line is that the active approach allows for the rapid design of stable, economical filters in a variety of applications.

Active filters are not perfect. First, by their very nature, active filters require a DC power supply whereas passives do not. This is usually not a problem, as the remainder of the circuit will probably require a DC supply anyway. Active filters are also limited in their frequency range. An op amp has a finite gain-bandwidth product, and the active filter produced can certainly not be expected to perform beyond it. For example, it would be impossible to design a filter that only passes frequencies above 10 MHz when using a standard ⋄ A741. Passive circuits do not have this limitation and can work well into the hundreds of MHz. Finally, active filters are not designed to handle large amounts of power. They are low signal-level circuits. With appropriate component ratings, passive filters may handle hundreds of watts of input power. A classic example of this is the crossover network found in most home loudspeaker systems. The crossover network splits the music signal into two or more bands and routes the results to individual transducers that are optimized to work within a given frequency range. Because the input to the loudspeaker may be as high as a few hundred watts, a passive design is in order.[1] Consequently, we can say that active filters are appropriate for designs at low to moderate frequencies (generally no more than 1 MHz with typical devices) that do not have to handle large amounts of power. As you might guess, that specification covers a great deal of territory, and therefore, active filters based on op amps have become rather popular.

1. In more advanced playback systems, active filters can be used. Examples include recording studio monitors and public address systems. We'll take a look at just how this is done in one of the upcoming examples.

15.4 FILTER ORDER AND POLES

The rate at which a filter's response falls in the transition band is determined by the filter's order. The higher the order of a filter, the faster its rolloff rate is. The order of a filter is given as an integer value and is derived from the filter's transfer function. As an example, all other factors being equal, a fourth-order filter will roll off twice as fast as a second-order filter, and four times faster than a first-order unit. The order of a filter also indicates the minimum number of reactive components that the filter will require. For example, a third-order filter requires at least three reactive components: one capacitor and two inductors, two capacitors and one inductor, or in the case of an active filter, three capacitors. Related to this is the number of poles that a filter utilizes. It is common to hear descriptions such as "a four-pole filter". For most general-purpose high- or low-pass filters, the terms pole and order may be used interchangeably and completely describe the rolloff rate. For more complex filters this isn't quite the case, and you may also hear descriptions such as "a sixpole, two-zero filter". Because this chapter is an introduction to filters, we will not detail the operation of these more esoteric types. Suffice it to say that when a circuit is described as an ◈◈*h* -order filter, you may assume that it is an ◈ -pole filter, as well.

A general observation can be given that the rolloff rate of a filter will eventually approach 6 dB per octave per pole (20 dB per decade per pole). Therefore, a thirdorder filter (i.e., three-pole) eventually rolls off at a rate of 18 dB per octave (60 dB per decade). We say "eventually" because the response around the break frequency may be somewhat faster or slower than this value. Figure 11.4.1 compares the effect of order on four otherwise identical low-pass filters. Note that the higher order filters offer greater attenuation at any frequency beyond the break point. As with most response plots, Figure 11.4.1 utilizes decibel instead of ordinary gain. Also, these filters are shown with unity gain in the pass band, although this doesn't have to be the case. High-order filters are used when the transition band needs to be as narrow as possible. It is not uncommon to see twelfth-order and higher filters used in special applications. As you might guess, higher order filters are more complex and costly to design and build. For many typical applications, orders in the range of two to six are common.

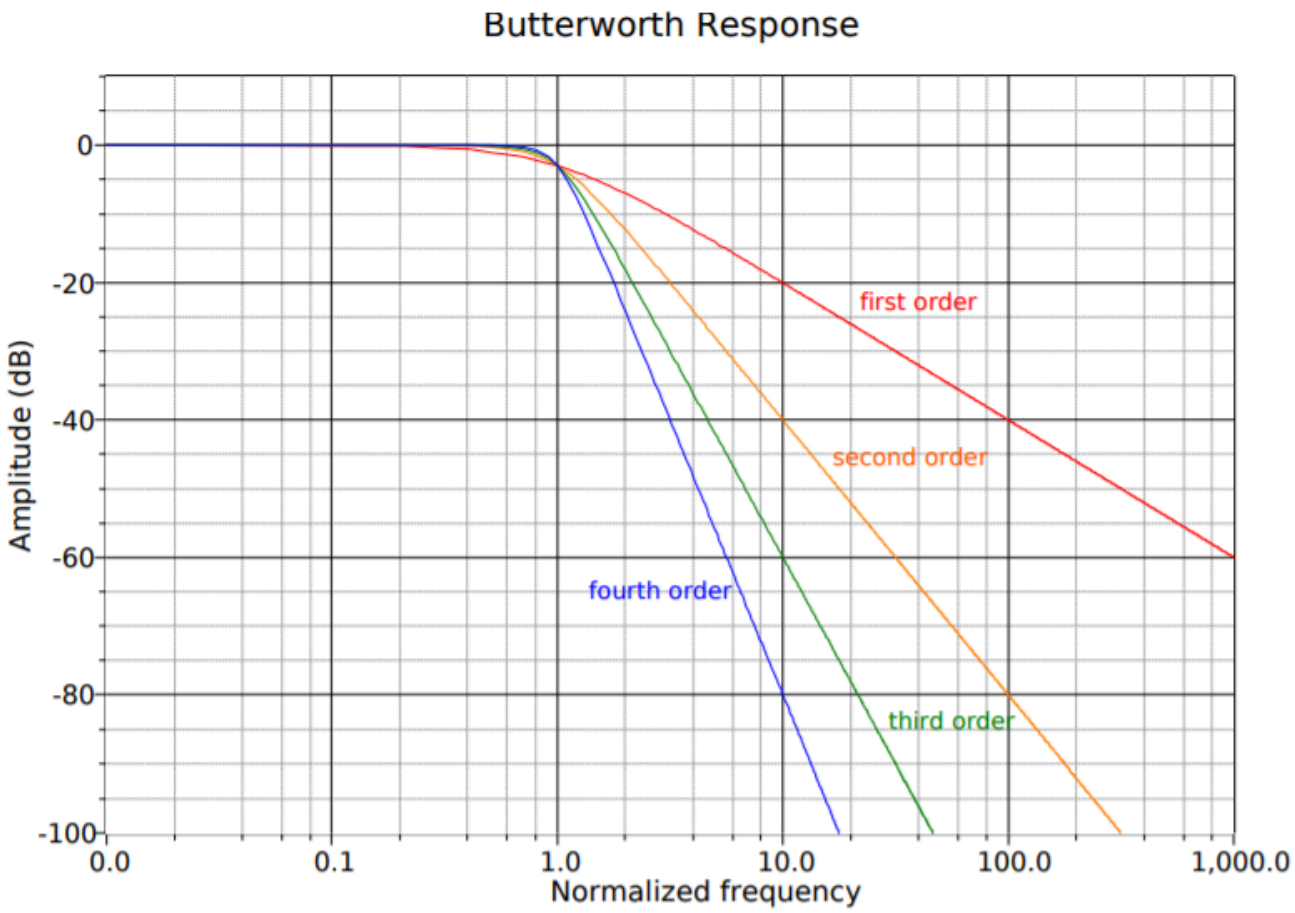


Figure 11.4.1 : Effect of order on low-pass filters.

15.5 FILTER CLASS OR ALIGNMENT

Besides order, the shape of the transition band is determined by a filter's class, or alignment. These terms are synonymous and reflect the filter's damping factor. Damping factor is the reciprocal of ◈ , the quality factor. You should be familiar with ◈ from earlier work with inductors and resonant circuits. The symbol for damping factor is alpha, ◈ . Alignment plays a key role in determining the shape of the transition region and, in some cases, the pass-band or stop-band shape also. There are a great number of possible filter alignments. We will look at a few of the more popular types. A graph comparing the relative responses of the major types is shown in Figure 11.5.1 . For simplicity, only second-order types are shown. Figure 11.5.1◈ shows filters with the same critical frequency and identical DC gains. In Figure 11.5.1◈ the responses have been adjusted for a peak gain of 0 dB and identical break frequencies (◈3◈◈) .

Figure 11.5.1◈ : A comparison of the major filter alignments: second order response.

Figure 11.5.1◇: A comparison of the major filter alignments: second order response (adjusted).

BUTTERWORTH

Perhaps the most popular alignment type is the Butterworth. The Butterworth is characterized by its moderate amplitude and phase response. It exhibits the fastest rolloff of any monotonic (i.e., single slope or smooth) filter. In the time-domain, moderate ringing on pulses may be observed. This is also the only filter whose 3 dB down frequency equals its critical frequency (◇3◇◇=◇◇) . The Butterworth makes an excellent general-purpose filter and is widely used.

BESSEL

Like the Butterworth, the Bessel is also monotonic, so it shows a smooth pass-band response. The transition region is somewhat elongated though, and the initial rolloff is less than 6 dB per octave per pole. The Bessel does exhibit a linear phase response that produces little ringing in the time domain. It is therefore a good choice for filtering pulses when the overall shape of the pulse must remain coherent (i.e., smooth and undistorted in time).

CHEBYSHEV

The Chebyshev is actually a class of filters all its own. They are based on Chebyshev polynomials. There are many possible variations on this theme. In general, the Chebyshev exhibits initial rolloff

rates in excess of 6 dB per octave per pole. This extra-fast transition is paid for in two ways: first, the phase response tends to be rather poor, resulting in a great deal of ringing when filtering pulses or other fast transients. The second effect is that the Chebyshev is non-monotonic. The passband response is not smooth; instead, ripples may be noticed. In fact, the height of the ripples defines a particular Chebyshev response. It is possible to design an infinite number of variations from less than 0.1 dB ripple to more than 3 dB ripple. Generally, the more ripple you can tolerate, the greater the rolloff will be, and the worse the phase response will be. The choice is obviously one of compromise. The basic differences between the various Chebyshev types are characterized in Figure 11.5.2 . Figure 11.5.2 compares two different low-pass Chebyshev filters of the same order. Note that only the height of the ripples is different. Figure 11.5.2 compares equal ripple Chebyshevs, but of different orders. Note that the higher-order filter exhibits a greater number of ripples. Also, note that even-order Chebyshevs exhibit a dip at DC whereas odd-order units show a crest at DC. The number of ripples in the passband is equal to the order of the filter divided by 2. When compared to the Butterworth and Bessel alignments in Figure 11.5.1 , it is apparent that heavy damping produces the smoothest curves. The final note on the Chebyshev concerns, of all things, its spelling. You will often see "Chebyshev" spelled in a variety of ways, including "Chebycheff" and "Tschebycheff". It merely depends on how the name of the Russian mathematician is transliterated from Russian Cyrillic into English. The spellings all refer to the same filter.

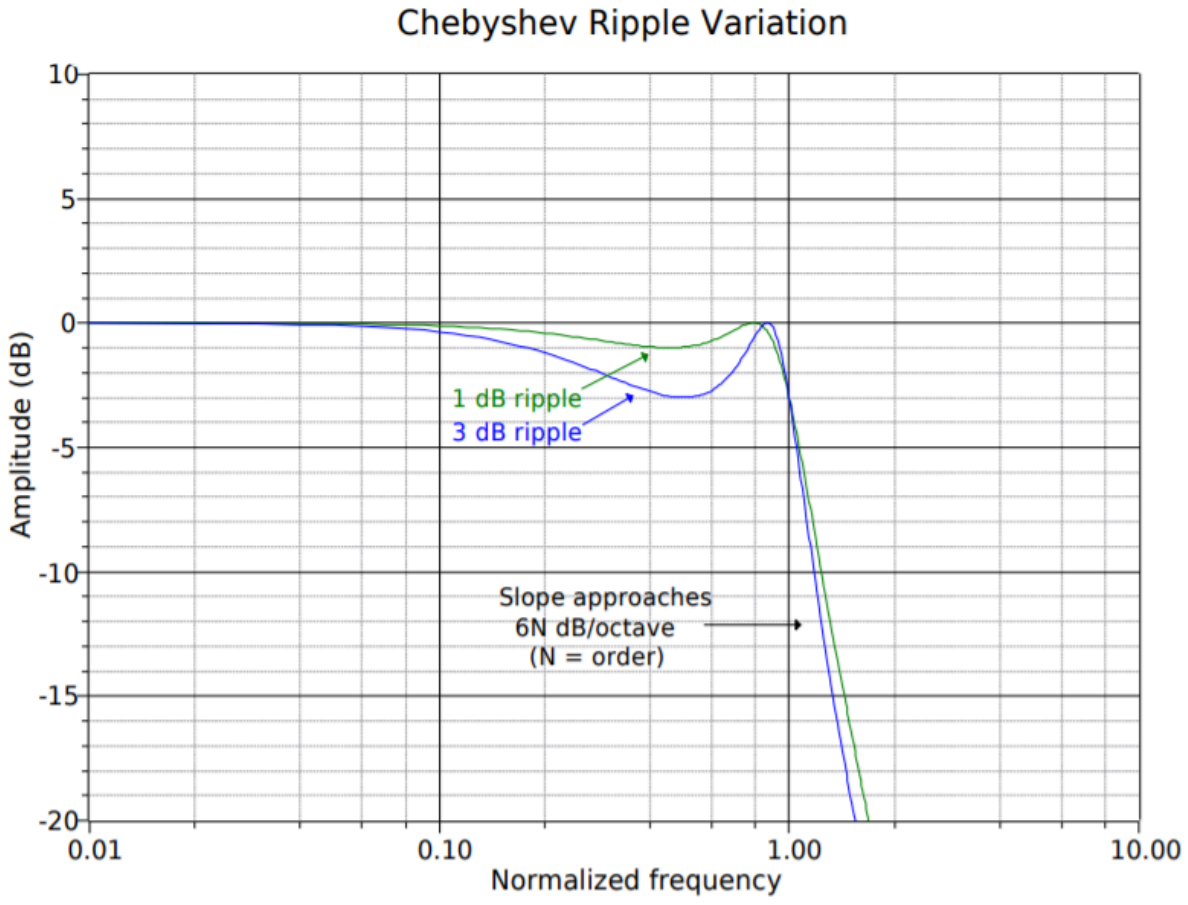


Figure 11.5.2 : Variation of Chebyshev ripple parameter.

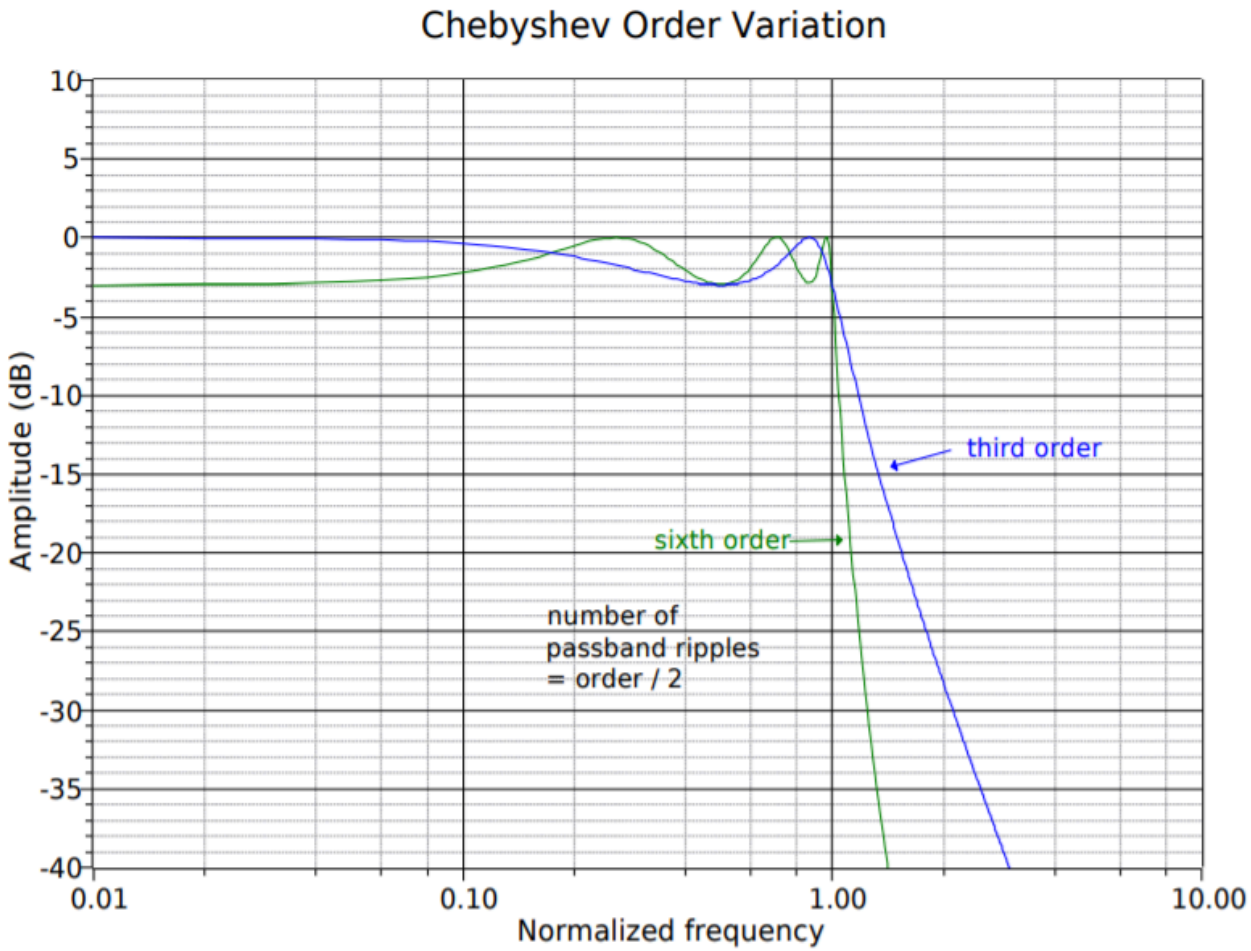


Figure 11.5.2 : Variation of Chebyshev order parameter.

ELLIPTIC

The elliptic is also known as the Cauer alignment. It is a somewhat more advanced filter. It achieves very fast initial rolloff rates. Unlike the other alignments noted above, the elliptic does not "roll off forever". After its initial transition, the response rises back up, exhibiting ripples in the stop band. A typical response is shown in Figure 11.5.3 . The design of elliptics is an advanced topic and will not be considered further. It should be noted, though, that they are popular in analog-todigital conversion systems that require very narrow transition bands.

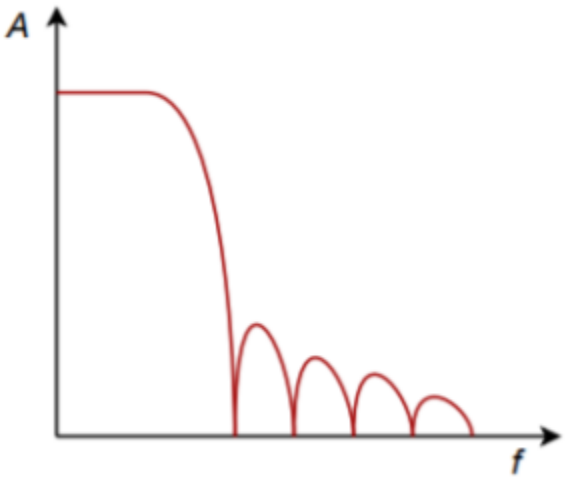


Figure 11.5.3 : Elliptic response

OTHER POSSIBILITIES

In order to optimize the time-domain and frequency-domain characteristics for specific applications, a number of other alignments may be used. These include such alignments as Paynter and Linkwitz-Reilly and are often treated as being midway between Bessel and Butterworth, or Butterworth and Chebyshev.

15.6 REALIZING PRACTICAL FILTERS

Usually, filter design starts with a few basic, desired parameters. This usually includes the break frequency, the amount of ripple that may be tolerated in the pass band (if any), and desired attenuation levels at specific points in the transition and stop bands. Phase response and associated time delays may also be specified. If phase response is paramount, the Bessel is normally chosen. Likewise, if pass-band ripple cannot be allowed, Chebyshevs are not considered. With the use of comparative curves such as those found in Figures 11.4.1 through 11.5.2, the filter order and alignment may be determined from the required attenuation values. At this point, the filter performance will be fully specified, for example, a 1 kHz, low-pass, third-order Butterworth. There are many ways in which this specification may be realized.

SALLEN AND KEY VCVS FILTERS

There are many possible ways to create an active filter. Perhaps the most popular forms for realizing active high- and low-pass filters are the Sallen and Key Voltage-Controlled Voltage Source models. As the name implies, the Sallen and Key forms are based on a VCVS; in other words, they use series-parallel negative feedback. A general circuit for these models is shown in Figure 11.6.1 . This circuit is a two-pole (second-order) section and can be configured for either high- or low-pass filtering. For our purposes, the amplifier block will utilize an op amp, although a discrete amplifier is possible. Besides the amplifier, there are four general impedances in the circuit. Usually, each element is a single resistor or capacitor. As we will see, the selection of the component type will determine the type of filter.

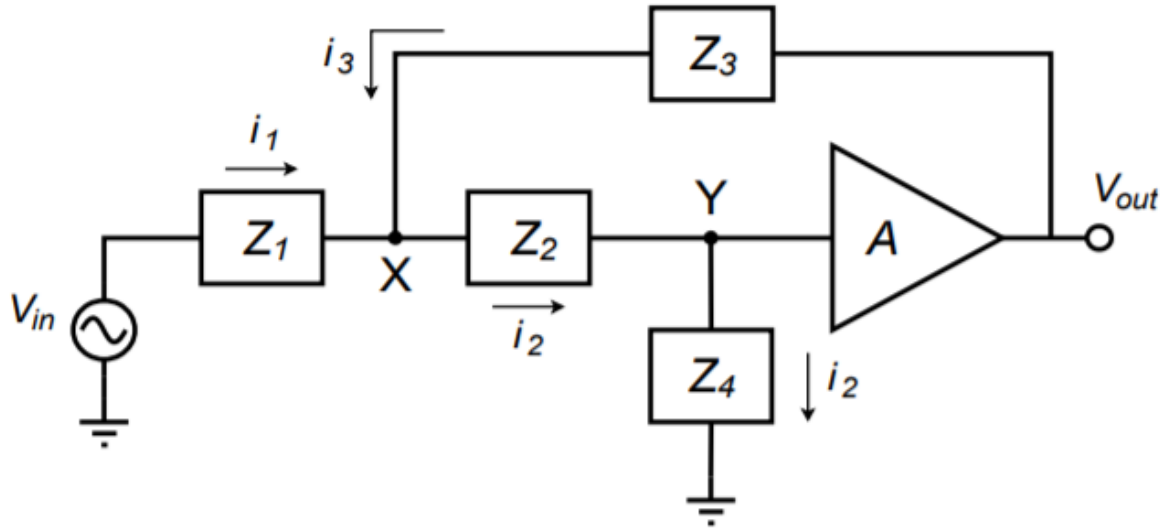


Figure 11.6.1 : General VCVS 2-pole filter section.

At this point, we need to derive the general transfer equation for the circuit. Once the general equation is established, we will be able to refine the circuit for special cases. Being a higher-order active circuit, this procedure will naturally be somewhat more involved than the derivations found in Chapter One

for the simple first-order lead and lag networks. The concepts are consistent though. This derivation utilizes a nodal analysis. In order to make the analysis a bit more convenient, it will help if we declare the voltage ⍰⍰ as 1 V. This will save us from carrying an input voltage factor through our calculations, which would need to be factored out of the general equation anyway. By inspection we note the following:

$$V_{out} = AV_y = A$$

(11.6.1)

$$i_1 = \frac{V_{in} - V_x}{Z_1}$$

$$i_2 = \frac{1}{Z_4}$$

(11.6.2)

$$i_3 = \frac{V_{out} - V_x}{Z_3} = \frac{A - V_x}{Z_3}$$

$$V_x = i_2 Z_2 + V_y = i_2 Z_2 + 1$$

(11.6.3)

By the substituting Equation 11.6.2 into Equation 11.6.3 , ⍰⍰ may be expressed as

$$V_x = \frac{Z_2}{Z_4} + 1$$

We now sum the currents according to the figure,

$$i_2 = i_1 + i_3$$

substitute our current equivalences,

$$\frac{1}{Z_4} = \frac{V_{in} - V_x}{Z_1} + \frac{A - V_x}{Z_3}$$

and solve the equation in terms of ⍰⍰⍰:

$$\frac{V_{in} V_x}{Z_1} = \frac{1}{Z_4} \frac{A V_x}{Z_3}$$

[V_{in} – V_x = \frac{Z_1}{Z_4} – \frac{Z_1 (A – V_x)}{Z_3} \]

[V_{in} = \frac{Z_1}{Z_4} – \frac{Z_1 (A – V_x)}{Z_3} +V_x \]

[V_{in} = \frac{Z_1}{Z_4} – \frac{Z_1 (A – V_x)}{Z_3} + \frac{Z_2}{Z_4} +1 \]

[V_{in} = \frac{Z_1}{Z_4} – \frac{Z_1}{Z_3} \left(A – \frac{Z_2}{Z_4} –1 \right) + \frac{Z_2}{Z_4} +1 \]

[V_{in} = \frac{Z_1}{Z_4} – \frac{Z_1}{Z_3} A+ \frac{Z_1 Z_2} {Z_3 Z_4} + \frac{Z_1}{Z_3} + \frac{Z_2}{Z_4} +1 \] (11.6.4)

We can now write our general transfer equation using Equations 11.6.1 and 11.6.4 .

$$\frac{V_{out}}{V_{in}} = \frac{A}{\frac{Z_1}{Z_4} + \frac{Z_1}{Z_3}(1 - A) + \frac{Z_1 Z_2}{Z_3 Z_4} + \frac{Z_2}{Z_4} + 1}$$

(11.6.5)

Equation 11.6.5 is applicable to any variation on Figure 11.6.1 that we wish to make. All we

need to do is substitute the appropriate circuit elements for ◈1 through ◈4 . As you might guess, direct substitution of resistance and reactance values would make for considerable work. In order to alleviate this difficulty, designers rely on the Laplace transform technique. This is also known as the ◈ domain technique. A detailed analysis of the Laplace transform is beyond the scope of this text, but some familiarity will prove helpful. The basic idea is to replace complex terms with simpler ones. This is performed by setting the variable ◈ equal to ◈◈ . As you will see, this substitution leads to far simpler and more generalized circuit derivations and equations. A capacitive reactance may be reduced as follows.

$$\text{Capacitive Reactance } = -jX_C$$

$$\text{Capacitive Reactance } = \frac{-j}{\omega C}$$

$$\text{Capacitive Reactance } = \frac{1}{j\omega C}$$

$$\text{Capacitive Reactance } = \frac{1}{sC}$$

Therefore, whenever a capacitive reactance is needed, the expression 1/◈◈ is used. Equations can then be manipulated using basic algebra.

SALLEN AND KEY LOW-PASS FILTERS

Let's derive a general expression based on Equation 11.6.5 for low-pass filters. A low-pass filter is a lag network, so to echo this, we will use resistors for the first two elements and capacitors for the third and forth. Using the ◈ operator we find, ◈1=◈1 , ◈2=◈2,◈3=1/◈◈1 , and ◈4=1/◈◈2 .

$$\frac{V_{out}}{V_{in}} = \frac{A}{\frac{Z_1}{Z_4} + \frac{Z_1}{Z_3}(1-A) + \frac{Z_1 Z_2}{Z_3 Z_4} + \frac{Z_2}{Z_4} + 1}$$

$$\frac{V_{out}}{V_{in}} = \frac{A}{sR_1C_2 + sR_1C_1(1-A) + s^2R_1R_2C_1C_2 + sR_2C_2 + 1}$$

$$\frac{V_{out}}{V_{in}} = \frac{A}{s^2R_1R_2C_1C_2 + s(R_1C_2 + R_2C_2 + R_1C_1(1-A)) + 1}$$

The highest power of ◈ in the denominator determines the number of poles in the filter. Because this is a 2 here, the filter must be a 2 pole-type (second-order), as expected. Usually, it is most convenient if the denominator coefficient for ◈2 is unity. This makes the equation easier to factor.

$$\frac{V_{out}}{V_{in}} = \frac{A/R_1R_2C_1C_2}{s^2 + s(\frac{1}{R_2C_1} + \frac{1}{R_1C_1} + \frac{1}{R_2C_2}(1-A)) + \frac{1}{R_1R_2C_1C_2}}$$

(11.6.6)

Second-order systems appear in a variety of areas including mechanical, acoustical, hydraulic, and electrical. They have been widely studied, and a generalized form of one group of second-order responses is given by

$$G = \frac{A\omega^2}{s^2 + \alpha\omega s + \omega^2}$$

(11.6.7)

where ◈ is the gain of the system, ◈ is the resonant frequency in radians, and ◈ is the damping factor.

By comparing the general form of Equation 11.6.7 to the low-pass filter Equation 11.6.6 , we find that

$$\omega^2 = \frac{1}{R_1 R_2 C_1 C_2}$$

(11.6.8)

$$\alpha\omega = \frac{1}{R_2 C_1} + \frac{1}{R_1 C_1} + \frac{1}{R_2 C_2}(1 - A)$$

(11.6.9)

Based on the general expression of Equation 11.6.7 , we can derive equations for the gain magnitude versus frequency and phase versus frequency, as we did in Chapter One for the first-order systems. For most filter work, it is convenient to work with normalized frequency instead of a true frequency. This means that the critical frequency will be set to 1 radian per second, and a generalized equation developed. For circuits using other critical frequencies, the general equation is used and its results are simply scaled by a factor equal to this new frequency. The normalized version of Equation 11.6.7 is

$$G = \frac{A}{s^2 + \alpha s + 1}$$

(11.6.10)

In order to determine the gain and phase expressions, Equation 11.6.10 must be split into its real and imaginary components. The first step is to replace ◈ with its equivalent, ◈◈ , and then group the real and imaginary components.

$$G = \frac{A}{(j\omega)^2 + j\alpha\omega + 1}$$

$$G = \frac{A}{-\omega^2 + j\alpha\omega + 1}$$

$$G = \frac{A}{(1 - \omega^2) + j\alpha\omega}$$

To split this into separate real and imaginary components, we must multiply the numerator and denominator by the complex conjugate of the denominator, (1–◈2)–◈◈◈ .

$$G = \frac{A}{(1\omega^2) + j\alpha\omega}\frac{(1\omega^2)j\alpha\omega}{(1\omega^2)j\alpha\omega}$$

[G = \frac{A((1–\omega^2) – j \alpha\omega)}{(1–\omega^2)^2+\alpha^2 \omega^2} \]

Applying Equation 11.6.11 to this relation yields

$$\text{Mag} = \sqrt{\left(\frac{A(1-\omega^2)}{(1-\omega^2)^2+\alpha^2\omega^2}\right)^2 + \left(-\frac{A\alpha\omega}{(1-\omega^2)^2+\alpha^2\omega^2}\right)^2}$$

$$\text{Mag} = \sqrt{\frac{A^2(1-\omega^2)^2}{((1-\omega^2)^2+\alpha^2\omega^2)^2} + \frac{A^2\alpha^2\omega^2}{((1-\omega^2)^2+\alpha^2\omega^2)^2}}$$

$$\text{Mag} = \frac{\sqrt{A^2((1\omega^2)^2+\alpha^2\omega^2)}}{((1-\omega^2)^2+\alpha^2\omega^2)^2}$$

$$\text{Mag} = \frac{A\sqrt{(1-\omega^2)^2+\alpha^2\omega^2}}{(1-\omega^2)^2+\alpha^2\omega^2}$$

$$\text{Mag} = \frac{A}{\sqrt{(1-\omega^2)^2+\alpha^2\omega^2}}$$

(11.6.12)

For the phase response, recall that ◇=tan–1(◇◇◇◇◇◇◇◇◇/◇◇◇◇) . Applying Equation 11.6.11 to this relation yields

$$\theta = \arctan \frac{-\frac{A\alpha\omega}{(1-\omega^2)^2+\alpha^2\omega^2}}{\frac{A(1-\omega^2)}{(1-\omega^2)^2+\alpha^2\omega^2}}$$

$$\theta = \arctan \frac{-A\alpha\omega}{A(1-\omega^2)}$$

(11.6.13)

To use Equations 11.6.12 and 11.6.13 with a particular alignment, substitute the appropriate damping value and simplify. Derivation of specific damping factors is beyond the scope of this chapter, but can be found in texts specializing on filter design. For our purposes, tables of damping factors for specific alignments will be presented. As one possible example, the damping factor for a second-order Butterworth alignment is 2‾√ . Substituting this into Equation 11.6.12 produces

$$\text{Mag} = \frac{A}{\sqrt{1+\omega^2(\alpha^2-2)+\omega^4}}$$

$$\text{Mag} = \frac{A}{\sqrt{1+\omega^2((\sqrt{2})^2-2)+\omega^4}}$$

$$\text{Mag} = \frac{A}{\sqrt{1+\omega^4}}$$

There are a large number of ways of configuring a low-pass filter given the above equations. So that

we might put some consistency into what appears to be a chaotic mess, we'll look at two distinct and useful variations. They are the equal component realization and the unity gain realization.

THE EQUAL-COMPONENT VERSION

Here, we will set ◈1=◈2 and ◈1=◈2 . To keep the resulting equation generic, we will use a normalized frequency of 1 radian per second. For other tuning frequencies, we will just scale our results to the desired value. From Equation 11.6.8 , if ◈=1 , then ◈1◈2◈1◈2=1 , and the most straightforward solution would be to set ◈1=◈2=◈1=◈2=1 (units of ohms and farads). Equation 11.6.6 simplifies to

$$\frac{V_{out}}{V_{in}} = \frac{A}{s^2 + s(1 + 1 + 1(1 - A)) + 1}$$

Note the damping factor is now given by

$$\alpha = 1 + 1 + 1(1 - A)$$

$$\alpha = 3 - A$$

We see that the gain and damping of the filter are linked together. Indeed, for a certain damping factor, only one specific gain will work properly:

$$A = 3 - \alpha$$

As the gain of a noninverting amplifier is

$$A = 1 + \frac{R_f}{R_i}$$

we may find the required value for ◈◈ by combining these two equations:

$$R_f = 2 - \alpha$$

The finished prototype is shown in Figure 11.6.2 .

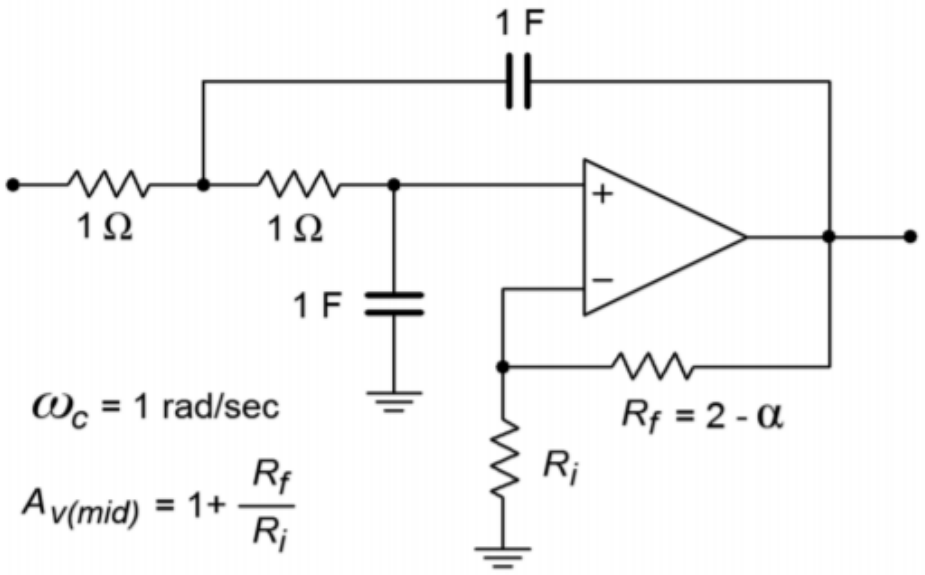


Figure 11.6.2 : Low-pass equal-component VCVS.

THE UNITY-GAIN VERSION

Here we will set ◈=1 , and ◈1=◈2 . From Equation 11.6.8 , if ◈=1 , then ◈1◈2◈1◈2=1 , and therefore ◈1=1/◈2 . In effect, the ratio of the capacitors will set the damping factor for the system. Equation 11.6.6 may be simplified to

$$\frac{V_{out}}{V_{in}} = \frac{A}{s^2 + s(\frac{1}{C_1} + \frac{1}{C_1}) + 1}$$

The damping factor is now given by

$$\alpha = \frac{1}{C_1} + \frac{1}{C_1}$$

$$\alpha = \frac{2}{C_1} \text{ or,}$$

$$C_1 = \frac{2}{\alpha}$$

Because ◈1=1/◈2 , we find

$$C_2 = \frac{\alpha}{2}$$

The finished prototype is shown in Figure 11.6.3 .

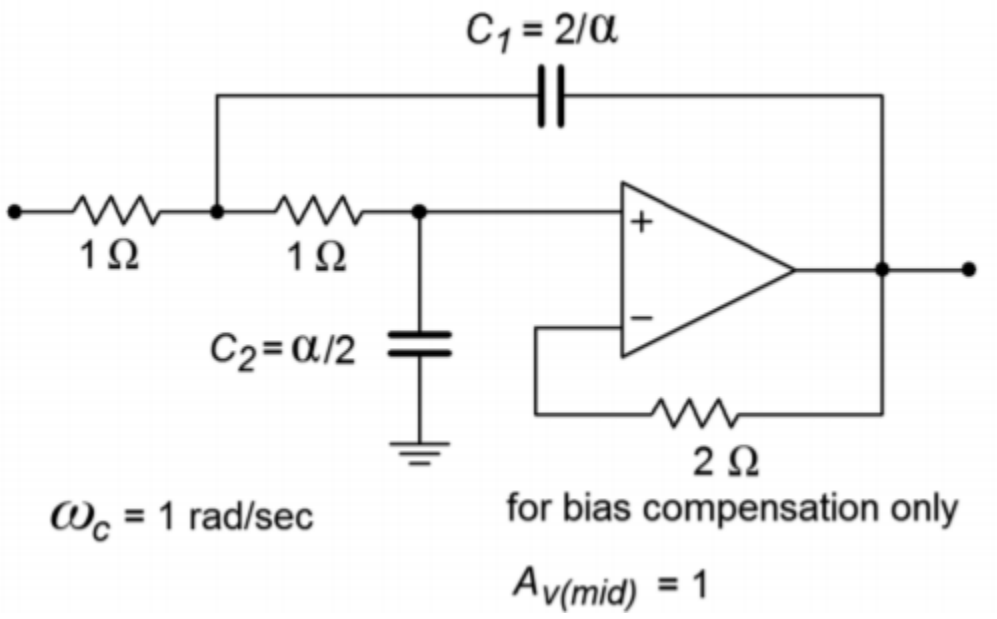


Figure 11.6.3 : Low-pass unity gain VCVS.

As you can see, there is quite a bit of similarity between the two versions. It is important to note that the inputs to these circuits must return to ground via a low-impedance DC path. If the signal source is capacitively coupled, the op amp's input bias current cannot be set up properly, and thus, some form of DC return resistor must be used at the source. Also, you can see that the damping factor (i.e., alignment) of the filter plays a role in setting component values. If the values shown are taken as having units of ohms and farads, the critical frequency will be 1 radian per second. It is an

accepted practice to normalize the basic forms of circuits such as these, so that the critical frequency works out to this convenient value. This makes it very easy to scale the component values to fit your desired critical frequency. Because the critical frequency is inversely proportional to the tuning resistor and capacitor values, you only need to shrink ◈ or ◈ in order to increase ◈◈ . (Remember, ◈◈=1/(2◈◈◈)) . A second scaling step normally follows this, in order to create practical values for ◈ and ◈ . This procedure is best shown with an example.

Example 11.6.1

Design a 1 kHz low-pass, second-order Butterworth filter. Examine both the equal-component and the unity-gain forms as drawn in Figure 11.6.2 and 11.6.3 , respectively. The required damping factor is 1.414.

Let's start with the equal-component version. First, find the required value for ◈◈ from the damping factor, as given on the diagram.

$$R_f = 2 - \alpha R_f = 21.414$$

[

$$R_f = 0.586\Omega$$

Note that this will produce a pass-band gain of

$$A_v = 1 + \frac{R_f}{R_i}$$

$$A_v = 1 + \frac{0.586}{1}$$

$$A_v = 1.586$$

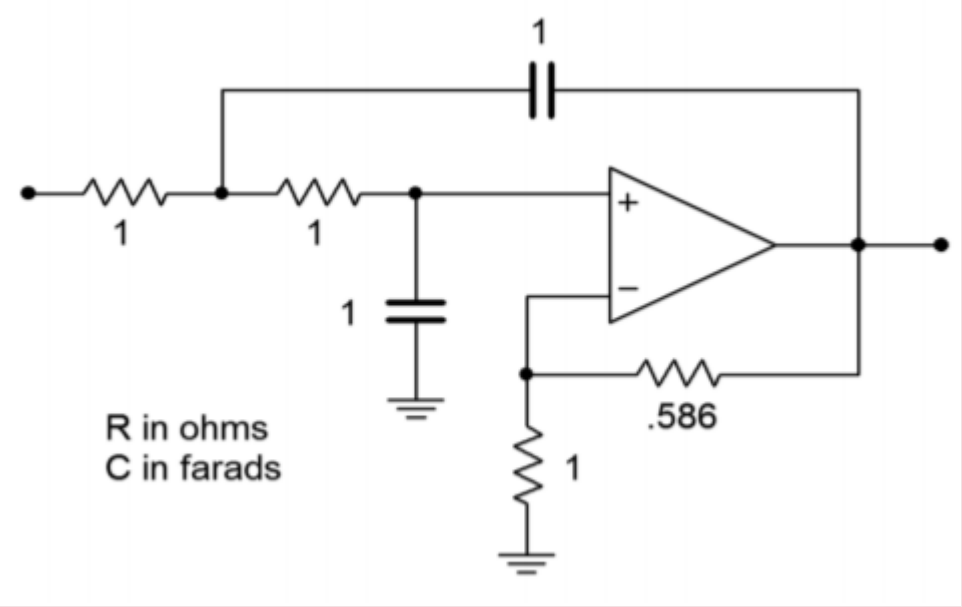


Figure 11.6.4 : Initial damping calculation for Example 11.6.1 (equal-component version).

Figure 11.6.4 shows the second-order Butterworth low-pass filter. Its critical frequency is 1 radian per second. We need to scale this to 1 kHz.

$$\omega_c = 2\pi f_c$$

$$\omega_c = 2\pi 1kHz$$

$$\omega_c = 6283 \text{ radians per second}$$

Our desired critical frequency is 6283 times higher than the normalized base. As ⬦⬦=1/⬦⬦ , to translate the frequency up, all we need to do is divide ⬦ or ⬦ by 6283. It doesn't really matter which one you choose, although it is generally easier to find "odd" sizes for resistors than capacitors, so we'll use ⬦ .

$$R = \frac{1}{6283}$$

$$R = 1.59 \times 10^{-4}\Omega$$

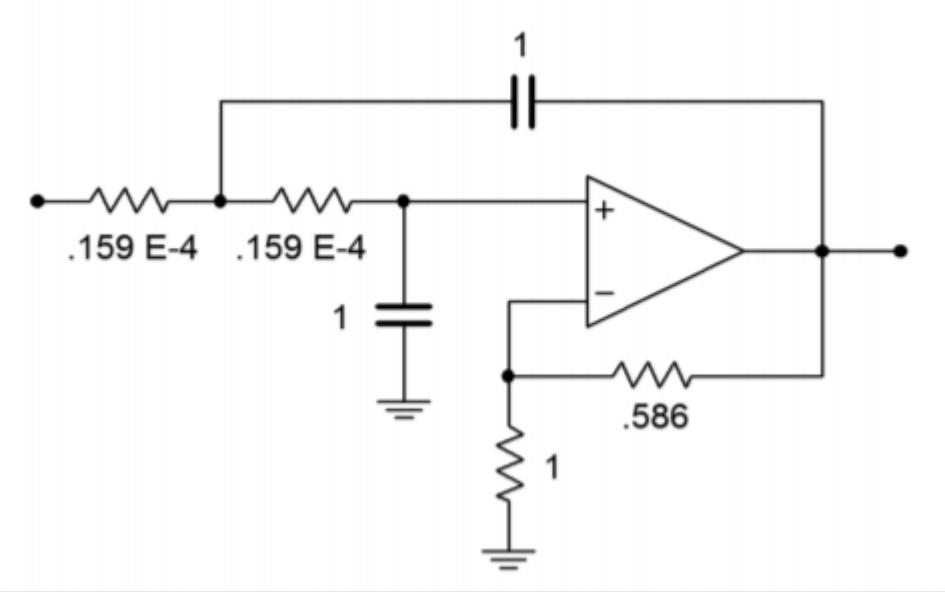


Figure 11.6.5 : The filter for Example 11.6.1 after frequency scaling (equal component version).

Figure 11.6.5 shows our 1 kHz, low-pass, second-order Butterworth filter. As you can see, though, the component values are not very practical. It is therefore necessary to perform the final scaling operations. First, consider multiplying ⬦⬦ and ⬦⬦ by 10 k. Note that this will have no effect on the damping, as it is the ratio of these two elements that determines damping. This scaling will not affect the critical frequency either, as ⬦⬦ is set by the tuning resistors and capacitors. Second, we need to increase ⬦ to a reasonable value. A factor of 107 will place it at 1.59⬦Ω . In order to compensate, the tuning capacitors must be dropped by an equal amount, which brings them to 100 nF. The completed design is shown in Figure 11.6.6 . Other scaling factors could also be used. Also, if bias compensation is important, the ⬦⬦ and ⬦⬦ values will need to be scaled further, in order to balance the resistance seen at the noninverting input.

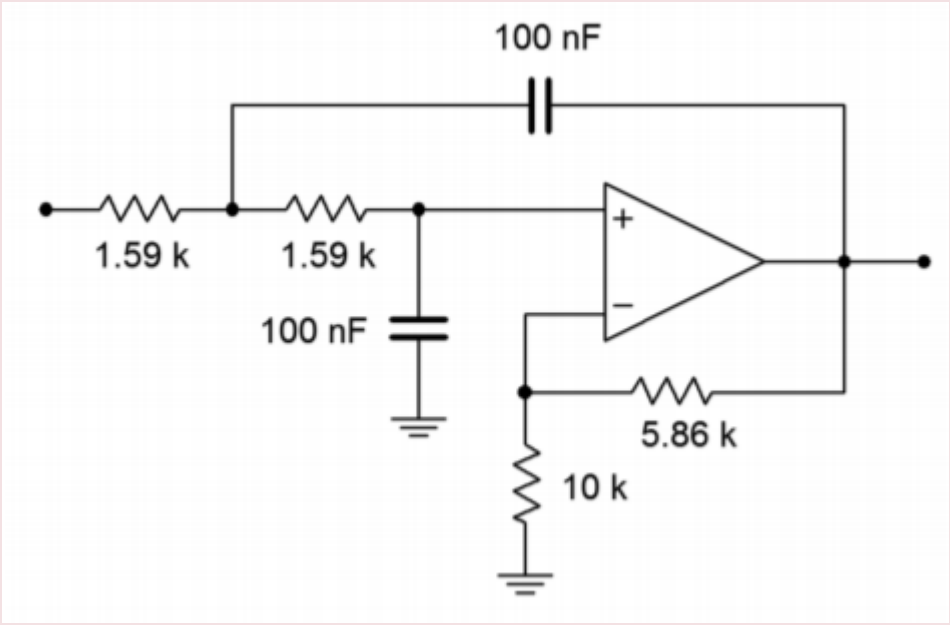


Figure 11.6.6 Final impedance scaling for Example 11.6.1 (equal component version).

The approach for the unity-gain version is similar. First, adjust the capacitor values in order to achieve the desired damping, as specified in Figure 11.6.3 .

$$C_1 = \frac{2}{\alpha}$$

$$C_1 = \frac{2}{1.414}$$

$$C_1 = 1.414F$$

$$C_2 = \frac{\alpha}{2}$$

$$C_2 = \frac{1.414}{2}$$

$$C_2 = 0.707F$$

The resulting circuit is shown in Figure 11.6.7 . The circuit must be scaled to the desired ⍰⍰ . The factor is 6283 once again, and the result is shown in Figure 11.6.8 . The final component scaling is seen in Figure 11.6.9 .

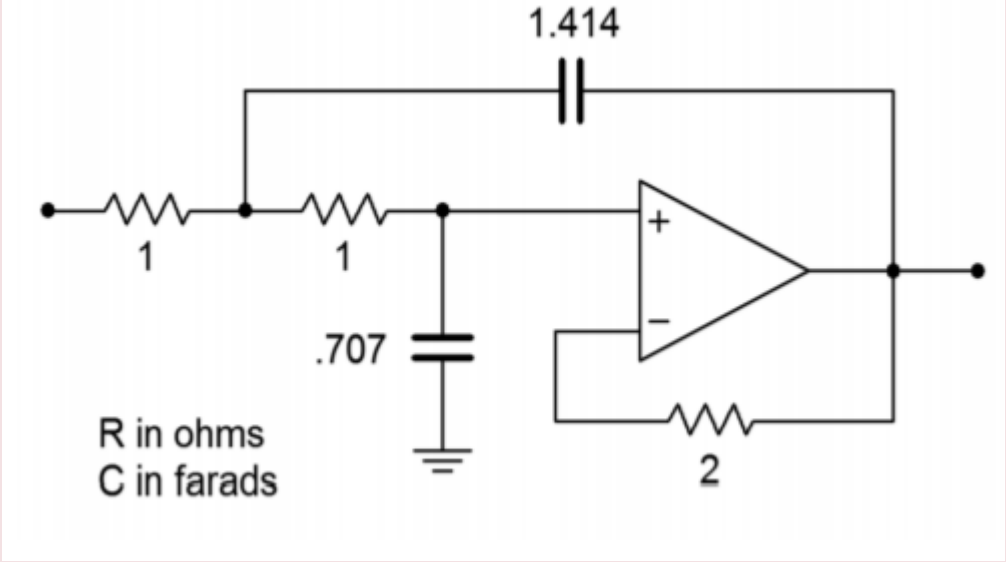


Figure 11.6.7 Initial damping calculation for Example 11.6.1 (unity-gain version).

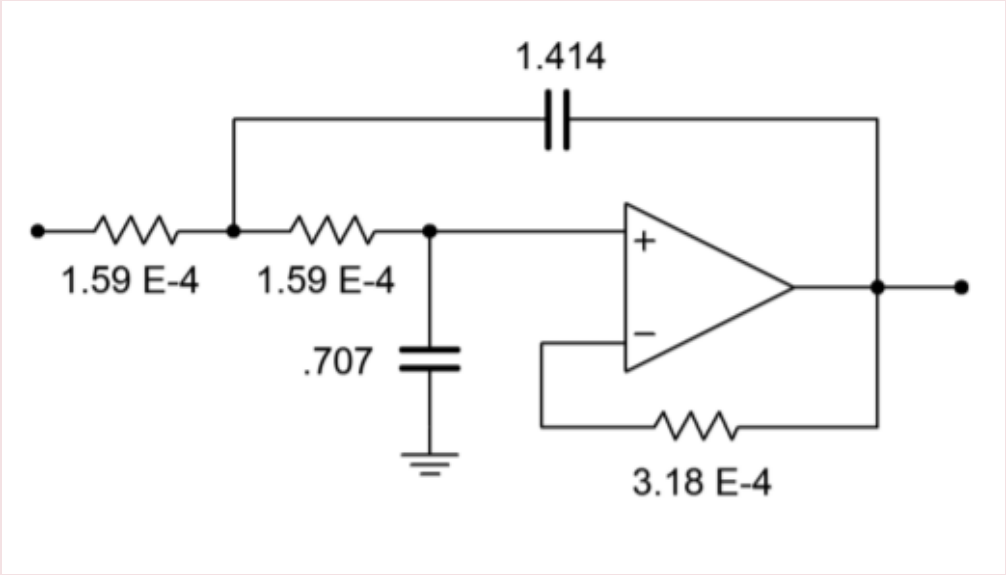


Figure 11.6.8 The filter for Example 11.6.1 after frequency scaling (unity-gain version).

