Laboratory Problems and Examples in Active, Voltage-Controlled, and Delay Line Networks

-by Bernie Hutchins

Published by Electronotes, 1 Pheasant Lane, Ithaca, NY 14850

LABORATORY PROBLEMS AND EXAMPLES IN ACTIVE, VOLTAGE-CONTROLLED, AND DELAY LINE NETWORKS

-by Bernie Hutchins

CONTENTS:

LOCATION OF EXPERIMENTS:

LOCATION OF EXERCISES:

ACKNOWLEDGMENT

The author has benefited greatly from many discussions with Professor Walter Ku and his students at Cornell University

GUIDE TO THE MANUAL - FOR INSTRUCTORS AND STUDENTS

This book is neither a text nor a lab experiment manual - it is somewhere in between, and perhaps could be used as either. It can be read as a text because there is sufficient experimental data given so that the reader will get the feel for the experiments even if he does not actually do them. It is also possible for a reader to do his own experiments at home based on the suggestions in the manual, so an actual classroom setup is not required. My own preference would be for the book to be used by the lab instructor before the lab to prepare the experiments to be performed in lab. The students will use the book in their write-up. It is unrealistic to suppose that students will have read the experiments before entering the lab, and much lab time will be wasted unless the instructor prepares some sort of "recipe" for the experiments. Furthermore, the instructor will be able to write up the lab instructions to fit exactly his available components and equipment (and available lab time).

Some instructors perhaps do not believe in making things too easy for the student in lab for fear the student will not learn anything. This view is utter nonsense. The student will learn much more from a well structured plan that is carefully thought out, one which will give him interesting results which he will enjoy presenting in his report. The "philosophical" approach, that things are not easy in the real world, and that the lab should reflect the real world, is not a valid one. In the first place, the student knows things are tough on the outside (that's probably why he is in school), and even the best laid plans .... so you need have no fear that the lab will not provide a degree of stimulating adversity. So make things easy for the student. Give him a reasonable recipe, and all the necessary parts in a neat little bag (which will also help control the mess in the lab and sticky fingers).

Much of the presentation is organized with a "look ahead" approach. We force the reader to become involved with something, often by step-by-step example, before stating the principles involved in a complete manner. This permits the reader to follow the text and perhaps grasp the principle himself before it is presented to him, which leads to lasting understanding. In the author's view, this is more useful that presenting the principle at one point and making the application appear later, or using a "mystery story" approach. Accordingly, the reader should not feel uneasy if his understanding is only superficial at first encounter. Complete understanding will often come later in the text, from the classroom, from other literature, and often, through experiences years later.

This manual is structured and paced with the idea in mind that the reader will also be involved in a classroom course on network theory. This means that there is much more detail in the presentations of theory in the earlier part of the manual, and less toward the end. Accordingly, the independent reader may have to spend more time working out details which the classroom student will get from lectures. Also, in the earlier part of the manual, exercises are scattered into the text and should be considered as the reader encounters them as a means of checking his understanding. Toward the end, exercises are omitted since the main presentation has plenty of details to be worked out as the reader goes along, or exercise material appears as part of the write up of experimental results. In regard to the lab problems, the first two experiments are fairly easy and are mainly to familiarize the student with the lab equipment and with IC op-amps. Later labs are more involved and will take more than one work session, giving the student a chance to play around the first session, and get the work done in the second or third session after he gets a feel for what is to be accomplished and knows what will be required to answer the questions presented. Instructors will probably want to go over the experiments suggested in this manual and rework the experimental instructions to fit the particular situation. Certainly some students with superior ability will be able to omit some of the experiments listed and spend more time on independent projects.

In addition to standard resistors, capacitors, and op-amps, the only other components required are the CA3080 "transconductors" for Experiment 5, and some sort of delay lines for Experiment 6. The delay lines should probably be available to the student on an "evaluation board", and not just as the IC's. At the time of this writing, it would be tedious to set up the associated clocks, level shifters, filters, and so on that are needed with these delay lines, so this should be done for the student. The instructor should consult the various manufacturers of delay lines to see what is currently available.

Students and instructors are invited to contact us at Electronotes if they have need of literature, supplies, problem solutions, or whatever might be useful which we can help with.

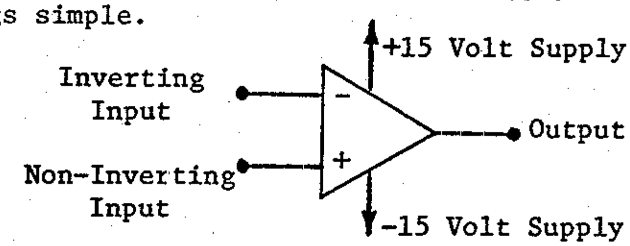
Bernie Hutchins
Electronotes
1 Pheasant Lane
Ithaca, NY 14850
607-273-8030

Chapter 1: Ideal Op-Amps And Basic Op-Amp Circuits

1A. THE IDEAL OP-AMP

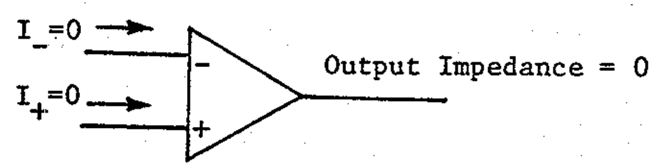
The ideal op-amp is a three terminal device which has an output, and two inputs. One input is inverting (-) and the other is non-inverting (+). The reader will note that the term "non-inverting" is in a sense a double negative, and this indicates that engineers have come to view "inverting" as somewhat the normal case, and have come up with a negative version of this term to describe what would otherwise be considered the normal case. The fact is that most amplifying circuits use some form of negative feedback to achieve the desired response and the necessary stability. Thus engineers come to think of inputs which cause an inverted response at the output to be the most useful. In Fig. 1 we show the op-amp as a three terminal device in its most common representation with a triangular symbol. We also show two additional terminals which provide the power to keep the device operating. These are commonly +15 and -15 as shown. It is common practice to leave these power supply terminals out of schemetic diagrams to keep things simple.

Fig. 1



The ideal op-amp has several additional idealized properties which are approached to a given degree by any real op-amp of the type you can go out and buy. We will first consider the property of infinite input impedance. Alternatively, we just say that the actual inputs (+ and -) of the op-amp draw no current, but just respond to the voltages applied to them. Still another way of saying the same general thing is to say that the inputs do not load the circuit driving them in any way. The main purpose of this assumption is that we can say where current flowing in the circuits around the op-amp does not go - it does not go in the inputs of the op-amp. Another assumption about ideal op-amps that we want to get out of the way is the assumption of infinite bandwidth and slew rate. This means that the device will work as expected regardless of input frequencies and waveforms. That is, the device has unlimited frequency response and power driving ability. Another common assumption is that the output impedance of the ideal op-amp is zero. That is, we would expect to be able to apply any load to the output of the ideal op-amp and have it drive that load. The implication is that the output can source or sink any necessary current to maintain the output voltage required. These assumptions are summarized in Fig. 2.

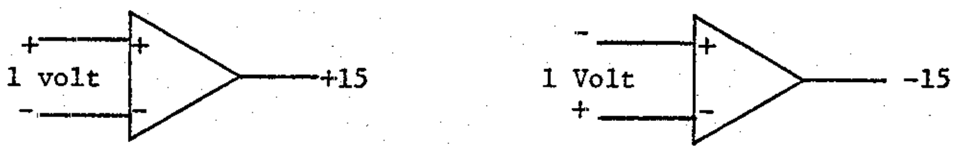
Fig. 2



By far, the most interesting property of the ideal op-amp (the one that has the most important consequences in the analysis of active networks) is the assumption that the amplifier has infinite gain. Infinite gain is a little hard to accept, and many designers will not find the actual gains of real op-amps, which may be a million or more, much easier to conceive of. This is not because the numbers are so large (except for the infinite one), but because circuits generally require finite gains and in fact, fairly low ones in the range of 10 or 100 or 1000, - something like that. Thus, using a very high gain stage may seem like the hard way of doing things. Why not just use what you need and stop there? Well, the point is that we will be looking at designs that use negative feedback to set the desired gains, and where the actual gain of the device is not important as long as it is large enough, and as long as we are only looking at the gain performance, and not at things like distortion.

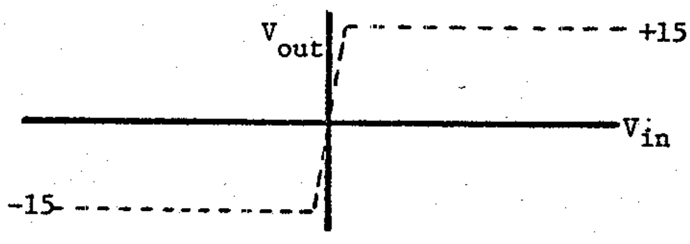
Now, the exact meaning of infinite gain is as follows. The output voltage of an amplifier with a differential (+ and -) input, as in the ideal op-amp, is the input voltage multiplied by some number A. The input voltage is the "differential input voltage," the difference between the voltage applied to the + and - inputs. The meaning of infinite gain is now seen to be the case where A goes to infinity. Right here it is best to jump in and say that we are not going to get any infinite voltages out of op-amps, even ideal ones. The op-amps are powered by real power supplies which have voltages like +15 and -15. Thus the actual limits of the output, even for infinite gain, are still the power supply voltages. But this limitation on the output still does not solve the difficulty with the concept of infinite gain. Suppose the differential input voltage is +1 volt as shown in Fig. 3. The infinite gain of the op-amp will drive the output to the positive supply voltage (+15 volts). Fig. 3 also shows the the case where the differential input voltage is negative (-1 volt) and the output is driven to -15 volts. This is in fact a useful circuit, although not generally for active networks. It is called a comparator, or a zero-cross detector in this case. If the differential input voltage is positive, the output is +15, and if the differential input voltage is negative, the output is -15. Thus we arrived at a circuit for making a decision of the either-or type, but how can this be an amplifier? At this point the reader must consider that since the output is either +15 or -15 for any non-zero input, that the output can be at a value in between +15 and -15 only if the differential input is zero. This is the most significant consequence of infinite gain that we will discover.

Fig. 3



Some may object to the fact that we have multiplied zero by infinity and arrived at some finite value (zero?) instead of at infinity and insist that to do this evaluation properly it is necessary to take limits as one value approaches zero and the other approaches infinity. Here we shall approach these limits by looking at what we might expect from a real op-amp. Let's assume the gain is 1,000,000. Thus, a differential input of 15 microvolts is enough to pin the output at the power supply limits. Thus there is a very small linear region around zero differential input. This is indicated by the curves of Fig. 4.

Fig. 4

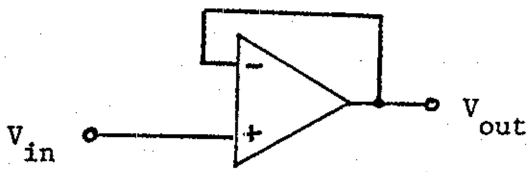


For a real op-amp, it will be virtually impossible to make the op-amp operate in the linear region by any manual means. If you try to supply small voltages about zero by turning a knob and varying the voltage, it will just jump away from you every time. It is likewise even less likely that you could get the amplifier operating in the linear region by trimming a circuit and then expect the circuit to be stable thereafter. Thus even with the real op-amp we are led to the conclusion that the output is either at +15 or - 15, or else the differential input voltage must be zero. We have not as yet devised a means for keeping the differential input voltage zero. In the real op-amp, we can not zero the differential input by shorting the inputs together, because there is always some real offset voltage between them. In the ideal op-amp, we can get the output to zero by shorting the inputs, but this is not very useful since we would then have only the choice of three voltages for the output (+15, 0, and -15). We will soon see that the differential input voltage can be easily maintained at zero by use of negative feedback, either in the case of the ideal op-amp, or to a good approximation with a real op-amp. Before doing this however, it is useful to speculate for a moment as to whether there is a linear region in the ideal op-amp. In fact, as we take the limit as A goes to infinity, we see that a linear region remains (assume the crossover region in Fig. 4 gets steeper and steeper as A increases). Thus, infinite gain can be thought of as implying an infinitesimal linear region. In fact, it won't matter much in what we do below whether we think of the op-amps shown as being ideal with infinite gain, or just real op-amps with very large gain.

1B. NEGATIVE FEEDBACK AND THE FOLLOWER CIRCUIT

The basic op-amp "follower" circuit is shown in Fig. 5. This circuit will be seen to provide at its output the same voltage that appears at its input. It is thus useful as a "buffer" since its input will not disturb the circuit it is connected to (due to the high input impedance) and the output will drive into whatever load is required (due to the very low output impedance). This circuit has 100% negative feedback - the output is connected directly back to the inverting input. We can now see how this negative feedback is used to maintain a zero differential input voltage. Suppose first that the output voltage is slightly more positive than the input. Thus a

Fig. 5



finite differential voltage appears at the input of the op-amp. Since this is a positive voltage applied to the inverting input, the very high gain of the op-amp will drive the input down, removing the small positive excursion. If on the other hand, the output wanders slightly negative, a similar argument shows how the output is driven positive. Thus, the only stable point is where the differential input is zero, and this means that the output follows the input.

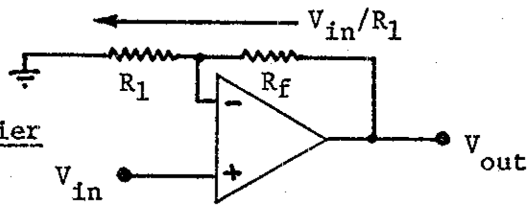
It will be left to the reader to show that similar correction processes take place in other op-amp configurations employing negative feedback. Thus, we will always assume that the differential input voltage is zero as long as there is negative feedback and as long as the negative feedback is allowed to work properly. There are cases with real op-amps where negative feedback is not allowed to work properly. For example, if the output must assume a value outside the limits of the power supply voltages, the negative feedback will fail. Likewise, if the output voltage can not change rapidly enough to assume the necessary value to zero the differential input, negative feedback will fail. We should always keep in mind that the op-amp action is a team effort between the high gain which amplifies the input differential voltage, and the output voltage which is thereby adjusted to zero this differential. However, it is the fact that the output assumes whatever value that is necessary to zero the differential input that is the easiest to accept. Also bear in mind that the two inputs are not necessarily zero, just that the two voltages are the same, so their difference is zero.

1C. THE NON-INVERTING AMPLIFIER

Next we want to get a circuit which will provide some actual voltage gain instead of just following the input voltage. We use the non-inverting amplifier shown in Fig. 6 for this purpose. The analysis of this circuit is simply a matter of applying the principle of zero differential input voltage, and using the fact that the actual op-amp inputs draw very little current (ideally zero). The (-) input is thus at the same voltage as the (+) input, which is the same as the actual input voltage. Thus the current flowing through R_1 to ground is V_{in}/R_1. This current must be flowing simply because Ohm's Law must be obeyed. This current is certainly not coming out of the (-) input of the op-amp, so the same current must be flowing into the node associated with the (-) input, entering by way of resistor R_f. This means that a voltage equal to $R_f(V_{in}/R_1)$ must be developed across the resistor R_f. The output voltage is thus determined by starting at the (+) input which is at voltage V_{in}, and the (-) input is also at V_{in}. The output is thus at $V_{in} + V_{in}(R_f/R_1)$ so the gain of the amplifier is:

$$G_{ni} = (1 + R_f/R_1)$$

Fig. 6 Non-Inverting Amplifier



We thus see how negative feedback is used to control the very large gain of the op-amp. We have only to set a resistance ratio to set the gain we want. Note that the non-inverting amplifier has a "residual" gain of one, which is the gain of the follower we studied above. Thus, gains of less than one are not possible unless an attenuator is placed on the + input of the op-amp. The input impedance of the non-inverting amplifier is very high since the signal is applied directly to an op-amp input.

1D. THE INVERTING AMPLIFIER

The inverting amplifier configuration of the op-amp is shown in Fig. 7. Note that in this circuit the (+) input is grounded. Thus when negative feedback is working properly, we expect that the (-) input will also be at ground potential. Yet it is also evident that there is no actual current path to ground since no current actually enters the (-) input of the op-amp, and the other two possible current paths go elsewhere (through resistors R_1 and R_f). Thus the (-) input is an unusual structure somtimes called a "virtual ground" and for reasons which will become apparent, it is also called a "summing node." The term "virtual ground" comes from the fact that the (-) input remains at ground potential even though no current actually flows to ground by this means. It should be realized that the "virtual ground" is just a special case of the more general principle of zero differential input voltage resulting from negative feedback.

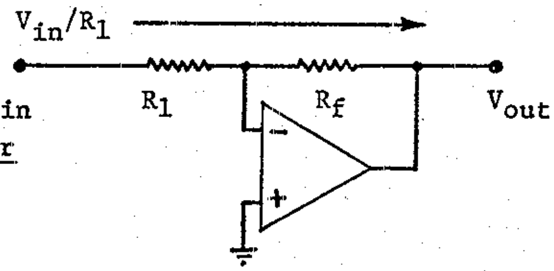


Fig. 7 Inverting Amplifier

Since the (-) input terminal is at ground potential, a current V_{in}/R_1 must be flowing through resistor R_1. This current has no place to go except out through the resistor R_f. Thus, a voltage $R_f(V_{in}/R_1)$ is developed across R_f. Since it is the usual convention to regard current as flowing from + to -, and one end of R_f is at ground potential, the output clearly assumes a negative potential, assuming V_{in} is positive. If V_{in} is negative, the direction of the current shown is reversed. Thus we have an output voltage of $-V_{in}(R_f/R_1)$ and the gain is:

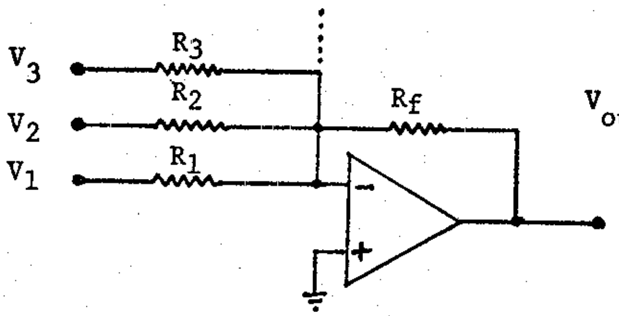
$$G_i = -R_f/R_1$$

It is essential to be very clear about what is actually going on in the inverting amplifier. Clearly, the (-) input is at virtual ground only because the output takes on whatever voltage is necessary to make this so. There is no reason to deal with the difficult impression of a current forcing its way into the (-) input and out through a feedback resistor, forcing the output to take on a certain value. It is in fact the output that does all the work. It takes on a value that keeps things "happy" at the (-) input, and this means a zero differential voltage. The reader can easily see for example that if the output voltage wanders positive, this will be fed back to add a positive increment to the voltage on the (-) input, and this increment will be amplified by the very high gain of the op-amp into a negative excursion at the output, thus driving the output back. A similar negative excursion at the output will be corrected in a similar manner. Thus, the negative feedback leads to a single stable state - that where the (-) input is at ground potential.

The reader will note that since the (-) input is a virtual ground, it is possible to attach more than one input resistor to it. In fact, it is only the current that enters the "summing node" that matters. This current can enter through resistors, or from any sort of current source. The total current into or out of the summing node must of course be zero (as with any node), and the excess current flows out through the feedback resistor. This makes possible the popular inverting summer configuration shown in Fig. 8.

Fig. 8

Inverting Summer



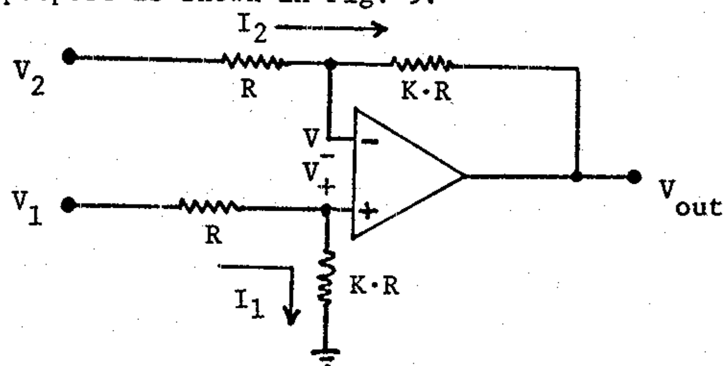
$$V_{out} = -V_1\frac{R_f}{R_1} - V_2\frac{R_f}{R_2} - V_3\frac{R_f}{R_3} - \cdots\cdots$$

1E. THE DIFFERENTIAL AMPLIFIER

The op-amp by itself is a differential amplifier of very high gain as we have seen above. In many cases, we want to have a circuit which takes the difference between two input voltages and multiplies this difference by a constant (often the constant is one). A circuit for this purpose is shown in Fig. 9.

Fig. 9

Differential Amplifier



Analysis of the circuit is similar to the analysis of the circuits above. The first thing to note is that the voltage V_+ is just determined by V_1 and the voltage divider on the (+) input. Thus $V_+ = V_1KR/(R+KR)$. Since there is negative feedback in the circuit, this is also the value for V_-, and consequently a current $(V_2-V_+)/R$ is flowing through the resistor R attached to V_2. This same current is then seen to be flowing out through the feedback resistor KR, and this allows us to calculate the output voltage as follows:

$$V_+ = \frac{V_1K}{1 + K}$$

$$I_2 = \left[V_2 - \frac{V_1K}{1+K}\right] \cdot \frac{1}{R}$$

$$\begin{aligned} V_{out} &= V_+ - I_2KR \\ &= \frac{V_1K}{1+K} - \left[\frac{V_2 + KV_2 - KV_1}{1+K}\right] \cdot \frac{KR}{R} \\ &= \frac{K(V_1-V_2) + K^2(V_1-V_2)}{1 + K} \\ &= K(V_1-V_2) \end{aligned}$$

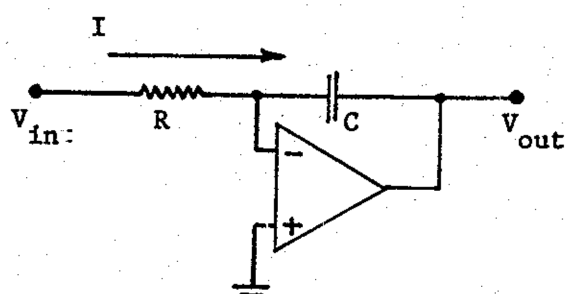
Thus, we arrive at a structure that gives a scaled value of the differential input voltage. Note that this amplifier requires that the ratio R:KR be accurately maintained in both the upper and lower branch of the circuit.

1F. THE INTEGRATOR - A FREQUENCY DEPENDENT RESPONSE

In the circuits above, frequency has not entered any of the circuit calculations. Since we assumed that the ideal op-amp had infinite bandwidth, it made no difference what the input waveform was. When we get to testing real op-amps, we will see that they do have certain limitations that depend on the input waveform. Now, even when we stay well below these limits of the op-amps, we may have a circuit that performs in a manner that varies with the input frequency. In fact, this is the idea we wish to exploit when we design electrical filters. To develope a filter circuit, it is necessary to add to the networks a circuit element that has frequency dependent properties. Yet, we want to restrict ourselves to linear elements since we want only linear distortion (change of amplitude and phases of frequency components, but no new frequency components). This means that we shall consider only capacitors and inductors. Then we will generally throw out inductors is being too large and expensive (at least at audio frequencies) in practical circuits. This leaves the capacitor. If we apply an AC voltage (e.g., an audio signal) to a network containing capacitors, in general the amplitude and phases of the voltages in the network will depend on the frequencies present. For a low-pass filter for example, we would hope the output amplitudes for high frequencies would be reduced relative to the input amplitude. Keeping track of phase is tedious at best. It is much simpler to declare the existence of a complex frequency "s" (the Laplace variable) where $s = \sigma + j\omega$, where ω is the ordinary radial frequency and j is the square root of minus one. Thus, s is a complex number. Readers unfamiliar with the Laplace technique will find that they can still use it by just thinking of capacitors as resistors of value (1/sC) and working out network functions with the usual Ohm's Law for resistors. Then when frequency response is needed, $j\omega$ is substituted for s in the equations. This will become clear later on. The circuit for the integrator is shown in Fig. 10 below:

Fig. 10

Integrator



In the circuit of Fig. 10, the (-) input of the op-amp is at virtual ground, and hence, as in the inverting amplifier:

$$I = V_{in}/R$$

Just as in the inverting amplifier, the current must be flowing out through the feedback element, which in this case is the capacitor, and we treat this capacitor like a resistor which has a value 1/sC. Hence:

$$V_{out} = -I\,(1/sC) = -V_{in}/sCR$$

It is common practice to write V_{out}/V_{in} as T(s) where T(s) is called the "transfer function" and the fact that T is a function of "s" means that a frequency dependent response will result. It is understood in such cases that V_{in} and V_{out} are the amplitudes of sinusoidal waveforms. We thus write T(s) for the integrator as:

$$T(s) = -1/sCR$$

Note that in arriving at T(s) all we do is just assign capacitors a value 1/sC and do ordinary circuit calculations (like current sums and voltage loops) to arrive at an expression for V_{out}/V_{in}. [While we will not be using inductors, we may have to recognize an inductive response, which would be sL. Thus, if we were using inductors, we would use a value sL and treat the element like a resistor, just as we do when we substitute 1/sC for capacitors. Thus we have an s term as an inductor, an $s^o=1$ term as a resistor, and s^{-1} as a capacitor.] While it was quite easy to get T(s) for the integrator, we shall be using the same general procedures for arriving at T(s) for more complicated structures.

Once we have T(s), we can use it to calculate the frequency response. To do this, we have to calculate the magnitude of the transfer function. Since $s = \sigma + j\omega$, T(s) is in general a complex number, so to get the magnitude of T(s), we need to multiply T(s) by its complex conjugate [for every j in T(s), you get the complex conjugate by substituting -j], and take the square root. In general, we are interested in the magnitude of T(s) along the $j\omega$ axis since this gives us the frequency response (where $\omega = 2\pi f$, f being the ordinary frequency in Hertz). To calculate $|T(s)|$, where the $|\ \ |$ mean "magnitude of" we thus take $T(j\omega)$ and the complex conjugate $T^*(j\omega) = T(-j\omega)$ and calculate:

$$T(j\omega) = -1/j\omega RC$$

$$T^*(j\omega) = -1/(-j\omega RC)$$

$$|T(j\omega)| = [T(j\omega)\cdot T^*(j\omega)]^{1/2} = \left[\frac{-1}{j^2\omega^2R^2C^2}\right]^{1/2} = \left[\frac{1}{\omega^2R^2C^2}\right]^{1/2} = 1/\omega RC = 1/2\pi fRC$$

Thus we see that the frequency response of the integrator falls off as 1/f. Doubling the frequency will halve the response. Doubling the frequency is an increase of one octave and halving the response is a 6db drop, hence we say the integrator has a roll-off of 6db/octave. This response is shown in Fig. 11, plotted on a log-log plot. A 45° slope on a log-log plot is 6db/octave, which is the same thing as 20db/decade.

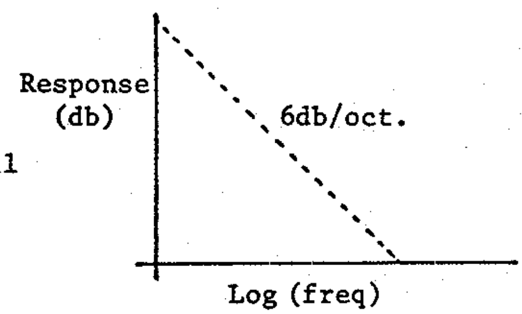
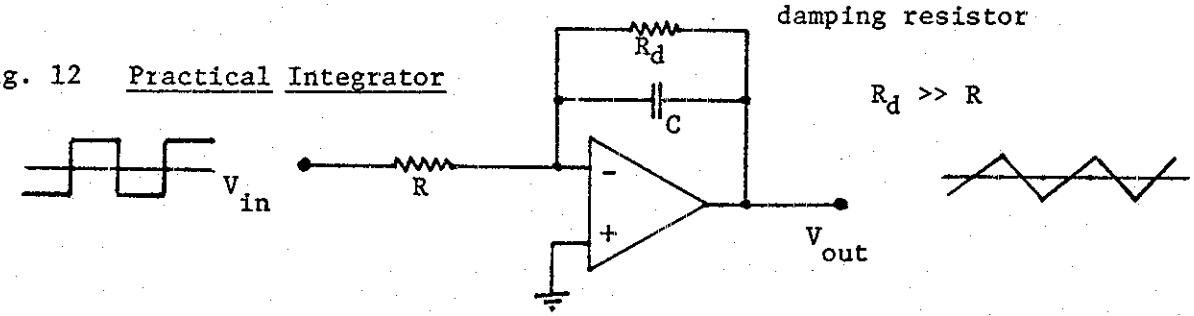


Fig. 11 <u>Frequency Response of Integrator</u>

Another feature of the integrator that is evident from the frequency response function is that the integrator has an infinite response to DC. This fact actually can be quite important since generally there is at least a small DC offset voltage present in any input waveform. Thus, in any practical op-amp integrator, it is necessary to make some provision to limit the DC response. Commonly, this is done by placing a very large resistor across the capacitor to "leak off" any DC charge that starts to accumulate. This in fact changes the circuit from an integrator to a single-pole filter, but it is a good approximation to an integrator as long as the leakage resistor is very large compared to the input resistor (R in Fig. 10). The circuit with the leakage resistor is called a damped integrator. What happens if the integrator is not properly damped is that the DC term in the input is integrated to the point where the output of the op-amp hits one power supply limit or the other. At this point, negative feedback fails, and the differential input is no longer zero.

