HANDBOOK - SHIPPING INFORMATION   Sept. 1975

1. Thank you for ordering the Musical Engineer's Handbook. Let us first apologize for any and all delays in getting this to you. Be assured that these delays have led to a bigger and generally more useful book than you would have received earlier.

2. BUSINESS: We have gone over your records fairly carefully and you will find attached a refund if any is due or a bill if it is the case that we have a debt recorded. This refund may be due to much older business, not necessarily the handbook order. The refund may be in stamps if it is a small amount.

3. CHECKING THE BOOK: Please take a few moments to check the book for missing pages or sheets that were misprinted, and let us know about any pages you need replaced as soon as possible. Note that the correct number of pages in each chapter is listed in the table of contents. We checked the pages fairly carefully when we got them from the printer, but there will likely be a few we missed and some collating errors. There is no need to return double pages - this is the most common collating error.

4. CORRECTIONS: Below are listed some errors we have found already. Surely there will be more, and these will be reported in ELECTRONOTES or in some manner. Please let us know about any you find if you get the chance to write.

5. REPRINT CHAPTERS: A few reprint chapters will likely be available in the near future. This will be announced in EN. Note that the book is copyrighted. In general however we will be glad to get additional exposure if anyone wants to use this material and acknowledge the source. Please write us first however.

Bernie Hutchins

Errors:

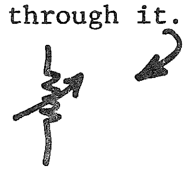
Pg. 1a (3), Line 6, Complete line as: proportional to the position of a key

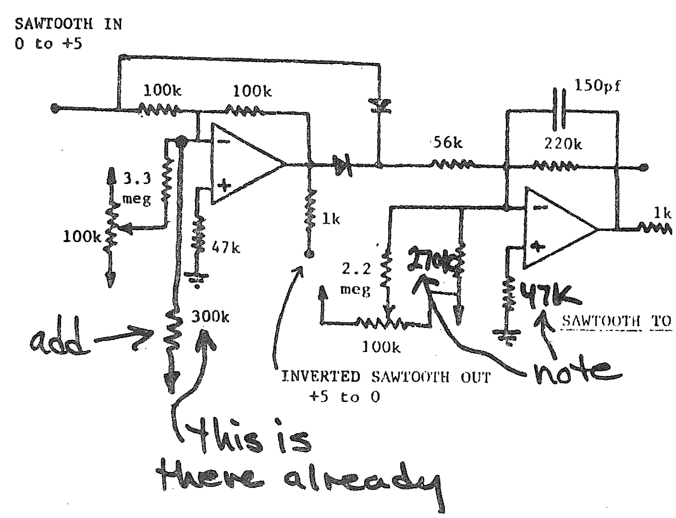
Pg. 4a (1), Diagram at bottom: change 1k to 10k and 56k to 560k, both cases.

Pg. 4b (7), Top Line: change first to second

Pg. 5b (16), Lower diagram: add a 300k resistor mark resistors 270k and 47k as shown at the right.

Pg. 5c (10), 25k resistor should be a pot (Exp. Gain), so draw an arrow through it.





MUSICAL ENGINEER'S HANDBOOK

FIRST EDITION

MUSICAL ENGINEERING FOR ELECTRONIC MUSIC

BY

BERNIE HUTCHINS 1975

PUBLISHED BY

ELECTRONOTES

DEDICATION:

This book is dedicated to musicians everywhere and of all times, without whom many of us would have only electronics to do.

ACKNOWLEDGMENT:

Much of the material in this book has been obtained from ELECTRONOTES and therefore the author is indebted to all those who have contributed to ELECTRONOTES over the years. Where possible, their ideas have been acknowledged in the text. There are numerous persons who have also contributed greatly to the field of musical engineering, either directly or indirectly, and where possible, we have acknowledged their efforts as well. We can also add to the list all those who have supported ELECTRONOTES over the years and who have helped make this book possible and worth writing.

TABLE OF CONTENTS

INTRODUCTION

It is probably most useful to think of this handbook as a reference rather than a text. The material consists of what can be considered groups of "articles" (marked as chapters) on various subjects. The technical level, necessary background, format, and degree of coverage of the subject matter varies greatly:

The technical level varies because, as Einstein said: "Things should be made as simple as possible, but not simpler."