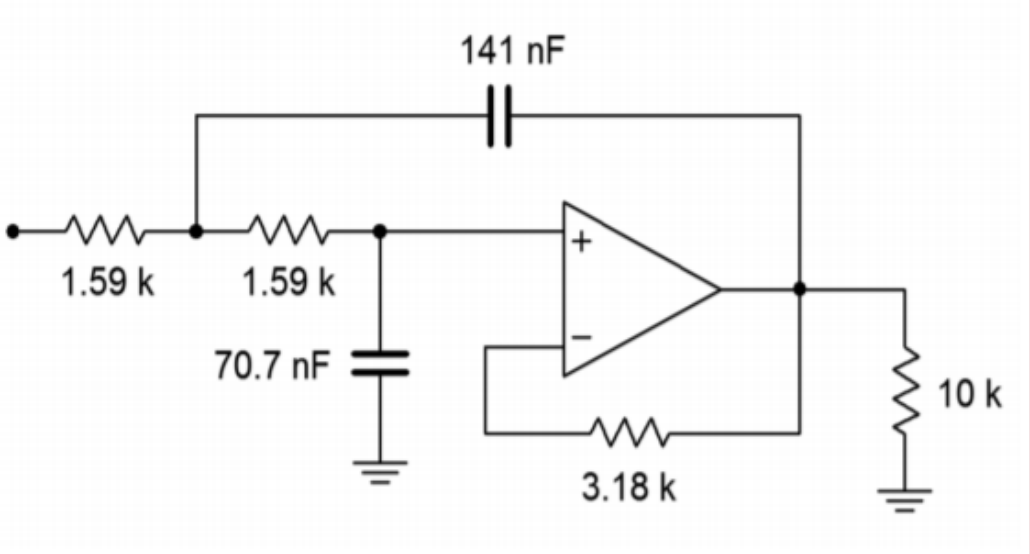


Figure 11.6.9 Final impedance scaling for Example 11.6.1 (unity gain version).

COMPUTER SIMULATION

A Multisim simulation of the filter design of Example 11.6.1 is shown in Figure 11.6.10 . The analysis shows the Bode plot, ranging from 50 Hz to better than 20 kHz. This yields over one decade on either side of the 1 kHz critical frequency. The graph clearly shows the –3 dB point at approximately 1 kHz, with an attenuation slope of –12 dB per octave. Since this is the unity-gain version, the low-frequency gain is set at 0 dB. Also, note that no peaking is evident in the response curve, as is expected for a Butterworth alignment. The phase response is also shown. Some graphing tools continue the phase shift below –180 degrees, whereas others will flip it back to +180, as Multisim does. Note that if the frequency plot range is extended, the phase shift starts to increase at the highest frequencies instead of leveling off. This is due to the extra phase shift produced by the op amp as the operating frequency approaches ◈◈◈◈◈ . A simpler op amp model would not create this real-world effect.

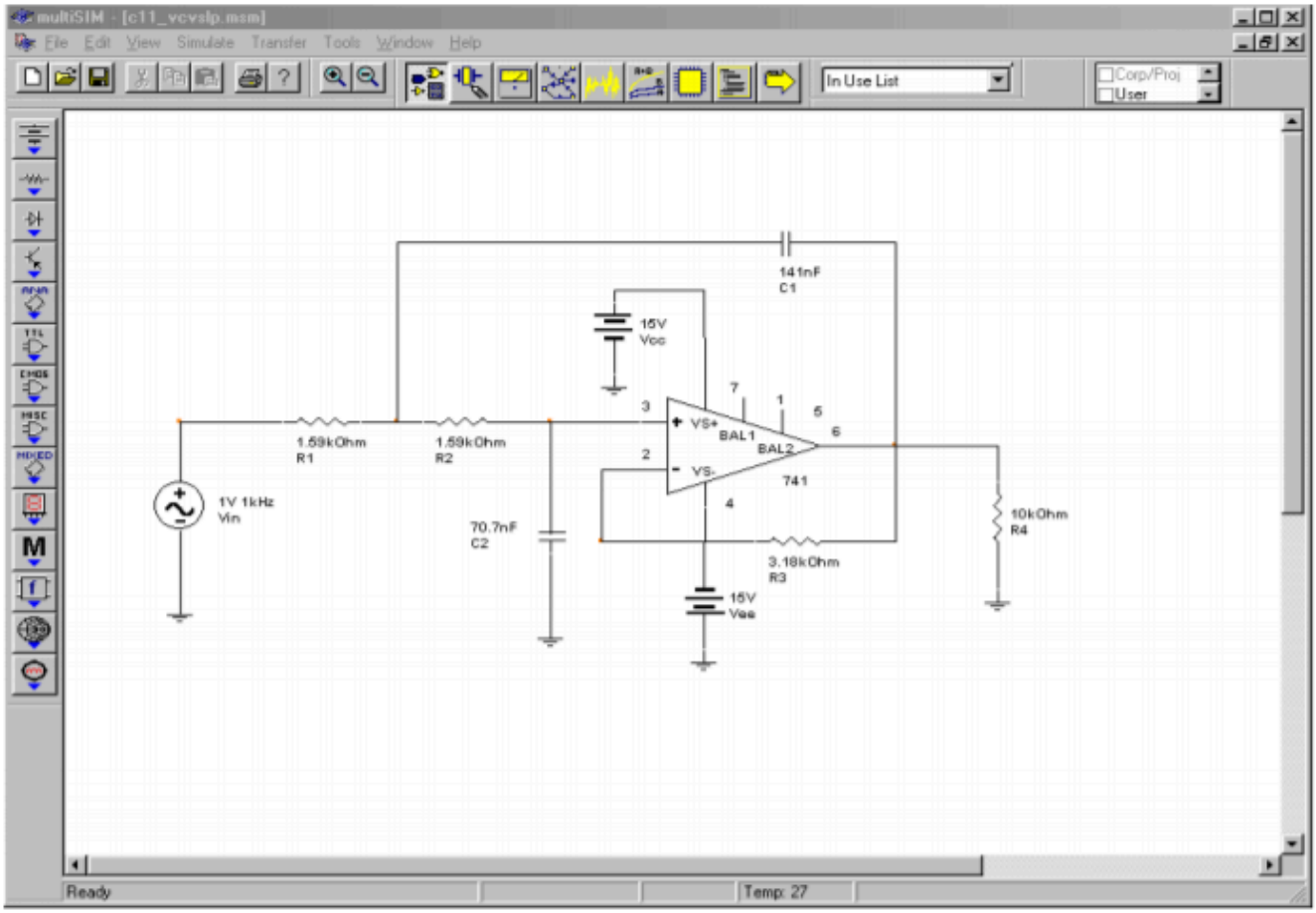


Figure 11.6.10 : VCVS low-pass filter in Multisim.

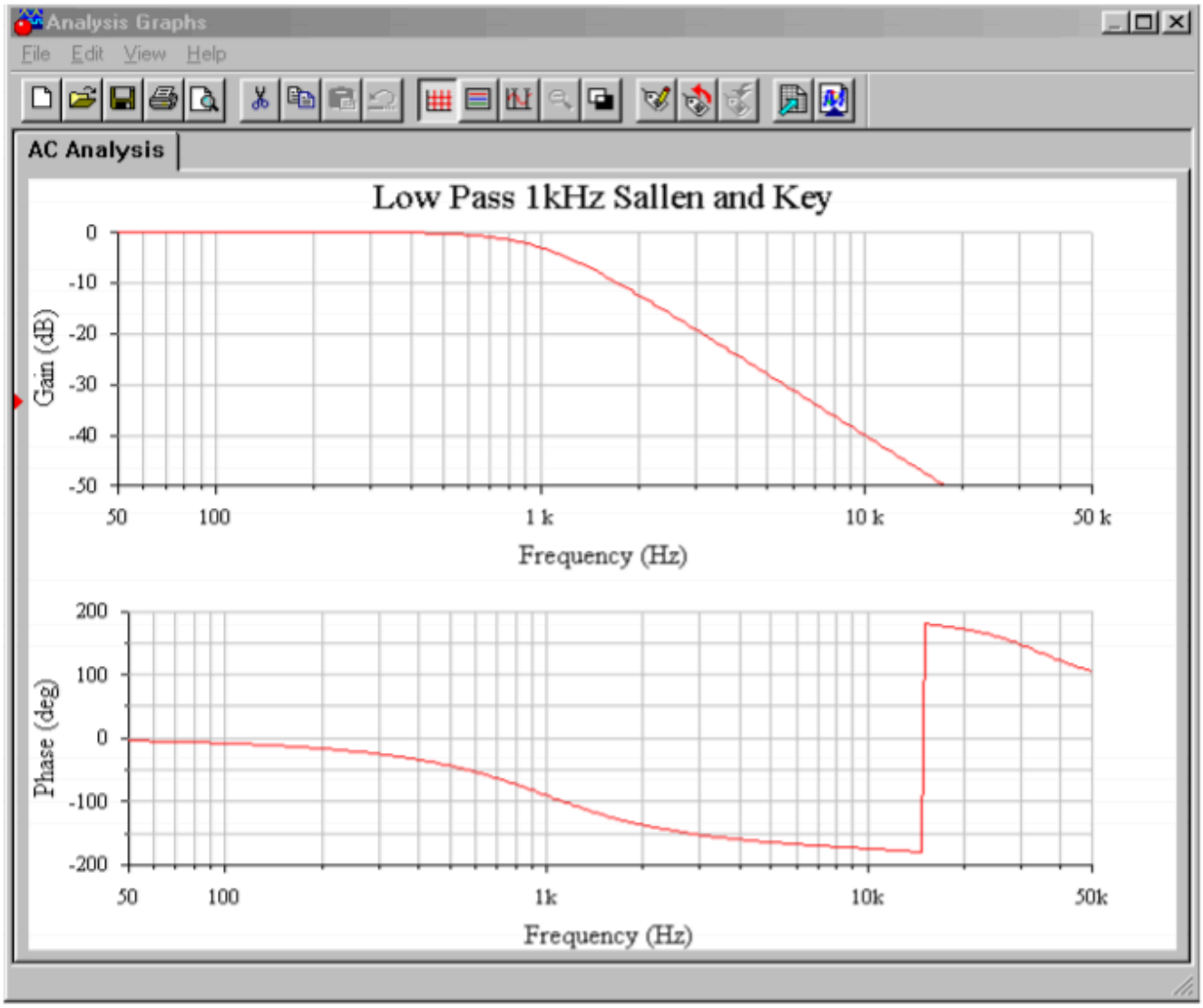


Figure Figure 11.6.10 : Gain and phase plots for VCVS low-pass filter.

Finally, a Monte Carlo analysis is used to mimic the effects seen due to component tolerance in a production environment (in this example, approximately 10 percent variation of nominal for each component). A total of 10 runs were generated. Although the overall shape of the curve remains consistent, there is some variation in the corner frequency, certainly more than the 10 percent offered by any single component. As you might guess, a Monte Carlo analysis is very tedious to do by hand, but quite straightforward to set up on a simulator.

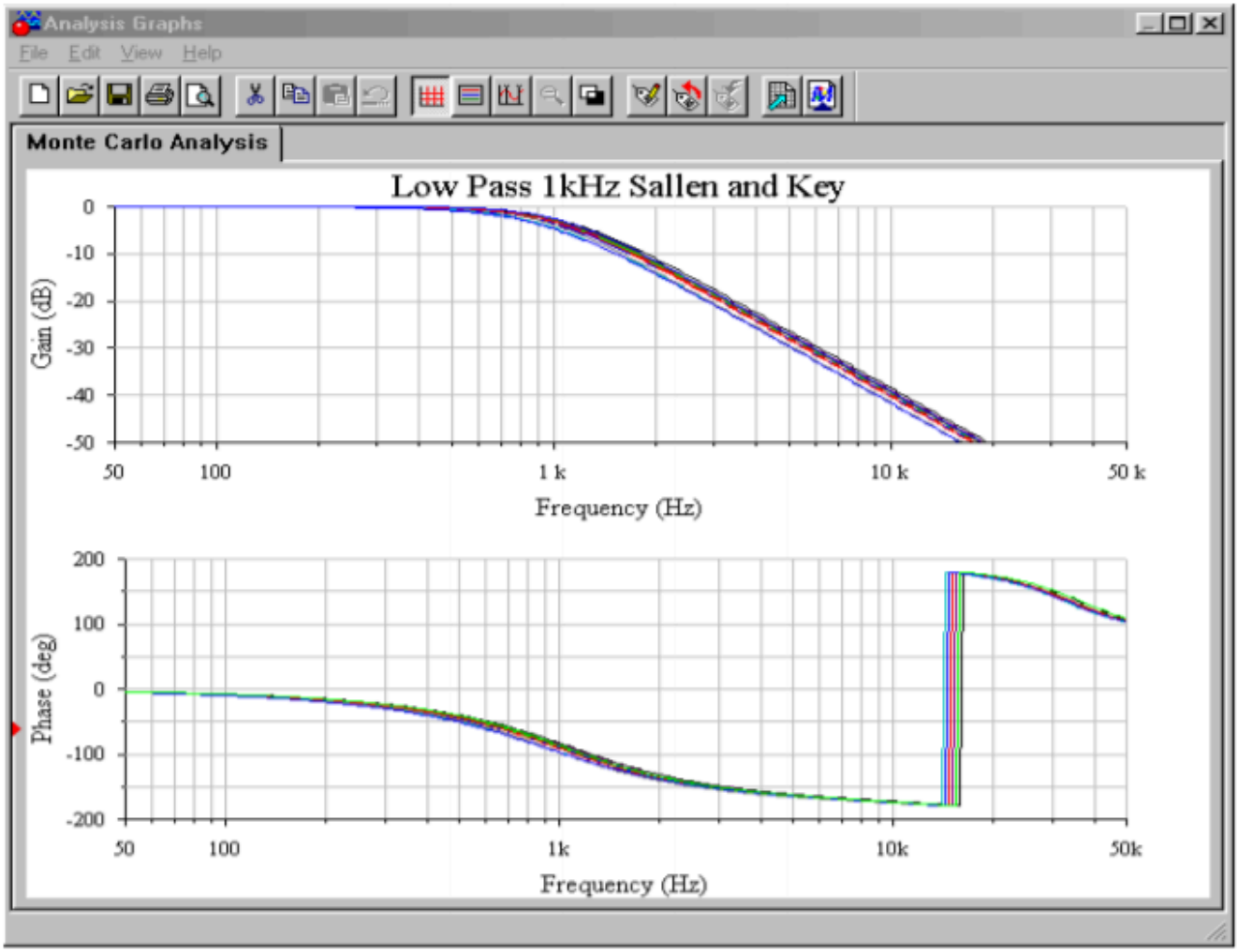


Figure 11.6.10 : Monte Carlo analysis for VCVS low-pass filter.

As you can see, the realization process is little more than a scaling sequence. This makes filter design very rapid. The operation for high-pass filters is essentially the same.

SALLEN AND KEY HIGH-PASS FILTERS

We can derive a general expression for high-pass filters, based on Equation 11.6.5 . A high-pass filter is a lead network, so to echo this, we will use capacitors for the first two elements and resistors for the third and forth. Using the operator, we find, 1=1/1 , 2=1/2 , 3=1 , and 4=2 .

$$\frac{V_{out}}{V_{in}} = \frac{A}{\frac{Z_1}{Z_4} + \frac{Z_1}{Z_3}(1-A) + \frac{Z_1 Z_2}{Z_3 Z_4} + \frac{Z_2}{Z_4} + 1}$$

$$\frac{V_{out}}{V_{in}} = \frac{A}{\frac{1}{sR_2C_1} + \frac{1}{sR_1C_1}(1-A) + \frac{1}{s^2R_1R_2C_1C_2} + \frac{1}{sR_2C_2} + 1}$$

$$\frac{V_{out}}{V_{in}} = \frac{As^2}{s^2 + s(\frac{1}{R_2C_1} + \frac{1}{R_2C_2} + \frac{1}{R_1C_1}(1A)) + \frac{1}{R_1R_2C_1C_2}}$$

(11.6.14)

As with the low-pass filters we have two basic realizations: equal-component and unity-gain. In both cases, we start with Equation 11.6.14 and use normalized frequency (1 radian per second). The derivations are very similar to the low-pass case, and the results are summarized below. The equal-component version:

$$\alpha = 3 - A$$

The unity-gain version:

$$R_2 = \frac{2}{\alpha}$$

$$R_1 = \frac{\alpha}{2}$$

These forms are shown in Figures Figure 11.6.11 and Figure 11.6.12 . You may at this point ask two questions: One, how do you find the damping factor, and two, what about higher-order filters?

In order to find the damping factor needed, a chart such as Figure Figure 11.6.13 may be consulted. This chart also introduces a new item, and that is the frequency factor, ⍰⍰ . Normally, the critical frequency and 3 dB down frequency (break frequency) of a filter are not the same value. They are identical only for the Butterworth alignment.

For any other alignment, the desired break frequency must first be translated to the appropriate critical frequency before scaling is performed. This is illustrated in the following example.

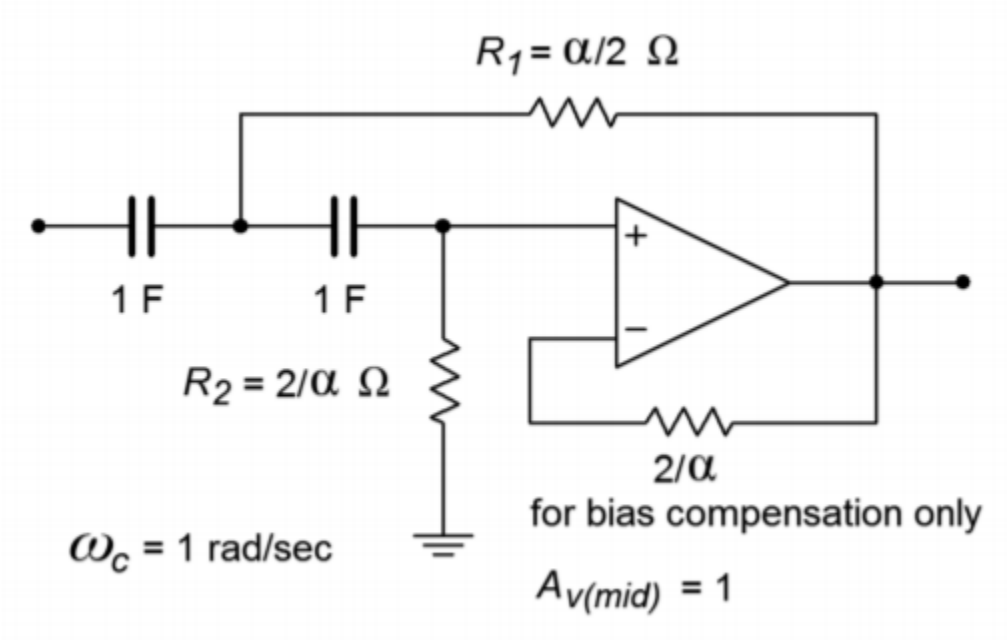


Figure 11.6.11 : High-pass unity-gain VCVS.

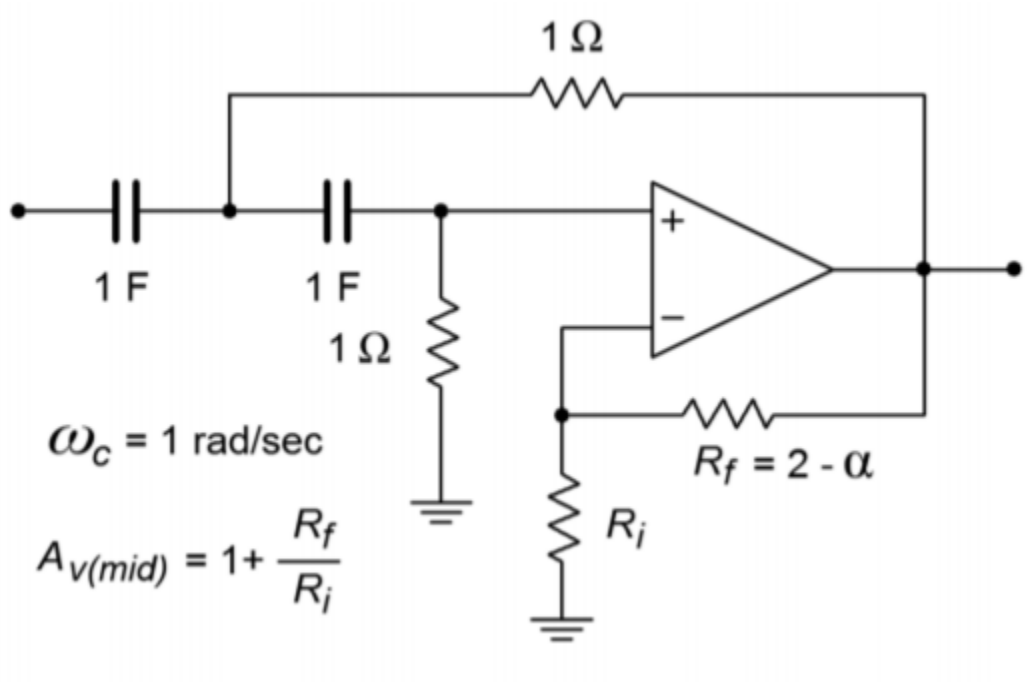


Figure 11.6.12 : High-pass equal-component VCVS.

Second-Order Filter Parameters

Type	α	k_f
Bessel	1.732	1.274
Butterworth	1.414	1.0
1 dB Chebyshev	1.045	0.863
2 dB Chebyshev	0.895	0.852
3 dB Chebyshev	0.767	0.841

$f_c = f_{3\,dB} / k_f$ for high-pass
$f_c = f_{3\,dB} \bullet k_f$ for low-pass

Figure 11.6.13 : Second-order filter parameters. From Lancaster, Don, Active Filter Cookbook, Second edition, Newnes1996. Reprinted with permission

Example 11.6.2

Design a second-order, high-pass Bessel filter, with a break frequency (◈3◈◈) of 5 kHz.

For this example, let's use the unity-gain form shown in Figure Figure 11.6.11 . First, obtain the damping and frequency factors from Figure Figure 11.6.13 .

$$k_f = 1.274, \text{ damping } = 1.732.$$

Using the damping factor, the two tuning resistors may be found:

$$R_1 = \frac{\alpha}{2}$$

$$R_1 = \frac{1.732}{2}$$

$$R_1 = .866\Omega$$

$$R_2 = \frac{2}{\alpha}$$

$$R_2 = \frac{2}{1.732}$$

$$R_2 = 1.155\Omega$$

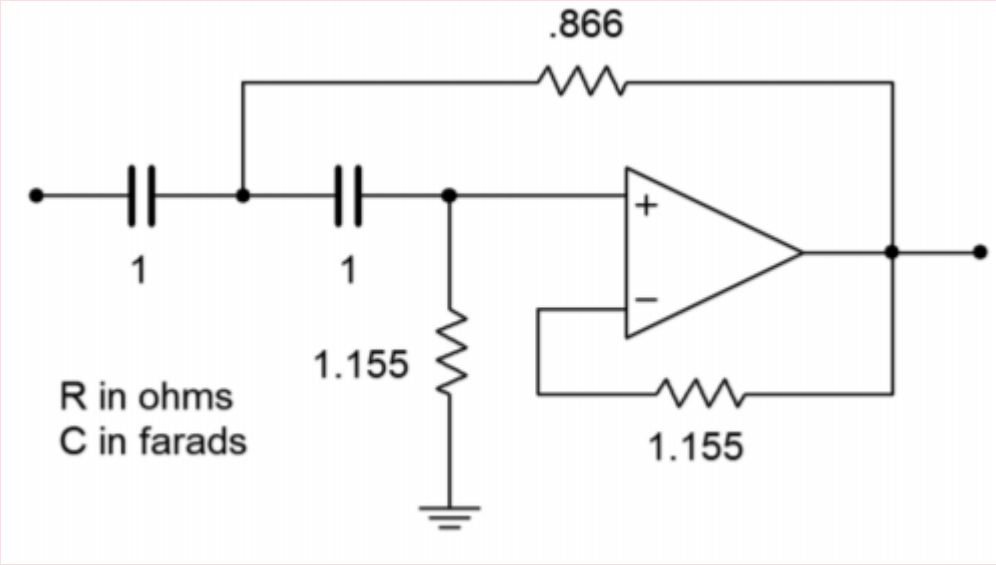


Figure 11.6.14 : Initial damping calculation for Example 11.6.2 .

The intermediate result is shown in Figure Figure 11.6.14 . In order to do the frequency scaling, the desired break frequency of 5 kHz must first be translated into the required critical frequency. Because this is a high-pass filter,

$$f_c = \frac{f_{3dB}}{k_f}$$

$$f_c = \frac{5kHz}{1.274}$$

$$f_c = 3925Hz$$

$$\omega_c = 2\pi f_c$$

$$\omega_c = 2\pi 3925$$

$$\omega_c = 24.66k \text{ radians per second}$$

Either the tuning resistors or capacitors may now be scaled.

$$R_1 = \frac{.866}{24.66k}$$

$$R_1 = 3.51 \times 10^{-5}\Omega$$

$$R_2 = \frac{1.155}{24.66k}$$

$$R_2 = 4.68 \times 10^{-5}\Omega$$

We now have a second-order, high-pass, 5 kHz Bessel filter. This is shown in Figure Figure 11.6.15 . A final scaling of 108 will give us reasonable values, and is shown in Figure Figure 11.6.16 .

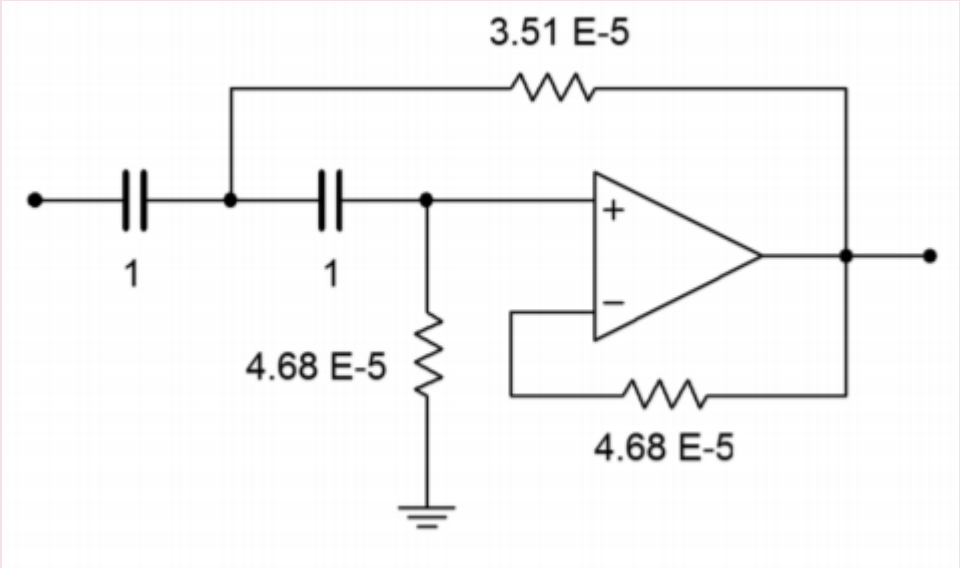


Figure 11.6.15 : Frequency scaling for Example 11.6.2 .

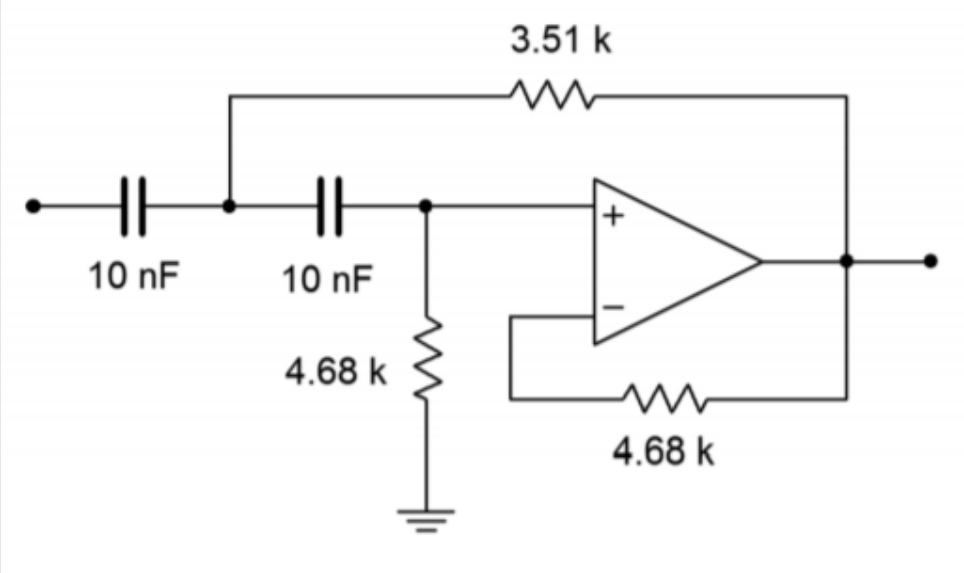


Figure Figure 11.6.16 : Impedance scaling for Example 11.6.2 .

FILTERS OF HIGHER ORDER

There is a common misconception among novice filter designers that higher-order filters may be produced by cascading a number of lower-order filters of the same type. This is not true. For example, cascading three second-order 10 kHz Butterworth filters will not} produce a sixth-order 10 kHz Butterworth filter. A quick inspection reveals why this is not the case: A single filter of any order will show a 3 dB loss at its break frequency by definition (in this case, 10 kHz). If three filters of the same type are cascaded, each filter will produce a 3 dB loss at the break frequency, which means an overall loss of 9 dB occurs. This much is true: a higher order filter will require a number of individual sections, each with specific damping and frequency factors. Each section will be based on the second-order forms already examined.[1] In order to make odd-ordered filters, we will introduce a simple single-pole filter. The high- and low-pass versions of this unit are shown in Figure Figure 11.6.17 . The damping factor for this circuit is always unity. When working with it, you need only worry about the frequency factor.

1. The transfer function of a higher-order filter contains a high-order polynomial in the denominator of the form ⬦⬦+⬦⬦–1⬦⬦–1+⋯+⬦1⬦+⬦0 . The ⬦ indicates the order of the filter and the ⬦ coefficients determine the alignment. This polynomial is factored into a product of second-order expressions (with a possible first-order unit for odd-ordered systems). Each of these expressions corresponds to a single section in the larger filter.

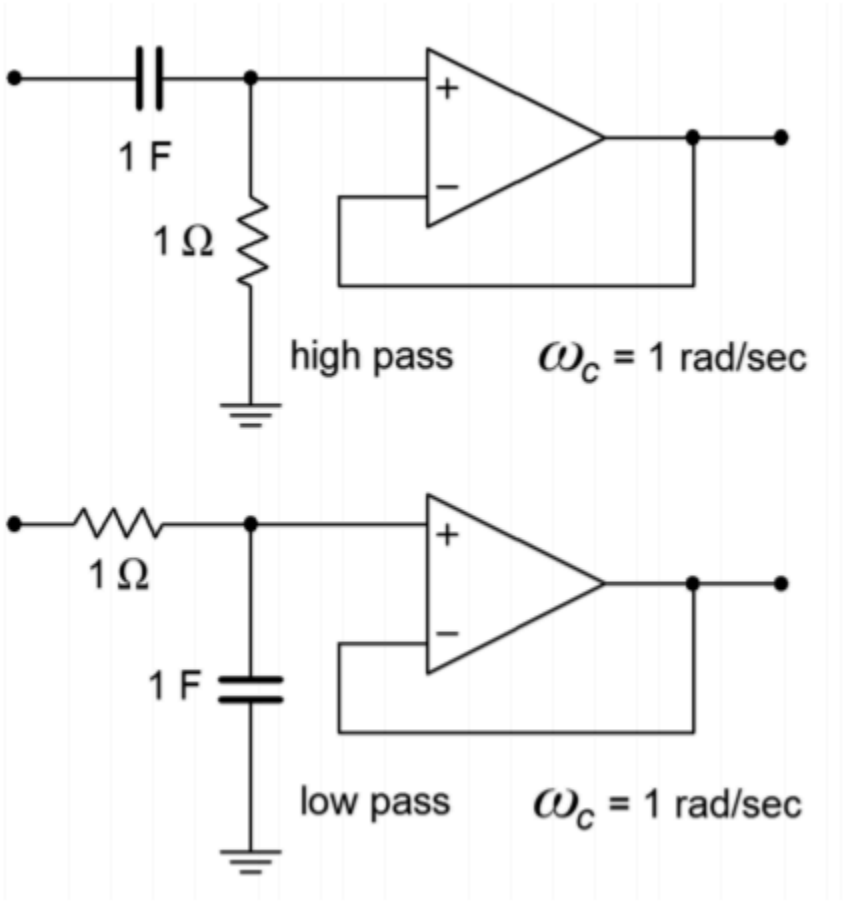


Figure 11.6.17 : Single-pole sections.

Designing higher-order filters is, conceptually, no different than designing second-order filters. The reality is that new charts are needed for the required damping and frequency factors. A set of compatible charts is shown in Figure Figure 11.6.18 for orders 3 through 6 on the following pages. An example is then presented to show the design sequence flow.

Third-Order Filter Parameters

	Second-order section		First-order section
Type	**α**	**k_f**	**k_f**
Bessel	1.447	1.454	1.328
Butterworth	1.0	1.0	1.0
1 dB Chebyshev	0.496	0.911	0.452
2 dB Chebyshev	0.402	0.913	0.322
3 dB Chebyshev	0.326	0.916	0.299

Fourth-Order Filter Parameters

	Second-order section		Second-order section	
Type	**α**	**k_f**	**α**	**k_f**
Bessel	1.916	1.436	1.241	1.610
Butterworth	1.848	1.0	0.765	1.0
1 dB Chebyshev	1.275	0.502	0.281	0.943
2 dB Chebyshev	1.088	0.466	0.224	0.946
3 dB Chebyshev	0.929	0.443	0.179	0.950

Fifth-Order Filter Parameters

	Second-order section		Second-order section		First-order section
Type	**α**	**k_f**	**α**	**k_f**	**k_f**
Bessel	1.775	1.613	1.091	1.819	1.557
Butterworth	1.618	1.0	0.618	1.0	1.0
1 dB Chebyshev	0.714	0.634	0.180	0.961	0.280
2 dB Chebyshev	0.578	0.624	0.142	0.964	0.223
3 dB Chebyshev	0.468	0.614	0.113	0.967	0.178

Sixth-Order Filter Parameters

	Second-order section		Second-order section		Second-order section	
Type	**α**	**k_f**	**α**	**k_f**	**α**	**k_f**
Bessel	1.959	1.609	1.636	1.694	0.977	1.910
Butterworth	1.932	1.0	1.414	1.0	0.518	1.0
1 dB Chebyshev	1.314	0.347	0.455	0.733	0.125	0.977
2 dB Chebyshev	1.121	0.321	0.363	0.727	0.0989	0.976
3 dB Chebyshev	0.958	0.298	0.289	0.722	0.0782	0.975

$f_c = f_{3\,dB} / k_f$ for high-pass $f_c = f_{3\,dB} \bullet k_f$ for low-pass

Figure 11.6.18 : Filter design tools Parameters for third- through sixth-order filters. From Lancaster, Don, Active Filter Cookbook, Second edition, Newnes1996. Reprinted with permission

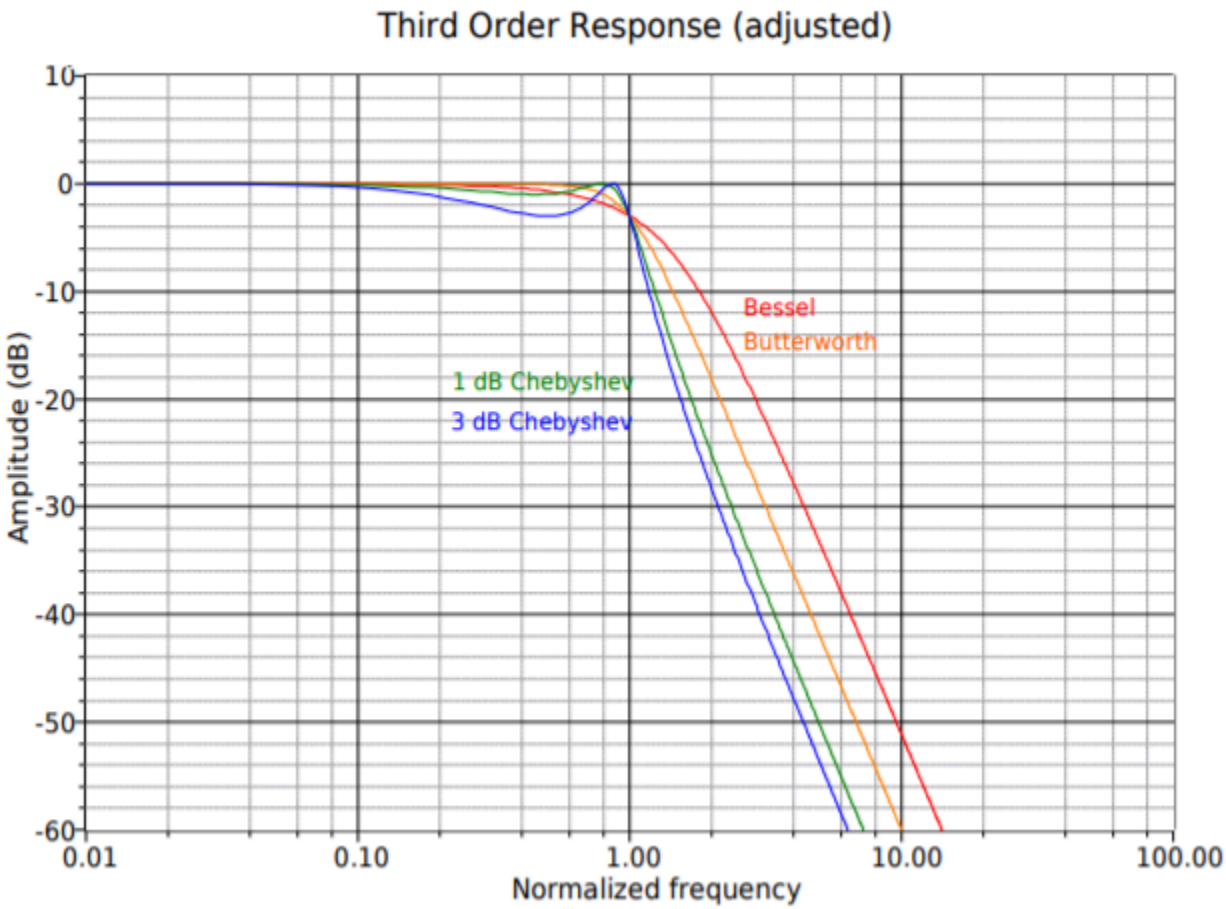


Figure 11.6.18 : Filter design tools, continued Third-order response (adjusted).

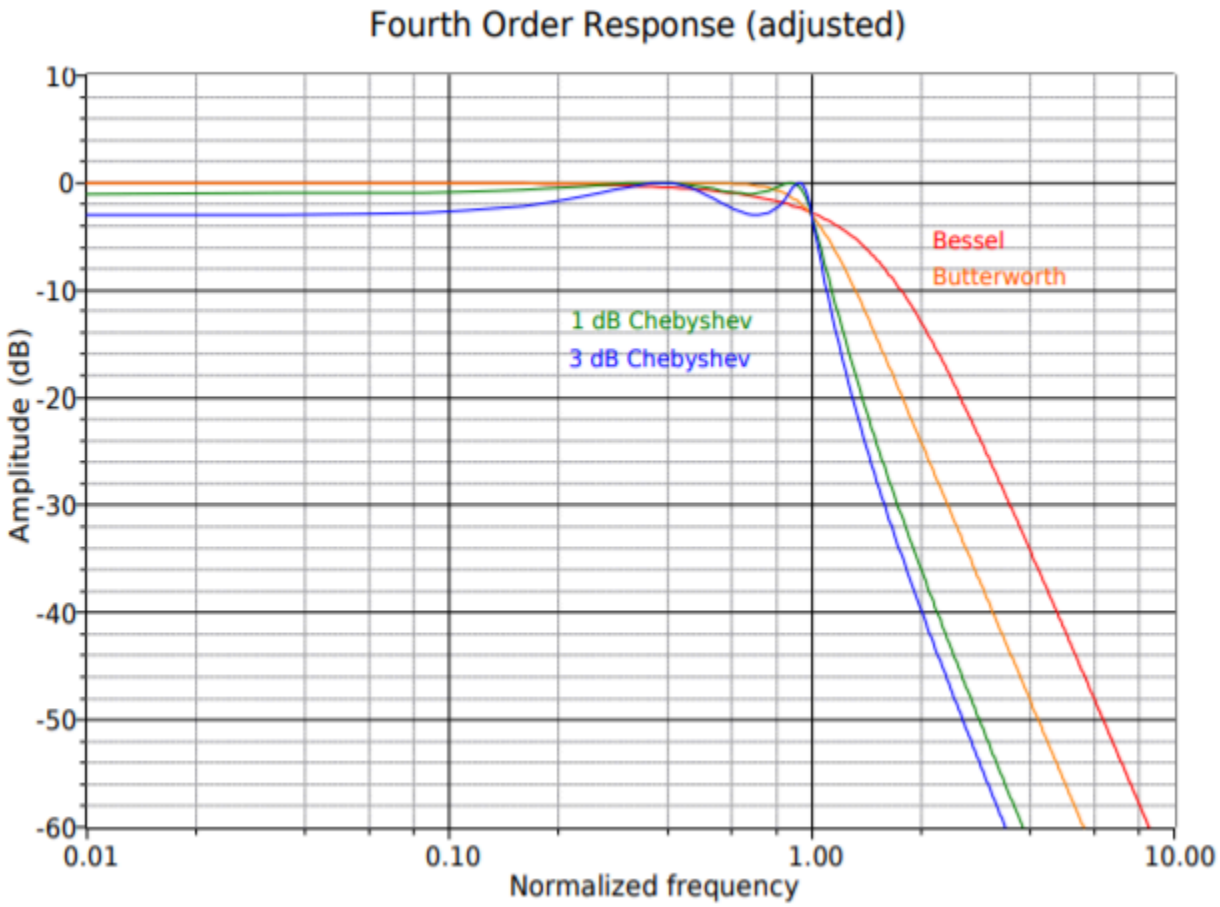


Figure 11.6.18 : Filter design tools, continued Fourth-order response (adjusted).

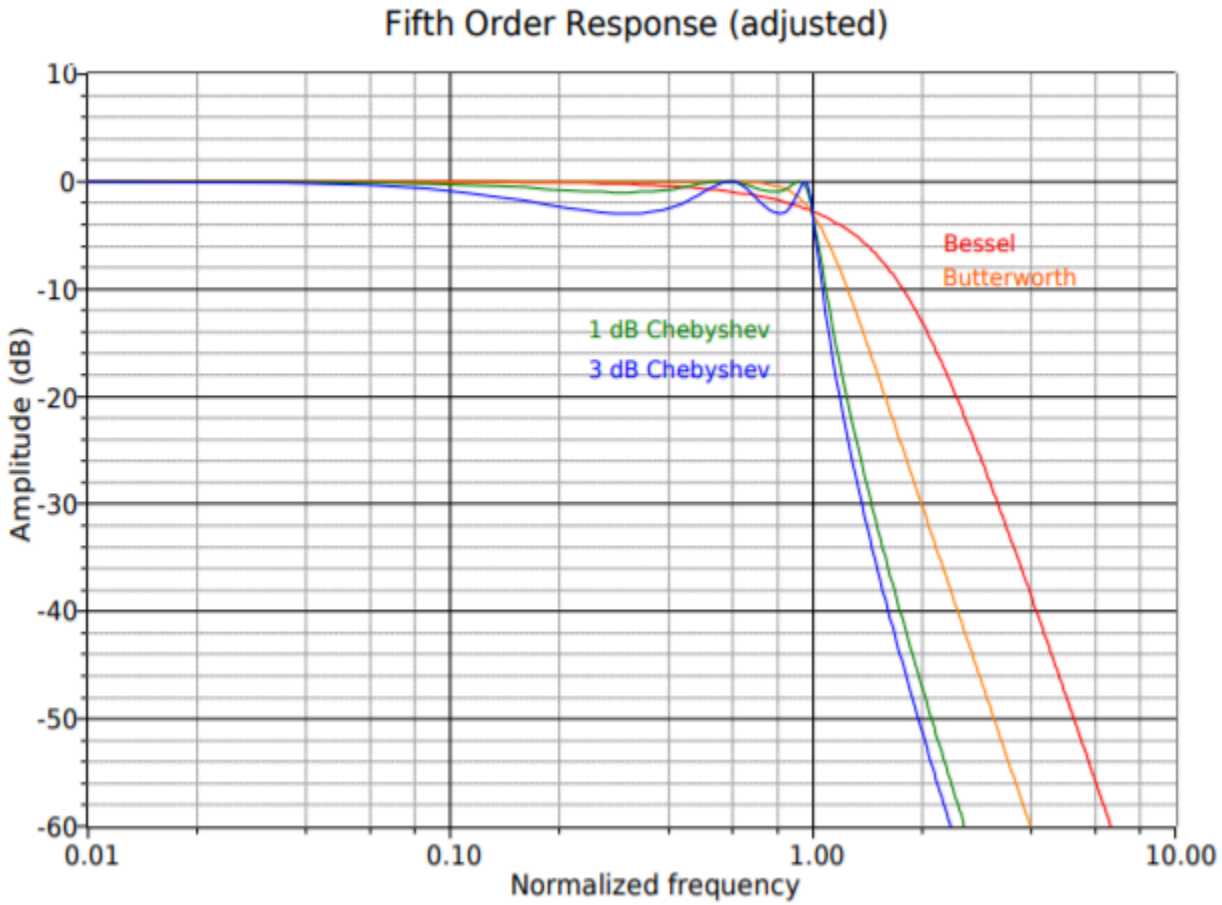


Figure 11.6.18 : Filter design tools, continued Fifth-order response (adjusted).

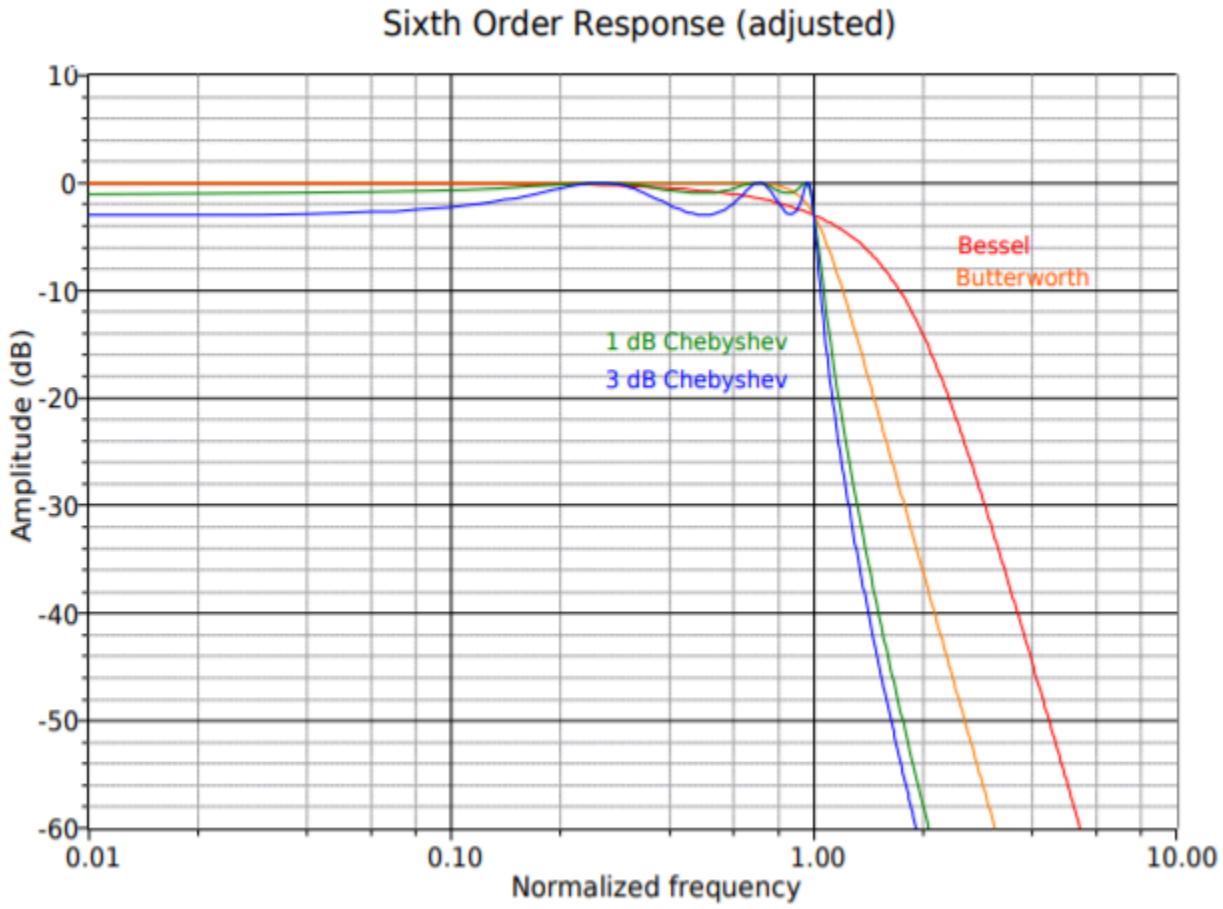


Figure 11.6.18 : Filter design tools, continued Sixth-order response (adjusted).

Example 11.6.3

We wish to design a filter suitable for removing subsonic tones from a stereo system. This could be used to reduce turntable rumble in a vintage hi-fi or DJ system, or to reduce stage vibration in a public address system. The filter should attenuate frequencies below the lower limit of human hearing (about 20 Hz), while allowing all higher frequencies to pass. Transient response may be important here, so we'll choose a Bessel alignment. We will also specify a fifth-order system. This will create an attenuation of about 15 dB, one octave below the break frequency.

First, note that the specification requires the use of a high-pass filter. This filter may be realized with either the equal-component or the unity-gain forms. As this design will require multiple sections, excessive gain may result from the equal-component version. Our fifth-order system will be comprised of two second-order sections and a first-order section. An overview of the design is shown in Figure 11.6.19 , with the appropriate damping and frequency factors taken from Figure 11.6.18 . We'll break the analysis down stage by stage.

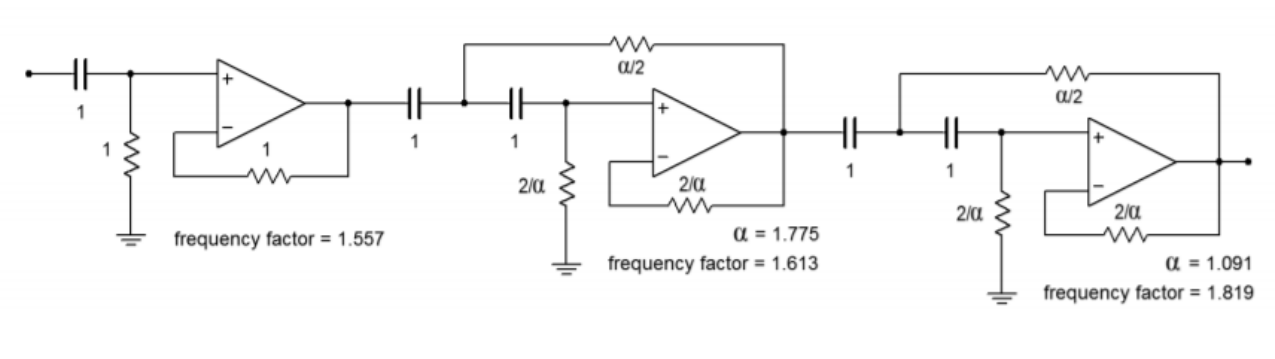


Figure 11.6.19 : Circuit outline (Note: all values in ohms and farads).