Above we considered only sine waves into the integrator. What if we put in a square wave for example? When the square wave is at its positive value, a constant current flows into the summing node. This current causes the output of the integrator to ramp negative. When the square wave goes to its negative value, the output starts to ramp back up in the positive direction. The result, for proper setting of time constants relative to the input frequency, is a triangular waveform. This is illustrated in Fig. 12.

Fig. 12 Practical Integrator



It is interesting that the conversion of the square wave to a triangle can be understood at least in part by the frequency response of the integrator. If we look at the Fourier series for the square wave, we see it has only odd harmonics, and that the amplitude of these fall off as $1/n$, where n is the order of the harmonic. If we apply this to the input of the integrator, each harmonic is treated separately. The first harmonic (F) is attenuated by a certain amount; the exact amount does not concern us. There is no second harmonic (2F) in either the input or the output. The third harmonic (3F) is present in the output, but since the frequency response of the integrator goes as $1/f$, the amplitude of the third harmonic is down by a factor of 3 relative to F, and since the third harmonic in the square wave was already down by $1/n = 1/3$, the third harmonic in the output of the integrator is down by a factor of 1/9 relative to the fundamental. Similar arguments show that the fifth harmonic is down by a factor of $(1/5)\cdot(1/5) = 1/25$, and so on, each of the harmonics is down by $1/n^2$. In fact, if we check the Fourier series of a triangle wave, we find that the harmonics do in fact fall off as $1/n^2$. To show that the square is actually converted to a triangle by means of the integrator, we would have to work out the phase response, which we will not do here.

There are two important points about the integrator that carry over to other filter circuits as well. First we saw that we could break an input waveform into components that are actually sine waves, and show how the filter acts on these separately. We then reconstruct the filtered waveform by combining filtered components at the output. This is an example of the principle of linear superposition. The second point is that while the filter alters the relative amplitudes of the components, and also the relative phase (we did not show this explicitely above), it does not create any additional frequencies in the output spectrum. This is a property of a linear system, as opposed to a non-linear system which can produce additional frequencies (harmonic distortion). The filters we shall be dealing with will be linear systems.

Chapter 2: Initial Experiments With Real Op-Amps

2A. INTRODUCTION

In this chapter, we begin a discussion of real op-amps and reach the point where the reader will find it helpful to perform a number of the suggested experiments. In most cases where we list a section as an experiment, we will give the expected results somewhere, so it is not essential that the reader perform the experiment, but he should study the results. In other cases, an experiment is not explicitely mentioned, but it should be obvious that one could be done.

In general, filter networks are designed on the assumption that op-amps are ideal. When it comes down to actually building a working filter, it is important to know how the real properties of real op-amps will alter the ideal performance of the network. In genral, it will be the frequency dependent properties of the op-amps that will be of the most interest. It is the purpose of this chapter to teach the reader which of the op-amp parameters to consider, how they are measured, and how to select a different device or alter a given one when performance falls short of that which is required.

2B. THE FREQUENCY RESPONSE OF OP-AMPS

When we speak of an op-amp, a real one that is, we have in mind a single chip (monolithic) device which is often bought in an eight pin mini-DIP package in the general form shown in Fig. 13. These op-amps cost from something like 25¢ up to a few dollars, depending on the type. Other types of op-amps are available, but these are the most common and the most generally used types. Let's suppose that we have obtained one of these and it is a type 741. How do we know it is a type 741? Well, most monolithic op-amps are made by several companies, one of which originated the thing, and the rest who "second source" it. It is the general practice when second sourcing a product to keep a portion of the type number put on by the original manufacturer. In the case of the 741, it is the sequence "741" that is the key, and may appear as µA741 (Fairchild - the originator), or as LM741 (National), MC1741 (Motorola), CA3741 (RCA), SN72741 (Texas Instruments), or in many other forms. The 741 is likely to be very inexpensive, so we would like to know just how good it is. Suppose we set it up as a follower as shown in Fig. 14. If the op-amp were ideal, we would expect we could put in any voltage (between the supply limits) at any frequency, and get the same thing out we put in with the added advantage of a zero output impedance. For a real op-amp, we know there has to be some upper limit beyond which the frequency response of the op-amp will drop off. All amplifiers have some upper limit beyond which their active elements will begin to fail, and in the case of the monolithic op-amp, these limits will be reached at relatively low frequencies. There are several reasons for this. For one, this "poor" response may have been built in as a means of compensating the op-amp so that it does not oscillate in a region where we would not expect to use it anyway. Also, even if the device is not intentionally compensated, there will be stray capacitances that will "short out" high frequency signals. The fact that the chip of silicon on which these amplifiers are fabricated is so small means that these capacitances will in general be significant. This is the reason that relatively simple op-amps made of discrete transistors may perform as well or better than many monolithic devices because the devices in them are widely separated. Of course, these discrete op-amps will be larger and more difficult to assemble, so it is often better to live with the limitations of the monolithic devices.

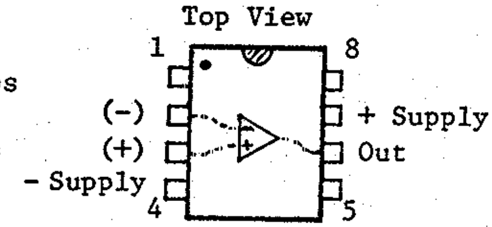
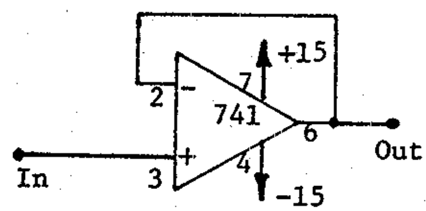


Fig. 13 Typical Mini-DIP Setup of Op-Amp

Fig. 14 Op-Amp Follower



2C. EXPERIMENT NO. 1, OP-AMP BANDWIDTH

The basic procedure of this experiment is as follows. First, we restrict the input signals to the op-amp so that the output signal level is always quite small. The reason we do this is so that we can avoid any complications that would arise from the limitations of the output stage, and can concentrate on inherent frequency response limitations of the op-amp as a whole. With this provision made, we will be setting up the op-amp for a certain gain, and then testing it to see what frequencies can be amplified to that gain, and where the amplifier starts to fail. The basic test circuit is shown in Fig. 15. Note that when B goes to zero, this circuit converts to the follower circuit of Fig. 14 (with the two R resistors acting as loads on the driving stage and the amplifier under test). Thus, by changing B, the circuit makes it possible to examine many different gain conditions. Note that the circuit is basically a non-inverting amplifier of gain 1 + B. There is also an attenuator on the (+) input of the op-amp which gives a loss of 1/(1+B) so the net gain of the circuit of Fig. 15 is one. Thus, as long as we input a small signal, we can be sure that the output will also be small. Input signals should have an amplitude of about 80 mV in this experiment (that's 160 mV peak-to-peak). The value of B·R should be kept in the range of 10k to 100k, and R is calculated back from the desired gain value. In practice, convenient resistor values are chosen, since the exact gain values measured are not too important. With this setup, the frequency response is measured. It should be relatively flat at low frequencies and at some upper frequency it will begin to fall at a relatively rapid rate. Plot these curves, and note the frequency at which the gain drops to about 0.7 (down 3 db) of its low-frequency value. An example curve is shown in Fig. 16.

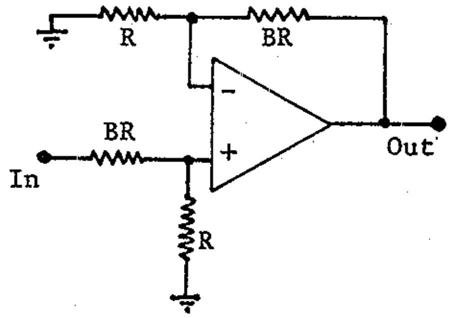


Fig. 15 Test Circuit

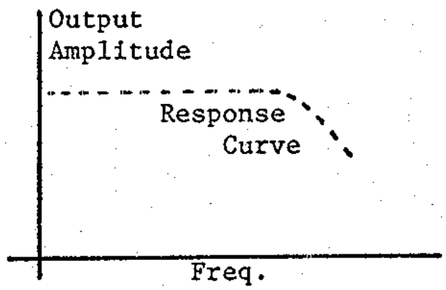


Fig. 16 Typical Response Curve

Keep in mind that although the circuit of Fig. 15 has an overall gain of unity, the amplifier is working at a gain of 1+B. We have just added the input attenuator as a means of avoiding tedious settings of input signal level during our experiment. Select one op-amp, say a 741 or an LM307, and make a set of curves of the type shown in Fig. 16 for various gains such as 1, 2, 5, 10, 100, 500, 1000, 5000, and so on. Plot these all on the same graph, being sure to plot the level portion (low frequency end) of the curves starting at the gain of 1+B, not at 1. If it is more convenient, plot the curves on log-log paper, using the actual voltage readings, and then trace these curves on a fresh piece of log-log paper, starting each at the appropriate gain 1+B by sliding the sheets up or down relative to each other. A full set of curves will look something like the one shown in Fig. 17. This is typical of the internally compensated, 1 MHz bandwidth type amplifiers of which the 741 and 307 types are the best known. The meaning of the 1 MHz bandwidth (also called more properly the "unity-gain bandwidth" or "gain-bandwidth product") will be clear if you simply multiply the gain (1+B) by the bandwidth, which is the flat portion of the response (from DC up to about the 3db points on the curves). For this type of internally compensated op-amp, this product will be approximately constant. This means that the gain-bandwidth product and the unity-gain bandwidth (Bandwidth x 1) are the same number. By quickly measuring the 3db points of other op-amps, determine the gain-bandwidth product of some other newer and improved internally compensated op-amps such as the MC1456, CA3140, LF351, etc.

EXERCISE: Using only a type 741 or type 307 op-amp, and resistors, build a low-pass filter that has a 3db frequency of 1000 Hz, and a DC gain of one.

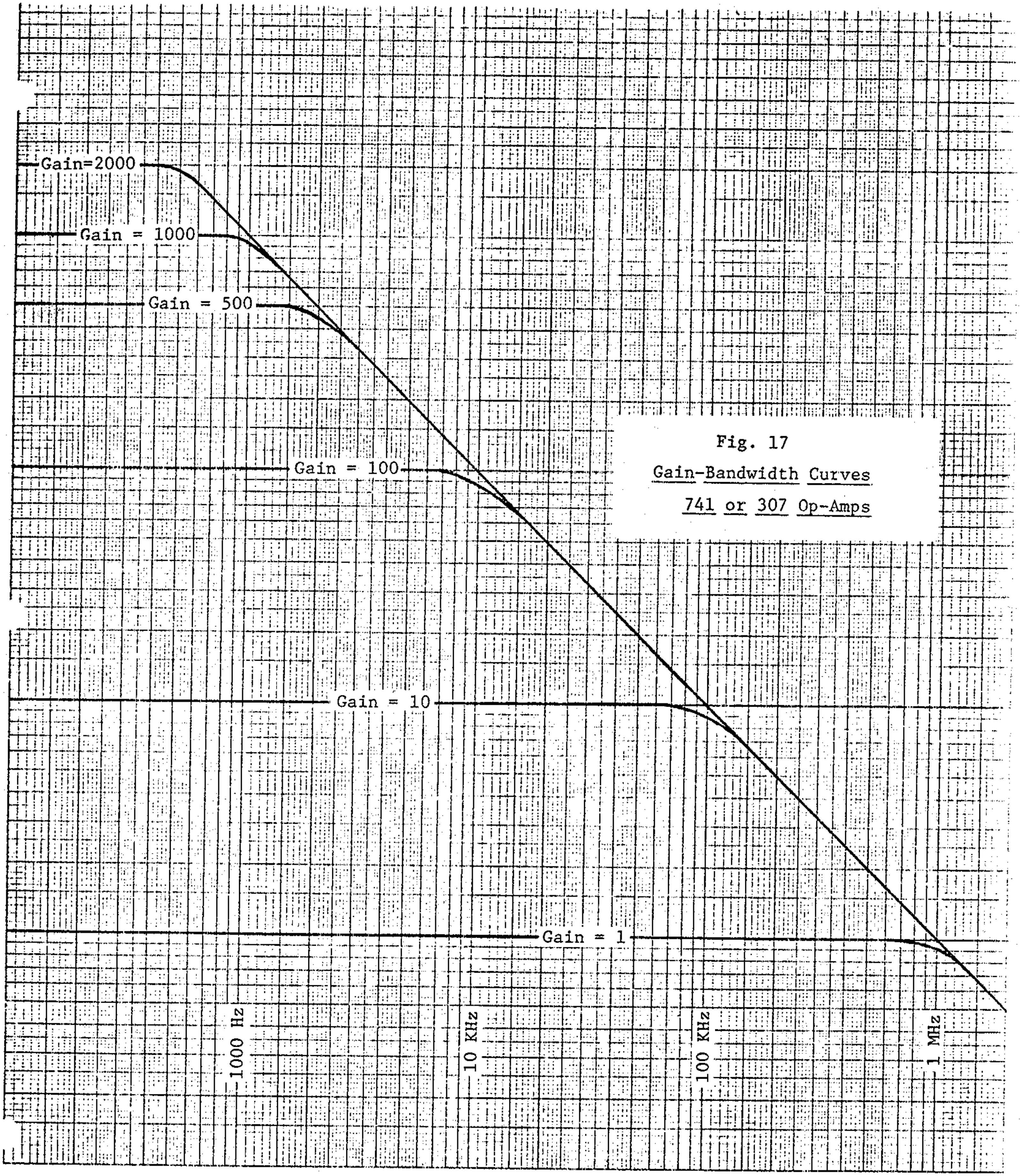


Fig. 17

Gain-Bandwidth Curves

741 or 307 Op-Amps

Before going any further, we should say a few brief words about the need for compensation. At the moment, it might seem to the reader like the designer of op-amp integrated circuits have gone to a lot of trouble to obtain what seems to be a very poor frequency response. It should be realized that in many op-amp applications, no compensation is needed, and in fact, op-amps work better in these applications if they are not compensated (comparators, for example). However, in active circuits, we are generally using negative feedback to obtain the desired gains or just to make the network work. Negative feedback is obtained by feeding back a portion of the output to the inverting input. There is an assumption here. We assume that the inverting input is inverting relative to the output with an instantaneous transfer across the op-amp from inputs to output. In all practical op-amps, there must be some delay, so the signal at the output is an inversion of the inverting input within a delay that can be represented as a phase shift $\phi(f)$, a frequency dependent phase angle. This phase shift is due to the small parasitic capacitances that are present in the chip, and as you might expect, this phase shift is greater at higher frequencies than at lower ones, because the parasitic capacitances are small enough to only be important at high frequencies. Thus, $\phi(f)$ starts at zero for $f = 0$ and increases as f increases. At some frequency, $\phi(f)$ will reach 180°, and this means that the inverting input is now a non-inverting input, and if the gain at this frequency is greater than one, you are going to get oscillation. One way to make sure there will be no oscillation is to make sure the gain of the op-amp is less than one at the frequency where the phase reaches 180°, and this is the purpose of the compensation.

While circuit oscillation is normally associated with high-gain, it is not high gain (but rather phase shift and feedback) that cause an op-amp to oscillate. It also seems strange at first that it is the lowest gain, unity gain, that presents the case requiring the most compensation. This is easy to understand if we realize that a unity gain follower uses the most feedback (to squelch the op-amp high gain) of any other case, and there is thus more of the output fed back. Thus if the output is badly phase shifted across the chip, it will do the most damage at unity gain.

As long as we are working well within the gain-bandwidth product, this unity gain compensation is a good idea, since the op-amp will be stable in most applications (not all applications - sometimes there is something in the feedback loop which changes things). However, internal unity-gain compensated op-amps will generally provide the fewest unpleasant surprises of any type. At other times, uncompensated (to be externally compensated) op-amps have an advantage. When only gains greater than one are to be used, less compensation and more bandwidth can be achieved. Perhaps more importantly in audio-frequency circuits, while more bandwidth may not be needed, less compensation results in higher slew rate, which in turn can lead to higher signal amplitudes and an improvement in signal-to-noise ratio. We will discuss slew rate in the next experiment.

2D. EXPERIMENT NO. 2, OP-AMP SLEW RATE

In experiment 1, we saw some of the limitations of real op-amps, the ones due to the small size and parasitic capacitances which make it necessary to compensate or "roll off" the high frequency end of the response. This was related to the delay across the op-amp. Here we want to consider the delayed response of the output, not relative to the input, but relative to how fast it would "like" to go. Suppose the op-amp's inputs are telling it to go as fast as possible, and we want to know how fast it is actually going. This is what we call the maximum slew rate. The main thing that limits slew rate is how fast the output stage can come up with the necessary current to drive whatever it needs to. In general, the actual load will not matter much, or at least if the load does slow the output down, it is probably a bad application. The major limitation is in the part of the output stage that compensates the op-amp. This contains a capacitor (internal or external) which must be charged by a finite current source, and this means that it can only go so fast. This slows down the entire output.

This is bad in that it causes a lower value of slew limiting than we might like to see, but it is good in that lower slew and compensation go together (as we shall see in the case of the 301 and 748 type op-amps), and we have to have the compensation in closed loop applications (which includes all the active circuits we will be looking at).

Perhaps the idea of slew rate limiting can be made clearest by describing three different ways in which it can be measured: the trapezoid method, the triangle method, and the sine wave method. The trapezoid method is straightforward, and is the one the reader would dream up himself if asked to test slew rate. The idea is to apply to the op-amp follower circuit a square wave from a generator that is known to have fast rise time. The setup is shown in Fig. 18. It is a simple matter to observe the output of the follower with an oscilloscope with an accurate time base, and essentially just read the slew rate of the edges in volts/second. Typical rates will be in the range of 0.5 volts/microsecond (types 741, 307) to 2.5 volts/microsecond (MC1456) and to 10 volts/microsecond and above for CA3140, LF351, etc. While this is an obvious way of measuring slew rate, there are several reasons it may not work well in a student lab. These reasons include the possibility that scopes may be too slow, or their time bases may be inaccurate, and signal generators may not have rise times much faster than the fastest op-amps to begin with.

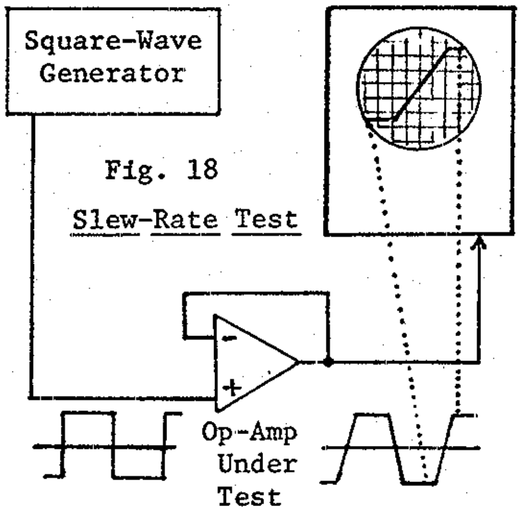


Fig. 18

Slew-Rate Test

The second method we will discuss relies very little on an accurate scope (as long as the voltage scale is accurate or can be calibrated (perhaps with a 15 volt power supply). It does require a known frequency, so an accurate generator dial or frequency counter is needed. This method uses the fact that an op-amp can be driven at a high enough frequency that it is constantly being driven to its slew limit. Such a situation results in a triangular waveform. The setup is shown in Fig. 19. In a sense this triangle method is an extension of the trapezoid method, except here the time factor is determined from the known frequency of the driving waveform. The basic procedure is to just turn up the frequency until a triangle waveform is obtained, and measure its amplitude and frequency. From Fig. 19 it is easily seen that the triangle covers a full voltage swing of 4A (where A is the amplitude) during one period (1/f). Thus the slew rate is:

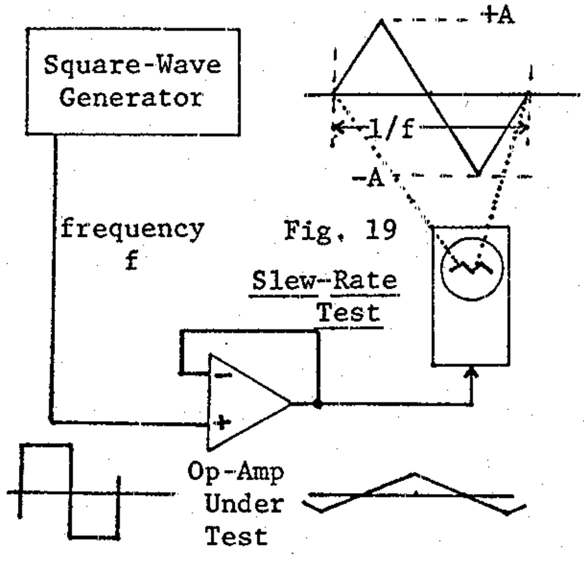


Fig. 19

Slew-Rate Test

$$SR = \text{volts/sec} = \frac{4A}{1/f} = 4Af$$

This is a very simple method and has several advantages. First, it is relatively independent of inadequacies in the scope and in the driving waveform (which could in fact be a sine, sawtooth, or even triangle if you think about what you are doing). You only need to know the frequency with some degree of accuracy. Typically the amplitude of the triangle should be down to 1/3 or so of the input waveform amplitude.

The third method of measuring slew rate (the sine method) is a little more subtle and probably less accurate than the first two methods, but it shows the practical consequences of slew limiting much better than the other two since we see the effect of slew limiting on a sine wave, and can better guess at the resulting distortion. We start by determining the maximum rate of change of output voltage needed to handle a sine wave of frequency f and amplitude A (that is, $V(t) = A \sin 2\pi ft$). The rate of change of output voltage is just the derivative $dV/dt = 2\pi fA \cos 2\pi ft$ and this function in turn has an extreme where its derivative equals zero. This occurs at:

$$d^2v/dt^2 = -4\pi^2 f^2 A \sin 2\pi ft = 0.$$

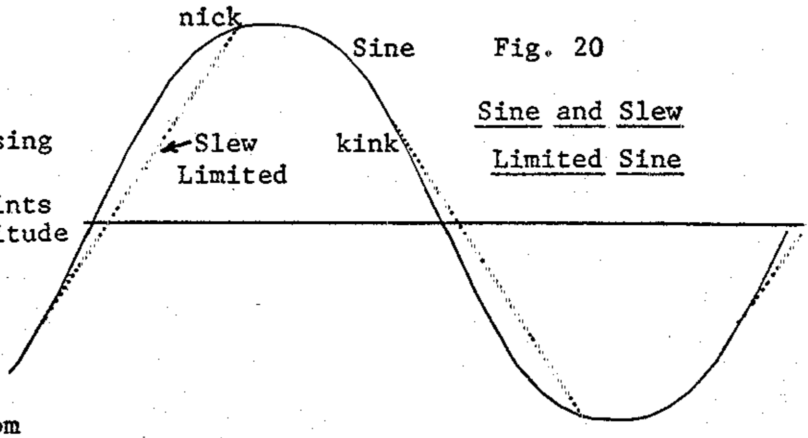


Fig. 20

Sine and Slew Limited Sine

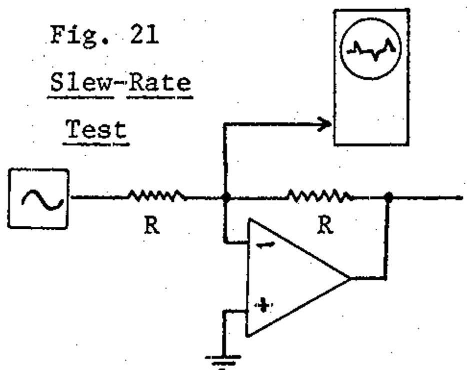
Thus dv/dt has its maximum at the zero crossing points of the sine wave, which is probably obvious from a study of Fig. 20. At the points where the Sine is zero, the Cosine has magnitude one, so the value of the slew rate is dv/dt evaluated with the Cosine = 1. Thus:

$$dv/dt_{max} = 2\pi fA$$

The trick is to detect the slew limiting from an experiment. The procedure is to input a sine wave of about 10 volts amplitude and turn up the frequency until some sign of slew limiting is seen, and then to apply the amplitude and frequency to the equation for dv/dt_{max}. The first sign of slew limiting will likely be a slight kink in the relatively straight portion of the waveform just before the zero crossings, and a corresponding nick as the voltage moves on up toward the peak, and the slew limited output catches up, and slows down to a following mode. This behavior is sketched in Fig. 20. Another way of using the sine method is illustrated by Fig. 21. Here we use a fairly sensitive scope scale to monitor the voltage on the (-) input of the op-amp which is this time configured as an inverter. As we discussed above in Chapter 1, this voltage should be zero since the (+) input is grounded. However, this assumes the feedback is from an ideal output. However, during slew limiting, the output is not where it should be relative to the input, and the (-) input will jump slightly off ground. These nicks as slew limiting sets in are thus the indication that the maximum frequency has been reached, and the amplitude and frequency can be plugged into the dv/dt_{max} equation.

Fig. 21

Slew-Rate Test



Once you understand the different methods of measuring the slew rate, try a few of them and compare the results against each other and against book values for slew rate from op-amp data books. A 741 or 307 type op-amp is suggested for this initial test since these slower op-amps will be easier to start with. Once a method is mastered, you can then try measuring other op-amps as well. Op-amps such as the MC1456, CA3140, and LF351 are good choices for additional measurments, although just about any op-amp will suffice.

In order to investigate the effect of compensation on both slew rate and on gain-bandwidth product, we have to start with an op-amp that can be custom compensated. Here types LM301 or 748 will be good choices. The compensation in these op-amps is accomplished by inserting a capacitor between pins 1 and 8. Start with a 30pf capacitor in this position with either the 301 or the 748 and verify by testing the slew rate, bandwidth, or both that the op-amp behaves very similar to the internally compensated types 741 or 307. [In fact, it is essentially an internal on-chip 30pf capacitor that is used in the 741 and 307].

Thus there is really no point in using a 748 plus an external 30pf capacitor just to get the performance of a 741 back. [All the experiments we are discussing here that use a 748 can also use a 301 with about the same results.] What happens when we remove the capacitor? Well the op-amp will take off into a nasty oscillation (high frequency and full amplitude). If we cut the compensation to about 10pf, we have a more (but not fully) stable situation. To see the real advantage of custom compensation, we have to consider something like the circuit of Fig. 15. Set up this circuit using a 748 with B = 1 (op-amp thinks the gain is 2) and 15pf for the compensating capacitor. Check the slew rate and bandwidth. Repeat this for B = 9 (gain of op-amp is 10) and a compensating capacitor of 3pf. In general, you should find that the compensation can be decreased from 30pf by a factor equal to the gain. This should extend the gain curves at the set gain so that they extend beyond the region reached in Fig. 17 (i.e., you can increase the gain bandwidth product, effectively sliding the 45° downslope of Fig. 17 to the right). You won't be able to use lower gains however unless the compensation is increased again, so a portion of the previously usable region is no longer usable. The slew rate at the same time increases by a factor proportional to the decrease in compensation.

Another interesting case is to compare the 748 with 30pf compensation (which you examined above and found to be essentially a 741) with the 748 with 15pf compensation used in the unity gain inverter of Fig. 22. While both of these circuits have unity gain, the 748 in the inverter slews twice as fast because it has half the compensation. This is often a useful trick to keep in mind.

