ELECTRONOTES 193

NEWSLETTER OF THE

MUSICAL ENGINEERING GROUP

1016 Hanshaw Rd., Ithaca, NY 14850

Volume 19, No. 193 March 2000

GROUP ANNOUNCEMENTS

CONTENTS OF EN#193

This issue covers Chapter 4 of Analog Signal Processing.

Much in keeping with the spirit of the Analog Signal Processing presentation in this series of issues, we also have, mostly by coincidence, a series of three items relating in analog-like issues, and familiar persons from the past. In this issue, immediately below, we have a discussion sparked by an inquiry from Ian Fritz, regarding a VCF design going back to the Musical Engineer's Handbook and even before. Future issues will show an improvement to a popular VCO design, and a more detailed active sensitivity study of the state-variable filter. Perhaps more similar items will surface.

A BUTTERWORTH THAT PEAKS!

Ian Fritz, long time designer of electronic music equipment, and contributor to Electronotes, recently contacted me regarding a voltage-controlled-filter design that originated with Sergio Franco [Hardware Design of a Real-Time Musical System, Dept of Comp Sci, U. of Ill, Urbana, UIUCDCS-R-74-677, Oct. 1974, pg 51] and which was also discussed in our own 1975 Musical Engineer's Handbook, pg 5d (5). In his discussion, Franco says "The circuit realizes a two-pole maximally-flat, low-pass

programmable filter." The MEH discussion calls this a two-pole Butterworth, and mentions a previous MEH circuit, pg 4b (6), that realizes a two-pole Butterworth with capacitors in a 2:1 ratio. Ian's experiment findings were that the response peaks (i.e., was not Butterworth / Maximally-Flat). A careful theoretical calculation indicates that Ian's experimental findings are in fact in agreement with the correct analysis.

Likely some of the confusion comes from the similarity of the voltage-controlled version (seen in Fig. 2) to the non-voltage-controlled version, which is shown in Fig. 1.

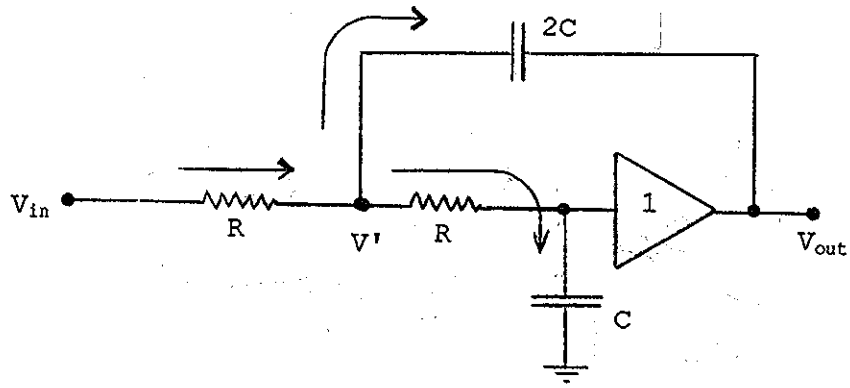


Fig. 1 A two-pole Butterworth using capacitors C and 2C

-continued on pg. 22

CHAPTER 4

ADDITIONAL FILTER TYPES:
NOTCH AND ALL-PASS:

4-1 FILTER TYPES AND THEIR POLE/ZERO PLOTS:

We need to be able to envision the frequency responses of different filter types in terms of their pole/zero plots. As we have seen, the pole/zero plot gives the filter's transfer function up to an arbitrary multiplicative constant. Further, we know that the frequency response is proportional to the product of the distances from the zeros, and inversely proportional to the product of the distances to the poles. This perspective permits us to look at the types we have already seen, and the types that will be added in this chapter, in terms of poles and zeros.

In saying that the frequency response is inversely proportional to a distance from a pole, it is clear that if we are close to a pole, the response should be large. Conversely, if we are close to a zero, the response should be small. We can also understand that at any one point in the s-plane (usually a frequency point on the $j\omega$-axis), it is possible that one or more poles or zeros, which are very close to this point, will dominate the response there. A good example would be a zero on the $j\omega$-axis (which is allowed, since zeros do not affect stability). This would clearly force the response to a null at the frequency of the zero (a "notch" filter, as we shall see). A second example would be a pole that is very close to the $j\omega$-axis, but in the left half-plane (as it must be for stability). For points on the $j\omega$-axis near this pole, we expect a very large value of frequency response.

We can also understand the "asymptotic" behavior of filters in terms of the poles and zeros. By this we mean the roll-off rate for very high frequencies. For frequencies very high as compared to the radii of all the finite poles and zeros of the filter, the distances from this high-frequency point on the $j\omega$-axis back to the poles and zeros are all approximately the same. If we put this fact into equation (1-35), any distance factors caused by finite zeros will cancel one of the distance factors caused by a pole. For example, if there are two finite zeros and three finite poles, and we are looking at a high frequency ω which is approximately a distance R from all the poles and zeros ($R \sim \omega$), then equation (1-35) gives the frequency response as:

$$|T(s)| = \frac{A \cdot R \cdot R}{R \cdot R \cdot R} = \frac{A}{R} \qquad (4\text{-}1)$$

which indicates that this filter is asymptotically rolling off as 1/R (as $1/\omega$), and is a first-order or 6db/octave roll-off. If there were two finite zeros and two finite poles, then:

$$|T(s)| = \frac{A \cdot R \cdot R}{R \cdot R} = A \qquad (4\text{-}2)$$

which means that the response will come flat to a value of A, which is the case with a high-pass filter, or with a notch (as we shall see). In general, the response rolls off at 6db/octave for every "excess pole" that is present. We can also understand that our transfer functions can not have "excess zeros" or they would blow up at high frequencies, and that the order of the numerator can not exceed the order of the denominator.

We can also understand a good deal about a frequency response by thinking in terms of $|T(s)|$ being a sort of "contour map" over the s-plane. In this view, there should be a very high "mountain" in the vicinity of any pole, and a "sinkhole" to zero at each zero. Envisioning this, we then contemplate a journey along the $j\omega$-axis as being what we look at in terms of frequency response. Along this axis, we cross "contour lines" as we move toward higher and lower ground. We understand that nearby mountains and sinkholes are more relevant than those that are far away. Ultimately, we want to read a pole/zero plot and be able to estimate the frequency response. Conversely, if we know of a response we need, we may be able to determine if we need extra poles and/or zeros, and approximately where.

Fig. 4-1 shows some "contour maps" for filter types which we have already looked at, including several low-pass types, high-pass, and bandpass. Fig. 4-1a shows a second-order low-pass where the poles are far enough back that there is no ripple in the frequency response (Butterworth-like). In this view, the jω-axis already exists as a straight line, and it is desired to make a perfectly flat "road" along this line for as long as is possible. Further, the way to do this is to "move the mountains" (poles) so that the "roadbed" becomes flat. This means that the contour lines shown must run parallel to the jω-axis for as long as possible, and then the jω-axis crosses the contour lines in a downward direction (low-pass roll-off). In contrast, Fig. 4-1b shows a corresponding "Chebyshev" case. We can think of this as effectively moving the jω-axis closer to the poles of Fig. 4-1a, and looking at the map on a slightly different scale. Now in the passband region the jω-axis crosses the contour lines going up (peaking) before going into the cutoff region.

In Fig. 4-1c, we have the poles moved back to their "Butterworth" positions and two zeros have been added at s=0. Here as we move up the jω-axis we first begin at s=0 crawling out of a "sinkhole." Notice that while our mountains here are infinitely high at their very peaks (which we never visit anyway), the sinkhole of the zero is not going to negative infinity, but only to zero. After getting away from s=0, we note that the contour lines become parallel to the jω-axis. [Note that this is exactly the case of equation (4-2).] This makes the response come level at high frequency, which is the response we want for high-pass.

Fig. 4-1d and Fig. 4-1e show two more low-pass cases, in this case 4th and 5th-order Chebyshev type responses, respectively. The rippling of the passband is caused by the jω-axis crossing over and back across contour lines. Note here in particular the difference between even and odd order. Even order has an elevated region at s=0 which corresponds to the real pole.

Finally, Fig. 4-1f shows a bandpass second-order case, which we need to compare with the high-pass of Fig. 4-1c. For bandpass, we have only the one zero at s=0, and this has two effects. First, the "climb" up out of the s=0 "sinkhole" is not so steep. Secondly, after we get over the peak adjacent to the pole, the contour lines bend back around and recross the jω-axis, and the response is headed down (which is the run away from one "net" pole). In the case of the high-pass, the second zero at s=0 held the contours parallel to the jω-axis.

Having now reviewed the past filter types within the context of pole/zero diagrams and contour maps, we want to look at the two new types that will become the major study of this chapter. These are the notch filter, and the all-pass filter. Above we saw zeros only at s=0 (or at infinity), and here we will look at cases where the zeros are at finite positions in the s-plane, but not at s=0. In the notch case, the zeros are on the jω-axis. In the all-pass case, the zeros are in the right half-plane (which is stable for zeros), and at positions that are mirror images of the pole positions, across the jω-axis.

Fig. 4-2a shows the notch filter pole/zero plot. Here the two zeros place sinkholes on the jω-axis, which the response must dip into. The poles in turn serve to hold up the region on either side of the notch. Far away from the center of the s-plane, the contour lines come parallel to the jω-axis, as in the high-pass case [again, see equation (4-2)], since there are no net poles. Fig. 4-2b is the all-pass case, and like the notch, it has no net poles, and the response comes flat at high frequencies. However, with the zeros placed exactly across from the poles, for any point on the jω-axis, corresponding pole and zero pairs are at exactly the same distances. Thus the response is everywhere flat, and hence the name "all-pass." It is clear that we are not going to be interested in the all-pass for its magnitude response, but may be for its phase response.