First, find the desired break frequency in radians.

$$\omega_{3dB} = 2\pi f_{3dB}$$
$$\omega_{3dB} = 2\pi 20Hz$$
$$\omega_{3dB} = 125.7 \text{ radians per second}$$

Stage 1:

The break frequency must be translated to the required critical frequency. Because this is a high-pass filter, we need to divide by the frequency factor.

$$\omega_c = \frac{\omega_{3dB}}{k_f}$$

$$\omega_c = \frac{125.7}{1.557}$$

$$\omega_c = 80.7 \text{ radians per second}$$

We will scale the resistor by 80.7 to achieve this tuning frequency.

$$R = \frac{1}{80.7}$$

$$R = .0124\Omega$$

The ◈ and ◈ values must now be scaled for practical values. A factor of 106 would be reasonable. The final result is

$$R = 12.4k$$
$$C = 1\mu F$$

Stage 2:

First, determine the values for the two resistors from the given damping factor.

$$R_1 = \frac{\alpha}{2}$$

$$R_1 = \frac{1.775}{2}$$

$$R_1 = 0.8875\Omega$$

$$R_2 = \frac{2}{\alpha}$$

[

$$R_2 = \frac{2}{1.775}$$

$$R_2 = 1.127\Omega$$

Now the break frequency must be translated to the required critical frequency.

$$\omega_c = \frac{\omega_{3dB}}{k_f}$$

$$\omega_c = \frac{125.7}{1.613}$$

$$\omega_c = 77.9 \text{ radians per second}$$

We will scale the resistors by 77.9 to achieve this tuning frequency.

$$R_1 = \frac{0.8875}{77.9}$$

$$R_1 = 0.0114\Omega$$

$$R_2 = \frac{1.127}{77.9}$$

$R_2 = 0.0145\Omega Again, and must be scaled for practical values. A factor of 106 would be reasonable. The final results 1 = 11.42 = 14.5 = 1 Stage 3 : First, determine the values for the two resistors from the given damping factor. R_1 = \frac{\alpha}{2}$

$$R_1 = \frac{1.091}{2}$$

$$R_1 = 0.5455\Omega$$

$$R_2 = \frac{2}{\alpha}$$

$$R_2 = \frac{2}{1.091}$$

$$R_2 = 1.833\Omega$$

Now the break frequency must be translated to the required critical frequency.

$$\omega_c = \frac{\omega_{3dB}}{k_f}$$

$$\omega_c = \frac{125.7}{1.819}$$

$$\omega_c = 69.1 \text{ radians per second}$$

We will scale the resistors by 69.1 to achieve this tuning frequency.

$$R_1 = \frac{.5455}{69.1}$$

$$R_1 = 7.89 \times 10^{-3}\Omega$$

$$R_2 = \frac{1.833}{69.1}$$

$$R_2 = 0.0265\Omega$$

Again, ◈ and ◈ must be scaled for practical values. A factor of 106 would be reasonable. The final result is

$$R_1 = 7.89k\Omega$$
$$R_2 = 26.5k\Omega$$
$$C = 1\mu F$$

The complete design is shown in Figure 11.6.20 . Note that all of the capacitors are set at 1◈◈ . This certainly helps to cut inventory and parts placement

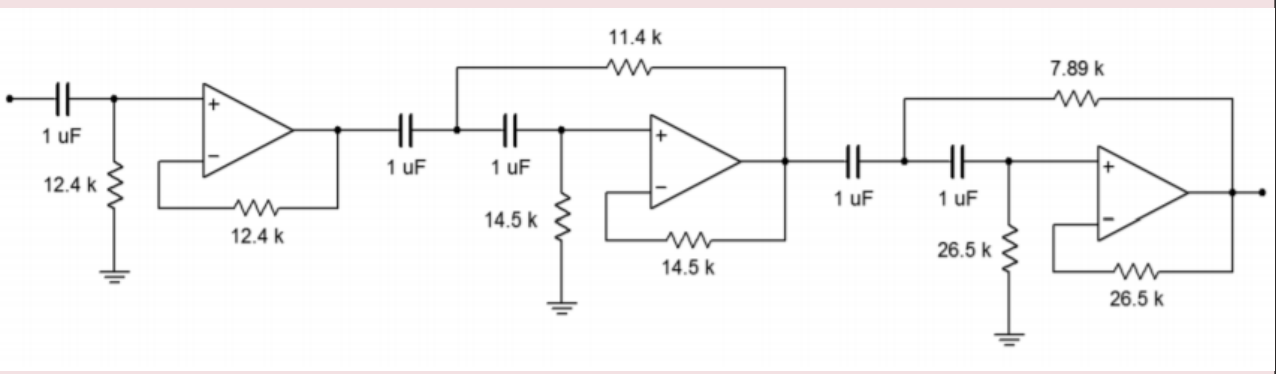


Figure 11.6.20 Completed filter for Example 11.6.3 .

We will complete our discussion of high- and low-pass VCVS filters with the following example.

Example 11.6.4

As mentioned earlier, it is common for loudspeaker systems to rely on passive filters to create their crossover networks. More demanding applications such as recording studio monitoring or large public-address systems (i.e., concert systems) cannot afford the losses associated with passive crossovers.

Instead, these applications utilize active crossovers composed of active filters, such as the one shown in Figure 11.6.21 . Before the audio signal is fed to a power amplifier, it is split into two or more frequency bands. The resulting signals each feed their own power amplifier/loudspeaker section.

Figure 11.6.21 : Commercial electronic crossover. Reprinted courtesy of Furman Sound, Inc.

A block diagram of this approach is shown in Figure 11.6.22 . This one is a two-way system. Large concert sound reinforcement systems may break the audio spectrum into four or five segments. The resulting system will be undoubtedly expensive, but will show lower distortion and higher output levels than a passively-crossed system. A typical two-way system might be crossed at 800 Hz. In other words, frequencies above 800 Hz will be sent to a specialized high frequency-transducer, whereas frequencies below 800 Hz will be sent to a specialized low-frequency transducer. In essence, the crossover network is a combination of an 800 Hz low-pass filter, and an 800 Hz high-pass filter. The filter order and alignment vary considerably depending on the application. Let's design an 800 Hz crossover with secondorder Butterworth filters.

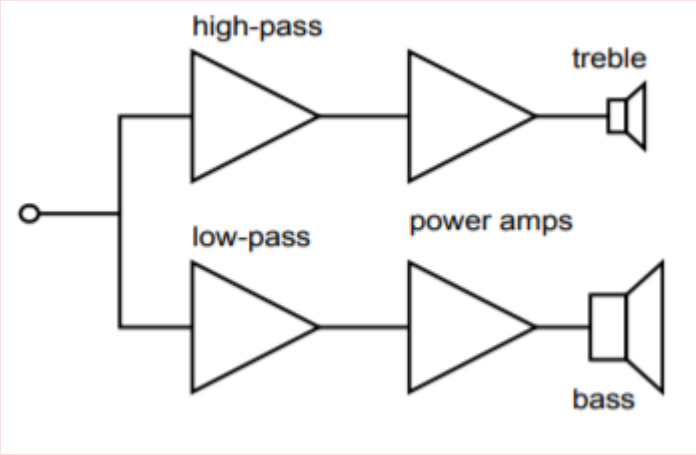


Figure 11.6.22 : Electronic crossover system.

The basic circuit layout is shown in Figure 11.6.23 . We're using the equalcomponent value version here. Exact gain is normally not a problem in this case, as some form of volume control needs to be added anyway, in order to compensate between the sensitivity of the low and high frequency transducers. (This is most easily produced by adding a simple voltage divider/potentiometer at the output of the filters.)

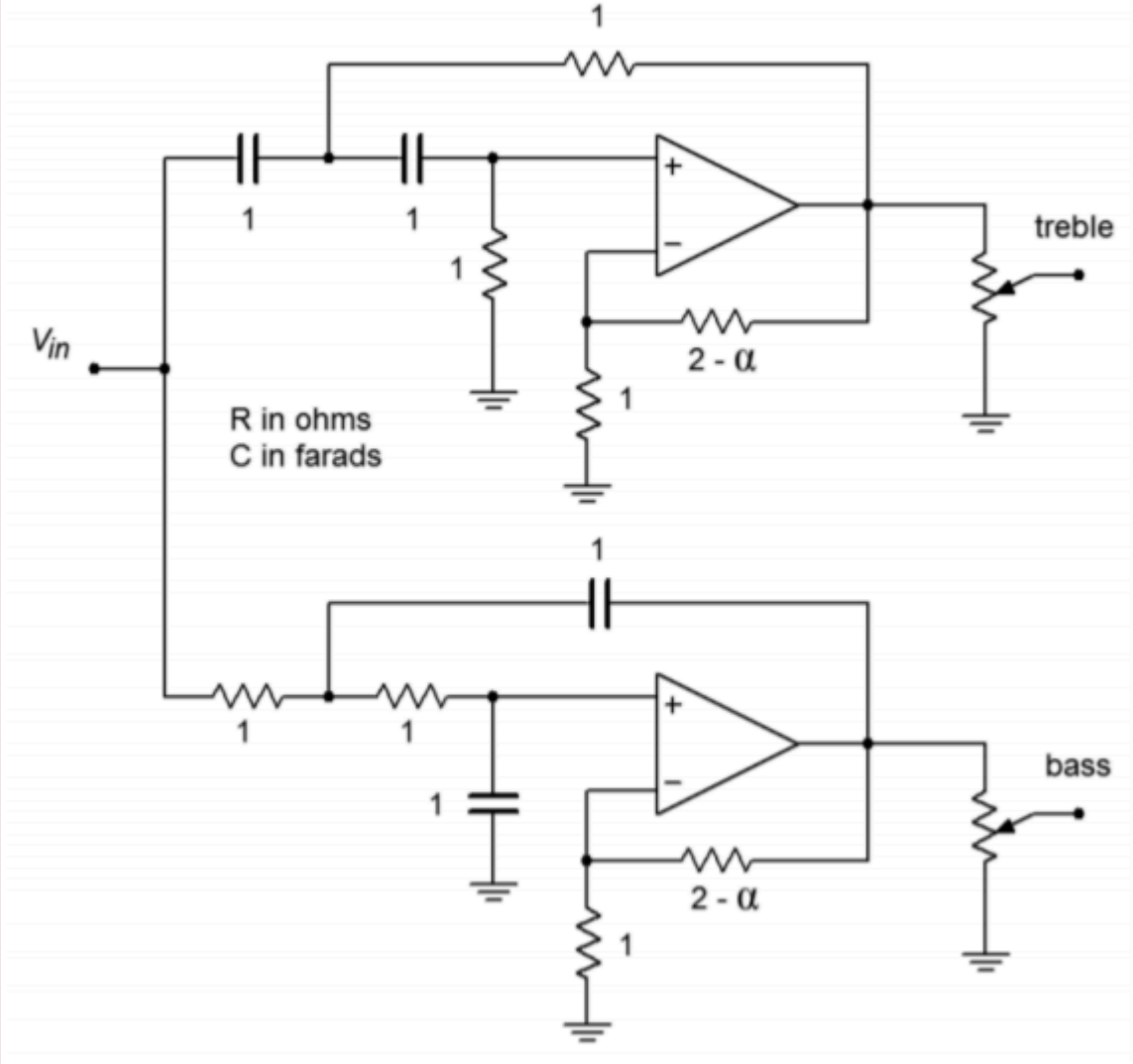


Figure 11.6.23 : Basic filter sections for crossover of Example 11.6.4 .

For second-order Butterworth filters, the damping factor is found to be 1.414, and the frequency factor is unity (indicating that ◈◈ and ◈3◈◈ are the same). Note that the design for both halves is almost the same. Both sections show an ◈◈ of 800 Hz, and a damping factor of 1.414. With identical characteristics, it follows that the component values will be the same in both circuits.

The required value for ◈◈ is

$$R_f = 2\alpha$$

$$R_f = 21.414$$

$$R_f = 0.586\Omega$$

The critical frequency in radians is

$$\omega_c = 2\pi f_c$$

$$\omega_c = 2\pi 800 Hz$$

$$\omega_c = 5027 \text{ radians per second}$$

Again, we will scale the tuning resistors to yield

$$R = \frac{1}{5027}$$

$$R = 1.989 \times 10^{-4}\Omega$$

A final ◈◈ scaling by 108 produces

$$R = 19.9k\Omega$$

$$C = 10nF$$

◈◈ and ◈◈ are scaled by 10◈ , and 10◈Ω log taper potentiometers may be used for the output volume trimmers. The resulting circuit is shown in Figure 11.6.24 .

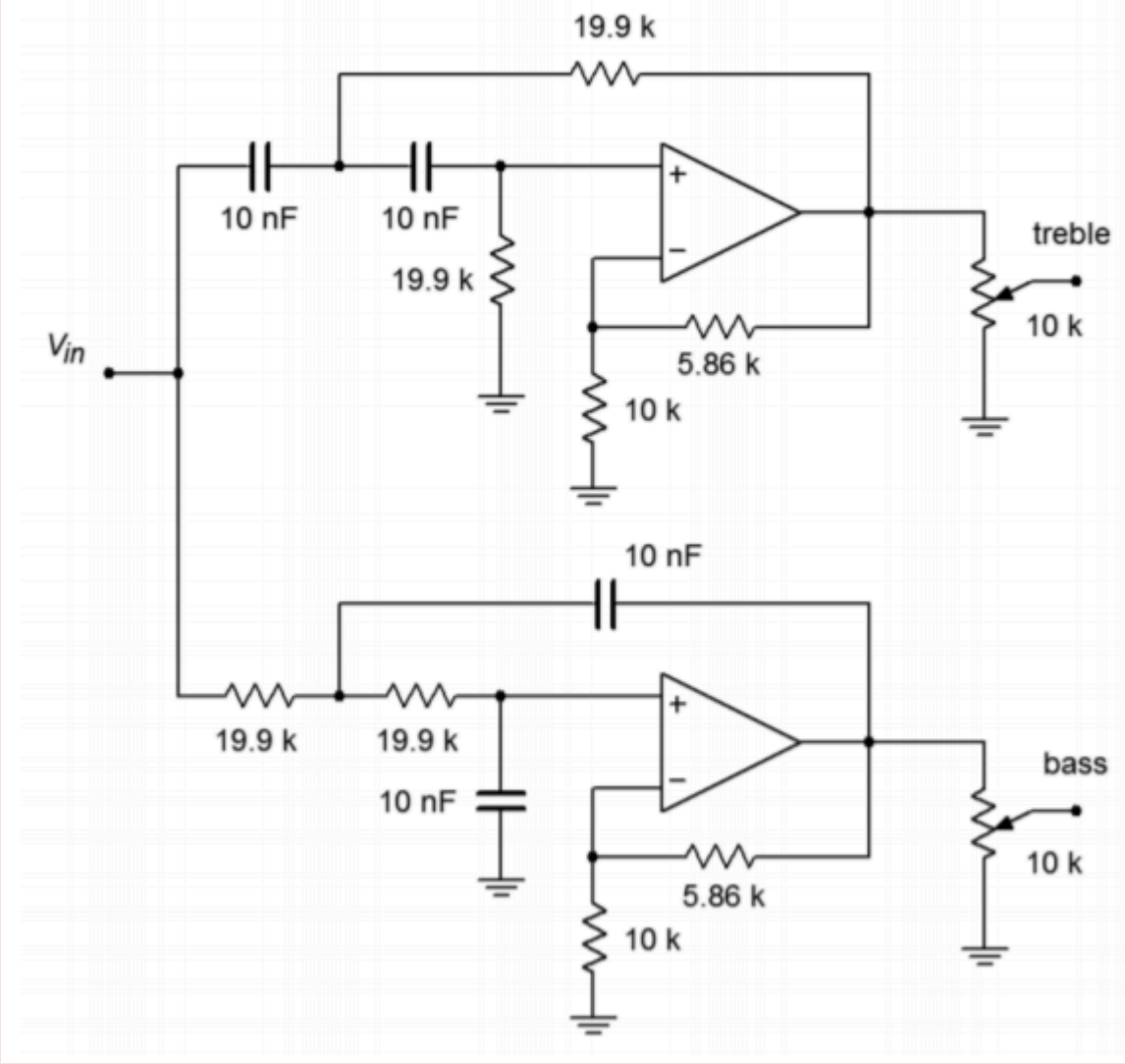


Figure 11.6.24 : Completed crossover design for Example 11.6.4 .

15.7 BAND-PASS FILTER REALIZATIONS

There are many ways to form a band-pass filter. Before we introduce a few of the possibilities, we must define a number of important parameters. As in the case of the high- and low-pass filters, the concept of damping is important. For historical reasons, band-pass filters are normally specified with the parameter ◈ , the quality factor, which is the reciprocal of the damping factor. Comparable to the break frequency is the center, or peak, frequency of the filter. This is the point of maximum gain. In RLC circuits, it is usually referred to as the resonance frequency. The symbol for center frequency is ◈◈ . Because a band-pass filter produces attenuation on either side of the center frequency, there are two "3 dB down" frequencies. The lower frequency is normally given the name ◈1 , and the upper is given ◈2 . The difference between ◈2 and ◈1 is called the bandwidth of the filter and is abbreviated as ◈◈ . The ratio of center frequency to bandwidth is equal to the filter's ◈ .

$$BW = f_2 - f_1$$

(11.7.1)

$$Q = \frac{f_o}{BW}$$

(11.7.2)

It is important to note that the center frequency is not equal to the arithmetic average of ◈1 and ◈2 . Instead, it is equal to the geometric average of ◈1 and ◈2 .

$$f_o = \sqrt{f_1 f_2}$$

(11.7.3)

These parameters are shown graphically in Figure 11.7.1 . If a filter requires a fairly low ◈ , say unity or less, the filter is best realized as a cascade of separate low- and high-pass filters. For higher ◈ s, we will examine two possible realizations. Multiple-feedback filters will be used for ◈ s up to about 10. For ◈ s above 10, the state-variable filter is presented.

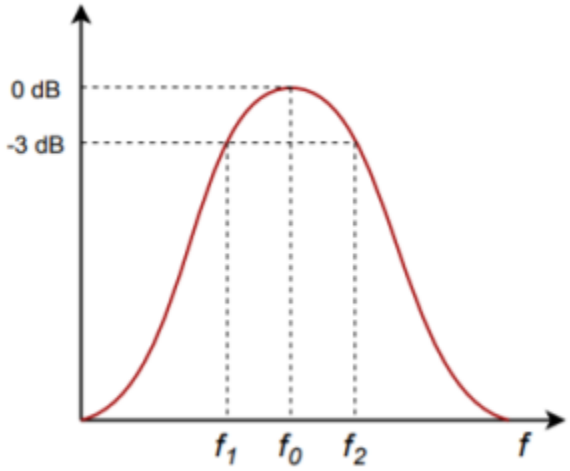


Figure 11.7.1 : Band-pass response.

MULTIPLE-FEEDBACK FILTERS

The basic multiple-feedback filter is a second-order type. It contains two reactive elements as shown in Figure 11.7.2 . One pair of elements creates the low-pass response (◈1◈1) , and the other pair creates the high-pass response (◈2◈2) . Because of this, the ultimate attenuation slopes are ±6 dB.

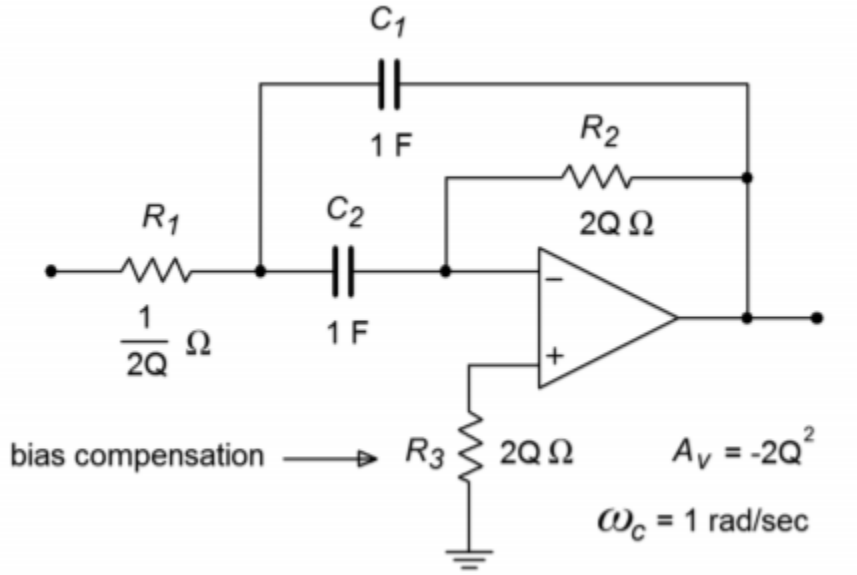


Figure 11.7.2 : Multiple feedback band-pass filter.

As with the VCVS high- and low-pass designs, the circuit of Figure 11.7.2 is normalized to a 1 radian per second center frequency. Extrapolation to new center frequencies is performed in the same manner as shown earlier. The peak gain for this circuit is

$$A_v = -2Q^2$$

(11.7.4)

You can see from Equation 11.7.4 that higher ◈ s will produce higher gains. For a ◈ of 10, the voltage gain will be 200. For this circuit to function properly, the open-loop gain of the op amp used must be greater than 200 at the chosen center frequency. Usually, a safety factor of 10 is included in order to keep stability high and distortion low. By combining these factors, we may determine the minimum acceptable ◈◈◈◈◈◈ for the op amp.

$$f_{unity} \geq 10 f_o A_v$$

(11.7.5)

or more directly,

$$f_{unity} \geq 20 f_o Q^2$$

(11.7.6)

For a ◈ of 10 and a center frequency of 2 kHz, the op amp will need an ◈◈◈◈◈◈ of at least 4 MHz. It is not possible to use this type of filter for high-frequency, high- ◈ work, as standard op amps soon "run out of steam". This difficulty aside, the high gains produced by even moderate values for ◈ may well be impractical. For many applications, a unity gain version would be preferred. This is not particularly difficult to achieve. All that we need to do is attenuate the input signal by a factor equal to the voltage gain of the filter. Because the gain magnitude of the filter is 2◈2 , the attenuation should be

$$Attenuation = \frac{1}{2Q^2}$$

(11.7.7)

Although it is possible to place a pair of resistors in front of the filter to create a voltage divider, there is a more efficient way. We can split ◈1 into two components, as shown in Figure 11.7.3 . As long as the Thevenin equivalent of ◈1◈ and ◈1◈ as seen from the op amp equals the value of ◈1 , the tuning frequency of the filter will not be changed. Also required is that the voltage divider ratio produced by ◈1◈ and ◈1◈ satisfies Equation 11.7.7 .

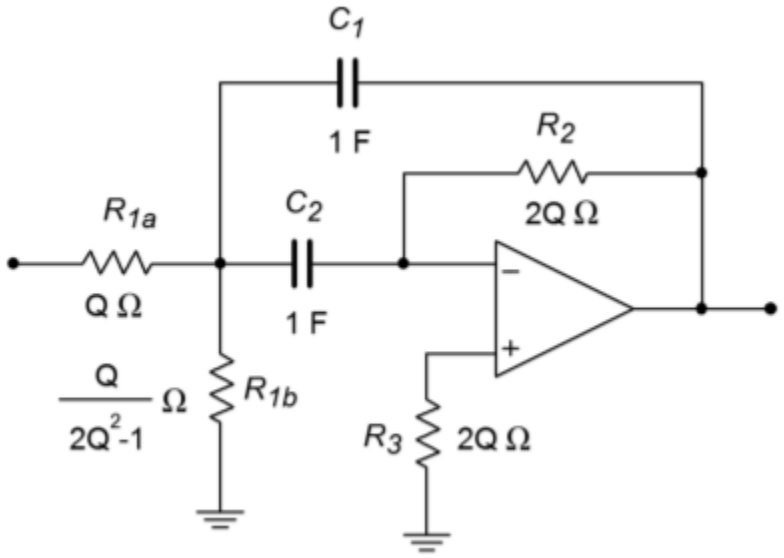


Figure 11.7.3 : Multiple feedback filter with unity-gain variation.

First, let's determine the ratio of the two resistors. We can start by setting ◈1◈ to the arbitrary value ◈ . Using the voltage divider rule and Equation 11.7.7 , ◈1◈ is found:

$$Attenuation = \frac{R_{1b}}{R_{1a} + R_{1b}}$$

$$\frac{1}{2Q^2} = \frac{K}{R_{1a} + K}$$

$$R_{1a} + K = K2Q^2$$

(11.7.8)

So, we see that ◈1◈ must be 2◈2–1 times larger than ◈1◈ . Now we must determine the value of ◈ which will set the parallel combination of ◈1◈ and ◈1◈ to the required value of 1/(2◈) , as based on Figure 11.7.2 .

$$R_{Thevenin} = R_{1a}||R_{1b}$$

$$R_{Thevenin} = \frac{R_{1a}R_{1b}}{R_{1a} + R_{1b}}$$

$$R_{Thevenin} = \frac{K^2(2Q^2 - 1)}{K(2Q^21) + K}$$

$$R_{Thevenin} = \frac{K^2(2Q^2-1)}{K2Q^2}$$

$$R_{Thevenin} = \frac{K(2Q^2-1)}{2Q^2}$$

Because ◈h◈◈◈◈◈◈=1/(2◈) ,

$$\frac{1}{2Q} = \frac{K(2Q^2-1)}{2Q^2}$$

$$1 = \frac{K(2Q^2-1)}{Q}$$

(11.7.9)

Because ◈1◈ was set to ◈ ,

```
*** QuickLaTeX cannot compile formula:
\[R_{1b} = \frac{Q}{2Q^2-1} \Omega \label] (11.7.10)
Substituting 11.7.9 into 11.7.8 yields
\[R_{1a} = Q \Omega \]

*** Error message:
Use of \df@label doesn't match its definition.
leading text: \[R_{1a} = Q \Omega \]
File ended while scanning use of \ltx@label.
Emergency stop.
```

(11.7.11)

By using these values for ◈1◈ and ◈1◈ , the filter will have a peak gain of unity. Note that as this scheme only attenuates the signal prior to gain, the ◈◈◈◈◈◈ requirement set in Equations 11.7.5 – 11.7.6 still holds true.

Example 11.7.1

Design a filter that will only pass frequencies from 800 Hz to 1200 Hz. Make sure that this is a unity-gain realization.

First, we must determine the center frequency, bandwidth, and ◈ .

$$BW = f_2 f_1$$

$$BW = 1200Hz - 800Hz$$

$$BW = 400Hz$$

$$f_o = \sqrt{f_1 f_2}$$

$$f_o = \sqrt{800Hz \times 1200Hz}$$

$$f_o = 980Hz$$

$$Q = \frac{f_o}{BW}$$

$$Q = \frac{980Hz}{400Hz}$$

$$Q = 2.45$$

The ◈ is too high to use separate high- and low-pass filters, but sufficiently low so that a multiple feedback type may be used. Before proceeding, we should check to make sure that the required ◈◈◈◈◈◈ for the op amp is reasonable.

$$A_v = -2Q^2$$

$$A_v = -2 \times 2.45^2$$

$$A_v = -12$$

$$f_{unity} \geq 10 A_v f_o$$

$$f_{unity} \geq 10 \times 12 \times 980Hz$$

$$f_{unity} \geq 117.6kHz$$

Just about any modern op amp will exceed the ◈◈◈◈◈◈ specification. As this circuit shows a gain of 12, the unity gain variation shown in Figure 11.7.3 will be used. The calculations for the normalized components follow.

$$R_2 = 2Q$$

$$R_2 = 2 \times 2.45$$

$$R_2 = 4.9\Omega$$

$$R_{1b} = \frac{Q}{2Q^2 - 1}]$$

$$R_{1b} = \frac{2.45}{2 \times 2.45^2 - 1}$$

$$R_{1b} = .2226\Omega$$

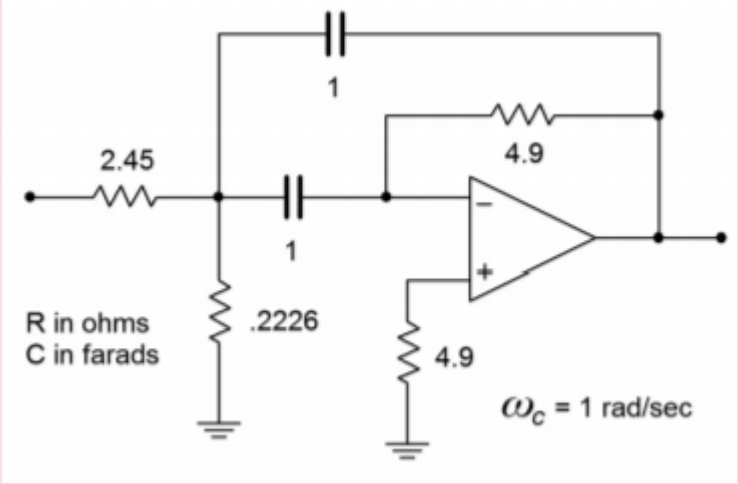


Figure 11.7.4 : Initial damping calculation for Example 11.7.1 .

The resulting normalized circuit is shown in Figure 11.7.4 . We must now find the frequency scaling factor.

$$\omega_o = 2\pi f_o$$

$$\omega_o = 2\pi 980 Hz$$

$$\omega_o = 6158 \text{ radians per second}$$

In order to translate our circuit to this frequency, we must divide either the resistors or the capacitors by 6158. In this example, let's use the capacitors.

$$C = \frac{1}{6158}$$

$$C = 162.4\mu F$$

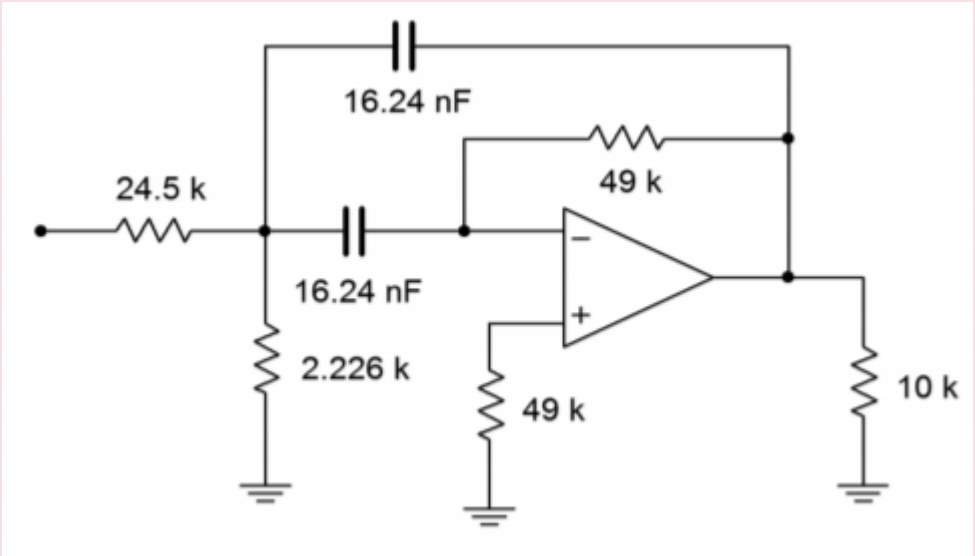


Figure 11.7.5 : Final impedance and frequency scaling for Example 11.7.1 .

A further impedance scaling is needed for practical component values. A factor of a few thousand or so

would be appropriate here. To keep the calculations simple, we'll choose 10 k. Each resistor will be increased by 10 k, and each capacitor will be reduced by 10 k. The final scaled filter is shown in Figure 11.7.5 .

COMPUTER SIMULATION

The Multisim simulation of the circuit of Example 11.7.1 is shown in Figure 11.7.6 . Note that the gain is 0 dB at the approximate center frequency (about 1 kHz). Also, the –3 dB breakpoints of 800 Hz and 1200 Hz are clearly seen. The phase response of this filter is also plotted. Note the very fast phase transition in the area around ◈◈ . If the ◈ of this circuit was increased, this transition would be faster still.

In simulations such as this, it is very important that realistic op amp models be employed. If an over-idealized version is used, non-ideal behavior due to a reduction of loop-gain will go unnoticed. This error is most likely to occur in circuits with high center frequencies and/or high ◈ s. You can verify this by translating the filter to a higher frequency and rerunning the simulation. For example, if ◈1 and ◈2 are decreased by a factor of 1000, the center frequency should move up to about 1 MHz. If the simulation is run again with an appropriate range of test frequencies, you will see that the limited bandwidth of the ◈ A741 op amp prematurely cuts off the filter response. The result is a peaking frequency more than one octave below target, a maximum amplitude several dB below 0, and an asymmetrical response curve. This response graph is shown in Figure 11.7.6◈ . The accompanying phase plot also shows a great deviation from the ideal filter. An excessive phase shift at the middle and higher frequencies is clearly evident.

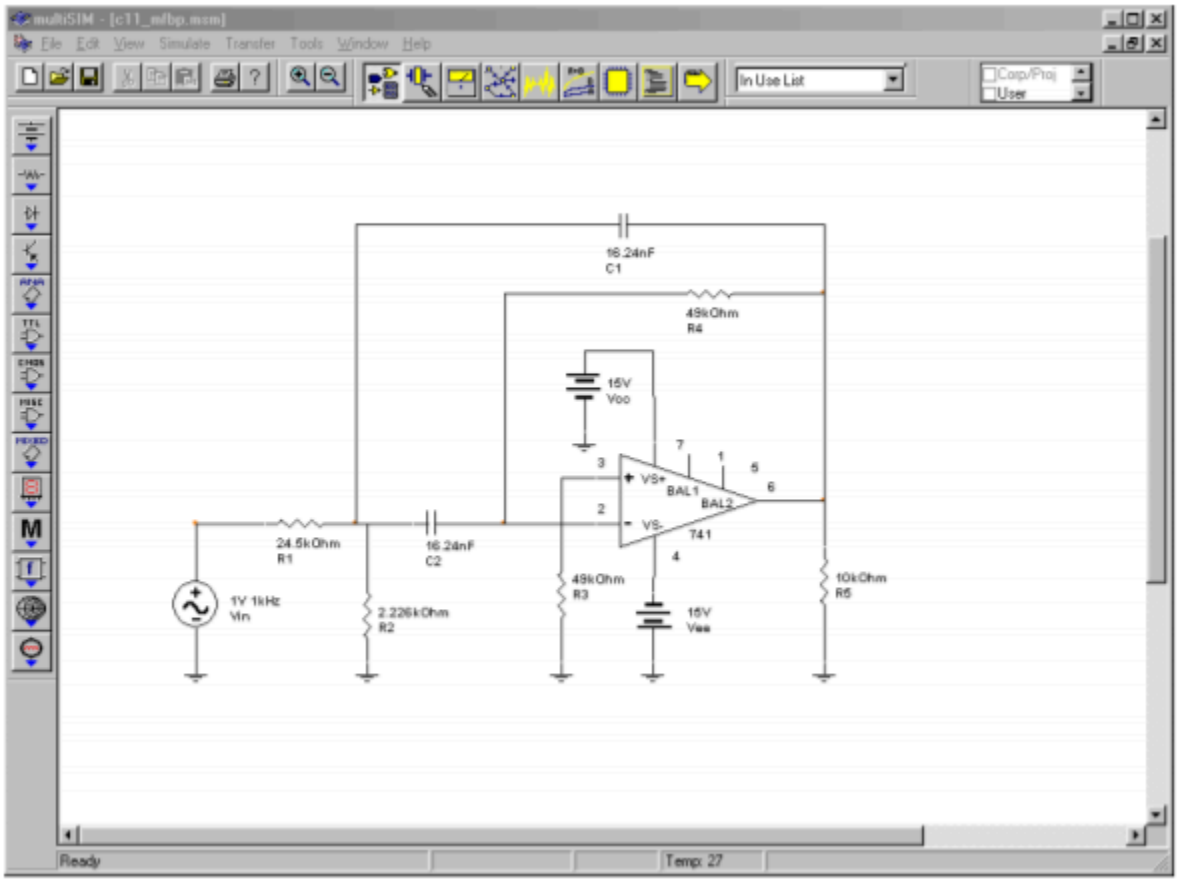


Figure 11.7.6 : Band-pass filter in Multisim.

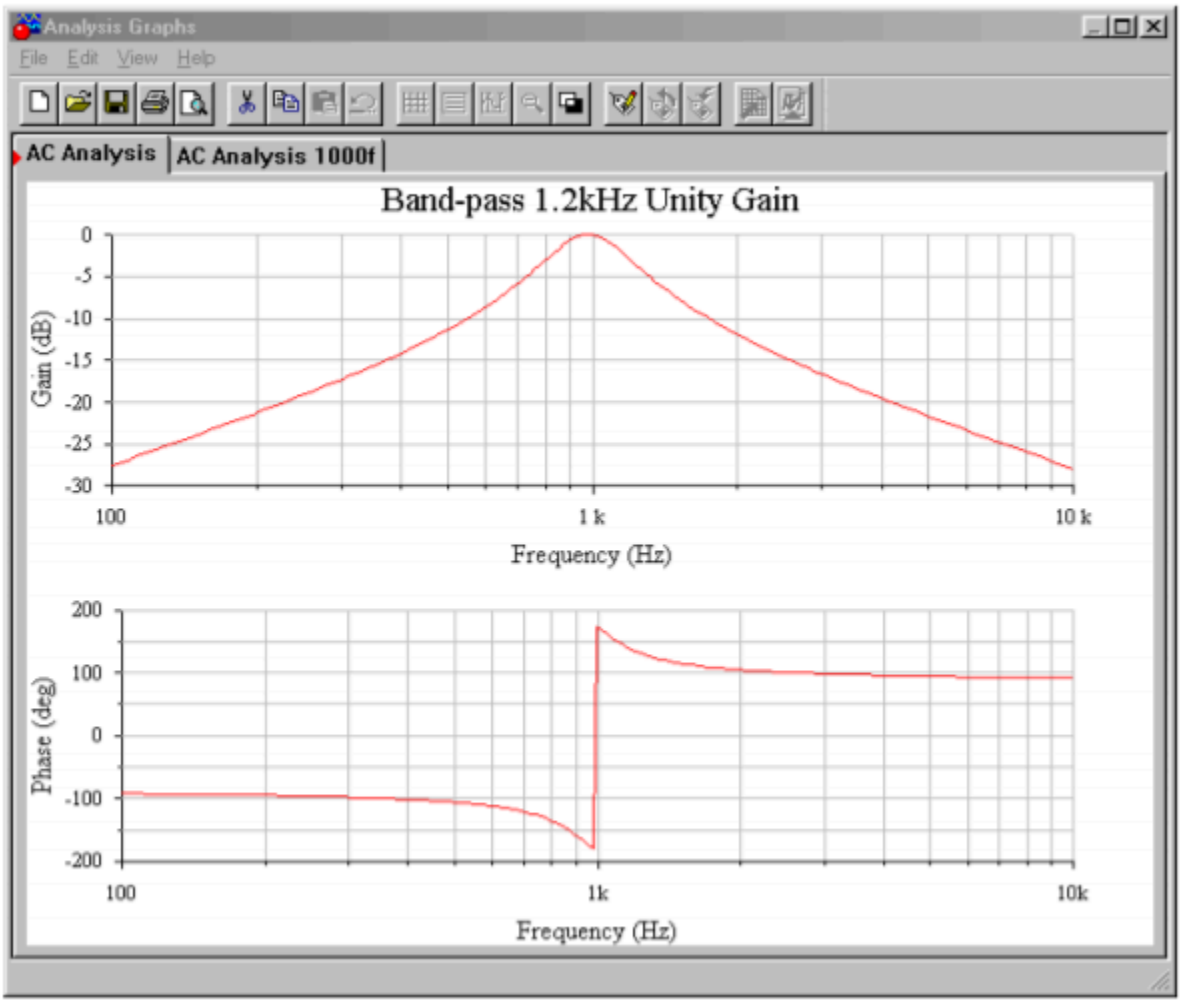


Figure 11.7.6 : Gain and phase plots for band-pass filter.

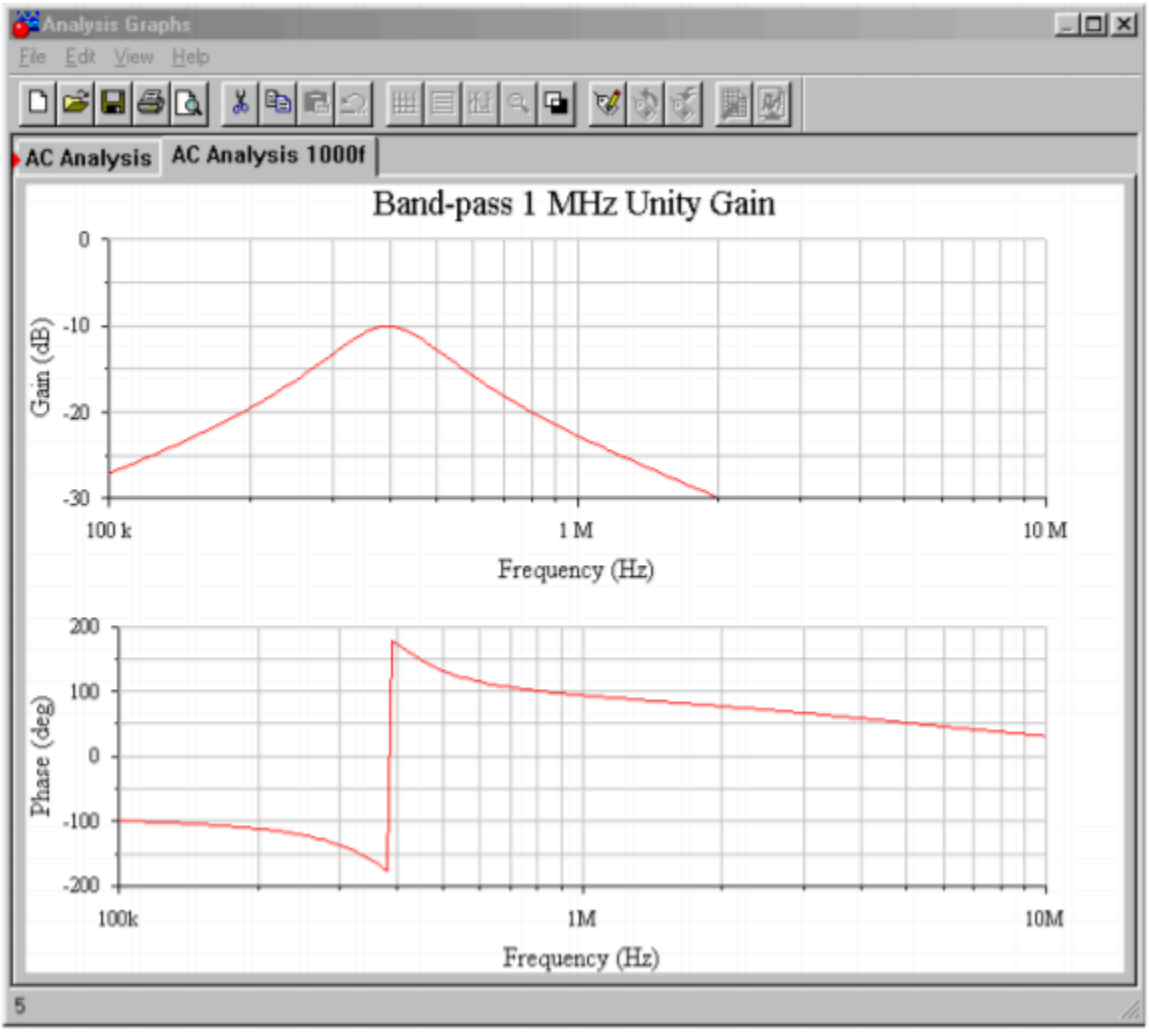


Figure 11.7.6 ◈ : Gain and phase plots for 1000 times frequency shift.

STATE-VARIABLE FILTER

As noted earlier, the multiple-feedback filter is not suited to high frequency or high ◈ work. For applications requiring ◈ s of about 10 or more, the state-variable filter is the form of choice. The state-variable is often referred to as the universal filter, as band-pass, high-pass, and low-pass outputs are all available. With additional components, a band-reject output may be formed as well. Unlike the earlier filter forms examined, the basic state-variable filter requires three op amps. Also, it is a second-order type, although higher-order types are possible. This form gets its name from state-variable analysis. One of the earliest uses for op amps was in the construction of analog computers (see Chapter Ten). Interconnections of differentiators, amplifiers, summers, and integrators were used to electronically solve differential equations that described physical systems. State-variable analysis provides a technique for solving involved differential equations. The equations may in fact, describe a required filter's characteristics. Although we will not examine state-variable analysis, this does not preclude a study of the state-variable filter. Designing with state-variable filters is really no more complex than our previous work.

Besides its ability to provide stable filters with relatively high ◈ s, the state-variable has other unique characteristics:

- It is relatively easy to tune electronically over a broad frequency range.
- It is possible to independently adjust the ◈ and tuning frequency.
- It offers the ability to create other, more complex, filters, as it has multiple outputs.

The state-variable filter is based on integrators. The general form utilizes a summing amplifier and two integrators, as shown in Figure 11.7.7 . To understand how this circuit works on an intuitive level, recall that integrators are basically first-order, low-pass filters. As you can see, the extreme right side output has passed through the integrators and produces a low-pass response. If the low-pass output is summed out of phase with the input signal, the low frequency information will cancel, leaving just the high frequency components. Therefore, the output of the summer is the high-pass output of the filter. If the high-pass signal is integrated (using the same critical frequency), the result will be a band-pass response. This is seen at the output of the first integrator. The band-pass signal is also routed back to the input summing amplifier. By changing the amount of the signal that is fed back, the response near the critical frequency may be altered, effectively setting the filter ◈ . Finally, the loop is completed by integrating the band-pass response, which yields the low-pass output. In effect, the second integrator's –6 dB per octave rolloff perfectly compensates for the rising band-pass response below ◈◈ . This produces flat response below ◈◈ . Above ◈◈ , the combination of the two falling response curves produces the expected second-order, low-pass response.

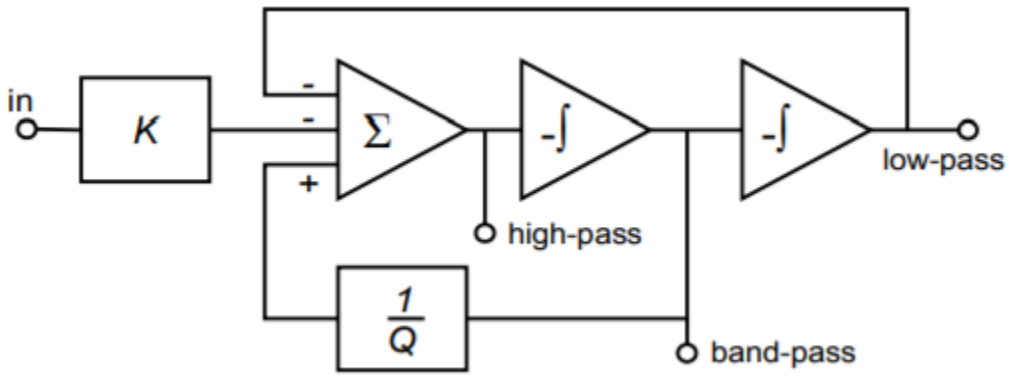


Figure 11.7.7 : Block diagram of state-variable filter.

Two popular ways of configuring the state-variable filter are the fixed-gain and adjustable-gain forms. The fixed-gain form is shown in Figure 11.7.8 . This circuit uses a total of three op amps. The ◈ of the circuit is set by a single resistor, ◈◈ . ◈ s up to 100 are possible with state-variable filters. For the high- and low-pass outputs, the gain of this circuit is unity. For the band-pass output, the gain is equal to ◈ .

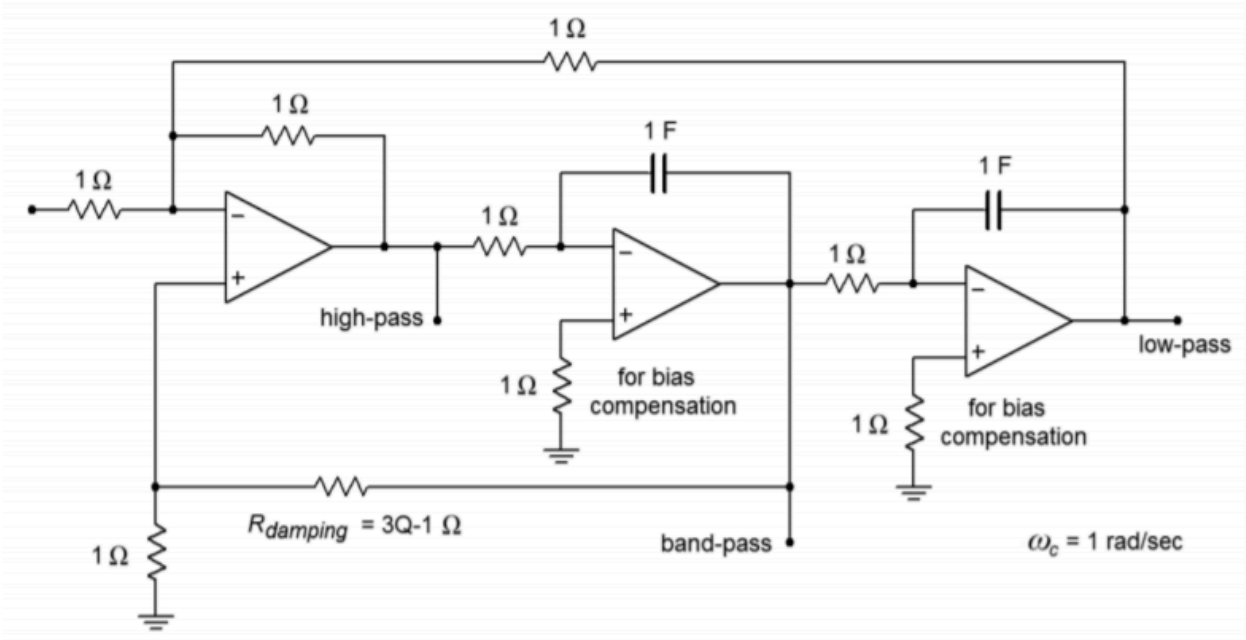


Figure 11.7.8 : Fixed-gain version of state-variable filter.

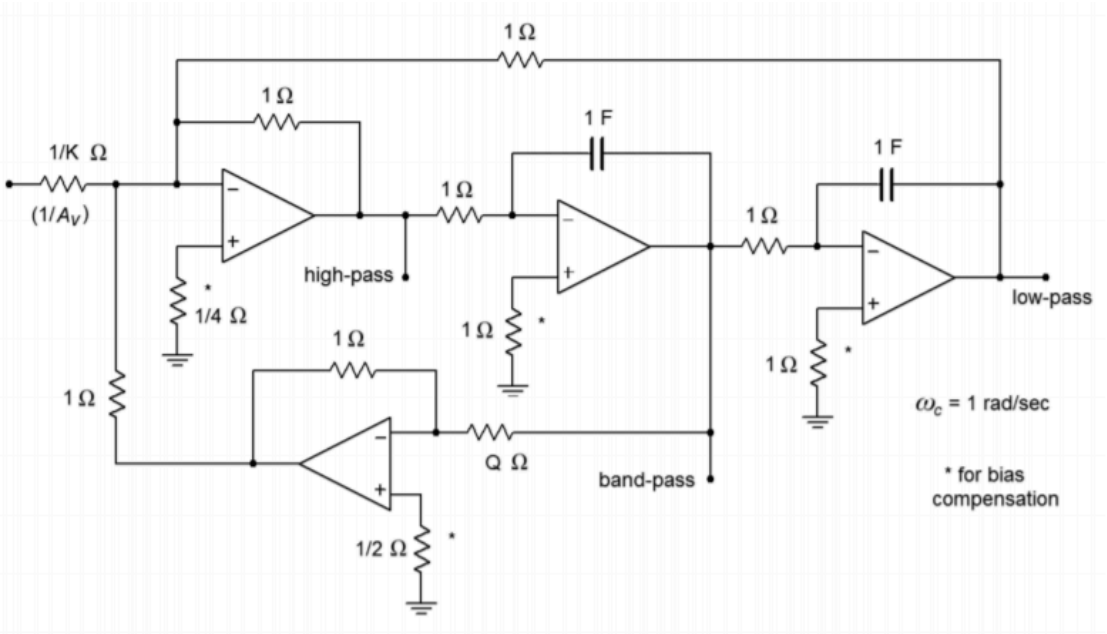


Figure 11.7.9 : Variable-gain version of state-variable filter.