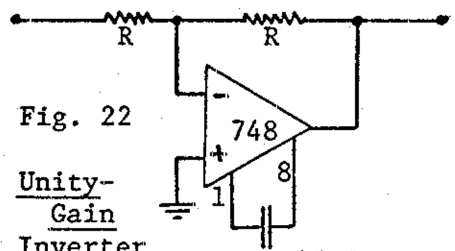


Fig. 22 Unity-Gain Inverter

A brief explanation of what is going on seems in order. We want to see why we can use less compensation with higher gain. We begin by considering the typical negative feedback stage as shown in Fig. 23. This consists of a voltage divider with the op-amp output at one end, the (-) input at the center, and something else at the other end. This "something else" may be a ground (non-inverting amp) or a signal input (inverting amp), but the important thing is that it is not the output voltage. The voltage divider thus attenuates the output signal fed back to the (-) input. If R1 is 100k for example and R2 is 200k, the signal fed back is attenuated to 1/3 the value it would be if a full feedback unity gain follower were being used. Thus, when we consider the possibility of oscillation, the signal at the output would have to be three times as large when the phase shift equals 180°. This means that we can achieve stability with only one third the compensation.

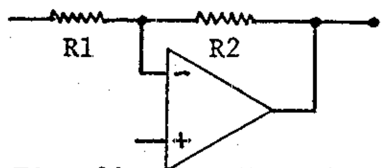


Fig. 23 Negative Feedback Voltage Divider

It might be supposed that additional slew rate can always be obtained at the output if we just attenuate signals and use additional gain in the op-amp. This is true up to a point, but there are always trade-offs. For example, attenuating the input signal before reaching the op-amp input means that the op-amp input noise level is going to be more important relative to the signal, and the signal-to-noise ratio will suffer. There are many considerations that come in when it comes to applying op-amps to the processing of high quality audio signals, and we will not attempt to cover these here. We just want to warn the reader that things can be a little trickier than they might appear at first.

EXERCISES:

1. Discuss the harmonic content of a severely slew limited waveform of any type.

2. Most modern op-amps will withstand a full supply voltage as the differential input voltage. That is, you can apply +15 to one input and -15 to the other without damage. Older op-amps could be damaged if the differential input were too large. If the maximum allowable differential input is something like 1 volt, tell how the op-amp could be damaged by slew rate limiting if the circuits described in this chapter are used directly. Tell how two simple diodes can be used to prevent damage.

3. Tell how the circuit of Fig. 15 or a simple modification of the basic structure can be used to make a simple test for adequate bandwidth and adequate slew rate simultaneously.

4. The circuit of Fig. 24 is apparently indaequately compensated since it is a unity gain inverter and the compensation capacitor is only 5pf. Compare this with Fig. 22. Tell why this circuit is in fact stable. What is the expected slew rate. Discuss any limitations of the circuit.

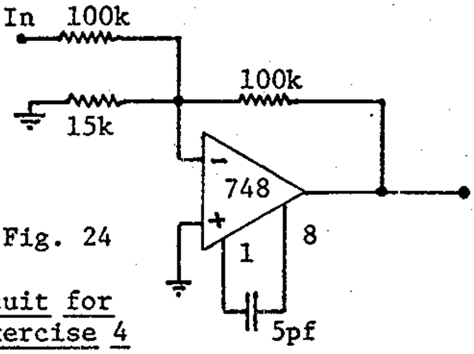


Fig. 24

Circuit for Exercise 4

CHAPTER 3: S-PLANE ANALYSIS TECHNIQUES AND ACTIVE FILTERS

3A. INTRODUCTION

In this chapter, we will be giving you some procedures that you will often use in the analysis of active networks. While there is a firm mathematical basis for the things we will be giving, we will not go into that here. Instead, here you will find more of a "cookbook" approach. You probably already have the general ideas about Laplace transforms and s-Plane analysis, or you can find it in many standard references With even a very sketchy idea about these methods, you will be able to follow our procedures and get useful answers.

3B. FREQUENCY SENSITIVE NETWORKS AND THE s FREQUENCY VARIABLE

In Chapter 1, Section 1F, we looked at the integrator, and found that while its behavior was frequency sensitive, the analysis could be handled in terms of a mathematically "complex" frequency which we call "s". Perhaps it would be a good idea if the reader goes back to that section for a review before going on here.

The steps in the analysis of a frequency sensitive active network, which is often called an "active filter" are as follows:

1) Each capacitor in the network should be considered mentally to be a resistor with value $1/sC$, where C is the capacitance.

2) The op-amp should be considered ideal. In active networks, this means two things: (a) The input currents are zero, and (b) The differential input voltage is zero.

3) Identify by appropriate symbols all unknown currents and voltages. You should also make sure you know which voltage is the input V_{in} and which is the output V_{out}.

4) Using standard Ohm's law type calculations, solve out the network. Of course you can set up n equations corresponding to the n unknown voltages and currents, and solve by standard methods for simultaneous equations, but it is usually the case that several of the equations are rather trivial. This makes the "by gosh and by gully" method more useful in most practical cases. In any case, you are trying to obtain an expression for the transfer function $T(s) = V_{out}/V_{in}$ in terms of s, the resistor values, and the capacitor values, with no unknowns.

5) From T(s), determine the network characteristics you want to know such as type of filter, cutoff frequencies, frequency response, Q, and so on. The way this is done will become clear later. For now we just want to point out that when the transfer function is obtained, you have essentially solved your problem.

3C. EXAMPLE 1, SINGLE-POLE LOW-PASS FILTERS

Fig. 25 shows a simple RC series circuit which is tapped in the center, and an op-amp follower has been added on to "buffer" the output. Here the op-amp is used because it has a high input impedance and a low output impedance. Without the buffer, we would expect our measuring instruments or following stages in our circuit to disturb the performance of the RC series.

Single-Pole Low-Pass

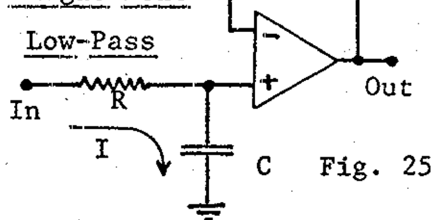
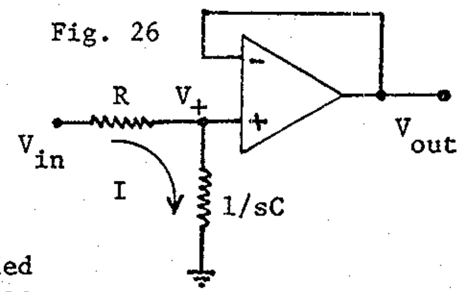


Fig. 25

This network is somewhat trivial, but trivial networks are the best for illustrating a method. We shall follow the steps above in 3B one by one.

1) We consider the capacitor C to be gone and replaced with a resistor 1/sC as in Fig. 26.

Fig. 26



2) The op-amp is ideal, the unity gain follower. The voltage V_{out} is the same as V_+.

3) The unknown current is I and the unknown voltage is V_+.

4) The current I is equal to the voltage V_{in} divided by the series "resistance" R + 1/sC. The voltage V_+ is the current I times the "resistor" 1/sC. Also the voltage V_+ is the same as V_{out} so:

$$T(s) = V_{out}/V_{in} = \frac{1/sC}{R + 1/sC} = \frac{1}{1 + sCR} \qquad (3C\text{-}1)$$

5) We note from T(s) that at low frequencies (s is small), T(s) approaches 1, and at high frequencies (s is large), T(s) approaches zero, hence we have a low-pass filter. If we want, we can get the frequency response function $|T(s)|$ by substituting $j\omega$ for s, and taking the magnitude of the complex number by multiplying $T(j\omega)$ by its complex conjugate $T(-j\omega)$.

$$T(j\omega) = 1/(1 + j\omega CR) \quad \text{and} \quad T(-j\omega) = 1/(1 - j\omega CR) \qquad (3C\text{-}2)$$

$$|T(s)| = [T(j\omega)\cdot T(-j\omega)]^{1/2} = [1/(1 + \omega^2C^2R^2)]^{1/2} \qquad (3C\text{-}3)$$

With this, we can make a plot of the frequency response as a function of frequency f where $f = \omega/2\pi$. If we take R to be 10k and C to be 0.01 mfd (10^{-8}) we arrive at a plot as shown in Fig. 27.

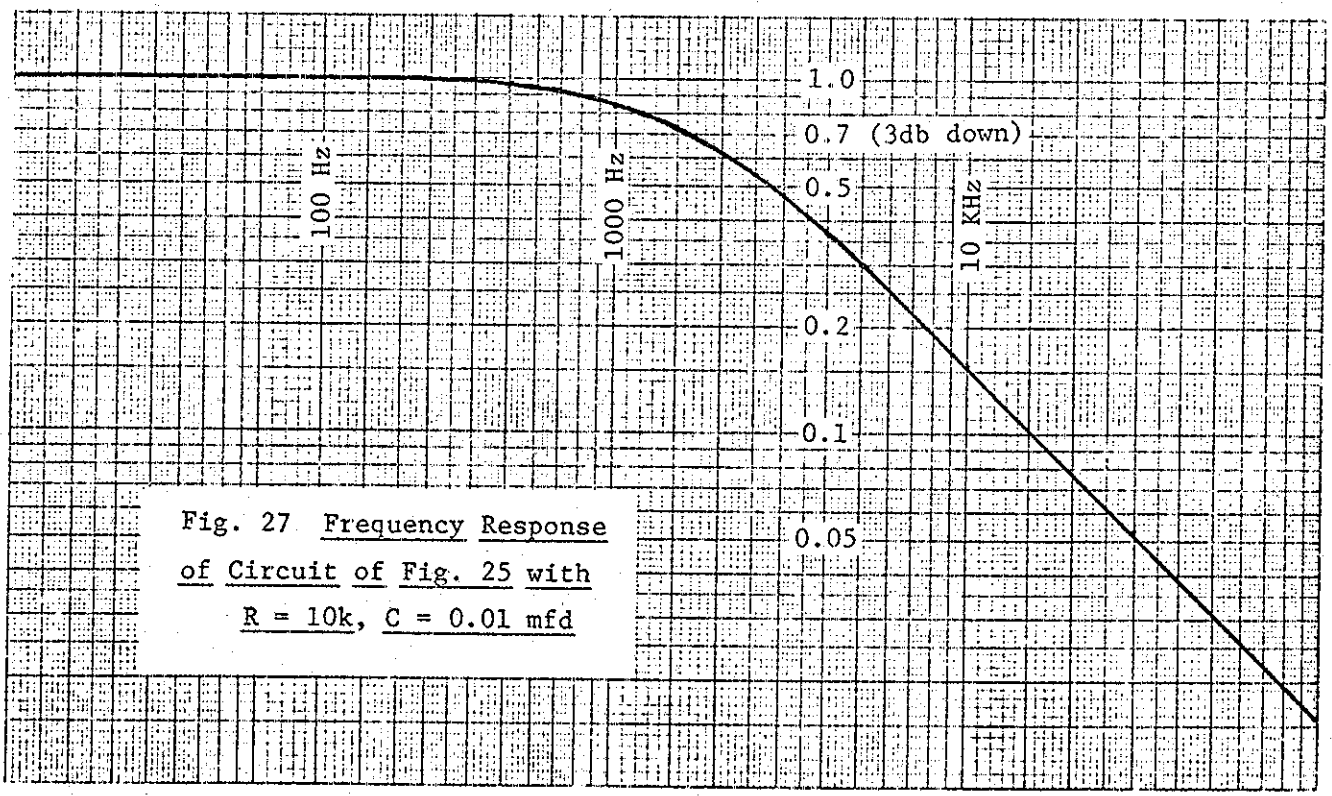
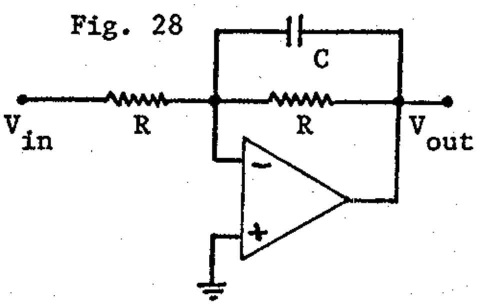


Fig. 27 Frequency Response of Circuit of Fig. 25 with R = 10k, C = 0.01 mfd

A frequency response graph of the type shown in Fig. 27 probably tells us more that we want to know about the filter than anything else. We can tell from this graph that the filter is low-pass and for example that a 100 Hz signal will be passed essentially unattenuated while a 10 KHz signal will be down by about 0.157.

EXERCISE: Analyze the circuit of Fig. 28 following the steps in Section 3B just as was done in the example in Section 3C. Show that the magnitude of the transfer function is the same as the magnitude of the transfer function of Fig. 25. Relate these results to the integrator and damped integrator of Section 1F. Discuss the differences between and relative advantages of the two circuits (Fig. 25 and Fig. 28).

Fig. 28

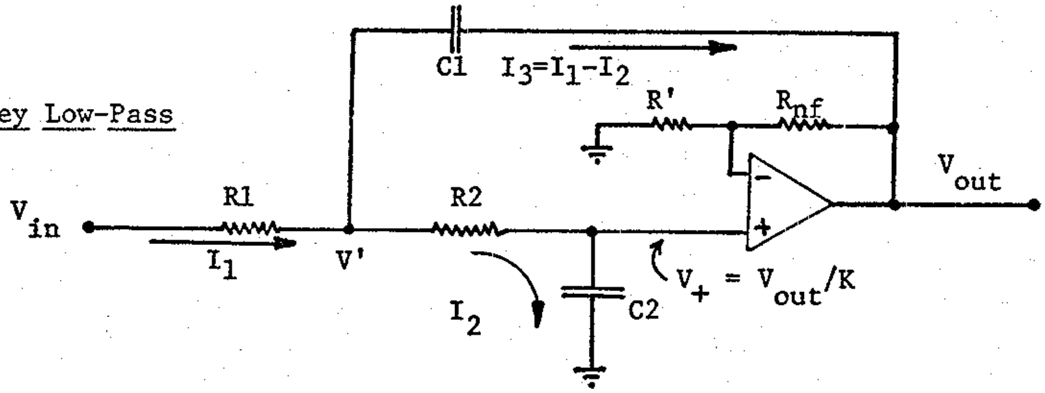


3D. EXAMPLE 2, SECOND-ORDER SALLEN-KEY LOW-PASS NETWORK

An extremely popular low-pass structure is the 2nd order "Sallen-Key" (named after the discoverers) low-pass filter. The structure is shown below in Fig. 29.

Fig. 29

Sallen-Key Low-Pass



We will follow the procedure of Section 3B.

1) We consider capacitor C1 to be replaced with a resistor 1/sC1 and capacitor C2 to be replaced with a resistor 1/sC2. It should not be necessary to redraw the network as was done in Fig. 26, but you can if you wish.

2) The ideal op-amp in this case is used to form a non-inverting amplifier with gain = $K = 1 + R_{nf}/R'$, and this amplifier has infinite input impedance.

3) There are three unknown currents (I_1, I_2 and I_3) and three unknown voltages V', V_+, and V_{out} in this network, so we would expect to need six equations to solve the network. However, rather than setting up six equations, we can get rid of two just by the way we represent our variables. This is a good example of the "by gosh and by gully" method. First note that with the ideal non-inverting op-amp gain stage, we can just write V_+ as V_{out}/K. Secondly, summing currents at the V' node tells us that $I_3 = I_1 - I_2$. These substitutions are useful because they replace one unknown (and the need for an equation) with at most two other symbols. It is a good idea to stop this sort of simplification with steps which replace one symbol with no more than two others, although the whole network could be solved by this method.

4) We now have to make some observations (i.e., write some equations). Our simplification in step 3 above has left us with only "Ohm's Law" type of calculations. These are:

$$I_1 = (V_{in} - V')/R_1$$

$$I_2 = V'/(R2 + 1/sC2) = (V_{out}/K)/(1/sC2)$$

$$I_1 - I_2 = (V' - V_{out})/(1/sC1)$$

This is really 4 equations, since the I_2 equation can be represented as two different equations. It is probably easiest to solve the two right hand members of the I_2 equation for V' in terms of V_{out}. Then the equations for I_1 and I_2 can be plugged into the one for $I_1 - I_2$ (using right most member of the I_2 equation). Solving for V_{out}/V_{in} we get:

$$T(s) = V_{out}/V_{in} = \frac{K/(R_1R_2C_1C_2)}{s^2 \quad + \quad s\left[\frac{(1-K)}{R_2C_2} + \frac{1}{R_1C_1} + \frac{1}{R_2C_1}\right] + \frac{1}{R_1R_2C_1C_2}} \tag{3D-1}$$

A useful simplification that is almost always used is to let $R_1 = R_2 = R$, and $C_1 = C_2 = C$, in which case $T(s)$ becomes:

$$T(s) = \frac{K/R^2C^2}{s^2 \ + \frac{s}{RC}[3 - K] + 1/R^2C^2} \tag{3D-2}$$

5) As before, we can find the magnitude of the transfer function $T(s)$ by substituting $j\omega$ for s and then $|T(s)| = [T(j\omega)\cdot T(-j\omega)]^{1/2}$. This done, a little algebra gives:

$$|T(s)| = K\left[\frac{1}{R^4C^4\omega^4 \ + \ R^2C^2\omega^2[7 - 6K + K^2] + 1}\right]^{1/2} \tag{3D-3}$$

In order to plot the magnitude of the transfer function, we have to select a value for K. To obtain appropriate values for K, designers generally use certain tables of filter parameters for different filter characteristics. Here we will just say that we have taken from these tables a value of $K = 2.233$, which is supposed to give a "3db Ripple, Chebyshev" characteristic to the filter. We can also choose $R = 1$ Meg and $C = 1$ mfd, so that $RC = 1$ and this will make our math easier. This done, the equation for the magnitude of the transfer function becomes:

$$|T(s)| = 2.233\left[\frac{1}{\omega^4 - 1.41\omega^2 + 1}\right]^{1/2} \tag{3D-4}$$

A plot of this function is shown in Fig. 30, where we have plotted $|T(s)|$ as a function of $f = \omega/2\pi$. We note that the response is basically low-pass, but there is a "bump" on the upper corner. This sort of bump can be useful because it results in a sharper initial cutoff. The bump is possible with this second-order structure, but need not be there for all values of K. The principal advantage of the second-order over the first-order is the sharper final cutoff rate (12db/octave as compared to 6db/octave for first-order), and the ability to control the bump on the corner.

The filter response of Fig. 30 is in fact that of a practical filter. It has a sub-audio cutoff frequency [usually taken to be the frequency 3db (0.707) down from the top of the bump, about 0.19 Hz in this case]. Such a filter would be useful for removing noise of say 5Hz and higher from a slowly varying DC level that we want to monitor. To construct the practical filter, we would have to set up the amplifier to give the gain K, and this is accomplished for example, with $R_{nf} = 27k$ and $R' = 22k$ because $1 + R_{nf}/R' = 2.227$ in this case which is within resistor tolerance of $K = 2.233$ for most practical circuits. A final consideration is to select components for the filter compatible with the low frequency. Clearly we do not need much slew rate at all here. But due to the low frequency, we have to keep bias currents and leakage currents to a minimum. Thus, we would choose a low bias current op-amp such as the CA3140 or LF351, and low leakage capacitors (not electrolytic) for the capacitors C.

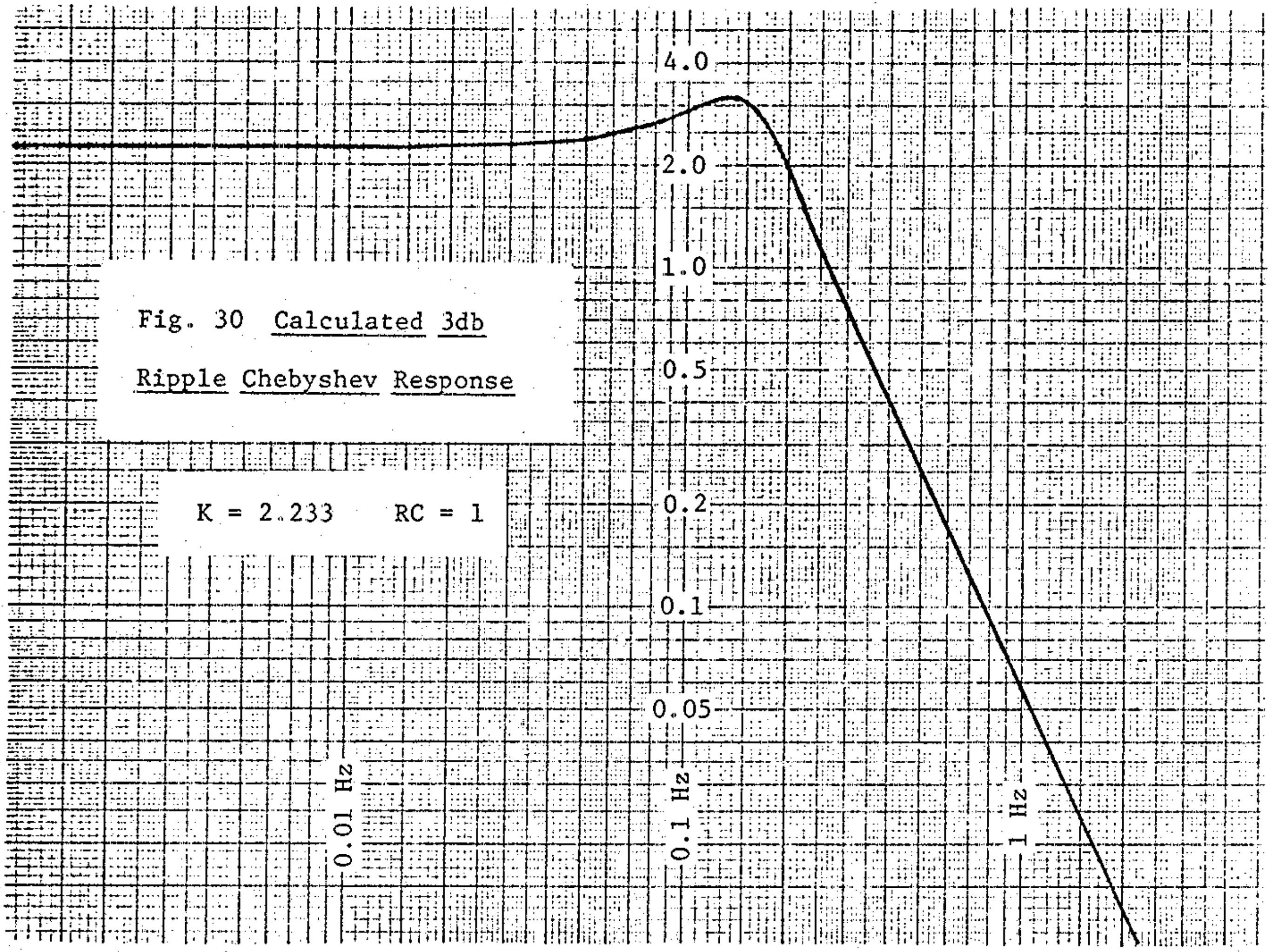


Fig. 30 Calculated 3db Ripple Chebyshev Response

Of course, in many cases we need to design filters with many different parameters and would like to avoid long calculations in each individual case. It is fortunate that in network design it is often possible to just transform available data into a new form. One good example is frequency scaling. The filter of Fig. 30 for example has a much lower cutoff frequency than is needed in many applications, and is not ideally suited for teaching purposes in labs due to its low frequency and the resulting need for special equipment. Suppose we want to have a higher cutoff, 1230 Hz for example. To scale a frequency up, we decrease the resistor R, or decrease the capacitor C, and the frequency scales upward inversely as the RC product. If we

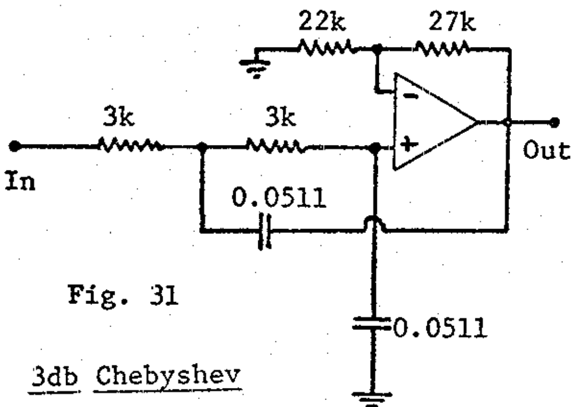


Fig. 31

3db Chebyshev

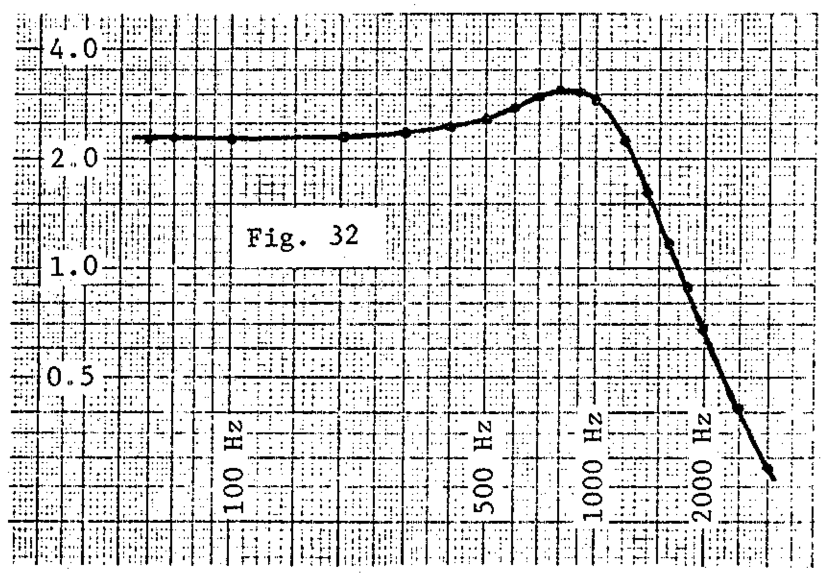


Fig. 32

wanted to make the scale from 0.19 Hz to 1230 Hz (a factor of 6474 up), we could decrease the capacitor to 154 pf down from 1 mfd. This capacitor is a little small for active filters because stray capacitances can easily amount to 5 pf or so. We could also change R to 154 ohms, but this is also a little small in op-amp circuits where we would prefer resistors in the range of 1k to 1 Meg. Thus, in this case, as in many cases, it is desirable to change both R and C. We know that in the case of Fig. 30, the RC product is 1, so to get the 1230 Hz cutoff, we need an RC product of 1/6474 = 0.000154. If we choose R at 3k for example, we would get C = 0.0514 mfd, which is a reasonable value. An experimental approximation to this filter is shown in Fig. 31 while the frequency response is shown in Fig. 32.

3E. GRAPHICAL INTERPRETATIONS

It is probably somewhat evident to the reader that things can get a lot more complicated very fast as we start to go over to more and more complicated networks. The math becomes so extensive that we need a computer to get our calculations done within a reasonable length of time. Conceputally, we would have a more and more difficult time telling just what is going on as the math passes from our direct grasp. Fortunately, there are some graphic aids that are a great help in letting us get a feel for what is going on.

So far, we have been looking at transfer functions that are of the general form (for the second order case):

$$T(s) = \frac{B}{s^2 + Ds + E} \tag{3E-1}$$

Where B, D, and E are constants. This is characteristic of a low-pass second-order filter. If it were a second-order bandpass, the numerator would be of the form Bs and the denominator would be the same. If it were a second-order high-pass, the numerator would be of the form Bs^2 and the denominator would be the same. A general response function would be of a form that combines high-pass, bandpass, and low-pass:

$$T(s) = \frac{As^2 + Bs + C}{s^2 + Ds + E} \tag{3E-2}$$

It turns out that this general form is to a large degree all we need to know how to handle, because higher order filters are generally built out of second-order sections because such structures are less sensitive to component variations than are their realizations with unique structures for the given order. The second-order numerators and denominators can be factored into two first order terms (by the quadratic formula if by no simpler means) such as:

$$T(s) = \frac{(s - z_1)(s - z_2)}{(s - p_1)(s - p_2)} \tag{3E-3}$$

Consider what happens when s takes on the value of any of the special numbers p_1, p_2, z_1, or z_2. If it takes on a z value (a zero), the numerator of T(s) becomes zero and hence T(s) becomes zero. If it takes on a p value (a pole), the denominator of T(s) becomes zero and T(s) becomes infinite. These are thus values of s that are of special interest to us. Consider also the solution to the quadratic equation:

$$ax^2 + bx + c = 0 \tag{3E-4}$$

which is by the quadratic formula:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{3E-5}$$

If we are using this quadratic equation to find the poles of a transfer function such as that of the Sallen-Key low-pass (setting RC = 1):

$$\text{Denominator} = s^2 + s[3-K] + 1 \tag{3E-6}$$

The roots of the denominator are the poles of T(s), and the quadratic formula gives:

$$p_1, p_2 = \frac{-[3-K] \pm \sqrt{5 - 6K + K^2}}{2} \tag{3E-7}$$

Thus, the poles have an imaginary part (the square root is taken of a negative number) when K becomes greater than one. Also, such complex poles must exist because we have already demonstrated an experimental filter (Fig. 31 and Fig. 32 where K = 2.233) where such poles are mathematically present. We can also note that the real part of the pole is negative until K becomes greater than 3. It will turn out that in order for a filter to be stable, the poles must have a real part that is negative. Thus, K in the Sallen-Key filter must be less than 3.

