

Feedback and Stability Theory

Ron Mancini

5.1 Why Study Feedback Theory?

The gain of all op amps decreases as frequency increases, and the decreasing gain results in decreasing accuracy as the ideal op amp assumption ($a \Rightarrow \infty$) breaks down. In most real op amps the open loop gain starts to decrease before 10 Hz, so an understanding of feedback is required to predict the closed loop performance of the op amp. The real world application of op amps is feedback controlled, and depends on op amp open loop gain at a given frequency. A designer must know theory to be able to predict the circuit response regardless of frequency or open loop gain.

Analysis tools have something in common with medicine because they both can be distasteful but necessary. Medicine often tastes bad or has undesirable side effects, and analysis tools involve lots of hard learning work before they can be applied to yield results. Medicine assists the body in fighting an illness; analysis tools assist the brain in learning/designing feedback circuits.

The analysis tools given here are a synopsis of salient points; thus they are detailed enough to get you where you are going without any extras. The references, along with thousands of their counterparts, must be consulted when making an in-depth study of the field. Aspirin, home remedies, and good health practice handle the majority of health problems, and these analysis tools solve the majority of circuit problems.

Ideal op amp circuits can be designed without knowledge of feedback analysis tools, but these circuits are limited to low frequencies. Also, an understanding of feedback analysis tools is required to understand AC effects like ringing and oscillations.

5.2 Block Diagram Math and Manipulations

Electronic systems and circuits are often represented by block diagrams, and block diagrams have a unique algebra and set of transformations[1]. Block diagrams are used because they are a shorthand pictorial representation of the cause-and-effect relationship between the input and output in a real system. They are a convenient method for characterizing the functional relationships between components. It is not necessary to understand the functional details of a block to manipulate a block diagram.

The input impedance of each block is assumed to be infinite to preclude loading. Also, the output impedance of each block is assumed to be zero to enable high fan-out. The systems designer sets the actual impedance levels, but the fan-out assumption is valid because the block designers adhere to the system designer's specifications. All blocks multiply the input times the block quantity (see Figure 5–1) unless otherwise specified within the block. The quantity within the block can be a constant as shown in Figure 5–1(c), or it can be a complex math function involving Laplace transforms. The blocks can perform time-based operations such as differentiation and integration.

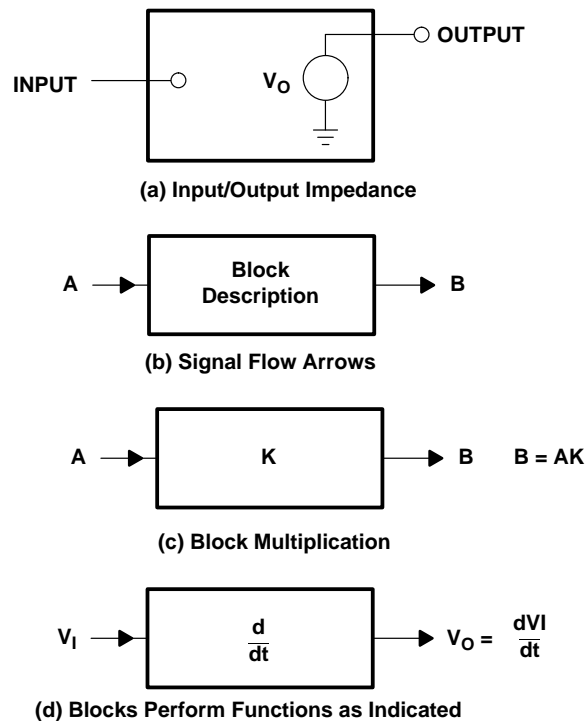


Figure 5–1. Definition of Blocks

Adding and subtracting are done in special blocks called summing points. Figure 5–2 gives several examples of summing points. Summing points can have unlimited inputs, can add or subtract, and can have mixed signs yielding addition and subtraction within a single summing point. Figure 5–3 defines the terms in a typical control system, and Figure 5–4 defines the terms in a typical electronic feedback system. Multiloop feedback systems (Figure 5–5) are intimidating, but they can be reduced to a single loop feedback system, as shown in the figure, by writing equations and solving for V_{OUT}/V_{IN} . An easier method for reducing multiloop feedback systems to single loop feedback systems is to follow the rules and use the transforms given in Figure 5–6.