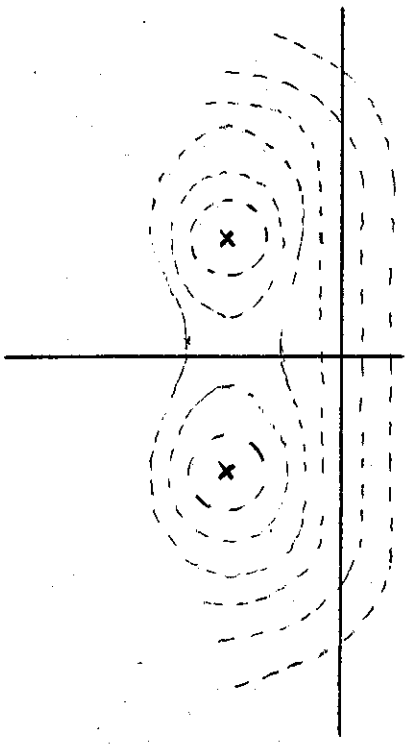


Fig. 4-1a "Butterworth" 2nd-Order Low-Pass

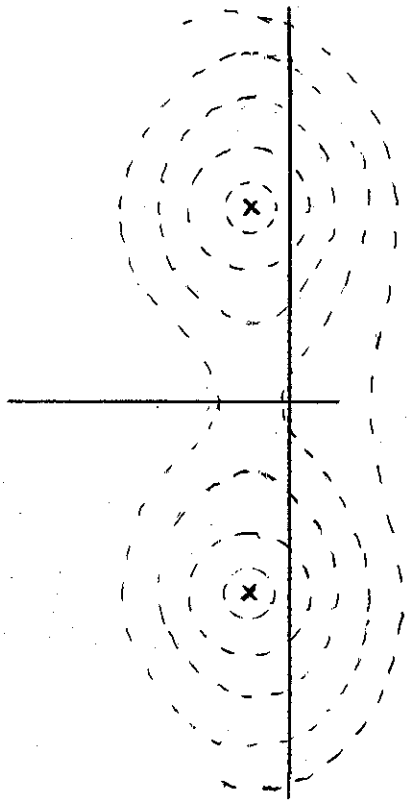


Fig. 4-1b "Chebyshev" 2nd-Order Low-Pass

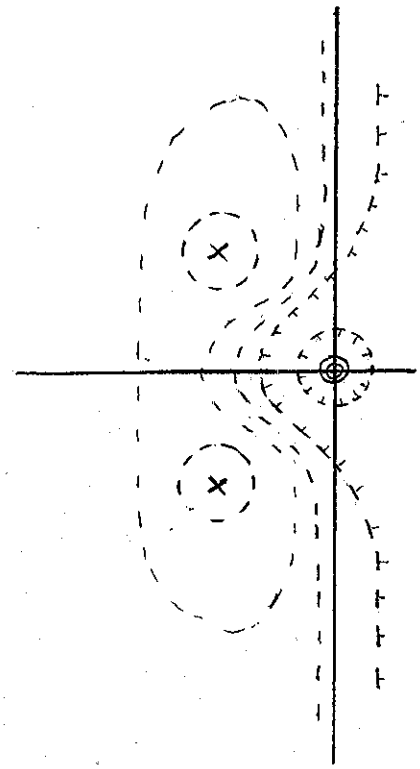


Fig. 4-1c "Butterworth" 2nd-Order High-Pass

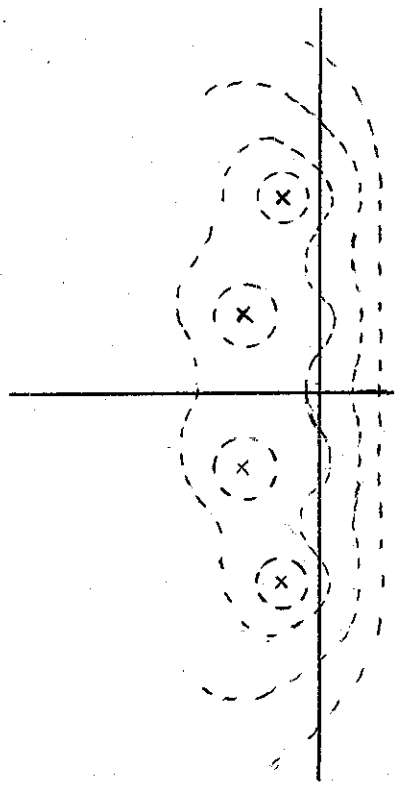


Fig. 4-1d "Chebyshev" 4th-Order Low-Pass

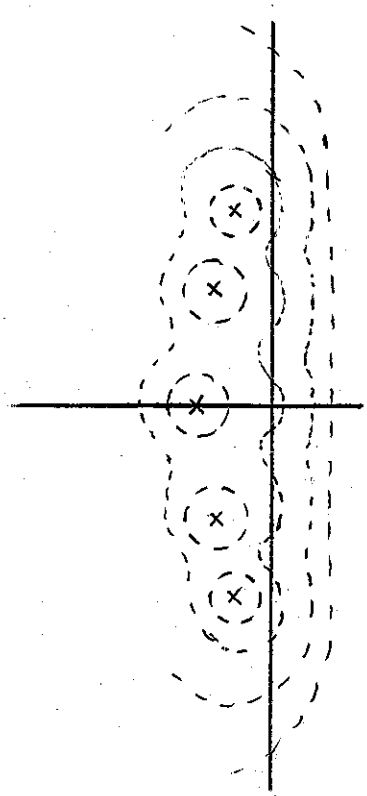


Fig. 4-1e "Chebyshev" 5th-Order Low-Pass

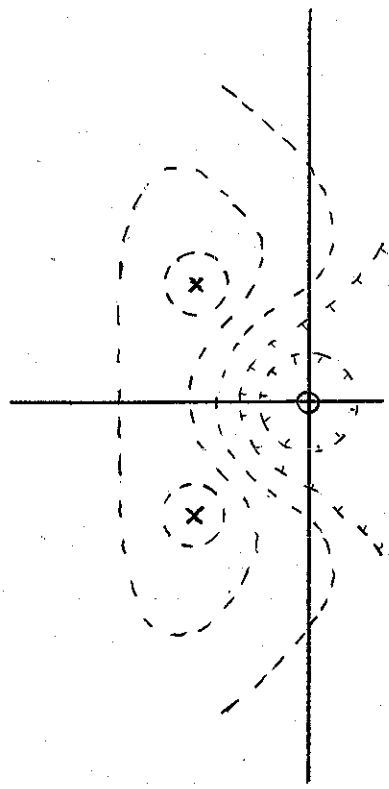


Fig. 4-1f 2nd-Order Bandpass

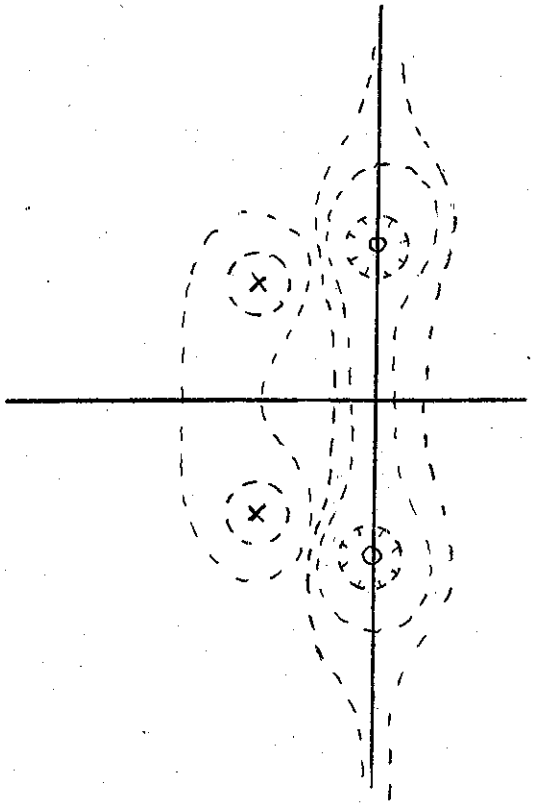


Fig. 4-2a "Notch"

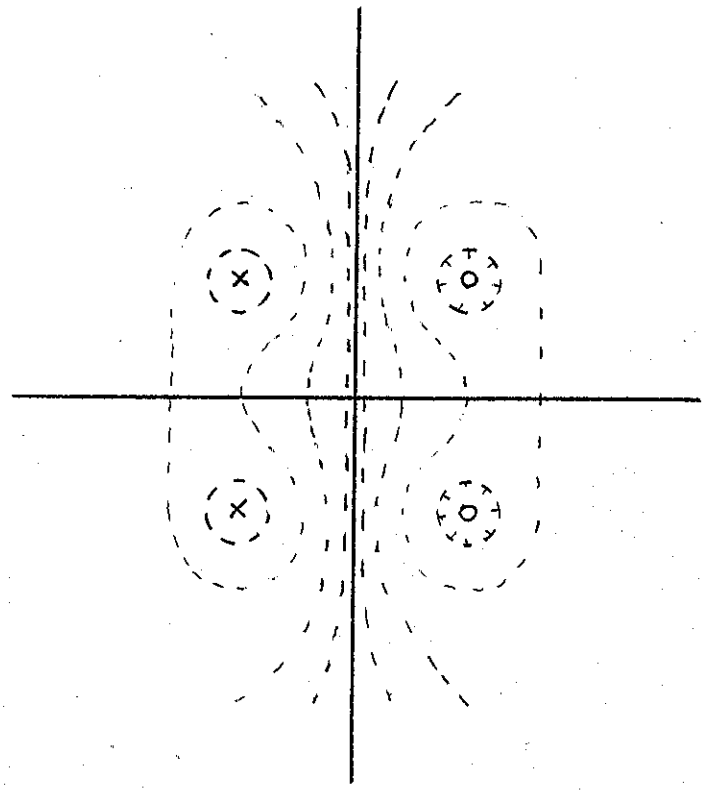


Fig. 4-2b 2nd-Order "All-Pass"

4-2 NOTCH FILTERING

Notch filters, like bandpass, must be at least second-order. In a notch filter we are trying to place a pair of complex conjugate zeros on the jω-axis, so that the frequency response is forced to zero at that frequency. This can be accomplished with a transfer function:

$$T_N(s) = \frac{s^2 + \omega_n^2}{s^2 + (\omega_o/Q)s + \omega_o^2} \qquad (4\text{-}3)$$

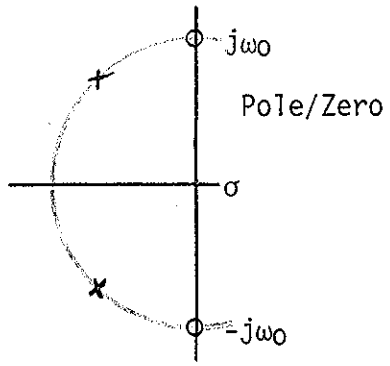
which has zeros at:

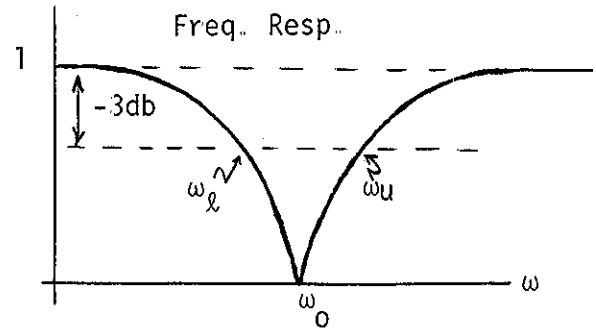
$$s_{z_1,z_2} = \pm j\omega_n \qquad (4\text{-}4)$$

plus the usual poles at radius ω_o. Because we have zeros at frequency ω_n, the response must go to zero there. However, usually implicit in the idea that a frequency component is to notched out, is the idea that all other frequencies should be passed. Accordingly we are concerned with the passbands on either side of the notch, and would usually like them to be as flat as possible, equal on both sides, and to extend up as close as possible to the notch itself. Fig. 4-3 shows a typical notch in terms of its pole/zero plot and its frequency response, for the case where $\omega_n = \omega_o$.

It can be seen from the response that we get a theoretically perfect notch at $\omega = \omega_o$. By placing the poles and zeros at exactly the same radii, the gain on both sides of the notch is the same, as can be seen by taking the limiting cases of equation (4-3), for dc and for infinite frequency. If $\omega_n \neq \omega_o$, the high frequency gain is 1, while the low-frequency gain is ω_n^2/ω_o^2. (It can be seen that this case corresponds to a second-order low-pass elliptic filter for $\omega_n > \omega_o$).