Since the poles and zeros of T(s) may be complex, and we know by our example that they can exist in a practical filter, we have to graph our poles and zeros as complex numbers with a real axis and an imaginary axis in a complex plane. This is called the complex frequency plane or the "s-Plane" in the literature. Note also that where poles and zeros are complex, they occur in complex conjugates as a result of the ± square root part of the quadratic formula. We can now make a two dimensional plot of the complex frequency in terms of a real part (σ) and an imaginary part (jω). Fig. 33 below shows such a plot, and we have added to the plot the pole positions for the 3db Chebyshev filter which are at:

$$p_1, p_2 = -0.384 \pm j(0.924) \tag{3E-8}$$

If we now take another look at equation (3E-3) for T(s) and think about how we take the magnitude of T(s), the frequency response, we can write:

$$|T(s)| = \frac{|s-z_1||s-z_2|}{|s-p_1||s-p_2|} \tag{3E-9}$$

Also, the magnitude of a complex number $a + bj$ is given as $[a^2+b^2]^{1/2}$ which is just the hypotenuse of a right triangle. It is the length of the line on the s-plane. For the 3db Chebyshev low-pass filter, we would have:

$$|T(s)| = \frac{1}{|s - p_1||s - p_2|} \tag{3E-10}$$

Thus if we want to know the value of T(s) at any point in the s-plane, we just measure the distances on the s-plane (with a ruler!) and do the division of equation (3E-10). The usual frequency response is obtained as the value of |T(s)| along the jω-axis from zero on up. [Remember that to obtain |T(s)| by direct mathematics we substituted jω for s]. We choose a point (j/2 in Fig. 33 for example) and measure the distances to the poles (dotted lines in Fig. 33). We then divide 1 by the product of these two distances. We record this value and

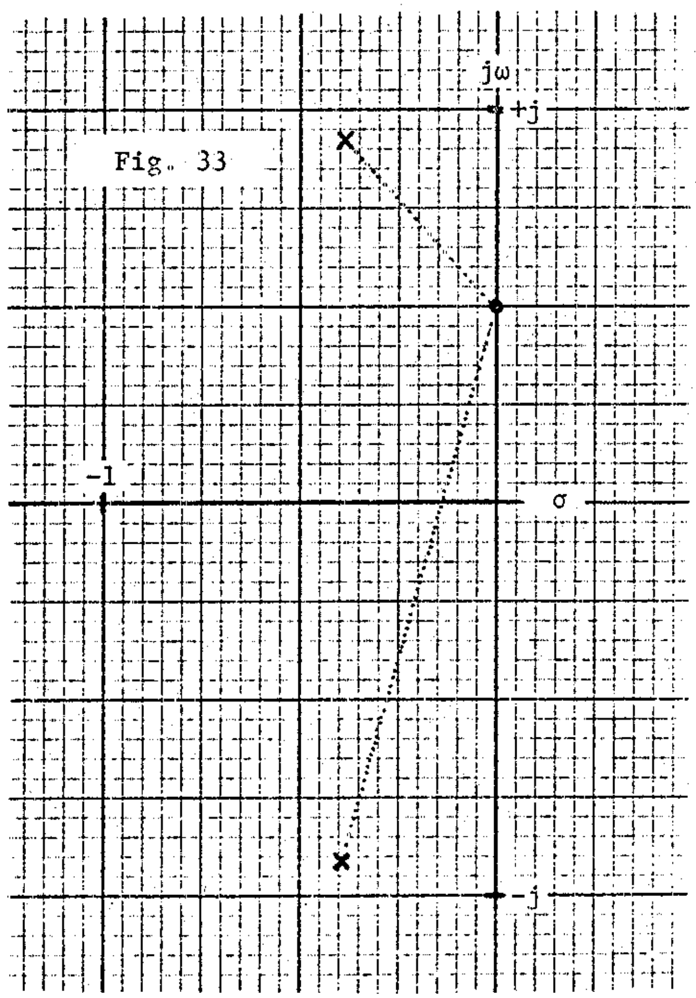


Fig. 33

go on to the next point that we need. When we have enough points, we can sketch in the full curve. It matters little exactly what units we measure the distances in (a centimeter ruler, or a strip of graph paper cut from the edge of the plot will work), as long as we plot our data on log-log graph paper. The curve that results from a complete graphical calculation on Fig. 33 will exactly overlay either Fig. 30 or Fig. 32. Often times it is possible to calibrate the response scale exactly by some observation on the filter structure being considered. In the Sallen-Key filter, we can easily see the DC gain is K (just pretend the capacitors are removed and see what the gain is). Thus, whatever number we obtained graphically for $s = 0$ in Fig. 33 corresponds to a gain of K. Another point to be made about the graphical interpretation is that it may be easier to write down a closed-form mathematical solution for the frequency response based on a graphical setup using standard methods of geometry and trigonometry than it is to calculate $|T(s)| = [T(j\omega)\cdot T(-j\omega)]^{1/2}$.

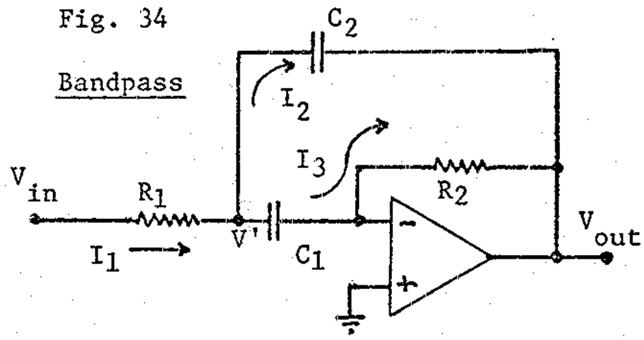
A graphical interpretation of phase response is also possible and useful. To obtain the phase response, we measure angles relative to the real axis as seen from the poles and zeros of the network, looking at the point on the $j\omega$-axis where the phase response is to be evaluated. The angles as seen from the zeros are added and the angles as seen from the poles are subtracted. The method will be illustrated in the next example.

3F. EXAMPLE 3, BANDPASS FILTER

A bandpass filter structure is shown in Fig. 34. Here we will derive the transfer function of the network and then use the graphical means of determining frequency and phase response. In our previous examples, we used op-amp structures that were examined in chapter 1 (buffers and non-inverting amplifiers), but here we go right back to basic ideal op-amp properties. We will be using the procedure of Section 3B but will not be numbering the steps.

Fig. 34

Bandpass



The first thing to note is that negative feedback is working, so it must be the case that the (-) input of the op-amp is at ground potential since the (+) input is grounded. This immediately tells us that the current I_3 must be $-V_{out}/R_2$. This same current must be flowing through capacitor C_1 since there is no current into the actual inputs of the op-amp. Thus we can get the voltage V' as:

$$V' = I_3(1/sC_1) = -V_{out}/sC_1R_2$$

It is then easy to see that $I_1 = (V_{in} - V')/R_1$ and $I_2 = (V' - V_{out})sC_2$, and since $I_1 = I_2 + I_3$, we can solve for the transfer function:

$$T(s) = \frac{-s/R_1C_2}{s^2 + s[1/R_2C_1 + 1/R_2C_2] + 1/R_1R_2C_1C_2} \qquad (3F - 1)$$

It is easier to take $C_1 = C_2 = C$ in which case T(s) is given by:

$$T(s) = \frac{-s/R_1C}{s^2 + [2s/R_2C] + 1/R_1R_2C^2} \qquad (3F - 2)$$

Unlike the low-pass transfer functions we have been discussing, the numerator of this transfer function has the first power of s in it. Suppose that frequency (s) is very small. In the low pass case, this meant that the s^2 and s terms in the denominator could be disregarded, and a constant response was the result. Here, this is still true as far as the denominator is concerned, but the s in the numerator will cause the transfer function to be zero at zero frequency. Thus, the response begins low at low frequency. At very high frequencies, the s^2 term in the denominator dominated the low-pass and caused the response to fall off. In this bandpass case, the s^2 still has its effect, but the s in the numerator means that the roll off is as $1/s$ and not as $1/s^2$ so it is more gradual. In between, there is a region (a frequency band) where there is significant response, and this is the bandpass region.

We have already argued that there is a zero at $s = 0$, and it is easy to find the poles by factoring the denominator, solving the quadratic formula in this case:

$$p_1, p_2 = \frac{1}{R_2 C} \left[-1 \pm \sqrt{1 - R_2/R_1} \right]$$

As long as R_2 is greater than R_1, then we have a pair of complex conjugate poles. Let's take an example where $R_2/R_1 = 37$ in which case the poles lie at:

$$p_1, p_2 = (1/R_2 C)[-1 \pm 6j]$$

A pole-zero plot for this set of components is shown in Fig. 35.

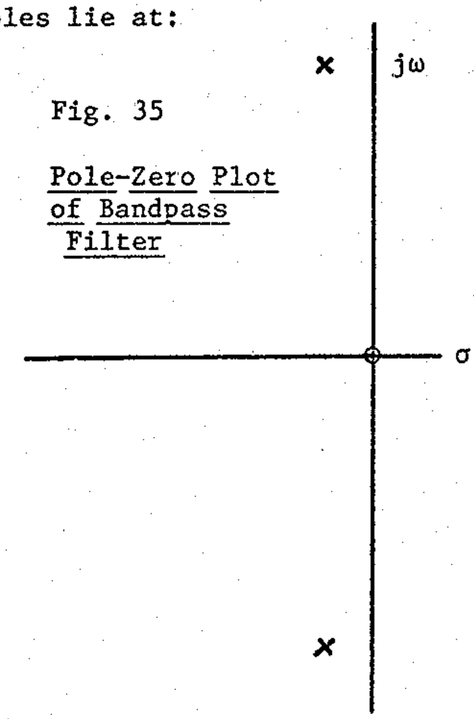


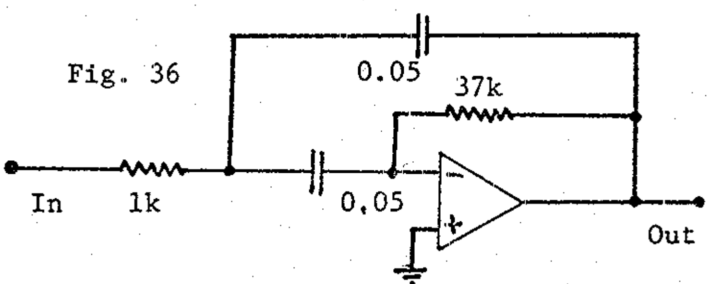
Fig. 35

Pole-Zero Plot of Bandpass Filter

Let's suppose we follow the jω-axis starting at 0 and moving in an upward direction. At $s = 0$, we are right on top of a zero. As we move away, we feel less infulence from the zero, and begin to feel higher ground due to the "mountain" ahead and to the left. When we are even with the pole, we are about at the point of maximum response. Moving even higher, we go down the slope and look back to see far in the distance two poles and one zero, all approximately at the same far away distance. If there were as many poles as there were zeros, they would cancel and we would be on flat ground (a high-pass filter) but since there is an extra pole back there, we are still going downhill (but not as fast as we would be if that zero weren't back there).

To graphically determine the frequency response, we select a point ω on the jω-axis and measure the distance to the zero (which is just ω of course), and the distances to the two poles we plotted in Fig. 35. We then divide the distance to the zero by the product of the distances to the poles, and this is the relative value of the response. We repeat this until we have enough points to sketch in the curve. Fig. 36 shows an experimental circuit. Fig. 37 shows the experimental data on this filter, and the theoretical, graphically determined points are plotted as large dots.

Fig. 36



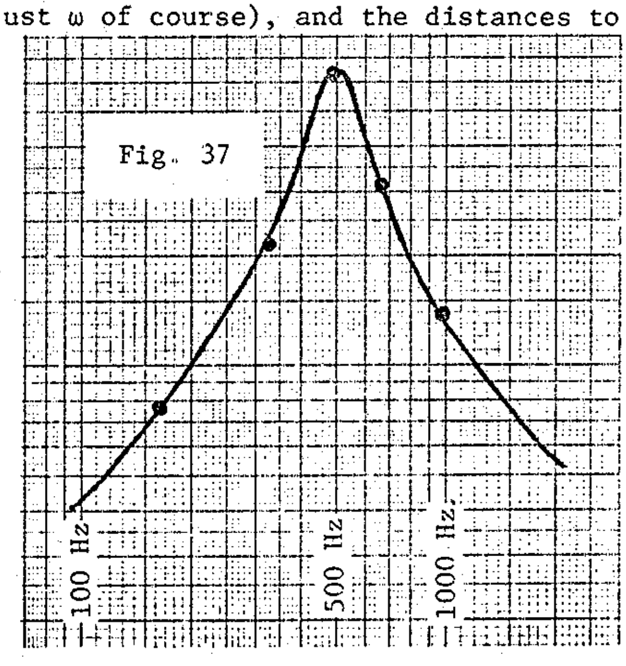
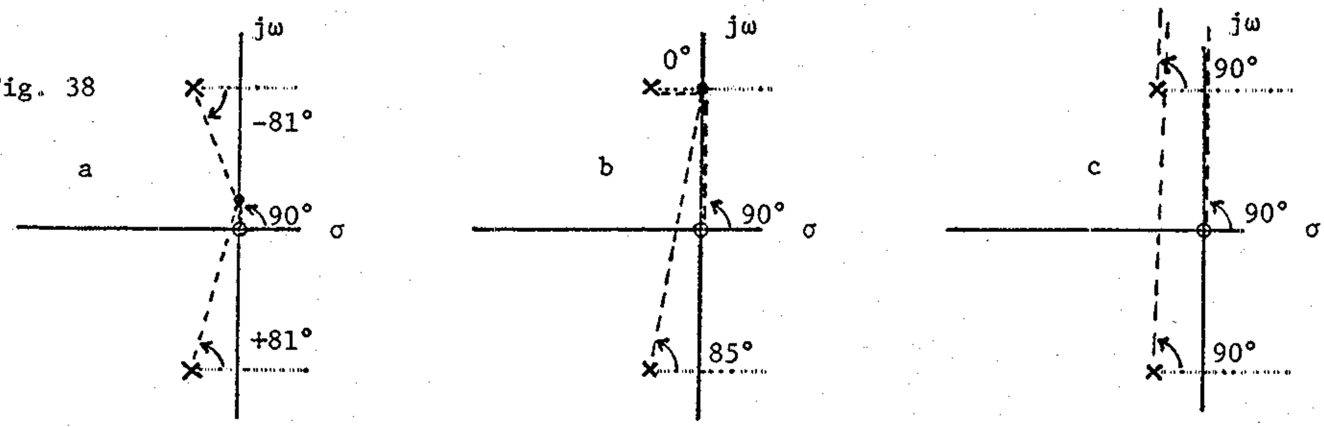


Fig. 37

Fig. 38



$90° - 81° - (-81°) = 90°$ $\quad 90° - 85° - 0° = 5°$ $\quad 90° - 90° - 90° = -90°$

To determine the phase response of the bandpass filter, we can use the graphical interpretation as applied to the pole-zero plot of Fig. 35. It is not difficult to set up analytical expressions for the phase response based on a graphical model, but here we will just look at three special cases as illustrated in Fig. 38. In 38a, the point of interest is assumed to be just slightly on the +j side of the jω-axis. If we take the position of the zero at s = 0, the point of interest is directly above, or at an angle of 90° relative to the +σ-axis. The angles of the point of interest as seen from the poles are +81° in the case of the lower pole, and -81° as seen from the upper pole. We add the angles as seen from the zero (+90°) and subtract the angles as seen from the poles (+81° and -81°) so that the total angle is +90°. In 38b the point of interest is even with the upper pole, and the total angle is about +5°. In 38c, the point of interest has moved far up the jω-axis, and the result is that all three angles are effectively 90°, and the total phase shift is -90°.

Often, bandpass filters are analyzed by comparing the denominator of T(s) with a standard denominator: $s^2 + (\omega_o/Q)s + \omega_o^2$. By comparing this with equation (3F-2), we determine that for the bandpass filter of Fig. 34:

$$f_o = \frac{1}{2\pi C\sqrt{R_1 R_2}} \qquad (3F - 3)$$

$$Q = (1/2)\sqrt{R_2/R_1} \qquad (3F - 4)$$

where $f_o = \omega_o/2\pi$. These two equations have a useful physical interpretation. The value f_o is the center frequency (maximum response) of the bandpass characteristic, and Q is a "quality" factor that tells us how sharp the bandpass is. We can show that Q is equal to:

$$Q = \frac{f_o}{f_{3db-u} - f_{3db-\ell}} \qquad (3F - 5)$$

where f_{3db-u} and $f_{3db-\ell}$ are the upper and lower 3db frequencies - the frequencies on either side of the peak at f_o where the response falls by 3db (down to 0.707) below the peak response. The difference between these two 3db frequencies is called the 3db bandwidth.

CHAPTER 4: CHARACTERIZATION OF ACTIVE FILTER RESPONSES

4A. INTRODUCTION

In this chapter we want to gather some of the ideas that we have touched on by means of example in the previous chapters. In particular, we will concentrate on terminology and the general relationships that exist among filter performance parameters.

4B. SOME TERMINOLOGY IN ACTIVE FILTERING

There are many terms that are used to describe active filters, and these can be confusing at times. Here we will list the more important terms and will attempt to describe their most important aspects and their interrelationships.

▶ BASIC TYPE OR BASIC FILTER FUNCTION: Selection of the basic filter type or basic function is determined by the application. That is, there is some basic job we want done. These may include low-pass, bandpass, high-pass, band-stop, all-pass, and notch. These basic response shapes are illustrated in Fig. 40. Note that the all-pass has a flat amplitude characteristic and is used for its phase properties.

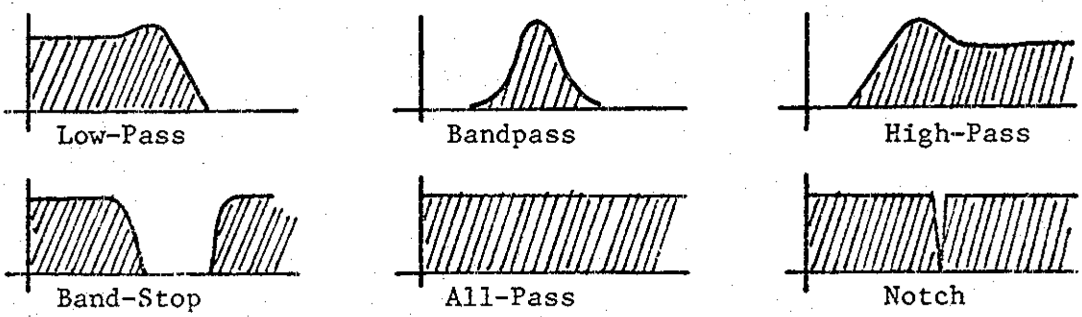


Fig. 40

▶ FILTER CHARACTERISTIC: The characteristic of the filter response is largely a matter of the amount of damping the filter has. This is determined by the positions of the poles of the network, which is in turn determined by the denominator of the transfer function. Certain types of filter responses are associated with certain mathematical polynomials that appear in the denominator of the transfer function, and since these special polynomials are named after mathematicians, the filter characteristic is given the same name. Thus we see filters with names like Gaussian, Bessel, Butterworth, and Chebyshev. These filter characteristics are associated with a certain desirable property of signal processing. The highest damping is generally that found in the Gaussian filter, which is "critically damped" (settles to a new voltage level as fast as possible without any overshoot at all). The roll-off corner of a Gaussian filter (or any filter damped more than Gaussian) is relatively poor. A much better corner is obtained with a Butterworth response, which is "maximally flat". The Butterworth filter does result in some overshoot however. In between Gaussian and Butterworth we find the Bessel characteristic, which has the useful property that it has linear phase (constant delay). This response is most useful when the shape of a desired waveform passing through must be preserved. There are a series of Butterworth-Thomson characteristics which lie between Butterworth and Bessel, and are a trade-off between the advantages of both. When damping is less than the Butterworth case, there is a series of characteristics called Chebyshev and this term is prefixed by the ripple in db. Chebyshev filters are those which give up the requirement that the passband response be flat. Instead, the poles are brought up close to the $j\omega$-axis and cause bumps or ripples in the response. The advantage is that if a final ripple occurs at the corner before cutoff, the response initially rolls off much faster than it would with other types of filter characteristic. The disadvantages of the Chebyshev filter are, in addition to the ripple, rather large overshoots when the voltage level changes.

TRANSFER FUNCTION: The transfer function, usually denoted $T(s)$ or $H(s)$ is a general output/input relationship. It relates the output $V_{out}(s)$ [the Laplace transform of the time output waveform $V_{out}(t)$] to the input $V_{in}(s)$ [the Laplace transform of the time input waveform $V_{in}(t)$]. The transfer function is obtained simply by considering all capacitors in the network to be "resistors" of "resistance" $1/sC$. The usual circuit laws are then used (sum of currents into a node = 0, sum of voltages around a loop = 0, and Ohm's Law) to arrive at $T(s) = V_{out}(s)/V_{in}(s)$. The mathematical form of $T(s)$ determines the basic type of filter and its characteristic. Most if not all useful transfer functions are in the form of ratios of polynomials in s, where the order of the numerator is equal to or less than that of the denominator. The simplest second-order forms are summarized by:

$$T(s) = \frac{s^n}{s^2 + Ds + 1} \qquad n = 0,\ 1,\ \text{or}\ 2$$

where D is related to the damping of the response. When $n = 0$, the filter's response for small s (low frequency) is just 1/1 because the s and s^2 terms in the denominator can be disregarded. For high frequencies, large values of s, the response is dominated by the $1/s^2$. Consider what happens when s doubles. The $1/s^2$ term decreases by a factor of 4. Now we know that 3db is $1/\sqrt{2} = 0.707$, so 6db would be 1/2 and thus 1/4 would have to be 12db. Since a 2:1 frequency ratio is called an octave, we say that a roll-off dominated by $1/s^2$ is 12db/octave. Note that a 1/s roll-off would be 6db/octave. Thus the $n = 0$ case corresponds to a low-pass response with a final 12db/octave roll-off. In the region between the low frequency response (where the 1 in the denominator is important) and the high frequency response (where the s^2 in the denominator is important), the Ds term has its effect. Thus the damping term is most important in the transition region between low and high frequency response. Fig. 41 illustrates these points.

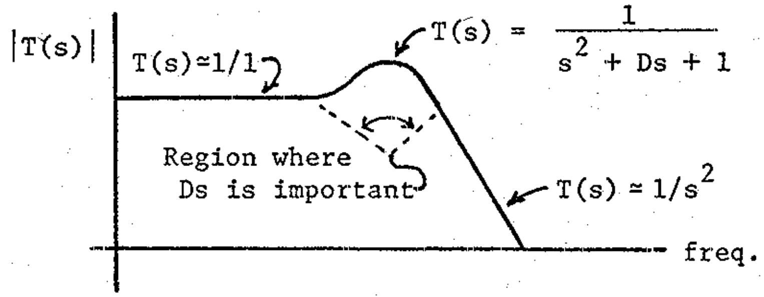


Fig. 41

Low-Pass Response

The case where $n = 2$ gives a high-pass response in a very similar manner (see Fig. 42). The bandpass case when $n = 1$ is similar as well (see Fig. 43) but note that the roll-up goes as s/1 and the roll-off as 1/s, and are 6db/octave slopes, not 12db/octave as we had in the low-pass and high-pass cases.

Fig. 42

High-Pass Response

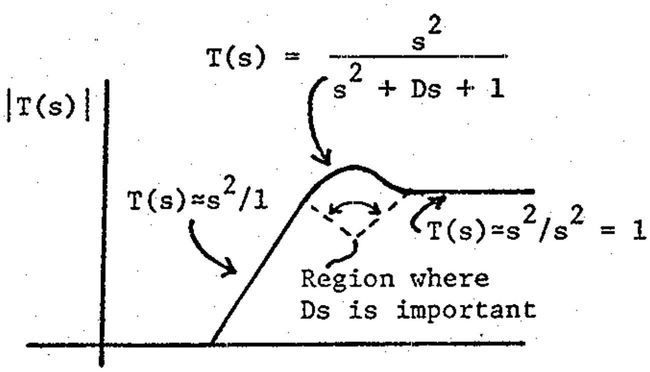
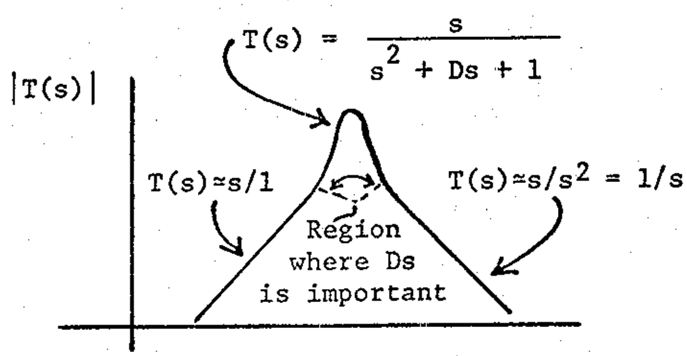


Fig. 43

Bandpass Response



FREQUENCY RESPONSE: The frequency response of a filter is the type of curve we get in a lab by putting in a sine wave of a certain frequency and seeing what portion of the signal comes out. The process is repeated until enough points are measured to plot a smooth curve. Mathematically, the frequency response is the magnitude of the transfer function denoted by $|T(s)|$ and we can substitute for s the value $j\omega$ where ω is radial frequency which is equal to $2\pi f$ where f is ordinary frequency like we read on the dial of a function generator. Since s is in general a complex number, and it certainly is when we set $s = j\omega$, a pure imaginary, to obtain the magnitude it is necessary to take the product of the complex number with its complex conjugate, and then take the square root. The complex conjugate of a complex number is obtained by replacing j by -j wherever it occurs. Hence, T(s) evaluated at $s = j\omega$ is obtained by substituting $j\omega$ for s in the expression for T(s), and the complex conjugate of T(s) is obtained by substituting $-j\omega$ for s. This is denoted as $T(j\omega)$ and the complex conjugate is $T(-j\omega)$. Hence, the frequency response is:

$$|T(s)| = [T(j\omega)\cdot T(-j\omega)]^{1/2}$$

PHASE RESPONSE: The transfer function T(s) is a complex number in general and it can be written as a magnitude and an angle in a complex plane. The angle is the phase shift which is a function of frequency. We can determine the phase angle $\phi(\omega)$ by knowing the real part of the transfer function and the imaginary part, and taking the inverse tangent, as with any complex number. Thus:

$$\phi(\omega) = \mathrm{Tan}^{-1}\frac{\mathrm{Im}[T(j\omega)]}{\mathrm{Re}[T(j\omega)]}$$

where Re means real part and Im means imaginary part. These are usually obtained simply by substituting $j\omega$ for s and separating out terms with j (imaginary) from those without j (real). Bear in mind that Im means a real number which when multiplied by j is imaginary. Thus Im[5 + 7j] = 7, not 7j. For example, we can look at the phase response of the first-order low-pass function of Fig. 25 with its transfer function:

$$T(s) = 1/[1 + sRC]$$

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

$$\mathrm{Re}[T(j\omega)] = 1/(1 + \omega^2R^2C^2) \qquad \mathrm{Im}[T(j\omega)] = -\omega RC/(1 + \omega^2R^2C^2)$$

$$\phi(\omega) = \mathrm{Tan}^{-1}[-\omega RC] = -\mathrm{Tan}^{-1}(\omega/\omega_o) \quad \text{where } \omega_o = 1/RC$$

Note that a graphical interpretation of the first-order low-pass phase function would give the phase as minus the angle of the point ω as seen from the pole at $-\omega_o$ as shown in Fig. 44. This is clearly the same answer we got by the mathematical procedure involving the real and imaginary parts of $T(j\omega)$.