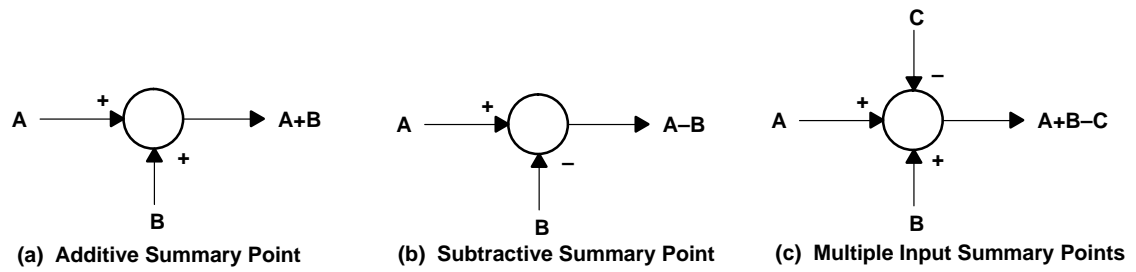


Figure 5-2. Summary Points

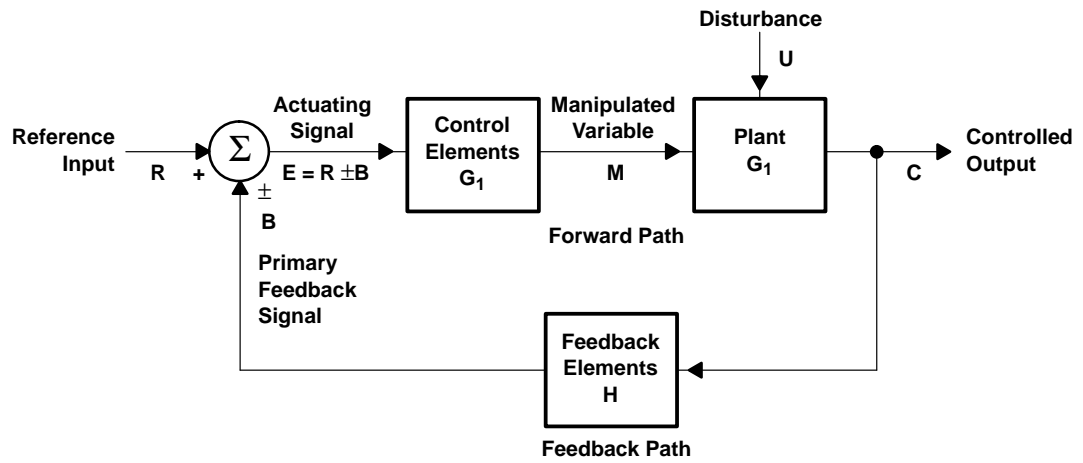


Figure 5-3. Definition of Control System Terms

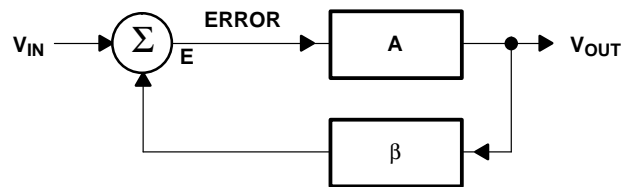


Figure 5-4. Definition of an Electronic Feedback Circuit

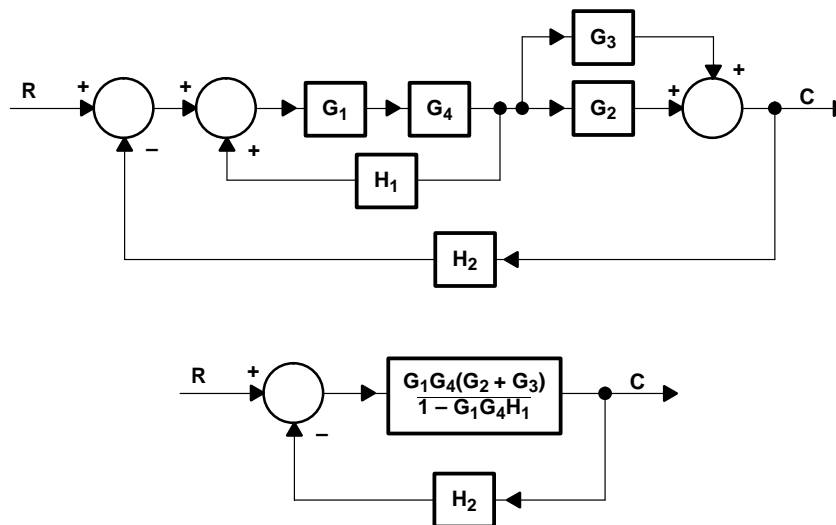


Figure 5–5. Multiloop Feedback System

Block diagram reduction rules:

- Combine cascade blocks.
- Combine parallel blocks.
- Eliminate interior feedback loops.
- Shift summing points to the left.
- Shift takeoff points to the right.
- Repeat until canonical form is obtained.

Figure 5–6 gives the block diagram transforms. The idea is to reduce the diagram to its canonical form because the canonical feedback loop is the simplest form of a feedback loop, and its analysis is well documented. All feedback systems can be reduced to the canonical form, so all feedback systems can be analyzed with the same math. A canonical loop exists for each input to a feedback system; although the stability dynamics are independent of the input, the output results are input dependent. The response of each input of a multiple input feedback system can be analyzed separately and added through superposition.