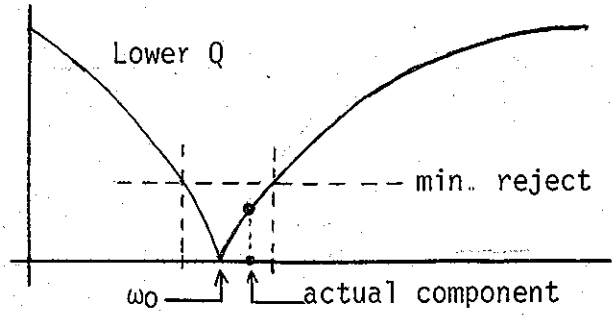
Fig. 4-3

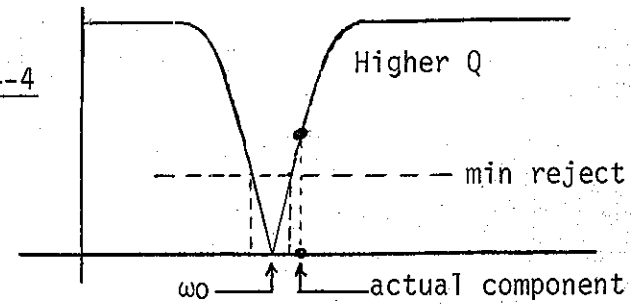




Thus placing the pole and the zero at exactly the same radius assures that the passbands are equal, but the Q of the poles also drastically influences the response shape. Poles that are real or highly damped will produce a wide and gradual notch, while poles approaching the jω-axis will produce sharp notches. This we can understand in terms of the pole close to the zero in effect "hiding" the notch from all but the frequencies closest to the notch. The notch then appears suddenly for a narrow frequency region. In general, we are looking for sharp notches, but this can also be overdone. This can be seen if we consider that in practice, we do not expect a perfect notch, but are satisfied with some sufficient measure of rejection. Any frequency region which is below this response level, and which includes the specific frequency to be notched out, will be satisfactory.

Fig. 4-4





In this view, a lower Q notch (poles back from zeros) is more forgiving of tuning errors than is a higher Q notch (poles up close to zeros), as is illustrated in Fig. 4-4. We note that an infinitely sharp notch would be infinitely difficult to tune.

Notch sharpness is measured in terms of the Q of the poles, much as bandpass sharpness is. In fact, the same "formula" for Q applies to the notch. We can show this by looking at the case of $\omega_o = 1$, and using equation (4-3), we have a frequency response of:

$$|T(s)| = (1 - \omega^2)\left[\frac{1}{\omega^4 + \omega^2[D - 2] + 1}\right]^{1/2} \qquad (4\text{-}5)$$

The 3db points are found by setting this equal to $1/\sqrt{2}$, which means that:

$$\omega^4 + D^2\omega^2 - 2\omega^2 + 1 = 2 - 4\omega^2 + 2\omega^4 \qquad (4\text{-}6)$$

This involves only even powers of ω, and can be solved using the quadratic equation to give:

$$\omega^2 = \frac{(D^2 + 2)}{2} \pm \frac{D}{2}\sqrt{D^2 + 4} \qquad (4\text{-}7)$$

which has the two solutions:

$$\omega_u = \sqrt{\frac{D^2+2}{2} + \frac{D}{2}\sqrt{D^2 + 4}} \qquad (4\text{-}8)$$

$$\omega_\ell = \sqrt{\frac{D^2+2}{2} - \frac{D}{2}\sqrt{D^2 + 4}} \qquad (4\text{-}9)$$

which are noted on Fig. 4-3. If the formula for Q here is to be the same as that for the bandpass, then it should be true that:

$$Q = \frac{\omega_o}{\omega_u - \omega_\ell} = \frac{\sqrt{\omega_u \omega_\ell}}{\omega_u - \omega_\ell} = \frac{1}{\omega_u - \omega_\ell} = \frac{1}{D} \qquad (4\text{-}10)$$

It is left to the reader as an exercise to show that if equations (4-8) and (4-9) are plugged into equation (4-10), that an identity results. Note however that the 3db side frequencies are measured down from the passbands, and not up from the bottom of the notch (as though the notch were an "upside-down bandpass"). We can now relate notch sharpness quantitatively to previous notions of the Q of the poles of the network.

No discussion of notch filters should omit the "classical" notch filter called the "Twin-T" which is shown in Fig. 4-5. This network is entirely passive, as shown, and yet does realize complex conjugate finite zeros on the $j\omega$-axis. [This can be done - it is complex conjugate poles that we can't get passively.]. The Twin-T is named because of the two T networks (one upside down as drawn here) that compose it. The transfer function can be found to be:

$$T(s) = V_{out}/V_{in} = \frac{s^2 + 1/R^2C^2}{s^2 + 4s/RC + 1/R^2C^2} \qquad (4\text{-}11)$$

which is seen to have zeros at $\pm j/RC$, and two real poles at $-0.268/RC$ and at $-3.732/RC$. Because of the real poles, this is not a very sharp notch. In fact, from equation (4-11) it is seen that the Q is only 1/4. In actual use, the network would probably require an op-amp follower at V_{out} to prevent loading.

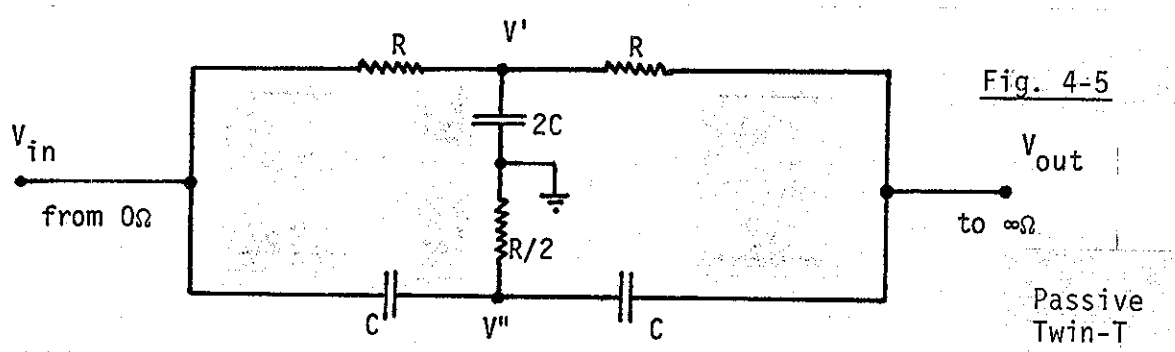


Fig. 4-5

The Q of the Twin-T notch can be enhanced by using positive feedback, which will require an active element, at least an op-amp buffer or two. This is illustrated in Fig. 4-6. The transfer function becomes:

$$T(s) = \frac{s^2 + 1/R^2C^2}{s^2 + s[4(1-K)](1/RC) + 1/R^2C^2} \qquad (4\text{-}12)$$

For this so-called "Bootstrapped Twin-T", the Q can now be enhanced, and is given by:

$$Q = 1/[4(1-K)] \qquad (4\text{-}13)$$

Thus K=0 gives the passive case, as expected, and K approaching 1 gives Q approaching infinity. In cases where K is fairly large, the output of A2 might be large enough to be used as the output, in which case, A1 might not be necessary, if R_1 and R_2 could be made large enough that they do not load down the passive part of the network. This method of using positive feedback to enhance the Q of a network is actually fairly general, and will be seen again for the case of bandpass filters (Section 5-2).

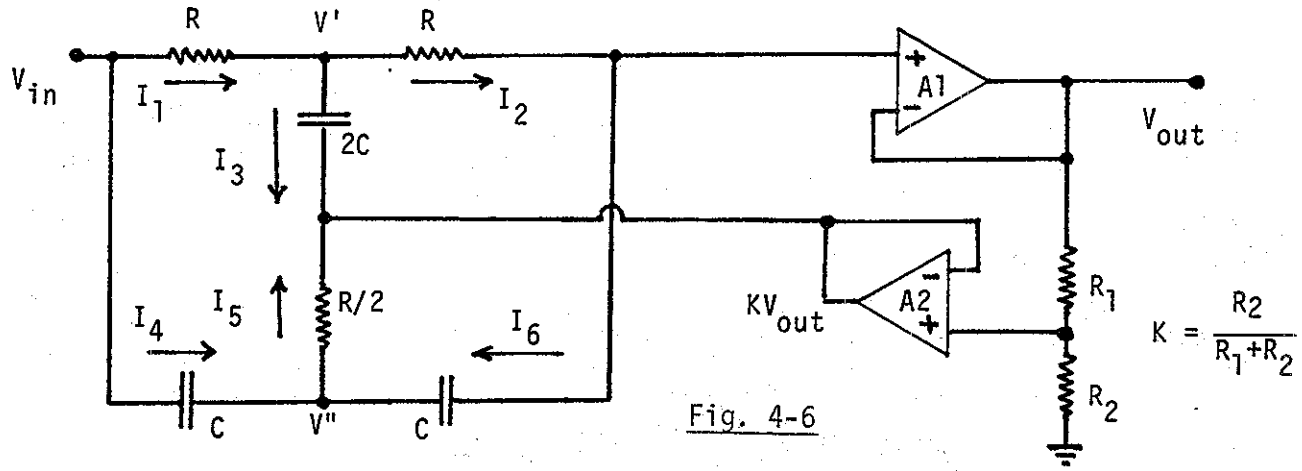


Fig. 4-6

Bootstrapped (Active) Twin-T

We will have a chance to look at more notch filter configurations in Chapters 5 and 6. Here we will look at an additional structure that also illustrates the use of a simulated inductance, a technique that will also be used in Chapter 5. Fig. 4-7 shows an op-amp structure that can be looked at based on the differential amplifier structure (Fig. 1-15). Here the inputs are connected together, and this would mean that there would be no output for any input. However, it is also seen that the lowest leg of the differential amplifier is not just a resistor R, but rather a R-L-C series circuit. We know that at the resonant frequency of the R-L-C series circuit, the impedances of the capacitor and inductor cancel, leaving only R. Thus we understand the notching out of the resonant frequency in terms of the R-L-C branch becoming purely resistive.

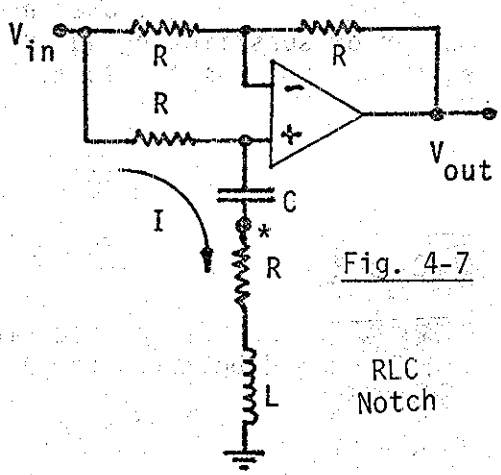


Fig. 4-7

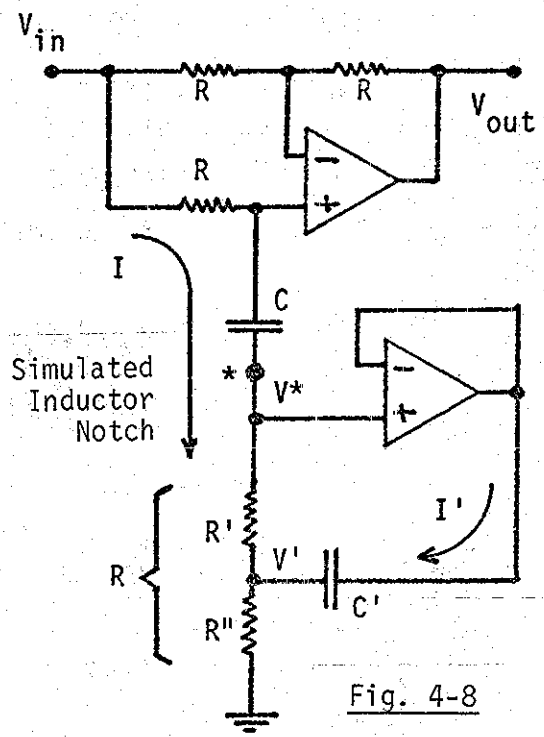


Fig. 4-8

Fig. 4-7 has transfer function:

$$T(s) = \frac{s^2 + 1/LC}{s^2 + (2R/L)s + 1/LC} \qquad (4\text{-}14)$$

which has a notch frequency of:

$$f_o = 1/2\pi\sqrt{LC} \tag{4-15}$$

and a Q of:

$$Q = (1/2R)\sqrt{L/C} \tag{4-16}$$

We can understand how the inductance is simulated here by examining Fig. 4-8, working with the structure below the node marked with a (*). We have three equations for the voltage V', as:

$$V' = (I+I')R'' \tag{4-17}$$

$$V' = V^* - IR' \tag{4-18}$$

$$V' = V^* - I'(1/sC') \tag{4-19}$$

Equations (4-18) and (4-19) are solved for I':

$$I' = IsC'R' \tag{4-20}$$

which can be put back into equation (4-17) to give:

$$V^* = I(R'+R'') + IsC'R'R'' \tag{4-21}$$

which leads to an impedance at the V* node of:

$$Z = V^*/I = (R'+R'') + sC'R'R'' = R + sC'R'R'' \tag{4-22}$$

This impedance has a resistive and an inductive term, since one term is multiplied by s. We can write this as:

$$Z = R + sL_{eq} \tag{4-23}$$

where the equivalent inductance is:

$$L_{eq} = C'R'R'' \tag{4-24}$$

Having found that the circuit below the (*) looks like a series R-L, we find that Fig. 4-8 is equivalent to Fig. 4-7, and we can plug L_{eq} into the design equations (4-15) and (4-16), giving.

$$f_o = 1/2\pi\sqrt{R'R''C'C} \tag{4-25}$$

$$Q = \frac{1}{2R}\sqrt{\frac{C'R'R''}{C}} \tag{4-26}$$

For example, if R'=R"=R/2, and C'=C, then the Q is only 1/4, the same as the passive Twin-T. In general, for higher Q's we must have C' greater than C.

