Chapter

12

Feedback-Loop Stabilization

12.1 Introduction

Before going into the details of stabilizing a feedback loop, it is of interest to consider in a semiquantitative way, why a feedback loop may oscillate.

Consider the negative-feedback loop for a typical forward converter in Fig. 12.1. The essential error-amplifier and PWM functions are contained in all pulse-width-modulating chip. The chip also provide many other functions, but for understanding the stability problem, only the error amplifier and pulse-width modulator need be considered.

For slow or DC variations of the output voltage V_o, the loop is, of course, stable. A small, slow variation of V_o due to either line input or load changes will be sensed by the inverting input of error amplifier EA via the sampling network $R1$, $R2$ and compared to a reference voltage at the noninverting EA input. This will cause a small change in the DC voltage level V_{ea} at the EA output and at the A input to the pulse-width-modulator PWM.

The PWM, as described heretofore, compares that DC voltage level to a roughly 0- to 3-V triangle V_t at its B input. It generates a rectangular pulse whose width t_{on} is equal to the time from the start of the triangle t_0 until t_1, the time the triangle crosses the DC voltage level at the B input of the PWM. That pulse fixes the on time of the output transistor of the chip and should also fix the on time of the power transistor.

Thus a slow increase (e.g.) in V_{dc} causes a slow increase in V_y and hence a slow increase in V_o since $V_o \cong V_y t_{on}/T$. The increase in V_o causes an increase in V_s and hence a decrease in V_{ea}. Since t_{on} is the time from the start of the triangle to t_1, this causes a decrease in t_{on} and restores V_o to its original value. Similarly, of course, a decrease in V_{dc} causes an increase in t_{on} to maintain V_o constant.

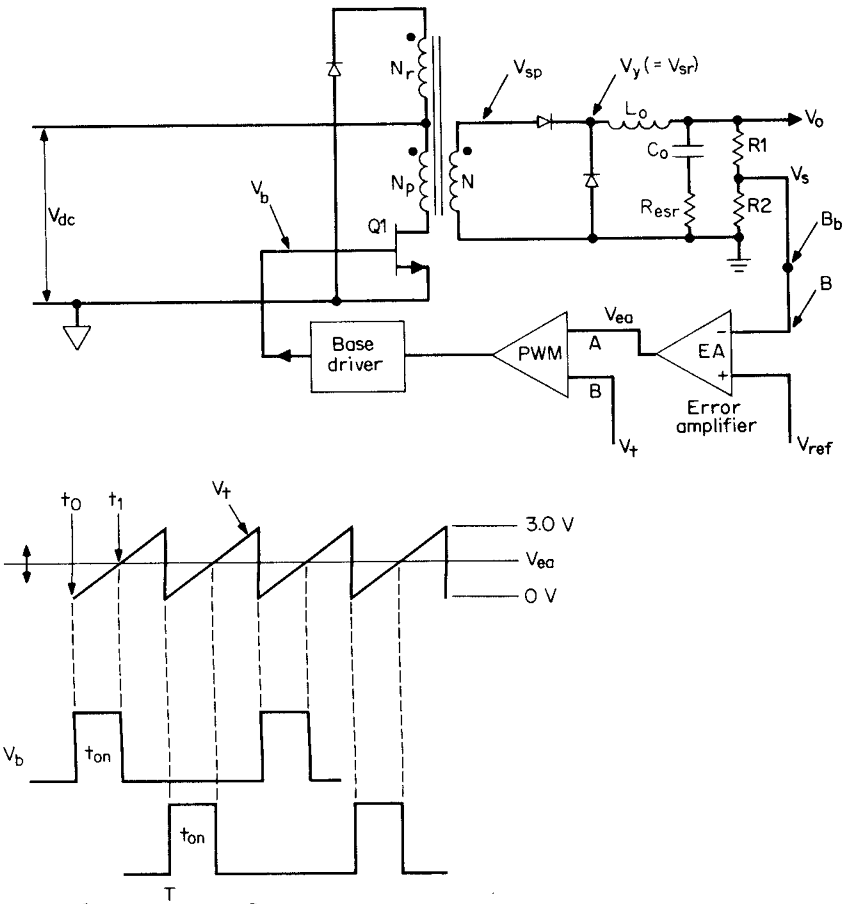


Figure 12.1 A closed feedback loop in a typical forward converter.

Drive to the power transistor may be taken from either the emitter or the collector of the chip's output transistor via a current-amplifying base driver. But from whichever point—emitter or collector—output is taken, it must be ensured that polarities are such that an increase in V_o causes a decrease in t_{on}.

Note that most PWM chips have their output transistors on for the time from t_0 to t_1. With such chips, V_s is fed to the inverting EA input and for an NPN power transistor, its base (or gate if a MOSFET) is driven from the emitter of the chip's output transistor.

In some PWM chips (TL494 family), however, the output transistor is on from the time the triangle at the V_t input to the PWM crosses the V_{ea} level (t_1) until the end of the triangle at t_2. With such chips, if the

NPN power transistor were to be turned on when the chip output transistor was on (drive from the chip transistor's emitter), that would cause the power transistor on time to increase as V_{dc} increased. This, of course, would be positive rather than negative feedback.

Thus, with chips of the TL494 family, V_s is fed to the noninverting input to the EA. This causes the output transistor on time to decrease as V_o increases and permits the power transistor to be driven from the chip transistor's emitter.

The circuit in Fig. 12.1 thus provides negative feedback and a stable circuit at low frequencies. But within the loop, there exist low-level noise voltages or possible voltage transients which have a continuous spectrum of sinusoidal Fourier components. All these Fourier components suffer gain changes and phase shifts in the L_o, C_o output filter, the error amplifier, and the PWM from V_{ea} to V_{sr}. At one of these Fourier components, the gain and phase shifts can result in positive rather than negative feedback and thereby result in oscillation as described below.

12.2 Mechanism of Loop Oscillation

Consider the forward converter feedback loop of Fig. 12.1. Assume for a moment that the loop is broken open at point *B*,the inverting input to the error amplifier. At any of the Fourier components of the noise, there is gain and phase shift from *B* to V_{ea}, from V_{ea} to the average voltage at V_{sr}, and from the average voltage at V_{sr} through the L_o, C_o filter around back to B_b (just before the loop break).

Now assume that a signal of some frequency f_1 is injected into the loop at *B* and comes back around as an echo at B_b. The echo is modified in phase and gain by all the previously mentioned elements in the loop. If the modified echo has returned exactly in phase with and is equal in amplitude to the signal which started the echo, if the loop is now closed (B_b closed to *B*) and the injected signal is removed, the circuit will continue to oscillate at the frequency f_1. The initial signal which starts the echo and maintains the oscillation is the f_1 Fourier component in the noise spectrum.

12.2.1 Gain criterion for a stable circuit

Thus the first criterion for a stable loop is that at the frequency where the total open-loop gain is unity (the *crossover frequency*), the total open-loop phase shift of all elements involved must be less than 360°. The amount by which the total phase shift is less than 360° (at the frequency where the total open-loop gain is unity) is called the *phase margin.*

To ensure a stable loop under worst-case tolerances of the associated components, the usual practice is to design for at least a 35° to 45° phase margin with nominal components.

12.2.2 Gain slope criteria for a stable circuit

At this point, a universally used jargon expression describing the gain slope is introduced. Gain versus frequency is usually plotted in decibels (dB) on semilog paper as in Fig. 12.2. If the scales are such that a linear distance of 20 dB (numerical gain of 10) is equal to the linear distance of a factor of 10 in frequency, lines representing gain variations of ±20 dB/decade have slopes of ±1. Circuit configurations having a gain variation of ±20 dB per decade are thus described as having " ± 1 gain slopes."

An elementary circuit having a gain slope of −1 (beyond the frequency $f_p = 1/2\pi R1C1$) between output and input is the *RC* integrator of Fig. 12.2*a*. The *RC* differentiator of Fig. 12.2*b* has a +1 gain slope (below the frequency $f_z = 1/2\pi R2C2$) or a gain variation of +20 dB/decade between output and input. Such circuits have only 20 dB/decade gain variations because as frequency increases or decreases by a factor of 10, the capacitor impedance decreases or increases by a factor of 10 but the resistor impedance remains constant.

A circuit which has a −2 or −40 dB/decade gain slope (beyond the frequency $F_o = 1/2\pi\sqrt{L_oC_o}$) is the output *LC* filter (Fig. 12.2*c*), which has no resistance (ESR) in its output capacitor. This, of course, is because as frequency increases by a factor of 10, the inductor impedance increases and the capacitor impedance decreases by a factor of 10.