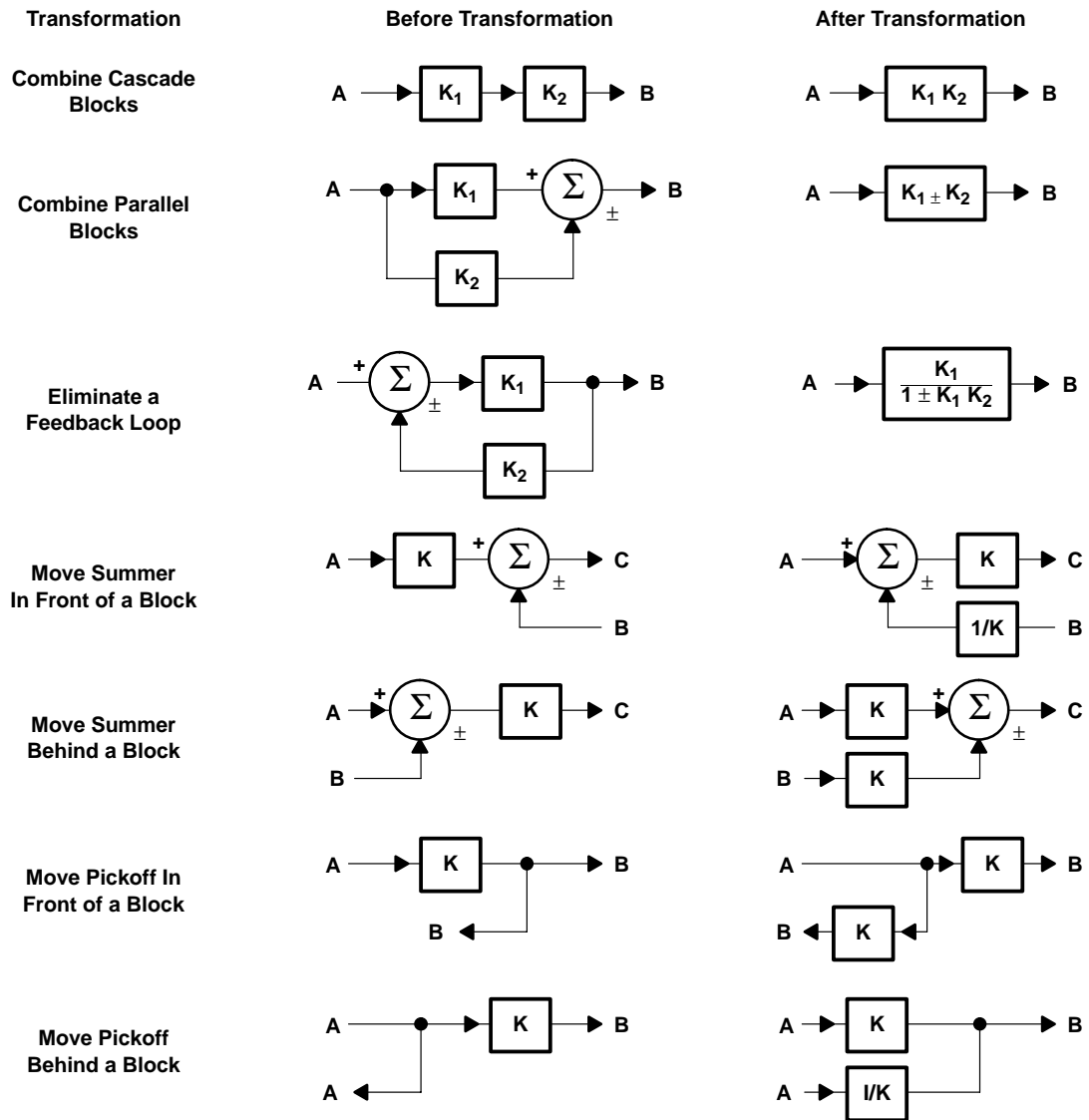


Figure 5–6. Block Diagram Transforms

5.3 Feedback Equation and Stability

Figure 5–7 shows the canonical form of a feedback loop with control system and electronic system terms. The terms make no difference except that they have meaning to the system engineers, but the math does have meaning, and it is identical for both types of terms. The electronic terms and negative feedback sign are used in this analysis, because subsequent chapters deal with electronic applications. The output equation is written in Equation 5–1.

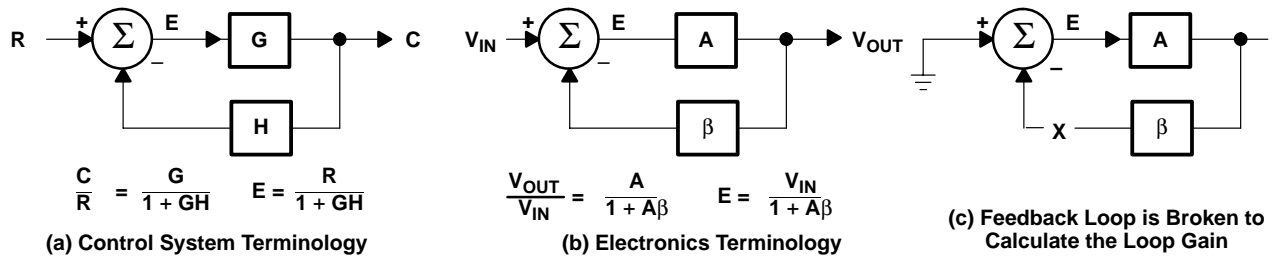


Figure 5–7. Comparison of Control and Electronic Canonical Feedback Systems

$$V_{OUT} = EA \tag{5-1}$$

The error equation is written in Equation 5–2.

$$E = V_{IN} - \beta V_{OUT} \tag{5-2}$$

Combining Equations 5–1 and 5–2 yields Equation 5–3.

$$\frac{V_{OUT}}{A} = V_{IN} - \beta V_{OUT} \tag{5-3}$$

Collecting terms yields Equation 5–4.