Above we have introduced the ideas of notch filtering, and have shown how a passive notch works, and how a notch can be achieved with inductance simulation. In Chapter 5, we will see a notch based on the idea of bandpass subtraction. In Chapter 6, we will see how a notch can be achieved using a summation of a low-pass and a high-pass, each having the same denominator, thereby achieving the transfer function of equation (4-3) directly, which is easily achieved with the biquad approach presented there. In cases where multiple notches are needed, to remove harmonics as well as a fundamental, the delay line methods of Chapter 9 are useful.

4-3 ALL-PASS FILTERING, PHASE SHIFTING, AND PHASE DIFFERENCING:

All-pass networks have zeros that are placed as mirror images to their poles. They may be first-order or any higher order. Placement of zeros in this way means that the amplitude response is constant. The all-pass filter is used only for its phase properties. They may be used to improve the phase linearity of some other filter without changing its magnitude response. They are also used in parallel networks to realize a phase difference over a wide bandwidth.

Fig. 4-9 shows a first-order low-pass filter, for which the transfer function is easily obtained as:

$$T(s) = \frac{1 - sCR}{1 + sCR} \tag{4-27}$$

which has a pole at -1/RC and a zero at +1/RC, as shown in Fig. 4-9. It is easy to show that $|T(s)| = 1$. The phase is given by equation (1-27) as:

$$\phi(\omega) = \text{Tan}^{-1}\frac{\text{Im}\{T(j\omega)\}}{\text{Re}\{T(j\omega)\}} \tag{4-28}$$

and we need to put $T(j\omega)$ in a form with a real denominator, as:

$$T(j\omega) = \frac{1 - j\omega RC}{1 + j\omega RC} = \frac{1 - j\omega RC}{1 + j\omega RC} \cdot \frac{1 - j\omega RC}{1 - j\omega RC} = \frac{1 - 2j\omega RC - \omega^2R^2C^2}{1 + \omega^2R^2C^2} \tag{4-29}$$

for which equation (4-28) becomes:

$$\phi(\omega) = \text{Tan}^{-1}\frac{-2\omega RC}{1 - \omega^2R^2C^2} = -\text{Tan}^{-1}\frac{2\omega RC}{1 - \omega^2R^2C^2} \tag{4-30}$$

This is the correct answer. However, we would like to relate it to the angle α shown in Fig. 4-9. Geometrically, the phase shift should be α off the poles, and α off the zero, for a total phase of -2α. Here we see from the pole/zero plot that:

$$\text{Tan}(\alpha) = \omega RC \tag{4-31}$$

Consulting trig tables we find the useful identity:

$$\text{Tan}(2\alpha) = 2\text{Tan}(\alpha)/(1 - \text{Tan}^2\alpha) \tag{4-32}$$

from which we establish that:

$$\phi(\omega) = -2\alpha = -2\text{Tan}^{-1}\omega RC \tag{4-33}$$

which is a simpler form than equation (4-30).

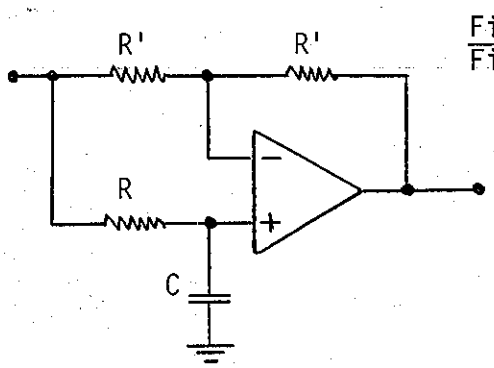
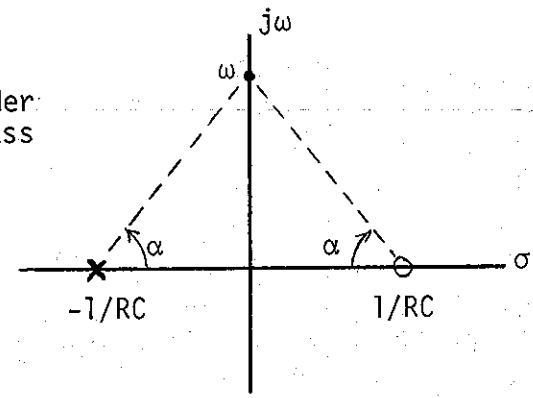


Fig. 4-9
First-Order
All-Pass



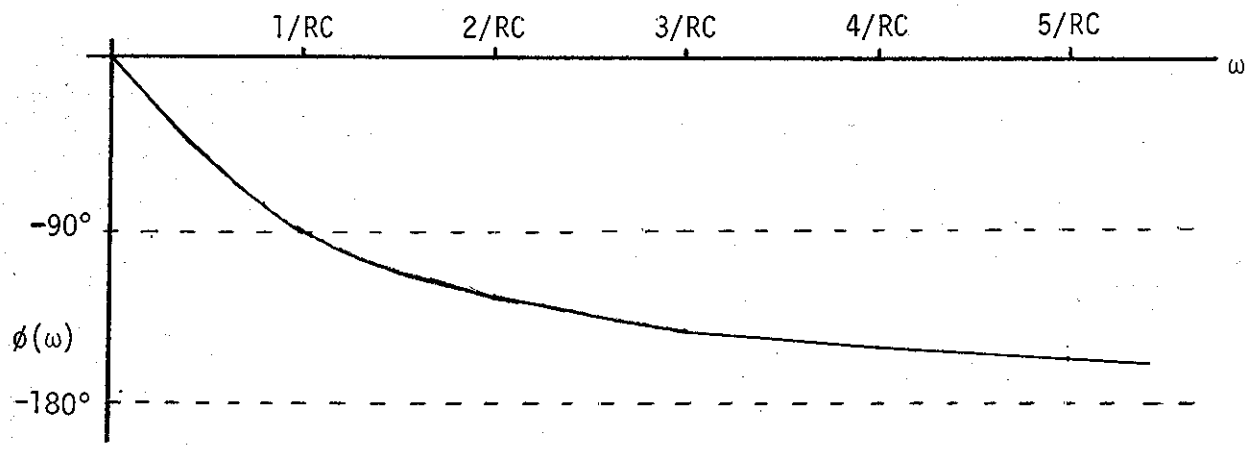


Fig. 4-10
Phase of 1st-order all-pass

This phase response is shown in Fig. 4-10. Above we have been a bit careless about the sign of the phase. This is a bit by intention by way of warning that we often have to specifically examine this issue. Some of the factors involved with determining phase are discussed below.

First, a pole/zero plot by itself only gives the phase to within an arbitrary inversion, much as a pole/zero plot by itself only gives the transfer function and frequency response to witnin an arbitrary multiplicative constant. This is because we can always put a (-) sign in front of T(s) and still get the same pole/zero plot. The second point follows directly on this inversion. Just what is an inversion? Is it a 180° phase shift? In fact, an inversion often looks like a 180° phase shift, and adds an apparent 180° to phase measurements, but they are not the same. This we can understand since a phase shift must always correspond to a time delay. If we simply invert a sine wave segment (Fig. 4-11), we get a response immediately. A device that provides a true 180° phase shift should also require some delay. However, it can also be seen that a comparision, in the absense of any "marker" that could be put on the waveform, looks like a 180° phase shift. A third factor is that there are usually different configurations that realize an all-pass of a given order, and that these often differ by an inversion. For example, we can modify Fig. 4-9 by switching the resistors R and capacitor C, and obtain a transfer function that is the negative of equation (4-27). Fourthly, there is often an ambiguity that results when we use a calculator or computer to calculate the inverse tangent function. For example, we may have a phase that exceeds 360°, and know that it is actually 395° but we still get 35° from the calculator. Alternatively, we may have discontinuities of 180° at places. Finally, conventions with respect to geometric interpretation of phase response may vary from author to author, and may be difficult to remember even at best.

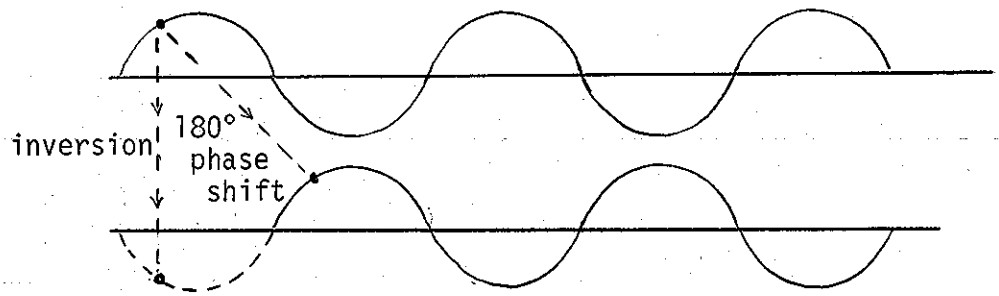


Fig. 4-11
Comparison of inversion and 180° phase shift

With all these possible sources of confusion, what can we rely upon? First, a phase shift should always be interpreted as a delay. We should never, for example, interpret an output as being ahead of an input by 25°, for example. It might be behind by 335°, or by 695°, for example, but never ahead. Secondly, it is often

useful, and sometimes absolutely necessary to take a careful look at the network to see what the phase is at zero frequency, to see if it is inverted to begin with. If so, we interpret our phase as starting at a 180° delay, and going from there. It is then wise to track this phase calculation for higher frequencies. Since phase can change very rapidly and very suddenly (as for example when we go by a high-Q all-pass pole/zero pairing), we have to be alert, and take phase measurement (or make calculations) at very small increments. In particular, we need to pick up the transitions from 360° back through 0°.

In addition to worrying about whether or not we have calculated, measured, or interpreted phase correctly, we need to talk a bit about when and why we should be worried about phase in the first place. This is probably particularly important at this point since in most other places we have concentrated on the magnitude response, and ignored the phase response.