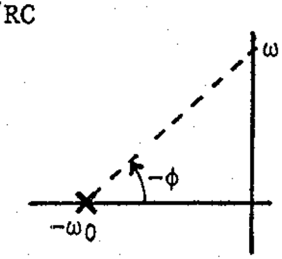


Fig. 44

IMPULSE RESPONSE: The impulse response of a filter is, much as the name implies, the response of the filter to a sharp impluse or spike. This response is a function that varies in time, and is thus an ordinary type of waveform, unlike $V_{out}(s)$ which is a Laplace transformed output of a time function. Mathematically an impulse is a delta function $\delta(t)$ and the Laplace transform of $\delta(t)$ is $\delta(s) = 1$. Since $V_{out}(s)/V_{in}(s) = T(s)$ and in this case $V_{in}(s) = \delta(s) = 1$, we get the Laplace transformed version of the response of the filter to an impulse as:

$$V_{out}(s)_{impulse} = T(s)\cdot V_{in}(s)_{impulse} = T(s)\cdot\delta(s) = T(s)\cdot 1 = T(s)$$

We have thus determined that the Laplace transformed version of the output which is in response to an impulse input is T(s). Stated another way, the inverse Laplace transform of T(s) is the impluse response. Thus, much as we see that T(s) is a description of the general response of a filter, so might the impulse response $h(t) = LT^{-1}[T(s)]$ be a complete description. In fact, this is true, and we might wonder why we don't just use the characterization of the filter as h(t), a time domain description, rather than jump to the less familiar Laplace or complex frequency s-domain. The answer to this question lies in the simplicity of an expression like $V_{out}(s) = T(s) \cdot V_{in}(s)$, which is a simple multiplication. According to Laplace transform theory, a multiplication in the s-domain corresponds to a convolution in the time domain, and convolution is more difficult to use than multiplication. To make the point in a slightly different way and at the same time give you some idea what is involved in convolution, let's think about how we would use the impulse response function to determine the response to an arbitrary input. We would do this by assuming the arbitrary input to be composed of an infinite set of different impulses, calculating the response to each input, and then summing all responses. Since the summation involves a continum of impulses, this must be done with integral calculus. The integral would look like:

$$V_{out}(t) = \int_{-\infty}^{\infty} V_{in}(t)\; h(t - x)\; dx$$

This may not be so bad, but if you use tables of transforms, it is usually easier to use T(s). Also, the determination of T(s) is straightforward, while the calculation of h(t) other than as the inverse Laplace transform of T(s) may be difficult.

▶ GROUP DELAY: Group delay may be defined as $\tau(\omega) = -d\phi(\omega)/d\omega$. The group delay is in units of time, and an important case is where $\tau(\omega)$ is a constant. This is the case of a Bessel filter. Note that if $\tau(\omega)$ is a constant then $\phi(\omega)$ must be a linear function of ω. This means that phase shift increases with frequency, and higher frequency components are shifted by larger phase angles. Linear phase thus appears as a constant time delay does.

▶ FILTER ORDER: The order of a filter is determined by the highest power of s that appears in the denominator of T(s). If the highest power is s^n, then the order is n. The numerator may have a power of n or any lesser powers, but none greater than n. Generally, a high-pass or low-pass filter of order n rolls up or off at 6n db/octave once it gets well beyond the corner (asymptotic roll-off as it is called). A bandpass filter must be of even order, and the response on either side of the peak (called the skirts of the bandpass) roll off at 3n db/octave. The filter order is also generally equal to the number of capacitors in the network. This means network capacitors and not any extra capacitors such as those that are for op-amp compensation. If the number of capacitors equals the order, the network is called canonic. A filter of order n also has n poles, some of which may have the same value. Often times a filter of order greater than two is realized by a cascade (series connection) of first and second-order sections. This generally gives a performance that depends less on small variations in component values than a "direct" realization does. The first-order sections realize any real poles while the second-order sections realize pairs of complex conjugate poles.

▶ CONFIGURATION: The configuration of a filter refers to the actual circuit that is comprised of resistors, capacitors, and op-amps that is used to realize a transfer function. The configuration generally determines the basic filter type, and a variation in some component or components within the configuration varies the characteristic. More than one configuration can be used to realize the exact same transfer function. Thus the advantage or disadvantage of one configuration or another has to do with such matters as the number of components involved (the cost and space required) and the sensitivity of the configuration to component variations. Sensitivity is a very important aspect of filter performance. It has to do with how much the filter's performance changes in proportion to a change in the value of certain components such as the variations due to the tolerance of a resistor or capacitor. The lower the sensitivity, the better. Direct realizations of order greater than two, functions with large passband ripples, and sharp bandpass or notch responses tend to present the greatest problems with sensitivity. When realizing higher order networks with combinations of second-order sections, some schemes called "leapfrog" structures which are based on passive ladder networks offer a lower sensitivity than the standard cascade.

4C. FILTER DAMPING AND ITS EFFECT ON CHARACTERISTIC

We have seen that the performance of a filter at extremes of frequency depends mainly on the basic type and the order, and relatively little on the characteristic of the filter. Here we will take a closer look at the effects of damping. It is the damping that controls the filter characteristic, and this characteristic is usually quite important as many frequencies of interest enter the filter and pass through a region where damping effects make a great difference.

First, we need a general idea of damping. Damping is basically just the reciprocal of resonance. While a resonant system tends to favor certain frequencies, a damped system tends to have a smoother response. It all depends on what you need for a certain application. At times you want more damping, and at others, you want more resonance. If you are familiar with the "Q" of a system, you will gain by realizing that essentially $Q = 1/D$ where D is the damping. We want to consider how damping relates to the positions of the poles of a system. We will look at the second order transfer function:

$$T(s) = \frac{s^n}{s^2 + Ds + 1} \qquad n = 0,\ 1,\ \text{or}\ 2 \qquad (4C\text{-}1)$$

The poles of the system are obtained by factoring the denominator, and are thus at:

$$p_1,\ p_2 = -D/2 \pm (1/2)\sqrt{D^2 - 4} \qquad (4C\text{-}2)$$

A plot of pole position as a function of D is shown in Fig. 45. By comparison with equation (3D-2) with RC=1 for the Sallen-Key low-pass, we see that D = 3-K in this case. Thus, we add to Fig. 45 the corresponding values of K which control the damping in the Sallen-Key filter.

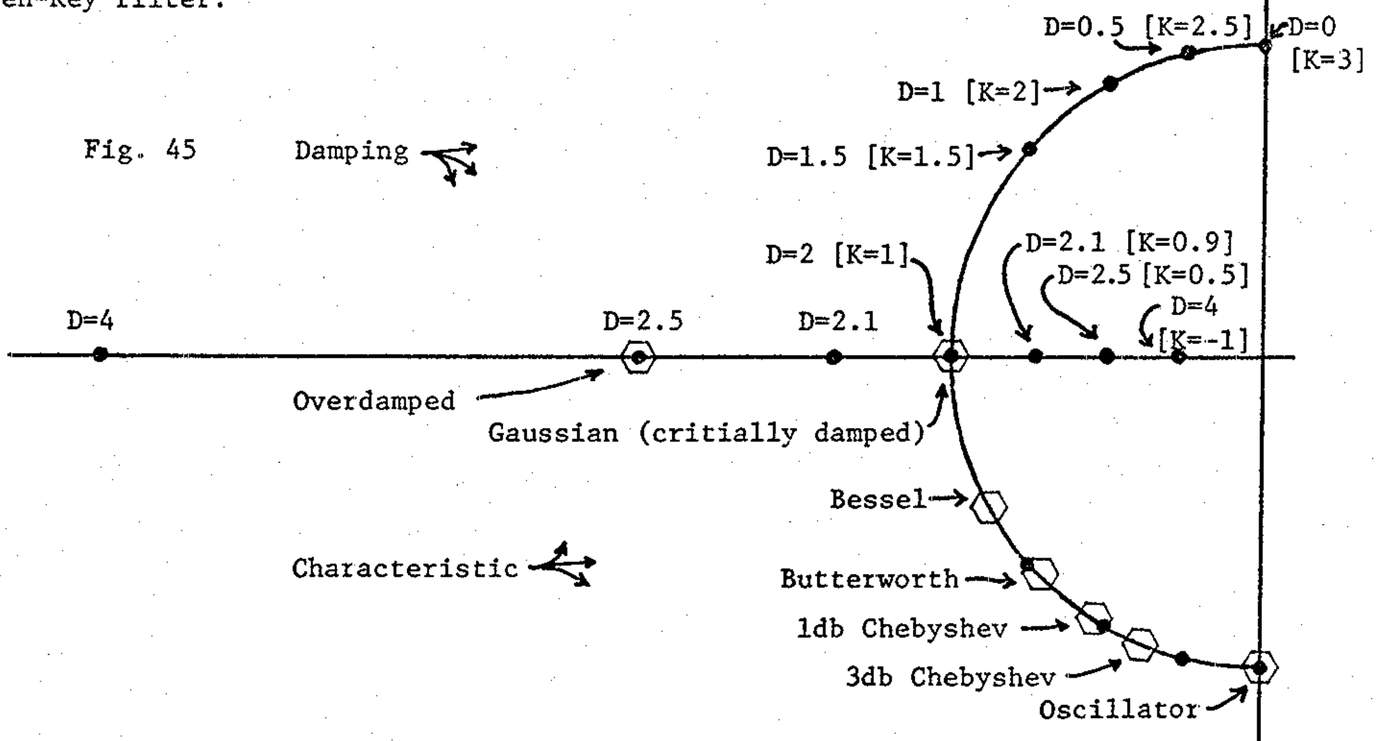


Fig. 45

Fig. 45 should be studied carefully. Let's begin at the top of the semicircle at D=0. Here we have a pole at +j and a corresponding pole at -j at the bottom of the semicircle. This is a totally undamped system (Q = infinity) and corresponds to oscillation. If we add some damping (D=0.5), the poles move back from the j-axis and oscillation stops, although the system remains quite resonant. By the time damping reaches 1.4, resonant behavior has ended and there are no bumps in the response. Between D = 1.4 and D = 2, the main difference in performance is seen in the step response which shows less and less overshoot and finally becomes critically damped at D = 2. Note that for D = 0 to D = 2 there are two complex conjugate poles, and that these come together at D = 2. Beyond D = 2, the system is overdamped. The two poles separate again and move in opposite directions but always on the real axis. According to equation (4C-2), one pole approaches but never reaches s = 0 while the other one moves out almost to -D as D gets very large. This results in what looks mainly like a one pole system except for very high frequencies. There is seldom any reason to use an overdamped filter, so we are mainly interested in filters which have complex conjugate poles on the semicircle between D = 0 and D = 2.

In order to give some meaning to the filter characteristics that are named after mathematicians, on the lower part of the semicircle we have labeled regions not by their D or K values (which are the same as the complex conjugate poles directly above) but by the filter characteristic. These regions are marked with hexigons. The extremes go from overdamped and critical damping (Gaussian) to oscillation. The main area of interest runs from Bessel (linear phase) through Butterworth (maximally flat) to 3db Chebyshev (which is about the maximum ripple that can be tolerated).

It may seem to the reader that filter characteristic is simply a matter of damping, and it might seem strange that there are a whole slew of names for characteristics that seem to differ only by the amount of one parameter. This is because we have only looked at second-order cases. [If we looked only at first-order, all the filter characteristics become the same - Gaussian.] In the case of fourth-order filters for example, a characteristic is determined by three parameters: the damping of the first two poles, the damping of the second two poles, and the ratio of the radii of the circles on which the poles lie. Thus, while saying that a characteristic is second-order Butterworth is saying only that the damping is a certain value, saying that it is tenth-order Butterworth tells us eight more parameters, and so on. While a second-order case may appear as a simple, trial-and-error setup, a tenth-order system is clearly not something we can adjust to our needs without aid of a mathematical theory, and this is where the mathematics pays off.

Chapter 5: Low-Pass Active Filter Examples And Experiments

5A. SECOND-ORDER LOW-PASS FILTERS

At this point, we are familiar with the second-order Sallen-Key low-pass filter. Basically, all we need now to realize a given filter characteristic is the required value of damping, and some means of determining where the cutoff frequency falls relative to the RC time constant of the filter. The table below provides the necessary design data.

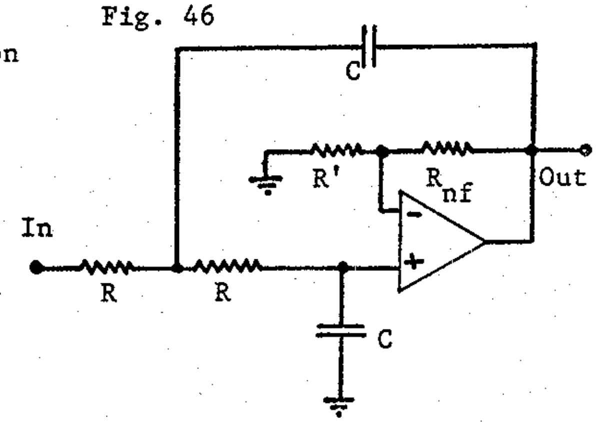
CHARACTERISTIC	D	K=3-D	f_{sd}/f_{3db}	R_{nf}/R'	Best 5% R_{nf}	Best 5% R'
Gaussian	2.00	1.00	1.55	0	0	omit
Bessel	1.73	1.27	1.27	0.268	15k	56k
Butter-Thomson	1.56	1.44	1.13	0.435	27k	62k
Butterworth	1.41	1.59	1.00	0.586	30k	51k
½db Chebyshev	1.22	1.78	0.93	0.784	30k	39k
1db Chebyshev	1.05	1.95	0.86	0.955	30k+1.5k	33k
2db Chebyshev	0.90	2.10	0.85	1.105	62k	56k
3db Chebyshev	0.77	2.23	0.84	1.233	27k	22k

TABLE 1

The data given in Table 1 refers to low-pass filters in general and the Sallen-Key configuration of Fig. 46 in particular. The transfer function of the Sallen-Key is:

$$T(s) = \frac{K/R^2C^2}{s^2 + \frac{s}{RC}[3 - K] + 1/R^2C^2} \qquad (5A-1)$$

Fig. 46



If you only intend to use the Sallen-Key, the only other thing you need to know is that f_{sd} is the "section design" frequency, and:

$$f_{sd} = 1/2\pi RC \qquad (5A-2)$$

We first choose the 3db frequency of the filter we need, and then select a characteristic we want to use. The ratio f_{sd}/f_{3db} from the tables gives us f_{sd} and the associated RC product. It is then only necessary to use the suggested resistor values, or other similar choice, to set the gain factor K (set the damping) to the proper value for the characteristic.

In the case of a low-pass filter with a different configuration, it is possible to put the transfer function in a form:

$$T(s) = \frac{A/\tau^2}{s^2 + (D/\tau)s + 1/\tau^2} \qquad (5A-3)$$

It is then possible to identify f_{sd} with $1/2\pi\tau$. Then the f_{sd}/f_{3db} values and the D values are obtained from Table 1, and the design can be completed. The factor A is an overall gain factor which is a constant, much as K is a gain constant in the case of the Sallen-Key filter.

At this point, an example will serve to clarify the design procedure. We will choose the design of a 2db ripple Chebyshev filter with 3db cutoff frequency of 1220 Hz. From Table 1 we see that we need (for the Sallen-Key configuration) K = 2.1 and $f_{sd} = 1220 \cdot 0.85 = 1037$ Hz. Thus we need $RC = 1/(2\pi \cdot 1037) = 154$ microseconds. If we choose C = 0.0511 mfd (just an arbitrary selection of a part on hand), we get R = 3k. Also from Table 1 we choose R_{nf} = 62k and R' = 56k. The circuit is shown in Fig. 47 and the experimentally measured response is shown in Fig. 48.

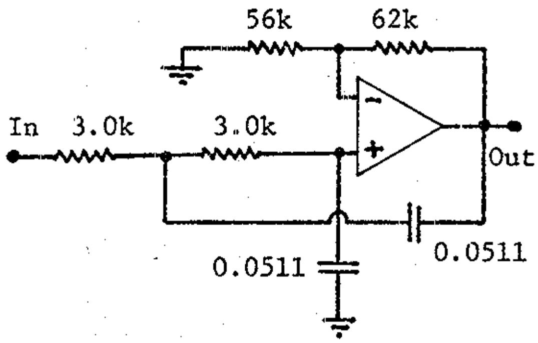


Fig. 47

2db Chebyshev

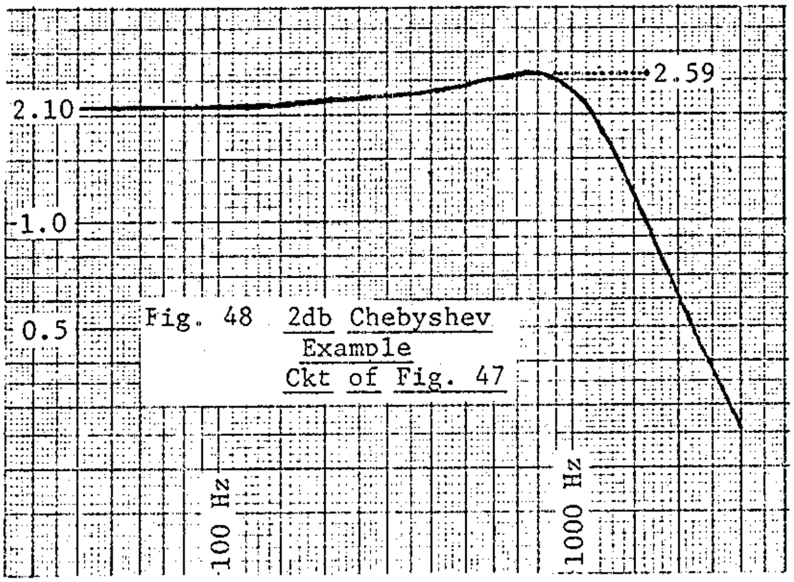


Fig. 48 2db Chebyshev Example Ckt of Fig. 47

TABLE 2	
+0.5db	= 1.06
-0.5db	= 0.94
+1.0db	= 1.12
-1.0db	= 0.89
+2.0db	= 1.26
-2.0db	= 0.79
+3.0db	= 1.41
-3.0db	= 0.71

There are several important points to note about the results shown in Fig. 48. First, the low frequency gain is 2.1, the same as the gain factor K. This can be understood if we realize that at low frequencies, the capacitors are effectively out of the circuit, and the two 3k resistors simply feed the signal through to the (+) input of the non-inverting amplifier which has gain K. Next note that the relatively low damping (0.9) leads to a peaking as we approach the corner. The peak is at 2.59, which is approximatley 2db above 2.1. It is this 2db of peaking that is called "ripple". In the case of higher order Chebyshev filters, there will be more than one ripple, but they are all of the same amplitude. For convenience, a table of decimal equivalents of the db values commonly used with Chebyshev filters is given in Table 2. Now, the data which we have given here give the 3db frequency as being down 3db from the peak value. Thus, the response has fallen to 3db below the peak when it reaches $2.59 \cdot 0.707 = 1.83$, and examination of the graph of Fig. 48 shows that this in fact does occur at approximately 1220 Hz, which was our design goal. Some tables of filter data may give different standards for the cutoff frequency. For example, it is also common to find the cutoff frequency of Chebyshev filters defined in terms of the ripple amplitude. For example, in such a system a 2db Chebyshev would have a cutoff frequency defined as being down 2db from the peak, and so on. We shall always use here 3db for all filters.

Before going on to fourth-order filters it will be useful to set up the equations for another second-order low-pass configuration commonly called "infinite gain, multiple feedback". The configuration as illustrated in Fig. 49 is not completely general because we find it simpler to just use equal valued capacitors. The points we want to show will be evident however. The analysis becomes very simple when we recognize that the op-amp, the capacitor between the output and the (-) input, and R_2 form the standard inverting integrator we studied in Section 1F. Hence we can easily see that the output is just the negative integral of the unknown voltage V', and thus:

$$V' = -V_{out} sCR_2 \qquad (5A\text{-}4)$$

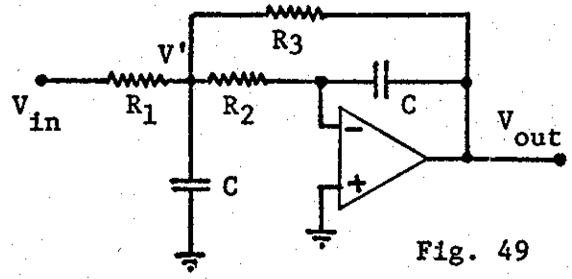


Fig. 49

With the unknown node voltage identified in this way, it is possible to write the current summing equation for the V' node:

$$\frac{V_{in} - V'}{R_1} = \frac{V' - V_{out}}{R_3} + \frac{V'}{R_2} + V'sC \qquad (5A\text{-}5)$$

and substitute equation (5A-4) into equation (5A-5). This done, we can solve for $T(s) = V_{out}(s)/V_{in}(s)$:

$$T(s) = \frac{-1/R_1R_2C^2}{s^2 + \frac{s}{C}\left[\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}\right] + \frac{1}{C^2R_2R_3}} \qquad (5A\text{-}6)$$

In order to read the characteristic frequencies, damping, and overall gain off the equation for T(s), we must achieve the form of equation (5A-3). With a little manipulation, equation (5A-6) takes on the required form:

$$T(s) = \frac{-(R_3/R_1)\ (1/R_2R_3C^2)}{s^2 + \frac{s}{C\sqrt{R_2R_3}}\left[\frac{\sqrt{R_2R_3}}{R_1} + \sqrt{R_3/R_2} + \sqrt{R_2/R_3}\right] + \frac{1}{R_2R_3C^2}} \qquad (5A\text{-}7)$$

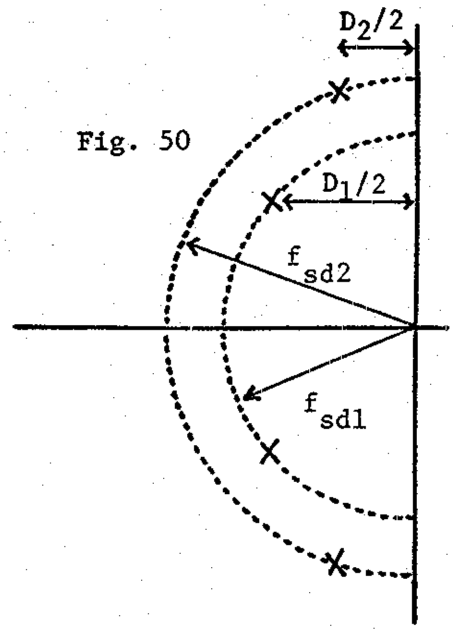
While the form of equation (5A-7) seems much more complicated than (5A-6), it is possible to read off the important parameters directly from (5A-7). The section design frequency is just $1/(2\pi\sqrt{R_2R_3}\,C)$, the gain is $-(R_3/R_1)$, and the damping D is equal to the quantity in the square brackets [ ] in the denominator.

<u>EXERCISE</u>: Show that if the gain of the circuit of Fig. 49 has a DC value of -1 that the filter is always overdamped. Rework equations (5A-4) through (5A-7) with all resistors of equal value, and capacitors αC and C/α (where α is a constant, so that the product of the two capacitances is C^2). Determine how the damping is controlled by the factor α. Discuss any advantages of this arrangement.

5B. FOURTH-ORDER LOW-PASS FILTERS

It is the usual practice to form fourth-order filters by cascading two second-order sections. In such a case, it is generally necessary that the two sections be different from each other, and the problem reduces to determining the paramaters of the two individual sections, and then constructing them just as we did in the second-order case. We can visualize the basic process by examining Fig. 50 which shows a four-pole system. These poles could be placed, as shown, by a direct fourth-order filter, but as shown by the circles, it is also possible to place the poles by two second-order pole sets each with its own section design frequency (also called pole frequency) and its own damping.

Fig. 50



CHARACTERISTIC	D	K=3-D	f_{sd}/f_{3db}	R_{nf}/R'	Best 5% R_{nf}	Best 5% R'
Gaussian	2.00	1.00	2.30	0	0	omit
Bessel	1.92	1.08	1.44	0.084	4.7k	56k
Butter-Thomson	1.88	1.12	1.20	0.119	5.6k	47k
Butterworth	1.85	1.15	1.00	0.152	15k	100k
½db Chebyshev	1.53	1.47	0.71	0.466	5.6k	12k
1db Chebyshev	1.28	1.72	0.50	0.725	24k	33k
2db Chebyshev	1.09	1.91	0.47	0.912	62k	68k
3db Chebyshev	0.93	2.07	0.44	1.071	16k	15k

TABLE 3 SECTION 1 SECTION 2 4th ORDER LOW-PASS

CHARACTERISTIC	D	K=3-D	f_{sd}/f_{3db}	R_{nf}/R'	Best 5% R_{nf}	Best 5% R'
Gaussian	2.00	1.00	2.30	0	0	omit
Bessel	1.24	1.76	1.61	0.759	47k	62k
Butter-Thomson	0.95	2.05	1.27	1.051	100k+5.1k	100k
Butterworth	0.77	2.23	1.00	1.235	27k	22k
½db Chebyshev	0.46	2.54	0.971	1.537	20k	13k
1db Chebyshev	0.28	2.72	0.943	1.719	62k	36k
2db Chebyshev	0.22	2.78	0.946	1.776	91k	51k
3db Chebyshev	0.18	2.82	0.950	1.821	20k	11k

As an example, suppose we want to design a fourth-order 3db ripple Chebyshev filter with a 3db cutoff frequency of 325 Hz. From the data in Table 3 we see that the first section should have a section design frequency $f_{sd} = 0.44 \cdot 325\text{Hz} = 143$ Hz and a damping of 0.93. The second section has $f_{sd} = 0.95 \cdot 325\text{Hz} = 309$ Hz and a damping of 0.18. With all capacitors taken to be 0.0511 mfd, equation (5A-2) gives the resistor values of 22k for the first section and 10k for the second section. The proper damping for each section is set using the suggested best 5% resistors from Table 3. The complete circuit is shown in Fig. 51. The pole-zero plot for the circuit is shown in Fig. 52 while the experimentally measured frequency response curve is shown in Fig. 53. Note from Fig. 53 that the low-frequency gain is approximately 5.8, the product of the two K values of the two sections. The ripple is also seen to be 3db and the 3db cutoff frequency falls very close to 325 Hz.