Figure 11.7.9 shows an adjustable-gain version. For high- or low-pass use, the gain is equal to the arbitrary value ◈ , whereas for band-pass use, the gain is equal to ◈◈ . This variation requires a fourth op amp in order to isolate the ◈ and gain settings. Although four op amps may sound like a large number of devices, remember that a variety of quad op amp packages exist, indicating that the actual physical layout may be quite small. Also, even though three different outputs are available, it is not possible to individually optimize each one for simultaneous use. Consequently, the state-variable is most often used as a stable and switchable high/low-pass filter, or as a high ◈ band-pass filter. Finally, in keeping with our previous work, the circuits are shown normalized to a critical frequency of one radian per second. Although we will concentrate on band-pass design in this section, it is possible to

use these circuits to realize various high- and low-pass filters, such as those generated with the Sallen and Key forms. The procedure is nearly identical and uses the same frequency and damping factors (Figures 11.6.13 and 11.6.18).

Example 11.7.2

Design a band-pass filter with a center frequency of 4.3 kHz and a ◈ of 25. Use the fixed-gain form.

First, determine the damping resistor value. Then, scale the components for the desired center frequency. Note that a ◈ of 25 produces a bandwidth of only 172 hertz for this filter (4.3 kHz/25).

$$R_{damping} = 3Q - 1$$

$$R_{damping} = 3 \times 25 - 1]$$

$$R_{damping} = 74\Omega$$

$$\omega_o = 2\pi f_o$$

$$\omega_o = 2\pi 4.3kHz$$

$$\omega_o = 27.02k \text{ radians per second}$$

In order to translate the filter to our desired center frequency, we need to divide either the resistors or the capacitors by 27,020. For this example, we'll use the capacitors.

$$C = \frac{1}{27.02k}$$

$$C = 37\mu F$$

A final impedance scaling is required to achieve reasonable component values. A reasonable value might be a factor of 5000.

$$C = \frac{37\mu F}{5000}$$

$$C = 7.4nF$$

$$R_{damping} = 74 \times 5000$$

$$R_{damping} = 370k\Omega$$

All remaining resistors will equal 5◈Ω .

Because this is a band-pass filter,

$$A_v = Q$$

$$A_v = 25$$

The completed filter is shown in Figure 11.7.10 . The value for ◈◈◈◈◈◈◈◈ is considerably larger than the other resistors. This effect gets worse as the required ◈ is increased. If this value becomes too large for practical components, it may be reduced to a more reasonable value as long as the associated divider resistor (from the noninverting input to ground) is reduced by the same amount. The ratio of these two resistors is what sets the filter ◈ , not their absolute values. Lowering these values will upset the ideal

input bias current compensation, but this effect can be ignored in many cases, or reduced through the use of FET input op amps.

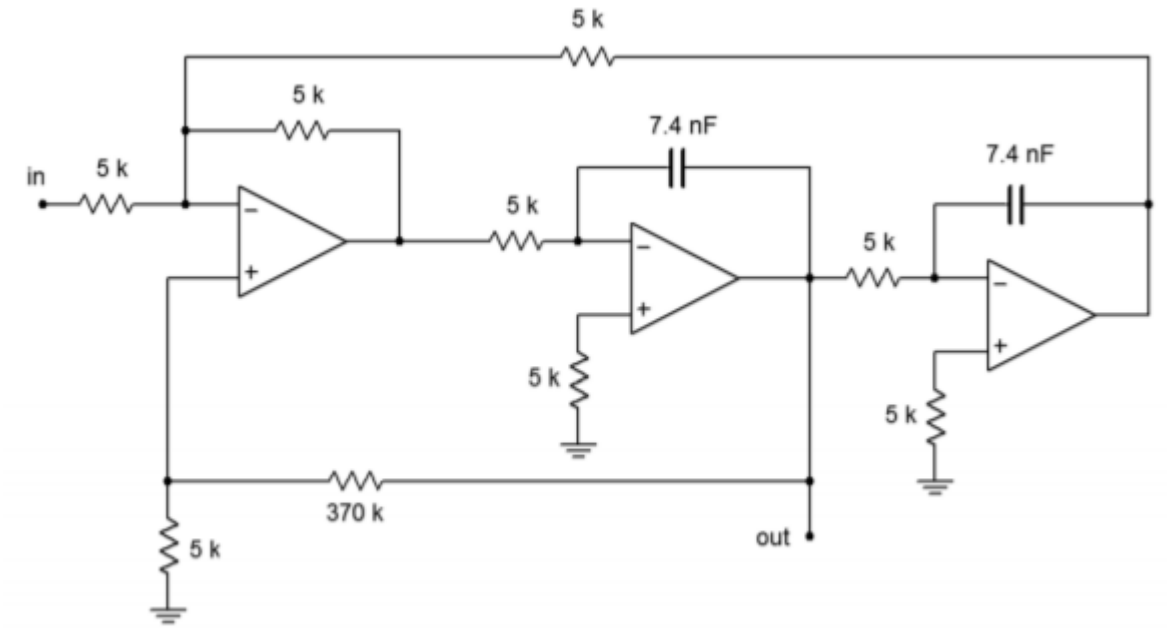


Figure 11.7.10 : Completed design of bandpass filter for Example 11.7.2 .

Altering this circuit for a variable-gain configuration requires the addition of a fourth amplifier as shown in Figure 11.7.9 . The calculation for the damping resistor is altered, and a value for the input gain determining resistor is needed. The remaining component calculations are unchanged from the example above. Note that by setting the gain constant ◈ to 1/◈ , the final filter gain may be set to unity.

15.8 NOTCH FILTER (BAND-REJECT) REALIZATIONS

By summing the high-pass and low-pass outputs of the state-variable filter, a notch, or band-reject, filter may be formed. Filters of this type are commonly used to remove interference signals. The summation is easily performed with a simple parallel-parallel summing amplifier, as shown in Figure 11.8.1 .

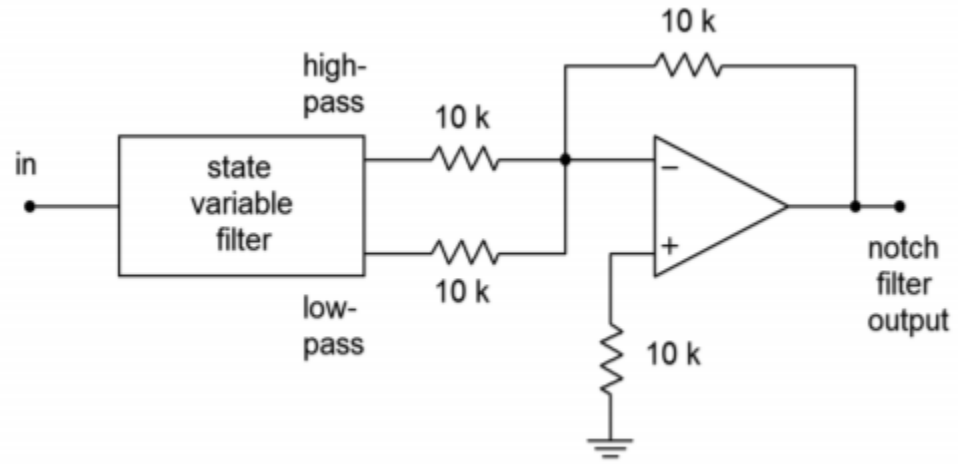


Figure 11.8.1 : Notch filter.

For reasonable ◈ values, there will be tight correlation between the calculated band-pass center and –3 dB frequencies, and the notch center and –3 dB frequencies. The component calculations proceed as in the band-pass filter.

Example 11.8.1

A filter is needed to remove induced 60 Hz hum from a transducer's signal. The rejection bandwidth of the filter should be no more than 2 Hz. From the specifications we know that the center frequency is 60 Hz and the ◈ is 60/2, or 30. For simplicity, we will use the fixed-gain form. (Note that the gain of the filter on either side of the notch will be unity.)

$$R_{damping} = 3Q - 1$$

$$R_{damping} = 3 \times 30 - 1$$

$$R_{damping} = 89\Omega$$

$$\omega_o = 2\pi 60$$

$$\omega_o = 377 rad/s$$

Scaling ⧫ produces

$$C = \frac{1}{377}$$

$$C = 2.65 \text{ milliFarads}$$

A practical value scaling of 104 produces the circuit of Figure 11.8.2 . Note that the damping resistors have only been scaled by 103 , as an ⧫⧫⧫⧫⧫⧫⧫⧫ value of 890⧫Ω might be excessive. Remember, it is the ratio of these two resistors that is important, not their absolute values.

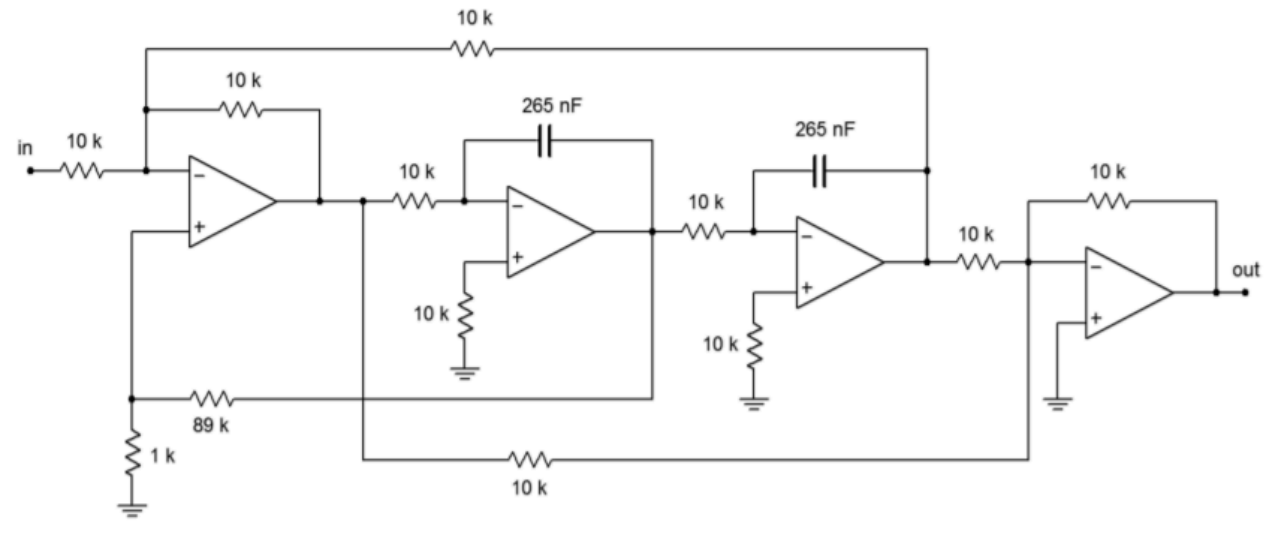


Figure 11.8.2 : Completed notch filter for Example 11.8.1 .

COMPUTER SIMULATION

Figure 11.8.3 shows a simulation of the circuit of Example 11.8.1 using Multisim. To keep the layout simple, an ideal op amp model was chosen. The AC analysis shows the very sharp notch centered at 60 Hz as expected. As filters of this type are designed to remove a single frequency without affecting surrounding material, high precision tuning components are required. To see the effects of even modest component deviations in a production run, a Monte Carlo analysis proves invaluable. In Figure 11.8.3⧫ a series of 10 runs is shown. Each resistor and capacitor in the filter has been given a 1 percent nominal tolerance. Further, the frequency plot range has been narrowed down to just 10 hertz on either side of the target frequency. Even with these relatively tight tolerances, tuning deviations of more than 1 hertz can be seen. Also, the response shape is not perfectly symmetrical in all cases.

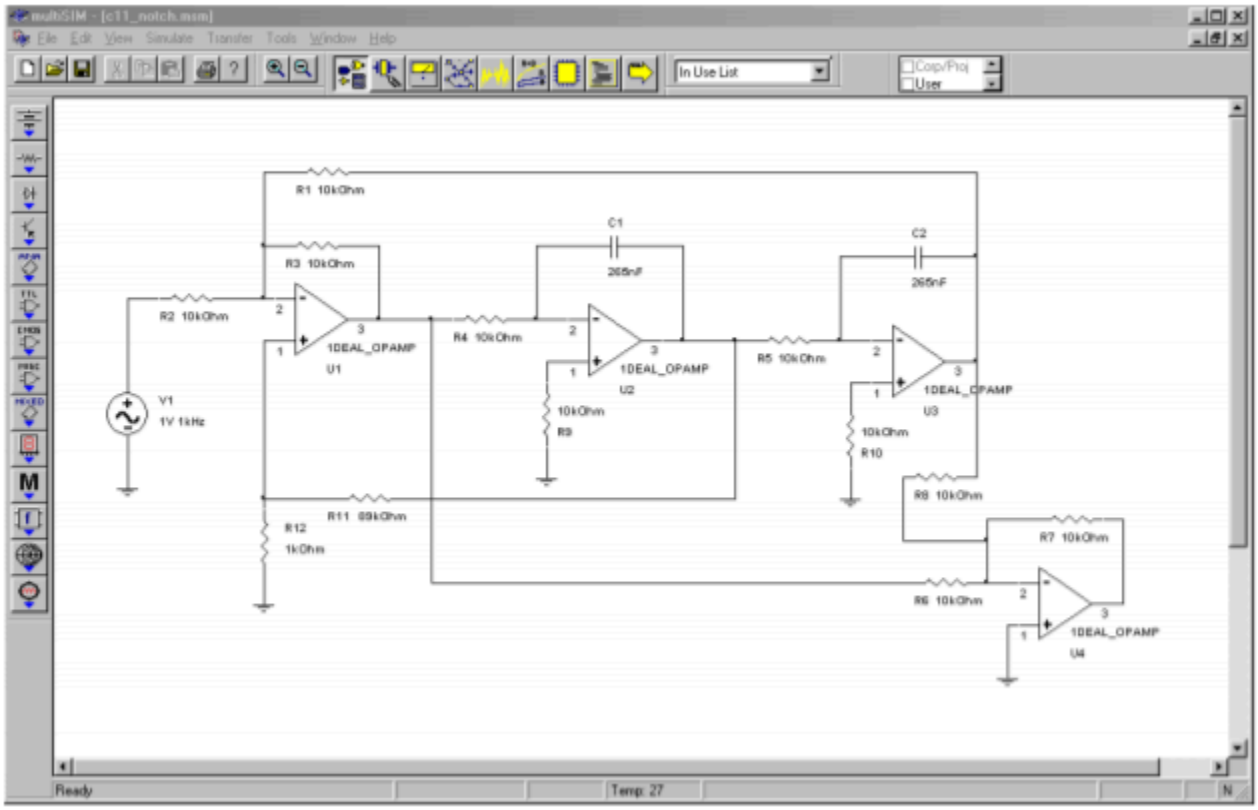


Figure 11.8.3 : State-variable notch filter in Multisim.

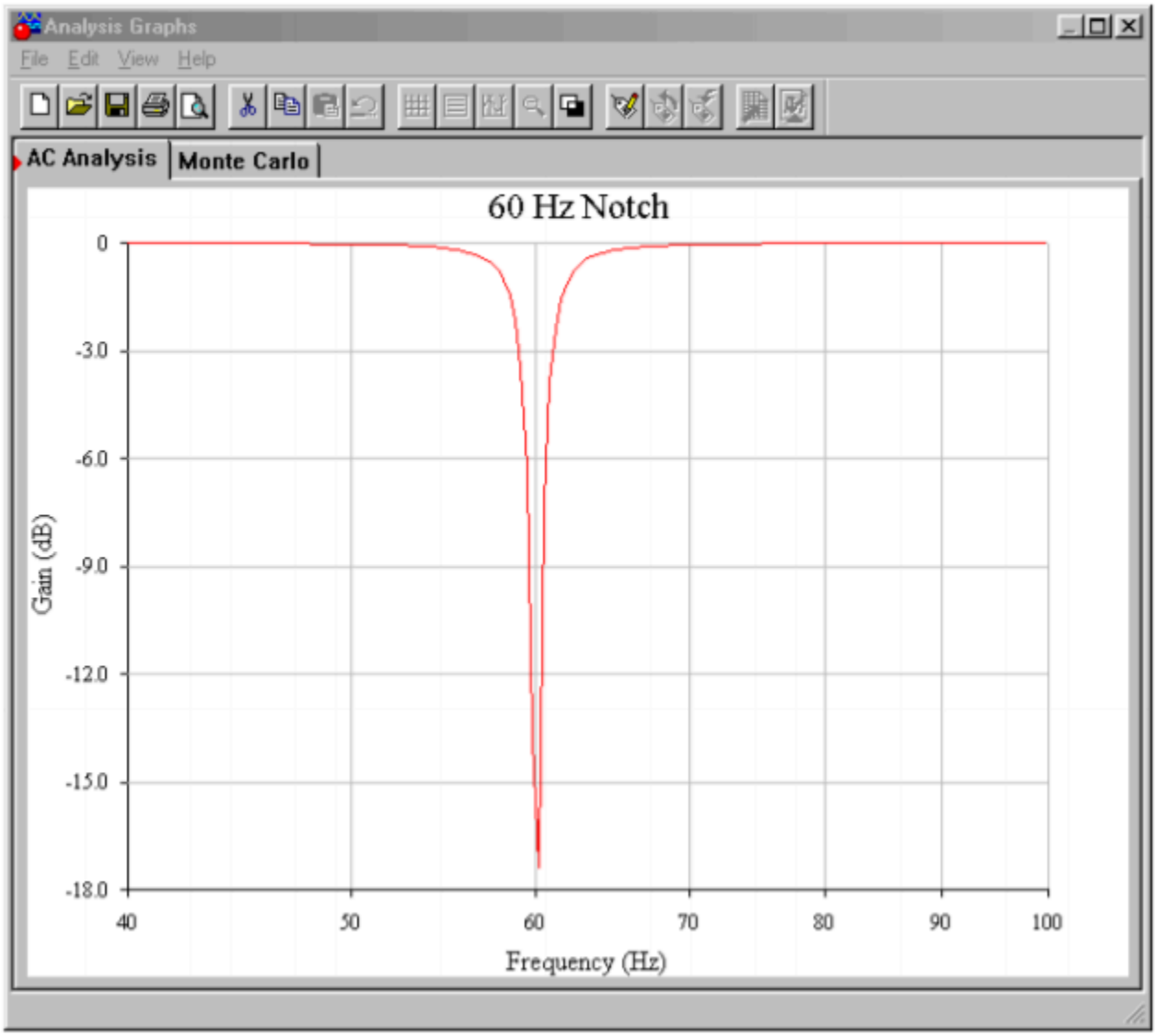


Figure 11.8.3 : Ideal notch response.

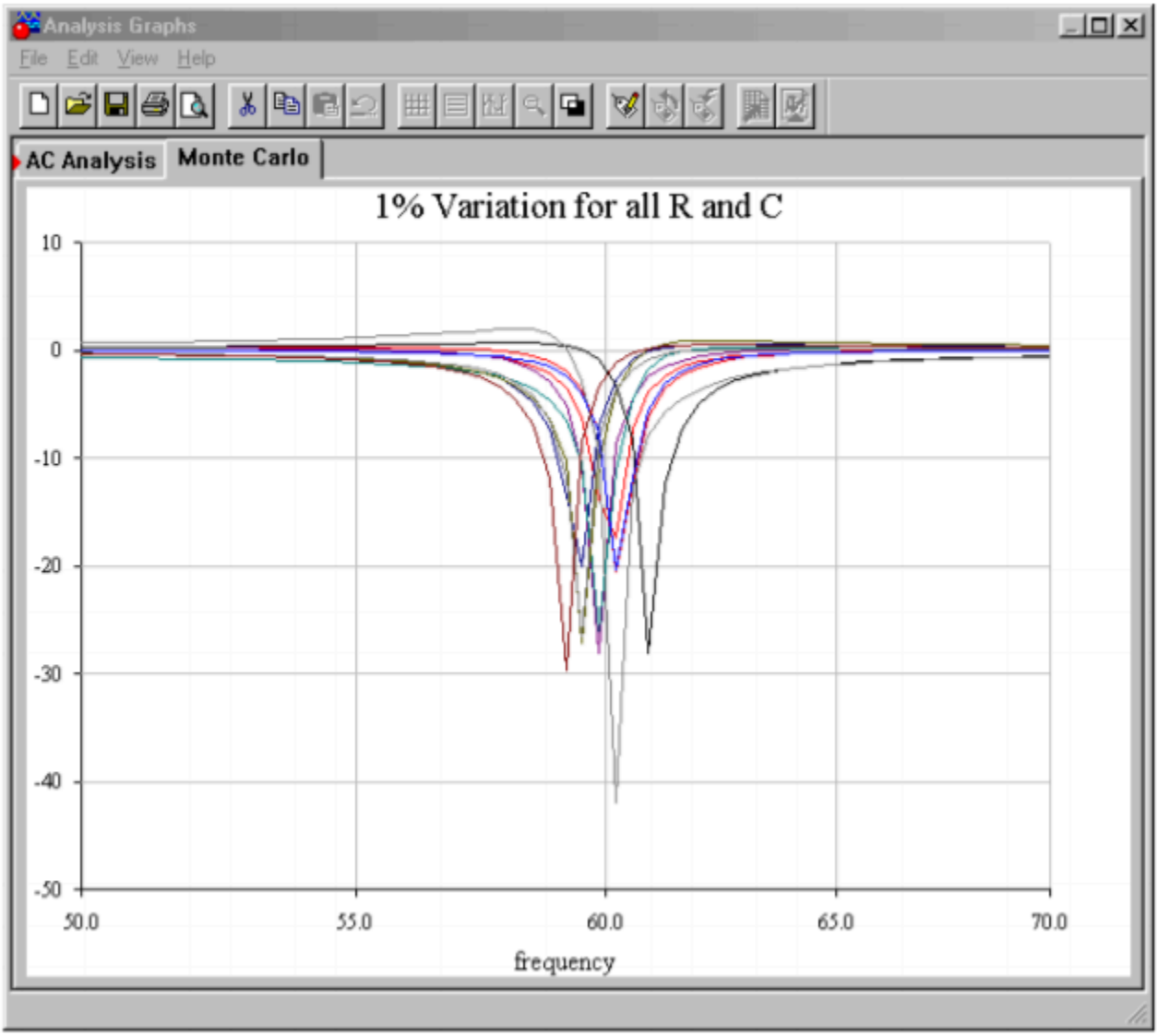


Figure 11.8.3 : Typical response with 1% component variations.

A Note on Component Selection

Ideally, the circuit of Example 11.8.1 will produce –3 dB points at approximately 59 Hz and 61 Hz and will infinitely attenuate 60 Hz tones. In reality, component tolerances may alter the response and, therefore, high-quality parts are required for accurate, high- circuits such as this. Even simpler, less-demanding circuits such as a second-order Sallen and Key filter may not perform as expected if lower quality parts are used. As a general rule, component accuracy and stability becomes more important as filter and order increase. One percent tolerance metal film resistors are commonly used, with 5 percent carbon film types being satisfactory for the simpler circuits. For capacitors, film types such as polyethylene (mylar) are common for general-purpose work, with polycarbonate, polystyrene, polypropylene, and teflon being used for the more stringent requirements. For small capacitance values (<100 pF), NPO ceramics may be used. Generally, large ceramic disc and aluminum electrolytic capacitors are avoided due to their wide tolerance and instability with temperature, applied voltage, and other factors.

FILTER DESIGN TOOLS

In order to further speed the process of active filter design, some manufacturers offer free filter design software. Examples include FilterCAD from Linear Technology and FilterPro from Burr-Brown. Some programs are rather generic and help you design Sallen and Key, multiple-feedback, and state-variable filters. Others are written expressly to support the manufacturer's specialized filter ICs. Typically, the programs will print out component values given desired filter types and break frequencies. Bode plots and pulse waveform simulations may also be available. Such programs can certainly reduce design tedium.

15.9 AUDIO EQUALIZERS

Another range of circuits that fall under the heading of filters are equalizers. Actually, equalizers are a class of adjustable filters that may produce gain as well as attenuation. Perhaps the most common uses for equalizers are in the audio, music, and communications areas. As the name suggests, equalizers are used to adjust or balance the input frequency spectrum. Equalizers range from complex 1/3 octave and parametric types for use in recording studios and large public-address systems, to the simpler bass and treble controls found on virtually all home and car stereos. Many of the more complex equalizers are based on extensions of the state-variable filter. On the other hand, some equalizers are little more than modified amplifiers. We'll take a look at the very common bass and treble controls, which adjust low and high audio frequencies, respectively. The purpose of bass and treble controls is to allow the listener some control of the balance of high and low tones. They may be used to help compensate for the acoustical shortcomings of a loudspeaker, or perhaps solely to compensate for personal taste. Unlike the filters previously examined, these controls will be manipulated by the user and must provide for signal boost as well as cut. A typical response curve is shown in Figure 1 . Normally, bass and treble circuits are realized with parallel-parallel inverting amplifiers. In essence, the feedback network will change with frequency.

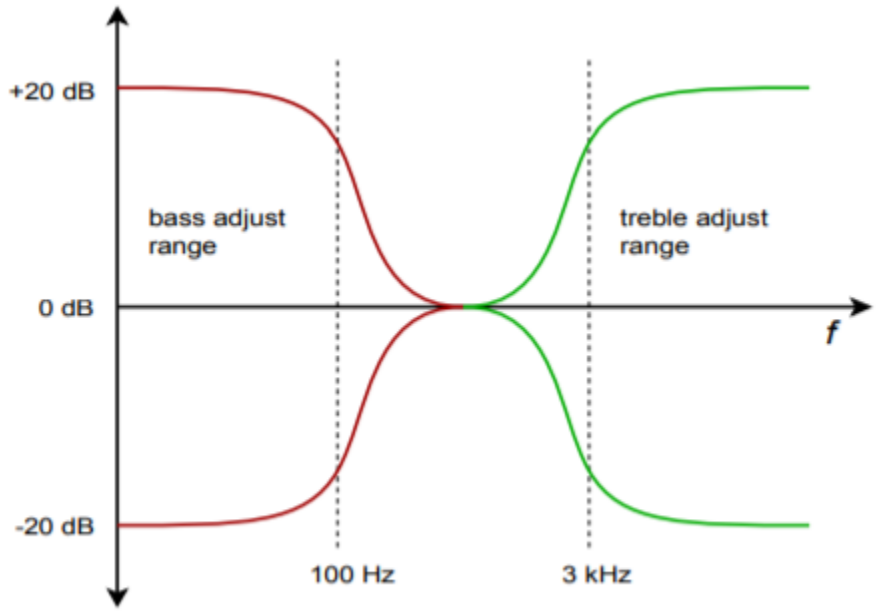


Figure 1 : Response of general bass/treble equalizer.

A simple bass control is shown in Figure 2 . To understand how this circuit works, let's look at what happens at very low and at very high frequencies. First of all, at very high frequencies, capacitor ⓿ is ideally shorted. Thus, the setting of potentiometer ⓿⓿ is inconsequential. In this case, the gain magnitude of the amplifier will be set at ⓿⓿/⓿⓿ . Normally, ⓿⓿ is equal to ⓿⓿ , so the gain is unity. At very low frequencies the exact opposite happens; the capacitor is seen as an open. Under this

condition the gain of the amplifier depends on the setting of potentiometer ⍰⍰ . If the wiper is set to the extreme right, the gain becomes ⍰⍰/(⍰⍰+⍰⍰) . If the wiper is moved to the extreme left, the gain becomes (⍰⍰+⍰⍰)/⍰⍰ . If ⍰⍰ is set to nine times ⍰⍰ , the total gain range will vary from 0.1 to 10 (–20 dB to +20 dB). A similar arrangement may be used to adjust the treble range.

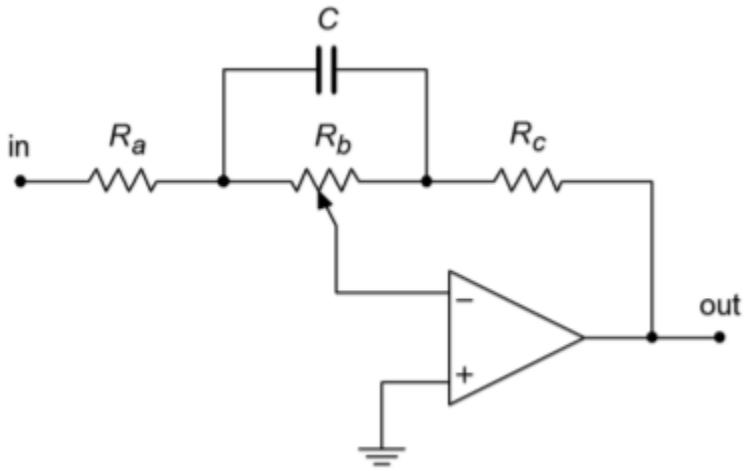


Figure 2 : Simple bass section.

When the bass and treble controls are combined, component loading makes the circuit somewhat more difficult to design. Basic models have already been derived by a number of sources, though, including the one shown in Figure 3 .

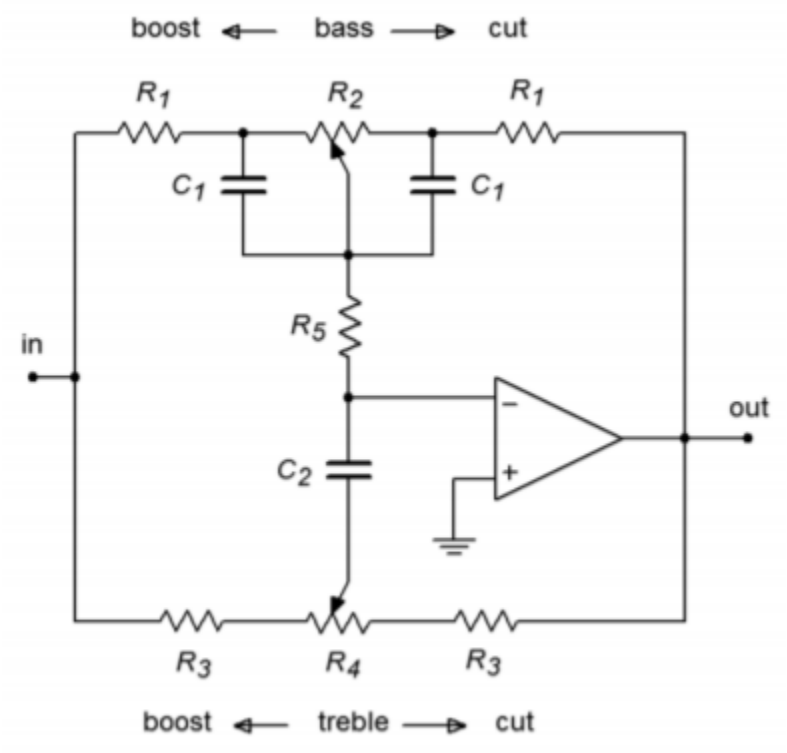


Figure 3 : Bass/treble equalizer.

The following equations are used to design the desired equalizer. (Refer to Figure 4 .)

Bass section (assumes ◈2≫◈1):

$$f_l = \frac{1}{2\pi R_2 C_1}$$

$$f_{lb} = \frac{1}{2\pi R_1 C_1}$$

$$A_{vb} = 1 + \frac{R_2}{R_1}$$

Treble section (assumes ◈4≫◈1+◈3+2◈5):

$$f_h = \frac{1}{2\pi R_3 C_3}$$

$$f_{hb} = \frac{1}{2\pi (R_1 + R_3 + 2R_5) C_3}$$

$$A_{vt} = 1 + \frac{R_1 + 2R_5}{R_3}$$

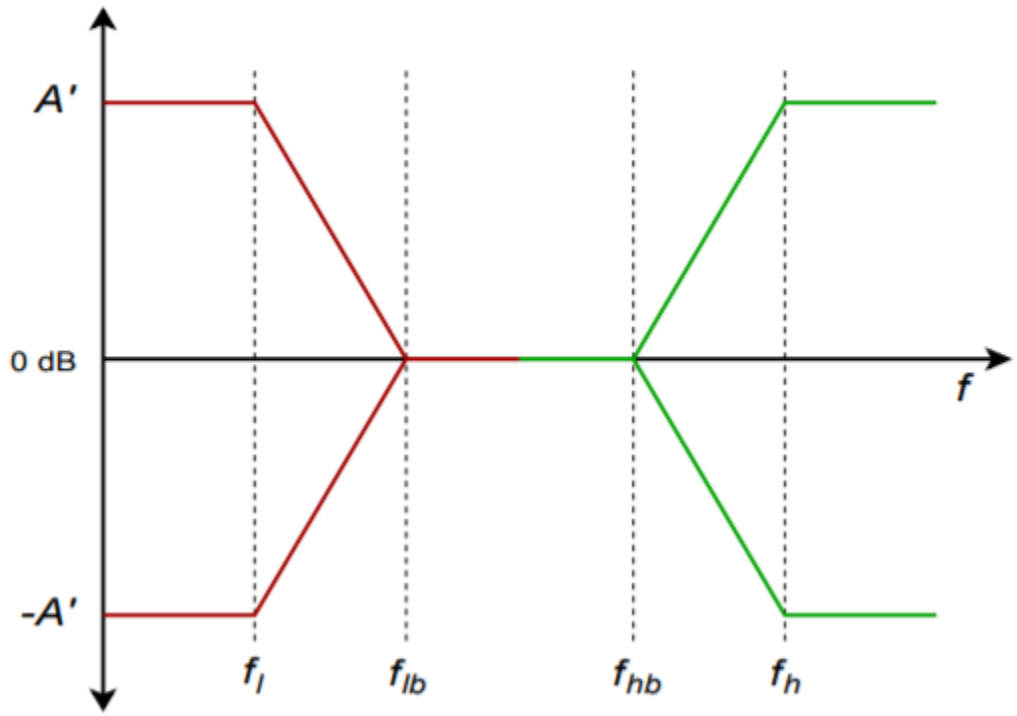


Figure 4 : Response of bass/treble equalizer.

In actuality, it is very common to design these sorts of circuits empirically. In other words, the given equations are used as a starting point, and then component values are adjusted in the laboratory until the desired response range is obtained. Circuits like this may be altered further to include a midrange control. Generally, three adjustments is considered to be the maximum for this type of

circuit. An example of a bass-midrange-treble equalizer is shown in the schematic for the Pocket Rocket amplifier in Chapter Six.

Computer Simulation A simple bass equalizer is simulated in Figure 5 . The maximum cut and boost is set to a factor of approximately 10, or 20 dB.

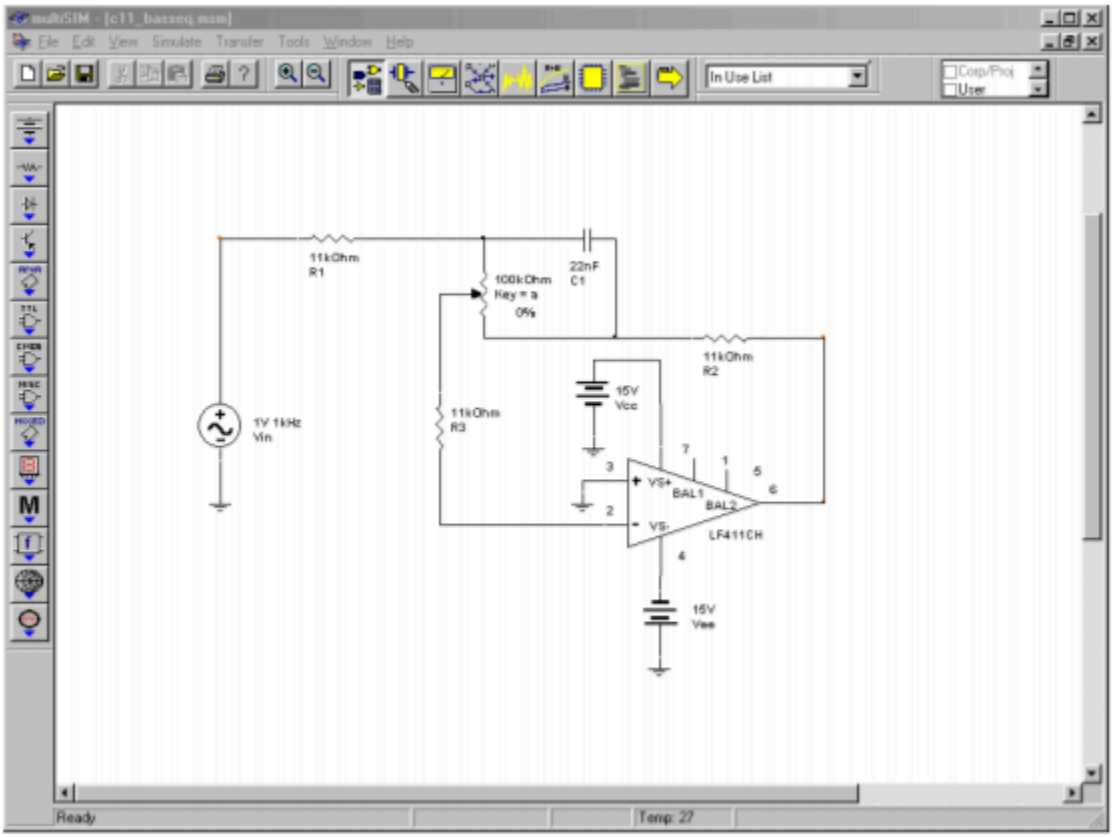


Figure 5 : Bass equalizer in Multisim.

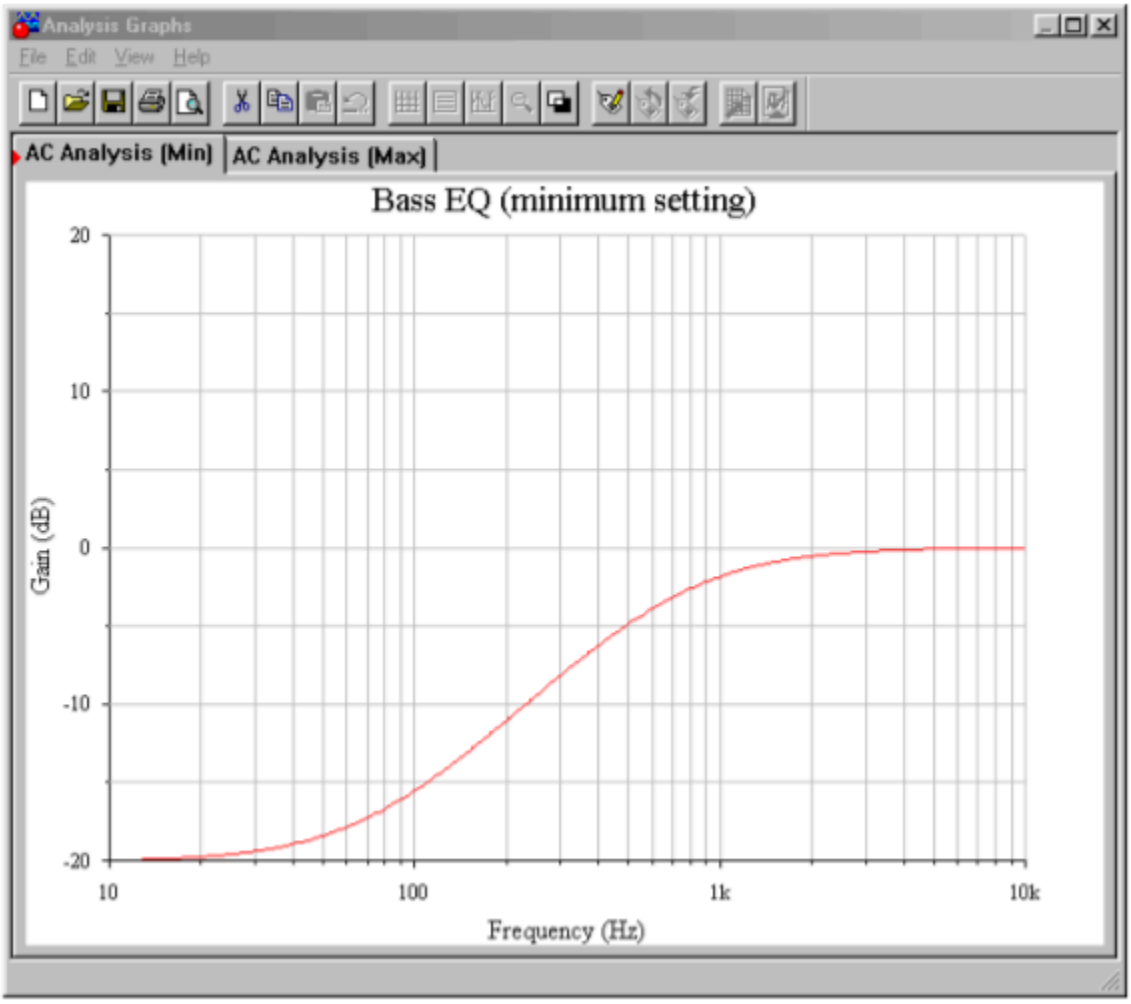


Figure 5 : Gain response at full cut.

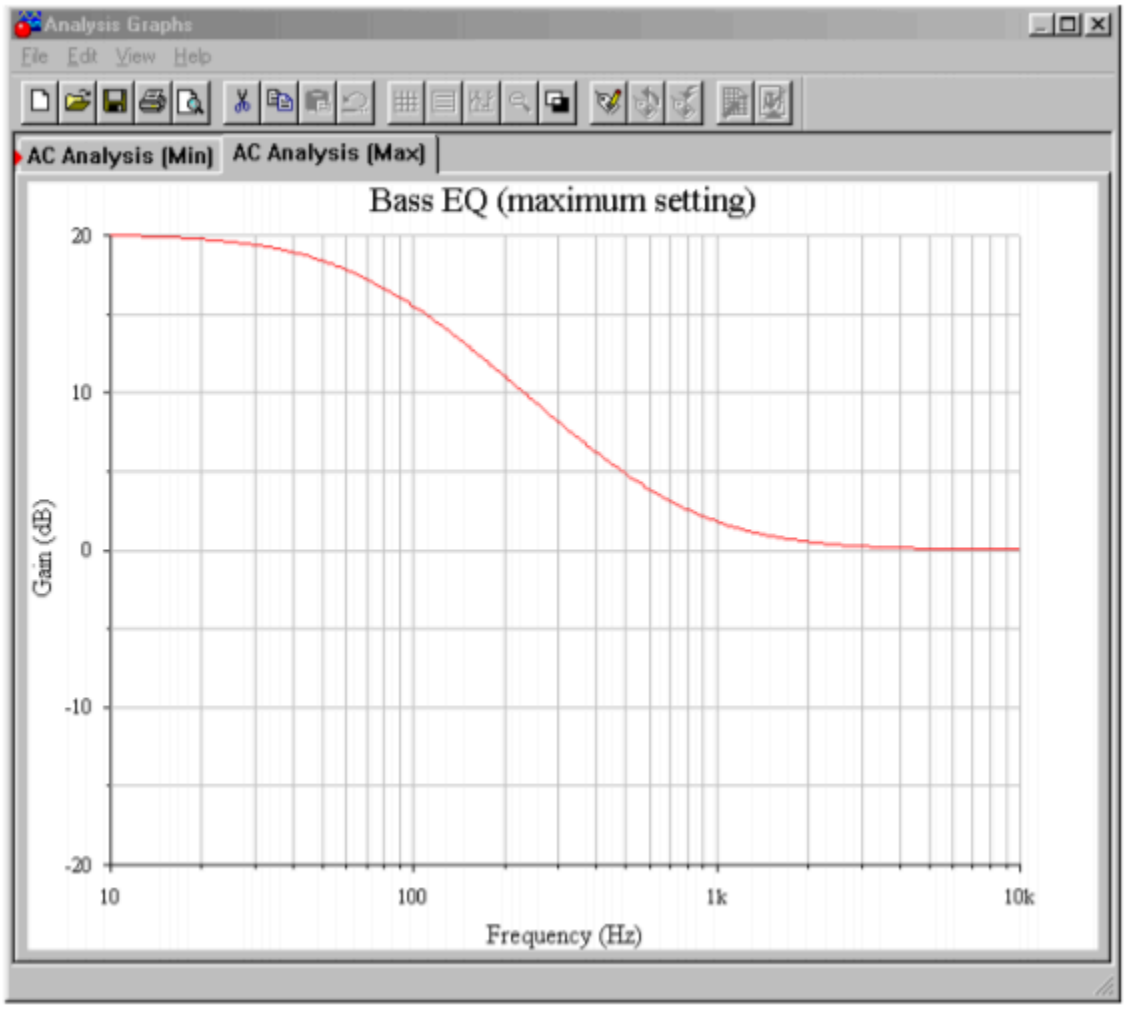


Figure 5 : Gain response at full boost.

In order to plot the response curves, two AC analysis runs are performed: the first with the potentiometer at maximum, and the second at minimum. The two curves are plotted from 10 Hz up to 10 kHz. These curves are essentially mirror images. In both cases the response above 1 kHz is smooth and reasonably flat. The transition to the cut/boost region occurs in the 100 Hz to 1 kHz range, with nearly full action by 40 Hz. If the plot is re-run with the potentiometer in some other position (say, 25 percent of rotation), some smooth, scaled curve within these two extremes will be plotted. If the potentiometer is set at mid-point, the response will be flat across the entire range.

15.10 SWITCHED-CAPACITOR FILTERS

Our final topic is the class of ICs known as switched-capacitor filters. These are just specific realizations of the types of filters that we have already examined. Generally, switched-capacitor filters come in two types: fixed order and alignment, and universal (state-variable based). A typical fixed IC might offer a sixth-order lowpass Butterworth filter. The number of external components required is minimal. The universal types offer most of the flexibility of the state-variable designs discussed previously. Both types are tunable and are relatively easy to use. Tuning is accomplished by adjusting the frequency of an external clock signal. The higher the clock frequency, the higher the resulting critical frequency. These ICs offer a convenient "black box" approach to general-purpose filter design. As is the case with most special-purpose devices, individual manufacturer's data sheets will give the specific application and design procedures for their parts.

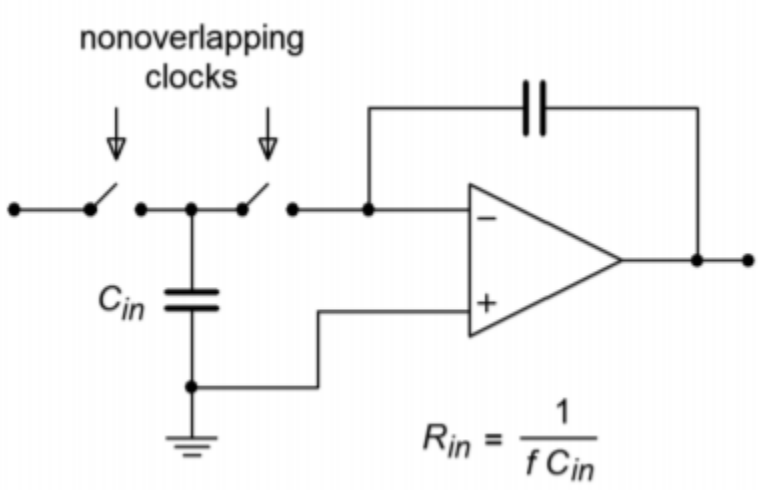


Figure 11.10.1 : Switched capacitor circuit.

The concept behind the switched-capacitor filter is quite interesting. The basic idea is to mimic a resistor through the use of a capacitor and a pair of alternating switches. As an example, a simple integrator is shown in Figure 11.10.1 . As you have already seen, integrators are little more than first-order, low-pass filters. In this circuit, the input resistor has been replaced with a capacitor, ⍰⍰⍰ , and a pair of switches. These switches are controlled by a non-overlapping bi-phase clock. This means that when one switch is closed, the other will be open, and that during switching, one switch will break contact before the other switch makes contact. (This is sometimes referred to as a "break before make" switch.)

During the first half of the clock cycle, ⍰⍰⍰ charges to the value of ⍰⍰⍰. During the second half of the cycle, this charge is transferred to the integration capacitor. Therefore, the total charge transferred during one clock cycle is

$$Q = C_{in}V_{in} \tag{11.10.1}$$

The flow of charge versus time defines current, so the average input current is

$$I_{in} = \frac{Q}{T_{clock}}$$

(11.10.2)

Substituting Equation 11.10.1 into Equation 11.10.2 yields

$$I_{in} = C_{in}\frac{V_{in}}{T_{clock}}$$

(11.10.3)

Because the input resistance is defined as the ratio of ◇◇◇to ◇◇◇ , and recognizing that ◇◇◇◇◇◇ is the reciprocal of ◇◇◇◇◇◇ , Equation 11.10.3 is used to find ◇◇◇ :

$$R_{in} = \frac{1}{C_{in}f_{clock}}$$

(11.10.4)

Equation 11.10.4 says two important things: first, because ◇◇◇ sets the input impedance, it follows that input impedance is inversely proportional to the clock frequency. Second, because ◇◇◇ is used to determine the corner frequency (in conjunction with the integration/feedback capacitor), it follows that the critical frequency of this circuit is directly proportional to the clock frequency. In other words, a doubling of clock frequency will halve the input impedance and double the critical frequency. Depending on the actual design of the IC, there will be a constant ratio between the clock frequency and the critical frequency. This is a very useful attribute. It means that you can make a tunable/sweepable filter by using one of these ICs and an adjustable square wave generator. For that matter, anything that can produce a square wave (such as a personal computer) can be used to control the filter response.

Typically, the ratio of clock frequency to critical frequency will be in the range of 50 to 100. The lower limit of clock frequency is controlled by internal leakage paths that create offset errors. The upper limit is controlled by switch settling time, propagation delays, and the like. A range of 100 Hz to 1 MHz is reasonable. This means that the entire audio frequency range is covered by these devices. For use at the highest frequencies, or with high impedance sources, an input buffer amplifier should be used.