Phase is important in that it corresponds to a time delay, and in most cases, a time delay that varies with frequency, and this can result in phase distortion. For example, it is not difficult to show that the phase contribution of a pair of poles, as represented by a denominator factor of $(s^2 + D\omega_0 s + \omega_0^2)$ is given by (see problems at end of chapter):

$$\phi(\omega) = \mathrm{Tan}^{-1}\left[\frac{D\omega_0\omega}{\omega_0^2 - \omega^2}\right] \qquad (4\text{-}34)$$

for which the corresponding time delay, called the "phase delay", is obtained by dividing by ω:

$$t_p(\omega) = \frac{1}{\omega}\mathrm{Tan}^{-1}\left[\frac{D\omega_0\omega}{\omega_0^2 - \omega^2}\right] \qquad (4\text{-}35)$$

Another useful notion of delay due to phase is the so-called "group delay" which is given by the derivative $d\phi(\omega)/d\omega$, or:

$$t_g(\omega) = d\phi(\omega)/d\omega = \frac{D\omega_0[\omega_0^2 + \omega^2]}{\omega_0^4 + (D^2-2)\omega_0^2\omega^2 + \omega^4} \qquad (4\text{-}36)$$

which gives a better idea how the phase varies with frequency at any particular frequency. It is easy to see that $t_p(\omega) = t_g(\omega)$ is a condition of linear phase.

One problem with phase distortion (see also the discussion of linear phase at the end of Chapter 3), is that certain sophisticated audio signals, such as speech and music, can take on what is often called a "muddy" sound. We can understand this if we consider that sounds such as music depend as much or more on transients as they do on the "steady state." There is a well-known "Ohm's Law of acoustics" (as opposed to the universally known electronic law by the same scientist!), which says essentially that the ear is "phase deaf." That is, the ear should hear the exact same thing regardless of the relative phase of the components making up a sound. Slight variations from this law are found even for steady state, but when it comes to transients, the law breaks down completely (Ohm of course had no way to test this). For example, a musical instrument might have a sound consisting of a fundamental and some harmonics, each of which are decaying exponentially, all being excited at the same initial instance, as seen in Fig. 4-12a. We come to identify the instrument with this particular "attack transient." If this signal is subjected to severe phase distortion, the upper harmonic may be so much delayed that the sound looses its identity (Fig. 4-12b). Accordingly, low phase distortion is something we often need in music handling circuits.

We thus sometimes need certain phase compensating circuits. These do not "reverse" any phase delay that has already occured, as this is not possible. Rather, they add phase, but in a way such that the total phase is now more linear. The total phase delay is longer, but it is more equal at the frequencies of most

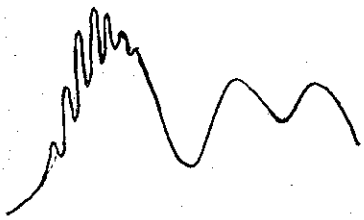


Fig. 4-12a A Normal Musical Transient

Fig. 4-12b A Phase Distorted or "Muddy" Musical Transient

interest to us. Such an improved phase is possible through the use of the various "multi-curved" phase responses available from different second-order all-pass networks, which we will look at shortly.

At other times, we are not particularly involved with audio circuits, and the phase response is less important. However, it must be remembered that in cases where signals are brought in reference to each other, as for example, when they are added or subtracted, the result depends critically on phase. Such a problem appears for example when one attempts to design an audio "graphic equalizer." The individual filters in parallel, which sort out the channels of the audio, may all have excellent phase properties. The assumption is then often made that the level in each of these channels can be adjusted individually, and the channels then recombined. In fact, when the signals are added, unless careful attention has been paid to the phase, the intersecting regions can have amplitude peaks and valleys that are unexpected and largely unacceptable. At other times, we need to pay attention to phase at summing points, as it will determine if we have a summation, or perhaps a subtraction. An example might be when a low-pass and a high-pass are added to form a notch. If the two are subtracted instead, real rather than imaginary zeros result.

One additional application of phase shifters or all-pass networks comes up in the need for phase-differencing networks. Such a network produces two different phases for the same input. Typically what is desired is a 90° phase shift that is used in certain modulation schemes. This is achieved with parallel all-pass networks, as will be discussed below.

As with previous filter types, we are interested in what can be done with second-order sections. A typical second-order all-pass transfer function is of the form:

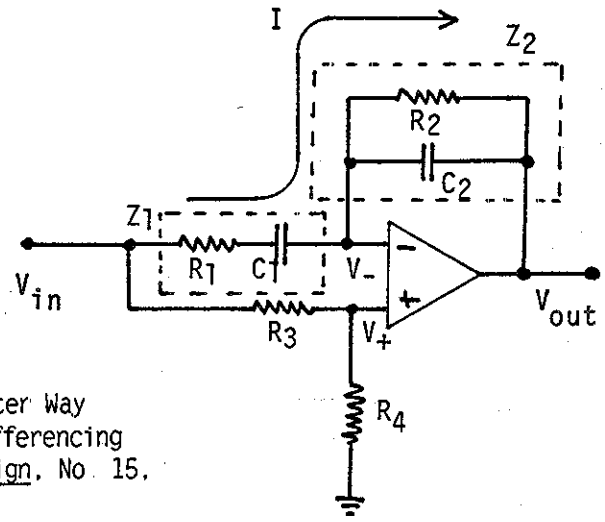
$$T(s) = \frac{s^2 - (D\omega_o)s + \omega_o^2}{s^2 + (D\omega_o)s + \omega_o^2} \qquad (4\text{-}37)$$

for which the numerator and denominator differ only in having middle terms of the opposite sign. If we just think about the quadratic equation, it can be seen that the poles and zeros will be mirror images in the case of complex conjugate poles, since the real part is determined by the middle term in such a case ($D<\sqrt{2}$). In cases where $D>\sqrt{2}$, where real poles occur, it can also be shown that the poles and zeros remain mirror images (see problems at end of chapter). As with the other filter types, there are several different possible configurations for second-order all-pass. We will look at one configuration in this chapter. Additional configurations are presented in Chapter 5 and Chapter 6.

Fig. 4-13 shows a second-order all-pass network due to A G. Lloyd which is capable of producing real mirror image poles and zeros. It is useful in the design of 90° phase differencing networks where it can replace two first-order networks, but otherwise use the same calculations and calculated component values. The network is in the form of a differential amplifier, with resistors R_3 and R_4 below, and the impedances Z_1 and Z_2 (as seen in Fig 4-13) above. We can easily write down the

Fig. 4-13

Lloyd's Second-Order All-Pass - Useful for real pole/zero sets.



from:
A.G.Lloyd, "Here's a Better Way to Design a 90° Phase-Differencing Network," Electronic Design, No. 15, Pg 78 (1971)

equations for the (+) and (-) inputs of the op-amp as:

$$V_+ = V_{in}R_4/(R_3+R_4) = KV_{in} \tag{4-38}$$

and:

$$V_- = \frac{V_{in}Z_2 + V_{out}Z_1}{Z_1+Z_2} \tag{4-39}$$

where:

$$Z_1 = R_1 + 1/sC_1 \tag{4-40}$$

and:

$$Z_2 = R_2/(1+sC_2R_2) \tag{4-41}$$

Setting $V_+ = V_-$, we can solve for T(s) as:

$$T(s) = \frac{K(1-sC_1R_1)(1-sC_2R_2) + [2sKC_1R_1 + 2sKC_2R_2 - (1-K)sC_1R_2}{(1+sC_1R_1)(1+sC_2R_2)} \tag{4-42}$$

The term in [ ] in the numerator disappears if:

$$K = \frac{1}{2(R_1/R_2) + 2(C_2/C_1) + 1} \tag{4-43}$$

Equation (4-43) shows that K will always be less than 1, and thus is always realizable with the voltage divider $K = R_4/(R_3+R_4)$. With K set according to equation (4-43), T(s) in equation (4-42) becomes:

$$T(s) = \frac{K(1 - sC_1R_1)(1 - sC_2R_2)}{(1 + sC_1R_1)(1 + sC_2R_2)} \tag{4-44}$$

What we thus obtain is the equivalent of two first-order all-pass networks in series, with an amplitude scaling (loss, in fact) of K. Note that equation (4-44) has a Q given by:

$$Q = \frac{\sqrt{C_1C_2R_1R_2}}{C_1R_1 + C_2R_2} \tag{4-45}$$

which has a maximum of Q=1/2 when $R_1C_1 = R_2C_2$. Accordingly only real poles are obtained (with corresponding mirror image real zeros) with this network. However, this is what we want with some applications (90° phase differencing networks). When complex pole/zero sets are needed, all-pass networks capable of producing Q's greater than 1/2 are found in Chapter 5 and Chapter 6.

A variety of phase curves can be obtained once we consider second-order all-pass networks with Q greater than 1/2, which have complex pole/zero sets, which are included in the general equation (4-37). Because of the mirror image zeros, the phase, phase delay, and group delay, are exactly twice the values given for the poles alone in equations (4-34), (4-35), and (4-36). Alternatively, phase curves can be obtained using the geometric method. In the geometric view, you begin at a point on the jω-axis where you need to know the phase, and measure angles to the poles and zeros. The reference line for angles is a line, starting at the jω point of interest, and running parallel to the real axis in the positive direction. Angles are measured counter-clockwise, and zero angles are added while pole angles are subtracted.

Fig. 4-14 shows curves for several second-order all-pass networks for a range of Q. A first-order all-pass is also given for comparison. We see that all the 2nd-order curves begin at 0, and run to 360°, all passing through 180° at $\omega = 1$. However, there is a substantial variation in the curves. In fact, curves for Q less than $1/\sqrt{3}$ all bend in only one direction, while curves for Q greater than $1/\sqrt{3}$ bend one way and then back. Note that the case $Q=1/\sqrt{3}$ corresponds to the linear phase case (see problems at end of chapter).

The sudden increase in phase that occurs near the pole/zero radius at higher Q's can be understood in terms of the very close pole/zero pairs that occur in such a case. At low frequencies, the poles and zeros tend to "hide" each other, and it is only when the frequency gets very close that they are suddenly "seen" and the major portion of their 360° phase shifting potential is used up suddenly. Thus a high Q "sharpens" a phase shift in the same general way it sharpens a bandpass or a notch. In particular, note the fact that in Fig. 4-14, within the rectangular region of