Now gain and phase shift versus frequency for an L_oC_o filter are plotted in Fig. 12.3*a* and 12.3*b* for various values of output resistance R_o^2. The gain curves are normalized for various ratios of $k_1 = f/F_o$ where $F_o = 1/2\pi\sqrt{L_oC_o}$ and for various ratios $k_2 = R_o/\sqrt{L_o/C_o}$.

Figure 12.3*a* shows that whatever the value of k_2, all gain curves, beyond the so-called corner frequency of $F_o = 1/2\pi\sqrt{L_oC_o}$ asymptotically approach a slope of −2 (−40 dB/decade). The circuit for $k_2 = 1.0$ is referred to as the *critically damped* circuit. The critically damped circuit has a very small resonant "bump" in gain and at the corner frequency F_o, starts immediately falling at a −2 slope.

For k_2 greater than 1 the circuit is described as *underdamped*. It is seen that underdamped *LC* filters can have a very large resonant bump in gain at F_o.

Circuits of k_2 less than 1.0 are *overdamped*. It is seen in Fig. 12.3*a* that overdamped *LC* filters also asymptotically approach a gain slope of −2. But for a heavily overdamped ($k_2 = 0.1$) filter, the frequency at

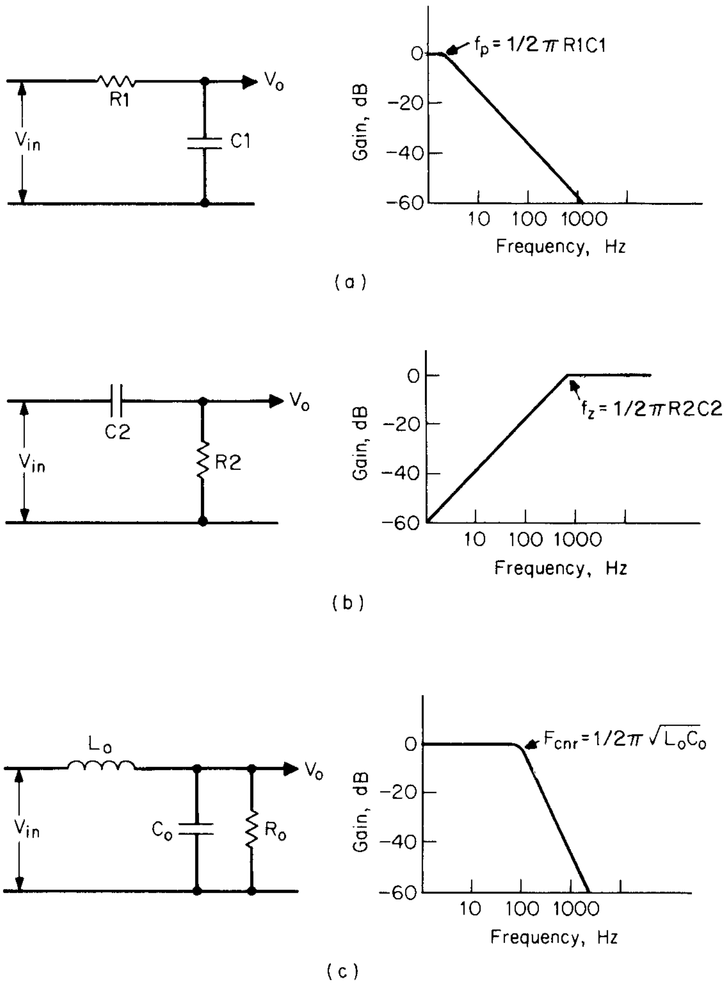
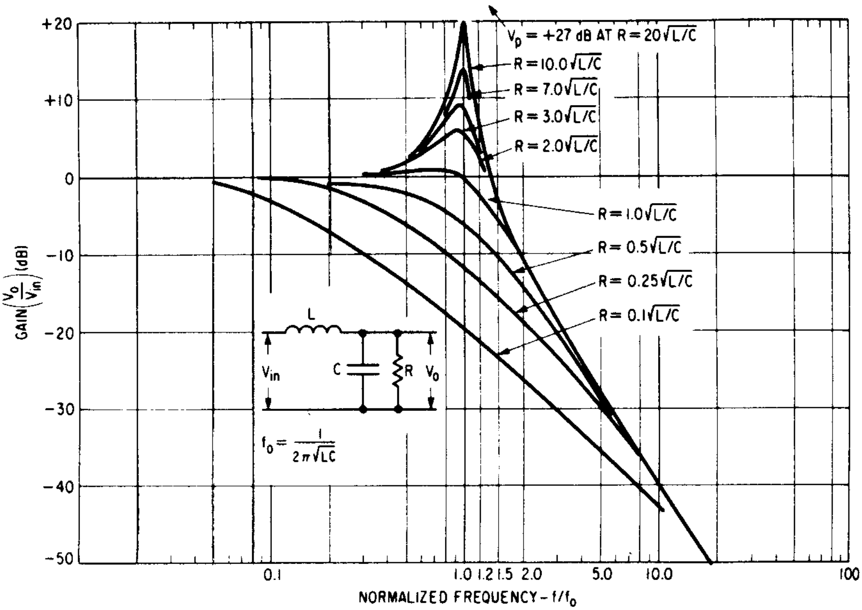
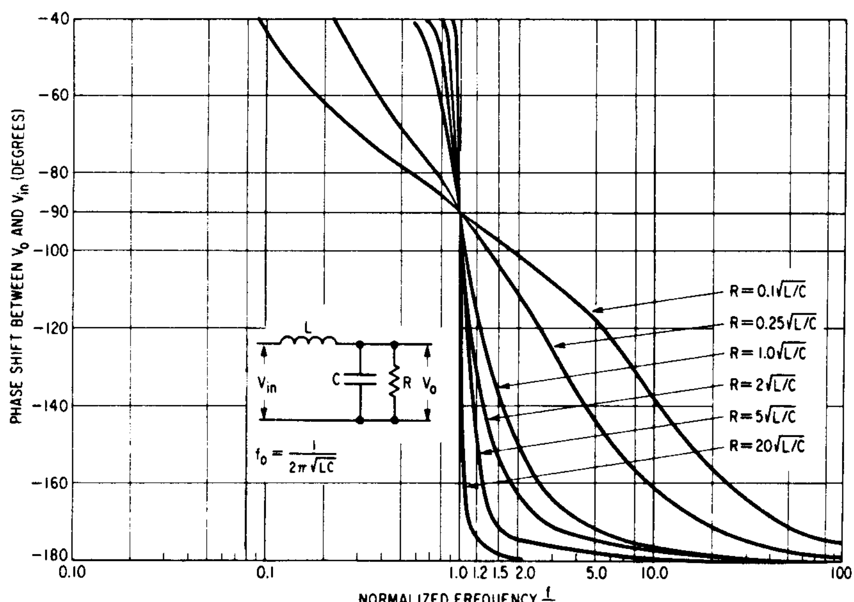


Figure 12.2 (*a*) An *RC* integrator has a gain dV_o/dV_{in} of −20 dB/decade beyond $F_p = 1/2\pi R1C1$. If the scales are such that 20 dB is the same linear distance as 1 decade in frequency, a gain slope of −20 dB/decade has a −1 slope. Such a circuit is referred to as a −1 *slope* circuit. (*b*) An *RC* differentiator has a gain of +20 dB/decade. At $F_z = 1/2\pi R2C2$, where $X_{C2} = R_2$, gain asymptotically approaches 0 dB. If scales are such that 20 dB is the same linear distance as 1 decade in frequency, a gain slope of +20 dB/decade has a +1 slope. Such a circuit is referred to as a +1 *slope* circuit. (*c*) An *LC* filter has a gain (dV_o/dV_{in}) of unity (0 dB) up to its corner frequency of $F_{cnr} = 1/2\pi\sqrt{L_oC_o}$ when critically damped ($R_o = \sqrt{L_o/C_o}$). Beyond F_{cnr} it commences falling at a rate of −40 dB/decade. This is so because for every decade increase in frequency, X_L increased and X_c decreases in impedance by a factor of 10. If scales are such that 20 dB is the same linear distance as 1 decade in frequency, a gain slope of −40 dB/decade has a −2 slope. Such a circuit is referred to as a −2 *slope* circuit.