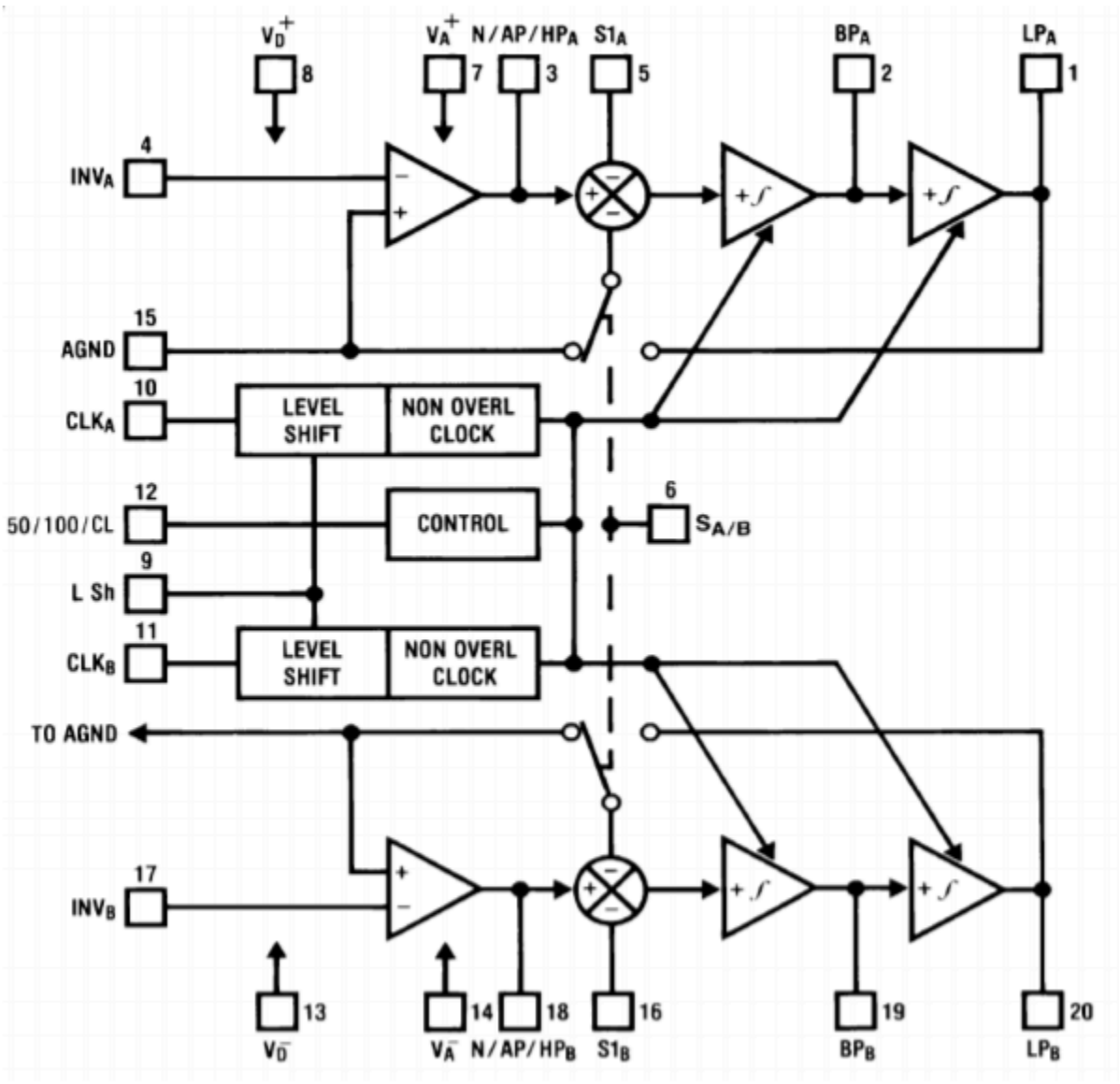


Figure 11.10.2 : Block diagram of the MF10. Reprinted courtesy of Texas Instruments

An example of a switched capacitor filter IC is the MF10, shown in Figure 11.10.2 . The MF10 is a dual second-order filter that can be connected in a variety of modes. Mode three is the general purpose state variable form and is shown in Figure 11.10.3 . Combining the two sections can produce fourth-order systems. In all cases, the external component count is minimal.

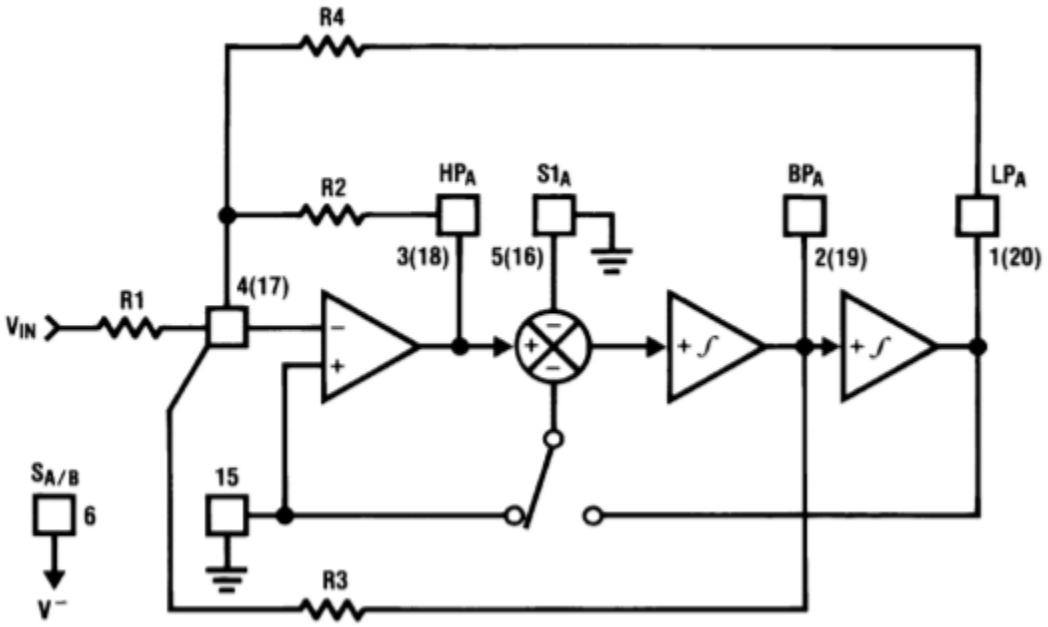


Figure 11.10.3 : MF10 mode 3 (state variable) equivalent. Reprinted courtesy of Texas Instruments

An example of a fourth-order filter is shown in Figure 11.10.4 . As you can see, the design is sparse, using only four resistors per section. In short, devices like the MF10 are good choices for general-purpose filter work, particularly when space and tuning considerations are important.

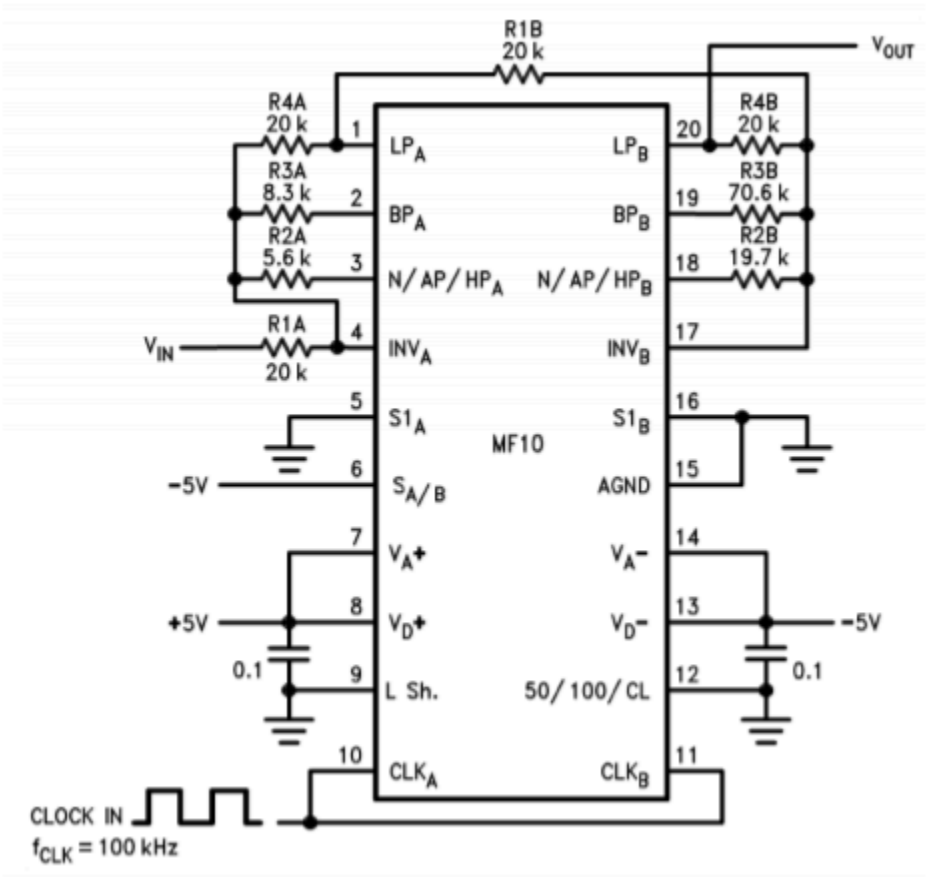


Figure 11.10.4 : MF10 fourth-order low-pass Chebyshev filter. Reprinted courtesy of Texas Instruments

Another device of interest is the LTC1068 from Linear Technology. The LTC1068 is a quad device, so a cascaded response of up to eighth-order is possible. Because these filters are basically little more than a switched capacitor version of the statevariable, a wide range of response types are possible. These include the high-pass, low-pass, band-pass and band-reject outputs, as well as a variety of alignments including Bessel, Butterworth, and Chebyshev. The LTC1068 is modular, so its internal op amps may be used as gain blocks or for creating a notch output without added external op amps (as seen in Example 11.8.1). Normally, no more than four or five external components (resistors and capacitors) are needed to realize a given filter function. Component calculation procedures are specified by the manufacturer.

Although switched-capacitor filters offer relatively quick and physically small realizations, they are not perfect. First of all, clock feedthrough is typically in the range of 10 mV, meaning that 10 mV of clock signal "leaks" into the output. Fortunately, this signal is much higher than the critical frequency, but may cause some problems for low-noise applications. Another problem arises from the fact that switched capacitor filters are actually sampled data devices. As you will see in the next chapter, sampled data devices may suffer from a distortion producing phenomenon called aliasing. In order to avoid aliasing, the input signal must not contain any components that are greater than one-half of the clock frequency. For example, if the MF10 is used to create a 1 kHz filter with a 50 kHz clock, no component of the input signal may exceed 25 kHz (one-half of the clock) if aliasing is to be avoided. If this requirement cannot be guaranteed, some form of pre-filtering is needed. Within these limits though, switched capacitor filters make light work out of many general-purpose applications.

15.11 VOLTAGE-CONTROLLED FILTERS (EXTENDED TOPIC)

A voltage-controlled filter, or VCF, is nothing more than a standard filter whose tuning frequency is controlled by an external voltage. You might think of this concept as an extension of the clock control aspects of the switched capacitor filter. VCFs are used in a wide range of applications including instrumentation devices such as swept frequency analyzers and music synthesizers. Any application that requires precise or rapid control of tuning frequency calls for a VCF. Virtually any of the filters presented in this chapter may be turned into VCFs. All you need to do is substitute the tuning elements of the filter with a voltage-controlled version. Typically, this means replacing the tuning resistors with voltage-controlled resistances.

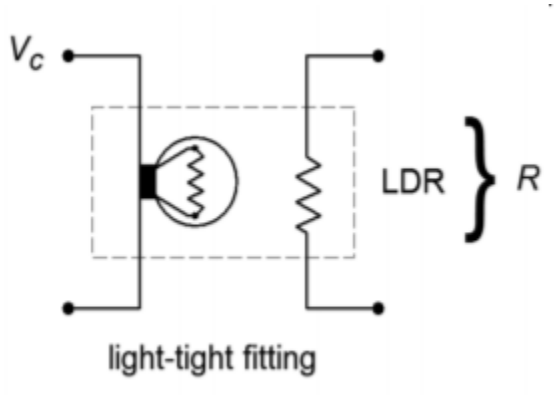


Figure 11.11.1 : A photoresistor/lamp used as a variable resistance.

Two possible ways of creating a voltage-controlled resistance include the photoresistor/lamp combination (Figure 11.11.1), and the use of an FET in its ohmic region (Figure 11.11.2). To use these items, simply remove the tuning resistor(s), and replace them with a voltage-controlled resistance. As an example, a simple single-pole high-pass VCF is shown in Figure 11.11.3 . As the control voltage (◈◈) increases, the lamp brightness increases causing the photoresistor's value to drop. Because the photoresistor sets the tuning frequency, the net result is an increase in ◈◈ . The FET version produces a resistance that is proportional to the magnitude of the gate voltage (◈◈).

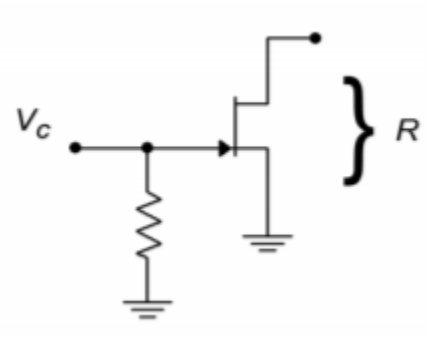


Figure 11.11.2 : Using a JFET in the ohmic region.

These two solutions are not without their problems. In the case of the lamp/photoresistor, response time is not very fast, and the lamp portion requires a fairly large drive current. The FET circuit eliminates these problems, but requires that the voltage across it remain fairly low (usually less than 100 mV). Larger signal swings will drive the FET out of the ohmic region, and distortion will increase dramatically. Also, the popular N channel variety requires a negative gate potential, which is generally not preferred. In both cases, one more problem remains: it is difficult to create a wide linear control range.

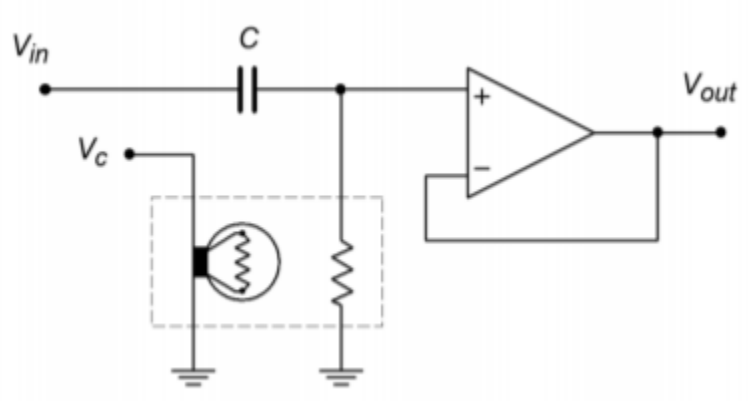


Figure 11.11.3 : Variable resistance connection for VCF.

Another way to create the effect of a voltage-variable tuning element is through the use of an operational transconductance amplifier, or OTA (see Chapter Six). Remember, this device is essentially a voltage-to-current converter. Its output current is a function of its control current. (The control currrent is easily derived from a control voltage and resistor.) This device is ideally suited to "inverting" type inputs, where an input resistor is used as a voltage-to-current converter. One possible example is shown in Figure 11.11.4 , a state-variable VCF. The boxed sections show where an OTA has replaced a standard single resistor. In this circuit, a large control voltage creates a large control current, thus increasing transconductance. This simulates a smaller tuning resistor value, and thus creates a higher tuning frequency. The OTA approach proves to be reliable, repeatable, and generally low in cost. It also offers a fairly wide linear tuning range.

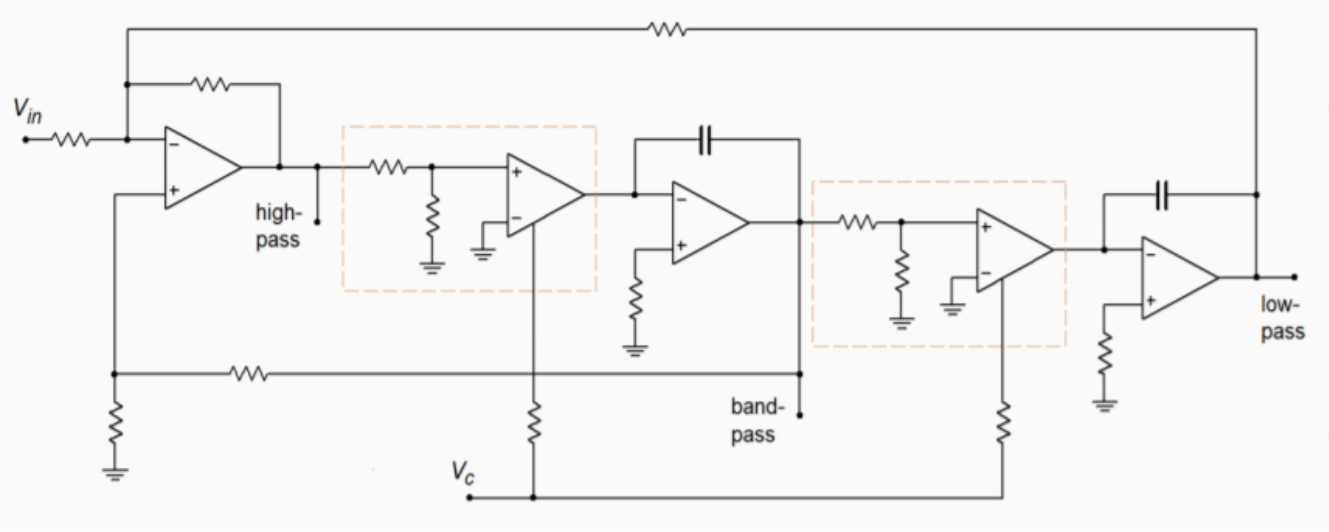


Figure 11.11.4 : Using OTAs as a controlled element in a VCF.

15.12 SUMMARY

Filters are frequency-selective circuits. The basic forms are high-pass, low-pass, band-pass, and band-reject. Although filters may be constructed solely from resistors, capacitors, and inductors, active filters using op amps offer many advantages. These advantages include: modest component size, control over impedances and loading effects, elimination of inductors, and gain (if desired). The negative aspects include: frequency range limited by op amps used, power supply required, and the inability to handle large input/output powers. For many applications the advantages far outweigh the disadvantages, and therefore, active filters are used in a wide variety of modern products.

Filters are further defined by order and alignment. Order indicates the steepness of the attenuation slope. As a general rule, the eventual rolloff rate will equal 6 dB times the order, per octave. Order also indicates the minimum number of reactive elements needed to realize the filter. Alignment indicates the shape of the filter response in the frequency domain. Popular alignments include Bessel (constant time delay), Butterworth (maximally flat response in the pass band), and Chebyshev (ripples in the pass band, but with faster roll off rates). There is generally a trade-off between fast attenuation rates and smooth phase response. Alignment is indicated by the damping or ◈ of the filter. ◈ is the reciprocal of damping. Filters with low damping factors (i.e., high ◈) tend to be "peaky" in the frequency domain and produce ringing on pulse-type inputs. (Chebyshevs are in this category.) The filter's critical frequency and 3 dB down frequency are not the same for alignments other than the Butterworth. The actual amount of "skew" depends on the alignment and order of the filter.

Once filter performance is specified, there are a number of ways in which the circuit can be physically realized. Common high- and low-pass realizations use the Sallen and Key VCVS approach. There are two variations on this theme: the unity-gain form and the equal-component form. Both forms use a second-order building block section. For higher orders, several second-order sections (and optionally, a first-order section) are combined to produce the final filter. It is important to remember that higher-order filters are not simple combinations of identical lower-order filters. For example, a fourth-order 1 kHz Butterworth filter is not made by cascading a pair of identical second-order 1 kHz Butterworth filters. Rather, each section requires specific damping and frequency factors. A common design procedure utilizes lookup tables for these factors. The filters are designed by first scaling the general filter to the desired cutoff frequency, and then scaling the components for practical values.

For relatively low ◈ s (<1), band-pass filters are best realized as a cascade of high- and low-pass filters. For higher ◈ s, this technique is not satisfactory. Moderate ◈ s (up to 10) may be realized with the multiple-feedback filter. Very high ◈ applications (up to 100) may be realized with the state-variable filter. The state-variable is often known as the universal filter, as it produces high-, low-, and band-pass outputs. With the addition of a fourth amplifier, a band-reject filter may be formed. Fixed and adjustable gain versions of the state-variable may be utilized by the designer.

A somewhat more specialized group of filters are the equalizers commonly employed in audio recording and playback equipment. Unlike traditional filters, equalizers offer both boost and

attenuation of frequencies. Generally, these circuits are based on parallel-parallel inverting amplifiers, utilizing an adjustable, frequency-selective feedback network.

Switched capacitor filter ICs offer the designer expedient solutions to general-purpose filter design. They are generally suited to the audio frequency range and require very few external components. The critical frequency is set by a clock input. The order and alignment may be either factory set or user adjustable (as in the universal state-variable types).

REFERENCES

Lancaster, Don Active Filter Cookbook, 2nd e, Woburn, Mass.: ButterworthHeinemann, 1996.

Bohn, Dennis ed. Audio Handbook, Santa Clara, Ca.: National Semiconductor Corporation, 1976.

Tedeschi, Frank P. The Active Filter Handbook, TAB Books, 1979.

Audio Handbook, Norwood, Mass.: Precision Monolithics Inc., 1990.

Jung, Walter G. Audio IC Op Amp Applications, 2nd e, Indianapolis, Ind.: Howard W. Sams, 1978.

15.13 PROBLEMS

ANALYSIS PROBLEMS

1. Using Figure 11.5.1b, determine the loss for a 1 kHz Butterworth secondorder low-pass filter at 500 Hz, 1 kHz, 2 kHz, and 4 kHz.
2. Using Figure 11.6.18, determine the loss for a 2 kHz 3 dB ripple Chebyshev third-order low-pass filter at 500 Hz, 1 kHz, 2 kHz, 4 kHz, and 6 kHz.
3. Using Figure 11.5.1b, determine the loss for a 500 Hz Bessel second-order high-pass filter at 200 Hz, 500 Hz, and 2 kHz.
4. Using Figure 11.6.18, determine the loss one octave above the cutoff frequency for a fourth-order low-pass filter of the following alignments: Butterworth, Bessel, 1 dB-ripple Chebyshev.
5. Repeat Problem 4 for high-pass filters.
6. Using Figures 11.5.1b and 11.6.18, determine the loss at 8 kHz for 3 kHz lowpass Butterworth filters of orders 2 through 6.
7. Using Figures 11.5.1b and 11.6.18, determine the loss at 500 Hz for 1.5 kHz high-pass Bessel filters of orders 2 through 6.
8. Repeat Problem 7 using 3 dB-ripple Chebyshevs.
9. Using Figures 11.5.1b and 11.6.18, determine the loss at 500 Hz for 200 Hz high-pass 3 dB-ripple Chebyshev filters of orders 2 through 6.
10. Using Figure 11.6.18, determine the loss one octave below the cutoff frequency for a third-order high-pass filter of the following alignments: Butterworth, Bessel, 1 dB-ripple Chebyshev.
11. Repeat Problem 10 for low-pass filters. 12. An application requires that the stop-band attenuation of a low-pass filter be at least –15 dB at 1.5 times the critical frequency. Determine the minimum order required for the Butterworth and 1 dB and 3 dB-ripple Chebyshev alignments.
12. An application requires that the stop-band attenuation of a high-pass filter be at least –20 dB one octave below the critical frequency. Determine the minimum order required for the Butterworth and 1 dB-ripple and 3 dB-ripple Chebyshev alignments.
13. A band-pass filter has a center frequency of 1020 Hz and a bandwidth of 50 Hz. Determine the filter ◈ .
14. A band-pass filter has upper and lower break frequencies of 9.5 kHz and 8 kHz. Determine the center frequency and ◈ of the filter.
15. Design a second-order Butterworth low-pass filter with a critical frequency of 125 Hz. The pass band gain should be unity.

16. Repeat Problem 16 for a high-pass filter.
17. Repeat Problem 16 using a Bessel alignment.

DESIGN PROBLEMS

19. 19. A particular application requires that all frequencies below 400 Hz should be attenuated. The attenuation should be at least –22 dB at 100 Hz. Design a filter to meet this requirement.
20. 20. Repeat Problem 19 for an attenuation of at least –35 dB at 100 Hz.
21. 21. Audiophile quality stereo systems often use subwoofers to reproduce the lowest possible musical tones. These systems typically use an electronic crossover approach as explained in Example 11.6.4. Design an electronic crossover for this application using third-order Butterworth filters. The crossover frequency should be set at 65 Hz.
22. 22. Explain how the design sequence of Problem 21 is altered if either a new crossover frequency is chosen, or a different alignment is specified.
23. 23. Design a band-pass filter that will only allow frequencies between 150 Hz and 3 kHz. The attenuation slopes should be at least 40 dB per decade. (A filter such as this is useful for "cleaning up" recordings of human speech.)
24. 24. Design a band-pass filter with a center frequency of 2040 Hz and a bandwidth of 400 Hz. The circuit should have unity gain. Also, determine the ◈◈◈◈◈◈ requirement of the op amp(s) used.
25. 25. Repeat Problem 24 for a center frequency of 440 Hz and a bandwidth of 80 Hz.
26. 26. Design a band-pass filter with upper and lower break frequencies of 700 Hz and 680 Hz.
27. 27. Design a notch filter to remove 19 kHz tones. The ◈ of the filter should be 25. (This filter is useful in removing the stereo "pilot" signal from FM radio broadcasts.)
28. 28. Design a second-order low-pass filter with a critical frequency of 30 kHz. Use a state-variable filter. The circuit should have a gain of +6 dB in the pass band.
29. 29. Design a bass/treble equalizer to meet the following specification: maximum cut and boost = 25 dB below 50 Hz and above 10 kHz.
30. 30. Using the MF10, design a fourth-order low-pass Butterworth filter with a critical frequency of 3.5 kHz.
31. 31. Using the MF10, design a low-pass filter that is adjustable from 200 Hz to 10 kHz. Do not ignore the oscillator design.

CHALLENGE PROBLEMS

31. 32. Design a low-pass second-order filter that may be adjusted by the user from 200 Hz to 2 kHz. Also, make the circuit switchable between Butterworth and Bessel alignments.
32. 33. Design a subsonic filter that will be 3 dB down from the pass band response at 16 Hz. The attenuation at 10 Hz must be at least 40 dB. Although pass band ripple is permissible, the gain should be unity.

33. 34. Design an adjustable band-pass filter with a ◈ range from 10 to 25, and a center frequency range from 1 kHz to 5 kHz.
34. 35. Modify the design of the previous problem so that as the ◈ is varied, the pass-band gain remains constant at unity.

COMPUTER SIMULATION PROBLEMS

35. Verify the magnitude response of the circuit designed in Problem 16 by using a simulator. Check both the critical frequency and the roll off rate.
36. Verify the magnitude response of the electronic crossover designed in Problem 21 by using a simulator. Plot both outputs simultaneously on one graph.
37. Verify the magnitude and phase of the filter designed in Problem 24 by using a simulator.
38. Compare the simulations of the circuit designed in Problem 28 using the relatively slow LM741, versus the medium-speed LF411. Is there any noticeable change? What can you conclude from this? Would the results be similar if the break frequency was increased by a factor of 50?
39. It is very common to plot the adjustment range of equalizers on a single graph, as shown in Figure 12-49. Use a simulator to create a plot of the adjustment range of the equalizer designed in Problem 29.
40. Simulate and verify the design of challenge Problem 32.
41. Simulate and verify the design of challenge Problem 33.
42. Verify the design of challenge Problem 34 using a simulator. Include four separate plots, showing maximum and minimum ◈ with maximum and minimum center frequency.
43. Verify the design of challenge Problem 35 using a simulator. Include two simultaneous plots, one showing minimum ◈ with maximum and minimum center frequency, and the other showing maximum ◈ with maximum and minimum center frequency

UNIT 16: ANALOG-TO-DIGITAL-TO-ANALOG CONVERSION

Learning Objectives

After completing this chapter, you should be able to:

- Outline the concept of pulse code modulation.
- Detail the advantages and disadvantages of signal processing in the digital domain.
- Define the terms resolution, quantization, and Nyquist frequency.
- Define an alias, and detail how it is produced and subsequently avoided.
- Explain the operation of an ◈/2◈R/2R digital-to-analog converter.
- Explain the need for anti-alias and reconstruction filters.
- Explain the operation of a successive approximation analog-to-digital converter.
- Detail the need for and operation of a track-and-hold amplifier.
- Explain the operation of a flash analog-to-digital converter.
- Compare the different analog-to-digital converters in terms of speed, size and complexity, and detail typical applications for each.

16.1 INTRODUCTION TO ANALOG TO DIGITAL TO ANALOG CONVERSION

Up to now, all of the circuits you have studied in this book were analog circuits. That is, the input waveforms were time-continuous and had infinite resolution along the time and amplitude axes. That is, you could discern increasingly smaller and finer changes as you examined a particular section. No matter whether the circuit was a simple amplifier, function synthesizer, integrator, filter, or what have you, the analog nature of the signal was always true. Fundamentally, the universe is analog in nature (at least as far as we can tell – until someone discovers a quantum time particle). Our only real deviation from the pure analog system was the use of the comparator. Although the input to the comparator was analog, the output was decidedly digital; its output was either a logic high (+[illegible]) or a logic low (–[illegible]) . You can think of the comparator's output as having very low resolution, as only two states are possible. The comparator's output is still time-continuous in that a logic transition can occur at any time. This is in contrast to a pure digital system where transitions are time-discrete. This means that logic levels can only change at specific times, usually controlled by some form of master clock. A purely digital system then, is the antithesis of a pure analog system. An analog system is time-continuous and has infinite amplitude resolution. A digital system is time-discrete and has finite amplitude resolution (two states in our example).

As you have no doubt noticed in your parallel work, digital systems have certain advantages and benefits relative to analog systems. These advantages include noise immunity, storage capability, and available numeric processing power. It makes sense, then, that a combination of analog and digital systems could offer the best of both worlds. This chapter examines the processes of converting analog signals into a digital format and turning digital words into an analog signal. A few representative examples of processing the signal in the digital domain are presented as well. Some examples with which you might already be familiar include the stereo compact disk (CD) and the digital storage oscilloscope. We will break down this topic into two broad sections: analog-to-digital conversion (AD) and digital-to-analog conversion (DA). Since many AD systems require digital-to-analog converters, we will examine DA systems first.

THE ADVANTAGES AND DISADVANTAGES OF WORKING IN THE DIGITAL DOMAIN

Given enough time, an analog circuit may be designed and manufactured for virtually any application. Why, then, would anyone desire to work in the digital domain? Perhaps the major reason for working in the digital domain is the flexibility it offers. Once signals are represented in a digital form, they may be manipulated by various means, including software programs. You have probably discovered that replicating a computer program is far easier than replicating an analog circuit. What's more, a program is much easier to update and customize than a hardware circuit. Because of this, it is possible to manipulate a signal in many different ways with the same digital/computer hardware; all that needs to be altered is the manipulation instructions (i.e., the program). The analog circuit, in contrast, needs to be re-wired, and extra components need to be added or old portions removed. This can be far more costly and time-intensive than just updating software. By working in the digital domain, processing circuits do not exist per se; rather, a generic IC such as a CPU is used to create a "virtual circuit".

With a certain amount of intelligence in the system design, the virtual circuit may be able to alter its own performance in order to precisely adapt to various signals. This all boils down to the fact that a digital scheme may offer much greater flexibility for involved tasks and allows a streamlined, generic hardware solution for complex applications. Because of this attribute, the digital solution may wind up being significantly less expensive than its analog counterpart.

When an analog signal is transferred to the digital domain, it is represented as a series of numbers (usually, high/low binary logic levels). One nice property of this representation is that it is exactly repeatable. In other words, an infinite number of copies of the data may be generated, and no distortions or deviations from the original will appear. The last copy will be identical to the first. Compare this to a simple analog copy. For example, if you were to record a song with a cassette recorder and then make a copy of the tape, the second-generation copy would suffer from increased noise and distortion. A copy of the second copy would produce even worse results. Every time the signal is copied, some corruption occurs. It is for this reason that early long distance telephone calls were of such low quality. Modern communications systems employ digital techniques that allow much higher quality, even if one person is in New York and the other is in Australia, halfway around the planet.

Besides being a desirable mathematical attribute, repeatability also lends itself to the problem of long term storage. A storage medium for a binary signal only needs to resolve two levels, whereas the analog medium needs to resolve very fine changes in signal strength. As you might guess, deterioration of the analog medium is a serious problem and results in information loss. The digital medium can theoretically survive a much higher level of deterioration without information loss. In a computerized system, data may be stored in a variety of formats including RAM (Random Access Memory) and magnetic tape or disk. For playback only (i.e., read only), data may be stored in ROM (Read Only Memory) or laser disk formats (such as DVD or audio CD).

As always, the benefits of the digital scheme arrive with specific disadvantages. First, for simpler applications, the cost of the digital approach is very high and cannot be justified. Second, the process of converting a signal between the analog and digital domains is an inexact one. Some information about the signal will be lost during the conversion. This is because the digital representation has finite resolution. This means that only signal changes larger than a certain minimum size (the resolution step size) are discernable, and therefore, some form of round-off error is inevitable. This characteristic helps determine the range of allowable signals, from the smallest detectable signal to the maximum signal before overload occurs. Third, analog systems are inherently faster than digital systems. Analog solutions can process input signals at much higher frequencies than digital schemes. Also, analog systems work in real-time, whereas digital systems might not. Digital systems can only perform in real-time if the input signal is not a very high frequency, if the processing task is not overly complex, or if specialized processing circuits are added. Not all applications require real-time performance, so this limitation is not always a problem. Also, because we can expect computing power to get less and less expensive in the coming years, cost-effective digital processing will undoubtedly expand into new areas.

16.2 THE SAMPLING THEOREM

There are many different ways in which an analog signal may be turned into a digital form. This process is referred to as AD conversion or, more simply, as digitization. We will only examine the most popular method, called pulse code modulation, or PCM for short. The reverse of this process, or turning the digital information back into analog form, is called DA conversion. The two most important charracteristics in the conversion process are sampling frequency and amplitude resolution. Let's look at the basic idea behind PCM.

In essence, PCM measures and encodes the value of the input signal at specific points in time. Normally, the time spacing is constant, and several points are used over the length of one input cycle. In other words, the process involves taking representative samples of the input signal over time. This is shown graphically in Figure 12.2.1 , and is referred to as sampling. The result of the sampling procedure is a list of times and corresponding amplitude values. A sequence might look something like: at ◇=1 ms, ◇◇◇=23 mV; at ◇=2 ms, ◇◇◇=45 mV; at ◇=3 ms, ◇◇◇=–15 mV; etc. If the time interval between samples is held constant (i.e., constant rate of sampling), then all we need to know is a starting time and the sampling rate in order to reconstruct the actual sample times.[1] This is much more efficient than recording each sample time. The resulting amplitude values may be manipulated in a variety of ways since they are now in a numeric form. The ultimate accuracy of this conversion will depend on two primary factors: how often we sample the signal and the accuracy and resolution of the sample measurement. Theoretically, the conversion will never be 100% accurate; that is, once converted, a finite amount of information will be lost forever. Another way of stating this is that when the digital representation is converted back to analog, the result will not be identical to the original waveform. In practical terms, it is possible to reduce the error to such small values that it may be ignored in many applications. It is important then, that we investigate the implications of sample rate and accuracy/resolution on the quality of conversion.

1. The terms sample rate and sample frequency are often used interchangeably, and are usually denoted by ◇◇ .

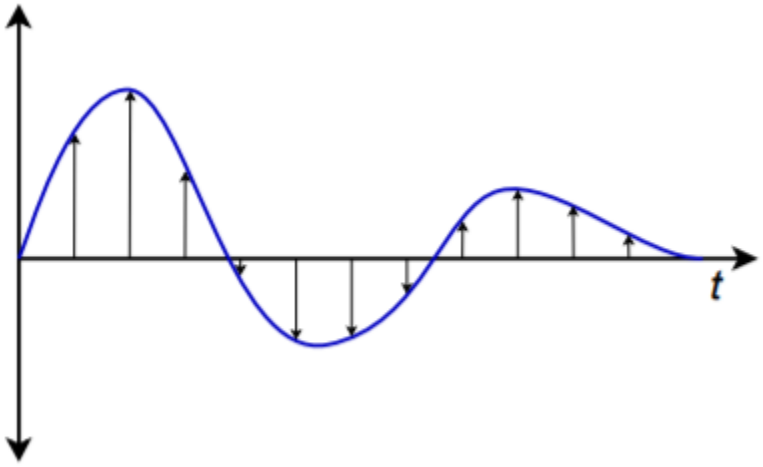


Figure 12.2.1 : Sampling input signals over time.

16.3 RESOLUTION AND SAMPLING RATE

Perhaps the most obvious source of error is the finite measurement accuracy of the individual signal levels. The major problem here is one of resolution. Resolution represents the finest discernible change in the signal and is often specified in terms of a number of bits, although a voltage specification is also possible. Because the signal level is represented with a binary number, it follows that a large number of bits are needed in order to achieve fine resolution. For example, if an eight-bit word is used, there will be 256 distinct values available. If the maximum peak-to-peak value of the input signal is 1 V, it works out that each step in the word represents about 3.9 mV (1 V/256). Under these conditions, it would be impossible to perfectly encode a value of 14 mV. The nearest values available would be binary 11, which yields 11.7 mV (3 times 3.9 mV), and binary 100, which yields 15.6 mV (4 times 3.9 mV). Obviously, the resulting round-off creates some error in the digital representation. This error can be reduced by increasing resolution so that finer steps may be detected. If 16 bits are used for the same 1 V range, a total of 65,536 values are available, with each step working out to 15.26 ◈ V. Now, although we still may not be able to exactly represent the 14 mV level, we are guaranteed to be within ± 7.63 ◈ V, instead of ± 1.95 mV as in the eight-bit case. Because the round-off errors tend to be random in magnitude and polarity, this effect may be viewed as a noise source. In other words, lower resolutions (i.e., fewer bits) produce noisier signals. The number of bits required for a particular application may vary from fewer than 6 for high speed video applications to more than 20 for high quality audio or measurement purposes.

It is important to note that once the size of the digital word is chosen and the peak amplitude fixed, the input signal must stay within specific bounds or gross distortion will occur. For example, if a peak input of 1 V produces the maximum numeric value, there is no way that the digital word can represent a level greater than 1 V. Likewise, if the step size is set at 1 mV, any signals less than 1 mV are lost. Also, low-level signals will suffer from reduced resolution. For best results, the signal peak should produce the maximum numeric value. If the peak is significantly less, the result is akin to using fewer bits in the representation. Finally, even if the resolution is adequate, the absolute accuracy of the conversion must be considered, as it is in any measurement device.

Examples

A certain system uses a 12-bit word to represent the input signal. If the maximum peak-to-peak signal is set for 2 V, determine the resolution of the system and its dynamic range.A 12-bit word means that 212 , or 4096, levels are possible. As these levels are equally spaced across the 2 V range, each step is

$$\text{Step Size} = \frac{2V}{4096}$$

$$\text{Step Size} = 488\mu V$$

Therefore, the system can resolve changes as small as 488 ◈ V.

Dynamic range represents the ratio of the largest value possible to the smallest.

$$\text{Dynamic Range} = \frac{2V}{488\mu V}$$

$$\text{Dynamic Range} = 4096$$

$$\text{Dynamic Range} = 20\log_{10} 4096$$

$$\text{Dynamic Range} = 72dB$$

In the preceding problem, note that the voltage range affects the step size, but does not affect the bit resolution. The actual number of discrete steps that may be resolved is set by the number of bits available. You may notice that each additional bit adds approximately 6 dB of range (a doubling of voltage). Consequently, the dynamic range calculation may be streamlined to

$$\text{Dynamic Range} \approx 6dB \times \text{Number of Bits} \tag{12.3.1}$$

Example 12.3.2

Audio compact disks use a 16-bit representation of the music signal. Determine the dynamic range. Also, if the maximum output level is 0.775 V peak, determine the step size.

$$\text{Dynamic Range} \approx 6dB \times \text{Number of Bits}$$

$$\text{Dynamic Range} \approx 6dB \times 16$$

$$\text{Dynamic Range} \approx 96dB$$

For 16 bits, the total number of steps is 216 , or 65536. Assuming that the signal is bipolar, the total signal range will be from –0.775 V to +0.775 V, or 1.55 V.

$$\text{Step Size} = \frac{1.55V}{65536}$$

$$\text{Step Size} = 23.65\mu V$$

At this point, we must consider the effect of sampling rate on the quality of the signal. It should be intuitively obvious that higher sampling rates afford greater overall conversion accuracy. Of course, there is a trade-off associated with high sampling rates, and that is the accompanying high data rate. In other words, greater resources will be required to store and process the larger volume of digital information. The real question is, just how fast does the sampling rate need to be, for optimum efficiency? The Nyquist-Shannon sampling theorem states that at least two samples are needed per cycle for proper signal conversion. If the signal is not a sinusoid, then at least two samples are required per cycle of the highest frequency component. For example, if a range of signals up to 10 kHz needs

to be digitized, then a sample rate of at least 20 kHz is required. Normally, a certain amount of "breathing room" is added to this figure. Another way of looking at this relationship is to state that the highest input-frequency component can be no more than one-half of the sampling rate. Because this is such an important parameter, the value of onehalf of the sampling rate is given the name Nyquist frequency.

$$\text{Nyquist frequency} = \frac{f_s}{2} \tag{12.3.2}$$

If an input frequency component is greater than the Nyquist frequency, a unique form of distortion called alias distortion, is produced. The resulting distortion product, called an alias, is a new signal at a frequency that is equal to the difference between the input and Nyquist frequencies. Normally, this new signal is not harmonically related to the input signal, and thus, is easily detected. The aliasing effect is shown graphically in Figure 12.3.1 .

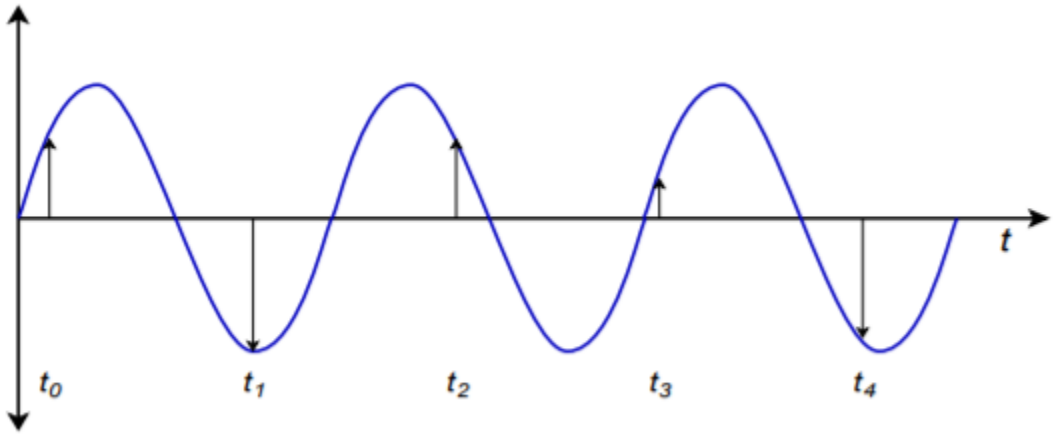


Figure 12.3.1 : The aliasing effect (sampling rate too low).

Here we see a sampling rate that is only about 1.5 times the input frequency, rather than the required factor of 2 times minimum. In Figure 12.3.2 the sample points are redrawn and connected as simply as possible. Note that the resulting outline is that of a lower-frequency wave. What we notice here is that the data points produced in Figure 12.3.1 are identical to the points produced by a lower-frequency input wave. When these data points are converted back to analog form, the DA converter will produce this lower frequency wave. Oddly enough, the original waveform has completely disappeared; hence the term alias. Any signal component that is greater than the Nyquist frequency will produce aliases.

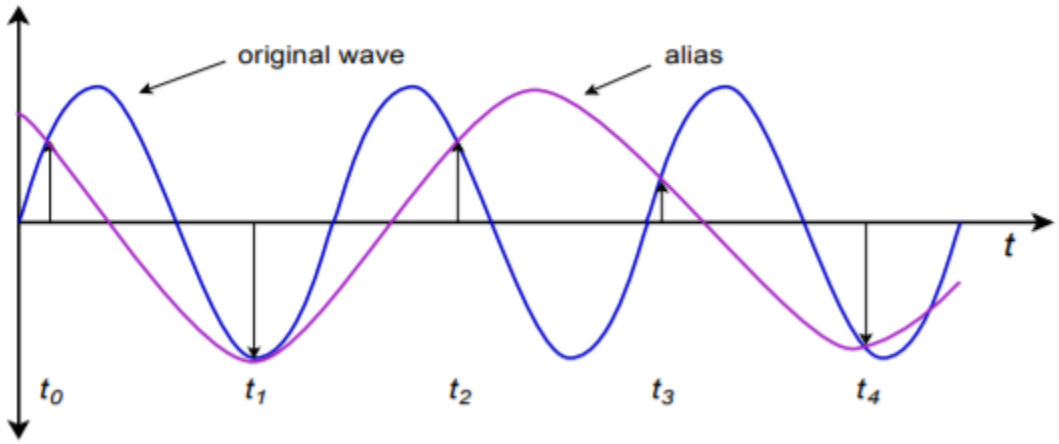


Figure 12.3.2 : Alias production.

When considering the possibility of alias distortion, it is worth repeating that the components of the signal must be investigated, not just the base frequency. For example, a 1 kHz square wave has a 1 kHz fundamental and an infinite series of odd-numbered harmonics (3 kHz, 5 kHz, 7 kHz, 9 kHz, etc.). If a 12 kHz sampling rate is used to digitize this signal, a fair amount of alias distortion will be seen. In this example, the Nyquist rate is 6 kHz. All of the harmonics above 6 kHz will produce an alias. In order to prevent this, the input signal must be frequency-band limited. That is, a low-pass filter must be used to attenuate all components above the Nyquist frequency before AD conversion takes place. The amount of filtering required depends on the resolution of the conversion and the relative strength of the above-band signals. Because filters cannot roll off infinitely fast, as noted in Chapter Eleven, sampling rates are normally set more than twice as high as the highest needed input component. In this way, the Nyquist frequency will be somewhat greater than the maximum desired input frequency. The low-pass filter (often referred to as an anti-alias filter) will use this frequency range as its transition band. Even though the attenuation in the transition band is less than optimum, alias distortion will not be a problem. This is shown graphically in Figure 12.3.3 . As you can see, high-roll-off rate filters are desirable in order to attenuate the out-of-band signals as quickly as possible. Very fast filter roll off rates mean that the sampling frequency need only be as little as 10% greater than the theoretical minimum in order to maintain sufficient alias rejection.

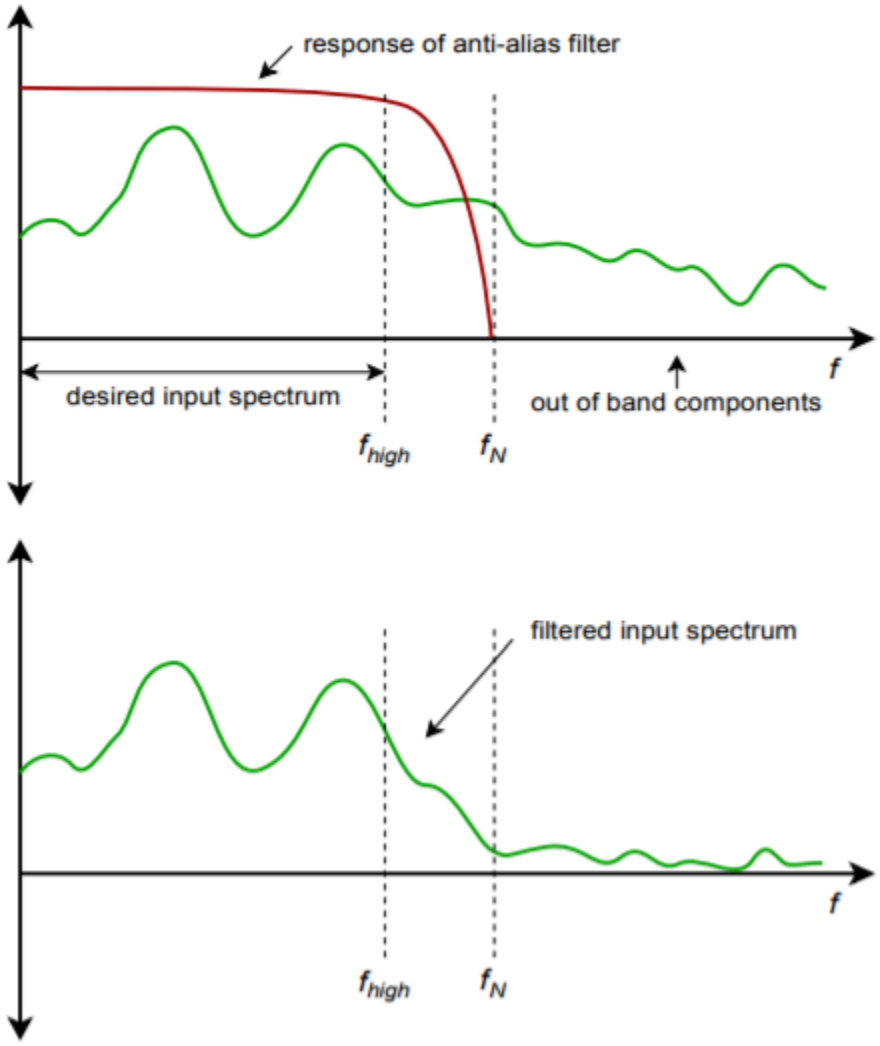


Figure 12.3.3 : Filtering to eliminate aliasing effects.

Example 12.3.3

Suppose that you need to digitize telephone signals. Assuming that you would like to maintain a dynamic range of at least 50 dB, with an upper frequency limit of 3 kHz, determine the minimum acceptable sampling rate and number of bits of required. The minimum Nyquist rate is equal to the highest desired input frequency. In this case, that's 3 kHz. Because the sampling rate is twice the Nyquist frequency, the sampling rate must be at least 6 kHz. In reality, if input components above 3 kHz exist, an anti-alias filter will be needed, and the sampling rate will have to be increased somewhat. Because dynamic range is set by the number of bits used, we find that

$$\text{Bits Required} = \frac{\text{Dynamic Range}}{6dB}$$

$$\text{Bits Required} = \frac{50dB}{6dB}$$

$$\text{Bits Required} = 9$$

Because we cannot have a fractional bit, the value is rounded up. The final system specification is a minimum rate of 6 kHz with a 9-bit resolution. Note that this represents a data rate of 9 bits per sample times 6000 samples per second, or 54000 bits per second (6750 bytes per second).

The basic digital-to-analog converter is little more than a weighted summing amplifier. Each successive bit in the digital word represents a level that is twice as large as the preceding bit. If each bit is taken as a given current or voltage, the increasing levels may be produced by using different gains in the summing inputs. A simple four-bit converter is shown in Figure 12.4.1 .

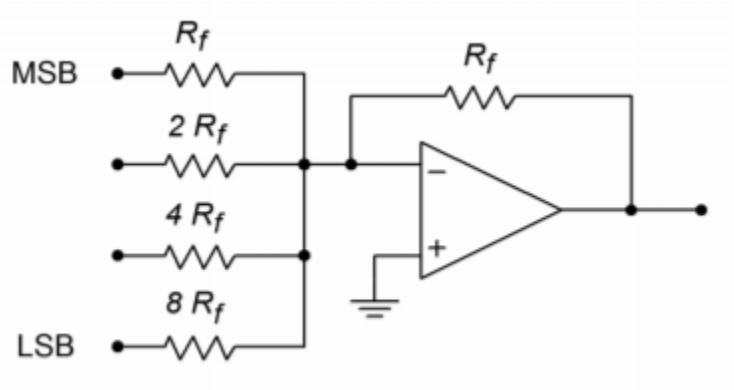


Figure 12.4.1 : A simple 4-bit converter.

This system can represent 24 , or 16, different levels. Each input is driven by a simple high/low logic level that represents a 1 or 0 for that particular bit. Note that the input resistors vary by factors of 2. The gain for the upper most path is ⓡⓕ/ⓡⓕ , or unity. This input is used for the most significant bit of the input word (MSB). The next input shows a gain of ⓡⓕ/(2ⓡⓕ) , or 0.5. The third input shows a gain of 0.25, and the final input shows a gain of 0.125. The final input has the lowest gain and is used for the least significant bit of the input word (LSB). If the input word had a higher resolution (i.e., more bits), extra channels would be added, each having half the gain of the preceding input. To better understand the conversion process, let's take a look at a few representative inputs and outputs.