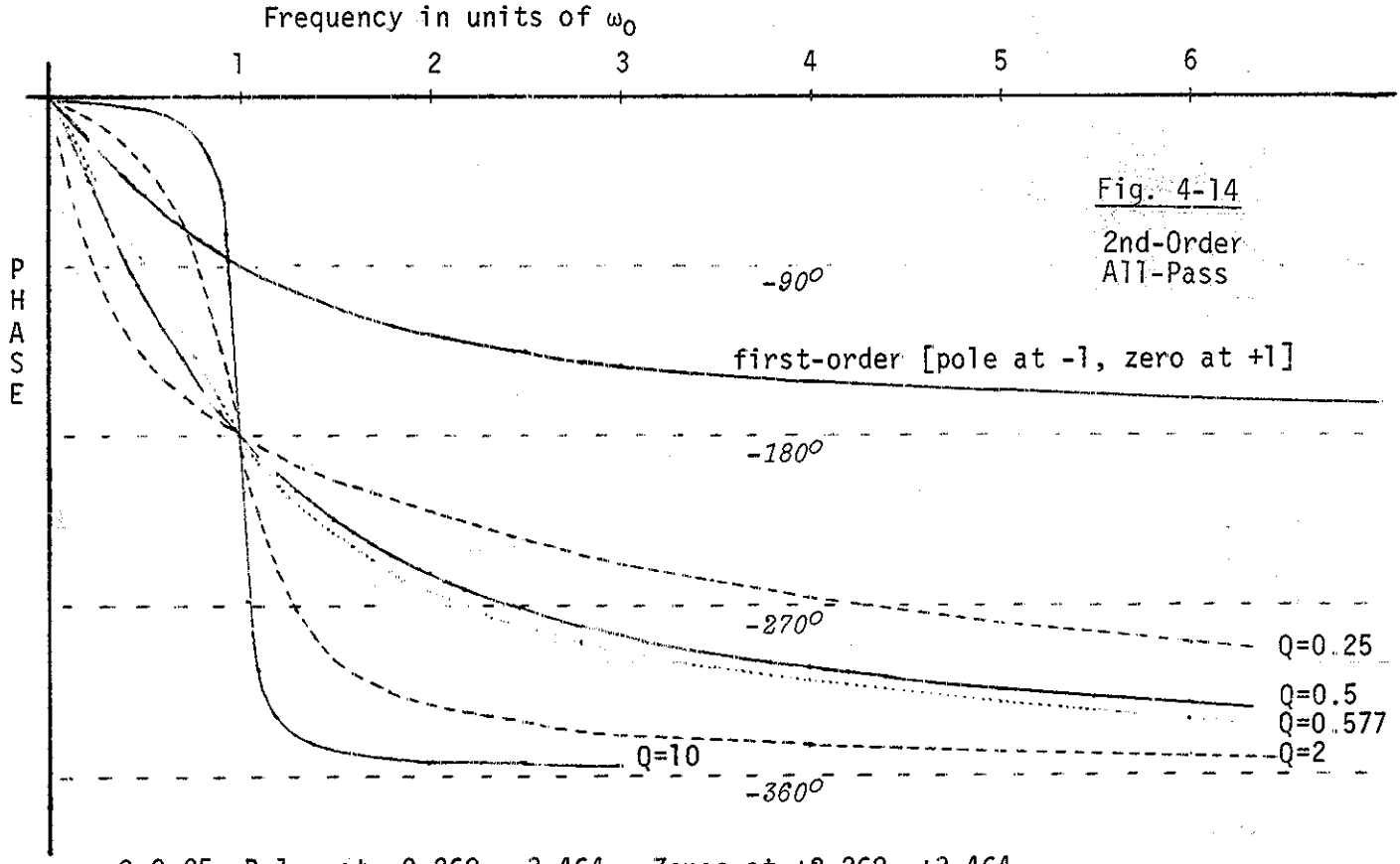


Fig. 4-14
2nd-Order
All-Pass

Q=0.25 Poles at -0.268, -3.464 Zeros at +0.268, +3.464
Q=0.50 Poles at -1.000, -1.000 Zeros at +1.000, +1.000
Q=0.577 Poles at -0.866 ± 0.500j Zeros at +0.866 ± 0.500j (Linear Phase)
Q=2.00 Poles at -0.250 ± 0.968j Zeros at +0.250 ± 0.968j
Q=10.0 Poles at -0.050 ± 0.999j Zeros at +0.050 ± 0.999j

frequency from 0 to 1, and of phase from 0 to 180°, that curves bending in both directions about linear phase are available to us for making phase corrections for other filters which use this passband.

An important application of all-pass networks is in the construction of 90° phase differencing networks. In many applications, we need to have sine and cosine versions of the same frequency component ("quadrature" components). With just a first-order all-pass, we can convert any single frequency component into a 90° phase shifted version simply by setting $1/2\pi RC$ to the frequency we are interested in (see Fig. 4-10). For a small bandwidth around this 90° frequency, the phase shift may be sufficiently close to 90° for some purposes. In many cases, we need a much wider bandwidth - perhaps a full audio bandwidth of say 1000:1 range. Integrators and differentiators can provide 90° shifts at wider bandwidths, but not with a flat amplitude response. We generally have to settle for a 90° phase differencing network. This is a network that takes an input and shifts its phase, often through many full cycles. In parallel with this first network is a second, which also provides a large shift, but which is adjusted to fall approximately 90° ahead or behind the first. Fig. 4-15a shows how two parallel all-pass networks are set up, and Fig. 4-15b shows the general idea of how the phase curves are arranged so as to produce an approximate 90° phase shift over a suitable bandwidth. Note that the phase shift from the input to the output is not 90° (in general) for either of the two networks, but the two outputs are approximately 90° apart.

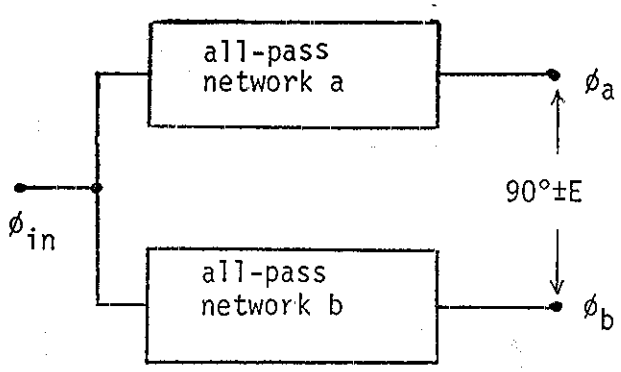


Fig. 4-15a

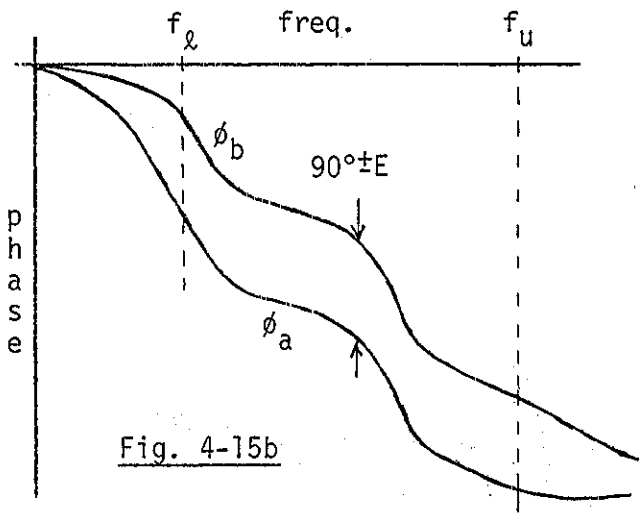


Fig. 4-15b

The achievement of a 90°±E phase difference over a certain bandwidth f_ℓ to f_u is not all that fundamentally different from other approximations that can be done, except here it is a phase difference that is of interest, and the phase error is more like the "ripple" in a case of amplitude approximation. A number of methods are available, and the one developed by Weaver* is widely used. The calculations needed for this method will be outlined here. One begins with the needed bandwidth ratio $B = f_u/f_\ell$, and then calculates a constant A as follows:

$$k = \sqrt{1-1/B^2} \tag{4-46}$$

$$L = \frac{1}{2}\frac{1-\sqrt{k}}{1+\sqrt{k}} \tag{4-47}$$

$$A' = L + 2L^5 + 15L^9 \tag{4-48}$$

$$A = e^{\pi^2/\mathrm{Log}_e A'} \tag{4-49}$$

From the value A, and the maximum error allowed (assumed here to be in degrees), the number of poles n that are required (for both networks) is given by:

*D K. Weaver. Proc IRE. 42 pg 671 April 1954

$$n > \ell n(E \cdot \pi/720) / \ell n\, A \qquad (4\text{-}50)$$

where E is the error and A comes from equation (4-49) above. Equivalently, the error E is given by:

$$E = 720A^n/\pi \qquad (4\text{-}51)$$

However, we can actually set aside the question of the error for a certain case as it is only the bandwidth B and the number of poles n that comes into our calculation of the pole positions. The poles are divided between the two networks, which we can denote as "a" and "b" by finding angles ϕ_a and ϕ_a' for network a and angles ϕ_b and ϕ_b' for network b, as follows:

$$\phi_a(r) = (45°/n)(4r-3) \qquad \text{for } r = 1,2,\ldots..(n/2) \qquad (4\text{-}52)$$

$$\phi_a'(r) = \text{Tan}^{-1}\frac{(A^2-A^6)\text{Sin}(4\phi_a)}{1 + (A^2+A^6)\text{Cos}(4\phi_a)} \qquad (4\text{-}53)$$

$$p_a(r) = \sqrt{B}\ \text{Tan}[\phi_a - \phi_a'] \qquad (4\text{-}54)$$

while the corresponding equations for the B network are:

$$\phi_b(r) = (45°/n)(4r-1) \qquad \text{for } r=1,2,\ldots..(n/2) \qquad (4\text{-}55)$$

$$\phi_b'(r) = \text{Tan}^{-1}\frac{(A^2-A^6)\text{Sin}(4\phi_b)}{1 + (A^2+A^6)\text{Cos}(4\phi_b)} \qquad (4\text{-}56)$$

$$p_b(r) = \sqrt{B}\ \text{Tan}[\phi_b - \phi_b'] \qquad (4\text{-}57)$$

Here we are assuming that n is even, which is the usual choice so that there are the same number of sections in each network. If n is odd, then equation (4-52) will run up to (n+1)/2 while equation (4-55) will run up to (n-1)/2. or vice versa. The poles as given in equations (4-54) and (4-57) are relative to f_ℓ, so for the actual pole positions, multiply by f_ℓ. Note that the poles are of course negative and real, and have corresponding zeros in their all-pass realizations.

Table 4-1 gives some pole position data for a number of representative cases, including the data for the example plot of Fig. 4-16, which also plots the error. Note that the bandwidth of Fig. 4-16 is 100:1 while the error is 1.64°.

Table 4-1

Data for 90° PDN's

Bandwidth: 1:10
Max Error: 1.1°
No. of Poles: 4

Poles A	Poles B
0.4788	1.8879
5.2968	20.888

Bandwidth: 1:100
Max. Error: 1.64°
No. of Poles: 6

Poles A	Poles B
0.5207	2.1246
6.0332	16.575
47.067	192.06

(example plotted in Fig. 4-16)

Bandwidth: 1:1500
Max. Error: 24°
No. of Poles: 4

Poles A	Poles B
1.3478	12.977
115.58	1112.9

Bandwidth: 1:1500
Max. Error: 0.26°
No. of Poles: 12

Poles A	Poles B
0.3846	1.3478
3.0076	6.2736
12.977	26.890
55.782	115.58
239.10	498.74
1112.9	3899.7