(a)



(b)

Figure 12.3 (*a*) Gain versus frequency for switching regulator *LC* filter. (*b*) Phase shift versus frequency for switching regulator *LC* filter. (*Courtesy Switchtronix Press*).

which the gain slope has come close to -2 is about 20 times the corner frequency F_o.

Figure 12.3*b* shows phase shift versus normalized frequency (f/F_o) again for various ratios of $k_2 = R_o/\sqrt{L_o/C_o}$. It is seen for any value of k_2, that the phase shift between output and input is 90° at the corner frequency $(F_o = 1/2\pi\sqrt{L_oC_o})$. And for highly underdamped filters (R_o greater than $\sim 5\sqrt{L_o/C_o}$), phase shift varies very rapidly with frequency. The shift is already 170° at a frequency of $1.5F_o$ for $R_o = 5\sqrt{L_o/C_o}$.

In contrast, a circuit with a -1 gain slope can never yield more than a 90° phase shift, and its rate of change of phase shift with frequency is far slower than that of a circuit -2 gain slope as exemplified in Fig. 12.3*b*.

This leads to the second criterion for a stable circuit. The first criterion was that the total phase shift at the crossover frequency (frequency where total open-loop gain is unity or 0 dB) should be short of 360° by the "phase margin," which is usually taken as at least 45°.

This second criterion for a stable circuit is that to prevent rapid changes of phase shift with frequency characteristic of a circuit with a -2 gain slope, the slope of the open-loop gain–frequency curve of the entire circuit (arithmetic sum in decibels of all the gain elements involved) as it passes through crossover frequency should be -1. This is shown in Fig. 12.4.

It is not an absolute requirement that the total open-loop gain curve must come through crossover at a -1 gain slope. But it does provide insurance that if any phase-shift elements have been overlooked, the small phase shift and relatively slow phase-shift–frequency curve characteristic of a -1 gain slope element will still preserve an adequate phase margin.

The third criterion for a stable loop is to provide the desired phase margin, which will be set at 45° herein (Fig. 12.4).

To satisfy all three criteria, it is necessary to know how to calculate gains and phase shifts of all the elements in Fig. 12.1. This is shown below.

12.2.3 Gain characteristic of *LC* output filter with and without equivalent series resistance (ESR) in output capacitor

Aside from the flyback (which has an output capacitor filter only), all topologies discussed herein have an output *LC* filter.[1] The gain-versus-frequency characteristic of this output *LC* filter is of fundamental importance. It must be calculated first as it determines how the gain and phase shift-versus-frequency characteristics of the error amplifier must be shaped to satisfy the three criteria for a stable loop.

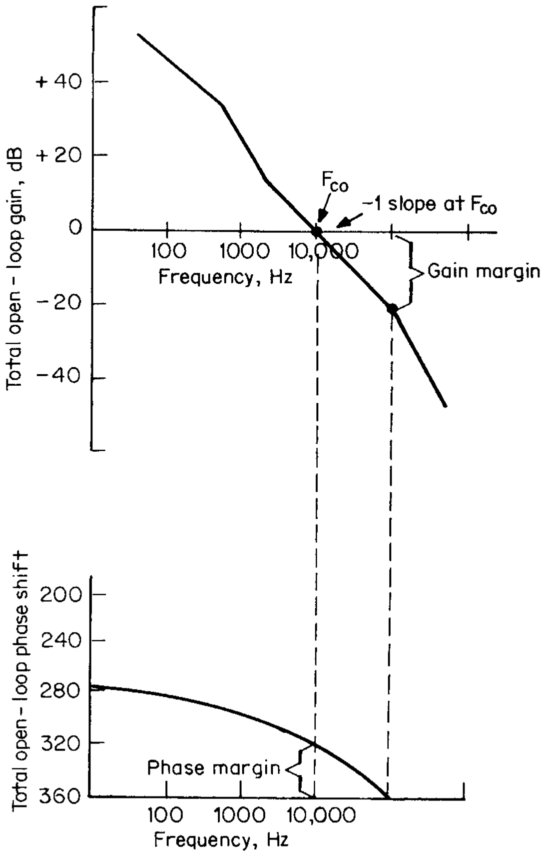


Figure 12.4 Total open-loop gain and phase shift. The frequency at which the total open-loop gain is 0 dB (F_{co}) is forced to be 0 dB is usually chosen one-fourth and one-fifth the switching frequency. For loop to be stable, total open-loop phase shift at F_{co} should be short of 360° by as large a value as possible. The amount by which it is short of 360° is the *phase margin*. A usual phase margin to strive for is 45°. A second criterion for a stable loop is that the total open-loop gain pass through F_{co} at a −1 slope.

The gain characteristic of an output *LC* filter with various output load resistances is shown in Fig. 12.3*a*. This curve assumes that the output capacitor has zero equivalent series resistance (ESR). For the purpose of this discussion, it is sufficiently accurate to assume that the filter is critically damped, that is, $R_o = 1.0\sqrt{L_o/C_o}$. If the circuit is made stable for the gain curve corresponding to $R_o = 1.0\sqrt{L_o/C_o}$, it will be stable at other loads. Nevertheless, the circuit merits examination for light loads ($R_o \gg 1.0\sqrt{L_o/C_o}$) because of the resonant bump in gain at the LC corner frequency $F_o = 1/2\pi\sqrt{L_oC_o}$. This will be considered below.

Thus the gain characteristic of the output LC filter with zero ESR will be drawn as curve 12345 in Fig. 12.5a. There it is seen that the gain is 0 dB (numerical gain of 1) at DC and low frequencies up to the

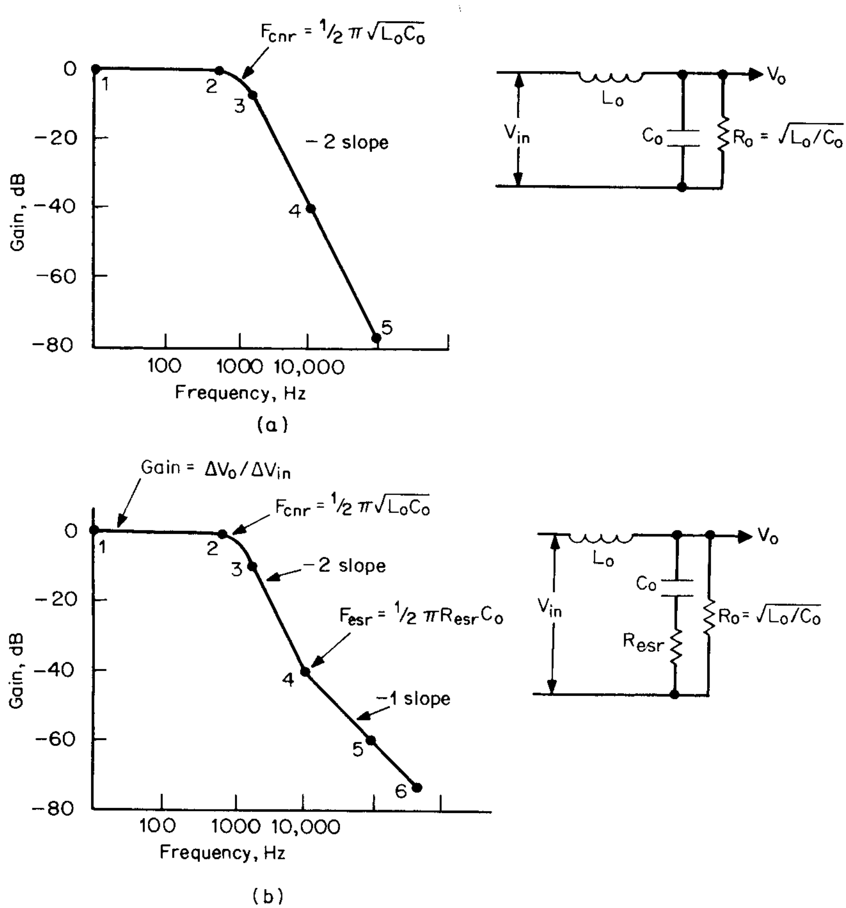


Figure 12.5 (a) Gain versus frequency for a critically damped LC filter in which the output capacitor has zero ESR. (b) Gain versus frequency for a critically damped LC filter in which the output capacitor has an equivalent series resistor (R_{esr}). When C_o has an ESR, the gain slope still breaks from horizontal to a −2 slope at F_{cnr}. But at a frequency $F_{esr} = 1/2\pi\ (R_{esr}C_o)$, it breaks into a −1 slope. This is because at F_{esr}, $X_{co} = R_{esr}$, the impedance of C_o becomes increasingly small compared to R_{esr}. The circuit is now an LR rather than an LC circuit. An LR circuit falls at a −1 slope because as frequency increases, the impedance of the series L increases, but that of the shunt R remains constant.

corner frequency $F_o = 1/2\pi\sqrt{L_oC_o}$. At DC and frequencies less than F_o, the impedance of C_o is much greater than that of L_o and the output-input gain is unity.