The circuit of Figure 12.4.1 may be driven by simple 5 V TTL-type logic circuits. 5 V represents a logical high, whereas 0 V represents a logical low. What is the output level if the input word is 0100? Because a logical high represents 5 V, 5 V is being applied to the second input. All other inputs receive a logical low, or 0 V. The output is the summation of the input signals (remember, this is an inverting summer, so the final output should have its sign reversed).

$$V_{out} = -(V_{in1}A_1 + V_{in2}A_2 + V_{in3}A_3 + V_{in4}A_4)$$
$$V_{out} = -(0V \times 1 + 5V \times 0.5 + 0V \times 0.25 + 0V \times 0.125)$$
$$V_{out} = -2.5V$$

So, a value of 4 (binary 100) is equivalent to a potential of 2.5 V. If we increase the word value to 9 (binary 1001), we see

$$V_{out} = -(V_{in1}A_1 + V_{in2}A_2 + V_{in3}A_3 + V_{in4}A_4)$$
$$V_{out} = -(5V \times 1 + 0V \times 0.5 + 0V \times 0.25 + 5V \times 0.125)$$
$$V_{out} = -5.625V$$

The minimum output occurs at binary 0000, (0 V) and the maximum at binary 1111 (–9.375 V). The

step size is equal to the logic level times the minimum gain; in this case that's 0.625 V. Notice that the output value may be found by simply multiplying the value of the input word by the minimum step size. Also, it is important to note that the output signal is unipolar (in this example, always negative).

A digital representation, of course, is made up of a sequence of words, not just one word. In reality, the logic circuits are constantly feeding the summing amplifier new words at a predetermined rate. Because of the changing inputs, the output of the converter is constantly changing as well. Using our previously calculated values, if the converter is fed the sequence 0000,0100,1001,1111, the output will move from 0 V to –2.5 V, to –5.625 V, to a final value of –9.375 V. This output is graphed in Figure 12.4.2 .

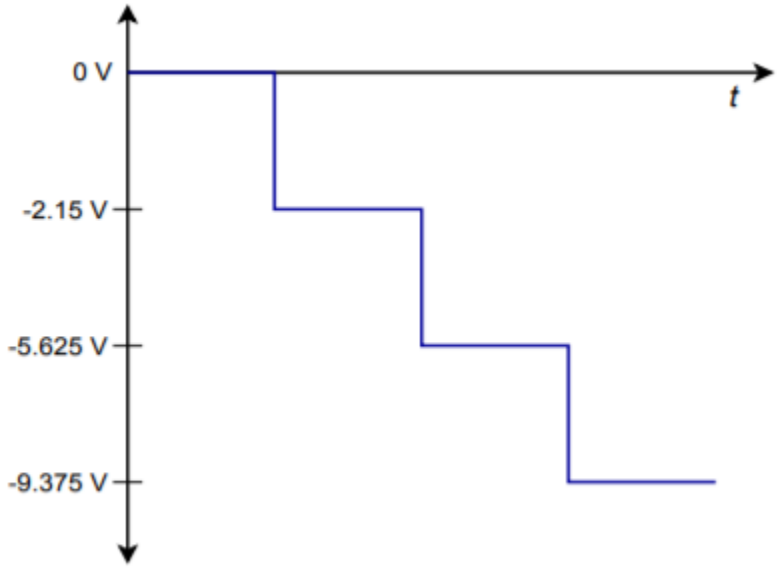


Figure 12.4.2 : Output with four digital words.

If this sequence is repeated over and over, the waveform of Figure 12.4.3 is the result. Note that a "stair-step" type wave is created. You might also think of this as a very rough form of a ramp function. A better ramp would be produced if we used all of the available values for the input sequence, as in 0000,0001,0010,0011,…,1111 . In order to remove the negative DC offset and make the signal bipolar, all we need to do is pass the signal through a coupling capacitor. The frequency of this waveform is controlled by the rate at which the words are fed to the converter. Note that by increasing the resolution and the number of words fed to the converter per cycle, a very close approximation to the ideal ramp function may be achieved. For that matter, by changing the input words to other sequences, we can create a wide variety of output wave shapes. This is the concept behind the digital arbitrary function generator. An arbitrary function generator allows you to create wave shapes beyond the simple sine/square/triangle found on the typical laboratory function generator. We'll take a closer look at this particular piece of test equipment a little later.

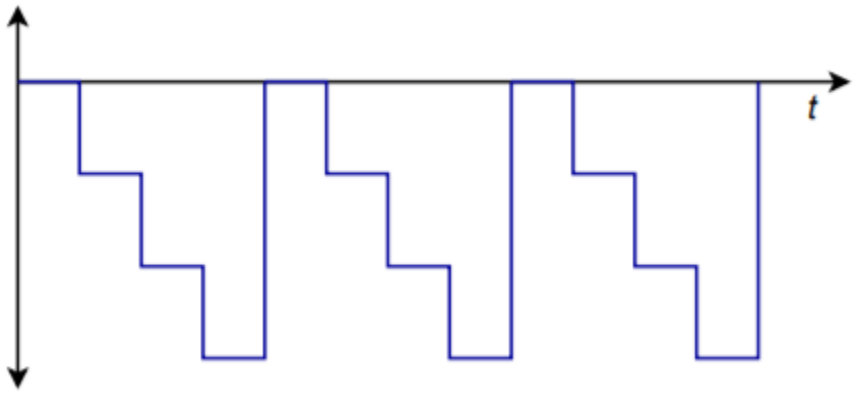


Figure 12.4.3 : Cycled output.

In order to increase resolution, it appears that all you need to do to the summing amplifier is add extra channels with larger and larger resistors. Unfortunately, the resistor sizes soon become impractical and another approach is required. For example, a 16-bit system would require that the LSB resistor be equal to 65,536 ◈◈ . One problem is that the resulting small input current may be dwarfed by input bias and offset currents. Also, high component accuracy is needed for the more significant inputs in terms of the input resistors and the drive signals. The excessively large resistors may also contribute added noise. The standard solution to this problem involves of the use of an ◈/2◈ resistive divider network.

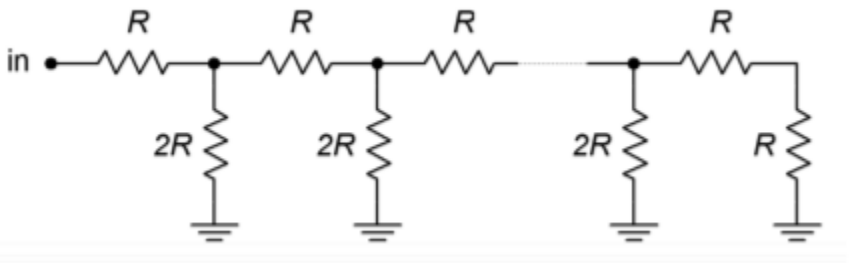


Figure 12.4.4 : ◈/2◈ ladder network

An ◈/2◈ network is shown in Figure 12.4.4 . This circuit exhibits the unique attribute of constant division by 2 for each stage. You may think of this as either a division of voltage at each successive node or a division of current in each successive leg. An example of a four stage (i.e., four-bit) network is shown in Figure 12.4.5 .

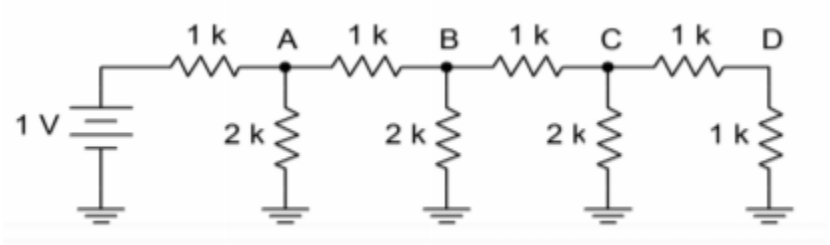


Figure 12.4.5 : A 4-stage ladder.

In order to find the voltage at any given node, the loading effects of the following stages must be taken into account. This is much easier to do than it first appears. If we need to find the voltage at point ⟐ , we must first find the resistance in parallel with the initial 2 ⟐Ω resistor. A quick inspection shows that each stage is loaded by the following stages, so it is easiest if we start at the last stage and work toward the input. The effective resistance to the right of node ⟐ is 1 ⟐Ω in series with 1 ⟐Ω , or 2 ⟐Ω . This resistance is placed in parallel with the 2 ⟐Ω resistor seen from node ⟐ to ground. The result is 1 ⟐Ω . In other words, from ⟐ to ground, we see 1 ⟐Ω . This creates a 2:1 voltage divider with the 1 ⟐Ω resistor placed from ⟐ to ⟐ , so the voltage at ⟐ must be half the voltage at ⟐ . This also points up the fact that the current entering node ⟐ splits into two equal portions: one that travels towards point ⟐ , and the other that travels through the 2 ⟐Ω resistor to ground. This is shown graphically in Figure 12.4.6 .

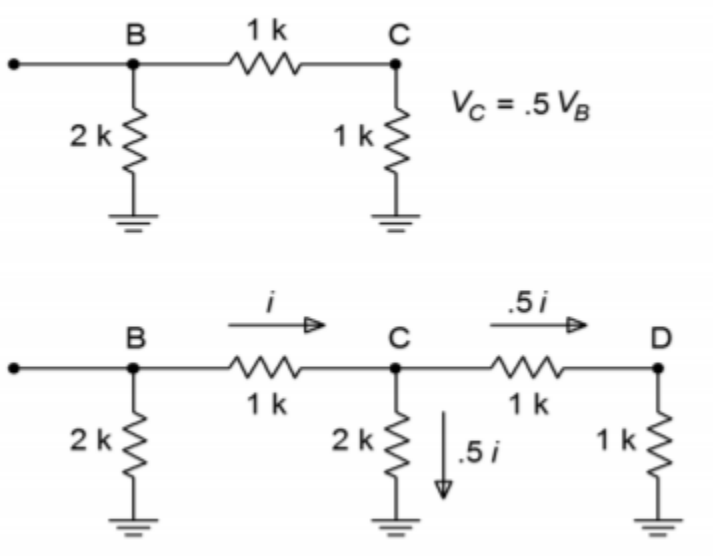


Figure 12.4.6 : Ladder analysis. a. Equivalent circuit (top). b. Current division (bottom).

If you look at the equivalent circuit section of Figure 12.4.6 , you will notice that this portion now looks exactly like the final portion of the original network. That is, every time a section is simplified and analyzed, the result will be a halving of voltage and current. It is already apparent that the voltage at ⟐ must be half the voltage at ⟐ , which in turn, must be half the voltage at ⟐ . As you can now prove, it follows that the voltage at ⟐ must be half the voltage at ⟐ . In a similar fashion, the current passing through each 2⟐ leg is half the preceding current. (For current division, the final section is not used to derive a current since it will be equal to the value in the preceding stage.) The halving of current is just what is needed for the binary representation of the digital input word.

Adapting the ⟐/2⟐ network to the DA converter is relatively easy. The network is fed from a stable current source, with each 2⟐ element feeding into a summing amplifier. In series with each 2⟐ element is a solid-state switch, which sets the appropriate logic level. This is shown in Figure 12.4.7 , with the network effectively on its side. When a logical high is presented to a given bit, the switch is closed and current flows through the 2⟐ element and into the op amp. Note that the right end of the resistor is effectively at ground, as the summing node of the op amp is a virtual ground. If a logical low is presented, the switch shunts the current to ground, bypassing the op amp. In this way, the appropriately weighted currents are summed and used to produce the output voltage.

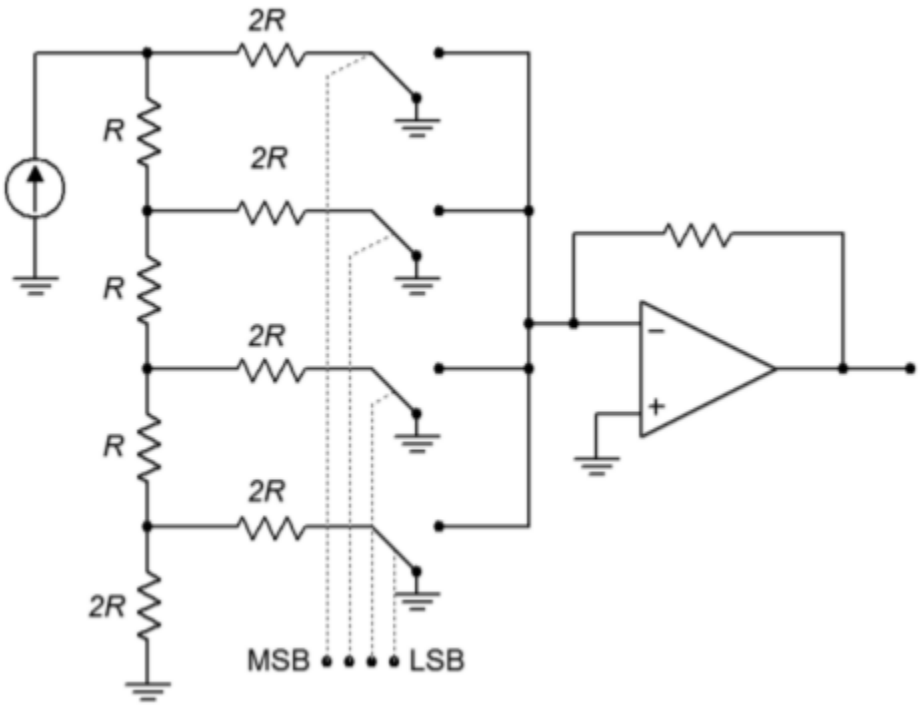


Figure 12.4.7 : Converter with R/2R ladder.

This technique offers several advantages over the simpler weighted gain version. First, all branches are fed by one common current source. Because of this, there is no need for output level matching. Second, only two different values of resistors are required for any number of bits used, rather than the impractically wide range seen earlier. It is more economical to control the tolerance of just two different parts than 12 or 16. Note that small input currents are still generated for the least significant bits, so attention to input bias and offset currents remains important.

PRACTICAL DIGITAL-TO-ANALOG CONVERTER LIMITS

Perhaps the most obvious limit associated with the DA converter is its speed. The op amp used in the DAC must be much faster than the final signals it is meant to produce. A given output waveform may contain several dozen individual sample points per cycle. The op amp must respond to each sample point. Consequently, wide bandwidth and high slew rates are required.

Integrated DAC spec sheets offer a few important parameters of which you should be aware. First of all, there is conversion speed. This figure tells how long it takes the DAC to turn the digital input word into a stable analog output voltage. This sets the maximum data rate. Next come the accuracy and resolution. Resolution indicates the number of discrete steps that may be produced at the output, and is set by the number of bits available. This is not the same as accuracy. Accuracy is actually comprised of several different factors including offset error, error, and nonlinearity. Offset error is normally measured by applying the all-zero input word and then measuring the output signal. Ideally, this signal will be zero volts. The deviation from zero is taken as the offset error. This has the effect of making all output levels inaccurate by a constant voltage. Offset error is relatively easy to compensate for in many applications by applying an equal offset of opposite polarity. Gain error is a deviation that affects each output level by a constant percentage. It is as if the signal were passed through a small

amplifier or attenuator. This error may be compensated for by using an amplifier with a gain equal to the reciprocal of the error. The two gains will effectively cancel. The effect of offset and gain error are shown in Figures 12.4.8 and 12.4.9 .

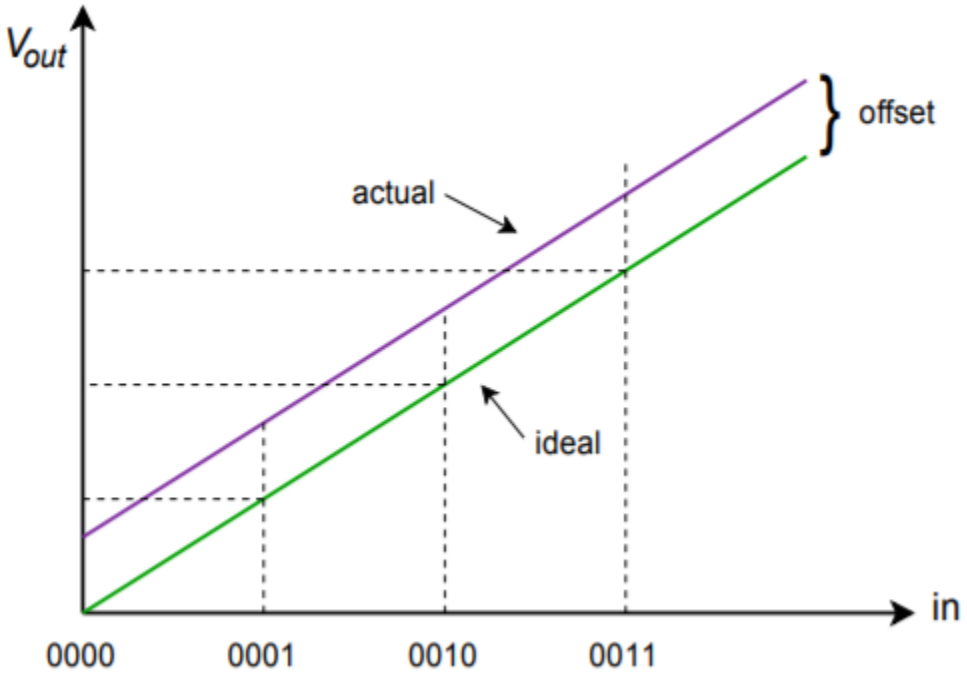


Figure 12.4.8 : Offset error only.

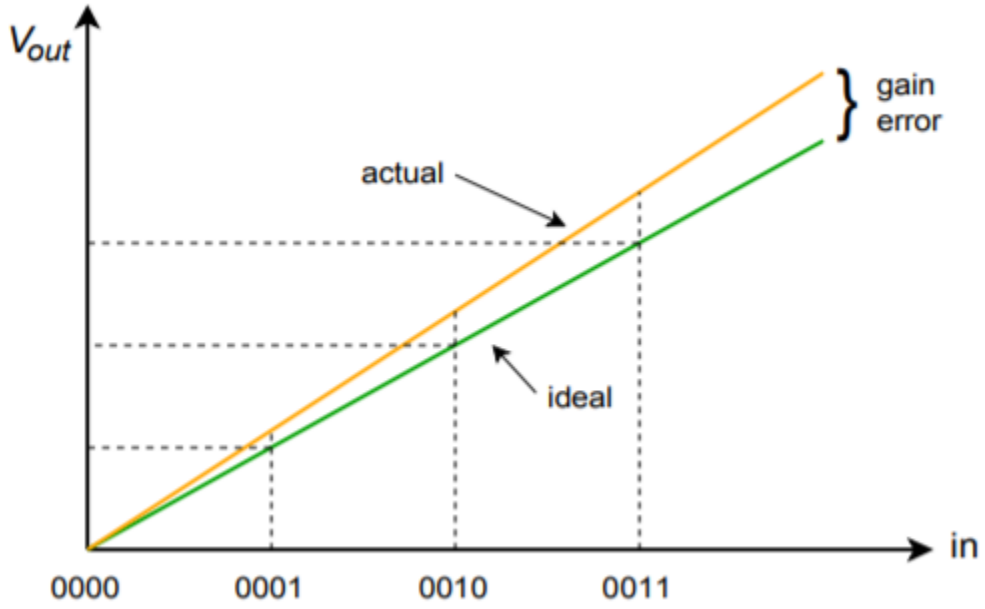


Figure 12.4.9 : Gain error only.

Nonlinearity errors may be broken into two forms: integral nonlinearity and differential nonlinearity. Integral nonlinearity details the maximum offset between the ideal outputs and the actual outputs for

all possible inputs. Differential nonlinearity details the maximum output deviation relative to one LSB caused by two adjacent input words. If differential nonlinearity is beyond ±1 LSB, the system may be non-monotonic. In other words, a higher digital input word may actually produce a lower analog output voltage. These two forms of error are shown in Figure 12.4.10 . Note that it is possible to have high integral nonlinearity and yet still have modest differential nonlinearity. This is the case in Figure 12.4.10 .

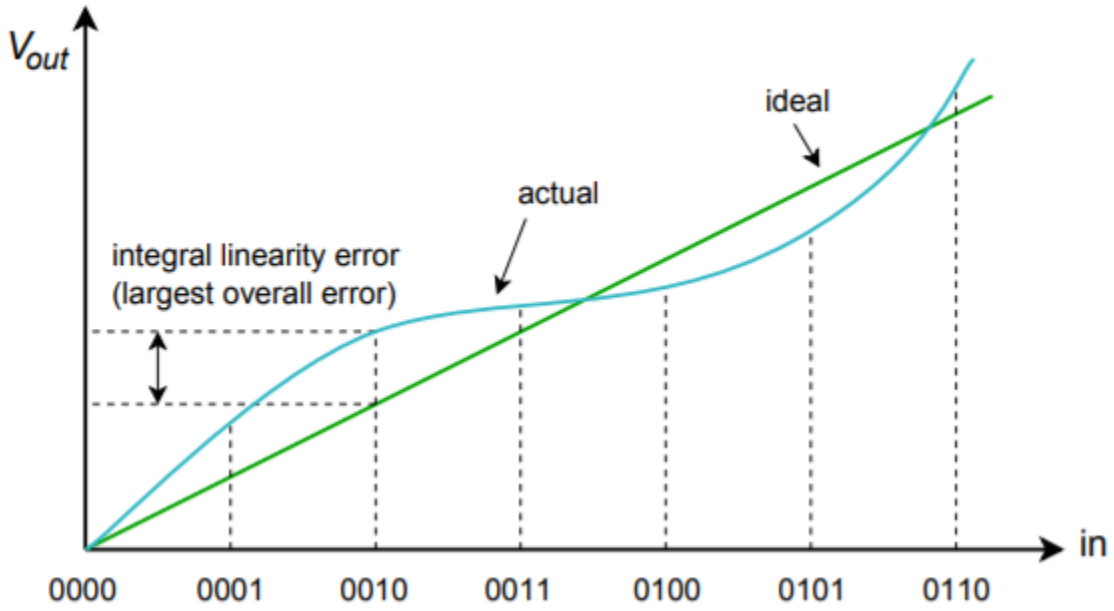


Figure 12.4.10 : Linearity error Integral linearity error.

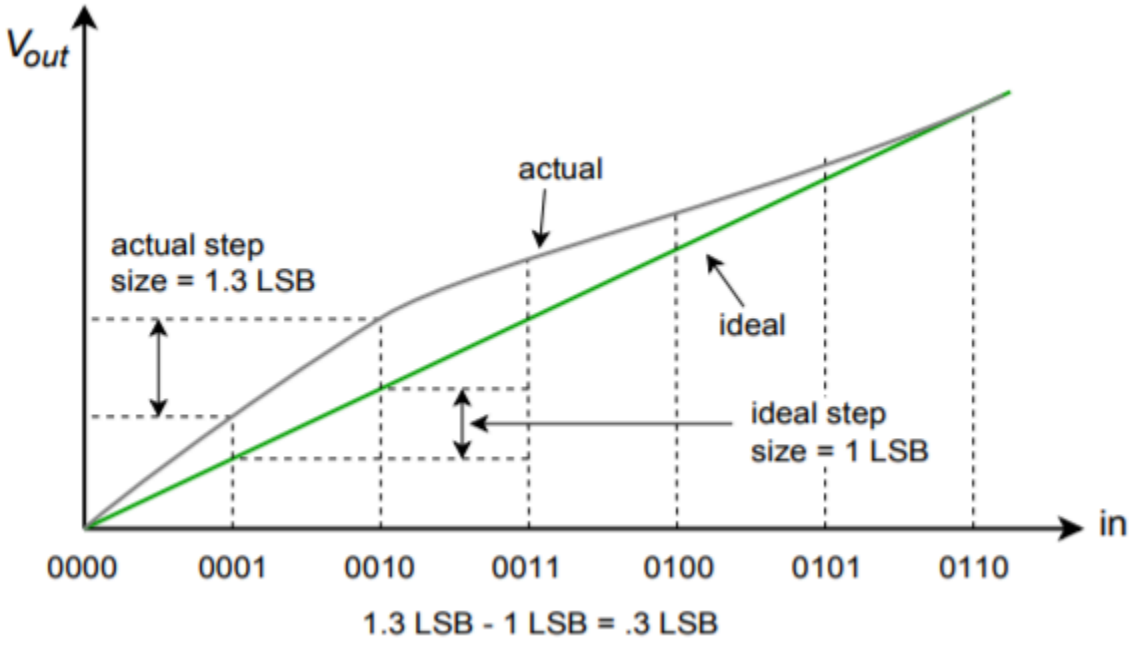


Figure 12.4.10 : Linearity error (continued) Differential linearity error (relative-adjacent error).

As you can see, accuracy is dependent on rather complex factors. In an effort to boil this down to a single number, some manufactures give an effective number of bits specification. For example, a 16-bit DAC may be specified as having 14-bit accuracy. This means that the 14 most significant bits behave in the idealized fashion, but the lowest 2 bits may be swamped out by linearity errors. Another spec that you will sometimes see is no missing codes. This means that for every increase in the input word, there will be an appropriate positive output level change.

In practice, the standard DA converter is used with an output filter. As you can see from the previous figures, the waveforms produced by the DAC contain a stair-step side effect. Generally, this is not desirable. The abrupt changes in output level indicate that higher frequency components are present. All components above the Nyquist rate should be filtered out with an appropriate low-pass filter. This filter is sometimes referred to as a reconstruction or smoothing filter. In an improperly designed system, the reconstruction filter will remove some of the highest in-band frequency components (i.e., components immediately below the Nyquist frequency). To compensate for this, logic levels are often latched to the DAC for shortened periods, thus creating a more spiked appearance, rather than the stair-step form. This effect is shown in Figure 12.4.11 . Although this spiked waveform appears to be less desirable than the stair-step form, it creates higher levels for the uppermost components, and after filtering, the result is a smoother overall frequency response.

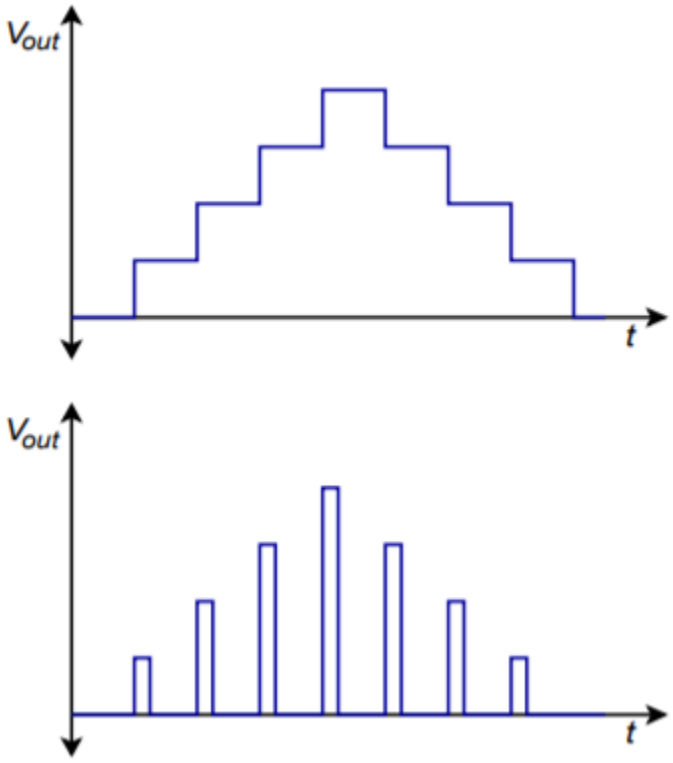


Figure 12.4.11 : Output reconstruction. a. Full period latch (top). b. Partial period latch (bottom).

To further increase the quality of the output waveform, a technique known as oversampling is sometimes employed. The basic idea is to create new sample points in between the existing ones. The result is a much denser data rate, which hopefully, will yield more exacting results after filtering. Also,

the higher data rate may loosen the requirements of the reconstruction filter. A typical system might use four times oversampling, meaning that the output data rate is four times the original. Therefore, for each input word, three new words have to be added. This effect is shown in Figure 12.4.12 .

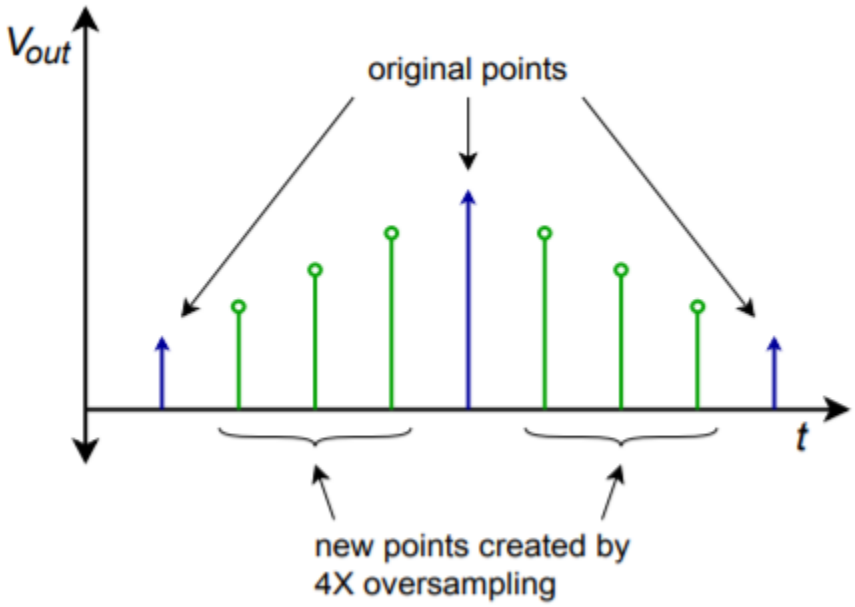


Figure 12.4.12 : Oversampled output.

There are a number of ways to create the new sample points. The most obvious way is via simple interpolation, but this does not achieve the best results. Another technique involves initializing the new values to zero and then passing the data stream through a digital low-pass filter, which effectively calculates the proper values. An extension of the oversampling principle is the delta-sigma technique. In delta-sigma, very high rates of oversampling are used in conjunction with specialized digital filter algorithms. The algorithms essentially trade the higher data rate for a slower rate with increased resolution. The design and analysis of delta-sigma systems is fairly advanced and is beyond the scope of this text. Suffice to say that these techniques can increase the quality of the output signal and are widely used in applications such as high-quality audio CD and DVD players.

DIGITAL-TO-ANALOG CONVERTER INTEGRATED CIRCUITS

There are many possible applications for digital-to-analog converters, and a number of different chips have evolved to meet specific needs. Generally, you can group these into specific classes, such as high speed, high resolution, or low cost. We will examine three representative types. The devices we will look at are the DAC0832; a basic 8-bit unit, the DAC7545; a microprocessor-compatible 12-bit unit, and the PCM1716; a 24-bit high-quality converter used in the audio industry.

DAC0832

This IC is a popular microprocessor-compatible 8-bit converter. The DAC0830 and DAC0831 are similar, but with somewhat reduced performance. It is a multiplying DAC. In other words, the output signal is a function of the digital input word and a reference input. In some applications, the reference input is not fixed, but rather, is a variable input signal. A feature list and pin-out are shown in Figure

12.4.13 . Notable items are a settling time of only 1 ◈ s, low power requirements, and high linearity. The DAC0832 may be used in either stand-alone mode or with a microprocessor. The switching waveforms are shown in Figure 12.4.14 .

National Semiconductor

DAC0830/DAC0832
8-Bit µP Compatible, Double-Buffered D to A Converters

General Description

The DAC0830 is an advanced CMOS/Si-Cr 8-bit multiplying DAC designed to interface directly with the 8080, 8048, 8085, Z80®, and other popular microprocessors. A deposited silicon-chromium R-2R resistor ladder network divides the reference current and provides the circuit with excellent temperature tracking characteristics (0.05% of Full Scale Range maximum linearity error over temperature). The circuit uses CMOS current switches and control logic to achieve low power consumption and low output leakage current errors. Special circuitry provides TTL logic input voltage level compatibility.

Double buffering allows these DACs to output a voltage corresponding to one digital word while holding the next digital word. This permits the simultaneous updating of any number of DACs.

The DAC0830 series are the 8-bit members of a family of microprocessor-compatible DACs (MICRO-DAC™).

Features

- Double-buffered, single-buffered or flow-through digital data inputs
- Easy interchange and pin-compatible with 12-bit DAC1230 series
- Direct interface to all popular microprocessors
- Linearity specified with zero and full scale adjust only — NOT BEST STRAIGHT LINE FIT.
- Works with ±10V reference-full 4-quadrant multiplication
- Can be used in the voltage switching mode
- Logic inputs which meet TTL voltage level specs (1.4V logic threshold)
- Operates "STAND ALONE" (without µP) if desired
- Available in 20-pin small-outline or molded chip carrier package

Key Specifications

- Current settling time: 1 µs
- Resolution: 8 bits
- Linearity: 8, 9, or 10 bits (guaranteed over temp.)
- Gain Tempco: 0.0002% FS/°C
- Low power dissipation: 20 mW
- Single power supply: 5 to 15 V_{DC}

Typical Application

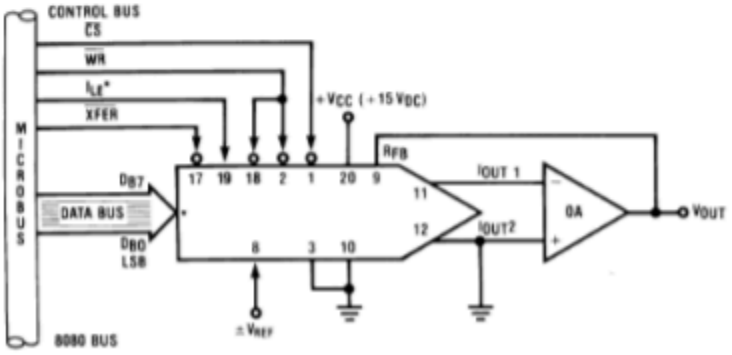
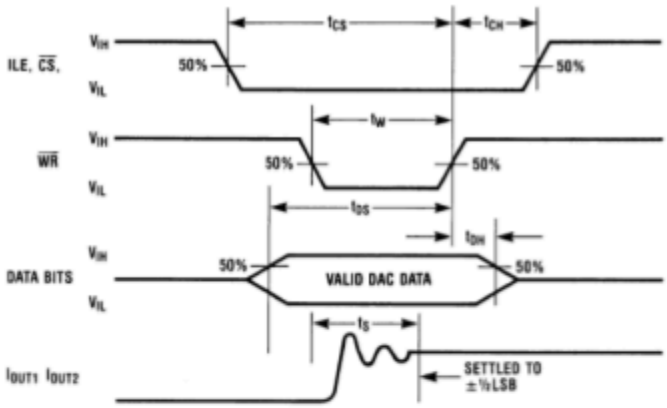


Figure 12.4.13 : DAC0832. Reprinted courtesy of Texas Instrument

Switching Waveform



Definition of Package Pinouts

Control Signals (All control signals level actuated)

$\overline{CS}$: **Chip Select** (active low). The $\overline{CS}$ in combination with ILE will enable $\overline{WR}_1$.

ILE: **Input Latch Enable** (active high). The ILE in combination with $\overline{CS}$ enables $\overline{WR}_1$.

$\overline{WR}_1$: **Write 1.** The active low $\overline{WR}_1$ is used to load the digital input data bits (DI) into the input latch. The data in the input latch is latched when $\overline{WR}_1$ is high. To update the input latch–$\overline{CS}$ and $\overline{WR}_1$ must be low while ILE is high.

$\overline{WR}_2$: **Write 2** (active low). This signal, in combination with $\overline{XFER}$, causes the 8-bit data which is available in the input latch to transfer to the DAC register.

$\overline{XFER}$: **Transfer control signal** (active low). The $\overline{XFER}$ will enable $\overline{WR}_2$.

Other Pin Functions

DI_0-DI_7: **Digital Inputs.** DI_0 is the least significant bit (LSB) and DI_7 is the most significant bit (MSB).

I_{OUT1}: **DAC Current Output 1.** I_{OUT1} is a maximum for a digital code of all 1's in the DAC register, and is zero for all 0's in DAC register.

I_{OUT2}: **DAC Current Output 2.** I_{OUT2} is a constant minus I_{OUT1}, or $I_{OUT1} + I_{OUT2}$ = constant (I full scale for a fixed reference voltage).

R_{fb}: **Feedback Resistor.** The feedback resistor is provided on the IC chip for use as the shunt feedback resistor for the external op amp which is used to provide an output voltage for the DAC. This on-chip resistor should always be used (not an external resistor) since it matches the resistors which are used in the on-chip R-2R ladder and tracks these resistors over temperature.

V_{REF}: **Reference Voltage Input.** This input connects an external precision voltage source to the internal R-2R ladder. V_{REF} can be selected over the range of +10 to −10V. This is also the analog voltage input for a 4-quadrant multiplying DAC application.

V_{CC}: **Digital Supply Voltage.** This is the power supply pin for the part. V_{CC} can be from +5 to +15V_{DC}. Operation is optimum for +15V_{DC}

GND: The pin 10 voltage must be at the same ground potential as I_{OUT1} and I_{OUT2} for current switching applications. Any difference of potential (V_{OS} pin 10) will result in a linearity change of

$$\frac{V_{OS} \text{ pin } 10}{3V_{REF}}$$

For example, if V_{REF} = 10V and pin 10 is 9mV offset from I_{OUT1} and I_{OUT2} the linearity change will be 0.03%.

Pin 3 can be offset ±100mV with no linearity change, but the logic input threshold will shift.

Figure 12.4.14 : DAC0832 switching waveforms Reprinted courtesy of Texas Instruments

An interesting application of the DAC0832 can be found in Figure 12.4.15 . Basically, this is a digitally-controlled state variable filter.

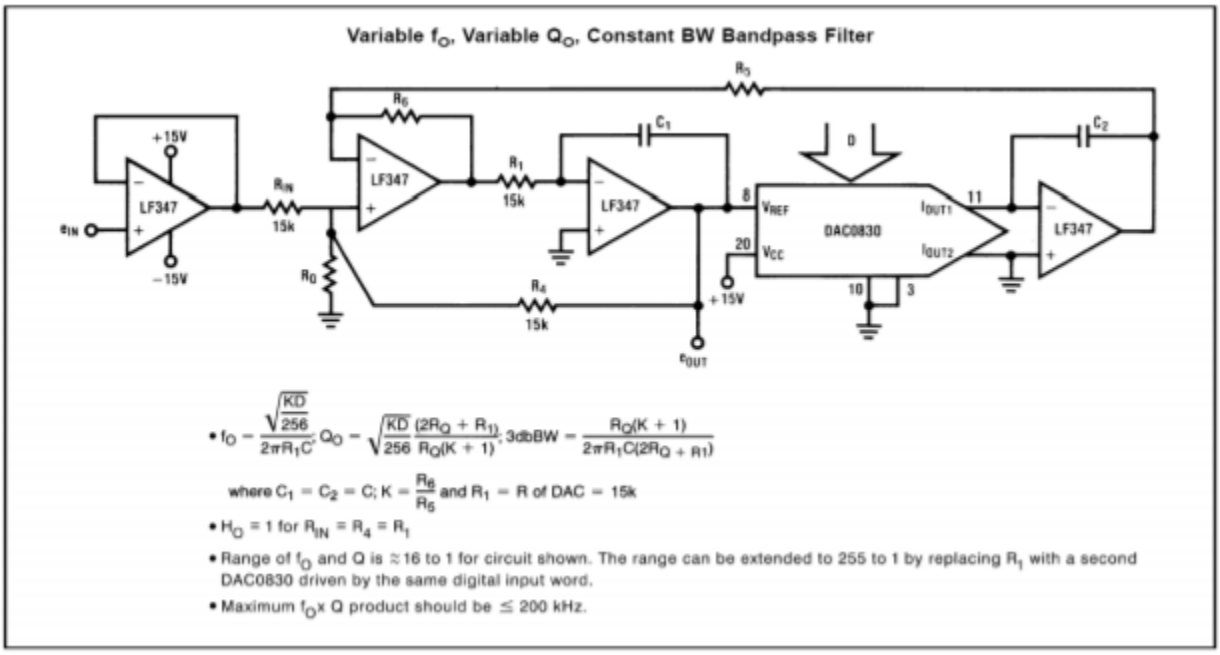


Figure 12.4.15 : DAC0832 state variable filter application. Reprinted courtesy of Texas Instruments

Note that the converter replaces the input resistor of the second integrator. Normally, that resistor would be used to convert the output voltage of the first integrator into an input current for the second integrator. This job is now handled by the DAC0832. The digital input word effectively sets the voltage-to-current conversion. Thus, a change in the input word alters the tuning frequency of the filter just as a potentiometer. Compare this circuit to the OTA-based voltage-controlled filter from Chapter Eleven. Conceptually, they are very similar.

12-Bit, Parallel Input, Multiplying Digital-to-Analog Converter

FEATURES

- 2.5V to 5.5V Supply Operation
- Fast Parallel Interface: 17ns Write Cycle
- Update Rate of 20.4MSPS
- 10MHz Multiplying Bandwidth
- ±10V Reference Input
- Low Glitch Energy: 5nV-s
- Extended Temperature Range: –40°C to +125°C
- 20-Lead TSSOP Packages
- 12-Bit Monotonic
- ±1LSB INL
- 4-Quadrant Multiplication
- Power-On Reset with Brownout Detection
- Readback Function
- Industry-Standard Pin Configuration

APPLICATIONS

- Portable Battery-Powered Instruments
- Waveform Generators
- Analog Processing
- Programmable Amplifiers and Attenuators
- Digitally-Controlled Calibration
- Programmable Filters and Oscillators
- Composite Video
- Ultrasound

DESCRIPTION

The DAC7821 is a CMOS 12-bit current output digital-to-analog converter (DAC). This device operates from a single 2.5V to 5.5V power supply, making it suitable for battery-powered and many other applications.

This DAC operates with a fast parallel interface. Data readback allows the user to read the contents of the DAC register via the DB pins. On power-up, the internal register and latches are filled with zeroes and the DAC outputs are at zero scale.

The DAC7821 offers excellent 4-quadrant multiplication characteristics, with a large signal multiplying bandwidth of 10MHz. The applied external reference input voltage (V_{REF}) determines the full-scale output current. An integrated feedback resistor (R_{FB}) provides temperature tracking and full-scale voltage output when combined with an external current-to-voltage precision amplifier.

The DAC7821 is available in a 20-lead TSSOP package.

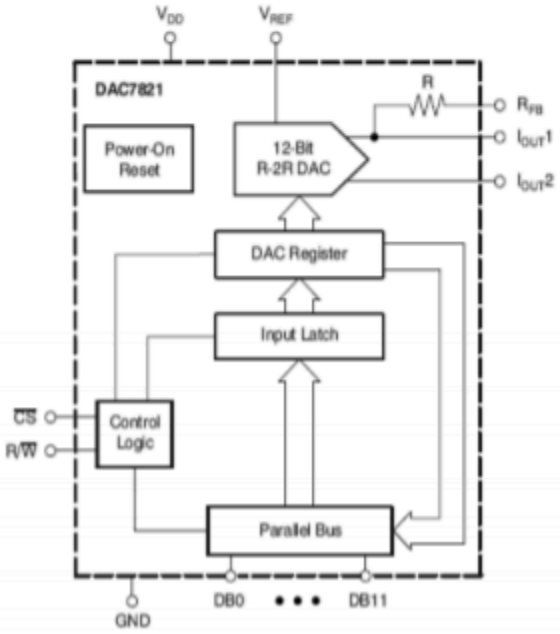


Figure 12.4.16 : DAC7821. Reprinted courtesy of Texas Instruments

DAC7821

The DAC7821 is a fairly standard 12-bit linear converter and is shown in Figure 12.4.16 . Its interesting aspects are that it is a multiplying converter and that it is microprocessor-compatible. The multiplying effect comes from the fact that a reference is used to drive the ◈/2◈ ladder network. If the reference is changed, the output is effectively rescaled. Consequently, you can think of the output signal as equal to the reference value times the digital input word. You may also think of this as a form of "digital volume control".

With the inclusion of a few extra logic lines, the IC has become microprocessor compatible. This means that the DAC7821 has chip select and read/write lines along with the 12 data input lines. This allows the converter to be connected directly to the microprocessor data bus. By using memory-mapped I/O, the microprocessor can write data to the converter just as it writes data to memory. A 16-bit microprocessor system can present the converter with all of the data it needs during one write cycle, however an 8-bit microprocessor will need two write cycles and some form of latch. One address may be used for the lower 8 bits, and another address for the remaining 4 bits. A simplified system is shown in Figure 12.4.17 using an 8-bit microprocessor.

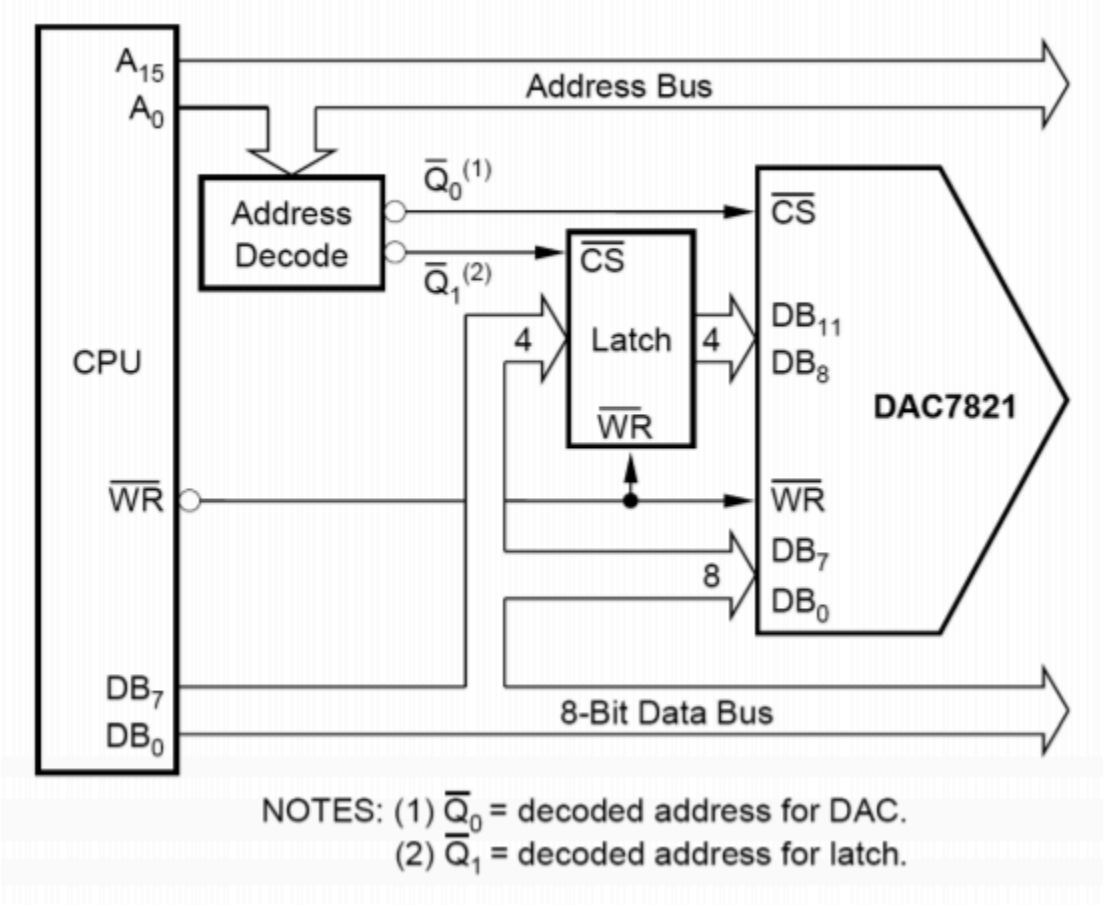
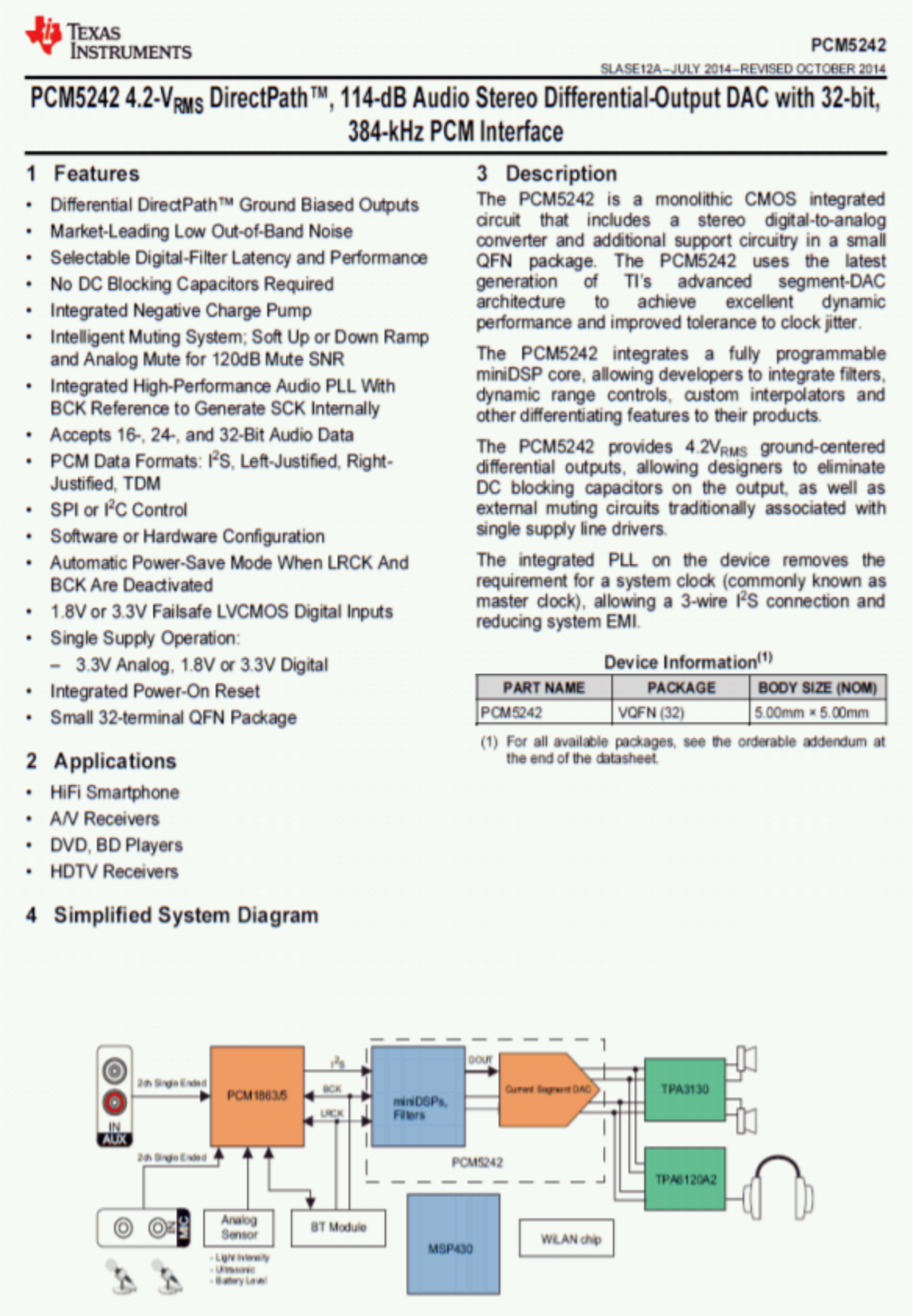


Figure 12.4.17 : Microprocessor to DAC7821 interface. Reprinted courtesy of Texas Instruments

PCM5242

The PCM5242 is a stereo 24-bit converter designed specifically for high-quality digital audio applications. It comes in a VQFN (Very thin Quad Flat No-lead) package. A block diagram and

feature list are shown in Figure 12.4.18◈ . Unlike the other converters, the PCM5242 features serial input of data, not parallel. It includes its own on-board serial conversion circuitry and logic. This technique helps to reduce system cost. It is also surprisingly convenient as many specialized digital signal processing ICs that might be used with the PCM5242 utilize a serial-type output. This may be fed directly into the PCM5242 in 16-, 24-, or 32-bit format.