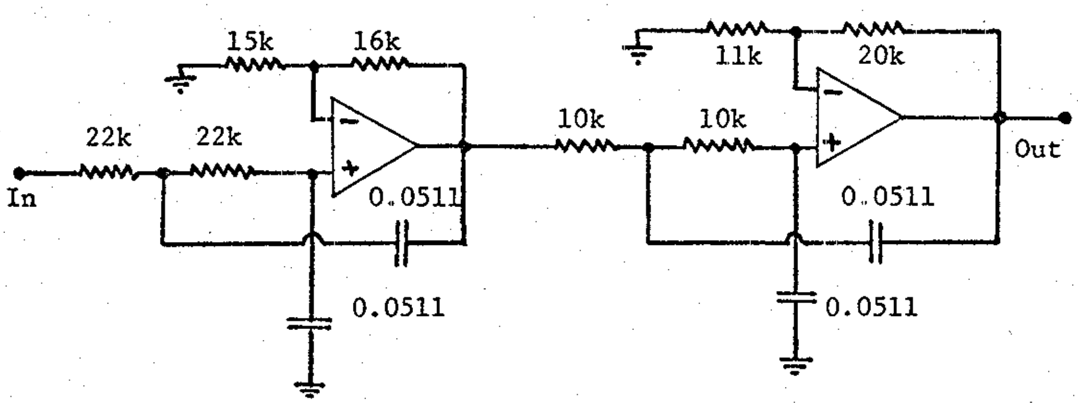


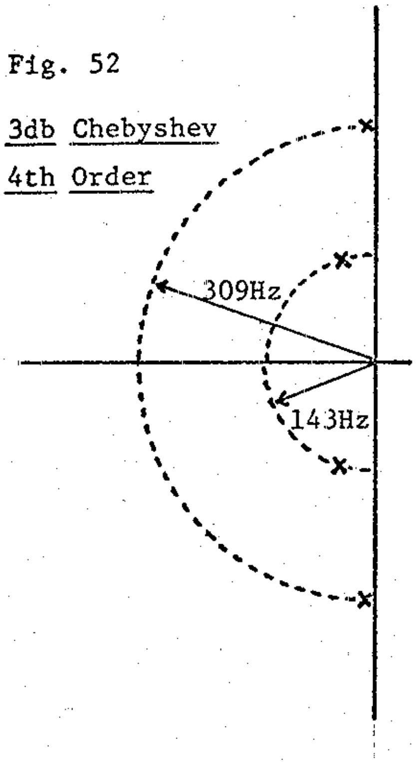
Fig. 51

Fourth-Order 3db Chebyshev 325 Hz Cutoff

Fig. 52

3db Chebyshev

4th Order



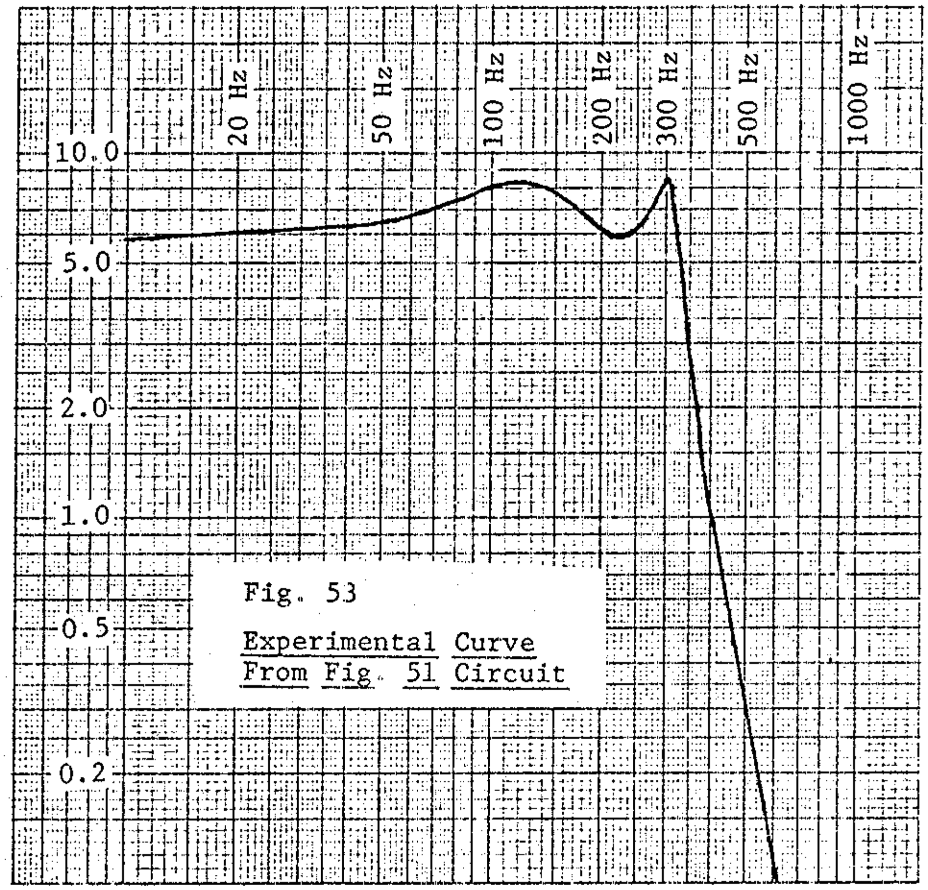


Fig. 53

Experimental Curve
From Fig. 51 Circuit

Note that the 4th-Order 3db Chebyshev represents about the limits of the Sallen-Key configuration. From Fig. 52 we see that the poles are quite close to the jω-axis and it is a factor of 3-K that moves the poles away from the axis. In the case of the poles on the outer radius of Fig. 52, K has reached 2.82, which is within 6% of 3.00, so we see that we are approaching the limits of ordinary component tolerances. In fact, the results shown in Fig. 53 are much better than we would expect with the 5% resistors used, and this can probably be attributed to the 1% capacitors that were used. Unless components with tolerances of 2% or less are available, the 4th-Order 3db Chebyshev is not suggested as a lab project. Butterworth and 1db Chebyshev are probably better choices for lab experiments.

5C. EXPERIMENT NO. 3, LOW-PASS FILTERS

1. Build and test the first-order low-pass filter in Fig. 28 of Chapter 3. Verify the cutoff frequency and the final roll-off rate.

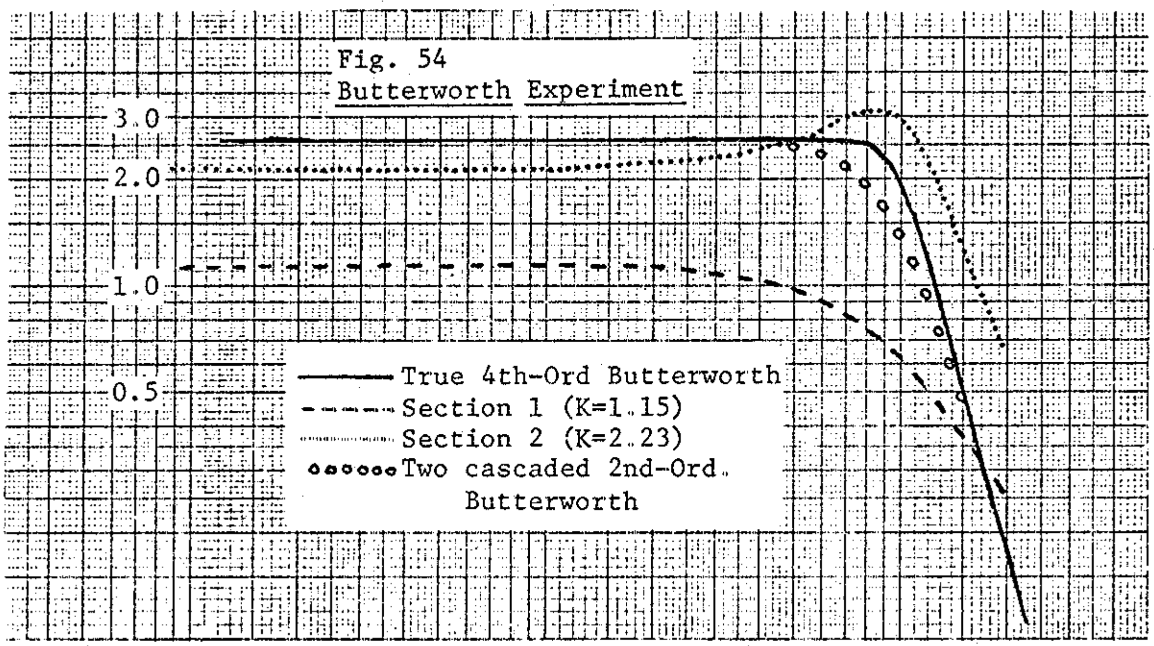
2. Build and test two or more of the second-order low-pass filters according to the data in Table 1 of this chapter using the circuit of Fig. 46. Verify the cutoff frequencies, final roll-off rates, and ripple (if any). By applying a low frequency square wave, measure the amount of overshoot of the filter relative to the level of the square wave peak-to-peak amplitude.

3. Build and test a fourth-order Butterworth filter according to the data in Table 3 of this chapter. Test each of the two sections separately. Compare the response of section 2 with your results for a 3db Chebyshev (or use the 3db Chebyshev of Chapter 3). Plot the response of the two cascaded sections and verify that it is the product of the response of the two sections taken individually. Build two second-order Butterworth filters, cascade them, and show that the resulting response, while fourth-order, does not have as sharp a corner as the true fourth-order Butterworth.

4. If 1% components are available, build and test a fourth-order 3db Chebyshev

filter and verify that the results are as predicted. Now, intentionally disturb one or more of the components by an amount in the range of 5% to 10% and remeasure the response curve. Describe the changes. Which components seem to be the most sensitive?

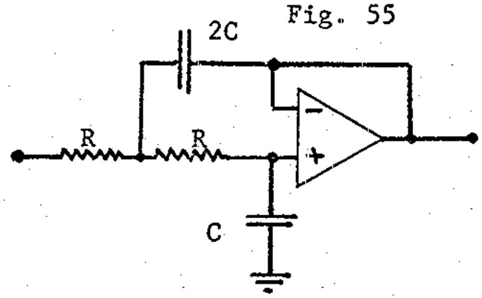
TYPICAL RESULTS: Some typical results for these experiments can be found in figures 27, 32, 48, 53, and in Fig. 54 below for the Butterworth case.



EXERCISES:

Show that the configuration in Fig. 55 with the component value relationship shown is a second-order Butterworth. Discuss the advantages and disadvantages of the circuit compared to the Sallen-Key structure used earlier.

Fig. 55



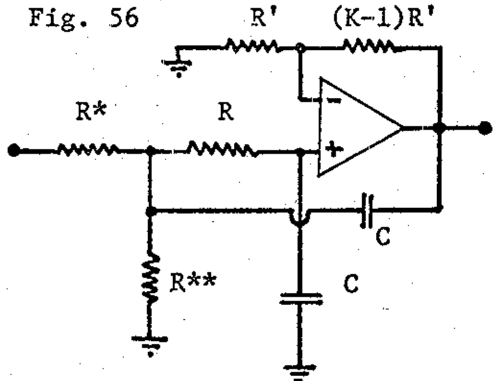
Show that if the input resistor R of a Sallen-Key filter is replaced with a voltage divider R*-R** as shown in Fig. 56 where:

$$R^* = KR$$

$$R^{**} = \frac{KR}{(K-1)}$$

that the response curve remains the same, except the low-frequency gain is now unity instead of K. In a fourth-order Sallen-Key filter, would there be an advantage to using this technique on each of the two stages separately rather than just using it on the first stage to compensate for the product of the two gain factors?

Fig. 56



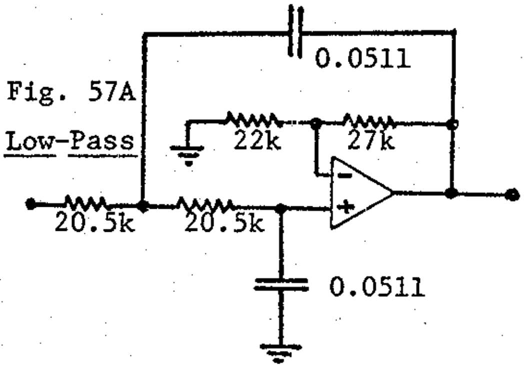
CHAPTER 6: HIGH-PASS, BAND-PASS, ALL-PASS, AND NOTCH FILTERS

6A. INTRODUCTION

Having now completed an extensive study of the low-pass active filter in various forms, we can now take up a study of filters of other basic types. A brief review of Chapter 4 might be useful at this point. Because much of the analysis of these other basic types is very similar to that of the low-pass, a good deal of detail can be left out here. Also, the bandpass filter was discussed in some detail in Chapter 3, Section 3F, so no additional analysis is needed here. Thus, we will be mainly concerned with the transformation of a low-pass to a high-pass filter, and will treat the all-pass and the notch filters as special cases. Finally, experiments on all four types of filters listed in this chapter title will be outlined.

6B. THE HIGH-PASS FILTER

Fig. 57A
Low-Pass



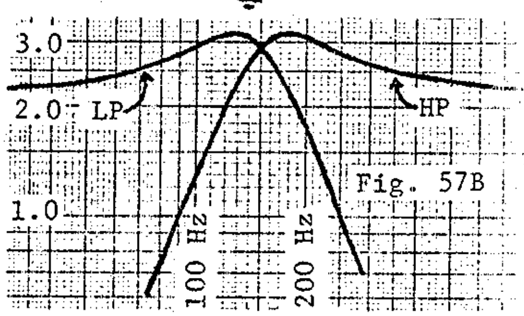
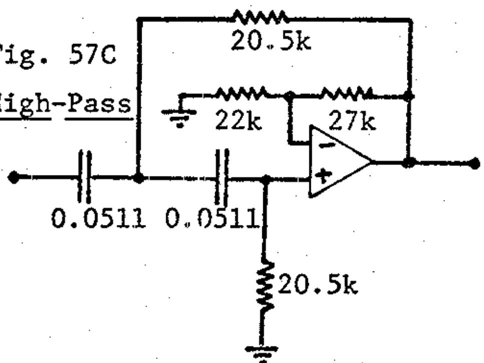


Fig. 57B

Fig. 57C
High-Pass



It is a relatively simple matter to transform a low-pass filter to a corresponding high-pass filter. A simple example is shown in Fig. 57. Fig. 57A shows our standard Sallen-Key low-pass set for a second-order 3db ripple Chebyshev characteristic. Fig. 57B shows the low-pass response. In Fig. 57C we show the same circuit with the resistors and capacitors interchanged (leaving the resistors determining K the same). The response of Fig. 57C is also plotted in Fig. 57B, and we can see that we have achieved a mirror image of the low-pass response. This is essentially all there is to it in this and many other cases, but we do want to be sure we know where the cutoff frequencies appear. From Fig. 57B it is clear that both filters have different 3db frequencies. The 3db frequency of the low-pass is about 180 Hz while the 3db frequency of the high-pass is about 130 Hz. What the two must have in common is the section-design (pole) frequency which is given by $1/2\pi RC = 152$ Hz, which is nearly in the middle of the two 3db cutoff frequencies, and by no coincidence, at the intersection of the two curves. In the low-pass case, the 3db frequency and section-design frequencies are related according to a constant (0.84) as given in Table 1 of Chapter 5. In the high-pass case, the relation between 3db frequency and the section design frequency is just the reciprocal of the corresponding low-pass case. Thus we would expect the high-pass 3db frequency to be at 0.84 times the section design frequency or at $0.84 \cdot 152 = 128$ Hz. In the low-pass case, the section design frequency is 0.84 times the 3db frequency or at $0.84 \cdot 180 = 151$ Hz. There is no need to construct tables such as Table 1 and Table 3 that appeared in Chapter 5 when we need data for the high-pass case. We can just use the low-pass data, inverting the relationship between 3db frequency and section design frequency. For example, suppose we want a 1db Chebyshev second-order high-pass with 3db cutoff frequency at 1000 Hz. Using Table 1 from Chapter 5, we see that we need $K = 1.95$. The low-pass case has $f_{sd}/f_{3db} = 0.86$, so for high-pass, $f_{3db}/f_{sd} = 0.86$, which means that we need a section-design frequency of $1000/(0.86) = 1163$ Hz in this case.

EXERCISE: Consider what happens when the input of the high-pass in Fig. 57C is allowed to float, and compare this to what happens when the input of Fig. 57A floats.

6C. THE ALL-PASS FILTER

The basic idea behind an all-pass filter design is that if we place a zero in the right half-plane that is the mirror image of a pole in the left half-plane, the frequency response due to this pair is a constant. This is because for any point on the $j\omega$-axis, the distance to the pole is exactly the same as the distance to the zero. Thus, since the frequency response can be determined by multiplying all distances to the zeros and dividing by all distances to the poles, as long as all poles and zeros exist in pairs mirrored about the $j\omega$-axis, the response is constant with frequency.

A popular all-pass filter that is often used for its phase shifting properties is shown in Fig. 58. The steps in the analysis are as follows:

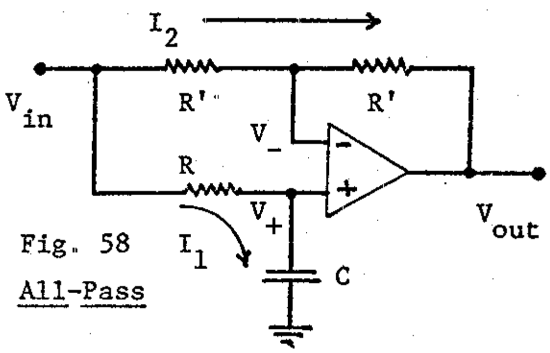


Fig. 58 All-Pass

$$I_1 = \frac{V_{in}}{R + 1/sC} = \frac{sCV_{in}}{1 + sCR}$$

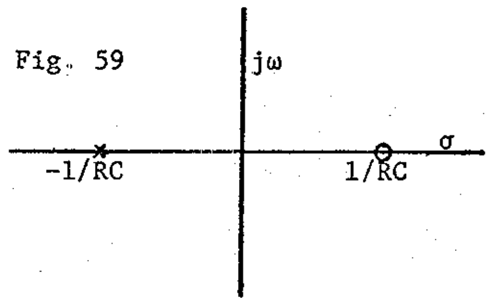
$$V_+ = V_- = I_1(1/sC) = \frac{V_{in}}{1 + sCR}$$

$$I_2 = (V_{in} - V_-)/R'$$

$$V_{out} = V_- - I_2R' = V_{in}\,\frac{1 - sCR}{1 + sCR}$$

$$T(s) = V_{out}/V_{in} = -\frac{s - 1/RC}{s + 1/RC}$$

Fig. 59



This circuit has a pole at $-1/RC$ and a zero at $+1/RC$ as shown in Fig. 59, so it is clear that the circuit is all-pass. Note that at low frequencies, the capacitor can be considered out of the circuit. The reader can verify that the resulting structure is actually a follower, and thus the phase response at DC is zero. At high frequencies, the phase lag goes to 180°.

6D. THE NOTCH FILTER

There are many circuits available for notch filters. The one we will discuss here is shown as much for its teaching value as its actual utility in practice. To obtain a notch response, it is desirable to have a zero occur on the $j\omega$-axis at some point other than at $s = 0$. Thus the numerator of T(s) will have to take on the form $s^2 + \omega_0^2$ where ω_0 is the frequency at which the response is to be zero. With this numerator, there is a zero at $\pm j\omega_0$. Fig. 60 shows a prototype of the active filter we wish to develope. It includes an inductor which we will eventually want to replace with a capacitor and an op-amp. Briefly we can see that the circuit will provide a notch because at a certain frequency, the LC combination will become resonant and take on a zero impedance, leaving only a resistor R in the ground leg. This leaves a differential amplifier with both inputs connected, and thus their difference should be zero. In the case where the LC combination is not resonant, we expect it to add an impedance much greater than R in the ground leg, effectively isolating this leg and giving us the same follower circuit that we saw in the all-pass filter above.

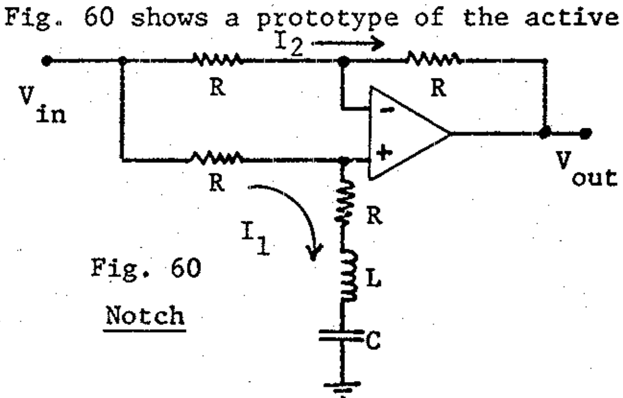


Fig. 60 Notch

We mentioned briefly that if you have an inductor, you mentally replace it with a "resistor" of resistance sL and solve the network as though it were composed of only resistors. With this knowledge, the reader should have no trouble following the general analysis of a differential amplifier and arriving at the transfer function of the network as:

$$T(s) = \frac{s^2 + 1/LC}{s^2 + (2R/L)s + 1/LC}$$

We note from this that T(s) has two zeros, at $\pm j/\sqrt{LC}$, so the notch appears at $1/2\pi\sqrt{LC}$ and the damping goes as R/L (the Q increases as the resistance decreases). Both these facts should be in line with what the reader probably knows about LC resonant circuits.

We have next to get rid of the inductor by using the circuit shown in Fig. 61, where we have only to show that the portion of the circuit below the node marked with a (*) is the same as a resistor and series inductor. By inspection, we can write three equations for the voltage V':

$$V' = (I_1 + I')(R/2)$$
$$V' = V^* - I_1(R/2)$$
$$V' = V^* - I'(1/sC')$$

We can solve these for $Z = V^*/I_1$ where Z is the input impedance as seen from the (*) node. The result is:

$$Z = V^*/I_1 = R + s[\frac{C'R^2}{4}]$$

Thus, the equivalent series inductance is given by:

$$L = C'R^2/4$$

This can be substituted back into the prototype.

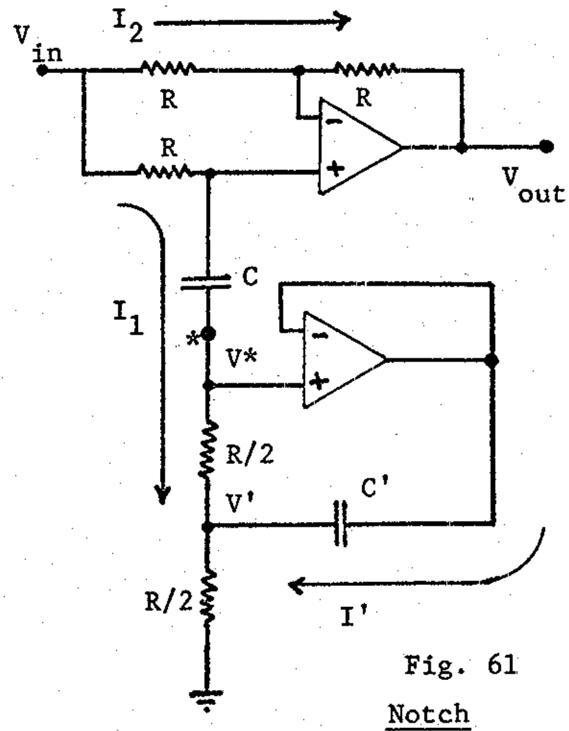


Fig. 61

Notch

6E. EXPERIMENT NO. 4, HIGH-PASS, BAND-PASS, ALL-PASS, AND NOTCH

1. Build and test a second-order 2db ripple Chebyshev low-pass with a 3db cotoff of 440 Hz, and construct and test the corresponding high-pass with the same cutoff frequency. The two filters should overlap at the 3db points, and not at the section design frequency as in Fig. 57B. Investigate the very high frequency behavior of the high-pass where the response starts to drop back down again, and associate this drop with the limitation of the op-amp you are using. Construct and test a fourth-order high-pass filter with a Butterworth characteristic, or with a 1 or 2 db ripple characteristic, by using the data in Table 3 of Chapter 5, converting the data for high-pass as needed.

2. Build and test the all-pass filter of Fig. 58 and examine the phase response using Lissajous figures or other phase measuring means. Discuss whether or not there is any reason to prefer that R' = 2R.

3. Complete all the necessary design equations for the notch filter as outlined in Section 6D. Build and test the filter, being careful to note the depth of the notch. Discuss how it might be possible to trim a single resistor

to obtain a deeper notch. Show what happens if the two resistors shown as R/2 are changed, but always so that their sum is R. What does this indicate about the possibility that the notch position may be tuned with a single potentiometer.

4. Build and test the bandpass circuit of Fig. 36, or with other values. Verify the equations (3F-3 and 3F-4) for the center frequency and the Q by using equation (3F-5) to determine Q experimentally. Apply the general technique associated with Fig. 56 to obtain filters with higher Q without going to unacceptable spreads in resistance between R1 and R2 (as in Fig. 34). For high Q filters, it is often more useful to measure the Q by means of "ringing" the filter as shown in Fig. 62. The ring time (T_r) is taken as the time for the envelope of the ring to decay from a starting value to $1/e = 0.37$ of that value. With the ringing frequency F_c (essentially taken as the center frequency of the bandpass), the Q is given by:

$$Q = \pi F_c T_r \qquad \text{(6E-1)}$$

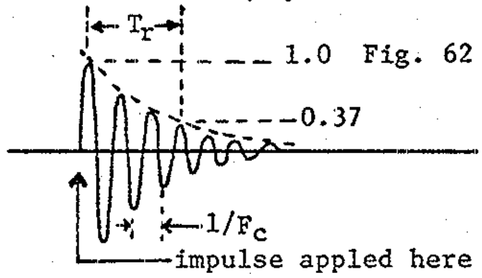


Fig. 62

TYPICAL RESULTS: Typical results for high-pass are shown in Fig. 57B, or by reflecting the data for the corresponding low-pass. Typical bandpass results are shown in Fig. 37. Typical results for the notch filter are shown in Fig. 63 below:

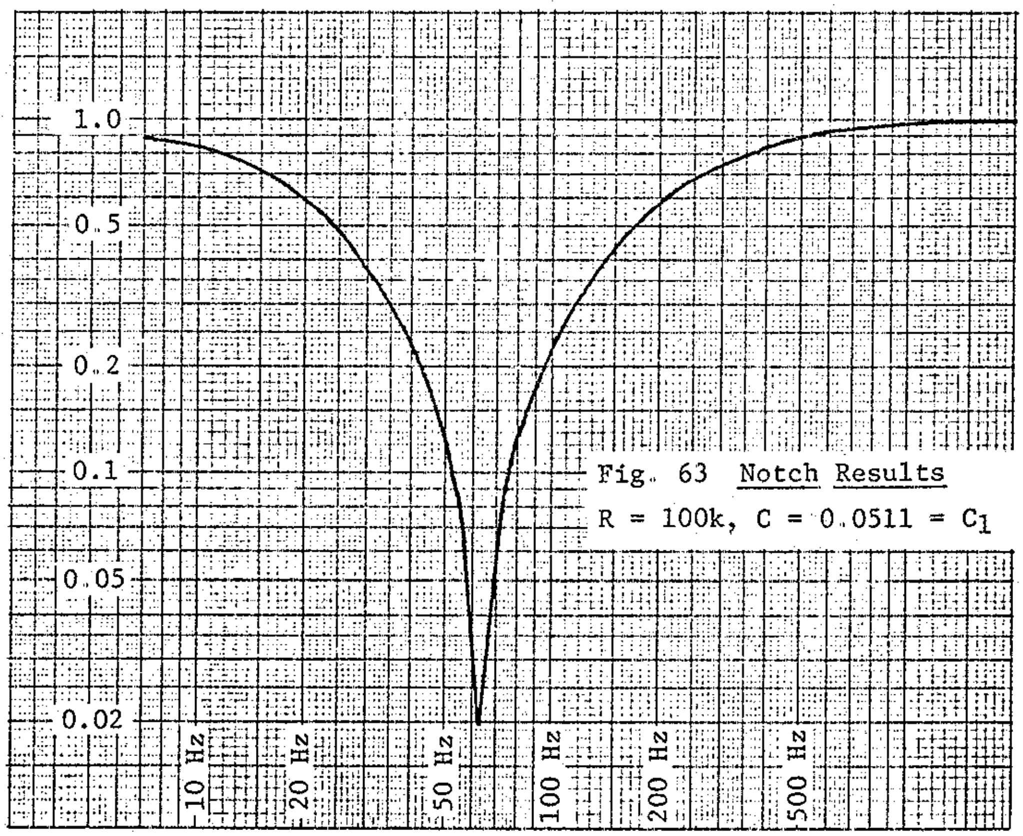


Fig. 63 Notch Results
R = 100k, C = 0.0511 = C_1

CHAPTER 7: STATE-VARIABLE AND VOLTAGE-CONTROLLED FILTERS

7A. INTRODUCTION

In this chapter, we will be covering two important subjects - the state-variable filter, and the idea of voltage-control of filters. In the course of the chapter, it should become clear why these go well together. We will begin with the basic theory of the state-variable filter, and give some actual state-variable active filter realizations. We will then discuss the transconductor as a filter control element (as a voltage-controlled resistor). As a simple voltage-controlled filter, the first-order controlled structure will be examined, and then we will move on to the voltage-controlled state-variable. In the third part of experiment 5 below, the state-variable will be refined so that in its control mode the Q remains constant dispite phase shifts across the control elements.

7B. BASIC THEORY OF THE STATE-VARIABLE FILTER

The name state-variable (s-v) comes from the fact that the filter is derived from a flow-graph between the state-variables of a simple passive network. The exact details of this derivation shall not concern us here. Instead we will consider the network as given, analyze it, and learn to control its parameters. The basic s-v filter consists of two integrators and a summer in a loop as shown in Fig. 64. The network is second-order (biquadratic) and is sometimes called a "Biquad" although this latter term is a more general term for a type of network of which the s-v is an example. The s-v is properly called biquadratic because it can be used to realize a numerator of the form $As^2 + Bs + C$, where A, B, and C are constants, which may be zero, so the s-v is capable of producing low-pass, high-pass, and bandpass functions, certain special combinations (such as notch), and in fact, any general biquadratic. It is also sometimes useful that the low-pass, bandpass, and high-pass outputs are available simultaneously.