$$V_{OUT} \left(\frac{1}{A} + \beta \right) = V_{IN} \tag{5-4}$$

Rearranging terms yields the classic form of the feedback Equation 5–5.

$$\frac{V_{OUT}}{V_{IN}} = \frac{A}{1 + A\beta} \tag{5-5}$$

When the quantity $A\beta$ in Equation 5–5 becomes very large with respect to one, the one can be neglected, and Equation 5–5 reduces to Equation 5–6, which is the ideal feedback equation. Under the conditions that $A\beta \gg 1$, the system gain is determined by the feedback factor β . Stable passive circuit components are used to implement the feedback factor, thus in the ideal situation, the closed loop gain is predictable and stable because β is predictable and stable.

$$\frac{V_{\text{OUT}}}{V_{\text{IN}}} = \frac{1}{\beta} \quad (5-6)$$

The quantity $A\beta$ is so important that it has been given a special name: loop gain. In Figure 5-7, when the voltage inputs are grounded (current inputs are opened) and the loop is broken, the calculated gain is the loop gain, $A\beta$. Now, keep in mind that we are using complex numbers, which have magnitude and direction. When the loop gain approaches minus one, or to express it mathematically $1\angle-180^\circ$, Equation 5-5 approaches $1/0 \Rightarrow \infty$. The circuit output heads for infinity as fast as it can using the equation of a straight line. If the output were not energy limited, the circuit would explode the world, but happily, it is energy limited, so somewhere it comes up against the limit.

Active devices in electronic circuits exhibit nonlinear phenomena when their output approaches a power supply rail, and the nonlinearity reduces the gain to the point where the loop gain no longer equals $1\angle-180^\circ$. Now the circuit can do two things: first it can become stable at the power supply limit, or second, it can reverse direction (because stored charge keeps the output voltage changing) and head for the negative power supply rail.

The first state where the circuit becomes stable at a power supply limit is named lockup; the circuit will remain in the locked up state until power is removed and reapplied. The second state where the circuit bounces between power supply limits is named oscillatory. Remember, the loop gain, $A\beta$, is the sole factor determining stability of the circuit or system. Inputs are grounded or disconnected, so they have no bearing on stability.

Equations 5-1 and 5-2 are combined and rearranged to yield Equation 5-7, which is the system or circuit error equation.

$$E = \frac{V_{\text{IN}}}{1 + A\beta} \quad (5-7)$$

First, notice that the error is proportional to the input signal. This is the expected result because a bigger input signal results in a bigger output signal, and bigger output signals require more drive voltage. As the loop gain increases, the error decreases, thus large loop gains are attractive for minimizing errors.

5.4 Bode Analysis of Feedback Circuits

H. W. Bode developed a quick, accurate, and easy method of analyzing feedback amplifiers, and he published a book about his techniques in 1945.[2] Operational amplifiers had not been developed when Bode published his book, but they fall under the general classification of feedback amplifiers, so they are easily analyzed with Bode techniques. The mathematical manipulations required to analyze a feedback circuit are complicated because they involve multiplication and division. Bode developed the Bode plot, which simplifies the analysis through the use of graphical techniques.

The Bode equations are log equations that take the form $20\text{LOG}(F(t)) = 20\text{LOG}(|F(t)|) +$ phase angle. Terms that are normally multiplied and divided can now be added and subtracted because they are log equations. The addition and subtraction is done graphically, thus easing the calculations and giving the designer a pictorial representation of circuit performance. Equation 5–8 is written for the low pass filter shown in Figure 5–8.

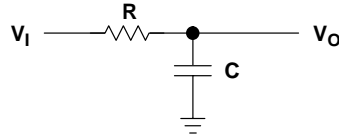


Figure 5–8. Low-Pass Filter

$$\frac{V_{\text{OUT}}}{V_{\text{IN}}} = \frac{\frac{1}{Cs}}{R + \frac{1}{Cs}} = \frac{1}{1 + RCs} = \frac{1}{1 + \tau s} \quad (5-8)$$

Where: $s = j\omega$, $j = \sqrt{-1}$, and $RC = \tau$

The magnitude of this transfer function is $|V_{\text{OUT}}/V_{\text{IN}}| = 1/\sqrt{1^2 + (\tau\omega)^2}$. This magnitude, $|V_{\text{OUT}}/V_{\text{IN}}| \cong 1$ when $\omega = 0.1/\tau$, it equals 0.707 when $\omega = 1/\tau$, and it is approximately = 0.1 when $\omega = 10/\tau$. These points are plotted in Figure 5–9 using straight line approximations. The negative slope is -20 dB/decade or -6 dB/octave. The magnitude curve is plotted as a horizontal line until it intersects the breakpoint where $\omega = 1/\tau$. The negative slope begins at the breakpoint because the magnitude starts decreasing at that point. The gain is equal to 1 or 0 dB at very low frequencies, equal to 0.707 or -3 dB at the break frequency, and it keeps falling with a -20 dB/decade slope for higher frequencies.