TEXAS INSTRUMENTS

PCM5242

SLASE12A–JULY 2014–REVISED OCTOBER 2014

PCM5242 4.2-V_{RMS} DirectPath™, 114-dB Audio Stereo Differential-Output DAC with 32-bit, 384-kHz PCM Interface

1 Features

- Differential DirectPath™ Ground Biased Outputs
- Market-Leading Low Out-of-Band Noise
- Selectable Digital-Filter Latency and Performance
- No DC Blocking Capacitors Required
- Integrated Negative Charge Pump
- Intelligent Muting System; Soft Up or Down Ramp and Analog Mute for 120dB Mute SNR
- Integrated High-Performance Audio PLL With BCK Reference to Generate SCK Internally
- Accepts 16-, 24-, and 32-Bit Audio Data
- PCM Data Formats: I^2S, Left-Justified, Right-Justified, TDM
- SPI or I^2C Control
- Software or Hardware Configuration
- Automatic Power-Save Mode When LRCK And BCK Are Deactivated
- 1.8V or 3.3V Failsafe LVCMOS Digital Inputs
- Single Supply Operation:
  - 3.3V Analog, 1.8V or 3.3V Digital
- Integrated Power-On Reset
- Small 32-terminal QFN Package

2 Applications

- HiFi Smartphone
- A/V Receivers
- DVD, BD Players
- HDTV Receivers

3 Description

The PCM5242 is a monolithic CMOS integrated circuit that includes a stereo digital-to-analog converter and additional support circuitry in a small QFN package. The PCM5242 uses the latest generation of TI's advanced segment-DAC architecture to achieve excellent dynamic performance and improved tolerance to clock jitter.

The PCM5242 integrates a fully programmable miniDSP core, allowing developers to integrate filters, dynamic range controls, custom interpolators and other differentiating features to their products.

The PCM5242 provides 4.2V_{RMS} ground-centered differential outputs, allowing designers to eliminate DC blocking capacitors on the output, as well as external muting circuits traditionally associated with single supply line drivers.

The integrated PLL on the device removes the requirement for a system clock (commonly known as master clock), allowing a 3-wire I^2S connection and reducing system EMI.

Device Information(1)

PART NAME	PACKAGE	BODY SIZE (NOM)
PCM5242	VQFN (32)	5.00mm × 5.00mm

(1) For all available packages, see the orderable addendum at the end of the datasheet.

4 Simplified System Diagram

Figure 12.4.18 : PCM5242. Reprinted courtesy of Texas Instruments

The PCM5242 specification sheet is shown in Figure 12.4.18 . Note that this device is specified

with sampling rates from 48 kHz to 192 kHz. Total harmonic distortion plus noise is typically 94 dB below a full-scale output when used with any of these sample rates. Due to its high resolution and 114 dB dynamic range, extra care must be taken during circuit layout to avoid hum pickup and RF interference.

7.5 Electrical Characteristics

All specifications at T_A = 25°C, AV_{DD} = CPV_{DD} = DV_{DD} = 3.3V, f_S = 48kHz, system clock = 512f_S and 24-bit data unless otherwise noted.

PARAMETER		TEST CONDITIONS	MIN	TYP	MAX	UNIT
	Resolution		16	24	32	Bits
Digital Input/Output						
	Logic Family: 3.3V LVCMOS compatible					
V_{IH}	Input logic level		0.7×DV_{DD}			V
V_{IL}					0.3×DV_{DD}	
I_{IH}	Input logic current	V_{IN} = V_{DD}			10	µA
I_{IL}		V_{IN} = 0V			–10	
V_{OH}	Output logic level	I_{OH} = –4mA	0.8×DV_{DD}			V
V_{OL}		I_{OL} = 4mA			0.22×DV_{DD}	
	Logic Family 1.8V LVCMOS compatible					
V_{IH}	Input logic level		0.7×DV_{DD}			V
V_{IL}					0.3×DV_{DD}	
I_{IH}	Input logic current	V_{IN} = V_{DD}			10	µA
I_{IL}		V_{IN} = 0V			–10	
V_{OH}	Output logic level	I_{OH} = –2mA	0.8×DV_{DD}			V
V_{OL}		I_{OL} = 2mA			0.22×DV_{DD}	
Dynamic Performance (PCM Mode)(1)(2)						
	THD+N at -1 dB(2)	f_S = 48kHz		–94	–87	dB
		f_S = 96kHz		–94		
		f_S = 192kHz		–94		
	Dynamic range(2)	EIAJ, A-weighted, f_S = 48kHz	108	114		
		EIAJ, A-weighted, f_S = 96kHz		114		
		EIAJ, A-weighted, f_S = 192kHz		114		
	Signal-to-noise ratio(2)	EIAJ, A-weighted, f_S = 48kHz		114		
		EIAJ, A-weighted, f_S = 96kHz		114		
		EIAJ, A-weighted, f_S = 192kHz		114		
	Signal to noise ratio with analog mute(2)(3)	EIAJ, A-weighted, f_S = 48kHz	113	123		
		EIAJ, A-weighted, f_S = 96kHz	113	123		
		EIAJ, A-weighted, f_S = 192kHz	113	123		
	Channel Separation	f_S = 48kHz	100 / 96	109 / 103		dB
		f_S = 96kHz	100 / 96	109 / 103		
		f_S = 192kHz	100 / 96	109 / 103		

(1) Filter condition: THD+N: 20Hz HPF, 20kHz AES17 LPF Dynamic range: 20Hz HPF, 20kHz AES17 LPF, A-weighted Signal-to-noise ratio: 20Hz HPF, 20kHz AES17 LPF, A-weighted Channel separation: 20Hz HPF, 20kHz AES17 LPF Analog performance specifications are measured using the System Two Cascade™ audio measurement system by Audio Precision™ in the RMS mode.
(2) Output load is 10kΩ, with 470Ω output resistor and a 2.2nF shunt capacitor (see recommended output filter).
(3) Assert XSMT or both L-ch and R-ch PCM data are BPZ

Figure 12.4.18 : PCM5242 specifications. Reprinted courtesy of Texas Instruments

Example 12.4.1

Perhaps the first thing many people think of when they hear the terms digital or digitized, is the audio compact disk, or CD for short. Home CD players are excellent examples of the use of precision DA circuitry in our everyday lives. Music data is stored on the CD with 16-bit resolution and a sampling rate of 44.1 kHz. This produces a Nyquist frequency of 22.05 kHz, which is high enough to encompass the hearing range of most humans. Error correction and auxiliary data is also stored on the disk. The data is stored on disk in the form of very tiny pits, which are read by a laser. The signal is then converted into the common electronic logic form where it is checked for error and adjusted as need be. The data stream is then fed to the DA converter for audio reconstruction. A single converter may be multiplexed between the two stereo channels, or two dedicated converters may be used. Oversampling in the range of 2X to 8X is often used for improved signal quality. A block diagram of the system is shown in Figure 12.4.19 . The actual DAC portion seems almost trivial when compared to some of the more sophisticated elements.

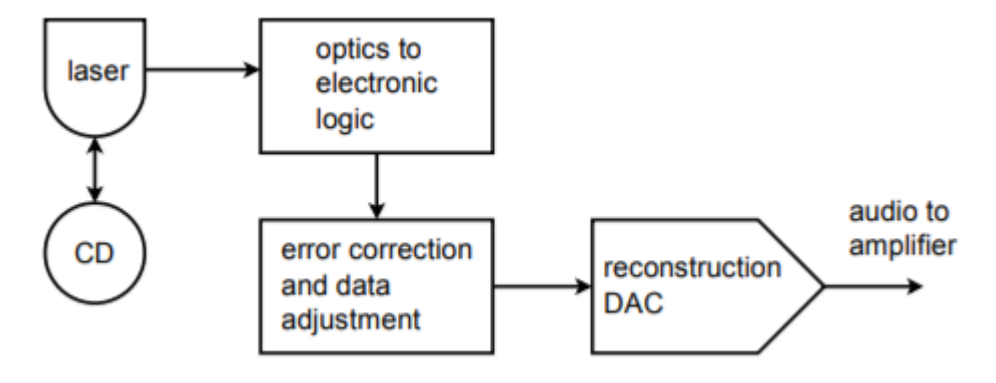


Figure 12.4.19 : Audio compact disk playback system.

The storage density of the optical CD is quite remarkable. This small disk (less than 5 inches in diameter) can hold 70 minutes of music. Ignoring the auxiliary data, we can quickly calculate the total storage. We have two channels of 16-bit data, or 32 bits (4 bytes) per sample point. There are 44,100 samples per second for 70 minutes, yielding 185.22 megasamples. The total data storage is 5.927 gigabits, or 741 megabytes.

Example 12.4.2

As we have already mentioned, it is possible to connect DACs directly to microprocessor systems. Furthermore, the microprocessor may write to the DAC with no more effort than writing to a memory location. The microprocessor can write any series of data words we desire out to the DAC and can repeat a sequence virtually forever. Given this ability, we can make an arbitrary waveform generator. Instead of being locked into a set of predefined wave shapes as on ordinary function generators, this system allows for all manner of wave shapes. The accuracy and flexibility of the system will depend on its speed and the available DAC resolution.

The basic idea is one of table look-up. For example, let's say that we have a 16-bit system. We will create a

table of data values for one cycle of the output waveform. For convenience sake, we might make the table size a handy power of 2, such as 256. In other words, a single output cycle will be chopped into 256 discrete time chunks. It is obvious, then, that the converter must be a few hundred times faster than the highest fundamental that we wish to produce. By increasing or decreasing the output data rate, we can change the frequency of the output fundamental. This is known as a variable sample rate technique. It is also possible to change the fundamental frequency with a fixed rate technique (this is somewhat more complex, but does offer certain advantages). An output flow chart is shown in Figure 12.4.20 .

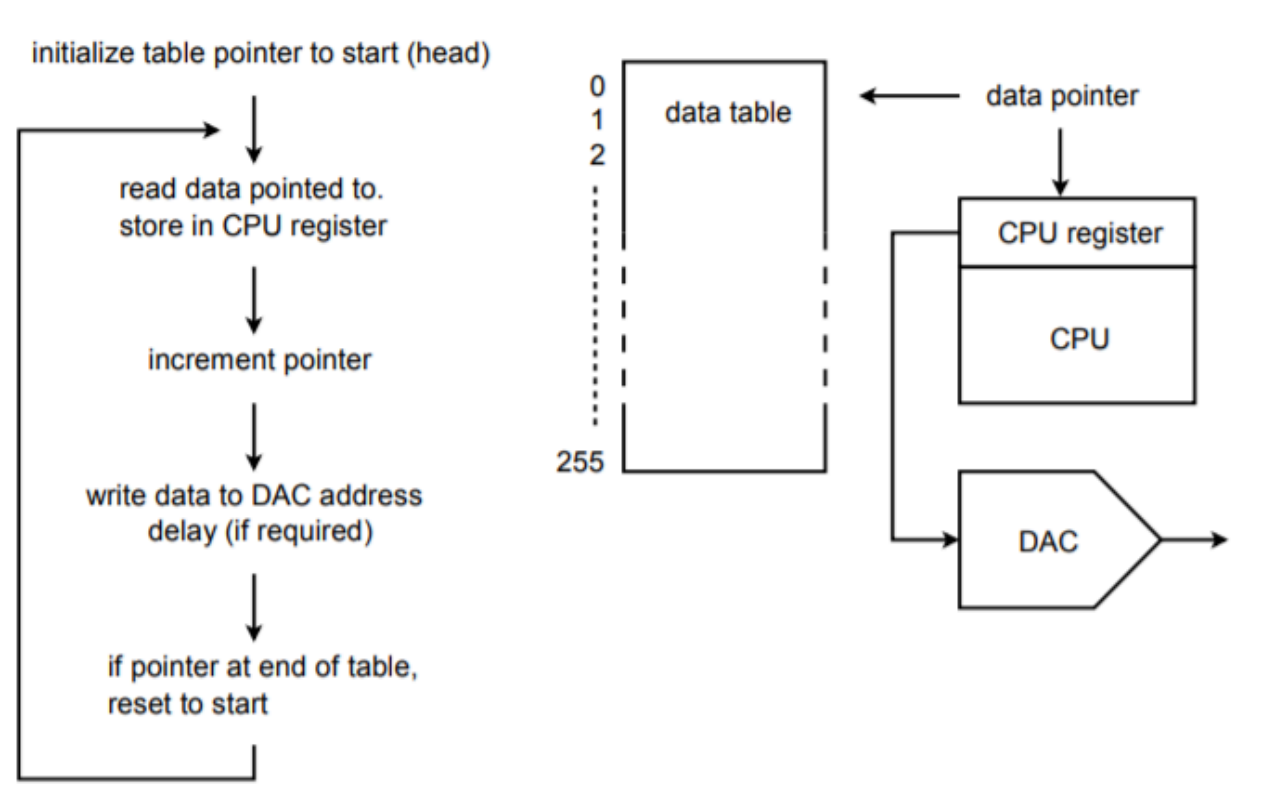


Figure 12.4.20 : Arbitrary waveform generator.

Upon initialization, an address pointer is set to the starting address of the data table. The CPU reads the data from the table via the pointer. The pointer is incremented so that it now points to the next element in the table. (Some CPUs offer a post-increment addressing mode so that both steps may be performed in a single instruction.) Next, the CPU writes the data to the special DAC address. At this point, some form of software/hardware delay is invoked that sets the output data rate. After the delay, the CPU reads the next data element via the pointer and continues as in the first run. Once the 256th element is sent out, the pointer is reset to the start of the table and the process continues on. In this way, the table can be thought of as circular, or never-ending. If the system software is written in a higher level language, the pointer/data table may be implemented as a simple array where the array index is set by a counter. This will not be as efficient as a direct assembly level approach, though.

The real beauty of this system is that the data table may contain virtually any sequence of data. The data could represent a sine, pulse, triangle, or other standard function. More importantly, the data could represent a sine wave with an embedded noise transient, or a signal containing a hum component just as easily. This data could come from three basic sources. First, the data table can be filled through direct computation if the time-domain equation of the desired function is known. Second, the data could be manufactured by the user through some form of interaction with a computer, perhaps with a mouse or

drawing pad. Finally, the data may be derived from a real-world signal. That is, an analog-to-digital converter may be used to record the signal in digital form. The data may then be loaded into the table and played back repeatedly. The arbitrary waveform generator allows its user to test circuits and systems with a range of wave shapes that would be impossible or impractical to generate otherwise.

Example 12.4.3

Under computer control, DA converters can be used as part of an automated test equipment system. In order to fully characterize an electronic product, a number of individual tests need to be run. Setting up each individual test can be somewhat time consuming and is subject to operator error. Automating this procedure may improve repeatability and decrease testing time. There are many ways in which this process may be automated. We'll look at one approach.

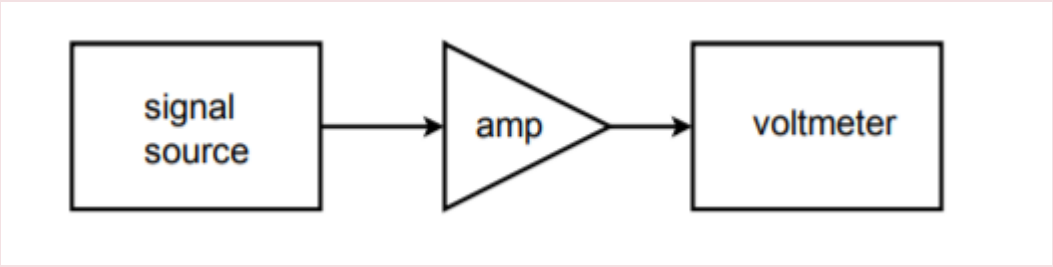


Figure 12.4.21 : Simple test setup.

Let's assume that we would like to make frequency response measurements for an amplifier at 20 different frequencies. A circuit test system appropriate for this job is shown in Figure 12.4.21 . To perform this test manually requires 20 distinct settings of the source signal, and 20 corresponding output readings. This can prove to be rather tedious if many units are to be tested. It would be very handy if there was some way in which the source frequency could be automatically changed to preset values. This is not particularly difficult. Most modern sources have control voltage inputs that may be used to set the frequency. The required control voltage can be created and accurately set through the use of a computer and DA converter. The computer can be programmed to send specific digital words to the DAC, which in turn feeds the signal source control input. In other words, the data word directly sets the frequency of the signal source. The computer can be programmed to send virtually any sequence of data words at almost any rate and do it all without operator intervention. All the operator needs to do is start the process. Test repeatability is very high with a system like this. A block diagram of this system is shown in Figure 12.4.22 .

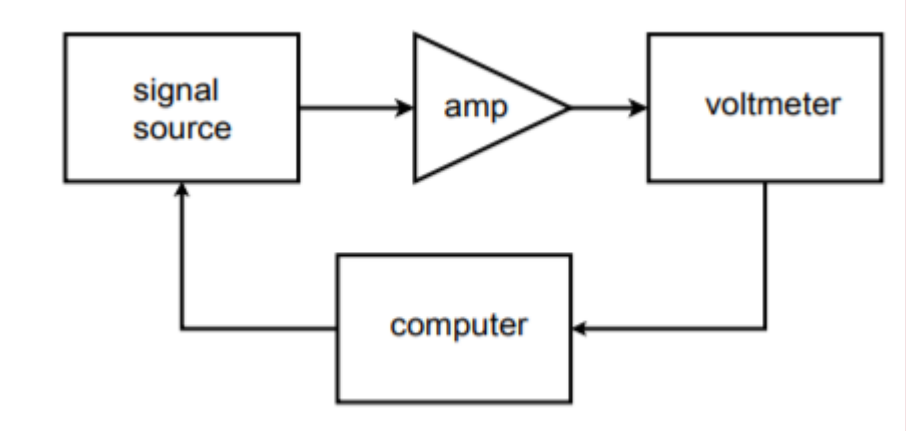


Figure 12.4.22 : Automated test setup.

In order to record the data, the voltmeter may be connected to a strip chart recorder, or better yet, back to the computer. The data may be sent to the computer in digital form if the voltmeter is of fairly advanced design, or, with the inclusion of an analog-to-digital converter, the output signal may be directly sampled and manipulated by the computer. In either case, data files may be created for each unit tested and stored for later use. Also, convenient statistical analysis may be performed quickly at the end of a test batch. Note that because the DA converter only generates a control signal, very high resolution and low distortion is normally not required. If a high resolution converter is used, it is possible to create the test signals in the computer (as in the arbitrary function generator), and dispense with the signal source.

The automated test system is only one possible application of instrument control. Another interesting example is in the generation of laser "light shows". A block diagram of a simplified system is shown in Figure 12.4.23 . In order to create the complex patterns seen by the audience, a laser beam is bounced off of tiny moving mirrors. The mirrors may be mounted on something as simple as a galvanometer. The galvanometer is fed by a DAC. The pattern that the laser beam makes is dependent on how the galvanometer moves the mirror, which is in turn controlled by the data words fed to the DAC. In practice, several mirrors may be used to deflect the beam along three axes.

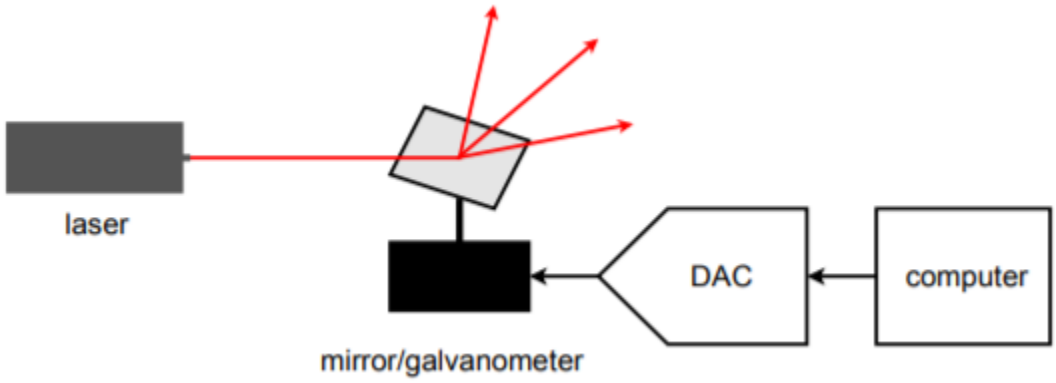


Figure 12.4.23 : Computer-controlled lighting.

DA converters can be used to adjust any device with a control-voltage type input. Also, they may be

used to control electromechanical devices that respond to an applied voltage. Their real advantage is the repeatability and flexibility they offer.

16.5 ANALOG-TO-DIGITAL CONVERSION

Now that you know the basics behind digital-to-analog conversion, we may examine the converse system, analog-to-digital conversion. The analog-to-digital conversion process is sometimes referred to as quantization, implying the individual discrete steps that the output assumes. There are several techniques to produce the conversion. Some techniques are optimized for fastest possible conversion speed, and some for highest accuracy. We will investigate the more popular types.

The concept of AD conversion is simple enough: you wish to measure the voltage of the incoming waveform at specific instances in time. This measurement will be translated into a digital word. One practical limitation is the fact that the conversion circuitry may require a small amount of computation or translation time. For ultimate accuracy, then, it is important that the measured waveform not change during the conversion interval. To do this, specialized subcircuits called track-and-hold (or sample-and-hold) amplifiers are used. They are usually abbreviated as T/H or S/H. Their job is to capture an input potential and produce a steady output to feed to the AD converter. The T/H schematic symbol is shown in Figure 12.5.1 . When the T/H logic is in track mode, the circuit acts as a simple buffer, so its output voltage equals its input. When the logic goes to the hold state, the output voltage locks at its present potential and stays there until the circuit is switched back to track mode. A representation of this operation is shown in Figure 12.5.2 .

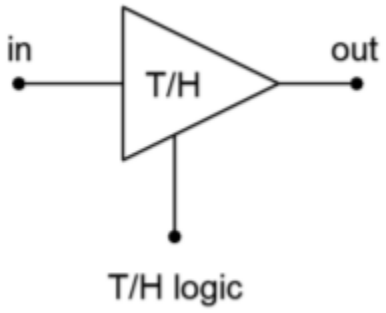


Figure 12.5.1 : Symbol for the track-and-hold amplifier.

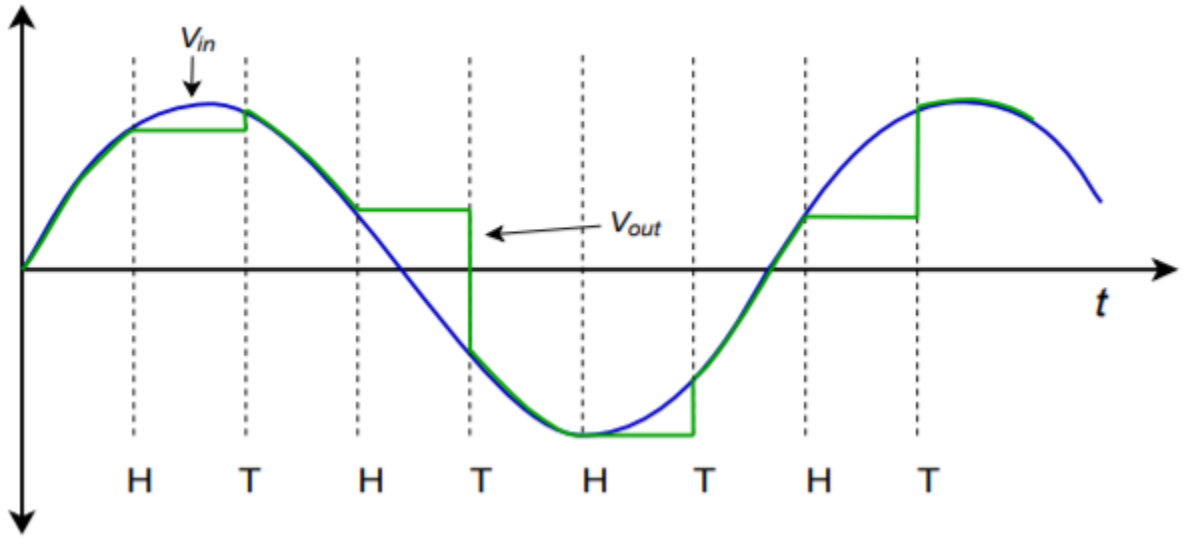


Figure 12.5.2 : Track-and-hold operation.

In reality, the track-and-hold process is not perfect, and errors may arise. These errors are shown magnified in Figure 12.5.3 . First of all, there is a small delay between the time the logic signal changes and the T/H starts to react. The exact amount of time is variable and is responsible for aperture error. This is equal to the voltage difference between the signals at the desired and actual times. Also, due to the dynamic nature of the process of switching from track to hold, some initial ringing may occur in the hold waveform.

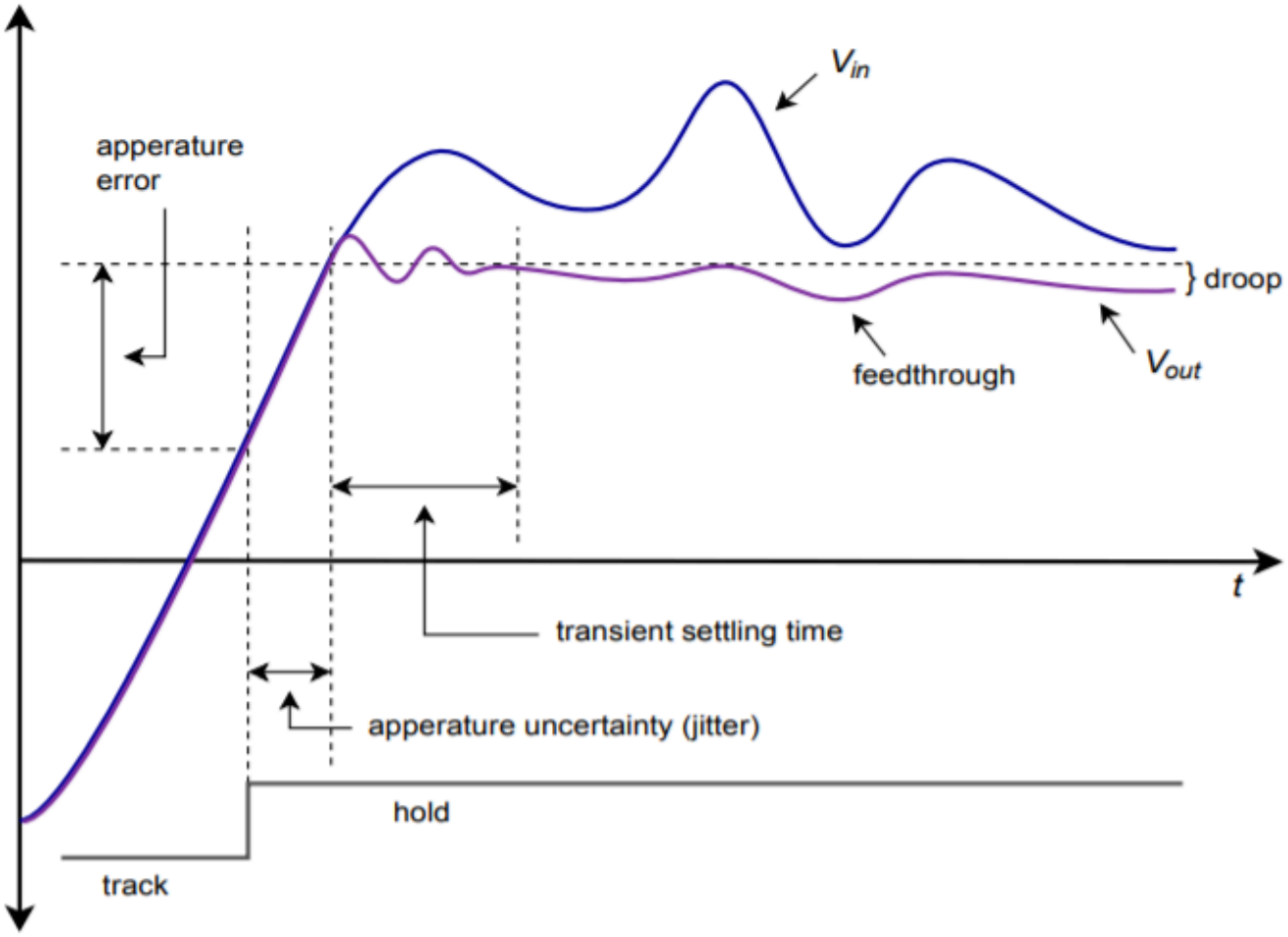


Figure 12.5.3 : Track-and-hold errors.

You can see that as time progresses, the hold waveform tends to decay toward zero. This is because the held voltage is usually formed across a capacitor. Although the discharge time constant can be very long, it cannot be infinite, and thus, the charge eventually bleeds off. This parameter is measured by the droop rate (fundamental units of volts per second). Finally, the possibility of feedthrough error exists. If a large change occurs on the input waveform during the hold period, it is possible that a portion of the signal may "leak" through to the T/H output. These errors are particularly troublesome when working with high-resolution converters. Lower-resolution systems may not be adversely affected by these relatively small aberrations.

In order to create a T/H, a pair of high-impedance buffers are normally used, along with some form of switch element and a capacitor to hold the charge. Examples are shown in Figure 12.5.4 . Figure 12.5.4◈ shows the general voltage type, open-loop T/H. The buffers exhibit high input impedance. When the switch is closed, op amp 1 directly feeds op amp 2, and therefore the output voltage equals the input voltage. Normally, the hold capacitor is relatively small and does not adversely affect the drive capability of op amp 1.

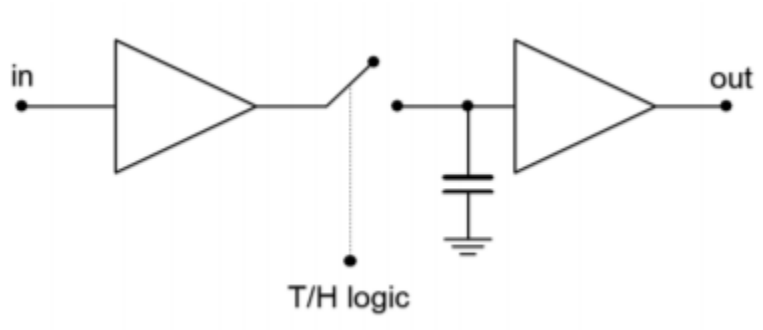


Figure 12.5.4 : Track-and-hold circuits. General open-loop voltage mode track-and-hold.

A more detailed version of this circuit is shown in Figure 12.5.4 . Each op amp is a FET input type for minimum input current draw. The switch is a simple JFET which is controlled by a comparator.

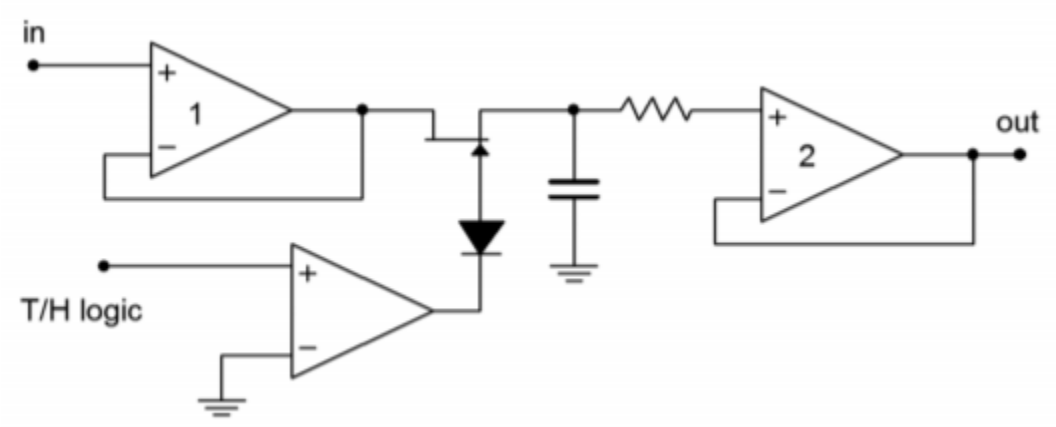


Figure 12.5.4 : Track-and-hold circuits. Op amp based track-and-hold.

When the T/H logic is high, the gate of the JFET is high, thus producing a low onresistance (i.e., a closed switch). When the comparator output goes low, the JFET is turned off creating a high impedance (i.e., open switch). In this state, op amp 2 is fed by the hold capacitor and buffers this potential to its output. For minimum droop, it is essential that the capacitor be a low leakage type and that op amp 2 have very low input bias current (e.g., FET input). The input resistor is used only to limit possible destructive discharge currents when the circuit is switched off. The diode positioned between the comparator and FET is used to prevent an excessively large positive comparator output potential from reaching the gate of the FET and possibly damaging it. (A comparator high will reverse-bias the series diode.) Figure 12.5.4 shows an alternate circuit using a closed-loop, current-mode approach. Note that the hold capacitor is now forming part of an integrator. While the open-loop form offers faster acquisition and settling times, the closed-loop system offers improved signal tracking.

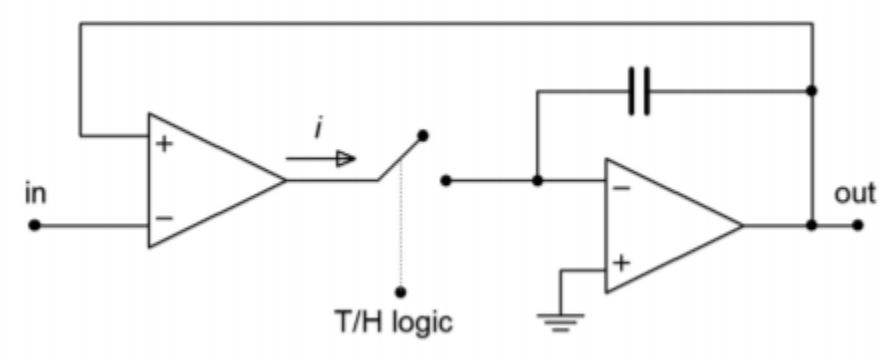


Figure 12.5.4 : Track-and-hold circuits. General closed-loop current mode track-and-hold.

A closed-loop voltage-mode is also possible, but the current form generally offers fewer problems with leakage and switching transients. For general-purpose work, a variety of track-and-hold amplifiers are available in IC form from several manufacturers.

COMPUTER SIMULATION

A simulation of a track-and-hold circuit similar to the one shown in Figure 12.5.4 is shown in Figure 12.5.5 . The 100 Hz input signal is being sampled at 2 kHz, or 20 times per cycle.

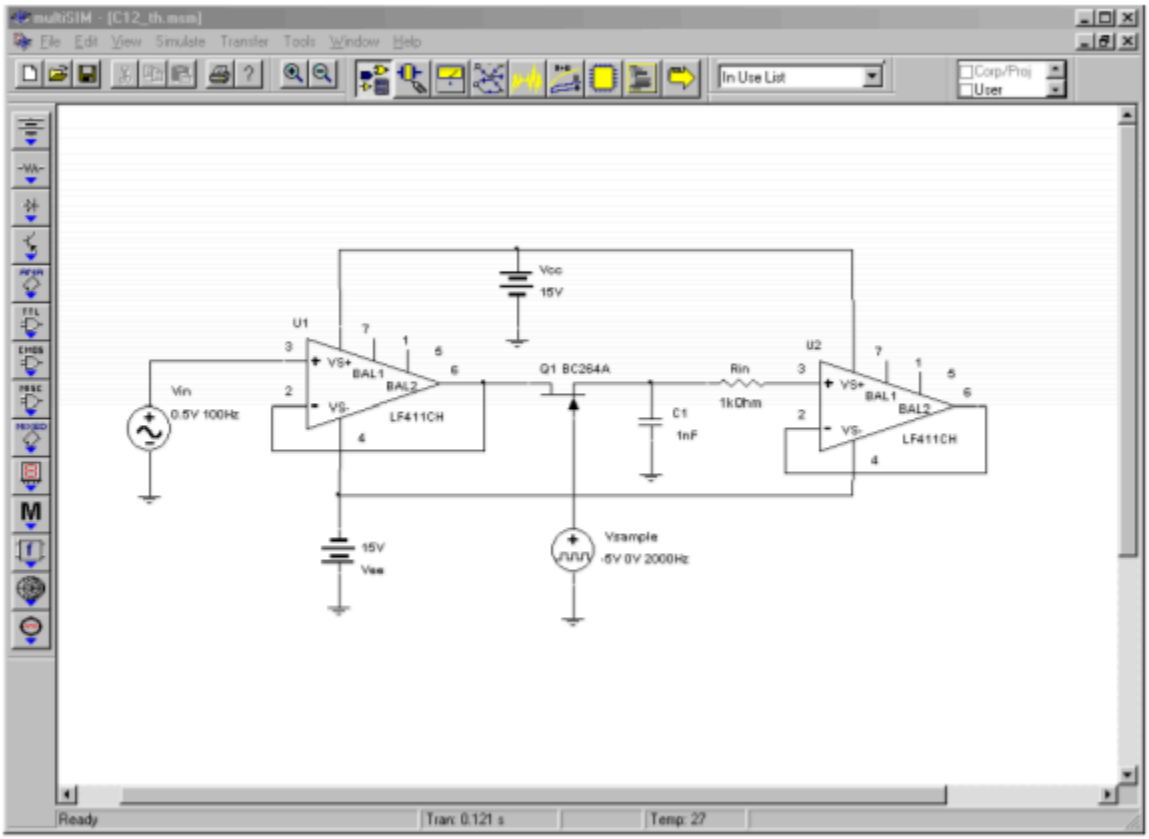


Figure 12.5.5 : Track and hold circuit in Multisim.

Three waveforms of interest are shown in the Transient Analysis graph. The input waveform is seen in the center as the smoothly varying sine wave. Along the bottom of the graph, narrow spikes can be seen. This is the T/H logic signal. The high portion of the pulse (at 0 volts) is the track logic, and

the wider low signal (dropping below the –1 volt limit of the graph) represents the hold logic. The stair-stepped sine wave is the output voltage. Note how the circuit acquires or tracks the input signal during the track pulse and then remains at that level when the hold logic is applied. It is during these hold times that the computations will be performed by the analogto-digital converter.

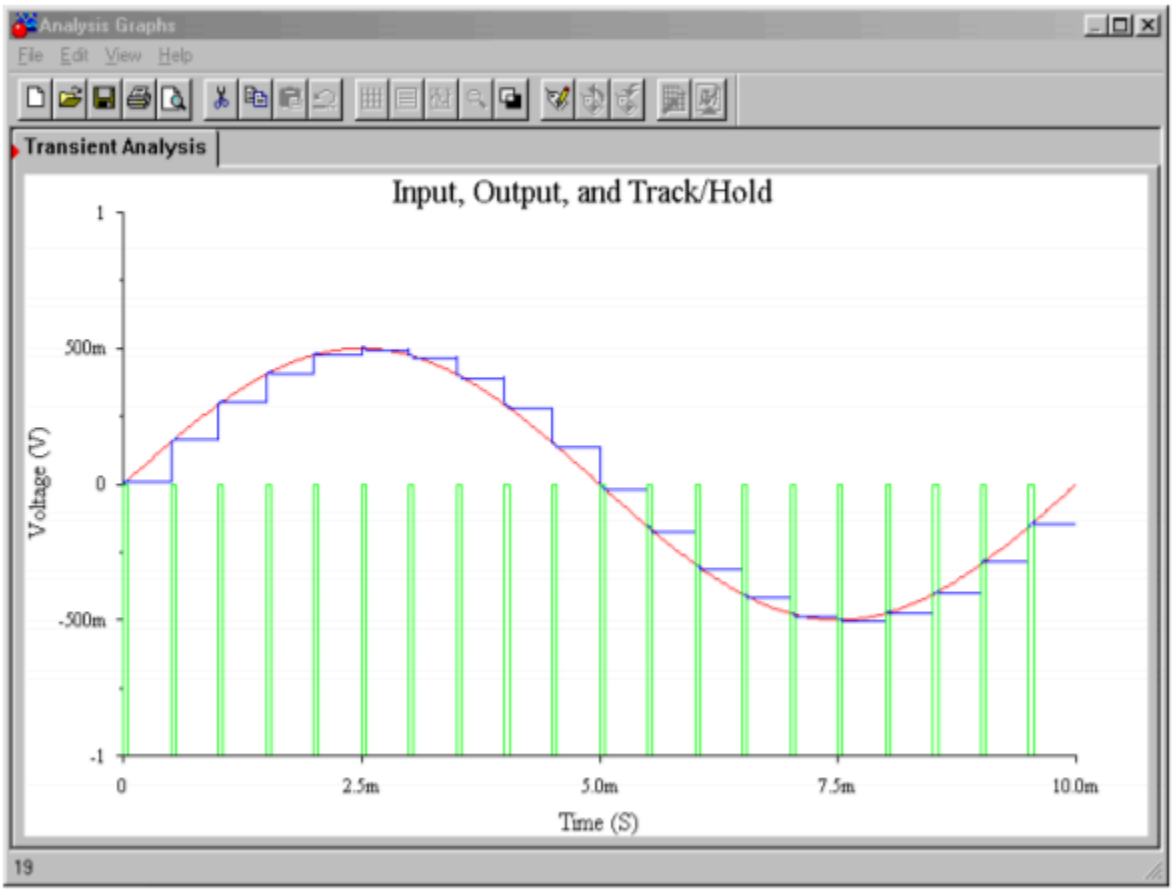


Figure 12.5.5 : Multisim waveforms for track-and-hold circuit.

ANALOG-TO-DIGITAL CONVERSION TECHNIQUES

Several different techniques have evolved for dealing with differing system requirements. These include flash, staircase, successive approximation, and delta-sigma forms.

Flash conversion is generally used for high-speed work, such as video applications. The circuits are usually low resolution. A flash converter is made up of a string of comparators as shown in Figure 12.5.6 . The input signal is applied to all of the comparators simultaneously. Each comparator is also tied into a reference ladder. Effectively, there is one comparator for each quantization step. When a given signal is applied, a number of comparators towards the bottom of the string will produce a high level, as will be greater than their references. Conversely, the comparators towards the top will indicate a low. The comparator at which the outputs shift from high to low indicates the step value closest to the input signal. The set of comparator outputs can be fed into a priority encoder that will turn this simple unweighted sequence into a normal binary word.

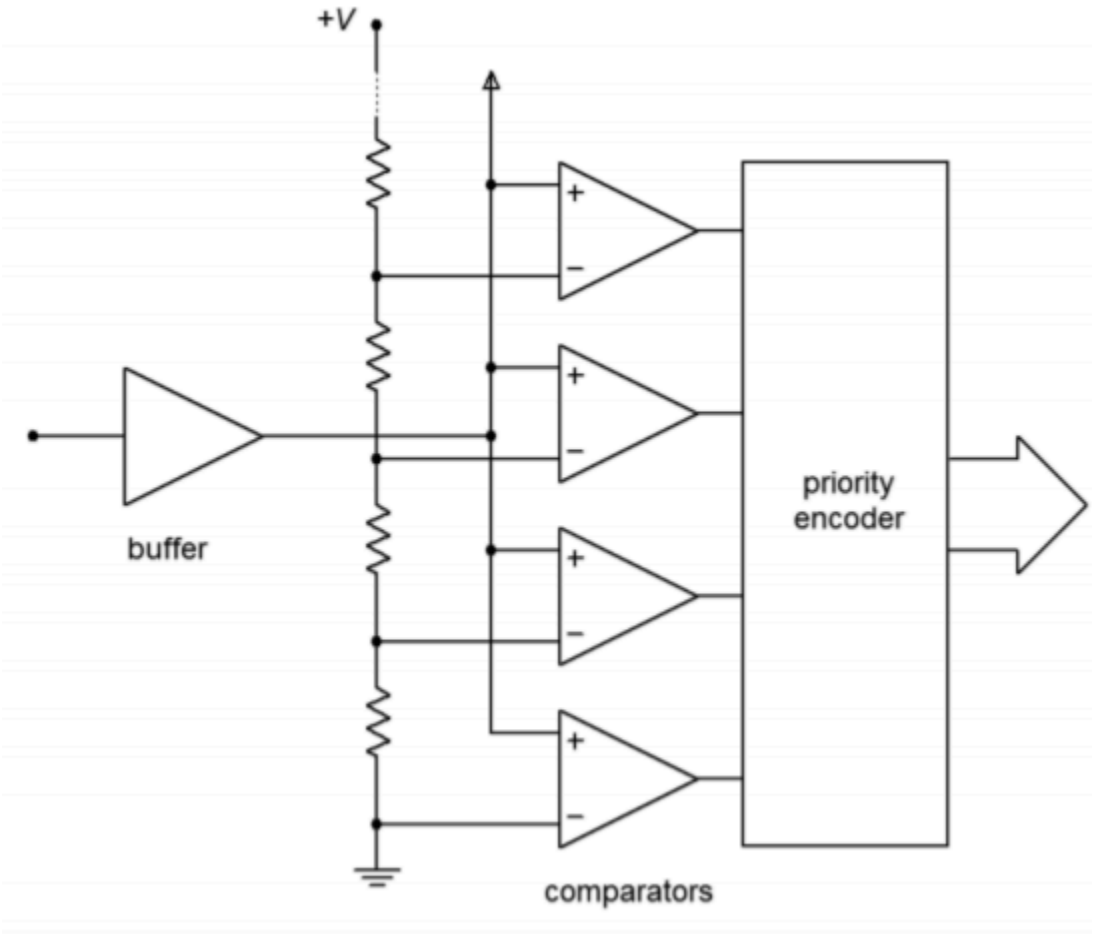


Figure 12.5.6 : Flash converter.

The only time delay involved in the conversion is that of logic propagation delay. Therefore, this conversion technique is quite useful for rapidly changing signals. Its downfall lies in the fact that one comparator is needed for each possible output step change. An 8-bit flash converter requires 256 comparators, whereas a 16-bit version requires 65,536. Obviously, this is rather excessive, and units in the 4- to 6-bit range are common. Although 6-bit resolution may appear at first to be too coarse for any application, it is actually quite useful for video displays.

For high-resolution work, some other technique must be used. One possibility is shown in Figure 12.5.7 . This is called a staircase converter. Its operation is fairly simple. When a conversion is first started, the output of the counter will be all zeros. This produces a DAC output of zero, and thus, the comparator output will be high. The next clock cycle will increment the up counter, causing the DAC output to increase by one step-level. This signal is compared against the input, and if the input is greater, the output will remain high. The clock will continue to increment the counter in this fashion until the DAC output just exceeds the input level. At this point the comparator output drops low, indicating that the conversion is complete. This signal can then be used to latch the output of the counter. This circuit gets its name from the fact that the waveform produced by the DAC looks like a staircase. The staircase technique can be used for very high-resolution conversion, as long as an appropriate high resolution DAC is used. The major problem with this form is its very low conversion speed, due to the fact that there must be time to test every possible bit combination. Consequently, a 16-bit system requires 65,536 comparisons. Even if a fast 1 microsecond DAC is used, this would limit

the sampling interval to nearly 66 milliseconds. This translates to a maximum input frequency before aliasing of only 7 Hz. Noting that the lowest frequency most humans can hear is about 20 Hz, this technique is hardly suitable for something like digital audio recording. In fact, for general-purpose work, the staircase system is avoided in favor of the successive approximation technique.

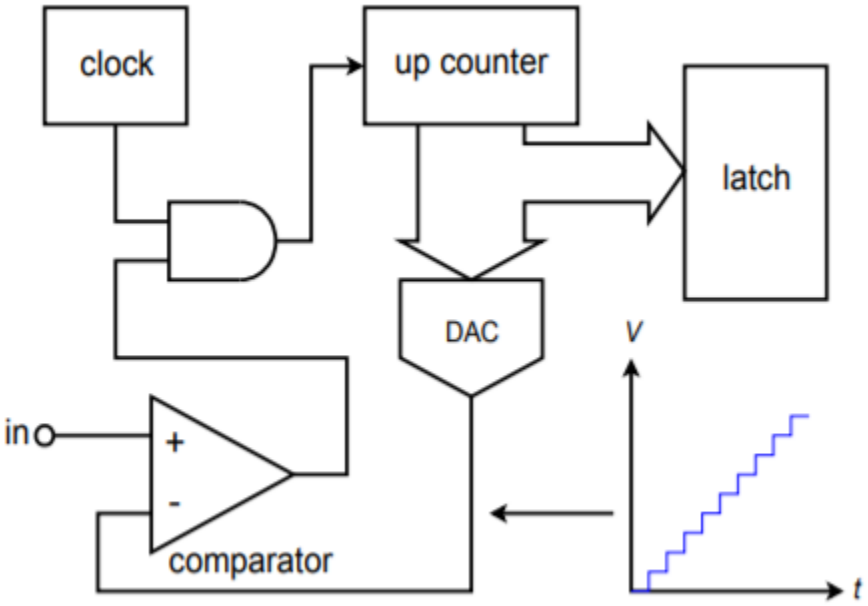


Figure 12.5.7 : Staircase converter.

Successive approximation is a good general-purpose solution suitable for systems requiring resolutions in the 16-bit area. Instead of trying to convert the input signal at one instant like the flash converter, this technique creates the output bit by bit. Unlike the staircase converter, each individual level does not need to be tested. Instead, a binary search algorithm is used. You might think of it as making a series of guesses, each time getting a little closer to the result. Each guess results in a simple comparison. For an ⍰ -bit output, ⍰ comparisons need to be made.

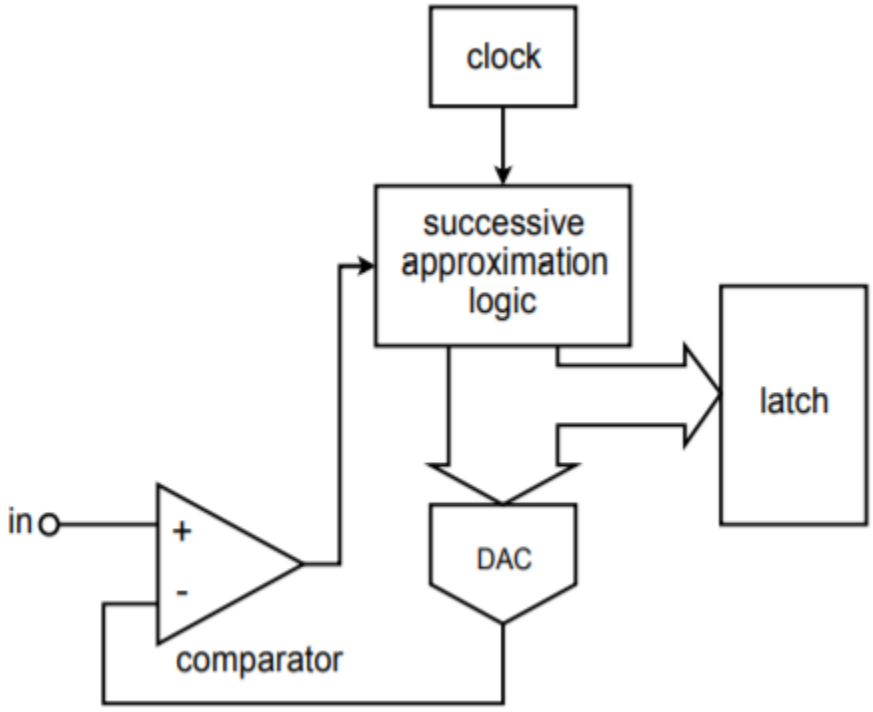


Figure 12.5.8 : Successive-approximation converter.

A block diagram of a successive approximation converter is shown in Figure 12.5.8 . Instead of a simple up counter, a circuit is used to implement the successive approximation algorithm. Here is how the circuit works. The first bit to be tested is the most significant bit. The DAC is fed a 1 with all of the remaining bits being set to 0 (i.e., 100000…) . This word represents half of the maximum value capable by the system. The comparator output indicates whether ◈◈◈is greater or less than the resulting DAC output. If the comparator output is high, it means that the required digital output must be greater than the present word. If the comparator output is low, then the present digital word is too large, so the MSB is set to 0. At this point, the next most significant bit is tested by setting it to 1. The new word is fed to the DAC, and again, the comparator is used to determine whether or not the bit under test should remain at 1 or be reset. At this point the two most significant bits have been determined. The remaining bits are individually set to 1 and tested in a similar manner until the least significant bit is determined. In this way, a 16-bit system only requires 16 comparisons. With a 1 microsecond DAC, conversion takes only 16 microseconds. This translates to a sampling rate of 62.5 kHz; thus a maximum input frequency of 31.25 kHz is allowed. As you can see, this is far more efficient than the staircase technique.