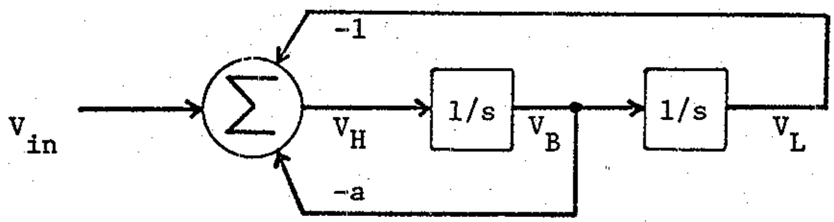


Fig. 64

Basic State-Variable Filter

For the moment, we will just use 1/s for the integrator, and thus we can see that the output of the summer is:

$$V_H = V_{in} - aV_B - V_L = V_{in} - aV_H/s - V_H/s^2 \tag{7B-1}$$

and this is easily solved for $T_H(s)$ as:

$$T_H(s) = V_H/V_{in} = \frac{s^2}{s^2 + as + 1} \tag{7B-2}$$

from which it is easily seen that:

$$T_B(s) = V_B/V_{in} = \frac{s}{s^2 + as + 1} \tag{7B-3}$$

and:

$$T_L(s) = V_L/V_{in} = \frac{1}{s^2 + as + 1} \tag{7B-4}$$

We thus have a basis for a filter that relies only on our ability to construct integrators and summers. We have shown also that the damping term is D = a for the s-v where a is just a feedback gain.

A popular configuration for the s-v filter formed from op-amps is shown in Fig. 65. In the analysis, it should be remembered that the integrators are inverting and have transfer functions $-1/sCR$. The first step in the analysis is to observe that the voltage on the inputs of A1 are determined by resistor voltage dividers:

$$V_- = \frac{V_L + V_H}{2} \qquad (7B\text{-}5)$$

$$V_+ = V_{in} - \frac{R'}{R'+R_Q}(V_{in} - V_B) \qquad (7B\text{-}6)$$

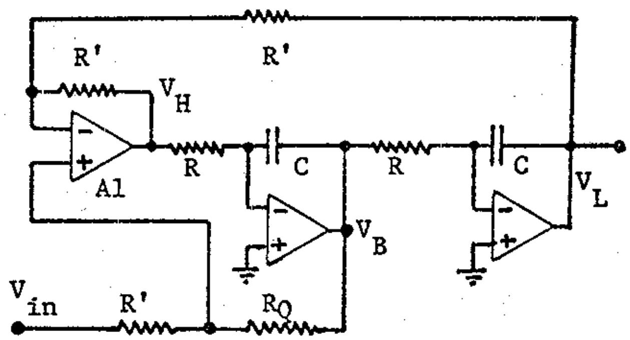


Fig. 65 One Form of State-Variable

As usual, we set $V_- = V_+$, and representing $R'/(R'+R_Q)$ by κ, and using the integrator result, we get:

$$V_{in}(1-\kappa) = \frac{V_H\kappa}{sCR} + \frac{V_H}{2s^2C^2R^2} + \frac{V_H}{2} \qquad (7B\text{-}7)$$

From this we obtain the transfer functions analogous to (7B-2), (7B-3), and (7B-4):

$$T_H(s) = V_H/V_{in} = (1-\kappa)\left[\frac{2s^2}{s^2 + (2\kappa/RC)s + 1/R^2C^2}\right] \qquad (7B\text{-}8)$$

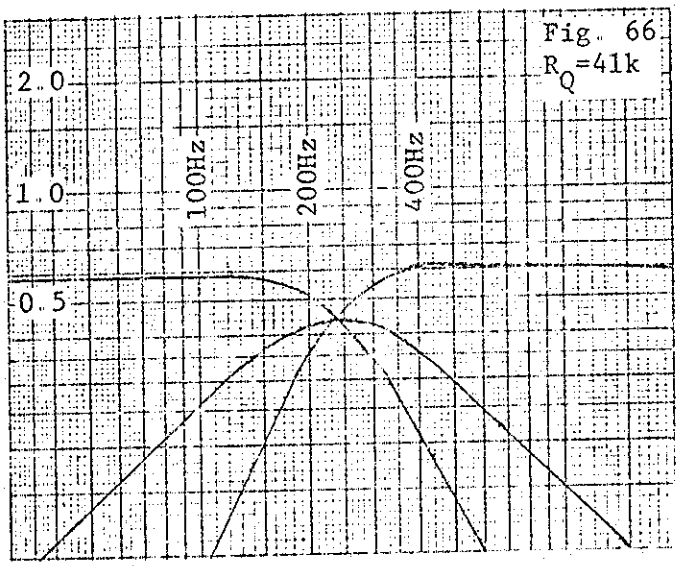
$$T_B(s) = V_B/V_{in} = (1-\kappa)\left[\frac{-2s/RC}{s^2 + (2\kappa/RC)s + 1/R^2C^2}\right] \qquad (7B\text{-}9)$$

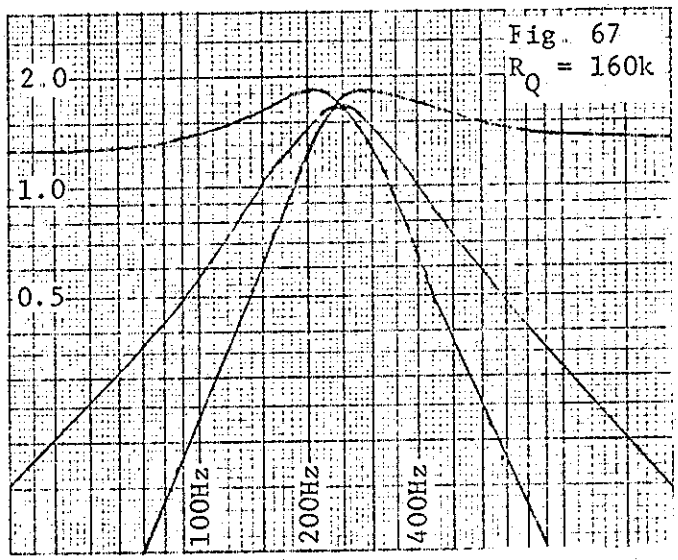
$$T_L(s) = V_L/V_{in} = (1-\kappa)\left[\frac{2/R^2C^2}{s^2 + (2\kappa/RC)s + 1/R^2C^2}\right] \qquad (7B\text{-}10)$$

From these equations, we note an overall gain of $2(1-\kappa)$, and a damping term 2κ, corresponding to a Q of $1/2\kappa$. We could have arrived at the expression for Q by means of the feedback gain using the general model of Fig. 64. To do this we would have to determine the value of the feedback path "a" in the general model. Clearly, the voltage V_B is fed back by voltage divider R_Q-R', and the proportion that reaches the (+) input of A1 is just κ. However, you should realize that the resistor R' from V_L back to the (-) input of A1 acts to make A1 a non-inverting amplifier of gain 2, just as though V_L were grounded. Thus, the feedback factor from V_B to V_H is $2\cdot\kappa$, a fact that should be carefully studied as a means of preparing for the analysis of the slightly different configuration given in the experiment that follows later. Thus, since the feedback gain $a = 2\kappa$, $Q = 1/2\kappa$, just as we get from analysis of T(s). The section design frequency of the state-variable is just $1/2\pi RC$. Thus note that control of the RC product of the integrators controls the frequency characteristics of this second-order section while R_Q determines the response characteristic independently. Thus, unlike some second-order structures (but like Sallen-Key), we have independent control over frequency and damping. It turns out, although we will not show it here, that one of the advantages of the s-v approach over Sallen-Key will be that the s-v has a lower sensitivity to component variations.

Knowing the relationship between damping and the resistor R_Q relative to R' allows us to set a response characteristic, which we might choose as Butterworth. Consulting Table 1 of Chapter 5 tells us that the damping should be 1.41. Thus $2\kappa = 1.41$ or $\kappa = 0.707$. Thus, choosing R' = 100k, we get $R'/(R'+R_Q) = 0.707$ or $R_Q = 41k$. We can run an experimental test choosing R = 12.48k and C = 0.0511 mfd (just values on hand chosen at random). The results are shown in Fig. 66. We note from Fig. 66 the Butterworth high-pass and low-pass with second-order 12db/octave slopes, as well as the bandpass with 6db/octave "skirts" on each side. The experimental Q (equation 3F-5) is close to the 0.707 predicted value. The overall gain is close to the

$2\cdot(1-\kappa) = 0.58$ predicted, as can be seen from the low frequency gain of the low-pass and the high-frequency gain of the high-pass. To check the peak gain of the bandpass it is a useful trick to realize that if we substitute $j\omega_0$ for s, where ω_0 is the center frequency = 1/RC, the s^2 term in the denominator will cancel the $1/R^2C^2$ term when it comes to taking the magnitude. Applying this to equation (7B-9) we see that we would have $2(1-\kappa)/2\kappa$ remaining, so the bandpass peak amplitude would be $(1-\kappa)/\kappa = 0.414$, close to the observed value. Finally, the center frequency is $1/2\pi RC$ = 248 Hz, close to the observed values. Note that all three curves pass through the same point, and that the 3db frequencies are the same for low-pass and high-pass. This is true only for this Butterworth example, as will be clear from the second example shown in Fig. 67 where R_Q = 160k (close to 3db Chebyshev). The reader may verify the other calculations on this second example against the experimental results.





7C. THE TRANSCONDUCTOR AS A CONTROL ELEMENT

In all the filters we have studied so far, we have seen that the frequency of operation and other parameters of the filter depended on the value of certain resistors (and on capacitors). Since variable capacitors of more than a few thousand pf are rare or unavailable, we resort to variable resistors when it is necessary to tune filters. Variable resistors, potentiometers, are easy to apply, but must be manually tuned. Here we will look at a control element that responds to voltage. This allows us to tune filters without manual control, and this can often be done automatically, at a faster rate, and with more accuracy than can be done with manual control. Such filters find numerous applications including the synthesis of music and speech, communications, adaptive processing, and adaptive control.

The development of a voltage-controlled filter (VCF) is a two stage process that generally starts with a control device, and then proceeds to the search for a network where this control device will work properly and naturally. Here we will choose as the control device the transconductor, of which the RCA type CA3080 is the best known example, and will see which networks can use this device. First, we will look at the transconductor itself.

Fig. 68 shows the CA3080 transconductor IC. It is in some ways much like an ordinary op-amp, except it has (1) a gain controlled by I_{ABC} (amplifier bias current), (2) a current output instead of a voltage output, and (3) an input that generally needs to be limited by an attenuator. In general, the control current

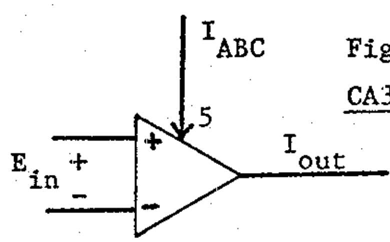
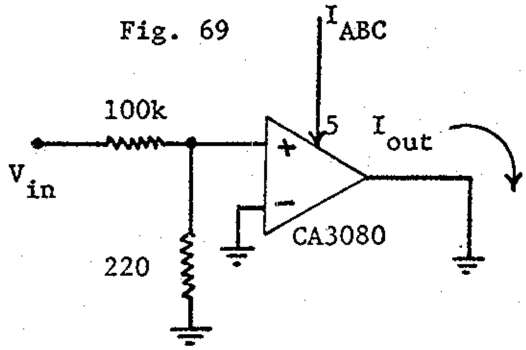


Fig. 68
CA3080

I_{ABC} should not exceed 1 to 2 ma. The differential input voltage should not exceed about 1 - 2 volts in all cases to prevent damage to the device, and here we have a further restriction that the input should not exceed ±20 mV, this latter restriction being added so that the device remains sufficiently linear. The basic equation for the CA3080 is:

$$I_{out} = 19.2 \cdot I_{ABC} \cdot E_{in} \qquad (7C\text{-}1)$$

Fig. 69



To assure that the input voltage does not exceed the ±20 mV we specified (with ±10 volt signals in the filter) or better still ±10 mV (with ±5 volts signals in the filter), the 100k to 220Ω attenuator shown in Fig. 69 is used. This gives: $E_{in} = 0.00220\ V_{in}$. Substituting this back into equation (7C-1), and solving for an equivalent resistance $R_{eq} = V_{in}/I_{out}$, we get:

$$R_{eq} = \frac{23.7}{I_{ABC}} \qquad (7C\text{-}2)$$

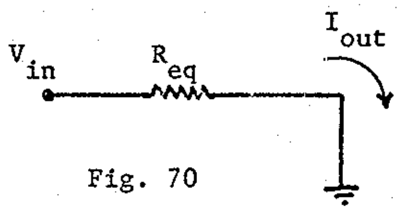


Fig. 70

To see the significance of this equivalent resistance, consider Fig. 70 where we show an ordinary resistor driven from the same voltage driving the same current into ground. It is not always necessary that the CA3080 drive current into ground potential, but here it is seen to work as an equivalent resistor driving into ground.

At this point, we have described a current-controlled resistance. It actually makes little difference what we use to control the controlling element - voltage or current - but voltage is often easier to design with and measure. For the purposes of this discussion and experiment, we need only know that the control terminal, pin 5, is at a voltage that is approximately 0.7 volts above the negative supply. On a ±15 power supply, pin 5 is at about -14.3 volts. Thus we can supply the control current by connecting a resistor between pin 5 and some other voltage more positive than -14.3. We will generally be supplying this current from the wiper of a pot connected between +15 and -15, and the resistor to pin 5 will be 33k, which safely limits the current even at control voltages of +15. Keep in mind that even though we are supplying the control voltage from a manual pot, we have the potential of supplying the voltage from any source, so we have a system that goes beyond manual tuning.

7D. A FIRST-ORDER VOLTAGE-CONTROLLED FILTER

As an example, we will consider the process of adding voltage-control to a first-order low-pass filter. We can use as a prototype either of the two circuits shown in Fig. 71, but for this discussion, we will choose the B circuit. It turns out that both circuits would give essentially the same voltage-controlled version, even though circuit 71B has one extra resistor. The voltage-controlled version of the B circuit has the slight advantage that the CA3080 is always driving into ground potential.

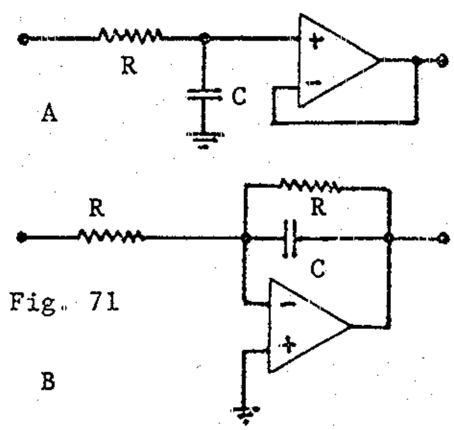


Fig. 71

Note first that in the A circuit, the voltage across the resistor is not a constant, but is the difference between the voltage on the input and the voltage on the capacitor. In the B version, the effective voltage across the resistor is decreased by a cancelling current fed back through the second resistor from the output. If we drive a standard inverting integrator with the voltage-controlled resistor of Fig. 69, we have a voltage-controlled integrator, a simple fact that we will exploit in the voltage-controlled version of the state-variable filter. But here, we must add provisions so that the decrease in the voltage across the filter resistor is accounted for.

The solution to the problem is seen in Fig. 72. We simply loop a portion of the output voltage back to the input of the CA3080, keeping in mind that the integrator is inverting. Since the 220 ohm resistor is so much smaller than the 100k resistors that bring in the input signal, and return the feedback, we can consider voltages to be summed (as well as attenuated) at the (+) input of the CA3080. In Fig. 73, the circuit of Fig. 72 is redrawn so as to look as much as possible like the circuit of Fig. 71B. The reader may want to show that the circuit of Fig. 74 is the proper voltage-controlled version of Fig. 71A, and note that here the CA3080 drives into a voltage equal to the output voltage of the filter, which it is capable of doing, as long as the power supply voltages to the CA3080 exceed the signal voltages of the filter. In all these filters, the cutoff frequency is given by $1/2\pi R_{eq}C$, where R_{eq} is given by equation (7C-2) with I_{ABC} approximately $(V_c + 14.3)/33k$.

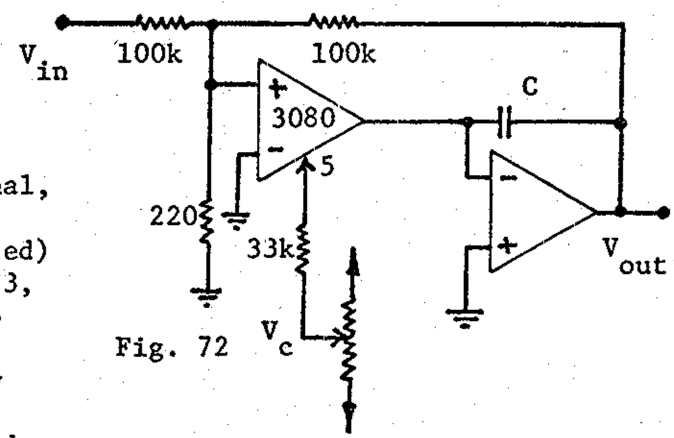


Fig. 72

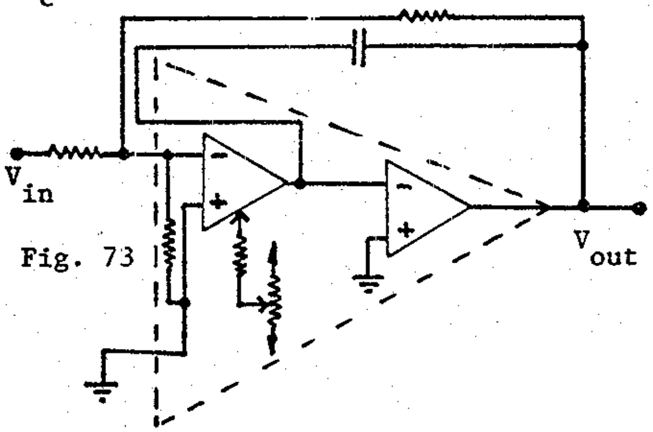


Fig. 73

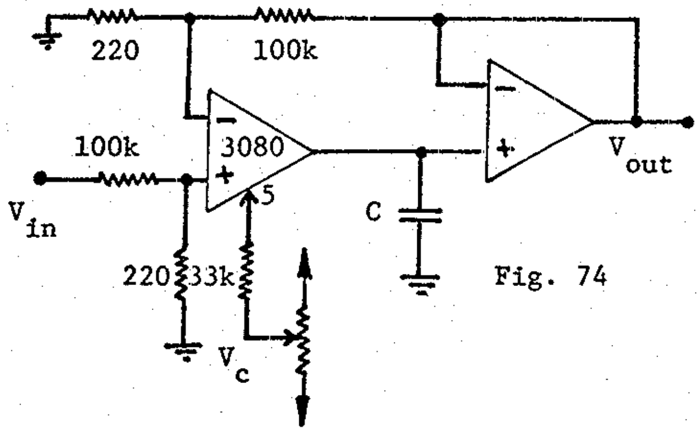


Fig. 74

7E. VOLTAGE-CONTROLLED STATE-VARIABLE FILTER

It is a simple matter to use the voltage-controlled resistor of Fig. 69 to drive into the ground potential (virtual ground) of the inverting integrator. Another advantage of this setup is that by using the (-) input of the CA3080 used in the voltage-controlled resistor, the overall integrator becomes positive, and the summer of the state-variable filter can be a simple inverting summer. The complete circuit is shown in Fig. 75. Note that both of the CA3080's in the two integrators can be controlled in parallel as shown. Another simple extension that we will not consider here is that a third CA3080 can be used to control the Q of the filter by controlling the feedback from the bandpass output, essentially using another voltage-controlled resistor (CA3080) to replace resistor R_q. Based on previous results, the reader should be able to show that the center frequency is:

$$F_c = 1/2\pi R_{eq}C = \frac{(V_c - 14.3)}{2\pi C\cdot 23.7\cdot 33k} = \frac{(V_c - 14.3)}{(4.9\times 10^{+6})C} \qquad (7E\text{-}1)$$

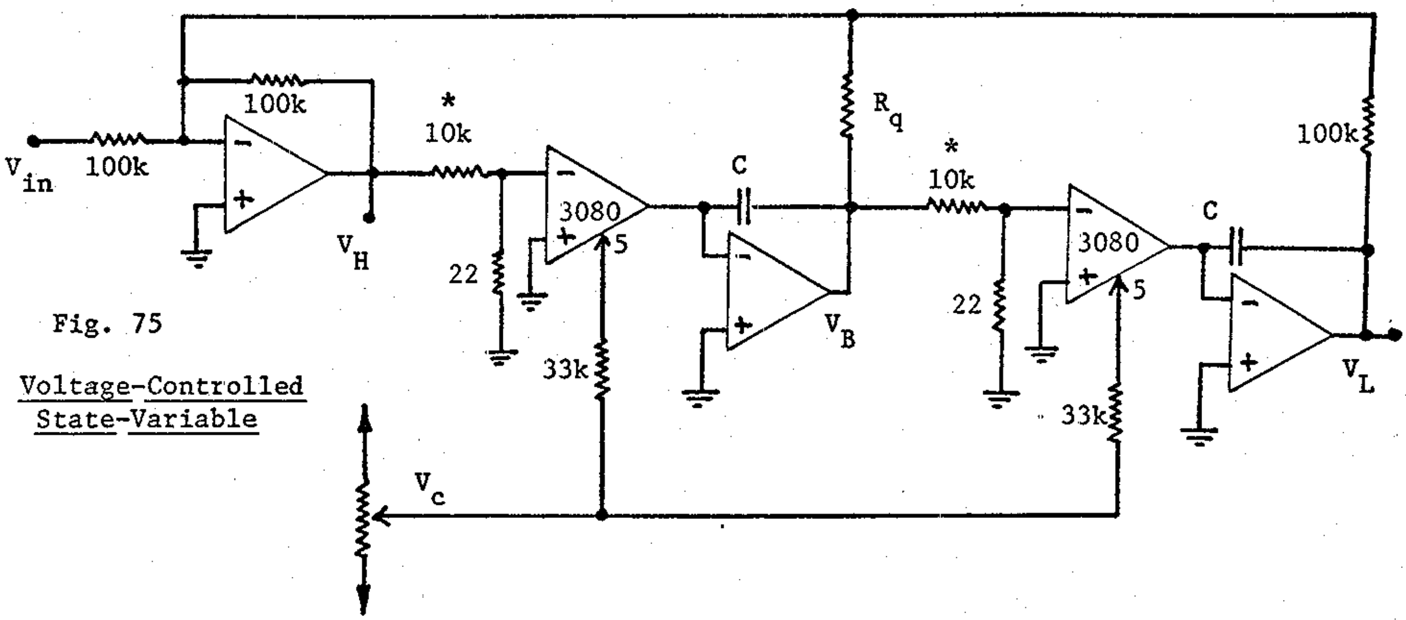


Fig. 75

Voltage-Controlled State-Variable

and that the Q is given by: $Q = R_q/100k$ (7E-2)

7F. EXPERIMENT NO. 5, VOLTAGE-CONTROLLED AND STATE-VARIABLE FILTERS

1. Build and test the state-variable filter shown in Fig. 76. This configuration is popular and about as common as the one in Fig. 65. Analyze the circuit to determine the relationship between filter Q and the resistor R_Q. Verify this relationship using equation (3F-5) to determine experimental Q's.

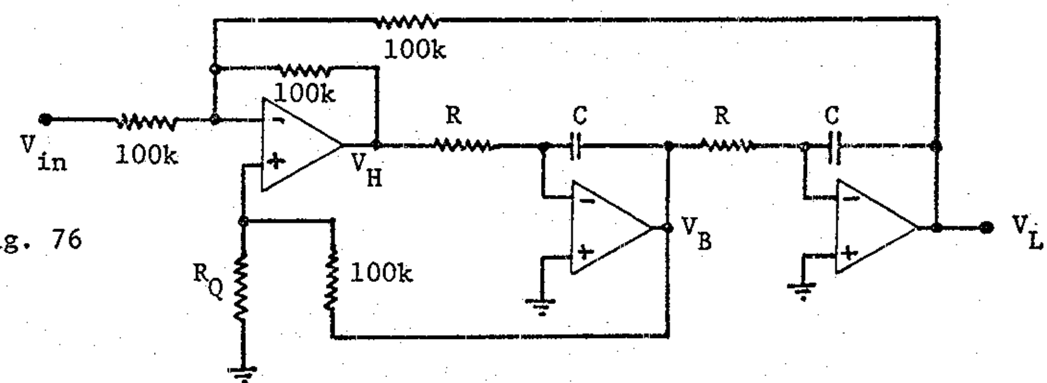


Fig. 76

Calculate and verify experimentally the overall gain factor for the configuration of Fig. 76. Compare the variations of overall gain and Q with the configuration of Fig. 65. Sum the high-pass and low-pass outputs using a simple inverting summer, and verify that a notch response results from this summation. Discuss the importance of relative phase in this summation.

2. Build and test the voltage-controlled single-pole low-pass of Fig. 72. Verify the relationship between cutoff frequency and control voltage. Alter the structure by considering the (+) input of the regular op-amp to be a second input as shown in Fig. 77. By analysis and by experiment determine what happens when the original input is grounded and the (+) input of the op-amp receives the input signal. Then see what happens when the input signal is applied to both the original input and to the (+) input of the op-amp.

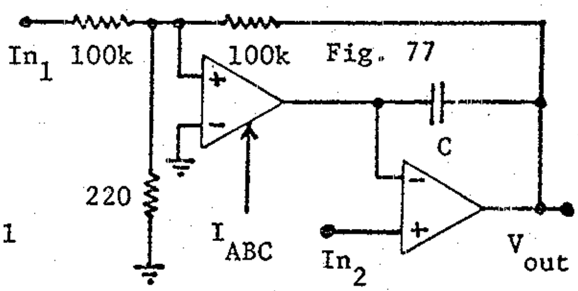


Fig. 77

3. Build and test the state-variable VCF of Fig. 75. A good starting value for C is 0.001 mfd. Note that here we have used attenuators on the CA3080 inputs that are formed from 10k-22ohm resistors instead of the previous 100k-220ohm setup. This makes no difference to the analysis to this point, as the attenuation is the same as before. Later we will want to be doing some special compensation where it will be convenient to have a smaller resistor in the (*) positions so that a needed RC constant can be obtained without having to locate a very small capacitor.

   By experimental measurements on the bandpass output, verify equations (7E-1) and (7E-2). Calculate the value of the overall gain of the network and verify by measuring the low-frequency gain of the low-pass output.

   Learn to use the "ringing" method of measuring Q [Fig. 62 and equation (6E-1)] and choose R_q so that the Q is about 10. Measure the Q at low-frequency (say 200 Hz) and at high frequency (say 10 kHz) and compare. You will probably find that the Q has increased at the upper frequency. This is due to phase shift across the control elements, and is called "Q-Enhancement." Plot a curve of Q as a function of frequency. Now, to compensate for this it is useful to add a "phase lead" section which is achieved simply by placing a small capacitor across the two 10k resistors marked with a (*). The capacitor should be something like 10 pf to start with. Remeasure the curve of Q as a function of frequency, and see if it is now flatter. Adjust the value of these capacitors up or down as needed to achieve a flat curve. The exact value will depend on the op-amps used, as these also contribute some phase shift to the loop. Discuss why this type of compensation is relatively important in VCF's or other tunable filters, and why it might not be so important in fixed filters.

TYPICAL RESULTS: Results of part 1 will be very similar to those shown in Fig. 66 and Fig. 67. Part 2 will result in a low-pass response similar to Fig. 27, and with the alterations described, a corresponding high-pass and all-pass response will appear. In part 3, the major new thing is the compensation for Q-Enhancement. Typical Q vs. freq. curves are shown in Fig. 78 below. The CA3140 was used for the op-amp for these. Slower op-amps such as the 307 may require up to 30 pf or more to flatten (and may be a better lab experiment for this reason).