Beyond F_o, the impedance of C_o decreases and that of L_o increases at the rate of 20 dB/decade, making the gain slope fall at the rate of −40 dB/decade or at a −2 slope. Of course, the transition to a −2 slope at F_o is not as abrupt as shown. The actual gain curve leaves 0 dB smoothly just before F_o and asymptotically approaches the −2 slope shortly after F_o. But for the purposes of this discussion, gain will be shown with a relatively abrupt transition as curve 12345 in Fig. 12.5*a*.

Most filter capacitor types have an internal resistance R_{esr} in series with their output leads as shown in Fig. 12.5*b*. This modifies the gain characteristic between the output and input terminals in a characteristic way.

Beyond F_o, in the lower frequency range where the impedance of C_o is much greater than R_{esr}, looking down to ground from V_o, the only effective impedance is that of C_o. In this frequency range, the gain still falls at a −2 slope. At higher frequencies, where the impedance of C_o is less than R_{esr}, the effective impedance looking down from V_o to ground is that of R_{esr} alone. Hence in that frequency range, the circuit is an *LR* rather than an *LC* circuit. In that frequency range, the impedance of L_o increases at the rate of 20 dB/decade but that of R_{esr} remains constant. Thus in that frequency range, gain falls at a −1 slope.

The break from a −2 to a −1 gain slope occurs at the frequency $F_{esr} = 1/2\pi R_{esr}C_o$, where the impedance of C_o is equal to R_{esr}. This is shown as G_{lc} in curve 123456 in Fig. 12.5*b*. The break in slope from −2 to −1 is, of course, asymptotic, but it is sufficiently accurate to assume it to be abrupt as shown.

12.2.4 Pulse-width-modulator gain

In Fig. 12.1, the gain from the error-amplifier output to the average voltage at V_{sr} (input end of the output inductor) is the PWM gain and is designated as G_{pwm}.

It may be puzzling how this can be referred to as a *voltage gain*. For at V_{ea}, there are DC voltage level variations proportional to the error-amplifier input at point *B*, and at V_{sr}, there are fixed-amplitude pulses of adjustable width.

The significance and magnitude of this gain can be seen as follows. In Fig. 12.1, the PWM compares the DC voltage level from V_{ea} to a 3-V triangle at V_t. In all PWM chips which produce two 180° out-of-phase adjustable-width pulses (for driving push-pulls, half or full bridges) these pulses occur once per triangle and have a maximum on or high time of a half period. After the PWM, the pulses are binarycounted and alternately routed to two separate output terminals (see Fig. 5.2*a*). In a forward converter, only one of these outputs is used.

Now (Fig. 12.1*b*), when V_{ea} is at the bottom of the 3-V triangle, on time or pulse width at V_{sr} is zero. The average voltage V_{av} at V_{sr} is then zero as $V_{av} = (V_{sp} - 1)(t_{on}/T)$, where V_{sp} is the secondary peak voltage. When V_{ea} has moved up to the top of the 3-V triangle, $t_{on}/T = 0.5$ and $V_{av} = 0.5\,(V_{sp} - 1)$. The modulator DC gain G_m, then, between V_{av} and V_{ea} is

$$G_m = \frac{0.5(V_{sp} - 1)}{3} \tag{12.1}$$

This gain is independent of frequency.

There is also a loss G_s due to the sampling network R_1, R_2 in Fig. 12.1. Most of the frequently used PWM chips cannot tolerate more than 2.5 V at the reference input to the error amplifier (point *A*). Thus, when sampling a +5-V output, $R_1 = R_2$ and gain G_s between V_s and V_o in Fig. 12.1 is −6 dB.

12.2.5 Total output *LC* filter plus modulator and sampling network gain

From the above, the total gain G_t (in decibels) of the output *LC* filter gain G_f plus modulator gain G_m plus sampling network gain G_s is plotted as in Fig. 12.6. It is equal to $G_m + G_s$ from DC up to $F_o = 1/2\pi\sqrt{L_oC_o}$. At F_o, it breaks into a −2 slope and remains at that slope up to the frequency F_{esr} where the impedance of C_o equals R_{esr}. At that frequency, it breaks into a −1 slope.

From this curve, the error-amplifier gain and phase-shift-versus-frequency characteristic is established to meet the three criteria for a stable loop as described below.

12.3 Shaping Error-Amplifier Gain-Versus-Frequency Characteristic

Recall that the first criterion for a stable loop is that at the frequency F_{co} where the total open-loop gain is unity (0 dB), total open-loop phase shift must be short of 360° by the desired *phase margin,* which will herein be taken as 45°.

The sequence of steps is then first to establish the crossover frequency F_{co}, where the total open-loop gain should be 0 dB. Then choose the error-amplifier gain so that the total open-loop gain is forced to be 0 dB at that frequency. Next design the error-amplifier gain slope so that the total open-loop gain comes through F_{co} at a −1 slope (Fig. 12.4). Finally, tailor the error-amplifier gain versus frequency so that the desired phase margin is achieved.

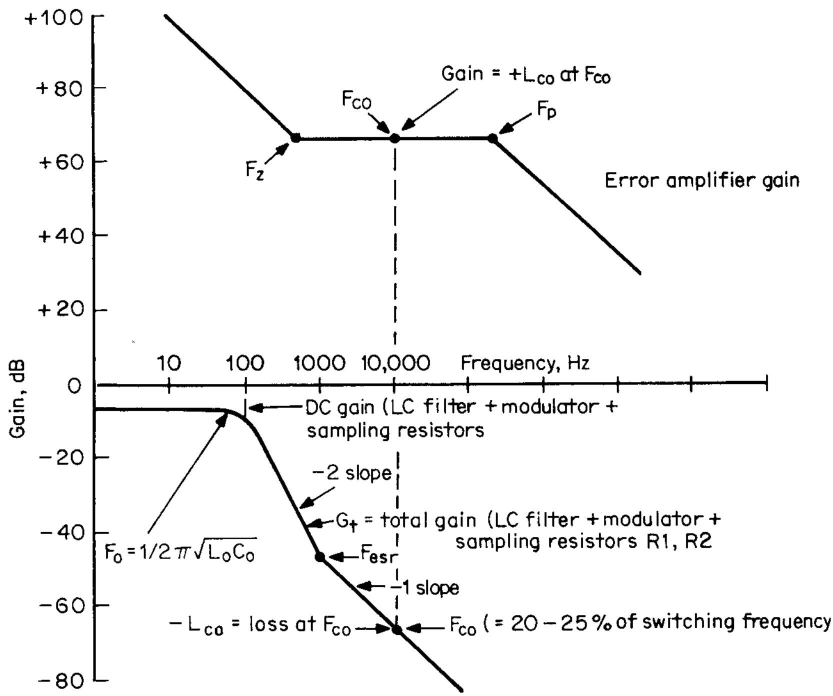


Figure 12.6 Gain G_t = sum of (*LC* filter + modulator + output voltage sampling resistors) gains determines error-amplifier gain. Error-amplifier gain at F_{co} is made equal and opposite to loss of G_t. Error-amplifier gain slope at F_{co} is made horizontal with upward and downward breaks at F_z and F_p. Location of F_z and F_p in frequency determines total circuit phase margin.

Sampling theory shows that F_{co} must be less than half the switching frequency for the loop to be stable. But it must be considerably less than that, or there will be large-amplitude switching frequency ripple at the output. Thus, the usual practice is to fix F_{co} at one-fourth to one-fifth the switching frequency.

Thus, refer to Fig. 12.6, which is the open-loop gain of the *LC* filter plus the PWM modulator plus the sampling network. The capacitor in the output filter is assumed in Fig. 12.6 to have an ESR which causes a break in the slope from -2 to -1 at $F_{esr} = 1/2\pi R_{esr}C_o$. Assume that F_{co} is one-fifth the switching frequency and read the loss in decibels at that point.

In most cases, the output capacitor will have an ESR and F_{esr} will come at a lower frequency than F_{co}. Thus at F_{co} the $G_1 = (G_{lc} + G_{pwm} + G_s)$ curve will already have a -1 slope.