The phase shift for the low pass filter or any other transfer function is calculated with the aid of Equation 5–9.

$$\phi = \text{tangent}^{-1}\left(\frac{\text{Real}}{\text{Imaginary}}\right) = -\text{tangent}^{-1}\left(\frac{\omega\tau}{1}\right) \quad (5-9)$$

The phase shift is much harder to approximate because the tangent function is nonlinear. Normally the phase information is only required around the 0 dB intercept point for an active circuit, so the calculations are minimized. The phase is shown in Figure 5–9, and it is approximated by remembering that the tangent of 90° is 1, the tangent of 60° is $\sqrt{3}$, and the tangent of 30° is $\sqrt{3}/3$.

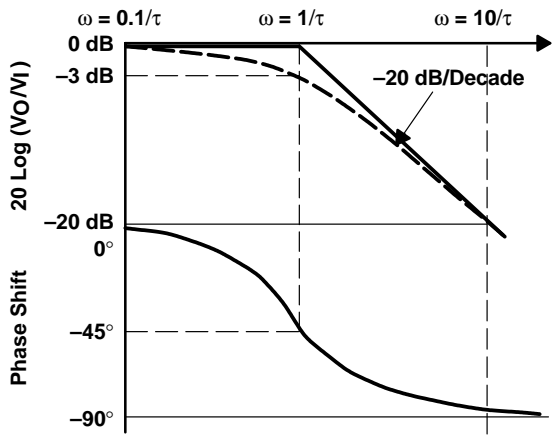


Figure 5–9. Bode Plot of Low-Pass Filter Transfer Function

A breakpoint occurring in the denominator is called a pole, and it slopes down. Conversely, a breakpoint occurring in the numerator is called a zero, and it slopes up. When the transfer function has multiple poles and zeros, each pole or zero is plotted independently, and the individual poles/zeros are added graphically. If multiple poles, zeros, or a pole/zero combination have the same breakpoint, they are plotted on top of each other. Multiple poles or zeros cause the slope to change by multiples of 20 dB/decade.

An example of a transfer function with multiple poles and zeros is a band reject filter (see Figure 5–10). The transfer function of the band reject filter is given in Equation 5–10.

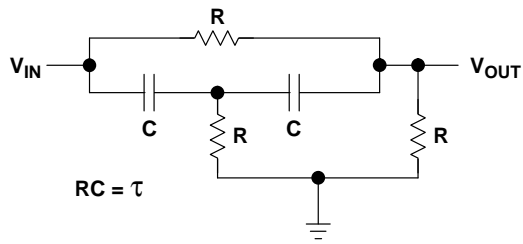


Figure 5–10. Band Reject Filter

$$G = \frac{V_{OUT}}{V_{IN}} = \frac{(1 + \tau s)(1 + \tau s)}{2 \left(1 + \frac{\tau s}{0.44}\right) \left(1 + \frac{\tau s}{4.56}\right)} \tag{5-10}$$

The pole zero plot for each individual pole and zero is shown in Figure 5–11, and the combined pole zero plot is shown in Figure 5–12.

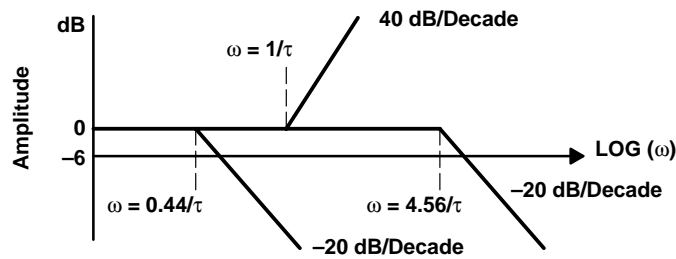


Figure 5–11. Individual Pole Zero Plot of Band Reject Filter

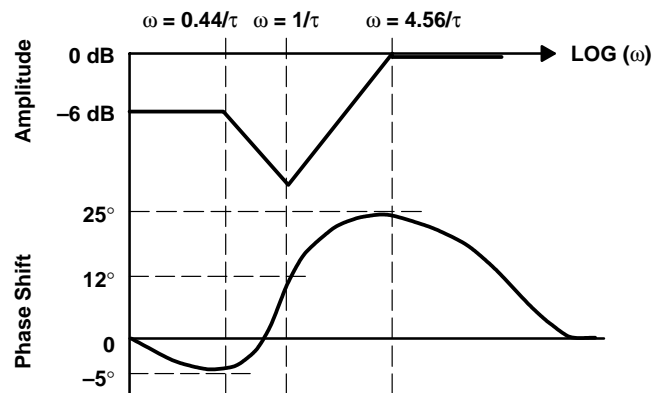


Figure 5–12. Combined Pole Zero Plot of Band Reject Filter

The individual pole zero plots show the dc gain of $1/2$ plotting as a straight line from the -6 dB intercept. The two zeros occur at the same break frequency, thus they add to a 40 -dB/decade slope. The two poles are plotted at their breakpoints of $\omega = 0.44/\tau$ and $\omega = 4.56/\tau$. The combined amplitude plot intercepts the amplitude axis at -6 dB because of the dc gain, and then breaks down at the first pole. When the amplitude function gets to the double zero, the first zero cancels out the first pole, and the second zero breaks up. The upward slope continues until the second pole cancels out the second zero, and the amplitude is flat from that point out in frequency.