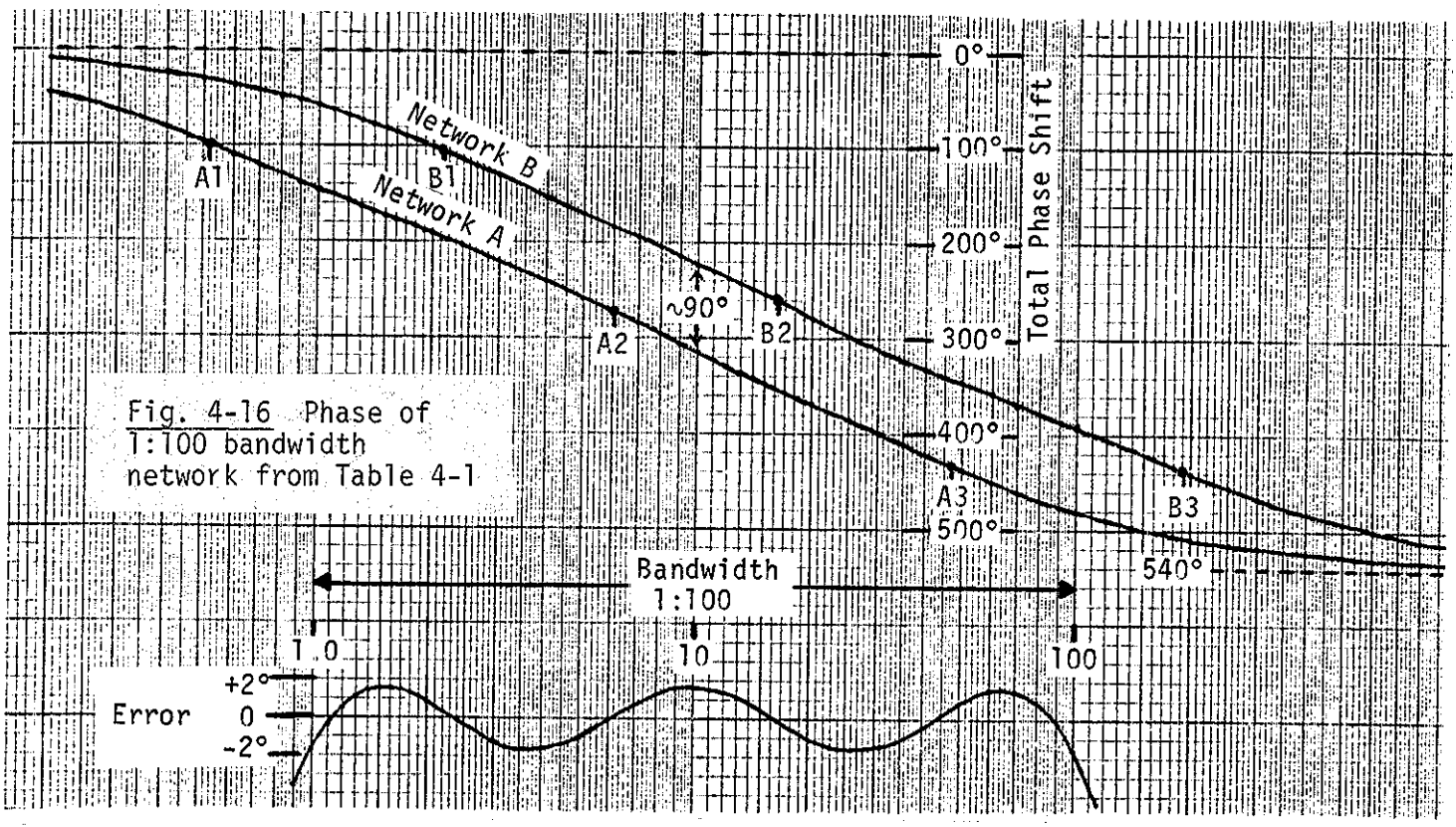
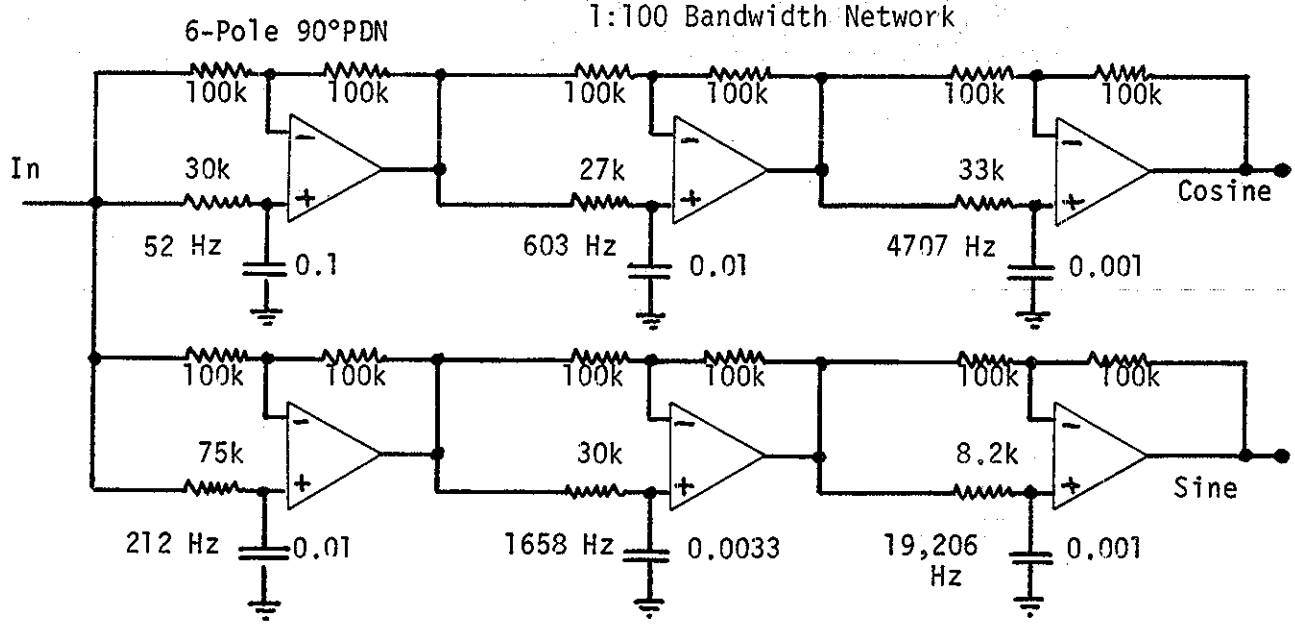


Fig. 4-16 Phase of 1:100 bandwidth network from Table 4-1

Fig. 4-17 shows a realization of the phase differencer using first-order all-pass networks. The bandwidth needed here is from 100 Hz to 10,000 Hz, so the poles given in Table 4-1, and the response curves of Fig. 4-16, should be scaled by a factor of 100. Note that the scaled pole frequencies are listed on Fig. 4-17. Fig. 4-18 shows a different network. This one is an eight-pole network intended for a bandwidth from 10 Hz to 15 kHz (an audio bandwidth) with a maximum error of 2.5°. Here, the second-order "Lloyd" all-pass is employed. Note that here it is necessary to group the poles into the networks shown for the most convenient values of K (note that the loss is only to about 3/5 of the original signal level).

Fig. 4-17 One Realization of 1:100 Bandwidth Network



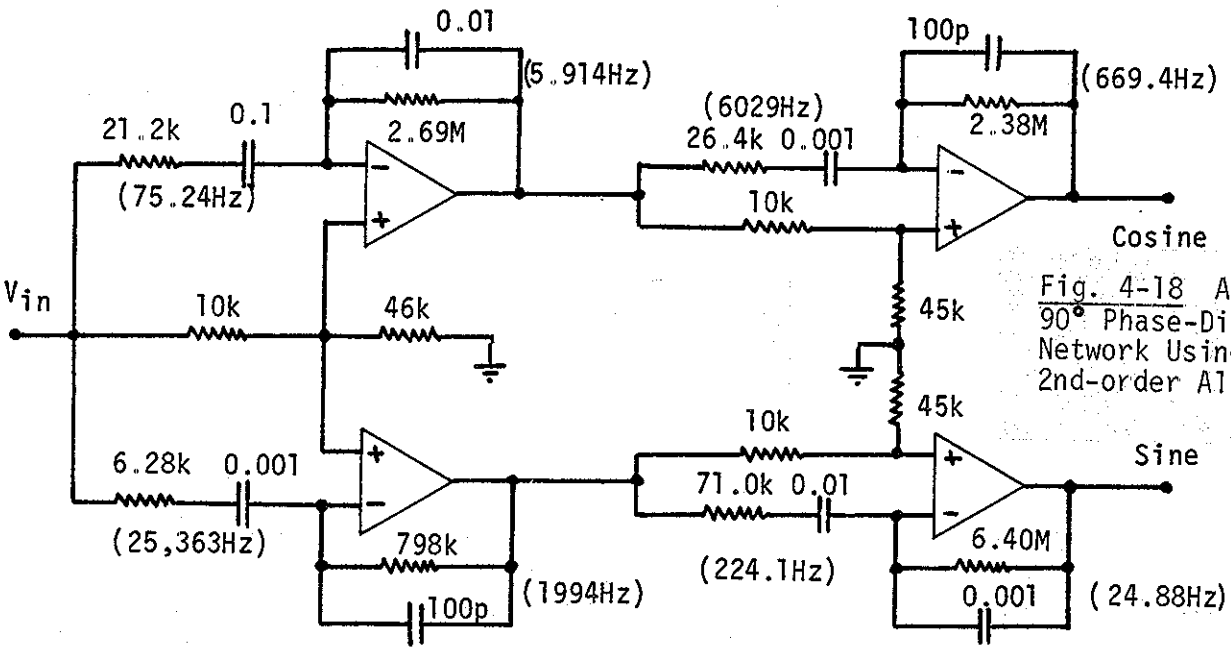


Fig. 4-18 An 8th-Order 90° Phase-Differencing Network Using Lloyd's 2nd-order All-Pass

One thing that is particularly notable about both the example networks, Fig. 4-17 and Fig. 4-18, is that there is a wide spread of capacitor values in the overall network, corresponding to the wide bandwidth and corresponding spread of pole radii. This can be a minor inconvenience in some circuit board construction, but it can be a major impediment to achieving a monolithic realization. One way to approach this problem is to group the poles into second-order sections in a particular way. This particular way is to combine the very lowest frequency pole with the very highest, the second from lowest with the second from highest, and so on. In this way, the ω_o term of these second-order sections becomes relatively constant. It then only remains to find a way of realizing this very low Q.

As an example, Table 4-2 gives the pole position data for the "a" network of the 12 pole network of Table 4-1, along with the values of ω_o and of Q.

Table 4-2

Grouping of Poles for Low Capacitance Spread

Poles Grouped	ω_o	Q
-0.3846, -1112.9	20.6	0.0186
-3.0076, -239.1	26.8	0.1108
-12.977 , -55.782	26.9	0.3913

This problem is a bit unusual first in that we are going to a second-order network not to obtain complex conjugate poles, but rather as a convenient means of achieving real poles. The requirement for low Q is also somewhat contrary to our usual filter design problems. In fact, the usual second-order all-pass networks do not lend themselves well to this grouping. However, a state-variable approach (Chapter 6) can be used to get the Q low enough.

This discussion of all-pass networks has prompted a general discussion of phase and why we are concerned with it. Before leaving the general subject it may be worth while to comment on how phase is measured. The Lissajous figure method is well known, and is reviewed in Fig. 4-19. Recall that the two phases are applyed to the X and Y inputs of a scope, and the resulting figure indicates the phase. The most notable Lissajous figures are the straight line at 45° (in phase), the straight line 45° backward (out of phase), and the circle (90° phase).

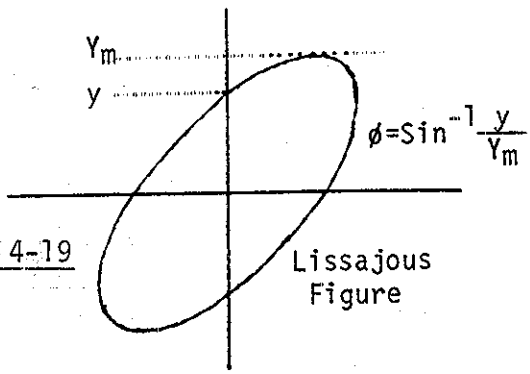


Fig. 4-19

It is rare to find a phasemeter in a lab, but it is very easy to construct one. Fig. 4-20 shows a simple phasemeter constructed from three op-amps, a CMOS Exclusive OR (EXOR or XOR) chip such as the type 4070, and a few other small parts. Fig. 4-21 shows how the circuit works. The first two op-amps square up the input waveforms, and the outputs are level shifted by the 10k resistors to a 0 and +15 level needed by the CMOS input. It can be seen by the example, that the difference in phase at the inputs results in a proportional pulse width at the output of the XOR gate. The RC low-pass then extracts the dc level from this pulse, and this is proportional to the phase. The phasemeter works over a range of 0 to 180°, giving 0 volts out for 0° phase shift, and +15 volts out for 180° of phase shift. Thus the constant is 12° per volt. Carefully done, the setup can be accurate to 1 or 2 degrees.