Now when gains are plotted in decibels, both gains and gain slopes of gain elements in cascade are additive. Hence, to force crossover frequency to be at the desired one-fifth the switching frequency, choose the error-amplifier gain at F_{co} to be equal and opposite in decibels to the $G_t = (G_{lc} + G_{pwm} + G_s)$ loss at that frequency.

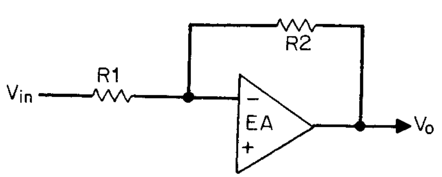
That forces F_{co} to occur at the desired point. Then, if the error-amplifier gain slope at F_{co} is horizontal, since the G_t curve at F_{co} already has a -1 slope, the sum of the error amplifier plus the G_t curve comes through crossover frequency at the desired -1 slope and the second criterion for a stable loop has been met.

Now the error-amplifier gain has been fixed as equal and opposite to the G_t loss at F_{co} and to have a horizontal slope as it passes through F_{co} (Fig. 12.6). Such a gain characteristic can be achieved with an operational amplifier with a resistor input and resistor feedback as in Fig. 12.7*a*. Recall that the gain of such an operational amplifier is $G_{ea} = Z2/Z1 = R2/R1$. But how far in frequency to the left and right of F_{co} should it continue to have this constant gain?

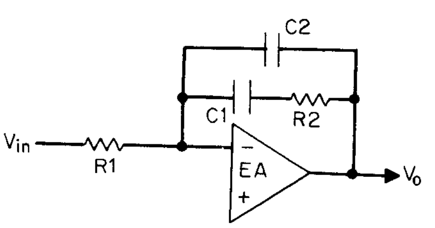
Recall that the total open-loop gain is the sum of the error-amplifier gain plus G_1 gain. If the error-amplifier gain remained constant down to DC, the total open-loop gain would not be very large at 120 Hz—the frequency of the AC power line ripple.

Yet it is desired to keep power line ripple attenuated down to a very low level at the output. To degenerate the 120-Hz ripple sufficiently, the open-loop gain at that frequency should be as high as possible. Thus at some frequency to the left of F_{co}, the error-amplifier gain should be permitted to increase rapidly.

This can be done by placing a capacitor $C1$ in series with $R2$ (Fig. 12.7*b*). This yields the low-frequency gain characteristic shown in Fig. 12.6. In the frequency range where the impedance of $C1$ is small compared to $R2$, the gain is horizontal and simply equal to $R2/R1$. At



(a)



(b)

Figure 12.7 (*a*) Error amplifier with resistor feedback $R2$ and resistor input $R1$ arms have a gain equal to $R2/R1$ which is independent of frequency up to the frequency where the open-loop error amplifier inside the loop (EA) commences falling off the gain. (*b*) Using complex feedback and input arms permits shaping the gain-versus-frequency and phase-shift-versus-frequency curves. The configuration above has the gain-versus-frequency characteristic of Fig. 12.6.

lower frequencies where the impedance of $C2$ is much higher than $R2$, effectively $R2$ is out of the circuit and the gain is X_{c1}/R_1. This gain increases at the rate of +20 dB/decade (+ 1 slope) toward lower frequencies and yields the higher gain at 120 Hz. Going in the direction of higher frequency, the −1 gain slope breaks and becomes horizontal at a frequency of $F_z = 1/2\pi R2C1$.

Now going to the right of F_{co} toward higher frequency (Fig. 12.6), if the error-amplifier gain curve were permitted to remain horizontal, total open-loop gain would remain relatively high at the high frequencies. But high gain at high frequencies is undesirable as thin, high-frequency noise spikes would be picked up and transmitted at large amplitudes to the output. Thus gain should be permitted to fall off at high frequencies.

This is easily done by placing a capacitor $C2$ across the series combination of $R2$ and $C1$ (Fig. 12.7*b*). At F_{co}, X_{c1} is already small compared to $R2$ and $C1$ is effectively out of the circuit.

At higher frequencies where X_{c2} is small compared to $R2$, however, $R2$ is effectively out of the circuit and gain is X_{c2}/R_1. Now the gain characteristic beyond F_{co} is horizontal up to a frequency F_p $(=1/2\pi R2C2)$, where it breaks and thereafter falls at a −1 slope as can be seen in Fig. 12.6. This lower gain at high frequency keeps high-frequency noise spikes from coming through to the output.

Now the break frequencies F_z and F_p must be chosen. They will be chosen so that $F_{co}/F_z = F_p/F_{co}$. The farther apart F_z and F_p are, the greater the phase margin at F_{co}. Large phase margins are desirable, but if F_z is chosen too low, low-frequency gain will be lower at 120 Hz than if a higher frequency were chosen (Fig. 12.8). Thus 120-Hz attenuation will be poorer. If F_p is chosen too high, gain at high frequencies is higher than if a lower F_p is chosen (Fig. 12.8). Thus high-frequency noise spikes would come through at a higher amplitude.

Thus a compromise between separating F_z and F_p by a large amount to increase phase margin and decreasing the separation to achieve better 120-Hz attenuation and lower-amplitude high-frequency noise spikes is sought.

This compromise and a more exact analysis of the problem is made easy by introducing the concept of transfer functions, poles, and zeros as shown below.

12.4 Error-Amplifier Transfer Function, Poles, and Zeros

The circuit of an operational amplifier with a complex impedance Z_1 input arm and a complex impedance Z_2 feedback arm is shown in Fig. 12.9. Its gain is Z_2/Z_1. If Z_1 is a pure resistor $R1$ and Z_2 is a pure re-

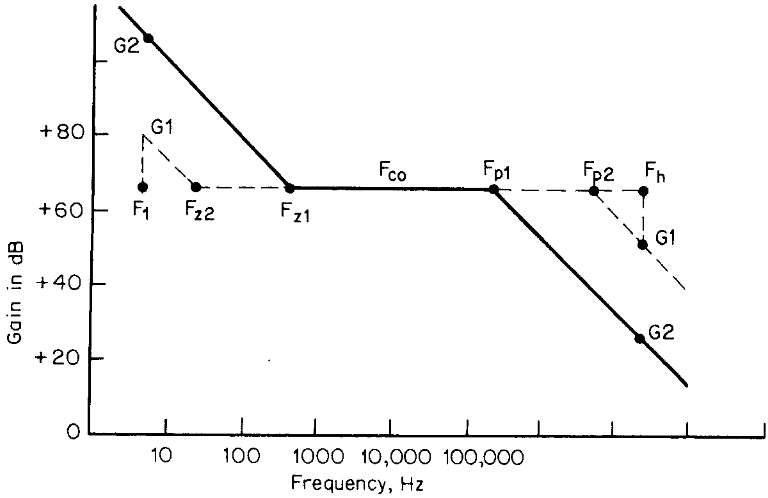


Figure 12.8 Where to locate break frequencies F_z and F_p. The farther apart F_z and F_p are spread, the greater the phase margin. But spreading them further apart reduces low-frequency gain, which reduces the degeneration of low-frequency line ripple. It also increases high gain, which permits high-frequency, thin noise spikes to come through at greater amplitude. If F_z were at F_{z2} instead of F_{z1}, gain at some low frequency F_1 would be G_1 instead G_2. And if F_p were at F_{p2} instead of F_{p1}, gain at some high frequency F_h would be G_1 instead of G_2.

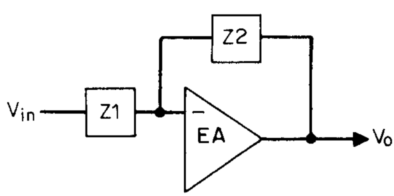


Figure 12.9 If inputs and feedback arms are various combinations of Rs and Cs, various gain-versus-frequency and phase-shift-versus-frequency curves are possible. By expressing impedances Z_1 and Z_2 in terms of an s operator ($=jw$), and performing a number of algebraic manipulations, a simplified expression for the gain arises. From this simplified gain expression (transfer function) the gain-versus-frequency and phase-shift-versus-frequency curves can be drawn at a glance.

sistor $R2$ as in Fig. 12.7*a*, gain is R_2/R_1 and is independent of frequency. Phase shift between V_o and V_{in} is 180°, as the input is to the inverting terminal.