When the separation between all the poles and zeros is great, a decade or more in frequency, it is easy to draw the Bode plot. As the poles and zeros get closer together, the plot gets harder to make. The phase is especially hard to plot because of the tangent function, but picking a few salient points and sketching them in first gets a pretty good approximation.[3] The Bode plot enables the designer to get a good idea of pole zero placement, and it is valuable for fast evaluation of possible compensation techniques. When the situation gets critical, accurate calculations must be made and plotted to get an accurate result.

Consider Equation 5–11.

$$\frac{V_{OUT}}{V_{IN}} = \frac{A}{1 + A\beta} \quad (5-11)$$

Taking the log of Equation 5–11 yields Equation 5–12.

$$20\text{Log}\left(\frac{V_{OUT}}{V_{IN}}\right) = 20\text{Log}(A) - 20\text{Log}(1 + A\beta) \quad (5-12)$$

If A and β do not contain any poles or zeros there will be no break points. Then the Bode plot of Equation 5–12 looks like that shown in Figure 5–13, and because there are no poles to contribute negative phase shift, the circuit cannot oscillate.

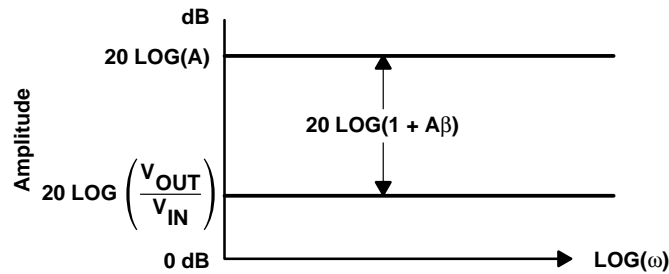


Figure 5–13. When No Pole Exists in Equation (5–12)

All real amplifiers have many poles, but they are normally internally compensated so that they appear to have a single pole. Such an amplifier would have an equation similar to that given in Equation 5–13.

$$A = \frac{a}{1 + j\frac{\omega}{\omega_a}} \quad (5-13)$$

The plot for the single pole amplifier is shown in Figure 5–14.

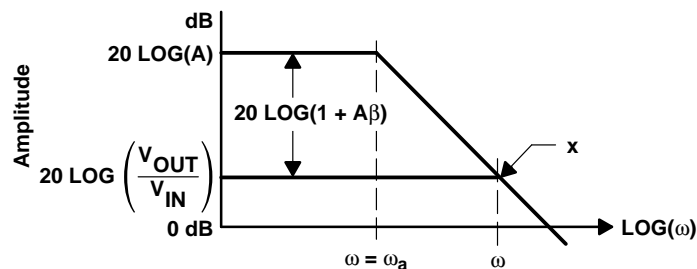


Figure 5–14. When Equation 5–12 has a Single Pole

The amplifier gain, A , intercepts the amplitude axis at $20\text{Log}(A)$, and it breaks down at a slope of -20 dB/decade at $\omega = \omega_a$. The negative slope continues for all frequencies greater than the breakpoint, $\omega = \omega_a$. The closed loop circuit gain intercepts the amplitude axis at $20\text{Log}(V_{\text{OUT}}/V_{\text{IN}})$, and because β does not have any poles or zeros, it is constant until its projection intersects the amplifier gain at point X. After intersection with the amplifier gain curve, the closed loop gain follows the amplifier gain because the amplifier is the controlling factor.

Actually, the closed loop gain starts to roll off earlier, and it is down 3 dB at point X. At point X the difference between the closed loop gain and the amplifier gain is -3 dB, thus according to Equation 5–12 the term $-20\text{Log}(1+A\beta) = -3$ dB. The magnitude of 3 dB is $\sqrt{2}$, hence $\sqrt{1 + (A\beta)^2} = \sqrt{2}$, and elimination of the radicals shows that $A\beta = 1$. There is a method [4] of relating phase shift and stability to the slope of the closed loop gain curves, but only the Bode method is covered here. An excellent discussion of poles, zeros, and their interaction is given by M. E Van Valkenberg,[5] and he also includes some excellent prose to liven the discussion.

5.5 Loop Gain Plots are the Key to Understanding Stability

Stability is determined by the loop gain, and when $A\beta = -1 = |1| \angle -180^\circ$ instability or oscillation occurs. If the magnitude of the gain exceeds one, it is usually reduced to one by circuit nonlinearities, so oscillation generally results for situations where the gain magnitude exceeds one.

Consider oscillator design, which depends on nonlinearities to decrease the gain magnitude; if the engineer designed for a gain magnitude of one at nominal circuit conditions, the gain magnitude would fall below one under worst case circuit conditions causing oscillation to cease. Thus, the prudent engineer designs for a gain magnitude of one under worst case conditions knowing that the gain magnitude is much more than one under optimistic conditions. The prudent engineer depends on circuit nonlinearities to reduce the gain magnitude to the appropriate value, but this same engineer pays a price of poorer distortion performance. Sometimes a design compromise is reached by putting a nonlinear component, such as a lamp, in the feedback loop to control the gain without introducing distortion.