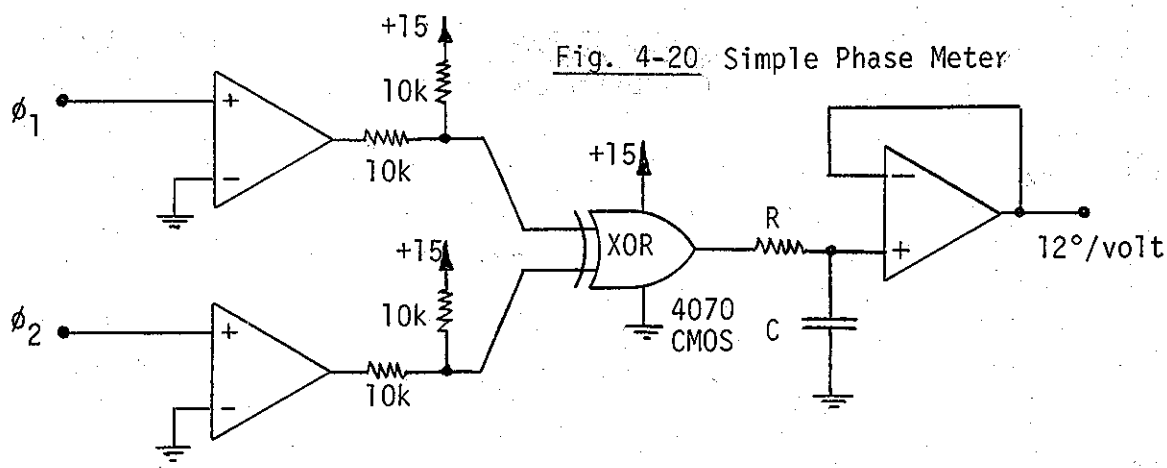


Fig. 4-20 Simple Phase Meter

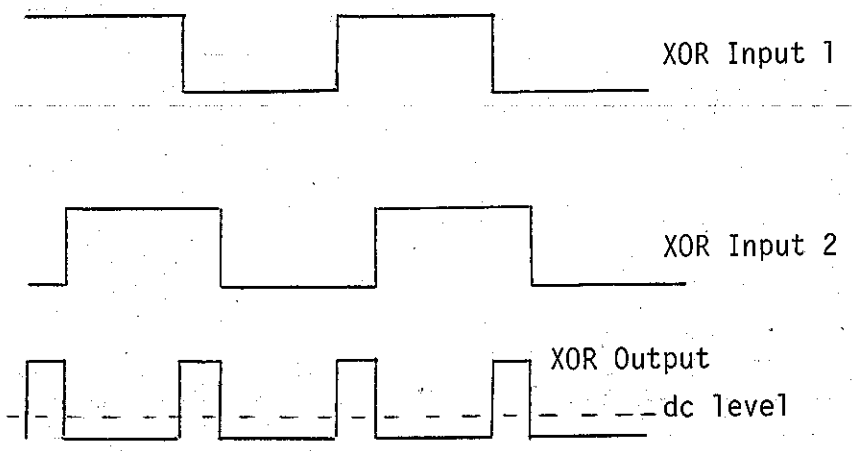


Fig. 4-21

Pulse "Duty Cycle" of XOR output is proportional to phase difference (0-180°) and this is proportional to dc level after low-pass filter.

---- A Butterworth that Peaks (Continued from pg. 2)

In the ASP notes, it has frequently been the case that we have noted that useful network configurations are easy to analyze because they inherently have recognizable, simpler, sub-networks. Here in the analysis of this fixed version, we recognize a familiar structure; in this case it is the first-order low-pass relating V' and V_{out}:

$$V_{out} = V'/(1+sCR) \qquad (1)$$

We next obtain the second of two necessary equation (two unknowns, V' and V_{out}) by summing currents at the V' node:

$$(V_{in}-V')/R = (V'-V_{out})/(1/2sC) + (V'-V_{out})/R \qquad (2)$$

It is then a simple matter of algebra to write down the correct transfer function:

$$T(s) = V_{out}(s)/V_{in}(s) = (1/2R^2C^2) \; / \; (s^2 + s/RC + 1/2R^2C^2) \qquad (3)$$

This has unity DC gain, as can easily be seen by looking at the limit as $s \rightarrow 0$. By comparing the denominator of equation (3) to the standard form for a second-order denominator:

$$s^2 + (\omega_0/Q)s + \omega_0^2 \qquad (4)$$

we see that the pole radius is $\omega_0 = 1/\sqrt{2}RC$, and the Q is $1/\sqrt{2}$ (damping $D = 1/Q = \sqrt{2}$) which is Butterworth. Thus the configuration of Fig. 1 is convenient in offering a Butterworth response at unity DC gain, at the price of having to find capacitors in a 2:1 ratio (or using three capacitors!). We only need to note that the 3db cutoff is $1/\sqrt{2}RC$ (not $1/RC$ as in the more familiar Sallen-Key structure). Of course, also keep in mind that the Butterworth cutoff, the -3db point, is the same as the pole radius, ω_0, but that this is in units of radians/sec. when you are talking about frequencies in Hertz, simply divide by 2π. In summary, this particular Butterworth implementation is a good, well understood circuit.

Fig. 2 shows Franco's voltage-controlled second-order section (Franco's unity gain buffers are shown here simply by the triangles with gain 1). The control elements are the popular CA3080 operational transconductance amplifiers (OTA's). These bear some resemblance to ordinary operational amplifiers, but must never be confused with them. Two things about OTA's that are fundamentally different are that the output is a current (not a voltage) and that the gain of the device is controlled by a controlling current I_C. Thus, we can use them as analog multipliers. The output current is obtained as a function of differential input voltage and control current as:

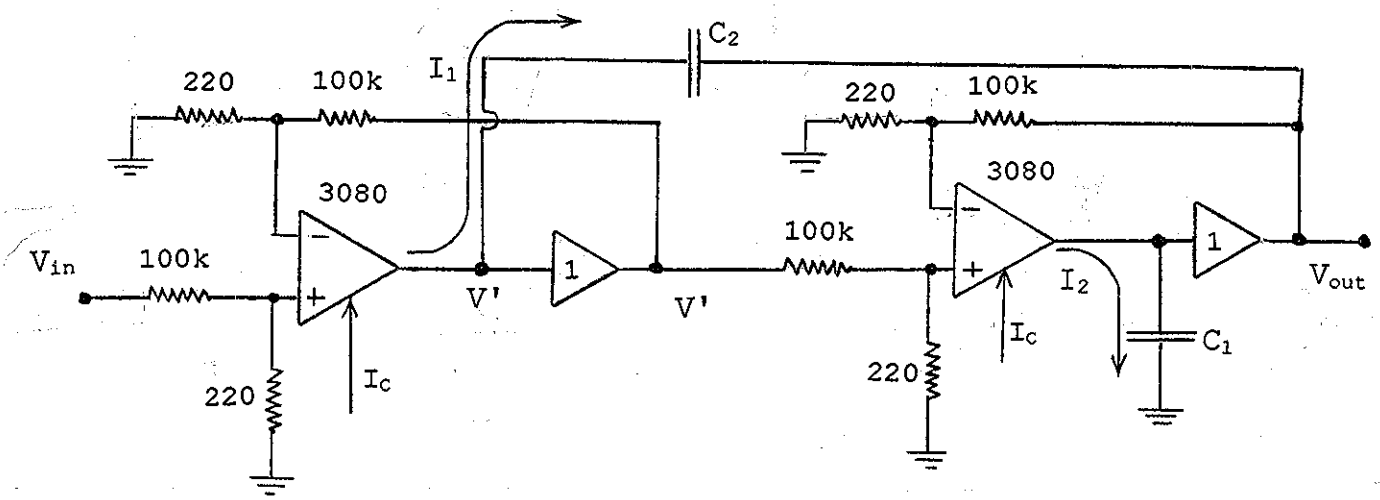


Fig. 2 A Voltage-Controlled Filter

$$I_{out} = 19.2 \cdot I_C \cdot E_{in} = V_{diff}/R_{eq} \quad (5)$$

where E_{in} is the differential input at the chip's pins, V_{diff} is the differential input at the attenuator inputs, and $R_{eq}=(19.2 \cdot I_C \cdot 220/100000)^{-1}$ is an equivalent resistance. The 220/100000 term in the expression for R_{eq} reflects the particular voltage divider resistors on the inputs. However, while we usefully employ the notation of an equivalent resistance, we need to note that the OTA configuration does not behave exactly like a resistor. The problem is that the OTA may source (current out) a current as though it had come through a resistor, but there is no corresponding current sink (current in) corresponding to the "other end" or an ordinary resistor. That is, a current must flow through an ordinary resistor, but the OTA is in general one sided.

Using equation (5) we can find the currents at the OTA outputs as:

$$I_1 = (V_{in} - V')/R_{eq} \quad (6)$$

and

$$I_2 = (V' - V_{out})/R_{eq} \quad (7)$$

These currents drive through capacitors to give:

$$V_{out} + I_1/sC_2 = V' \quad (8a)$$

$$I_1 = (V'-V_{out})sC_2 \quad (8b)$$

and

$$V_{out} = I_2/sC_1 \quad (9a)$$

$$I_2 = sC_1V_{out} \quad (9b)$$

Combining equations (6) with (8b) and (7) with(9b) we have two equations in two unknowns (V' and V_{out}) which can be solved for the transfer function:

$$T(s) = V_{out}(s)/V_{in}(s) = 1/R_{eq}^2C_1C_2 \;/\; (s^2 + s/R_{eq}C_2 + 1/R_{eq}^2C_1C_2) \qquad (10)$$

From equation (10) we see that the pole frequency is:

$$\omega_0 = 1/R_{eq}\sqrt{C_1C_2} \qquad (11)$$

and the all-important Q is given by:

$$Q = \sqrt{C_2/C_1} \qquad (12)$$

This answers our questions. In the original circuit, C_1=C and C_2=2C. Accordingly the Q is $\sqrt{2}$, and this would correspond to a second-order section with a peaking at about 1.5 times the DC gain. We might wonder if making the two capacitors equal might result in a Butterworth response, but this gives Q = 1 which has a peaking at about 1.15. Clearly, it is the case where C_1=2C and C_2=C that give $Q=1/\sqrt{2}$ that gives Butterworth. Thus Getting Butterworth in the voltage-controlled case involves a placement of capacitors C and 2C that is exactly the opposite of the placement in the non-voltage-controlled case.

What went wrong in 1974 and 1975? Who can say for sure? Did Sergio simply mislabel the capacitors? Was he fooled by the active filter case? Was it the case that the peaking just wasn't noticed in a musical test (indeed, the peaking generally has a favorable effect during synthesis - witness the Moog four-pole low-pass). As for the repeat of the mistake in the MEH, I can say for sure that the active version was in my mind, and that this was probably not tested on the bench. Further, the theoretical analysis, as done so easily above, was not as automatically exercised in 1975 as it was a few years later as we learned more about using OTA's.

Does it matter? Well, apparently not much - people were using, almost exclusively, four-pole low-pass and state-variable configurations. I suppose it is possible that no one tried it, or at least noticed the peaking, until Ian did. What we can do is make a note that the voltage-controlled resistors are sometimes one way (as noted above) and that that makes a difference. Generally, nothing is as easy as it seems at first! Thanks for reminding us of this, Ian.

Electronotes, Vol. 19, No. 193, March 2000
Published by B. Hutchins, 1016 Hanshaw Rd., Ithaca, NY 14850
(607)-257-8010