Now impedances in Z_1, Z_2 are expressed in terms of the complex variable $s = j(2\pi f) = jw$. Thus the impedance of a capacitor $C1$ is then $1/sC1$ and that of a resistor $R1$ and capacitor $C1$ in series is $(R1 + 1/sC1)$.

The impedance of an arm consisting of a capacitor $C2$ in parallel with a series combination of $R1$ and $C1$ is then

$$Z = \frac{(r + 1/sC1)(1/sC2)}{r + 1/sC1 + 1/sC2} \qquad (12.2)$$

Now the gain or transfer function of the operational error amplifier is written in terms of its Z_1, Z_2 impedances, which are expressed in terms of the complex variable s. Thus $G(s) = Z_2(s)/Z_1(s)$, and by algebraic manipulation, $G(s)$ is broken down into a simplified numerator and denominator which are functions of s: $G(s) = N(s)/D(s)$. The numerator and denominator, again by algebraic manipulation, are factored and $N(s)$, $G(s)$ are expressed in terms of these factors. Thus

$$G(s) = \frac{N(s)}{D(s)} = \frac{(1 + sz_1)(1 + sz_2)(1 + sz_3)}{sp_0(1 + sp_1)(1 + sp_2)(1 + sp_3)} \qquad (12.3)$$

These z and p values are RC products and represent frequencies. These frequencies are obtained by setting the factors equal to zero. Thus

$$1 + sz_1 = 1 + s(j2\pi fz_1) = 1 + j2\pi fR1C1 = 0 \qquad \text{or} \qquad f_1 = 1/2\pi R1C1$$

The frequencies corresponding to the z values are called *zero* frequencies, and those corresponding to the p values are called *pole* frequencies. There is always a factor in the denominator which has the "1" missing (note sp_0 above). This represents an important pole frequency, $F_{po} = 1/2\pi R_oC_o$, which is called *the pole at the origin.*

From the location of the pole at the origin and the zero and pole frequencies, the gain-versus-frequency characteristic of the error amplifier can be drawn as discussed below.

12.5 Rules for Gain Slope Changes Due to Zero and Pole Frequencies

The zero and pole frequencies represent points where the error-amplifier gain slope changes.

A zero represents a +1 change in gain slope. Thus (Fig. 12.10*a*), if a zero appears at a point in frequency where the gain slope is zero, it turns the gain into a +1 slope. If it appears where the original gain slope is −1 (Fig. 12.10*b*), it turns the gain slope to zero. Or if there are two zeros at the same frequency (two factors in the numerator of Eq. 12.3 having the same RC product) where the original gain slope is −1,

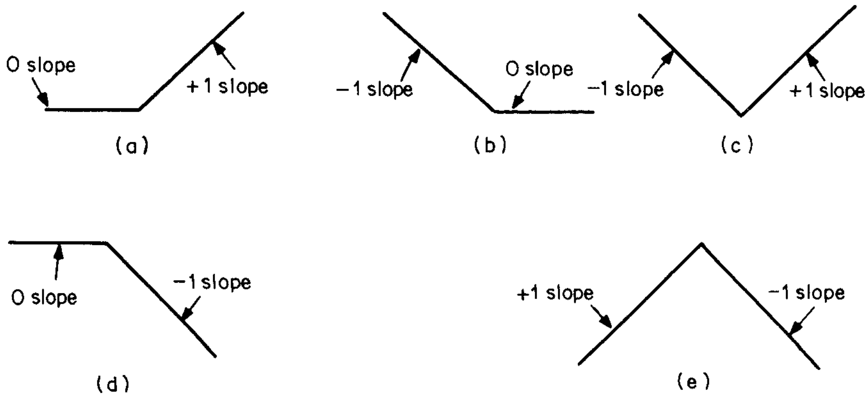


Figure 12.10 (*a*) A zero appearing on a gain curve where the original slope was horizontal turns that gain slope to +1 or +20 dB/decade. (*b*) A zero appearing on a gain curve where the original slope was −1, terms that gain slope horizontal. (*c*) Two zeros at the same frequency appearing on a gain curve where the original slope was −1, turns that slope to +1. The first zero turns the −1 slope horizontal; the second zero at the same frequency turns the horizontal slope to +1. (*d*) A pole appearing on a gain curve where the original slope is horizontal turns that slope to −1 or 20 dB/decade. (*e*) Two poles at the same frequency appearing on a gain curve where the original slope is +1 turns that slope to −1. The first pole turns the original +1 slope horizontal; the second one at the same frequency turns the horizontal slope to −1 or −20 dB/decade.

the first zero turns the gain slope horizontal, and the second zero at the same frequency turns the gain into a +1 slope (Fig. 12.10*c*).

A pole represents a −1 change in gain slope. If it appears in frequency where the original gain slope is horizontal (slope is zero), it turns the gain into a −1 slope (Fig. 12.10*d*). Or if there are two poles at the same frequency at a point where the original gain slope is +1, the first pole turns the slope horizontal and the second at that same frequency turns the slope to −1 (Fig. 12.10*e*).

The pole at the origin, as does any pole, represents a gain slope of −1. It also indicates the frequency at which the gain is 1 or 0 dB. Thus, drawing the total gain curve for the error amplifier starts with the pole at the origin as follows. Go to 0 dB at the frequency of the pole at the origin $F_{po} = 1/2\pi R_o C_o$. At F_{po}, draw a line backward in frequency with a slope of +1 (Fig. 12.11). Now if somewhere on this line, which has a −1 slope in the direction of higher frequency, the transfer function has a zero at a frequency $F_z = 1/2\pi R1C1$, turn the gain slope horizontal at F_z. Extend the horizontal gain indefinitely. But if at some higher frequency there is a pole in the transfer function at a frequency $F_p = 1/2\pi R2C2$, turn the horizontal slope into a −1 slope at F_p (Fig. 12.11).

The gain along the horizontal part of the transfer function is $R2/R1$

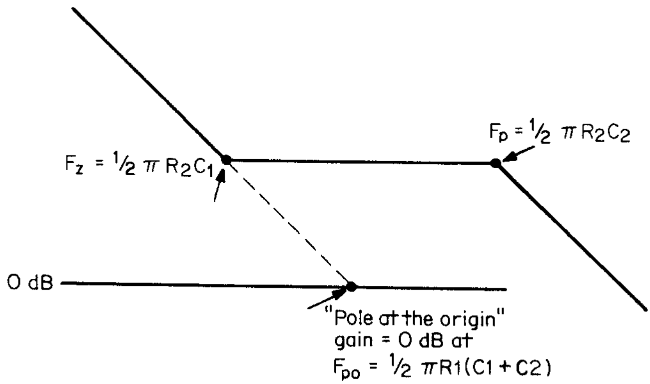


Figure 12.11 Drawing the gain curve of the error amplifier of Fig. 12.7*b* directly from its transfer function of Eq. 12.3.

and is made (in decibels) equal and opposite to the loss of the G_t curve (Fig. 12.6) at F_{co}.

Thus an error-amplifier gain curve having a single pole at the origin, a single zero, and another single pole has the desired shape shown in Fig. 12.11. It is implemented by the circuit of Fig. 12.7*b*. It remains only to select the location of the zero and pole frequencies which yield the desired phase margin. This will be discussed below.

12.6 Derivation of Transfer Function of an Error Amplifier with Single Zero and Single Pole from Its Schematic

It has been shown above that if an error amplifier had a single zero, a single pole, and a pole at the origin, its gain-versus-frequency curve would be as in Fig. 12.11.

Now it will be demonstrated how the transfer function of an error amplifier is derived and that the circuit of Fig. 12.7*b* does truly have a single zero, a single pole, and a pole at the origin. Gain of the circuit in Fig. 12.7*b* is

$$G = \frac{dV_o}{dV_i}$$

$$= \frac{Z_2}{Z_1}$$

$$= \frac{(R2 + 1/jwC1)(1/jwC2)}{R1(R2 + 1/jwC1 + 1/jwC2)}$$

Now introduce the complex variable $s = jw$. Then

$$G = \frac{(R2 + 1/sC1)(1/sC2)}{R1(R2 + 1/sC1 + 1/sC2)}$$

And by algebraic manipulation

$$G = \frac{1 + sR2C1}{sR1(C1 + C2)(1 + sR2C1C2/C1 + C2)}$$

And since generally $C2 << C1$

$$G = \frac{1 + sR2C1}{sR1(C1 + C2)(1 + R2C2)} \tag{12.4}$$

The error amplifier of Fig. 12.7*b*, having the transfer function of Eq. 12.4, is commonly referred to as a *Type 2 amplifier* in conformance with the designation introduced by Venable in his classic paper.[1] A Type 2 error amplifier is used when the output filter capacitor has an ESR so that F_{co} lies on a -1 slope of the G_1 curve (Fig. 12.6).