Some high gain control systems always have a gain magnitude greater than one, but they avoid oscillation by manipulating the phase shift. The amplifier designer who pushes the amplifier for superior frequency performance has to be careful not to let the loop gain phase shift accumulate to 180° . Problems with overshoot and ringing pop up before the loop gain reaches 180° phase shift, thus the amplifier designer must keep a close eye on loop dynamics. Ringing and overshoot are handled in the next section, so preventing oscillation is emphasized in this section. Equation 5–14 has the form of many loop gain transfer functions or circuits, so it is analyzed in detail.

$$(A)\beta = \frac{(K)}{(1 + \tau_1(s))(1 + \tau_2(s))} \quad (5-14)$$

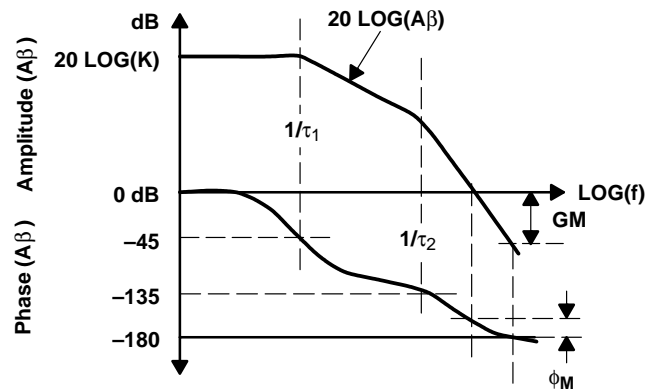


Figure 5–15. Magnitude and Phase Plot of Equation 5–14

The quantity, K , is the dc gain, and it plots as a straight line with an intercept of $20\text{Log}(K)$. The Bode plot of Equation 5–14 is shown in Figure 5–15. The two break points, $\omega = \omega_1 = 1/\tau_1$ and $\omega = \omega_2 = 1/\tau_2$, are plotted in the Bode plot. Each breakpoint adds -20 dB/decade slope to the plot, and 45° phase shift accumulates at each breakpoint. This transfer function is referred to as a two slope because of the two breakpoints. The slope of the curve when it crosses the 0 dB intercept indicates phase shift and the ability to oscillate. Notice that a one slope can only accumulate 90° phase shift, so when a transfer function passes through 0 dB with a one slope, it cannot oscillate. Furthermore, a two-slope system can accumulate 180° phase shift, therefore a transfer function with a two or greater slope is capable of oscillation.

A one slope crossing the 0 dB intercept is stable, whereas a two or greater slope crossing the 0 dB intercept may be stable or unstable depending upon the accumulated phase shift. Figure 5–15 defines two stability terms; the phase margin, ϕ_M , and the gain margin, G_M . Of these two terms the phase margin is much more important because phase shift is critical for stability. Phase margin is a measure of the difference in the actual phase shift and the theoretical 180° required for oscillation, and the phase margin measurement or calculation is made at the 0 dB crossover point. The gain margin is measured or calculated at the 180° phase crossover point. Phase margin is expressed mathematically in Equation 5–15.

$$\phi_M = 180 - \text{tangent}^{-1}(A\beta) \quad (5-15)$$

The phase margin in Figure 5–15 is very small, 20° , so it is hard to measure or predict from the Bode plot. A designer probably doesn't want a 20° phase margin because the system overshoots and rings badly, but this case points out the need to calculate small phase margins carefully. The circuit is stable, and it does not oscillate because the phase margin is positive. Also, the circuit with the smallest phase margin has the highest frequency response and bandwidth.

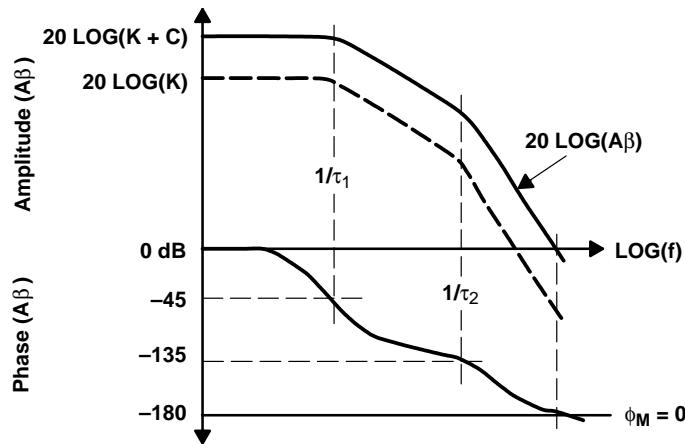


Figure 5–16. Magnitude and Phase Plot of the Loop Gain Increased to (K+C)

Increasing the loop gain to (K+C) as shown in Figure 5–16 shifts the magnitude plot up. If the pole locations are kept constant, the phase margin reduces to zero as shown, and the circuit will oscillate. The circuit is not good for much in this condition because production tolerances and worst case conditions ensure that the circuit will oscillate when you want it to amplify, and vice versa.