For the highest resolutions combined with high sample rates, delta-sigma conversion techniques are popular. Basically, a low-resolution converter is run at a rate many times higher than the Nyquist frequency (perhaps 256 times higher). A special digital filter called a decimator converts the low-resolution high sample-rate data stream into a lower rate with higher resolution. This technique can achieve conversions in the 50 kHz range with 20-bit resolution. Design and analysis of delta-sigma modulators and digital filters is an advanced topic beyond the scope of this text. We will, however, look at a representative IC in the next section.

ANALOG-TO-DIGITAL CONVERTER INTEGRATED CIRCUITS

In this section we will examine a few specific ADC ICs, along with selected applications. The ICs include the ADC0844, an 8-bit microprocessor-compatible unit; the ADC12181, a very fast 12-bit converter utilizing pipeling techniques; and the CS5396, a 24-bit converter designed primarily for audio applications.

ADC0844

A block diagram of the ADC0844 is shown in Figure 12.5.9 . Along with the built-in clock and 8-bit successive approximation register (SAR), the IC also includes an input multiplexer, tri-state latches and read, write, and chip select logic pins. This means that the ADC0844 is easily interfaced to a microprocessor data bus and may be used as a memory-mapped I/O device. This IC is a moderate-speed device, showing a typical conversion speed of 40 microseconds. The timing diagram for the ADC0844 is shown in Figure 12.5.10 . A conversion is initiated by bringing both the chip-select and write-logic lines low. The falling edge of write resets the converter and its rising edge starts the actual conversion. After the conversion period, which is set internally, the digital data may be transferred to the output latches with the readlogic line. Note that the pulse repetition rate of the write-logic line sets the sampling rate. Therefore, a small program running on the host microprocessor that reads and writes to the ADC may be used to control the sampling rate and store the data for later use.

National Semiconductor

ADC0844/ADC0848
8-Bit µP Compatible A/D Converters with Multiplexer Options

General Description

The ADC0844 and ADC0848 are CMOS 8-bit successive approximation A/D converters with versatile analog input multiplexers. The 4-channel or 8-channel multiplexers can be software configured for single-ended, differential or pseudo-differential modes of operation.

The differential mode provides low frequency input common mode rejection and allows offsetting the analog range of the converter. In addition, the A/D's reference can be adjusted enabling the conversion of reduced analog ranges with 8-bit resolution.

The A/Ds are designed to operate from the control bus of a wide variety of microprocessors. TRI-STATE® output latches that directly drive the data bus permit the A/Ds to be configured as memory locations or I/O devices to the microprocessor with no interface logic necessary.

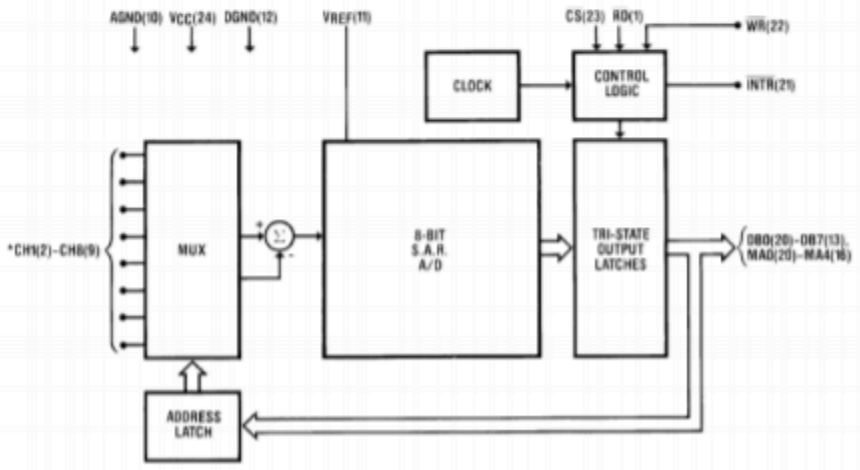
Features

- Easy interface to all microprocessors
- Operates ratiometrically or with 5 V_{DC} voltage reference
- No zero or full-scale adjust required
- 4-channel or 8-channel multiplexer with address logic
- Internal clock
- 0V to 5V input range with single 5V power supply
- 0.3" standard width 20-pin or 24-pin DIP
- 28 Pin Molded Chip Carrier Package

Key Specifications

■ Resolution	8 Bits
■ Total Unadjusted Error	±½ LSB and ± 1 LSB
■ Single Supply	5 V_{DC}
■ Low Power	15 mW
■ Conversion Time	40 µs

Block and Connection Diagrams



*ADC0848 shown in DIP Package CH5-CH8 not included on the ADC0844

Figure 12.5.9 : The ADC0844. Reprinted courtesy of Texas Instruments

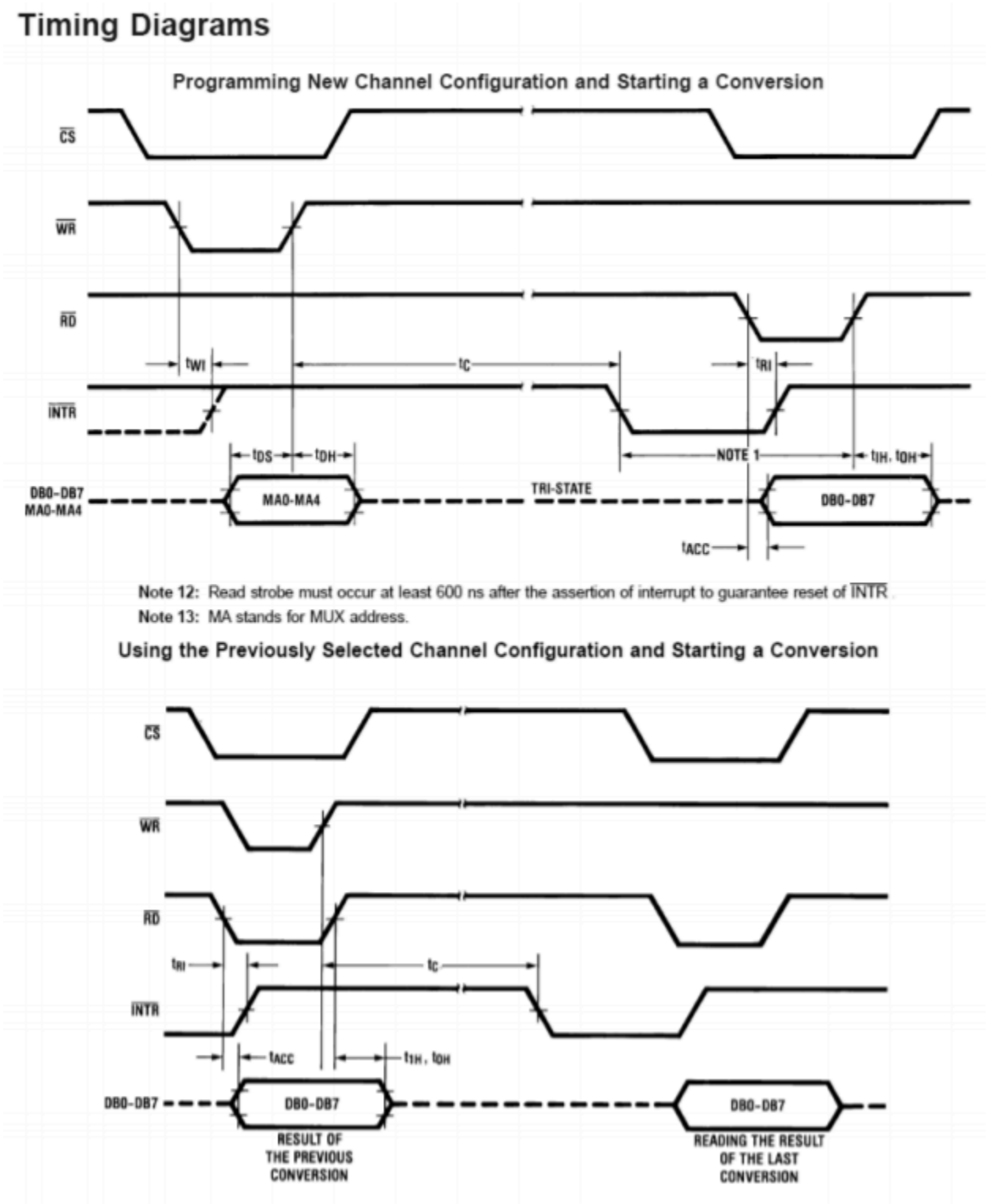


Figure 12.5.10 : Timing diagram for the ADC0844. Reprinted courtesy of Texas Instruments

LTC2282

The LTC2282 is a 12-bit, 105 mega-samples per second converter with an internal sample-and-hold. Its data sheet is shown in Figure 12.5.11 . In order to achieve its combination of high sample rate with

relatively high resolution, the LTC2282 relies on a technique known as pipelining. In this particular chip, the pipeline consists of six stages. Each stage produces a digital signal of just three bits and an error signal known as a residual. The residual is passed to the following stage where it is multiplied by a fixed gain, thus bringing it up to the former bit-weight. The newly resulting residue is passed to the next stage where the process is repeated. In essence, the signal propagates down the pipeline in a fashion conceptually similar to the successive approximation technique. Note that pipelining speeds the conversion process because once the residual is passed to the next stage, the prior stage(s) can begin work on the following sample(s).

LINEAR TECHNOLOGY

LTC2282

Dual 12-Bit, 105Msps Low Power 3V ADC

FEATURES

- **Integrated Dual 12-Bit ADCs**
- **Sample Rate: 105Msps**
- **Single 3V Supply (2.85V to 3.4V)**
- **Low Power: 540mW**
- **70.1dB SNR, 88dB SFDR**
- 110dB Channel Isolation at 100MHz
- Flexible Input: $1V_{P-P}$ to $2V_{P-P}$ Range
- 575MHz Full Power Bandwidth S/H
- Clock Duty Cycle Stabilizer
- Shutdown and Nap Modes
- Pin Compatible Family
  105Msps: LTC2282 (12-Bit), LTC2284 (14-Bit)
  80Msps: LTC2294 (12-Bit), LTC2299 (14-Bit)
  65Msps: LTC2293 (12-Bit), LTC2298 (14-Bit)
  40Msps: LTC2292 (12-Bit), LTC2297 (14-Bit)
  25Msps: LTC2291 (12-Bit), LTC2296 (14-Bit)
  10Msps: LTC2290 (12-Bit), LTC2295 (14-Bit)
- 64-Pin (9mm × 9mm) QFN Package

DESCRIPTION

The LTC®2282 is a 12-bit 105Msps, low power dual 3V A/D converter designed for digitizing high frequency, wide dynamic range signals. The LTC2282 is perfect for demanding imaging and communications applications with AC performance that includes 70.1dB SNR and 85dB SFDR for signals at the Nyquist frequency.

DC specs include ±0.4LSB INL (typ), ±0.2LSB DNL (typ) and no missing codes over temperature. The transition noise is a low $0.3LSB_{RMS}$.

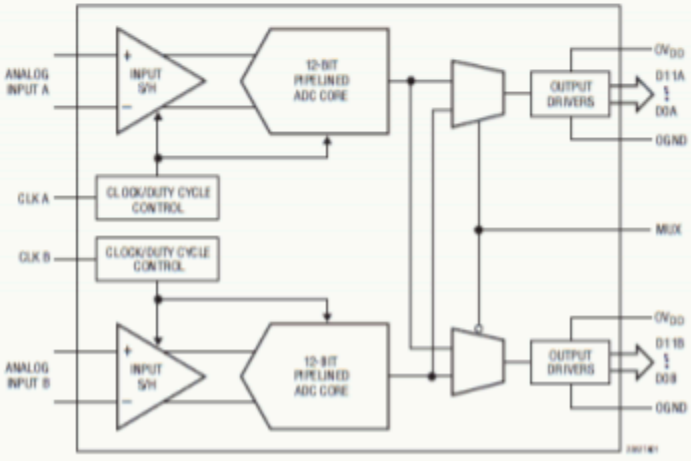
A single 3V supply allows low power operation. A separate output supply allows the outputs to drive 0.5V to 3.6V logic.

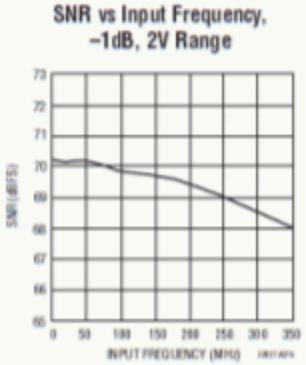
A single-ended CLK input controls converter operation. An optional clock duty cycle stabilizer allows high performance at full speed for a wide range of clock duty cycles.

APPLICATIONS

- Wireless and Wired Broadband Communication
- Imaging Systems
- Spectral Analysis
- Portable Instrumentation

TYPICAL APPLICATION





22852fb

Figure 12.5.11 : LTC2282. Reprinted courtesy of Linear Technology

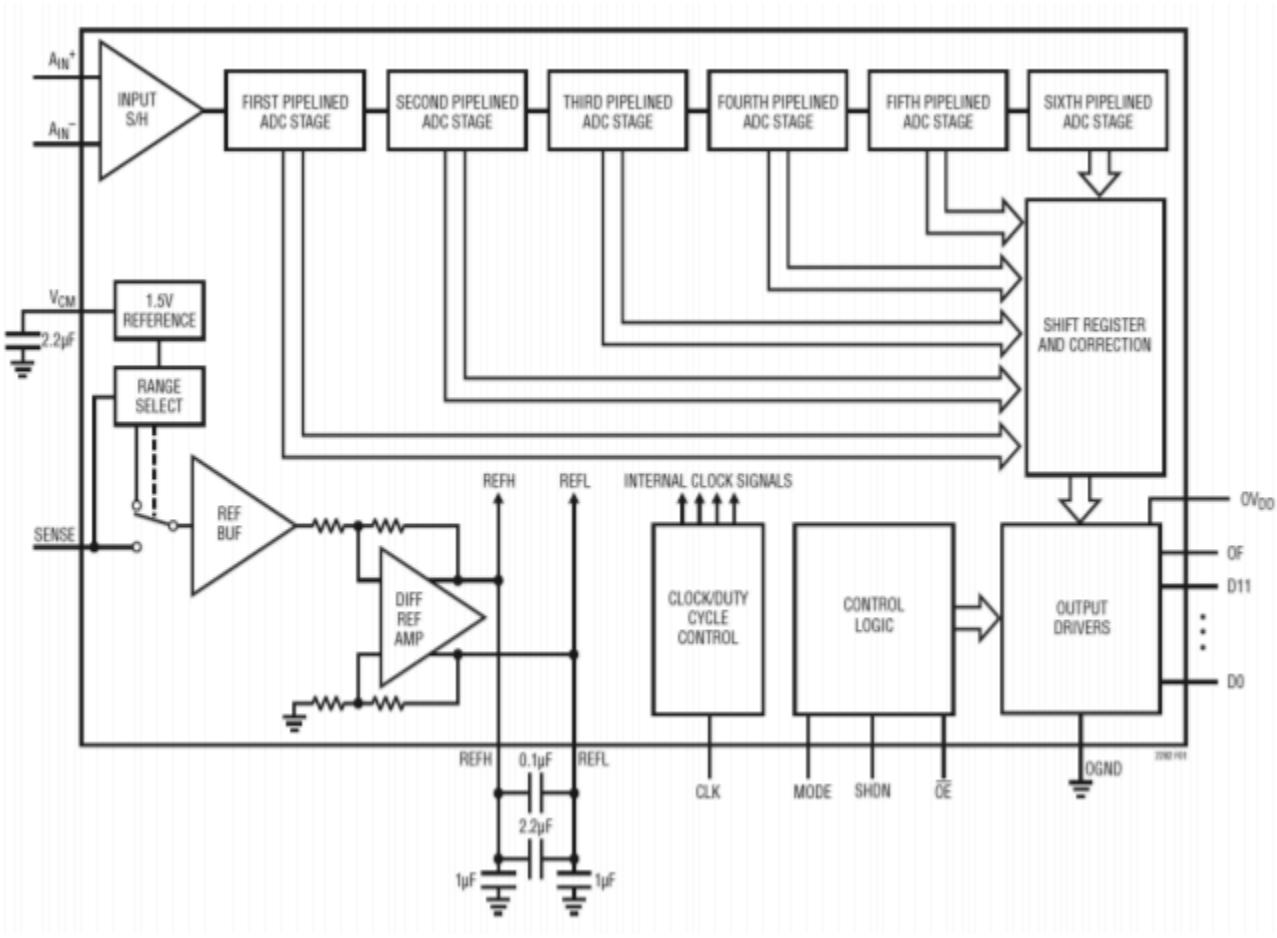


Figure 12.5.11 : LTC2282 block diagram. Reprinted courtesy of Linear Technology

The LTC2282 is used for applications requiring both high speed and high resolution. It also offers self-calibration, single +5 V power supply operation, and low power consumption.

CS5381

The CS5381 offers stereo 24-bit resolution with a maximum sampling rate of 192 kHz. This is ideal for a wide range of high-quality digital audio systems. The CS5381 data sheets are shown in Figure 12.5.12 . This IC uses the delta-sigma conversion technique and achieves a dynamic range of 120 dB. The signal-to-noise plus distortion rating (THD+N) is typically 110 dB. The device over-samples the input at 6.144 MHz and has logic settings for 1X, 2X and 4X effective conversion rates at its output. For example, the standard pro-audio sampling rate is 48 kHz with double rate at 96 kHz and quad rate at 192 kHz.

As with any high-resolution converter, you must be particularly careful with the circuit layout or excessive noise may result. Attention must be paid to proper power supply bypassing. The CS5381, like many ICs in this category, requires separate analog and digital power supplies. Although this IC only requires 5 V, note that there are separate physical connections for analog and digital power. Standard logic supplies are often too noisy for use on such precision devices and will degrade overall audio performance when applied to the analog portion of the IC.

CIRRUS LOGIC®

CS5381

120 dB, 192 kHz, Multi-Bit Audio A/D Converter

Features

- Advanced Multi-bit Delta-Sigma Architecture
- 24-bit Conversion
- 120 dB Dynamic Range
- -110 dB THD+N
- Supports All Audio Sample Rates Including 192 kHz
- 260 mW Power Consumption
- High-Pass Filter or DC Offset Calibration
- Supports Logic Levels between 5 and 2.5 V
- Differential Analog Architecture
- Low-Latency Digital Filtering
- Overflow Detection
- Pin-Compatible with the CS5361

General Description

The CS5381 is a complete analog-to-digital converter for digital audio systems. It performs sampling, analog-to-digital conversion, and anti-alias filtering - generating 24-bit values for both left and right inputs in serial form at sample rates up to 216 kHz per channel.

The CS5381 uses a 5th-order, multi-bit delta-sigma modulator followed by digital filtering and decimation, which removes the need for an external anti-alias filter. The ADC uses a differential architecture which provides excellent noise rejection.

The CS5381 is available in 24-pin TSSOP and SOIC packages for Commercial grade (-10° to +70° C). The CDB5381 Customer Demonstration board is also available for device evaluation and implementation suggestions. Please refer to the "Ordering Information" on page 22.

The CS5381 is ideal for audio systems requiring wide dynamic range, negligible distortion, and low noise - such as A/V receivers, DVD-R, CD-R, digital mixing consoles, and effects processors.

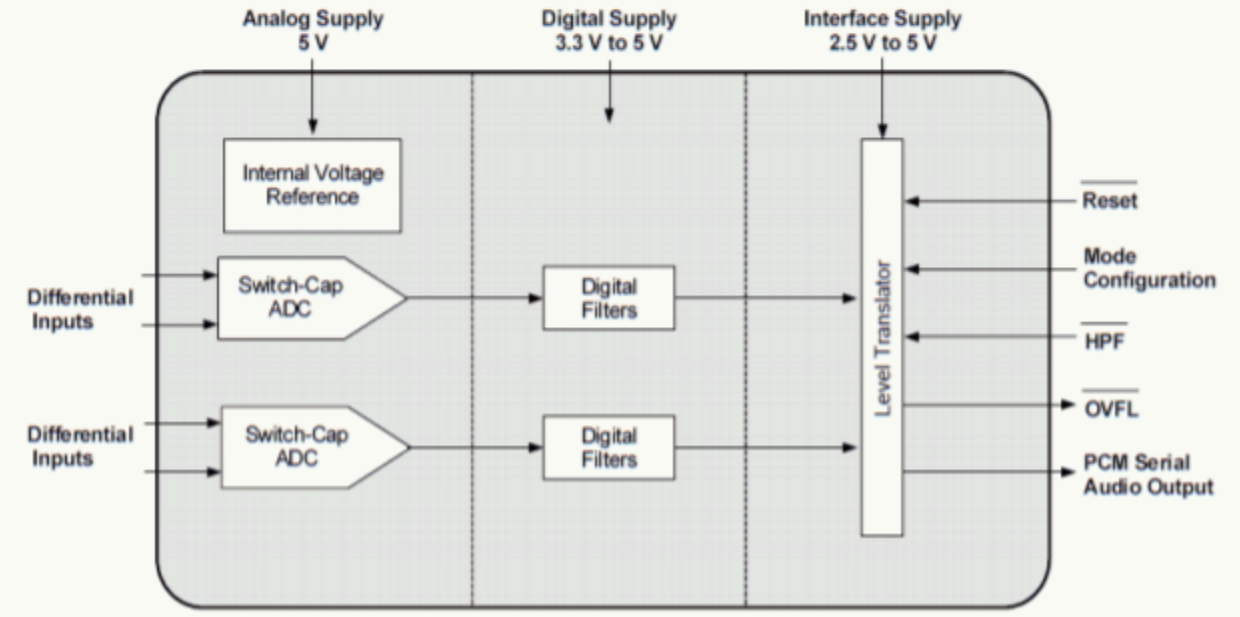


Figure 12.5.12◈: CS5381. Reprinted courtesy of Cirrus Logic, Inc.

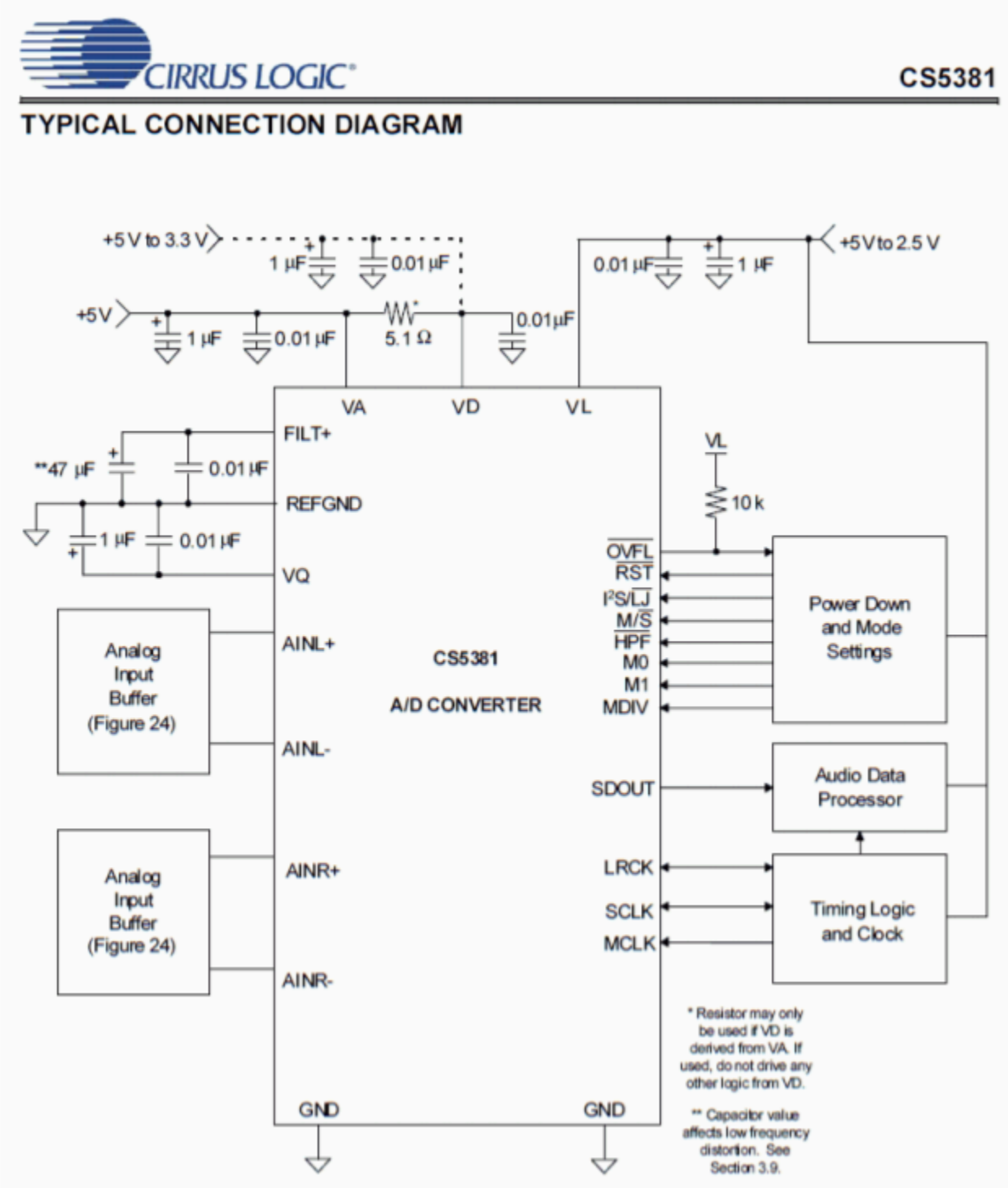


Figure 12.5.12 : CS5381 connection diagram. Reprinted courtesy of Cirrus Logic, Inc.

APPLICATIONS OF ANALOG-TO-DIGITAL CONVERTER INTEGRATED CIRCUITS

Example 12.5.1

Perhaps the most straightforward application of analog-to-digital conversion is the acquisition of a signal. Once in the digital domain, the signal may be processed in a variety of ways. An example of this is the Digital Sampling Oscilloscope (DSO). A simplified block diagram of a DSO is shown in Figure 12.5.13 . Since the final output of the system is a simple graph, it generally does not make sense to resolve the input beyond 8 bits (256 steps), and often even fewer bits may be used. Typically, high sampling rates are more important than fine resolution in this application. Consequently, 12- and 16-bit converters are not found here. Instead, lower-resolution converters capable of sampling at hundreds of MHz are used.

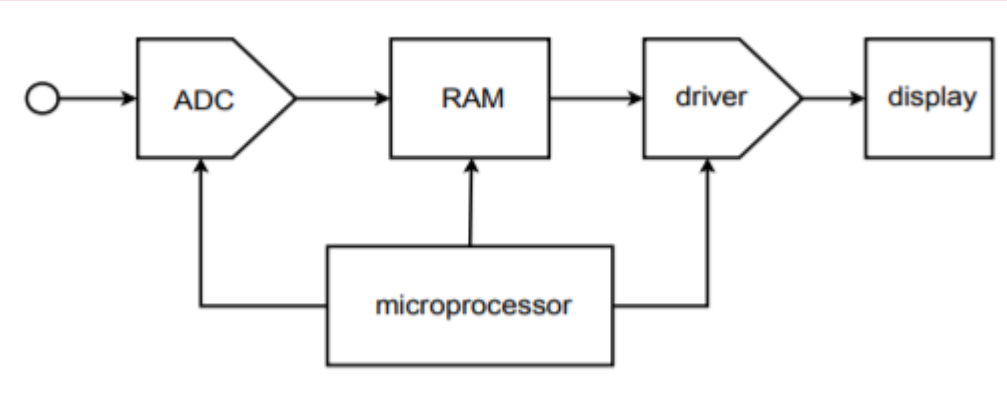


Figure 12.5.13 : Basic digital sampling oscilloscope.

DSOs can usually be run in one of two modes: continuous, or single-shot. In single-shot mode, the ADC acquires the signal and stores it in memory. This data can then be routed to the display circuits continuously in order to create a trace on the display. The signal is captured and stored in computer memory, thus it can be replayed virtually forever without a loss of clarity. This is very useful for catching quick, non-repetitive transients. Once the signal is captured, it may be examined at leisure. For that matter, if some form of level-sensing logic is included, sampling can be initiated by specific transient events. This is known as baby-sitting. For example, you might suspect that a circuit you have designed occasionally emits an undesirable voltage spike. It is not practical for you to hook up the circuit and stare at the face of an oscilloscope, perhaps for hours, waiting for the system to misbehave. Instead, the DSO can be programmed to wait for the transient event before recording. In this way, you can leave the system, and when you return some time later, the spike will have been recorded and will be waiting your inspection.

Figure 12.5.14 : Commercial MDO. Copyright © Tektronix. Reprinted with permission. All rights reserved

An extension of this concept is pre-trigger recording. In this variation, the DSO is constantly recording and "throwing away" data. When the transient spike finally occurs, the DSO has a snapshot of the events leading up to the spike, as well as the spike itself. This can provide very useful information in some applications. In either case, since the data is in digital form, it may be off-loaded to a computer for further analysis. Indeed, many DSOs offer some interesting on-board analysis functions, including signal smoothing (noise reduction), trace cursors which allow for easy delta time/delta voltage computation, basic math functions, spectrum computation via FFT and more.

In continuous mode, a DSO appears to operate in much the same fashion as an ordinary analog oscilloscope. In this mode, the DSO samples the input signal and passes the data out to memory. From here the data is relayed to the display circuitry where a trace is produced. The trace is being constantly updated, so any change in the input will be quickly displayed.

The basic concept of the DSO has been expanded to the MSO and the MDO. The MSO, or Mixed Signal Oscilloscope, adds digital domain measurements while the MDO, or Mixed Domain Oscilloscope, extends the MSO into the frequency domain as well.

Example 12.5.2

Another interesting application of the analog-to-digital converter is in the digital sampling music keyboard or drum computer (generally referred to as a sampler). The usage of a sampler is quite straightforward, actually. The idea is to mimic the sound of a given instrument (such as a trumpet or flute) from the keyboard. Before the advent of the sampler, this was done by properly setting the filters,

amplifiers and oscillators of a keyboard synthesizer. Although the resulting sounds were reasonably close to the desired instrument, they usually weren't close enough to fool the average listener. The sampler bypasses the problems of synthesizers by directly recording an instrument with an analog-to-digital converter. For example, a trumpet player might sound an ♢ into a microphone that is connected to the sampler. This note is digitized and stored in RAM. Now, when the keyboard player hits the ♢ key, the data is retrieved from RAM and fed to a DAC where it is reconstructed. The result is the exact same note that the trumpet player originally produced. By recording several different pitches from many different instruments, the keyboard player literally has the power of an entire orchestra at his fingertips. Of course, there is no limit to the sounds that might be sampled, and the keyboardist could just as easily use the sampler to "play" a collection of dog barks, door slams, and bird calls. A sampling drum computer is similar to a sampling keyboard, but replaces the standard musical keyboard with a series of buttons that allow the musician to create a programmed sequence of notes. This sequence can be played back at any time, and at virtually any tempo. In this manner, an entire percussion section may be simulated using digital recordings of real drums. A block diagram of a typical musical sampling system is shown in Figure 12.5.15 .

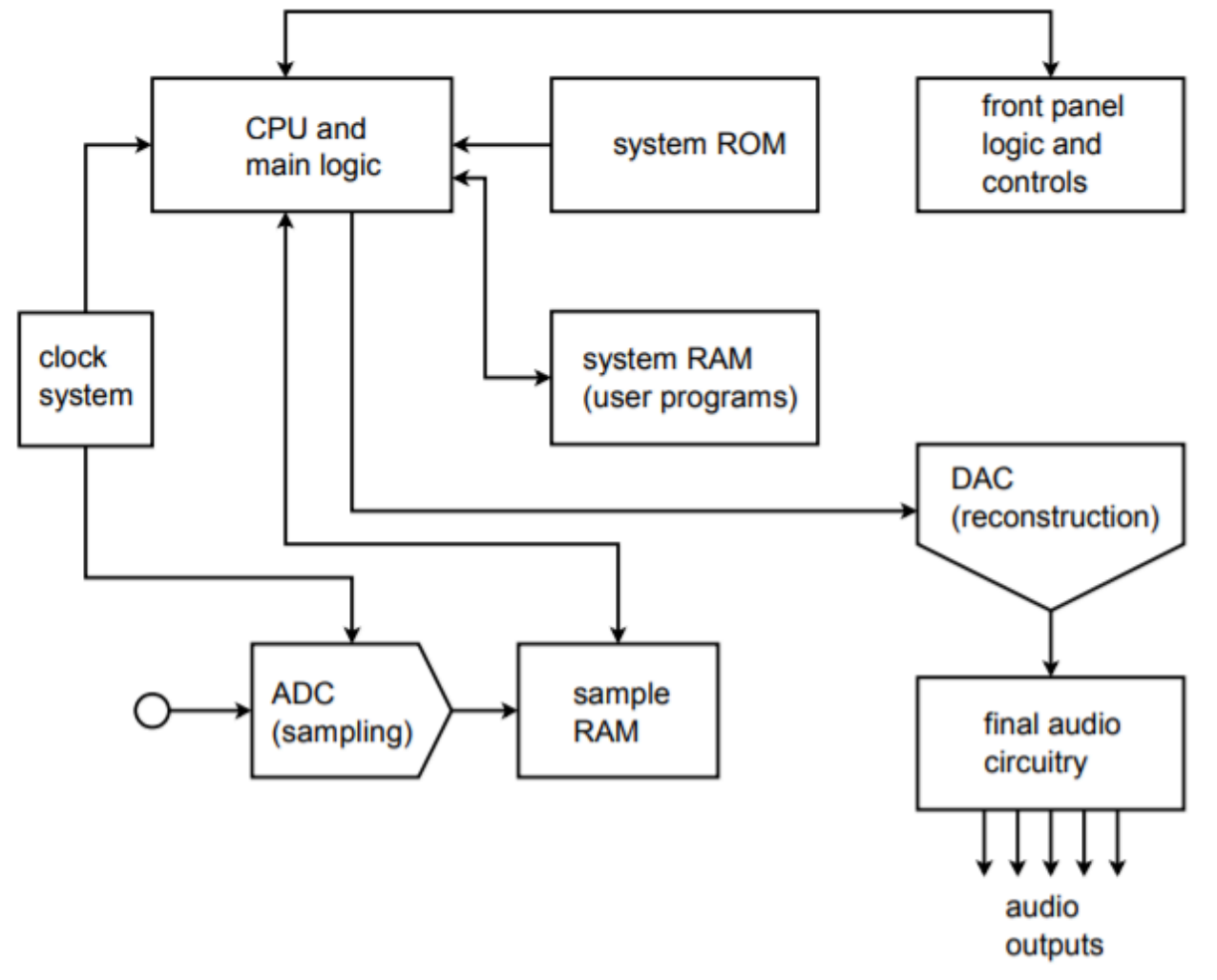


Figure 12.5.15 : Block diagram of musical instrument sampling and playback device.

Further enhancements to the digital sampling musical instrument include direct connections to personal computers. In this way, the computer can be used to analyze and augment the sound samples in new ways. For example, the personal computer may be used to create a graphic display of the waveform, or compute

and display the spectral components of the sound using a popular technique called the Fast Fourier Transform. An example is shown in Figure 12.5.16 . Finally, the computer can be programmed to “play” the sampler. In this way, even non-keyboardists can take advantage of the inherent musical flexibility this system offers.

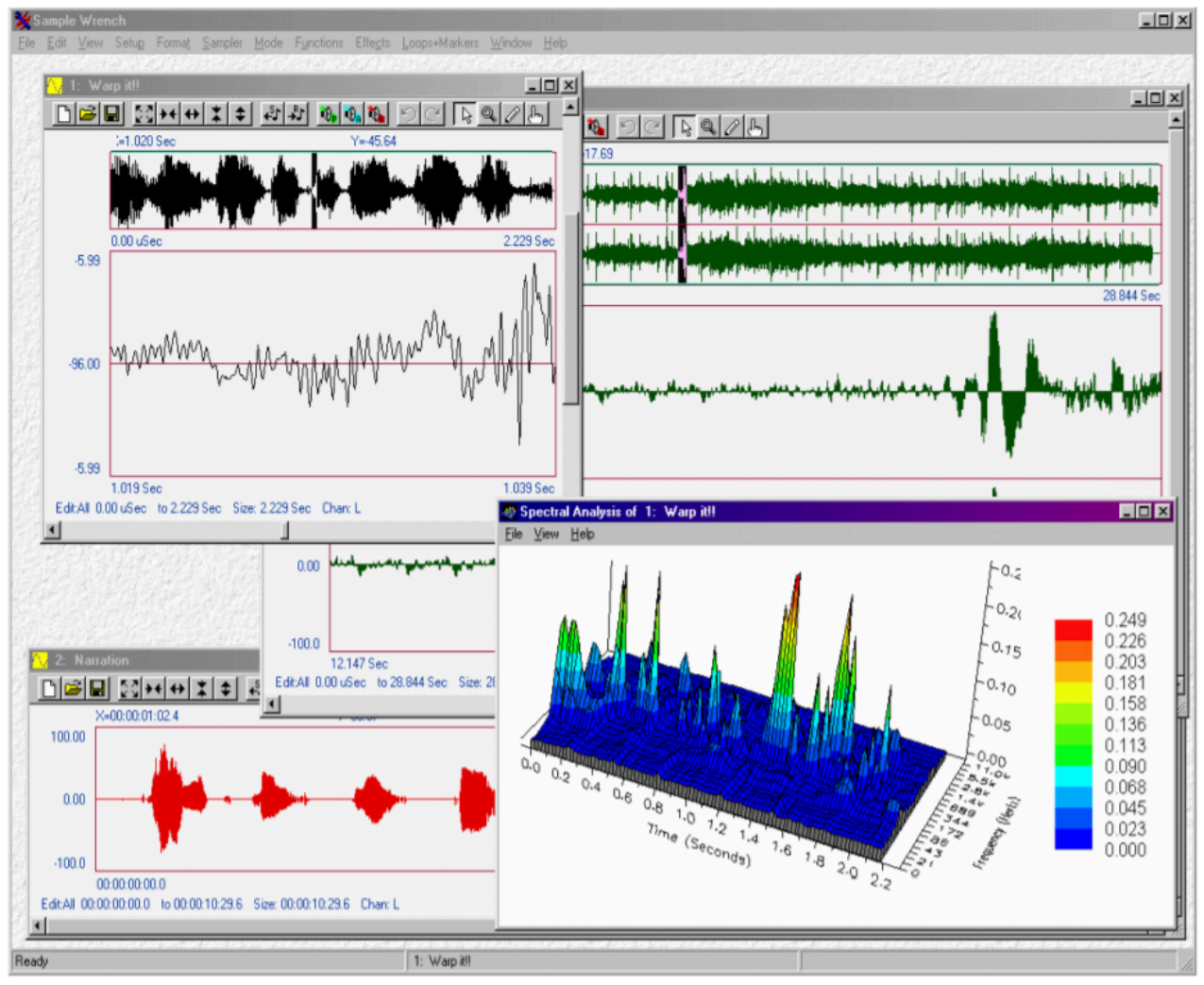


Figure 12.5.16 : Commercial waveform display and editing software. Reprinted courtesy of dissidents

16.6 EXTENDED TOPIC - DIGITAL SIGNAL PROCESSING

Much has been said in this chapter concerning the ability of personal computers to store and manipulate data in the digital domain. This process is called digital signal processing (DSP). The sampled data is generally stored in RAM as a large array of values. Normally, the data is in integer form. This is because microprocessors are generally faster at performing integer calculations than floating point calculations. Also, AD converters produce integer values, so this is doubly convenient. As an example, if 8-bit resolution is used, each sample point will require one byte (8 bits). The byte is the fundamental unit of computer memory storage, and thus storage as a simple byte multiple (8 bits, 16 bits, 32 bits) is straightforward. The data may be stored in either unipolar (unsigned) or bipolar (signed) formats. For unsigned systems, the data value "0" means "most negative peak". In signed systems, a two's compliment form is normally used. Here, a data value of "0" means "0 V". All negative values have their most significant bit set (for an 8 bit representation, –1=11111111,–2=11111110, etc.). Input signals are often AC, thus signed representations are quite common.

In BASIC, the data array might be called `data()`, and individual elements are accessed by properly setting the array index. The first element is `data(1)`, the second is `data(2)`, and the `N`th element is `data(N)`. Although BASIC is useful, it is not as powerful as languages such as C, which offer direct manipulation of memory addresses via pointers. In C, the starting, or base, address of the data might be given the name `data\_ptr`. To access different elements in the array, an offset is added to the base. The address of the second element is `data\_ptr + 1`, the third is `data\_ptr + 2`, and so on. Alternately, we could just increment the pointer continuously, as in `data\_ptr++`. Direct manipulation of pointers is generally faster for the microprocessor than array indexing. To find the value stored at a given address, C uses the indirection operator, `*`. To assign the second data value to the variable `x`, you would say `x = *(data\_ptr + 1)` or simply `x = *data\_ptr++` if the pointer is being continuously incremented. The BASIC equivalent is `x = data(2)`.

No matter how the data is accessed, any mathematical function may be applied to the data elements. In the examples below, C is used to illustrate the concepts. Only the processing portion of the code is shown. The source array is called `src()` and has `TOTAL` number of elements. The destination array is called `dest()`. Also, questions concerning integer versus floating point representation are bypassed in order to keep the examples as straightforward as possible. In the real world, problems such as execution speed, round-off error, and overflow cannot be ignored.

Let's start with something simple. Suppose we would like to simulate the operation of an inverting buffer. An inverting buffer simply multiplies the input signal by –1. Here is how we do this in the digital domain: each point is multiplied by –1.

```
for( i=0; i<TOTAL; i++ ) { *dest++ = -1 * (*src++); }
```

We can extend this concept a bit further by replacing "–1" with a variable called "gain". In this way, we can alter the signal amplitude digitally.

Another useful calculation is average value. For signed data, the average value is equal to the DC offset component.

```
sum = 0;
for( i=0; i<TOTAL; i++ )
{
     sum += *src++;
}
average = sum / TOTAL;
```

In order to add a DC offset, a constant is added to each data element in turn.

```
for( i=0; i<TOTAL; i++ )
{
    *dest++ = offset + (*src++);
}
```

We might wish to combine two different waveforms into a third waveform. Simple addition is all that is required. (The example assumes that both arrays are the same size).

```
for( i=0; i<TOTAL; i++ )
{
    *dest++ = (*src++) + (*src2++);
}
```

If the two source arrays are multiplied instead of added, signal modulation may be produced.

These examples are relatively simple. With proper programming, a wide variety of functions may be simulated including spectral analysis, translation to new sampling rates, and the realization of complex filter functions that are impractical to implement in the analog domain. The major drawback of DSP techniques is the inherent computing power required. Thus, real-time application of complex DSP principles is not possible without using powerful computers. Specialized ICs have been designed to implement DSP sub-functions very quickly. These chips make realtime DSP practical and ecomical. The DSP ICs are programmable, thus they offer the distinct advantage of being able to produce different functions at different times by loading new programs. In this way, they are far more flexible than dedicated analog circuits. For example, an all-DSP guitar effects unit encompassing a broad range of functions such as equalization, pitch shifting, chorusing, reverberation and other effects can be designed. The amplitude and spectral shaping is performed directly in the digital domain instead of using dedicated analog circuitry.

16.7 SUMMARY

In this chapter we have covered the basic concepts of analog-to-digital and digitalto-analog conversion. By placing a signal in the digital domain, a variety of new analysis, storage, and transmission techniques become available.

The process of transforming an analog signal into a digital representation is referred to as quantization, or more simply, as digitizing or sampling. One of the most popular methods used is PCM, pulse code modulation. In this scheme, the input signal is measured, or sampled, at a constant rate. Each sample point is represented as a digital word. The sequence of words describes the input signal and is used to recreate it, if necessary. The ultimate accuracy of the conversion is dependent on the resolution and sampling rate of the system. Resolution refers to the number of bits present in the digital word. An 8-bit word can represent 256 steps, whereas a 16-bit word is much finer, offering 65,536 discrete steps. With so many steps, round-off error is much less of a problem in 16-bit systems than in, say, 8- or 12-bit systems.

The minimum allowable sampling rate is twice the highest input frequency. In other words, at least two samples per cycle are required for the highest harmonic in the input signal. Another way of stating this is that no input frequency component can exceed the Nyquist frequency, which is defined as one-half of the sampling frequency. If the input exceeds this limit, alias distortion may occur. Aliases are nonharmonically-related frequencies that are effectively created by improper sampling. In order to remove all possibility of alias distortion, special high-order low-pass filters, called anti-alias filters, are normally placed before the sampling circuitry.

Two popular methods of analog-to-digital conversion are the flash and successive approximation techniques. Flash converters are very fast, but require one comparator per output step, so they are not normally used where high resolution is required. Successive approximation takes longer than flash conversion, but can produce resolutions in excess of 16 bits. In either case, the conversion process is not instantaneous, and any fluctuation of the input signal during the conversion can produce errors. In order to alleviate this difficulty, special track-and-hold amplifiers are used between the anti-alias filters and the AD converter to create a non-varying signal.

Once the signal has been digitized, it may be stored in RAM or some other media for future use, or directly analyzed. Often, a personal computer can prove to be very useful for waveform analysis.

To reconstruct the waveform, a digital-to-analog converter is used. In essence, this is usually little more than a weighted summing amplifier. Due to accuracy and construction constraints, an ◈/2◈ ladder technique is often employed. In order to smooth out the resulting waveform and remove any remaining digital "glitches", the signal is passed through a reconstruction filter. This is a low-pass filter and is often called a smoothing filter. Generally, the process of digital-to-analog conversion is much faster than analog-to-digital conversion. Indeed, the successive approximation analog-to-digital scheme requires an internal digital-to-analog converter.

Applications for AD and DA converters range from laboratory instruments such as digital sampling oscilloscopes and hand-held digital multimeters, to industrial instrument and device control. AD and

DA systems have found a large application market in the commercial sector. Uses include the popular music CD player and musical keyboard equipment.

16.8 PROBLEMS

REVIEW QUESTIONS

1. What is PCM?
2. Define the term resolution.
3. What is quantization?
4. What is an alias, and how is it produced? How is an alias avoided?
5. Define Nyquist frequency, and discuss the importance of this parameter.
6. Explain how a summing amplifier may be used to create a digital-to-analog converter.
7. Explain the difference between integral nonlinearity and differential nonlinearity.
8. What is a smoothing (reconstruction) filter?
9. What is the purpose of a track-and-hold amplifier?
10. Detail the differences between flash conversion and the successive approximation technique. Where would each type be used? What are their limitations?
11. Give several examples of possible DSP functions.

PROBLEMS

Analysis Problems

1. Determine the number of quantization steps for a 10-bit system.
2. A 14-bit converter produces a maximum peak-to-peak output of 2.5 V. What is the step size?
3. Determine the dynamic range of the converters in Problems 1 and 2.
4. We wish to resolve a 1 V peak-to-peak signal to at least 1 mV. What is the minimum allowable number of bits in the converted data?
5. We wish to create analog signals using an arbitrary waveform generator. If we send out digital words at the rate of 50 kHz, what is the maximum allowable conversion speed for the DAC?
6. Assume that we are trying to digitize ultrasonic signals lying between 25 kHz and 45 kHz.
    A. What is the Nyquist frequency?
    B. What is the minimum acceptable sampling rate?
7. 7. Determine the maximum allowable conversion time for the ADC of Problem 6.
8. 8. Given a 14-bit ADC,
    A. Determine the number of comparators required for the flash technique.

B. Determine the number of comparisons required if the successive approximation technique is used.

9. Assume that a single 16-bit ADC is connected to an embedded computer. The sampling rate is 10 kHz. Determine the data rate in bytes per second.
10. Referring to Problem 9, if the computing device has 350 k bytes of RAM available for data storage, how much time does this represent?
11. DAT (digital audio tape) recorders normally use a 16 bit representation with a sampling rate of 48 kHz. If the unit is used to record a performance of Stravinsky's "Rite of Spring" (35 minutes total), what is the required storage capacity in bytes?
12. If the data is transferred serially from the DAT of Problem 11 to a digital signal processing IC in real time, what is the width of each individual pulse?
13. A 12-bit 2-microsecond DAC is used as part of a discrete successive approximation analog-to-digital converter. Assuming that logic delays and signal settling times are negligible, determine:
    A. The minimum time allowable between sample points.
    B. The maximum input signal frequency without aliasing.
14. A 6-bit video DA converter produces a maximum output swing of approximately 1.25 V (unipolar). Determine the output voltage for the following digital input words.
    A. 000001
    B. 100000
    C. 111111
    D. 011101
15. A 10-bit instrumentation DAC produces an output of 16 mV with an input of 0000000100. Determine:
    A. The step size
    B. The maximum output signal.
16. An 8-bit ADC produces a full scale output of 11111111 with a 2 V input signal. Determine the output word given the following inputs. (Assume that this converter rounds to the nearest output value and is unipolar.)
    A. 100 mV
    B. 10 ◈μV
    C. 0 V
    D. 1.259 V
17. Assume that comparator/logic delays, amplifier settling times, and other factors require 400 ns total in a particular IC fabrication technique. If this technology is used to create AD converters, determine the maximum conversion time for the following 8-bit converters:
    A. Flash

B. Successive approximation

C. Staircase/Ramp type

Design Problems

18. We wish to digitize human voice signals. The maximum input frequency is to be limited to 3 kHz and resolution to better than 0.5% of the maximum input value is required.
    A. Draw a block diagram of the complete system.
    B. Determine the minimum bit requirement.
    C. Determine the minimum sampling rate if the Nyquist rate is set to 25% greater than the theoretical minimum.
    D. Determine the anti-alias filter tuning frequency.
    E. Determine the preferred conversion technique.
    F. Determine which of the ICs presented in the chapter is best suited for this system.
19. We wish to design a system capable of digitizing complex signals with a spectrum ranging from DC to 400 kHz. Accuracy must be at least 0.2% of full scale.
    A. Draw a block diagram of the complete system.
    B. Determine the minimum bit requirement.
    C. Determine the minimum sampling rate if the Nyquist rate is set to 20% greater than the theoretical minimum.
    D. Determine the anti-alias filter tuning frequency.
    E. Determine the preferred conversion technique.
    F. Determine which of the ICs presented in the chapter is best suited for this system.
20. Write a computer algorithm that can be used to "flip" digital data back to front. (i.e., play it backwards).
21. Write a computer algorithm that will determine the maximum peak value of the digital data.
22. Write a computer algorithm that will determine the RMS value of the digital data.

VERSIONING HISTORY

Resource 1: Semiconductor Devices: Theory and Application by James M. Fiore, Mohawk Valley Community College, CC BY-NC-SA.

Resource 2: Operational Amplifiers & Linear Integrated Circuits: Theory and Application by James M. Fiore, Mohawk Valley Community College, CC BY-NC-SA.
https://milneopentextbooks.org/operational-amplifiers/

NSCC Version	Semiconductor Devices: Theory and Application	Operational Amplifiers & Linear Integrated Circuits: Theory and Application/ 3E
Chapter 1	Chapter 10	
Chapter 2	Chapter 11	
Chapter 3	Chapter 12	
Chapter 4	Chapter 13	
Chapter 5	Chapter 14	
Chapter 6	Chapter 15	
Chapter 7		Chapter 1
Chapter 8		Chapter 2
Chapter 9		Chapter 3
Chapter 10		Chapter 4
Chapter 11		Chapter 5
Chapter 12		Chapter 8
Chapter 13		Chapter 9
Chapter 14		Chapter 10
Chapter 15		Chapter 11
Chapter 16		Chapter 12