Examination of this transfer function for the circuit of Fig. 12.7*b* permits immediate drawing of its gain characteristic as follows (Fig. 12.11). Equation 12.4 shows that this circuit (Fig. 12.7*b*) has a pole at the origin at a frequency of $F_{po} = 1/2\pi R1(C1 + C2)$. Thus, go to 0 dB at that frequency and draw a line backward toward lower frequency with a slope of $+1$.

From Eq. 12.4, the circuit has zero at a frequency of $F_z = 1/2\pi R2C1$. Go to that sloped line and turn it horizontal at F_z. Again from Eq. 12.4, the circuit has a pole at a frequency of $F_p = 1/2\pi R2C2$. Now go out along that horizontal line and turn it into a -1 slope at F_p.

Now that the transfer function of the Type 2 error amplifier can be drawn from its pole and zero frequencies, it remains to be able to locate them (choosing $R1$, $R2$, $C1$, $C2$) so as to achieve the desired phase margin. This is demonstrated below.

12.7 Calculation of Type 2 Error-Amplifier Phase Shift from Its Zero and Pole Locations

Adopting Venable's scheme,[1] the ratio $F_{co}/F_z = K$ will be chosen equal to $F_p/F_{co} = K$.

Now a zero, like an *RC* differentiator (Fig. 12.2*b*), causes a phase lead. A pole, like an *RC* integrator (Fig. 12.2*a*), causes a phase lag.

The phase lead at a frequency F due to a zero at a frequency F_z is

$$\Theta_{ld} = \tan^{-1} \frac{F}{F_z}$$

But we are interested in the phase lead at F_{co} due to a zero at a frequency F_z. This is

$$\Theta_{ld}\,(\text{at } F_{co}) = \tan^{-1} K \tag{12.5}$$

The phase lag at a frequency F due to a pole at a frequency F_p is

$$\Theta_{lag} = \tan^{-1}\frac{F}{F_p}$$

and we are interested in the lag at F_{co} due to the pole at F_p. This is

$$\Theta_{lag}\,(\text{at } F_{co}) = \tan^{-1}\frac{1}{K} \tag{12.6}$$

The total phase shift at F_{co} due to the lead of the zero at F_z and the lag due to the pole at F_p is the sum of Eqs. 12.5 and 12.6.

These shifts are in addition to the inherent low-frequency phase shift of the error amplifier with its pole at the origin. The error amplifier is an inverter and at low frequency causes a 180° phase shift.

At low frequencies, the pole at the origin causes a 90° phase shift. This is just another way of saying that at low frequencies, the circuit is just an integrator with resistor input and capacitor feedback. This is seen from Fig. 12.7*b*. At low frequencies, the impedance of *C*1 is much greater than *R*2. The feedback arm is thus only *C*1 and *C*2 in parallel.

Thus the inherent low-frequency phase lag is 180° because of the phase inversion plus 90° due to the pole at the origin or a total lag of 270°. Total phase lag, including the lead due to the zero and lag due to the pole, is then

$$\Theta_{(\text{total lag})} = 270° - \tan^{-1} K + \tan^{-1}\frac{1}{K} \tag{12.7}$$

Note that this is still a net phase lag as when *K* is a very large number (zero and pole frequencies very far apart), the lead due to the zero is a maximum of 90° and the lag due to the pole is 0°.

Total phase lag through the error amplifier, calculated from Eq. 12.7 is shown in Table 12.1.

12.8 Phase Shift through *LC* Filter Having ESR in Its Output Capacitor

The total open-loop phase shift consists of that through the error amplifier plus that through the output *LC* filter. Figure 12.3*b* showed for $R_o = 20\sqrt{L_o/C_o}$ and no ESR in the filter capacitor, the lag through the filter itself is already 175° at $1.2F_o$.

This lag is modified significantly if the output capacitor has an ESR

TABLE 12.1 Phase Lag through a Type 2 Error Amplifier for Various Values of $K(= F_{co}/F_z = F_p/F_{co})$

K	Lag (from Eq. 12.7)
2	233°
3	216°
4	208°
5	202°
6	198°
10	191°

as in Fig. 12.5*b*. In that figure, the gain slope breaks from a −2 to a −1 slope at the so-called ESR zero frequency of $F_{esr} = 1/2\pi R_{esr}C_o$. Recall that at F_{esr}, the impedance of C_o equals that of R_{esr}. Beyond F_{esr}, the impedance of C_o becomes smaller than R_{esr} and the circuit becomes increasingly like an *LR* rather than an *LC* circuit. Moreover, an *LR* circuit can cause only a 90° phase lag as compared to the possible maximum of 180° for an *LC* circuit.

Thus the ESR zero creates a boost in phase over a possible maximum of 180°. Phase lag at a frequency F due to an ESR zero at F_{esro} is

$$\Theta_{lc} = 180° - \tan^{-1}\frac{F}{F_{esro}}$$

and since we are interested in the phase lag at F_{co} due to the zero at F_{esro}

$$\Theta_{lc} = 180° - \tan^{-1}\frac{F_{co}}{F_{esro}} \tag{12.8}$$

Phase lags through the *LC* filter (having an ESR zero) are shown in Table 12.2 for various values of F_{co}/F_{esro} (from Eq. 12.8).

TABLE 12.2 Phase Lag through an *LC* Filter at F_{co} Due to a Zero at F_{esro}

F_{co}/F_{esro}	Phase lag	F_{co}/F_{esro}	Phase lag
0.25	166°	2.5	112°
0.50	153°	3	108°
0.75	143°	4	104°
1.0	135°	5	101°
1.2	130°	6	99.5°
1.4	126°	7	98.1°
1.6	122°	8	97.1°
1.8	119°	9	96.3°
2.0	116°	10	95.7°

Thus, by setting the error-amplifier gain in the horizontal part of its gain curve (Fig. 12.6) equal and opposite to the G_t (Fig. 12.6) loss at F_{co}, the location of F_{co} is fixed where it is desired. Since F_{co} will in most cases be located on the -1 slope of the G_t curve, the total gain curve will come through F_{co} at a -1 slope. From Tables 12.1 and 12.2, the proper value of K (locations of the zero and pole) is established to yield the desired phase margin.

12.9 Design Example—Stabilizing a Forward Converter Feedback Loop with a Type 2 Error Amplifier

The design example presented below demonstrates how much of the material discussed in all previous chapters is interrelated.

Stabilize the feedback loop for a forward converter with the following specifications:

V_o	5.0 V
$I_{o(nom)}$	10 A
Minimum I_o	1 A
Switching frequency	100 kHz
Minimum output ripple (peak to peak)	50 mV

It is assumed that the filter output capacitor has an ESR and F_{co} will occur on the -1 slope of the LC filter. This permits the use of a Type 2 error amplifier with the gain characteristics of Fig. 12.6. The circuit is shown in Fig. 12.12.

First L_o, C_o will be calculated and the gain characteristic of the output filter will be drawn. From Eq. 2.47

$$L_o = \frac{3V_oT}{I_{on}}$$

$$= \frac{3 \times 5 \times 10^{-6}}{10}$$

$$= 15 \times 10^{-6}\ \text{H}$$

and from Eq. 2.48

$$C_o = 65 \times 10^{-6}\frac{dI}{V_{or}}$$

where dI is twice the minimum output current $= 2 \times 1 = 2$ A and V_{or} is the output ripple voltage $= 0.05$ V. Then $C_o = 65 \times 10^{-6} \times 2/0.05 =$ 2600 microfarads.