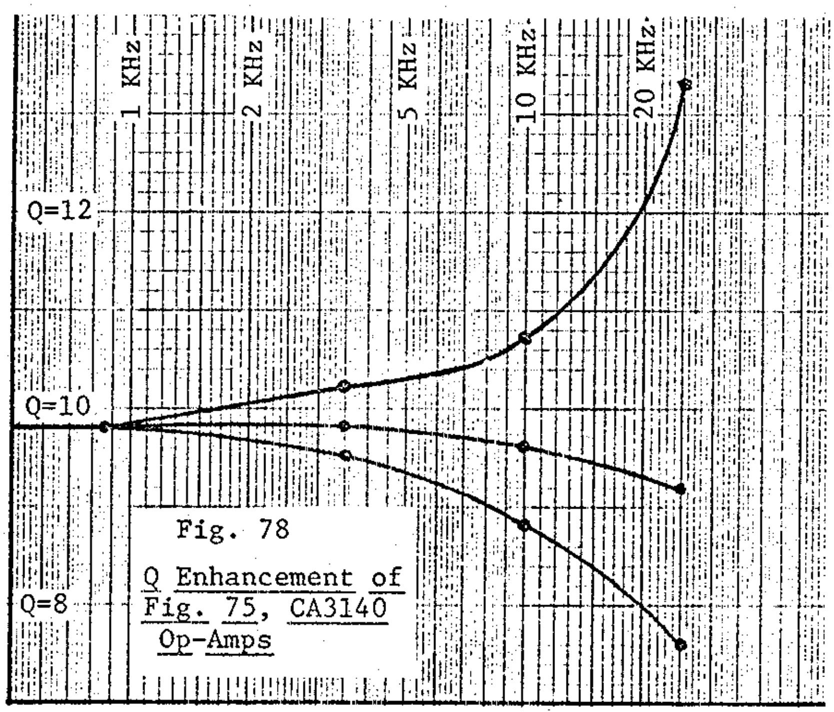


Fig. 78

Q Enhancement of Fig. 75, CA3140 Op-Amps

CHAPTER 8: DISCRETE-TIME FILTERING WITH DELAY LINE NETWORKS

8A. INTRODUCTION

In this chapter, we depart from the general flow that we have been following, and will jump into the domain of discrete time - the world of the so-called "digital filter" and its relatives. It will not be possible here to go into a lot of detail. Instead, we will give only enough background so that the reader will be able to perform the experiments. There is probably no need to discuss the emerging importance of digital filters. They must be studied, and it is probably the case that the best way to learn about digital filters is to work with them. As of this writing there are several good ways to work with digital filters, and actually building one as a lab exercise is not one of these good ways. Among the good ways are: computer simulation, manipulation of special purpose digital signal processors, use of a programmable calculator, and experimenting with delay line networks.

The delay lines we have in mind are analog delay lines of the "Bucket brigade" or "charge-coupled device" type. These are basically integrated circuits that are designed to represent an analog voltage by a proportional amount of charge, pass this charge sample down a line, and reconstruct the analog voltage on the other end. This process involves time sampling, so the reconstructed waveform will be of a stepped approximation nature, and the limitations of sampled data systems apply here. However, we will assume, and it can be made the case, that the sampling rate on the actual IC delay line is sufficiently fast, and that the proper input guard and output smoothening filters are present so that we may regard the device as an ideal fixed delay. We are then free to design with a delay element which we can call τ.

Before starting in, we want to point out that digital filters are not quite as unfamiliar as many readers first think. It would seem likely that nearly everyone has at one time or another used a digital filter, probably without knowing that what he was doing was a form of digital filtering. For example, you may have worked in a store and made a chart of sales. If this was done on a daily basis, you may have found the chart to have too many irregular variations to be useful for detecting the trends you are interested in. So in a very natural way, you average data for several days and replot, and arrive at a smoother sales curve. If you think about it, this is a type of low-pass filtering, one which removes high frequency random events and returns only low frequency trends. Some information is intentionally discarded so that the remaining information is clearer. This is a type of digital filtering. It is digital because the data is digital (discrete amplitudes) since they are in amounts of dollars and cents - equally spaced numbers. Note that it is also a discrete time process because of the daily sampling. The analog delay lines we will be using will be discrete time (time exists at the input, and at a delayed point or at delayed points in the case of multiple delays), but not discrete amplitude since the charge packets that store the samples are proportional to the input voltage, and are not rounded off to the closest available discrete value. This is an improvement over a fully digital system, at least in theory. In practice, amplitude samples capable of taking on any of a continuous set of values are subject by their very nature to errors because the system receiving them knows nothing of them except what is found when they arrive. A digital receiving system on the other hand knows that only a finite set of values may be coming it. Thus, even with small variations that must exist, the receiving system is able to regenerate the original value, and only very gross errors will result in a sample jumping into a different bin so to speak.

8B. SOME THEORY OF DISCRETE-TIME SIGNAL PROCESSING

It is natural when starting in on a new type of analysis to try to hang on to something familiar. We all have a good feeling for time delay, and thus would be inclined to work in the time domain first. Yet, we should also keep in the back of our minds that in the active filter case, things got really easy only when we went over to the frequency domain and started working with the s variable. In the case of discrete time filtering, we will have to resort to a complex z variable eventually, and many of the things we do will follow closely the procedures that went on with the s variable. First however, we will do one example where time domain analysis works rather well.

Fig. 79 shows a simple setup where a sine wave is summed (inverted) with a delayed version of the input. We will apply time domain analysis here. We see that the output is the sum:

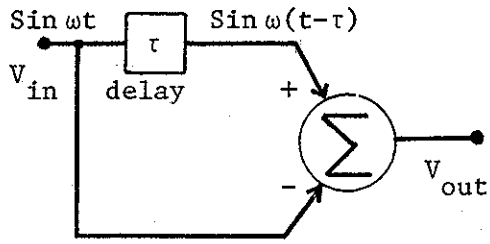


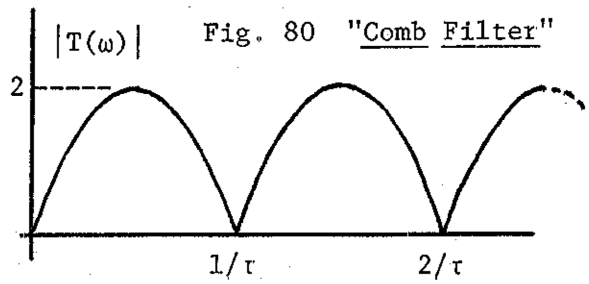
Fig. 79

$$V_{out} = \mathrm{Sin}(\omega t - \omega\tau) - \mathrm{Sin}(\omega t)$$

$$= -2\,\mathrm{Sin}(\omega\tau/2)\ \mathrm{Cos}[\omega t - (\omega\tau)/2] \qquad (8B\text{-}1)$$

were the second line is obtained simply by applying the trig formula for the sum of two sines. The first term in the result, $\mathrm{Sin}(\omega\tau/2)$ is independent of time t, but depends on the delay time τ. The second term is a sinusoidal that is identical to the input except for the phase shifts [changing to Cosine, and the $(\omega\tau/2)$]. The magnitude of the frequency response is thus as indicated in Fig. 80. This is easily understood in terms of the delayed signal cancelling and then reinforcing the input as the frequency changes, changing the relative phase at the summer. The result is interesting in that we have not yet studied any structures with multiple notches (or any other multiple feature), and this has the obvious application of removing harmonically related signals since the notches can be tuned by adjusting the delay time.

Fig. 80 "Comb Filter"

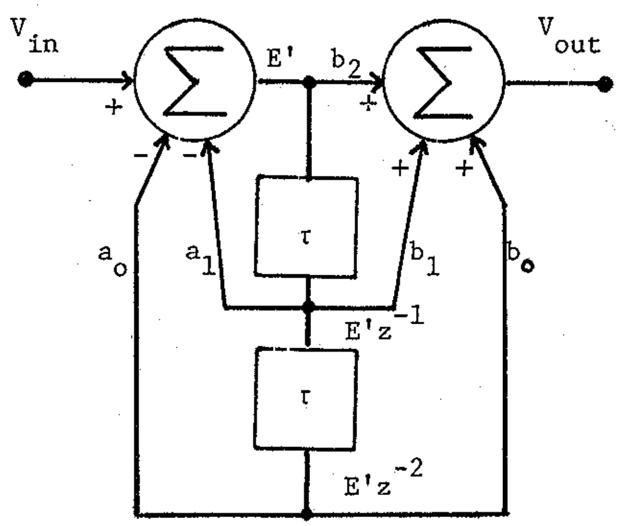


zero when $\omega\tau/2 = n\pi$,
thus when $f = n/\tau$

It becomes obvious that when we consider a more complex delay line network such as the "second-order" (two delay lines) structure in Fig. 81, that time domain analysis is going to get very tedious. Not only is there the feed-forward summer as in the above "comb filter", but we also have feedback paths (making the structure "recursive"). Based on our success with frequency domain analysis with active filters, we might want to jump right to the s-domain. This could be done, but a somewhat different solution leads to results that are more direct, and we will come to use a "z" variable here. The first thing to observe is that in the frequency domain, we are really performing operations on quantities that are Laplace transforms of their time domain counterparts. We get away with being a little "loose" with our thinking because we are always thinking in terms of response to sinusoidals, and it is the sinusoidals that are the "common denominator" between the time and frequency worlds. In any case, we want to know how a delay effects a Laplace transformed quantity.

Fig. 81

Second-Order Delay Line Network



τ = delay $\quad z^{-1} = e^{-s\tau}$

It is a standard exercise to show that the Laplace transform of a delayed signal is just $e^{-s\tau}$ times the Laplace transform of the undelayed signal. It is standard practice to represent $e^{-s\tau}$ by the notation z^{-1}, which is, as it will turn out, more than just a change of notation. We now are in a position to write a transfer function for the second-order structure of Fig. 81. Now, just as we have learned to treat capacitors as "resistors" of value (1/sC), we will treat delays as devices which simply multiply a signal by z^{-1} (in the Laplace transformed domain, we remind you once again). Thus, the signal at the output of the first summer can be represented as E', and thus the first delay changes this to $E'z^{-1}$ (that is, E'/z) and the second delay multiplies by an additional z^{-1}, giving $E'z^{-2}$ (that is, E'/z^2). We then can just write:

$$E' = V_{in}(z) - a_1E'z^{-1} - a_oE'z^{-2} \qquad (8B\text{-}2)$$

and that:

$$V_{out}(z) = b_2E' + b_1E'z^{-1} + b_oE'z^{-2} \qquad (8B\text{-}3)$$

We can solve for E' in equation (8B-2) and plug into (8B-3), arriving at a transfer function which we will denote H(z) [you may use T(z) if you prefer].

$$H(z) = V_{out}(z)/V_{in}(z) = \frac{b_2z^2 + b_1z + b_o}{z^2 + a_1z + a_o} \qquad (8B\text{-}4)$$

Equation (8B-4) is something we can easily work with in terms of poles and zeros and related techniques. Notice that it would have been awkward to have carried $e^{-s\tau}$ instead of z^{-1}. Yet, when it comes to finding the response of a discrete time filter to ordinary sinusoidals, we will end up substituting $e^{-j\omega\tau}$ into equation (8B-4) or into an expression derived from it.

Since we are now using a variable z, which is in general a complex number, we have to work in the z-plane. The form of the z-plane is mathematically identical to the s-plane. A calculation that is valid in the s-plane (such as the distance between two points) is also valid in the z-plane. What will be different are the physical consequences of the mathematics. For example, we know that in the s-plane, poles could not appear in the right half-plane if the system was to be stable. If you transform a few points from the s-plane to the z-plane using $z = e^{s\tau}$, you will discover that the left half-plane transforms or maps entirely into the interior of the unit circle of the z-plane, and that the right half-plane maps into the exterior of the unit circle. Thus, stability in the z-plane is a matter of keeping poles inside the unit circle. Perhaps of the most interest is the fact that the $j\omega$-axis of the s-plane maps into the unit circle, but in a curious way. It wraps itself around and around the unit circle repeating ever $1/\tau$ units of frequency, with zero frequency at z = +1. Yet the point z = +1 also represents frequencies $1/\tau$, $2/\tau$, $3/\tau$, as well as $-1/\tau$, -2τ, $-3/\tau$, and likewise, any other point on the unit circle in the z-plane represents an infinite number of equally spaced frequencies.

With this in mind, consider the evaluation of a frequency response using a z-plane pole-zero model. We start our evaluation at zero frequency (at z = +1), and start counterclockwise around the circle (see Fig. 82). Just as we did in the s-plane, we evaluate the magnitude of the response by multiplying the distances to the zeros and dividing by the distances to the poles. The frequency going round the circle is just $(D/360)\cdot(1/\tau)$, where D is the angle with respect to the positive real axis. The frequency $(1/\tau)$ is often called the "sampling frequency" f_s.

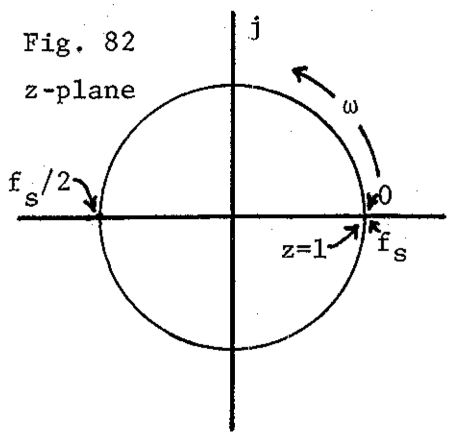


Fig. 82
z-plane

When the angle around the circle reaches 180°, the response function will start to back up (because the poles and zeros here must be complex conjugates). Eventually, we get back to the starting point at 360°. Going beyond 360°, we get the same values all over again. We have already seen an example of this by the response shown in Fig. 80. We now see that the "comb filter" response is something that comes in a very natural way, and in fact, we have to take steps to get rid of it. The way to get rid of the repeating response is to obey the "sampling theorem." The sampling theorem tells us that we should not input to the system any frequency components that are higher than $f_s/2$. This would be done by applying a "guard filter", a low-pass cutting off sharply by the time it reaches $f_s/2$. [Note that this has probably been done in analog delay lines, but working very much higher at the clocking frequency of the delay line]. This clocking frequency should not be confused with the sampling frequency we are talking about here. In fact, it may be best to forget about the actual clocking frequency on the delay line, and just consider the delay as a continuous analog delay.] In the comb filter, we intentionally violate the sampling theorem so that we do get a repeating response.

8C. EXPERIMENT NO. 6, DELAY LINE FILTERS

1. Set up and test the delay lines available to you. It is important that the frequency response, overall gain, upper frequency limitations, etc. be understood before attempting to apply the line to these experiments. If the gain of the line is not unity, this is not important, as long as it is allowed for when setting network coefficients. Run the input frequency up until either the response drops off, until steps appear in the output waveform, or until the output frequency goes down as the input frequency increases (frequency aliasing, or violation of the sampling frequency on the delay line itself). Whichever of these limitations occurs first depends on the exact setup of the delay line IC you are using. When you have determined the upper frequency limit, plan your experiments so that you work at frequencies that are about ten times lower than this limit.

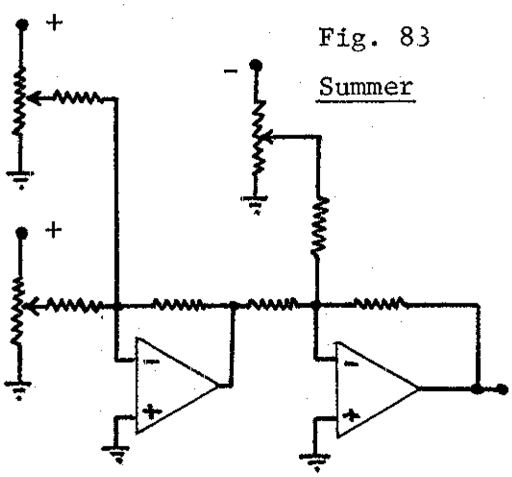


Fig. 83

Summer

2. Set up and test the comb filter of Fig. 79. A suggested summing network is shown in Fig. 83. Find the spacing between the nulls and show that this is $1/\tau$, where τ is the delay time as obtained by direct measurement with an accurate scope, or by a knowledge of the frequency of the delay line clock and its operation. If necessary, just use the spacing between the nulls of this comb filter as your measured delay time. Change the experiment by setting up the summer with both inputs positive, measure the frequency response with this setup, and compare with the setup of Fig. 79. Do a z-plane analysis of both these structures [a special case of Fig. 81 and equation (8B-4)] and understand the network response in terms of the zeros which you locate by this means. Suppose the system is a digital filter. How would you place a "guard filter" so that the comb filter with both summer inputs positive is a low-pass filter. [Ignore the fact that this is not a practical filter in this case.]

3. Using two delay lines that have identical delay, realize the system of Fig. 81. The summer stage suggested in Fig. 83 should be modified to provide any additional inputs necessary, and a second summer should be built. By factoring equation (8B-4) using the quadratic formula, show that the poles of (8B-4) are at a radius $\sqrt{a_o}$ and at angles (see Fig. 84) ϕ_1 and ϕ_2 given by:

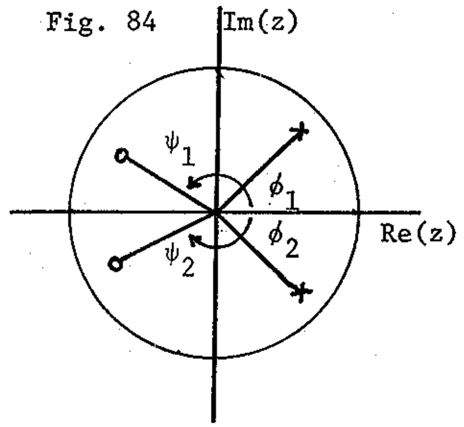
$$\phi_1,\ \phi_2 = \pm \operatorname{Cos}^{-1} \frac{-a_1}{2\sqrt{a_o}}$$

and that the zeros are at a radius $\sqrt{b_o/b_2}$ and at angles ψ_1 and ψ_2 given by:

$$\psi_1, \psi_2 = \pm \mathrm{Cos}^{-1} \frac{-b_1/b_2}{2\sqrt{b_o/b_2}}$$

By using these formulas, you can locate poles in the z-plane by adjusting the summing coefficients. Note that the requirement that the system be stable will restrict the poles to the interior of the unit circle, and hence a_o must be less than one.

Fig. 84



As a starting point, set the poles at approximately the locations shown in Fig. 84, and place two zeros at z = -1. Measure the response of the filter, and verify the response curve by using a graphical method. Remove the zeros (by setting b_1 and b_o = 0) and compare the stopband rejection with the case where these zeros are present.

Often in the design of discrete-time filters, data borrowed from the world of active filters is used with the idea that one or more of the properties will be transformed and maintained in the discrete-time design. Here we will give an example using the "Matched z-Transform". The method simply transforms the poles of a filter in the s-plane to the z-plane through the mapping $z = e^{s\tau}$, the same mapping we used to set up the z-plane in the first place. We will see that this can be made to work under certain conditions. First, we can use Table 1 of Chapter 5 to see that the damping for a 3db ripple Chebyshev filter is 0.77, and thus the denominator in the s-plane is $s^2 + 0.77s + 1$ (we are not concerned here with the exact placement of cutoff frequencies). This gives poles at $-0.385 \pm 0.923j$. When we transform these, we have to choose a value for τ relative to the s-plane dimensions. Let's start with $\tau = 1$. With this done, the poles transform to:

$$z = e^{-0.385 \pm 0.923j} = 0.68[\mathrm{Cos}(0.923) \pm j\mathrm{Sin}(0.923)] = 0.410 \pm 0.542\,j$$

Set up the experimental filter to give these poles. You should observe a 3db ripple, but relatively poor stopband rejection, and of course the response will repeat at higher frequencies. If you add in the zeros at z = -1, you will get a much better stopband rejection, but the ripple in the passband will be altered. Another approach to improving the situation is to change the sampling frequency to a higher value. We can make this four times higher if we set $\tau = 0.25$ instead of one. The poles in this case are:

$$z = e^{(-0.385 \pm 0.923j)(0.25)} = 0.884 \pm 0.208\,j$$

Set the filter for these poles, remove any zeros, and remeasure the response. You should find the 3db ripple, and a much improved stopband. Plot and compare the active filter 3db Chebyshev, the discrete time filter with $\tau = 1$, and for $\tau = 0.25$.

If you are familiar with other digital filter design methods (such as the Bilinear z-Transform) transform these s-plane filters to the z-plane and examine them experimentally.

<u>TYPICAL RESULTS</u>: Typical results for the comb filters are as shown in Fig. 85 and Fig. 86. Note that the spacing between nulls is the same in both cases, but one starts as a high-pass filter while the other starts as a low-pass filter. In the case of the

Fig. 85 Out of Phase Summing

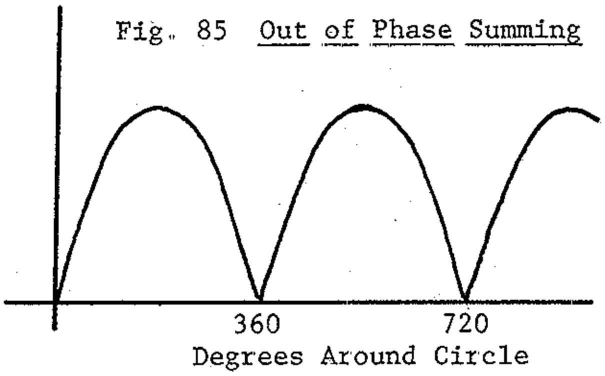
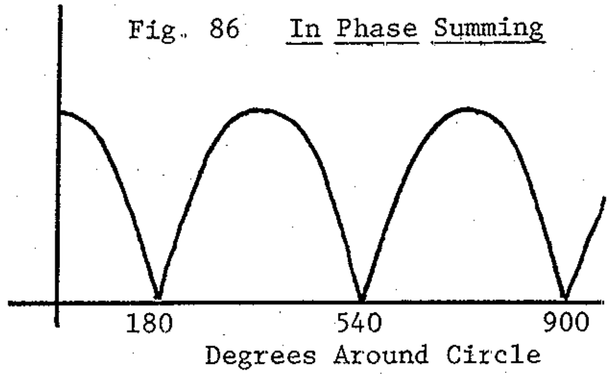


Fig. 86 In Phase Summing



second-order filter, typical results for poles placed approximately as in Fig. 84 are as shown in Fig. 87. In Fig. 87, the poles were placed at $+0.6 + 0.6j$ and $+0.6 - 0.6j$. The dotted line in Fig. 87 shows the results of placing a double zero at $z = -1$. Note the much sharper cutoff as a result of these zeros. The data on the case with zeros was lowered by a factor of four to make up for the gain of four that results at DC.

Fig. 88 shows the result of the matched z-transform applied to the 3db ripple Chebyshev. The filter shows the 3db ripple. The solid line is the case where $\tau = 1$ and shows the poor stopband rejection. The dotted line on the left shows both the analog Chebyshev, and the case where $\tau = 0.25$. The dotted line on the right shows the return of the $\tau = 0.25$ case. The curves have been slid left-right and/or up-down as necessary to make comparison easy.

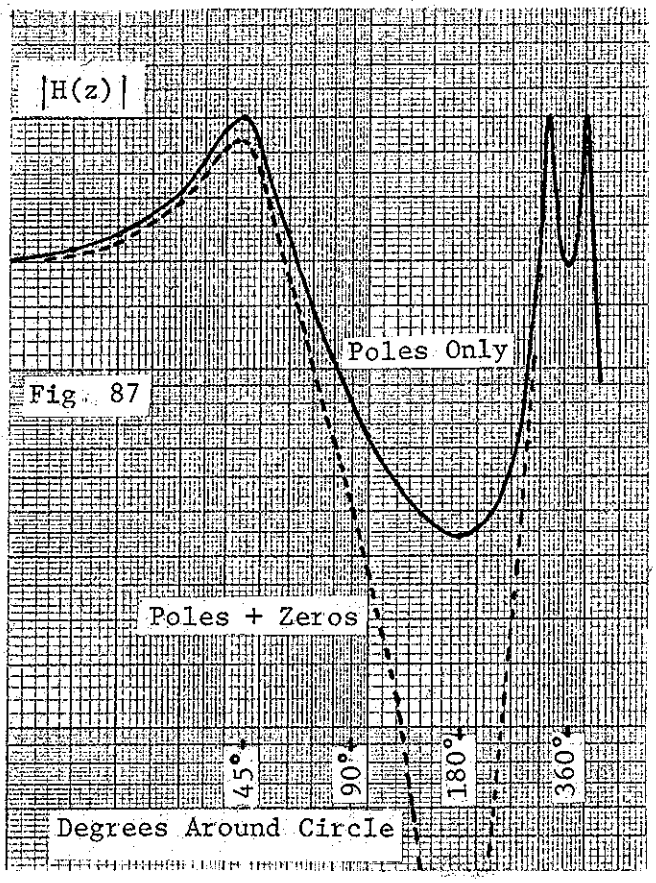


Fig. 87

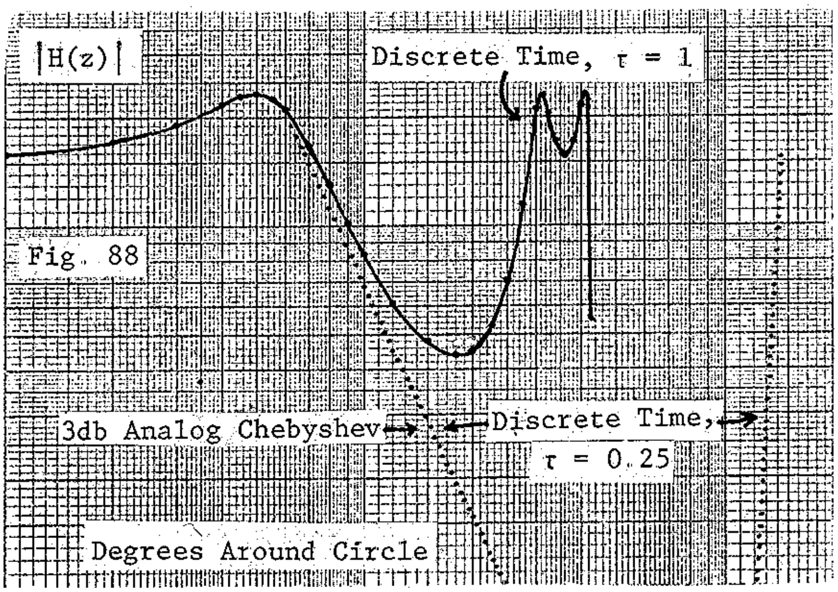


Fig. 88

INDEX

REFERENCES

The following Application Notes published by ELECTRONOTES will provide additional information on topics related to those in this manual:

AN-5 Notch Filters
AN-12 The Biquad Active Filter Circuit
AN-13 Equalizers and Tone Controls
AN-23 The CA3080 as a Voltage-Controlled Resistor
AN-25 Low-Q Bandpass Filter
AN-26 Bandpass Filter Examples
AN-34 Delay Line Setup Using the MN3001
AN-35 Delay Line Setup Using the SAD-1024
AN-36 Applying Delay Line Evaluation Setups
AN-37 High-Q Bandpass Filter
AN-38 High-Q Bandpass Filter Example
AN-47 Graphical Relationships in the z-Plane
AN-48 z-Plane Graphical Relationships for First-Order Systems
AN-49 z-Plane Graphical Relationships - A Second-Order Example
AN-50 Discrete Time All-Pass Networks
AN-51 Discrete Time Notch Filters
AN-52 Relationships for Higher Order Discrete Time Filters
AN-55 R-Filters
AN-56 R-Filter Method of Op-Amp Frequency Response Testing
AN-66 Lissajous Figures
AN-71 Multi-Mode Filter Based on First-Order Low-Pass
AN-72 The Twin-T Filter
AN-75 Designing Butterworth Filters Without Data
AN-76 Graphical Methods For Chebyshev Pole Placement
AN-77 TI-58/59 Program for $|T(s)|$
AN-78 TI-58/59 Program for $|H(z)|$

Additional information on filters, and particularly voltage-controlled filters applied to electronic music, can be found in regular issues of ELECTRONOTES NEWSLETTER.

Persons needing design data on filters which cover order three, or for order five or higher, will find this data in the Active Filter Cookbook by Don Lancaster, Howard W. Sams Co. Inc. (1975)

There are numerous good books on active filters. One of the few books that covers the most recent developments in active filters is Principles of Active Network Synthesis and Design by Gobind Daryanani, John Wiley & Sons (1976).

A good overview of filters of all types is found in Modern Filter Theory and Design edited by G. C. Temes and S. K. Mitra, John Wiley & Sons (1973)

At the time of this writing, digital filter books are appearing on a regular basis. We will leave it to the individual reader or his instructor to select an appropriate book on digital filters.