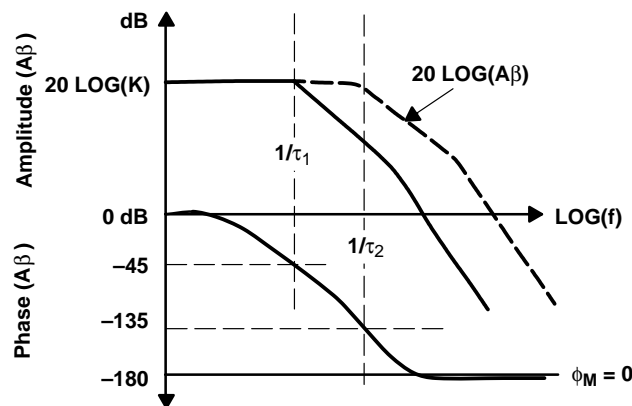


Figure 5–17. Magnitude and Phase Plot of the Loop Gain With Pole Spacing Reduced

The circuit poles are spaced closer in Figure 5–17, and this results in a faster accumulation of phase shift. The phase margin is zero because the loop gain phase shift reaches 180° before the magnitude passes through 0 dB. This circuit oscillates, but it is not a very stable oscillator because the transition to 180° phase shift is very slow. Stable oscillators have a very sharp transition through 180°.

When the closed loop gain is increased the feedback factor, β , is decreased because $V_{OUT}/V_{IN} = 1/\beta$ for the ideal case. This in turn decreases the loop gain, $A\beta$, thus the stability increases. In other words, increasing the closed loop gain makes the circuit more stable. Stability is not important except to oscillator designers because overshoot and ringing become intolerable to linear amplifiers long before oscillation occurs. The overshoot and ringing situation is investigated next.

5.6 The Second Order Equation and Ringing/Overshoot Predictions

The second order equation is a common approximation used for feedback system analysis because it describes a two-pole circuit, which is the most common approximation used. All real circuits are more complex than two poles, but except for a small fraction, they can be represented by a two-pole equivalent. The second order equation is extensively described in electronic and control literature [6].

$$(1 + A\beta) = 1 + \frac{K}{(1 + \tau_1 s)(1 + \tau_2 s)} \quad (5-16)$$

After algebraic manipulation Equation 5–16 is presented in the form of Equation 5–17.

$$s^2 + s \frac{\tau_1 + \tau_2}{\tau_1 \tau_2} + \frac{1 + K}{\tau_1 \tau_2} = 0 \quad (5-17)$$

Equation 5–17 is compared to the second order control Equation 5–18, and the damping ratio, ζ , and natural frequency, ω_N are obtained through like term comparisons.

$$s^2 + 2\zeta\omega_N s + \omega_N^2 \quad (5-18)$$

Comparing these equations yields formulas for the phase margin and per cent overshoot as a function of damping ratio.

$$\omega_N = \sqrt{\frac{1 + K}{\tau_1 \tau_2}} \quad (5-19)$$

$$\zeta = \frac{\tau_1 + \tau_2}{2\omega_N \tau_1 \tau_2} \quad (5-20)$$

When the two poles are well separated, Equation 5–21 is valid.

$$\phi_M = \tan^{-1}(2\xi) \quad (5-21)$$

The salient equations are plotted in Figure 5–18, which enables a designer to determine the phase margin and overshoot when the gain and pole locations are known.

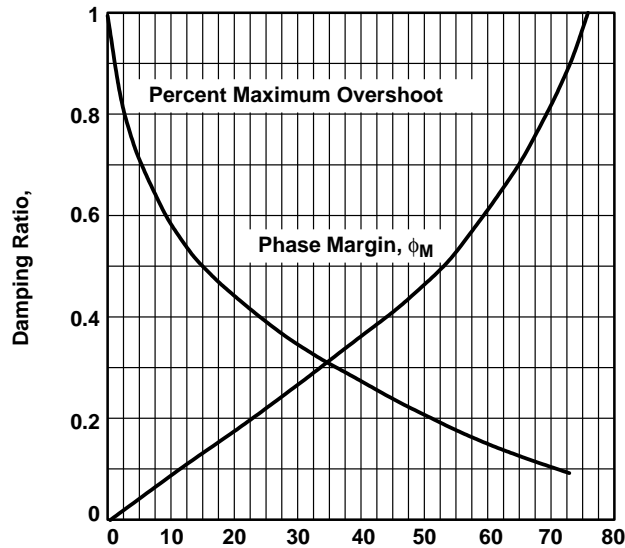


Figure 5–18. Phase Margin and Overshoot vs Damping Ratio

Enter Figure 5–18 at the calculated damping ratio, say 0.4, and read the overshoot at 25% and the phase margin at 42°. If a designer had a circuit specification of 5% maximum overshoot, then the damping ratio must be 0.78 with a phase margin of 62°.

5.7 References

1. DiStefano, Stubberud, and Williams, *Theory and Problems of Feedback and Control Systems, Schaum's Outline Series*, McGraw Hill Book Company, 1967
2. Bode, H. W., *Network Analysis And Feedback Amplifier Design*, D. Van Nostrand, Inc., 1945
3. Frederickson, Thomas, *Intuitive Operational Amplifiers*, McGraw Hill Book Company, 1988
4. Bower, J. L. and Schultheis, P. M., *Introduction To The Design Of Servomechanisms*, Wiley, 1961
5. Van Valkenberg, M. E., *Network Analysis*, Prentice-Hall, 1964
6. Del Toro, V., and Parker, S., *Principles of Control Systems Engineering*, McGraw–Hill, 1960.