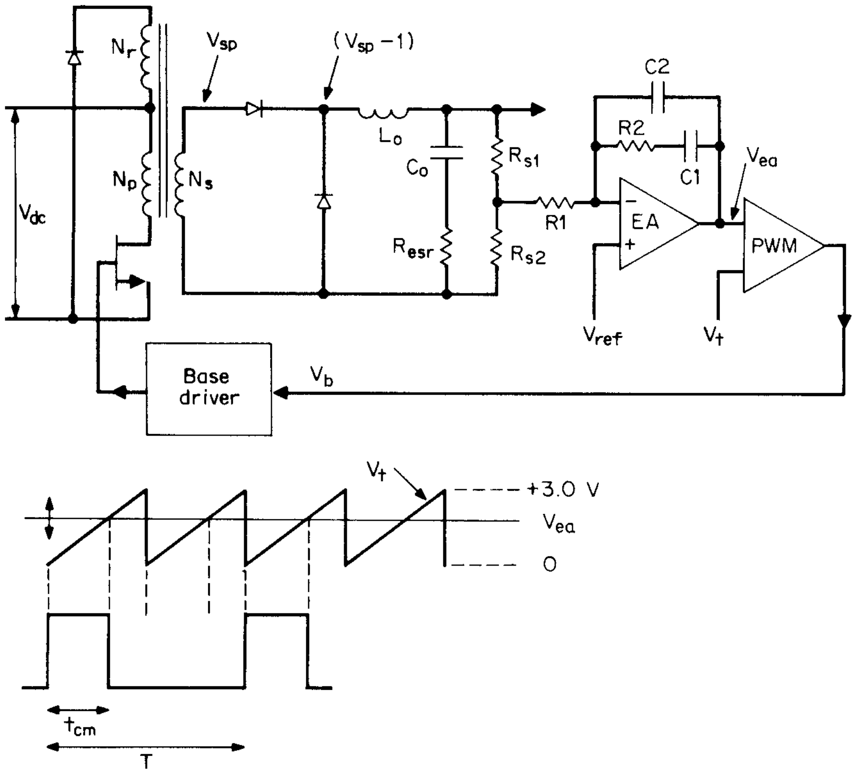


Figure 12.12 Forward converter design example schematic stabilizing the feedback loop.

Corner frequency of the output *LC* filter, from Sec. 12.2.3, is

$$F_o = 1/2\pi\sqrt{L_oC_o}$$

$$= 1/2\pi\sqrt{15 \times 10^{-6} \times 2600 \times 10^{-6}}$$

$$= 806 \text{ Hz}$$

Again from Sec. 12.2.3, the frequency of the ESR zero (frequency where the gain slope abruptly changes from a −2 to a −1 slope) is

$$F_{esr} = 1/2\pi R_{esr}C_o$$

$$= 1/2\pi(65 \times 10-6)$$ (assuming, as in Sec. 2.3.11.2, that over a large range of aluminum electrolytic capacitor magnitudes and voltage ratings, that $R_{esr}C_o$ is constant and equal to 65 × 10−6)

$$= 2500 \text{ Hz.}$$

From Eq. 12.1, the modulator gain G_m is $G_m = 0.5(V_{sp} - 1)/3$, and when the duty cycle is 0.5, for $V_o = 5V$, $V_{sp} = 11$ V since $V_o = (V_{sp} - 1)T_{on}/T$. Then $G_m = 0.5(11 - 1)/3 = 1.67 = +4.5$ dB.

For the usual SG1524-type PWM chip, which can tolerate only 2.5 V at the reference input to the error amplifier, for $V_o = 5$ V, $R_{s1} = R_{s2}$. Sampling network gain (loss) is then $G_s = -6$ dB. Then $G_m + G_s = +4.5 - 6.0 = -1.5$ dB.

The open-loop gain curve of everything but the error amplifier is then $G_t = G_{lc} + G_m + G_s$ and is drawn in Fig. 12.13 as curve *ABCD*. From *A* to the corner frequency at 806 Hz (*B*) it has a value of $G_m + G_s = -1.5$ dB. At *B*, it breaks into a −2 slope and continues at that slope up to the ESR zero at 2500 Hz (*C*). At point *C*, it breaks into a −1 slope.

Now crossover frequency is taken at one-fifth the switching frequency or 20 kHz. From the G_t curve, loss at 20 kHz is −40 dB (numerical loss of 1/100). Hence, to make 20 kHz the crossover frequency, the error-amplifier gain at that frequency is made +40 dB. Since the total open-loop gain of the error amplifier plus curve *ABCD*) must come through crossover at a −1 slope, the error-amplifier gain curve must have zero slope between points *F* and *G* in curve *EFGH* of Fig. 12.13 since *ABCD* already has a −1 slope at 20 kHz.

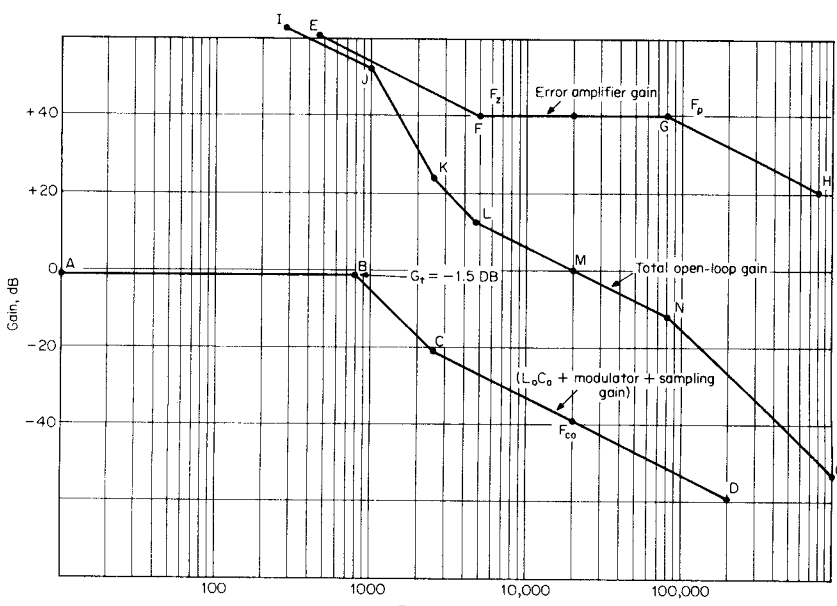


Figure 12.13 Design example stabilizing the feedback loop for Fig. 12.12.

This horizontal gain slope between points F and G is obtained as described above with a Type 2 error amplifier. The gain of the Type 2 error amplifier in the horizontal part of its slope is $R2/R1$. If $R1$ is arbitrarily taken as 1000 Ω, R_2 is 100,000 Ω.

Now a zero is located at F_z to increase low-frequency gain to degenerate 120-Hz line ripple and a pole is located at G to commence decreasing high-frequency gain so as to minimize thin noise spikes at the output. The zero and the pole will be located to give the desired phase margin.

Assume a 45° phase margin. Then total phase shift around the loop at 20 kHz is 360 − 45 = 315°. But the LC filter by itself causes a phase lag given by Eq. 12.7. From that equation, the lag for F_{co} = 20 kHz and F_{esro} = 2500 Hz is 97° (Table 12.2). Thus the error amplifier is permitted only 315 − 97 or 218°. Table 12.1 shows that for an error-amplifier lag of 218°, a K factor of slightly less than 3 would suffice.

To provide somewhat more insurance, assume a K factor of 4 which yields a phase lag of 208°. This, plus the 97° lag of the LC filter, yields a total lag of 305° and a phase margin of 360 − 305° or 55° at F_{co}.

For a K factor of 4, the zero is at $F_z = 20/4 = 5$ kHz. From Eq. 12.3, $F_z = 1/2\pi R2C1$. For $R2$ determined above as $100K$, $C1 = 1/2\pi(100{,}000)(5000) = 318 \times 10^{-12}$.

Again for the K factor of 4, the pole is at $F_{po} = 20 \times 4 = 80$ kHz. From Eq. 12.3, $F_{po} = 1/2\pi R2C2$. For $R2 = 100K$, $F_{po} = 80$ kHz, $C2 = 1/2\pi(100{,}000)(80{,}000) = 20 \times 10^{-12}$. This completes the design; the final gain curves are shown in Fig. 12.13. Curve *IJKLMO* is the total open-loop gain. It is the sum of curves *ABCD* and *EFGH*.

12.10 Type 3 Error Amplifier—When Used and Transfer Function

In Sec. 2.3.11.2, it was pointed out that the output ripple $V_{or} = R_o\, dI$ where R_o is the ESR of the filter output capacitor C_o and dI is twice the minimum DC current. Now most aluminum electrolytic capacitors do have an ESR. Study of many capacitor manufacturers' catalogs indicates that for such capacitors, R_oC_o is constant and equal to an average value of 65×10^{-6}.

Thus, using conventional aluminum electrolytic capacitors, the only way to reduce output ripple is to decrease R_o, which can be done only by increasing C_o. This, of course, increases size of the capacitor, which may be unacceptable.

Within the past few years, capacitor manufacturers have been able (at considerably greater cost) to produce aluminum electrolytic capacitors with essentially zero ESR for those applications where output ripple must be reduced to an absolute minimum.