The necessary background required varies because it is necessary to start somewhere. Since the background of readers of this book will vary greatly, it is reasonable to start at the most convenient place. This means that some readers will not have the necessary background for some chapters, but will for later chapters. At time reference can be made to different chapters in the book, but since the book was not prepared from the first page on, but rather as material developed, exact references are sometimes not possible and the phrase "the chapter on ...." frequently had to be used.

The format varies because chapters were prepared at different times and with different degrees of care to the diagrams. In places, material has been used directly from ELECTRONOTES.

The degree of coverage varies not necessarily with respect to what I consider the degree of importance, but rather with respect to what I felt was well established and useful. Some chapters on important subjects are short because the material is available in a better form elsewhere, or we couldn't say exact things about the subject and give practical circuits at the time of writing. The handbook is intended for proven, useful information. Developing designs and information appear in ELECTRONOTES.

All of the above can be thought of as an excuse for shoddy organization, if you wish. None the less, many readers may want to take "the grand tour" and will find the conventional front-to-back reading style to be appropriate. It is suggested that everyone skim the entire book at least once as some items may be found in unexpected places.

The chapters are arganized into sections. Sections 1 and 2 are basics with Section 1 being more general while section 2 gets down to more specific synthesis techniques and methods. Section 3 deals with electronic components and their basis electronic music applications while section 4 treats some special circuit designs that are extremely useful for electronic music but which are still useful in other fields of electronics. Sections 5, 6, and 7 deal extensively with the design of electronic music equipment. Section 5 *deals with the "conventional modules" while Section 6 deals with the less common devices. Section 7 deals with control devices. In these sections, we have purposly avoided a "cookbook" or "circuits sourcebook" approach. Instead we have tried to show how designs are motivated, giving the designs in bits and pieces. The actual design examples are considered to be reliable designs and do illustrate many of the points of the chapters. Note however that they are not to be considered the last word in the designs. In many cases we have indicated in the chapter how improvements can be made (e.g., high frequency tracking of VCO's, linerization of VCA's by predistortion) but do not include these in the design examples. The individual designer may want to include these, but they may be unnecessary refinements in many cases. Other examples appear in ELECTRONOTES. Bear in mind also that improved designs will likely appear in ELECTRONOTES from time to time. Thus, these chapters of the handbook will serve to show the principle considerations of current designs and to free future articles in ELECTRONOTES from additional repetition of basic ideas.*

Section 8 is the "Operational" section and treats actual system design, construction, and troubleshooting. The emphesis here is on the "quick and dirty," - getting things done. Those readers whose work generally demands more refinement should have no problem working with "neater" techniques.

Section 9 is the reference section. Note that this includes a bibliography for all the other chapters and includes useful references that are not specifically referenced in the text material. Also listed are references from ELECTRONOTES that were not transferred to the handbook for reasons of space. The reader should be aware of these articles.

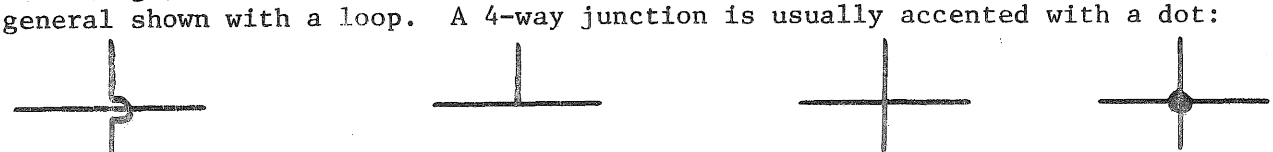
The handbook is intended to serve as both an equipment reference and a theoretical reference. Musical engineers will often be called upon to handle theoretical problems as well as circuit problems. For this reason, many mathematical and other models have been included.

As this is a first edition, there will likely be numerous errors, and we will appreciate your bringing them to our attention and making suggestions about how the book could be improved.

Bernie Hutchins
Ithaca, NY
July 1975

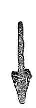
CONVENTIONS USED IN THE BOOK:

For the most part, a general part type number is used instead of an exact manufacturer's type number (e.g., 741 instead of MC1741 or μA741, etc.). Wires which cross on a schematic are in general shown with a loop. A 4-way junction is usually accented with a dot:



NO CONNECTION | CONNECTION | CONNECTION | CONNECTION

Standard +15 volt and -15 volt supplies, along with a +5 volt TTL supply are generally used. These will be denoted as follows.

+15  -15  +5  Common Ground 

In many cases, (CMOS, CDA's), the exact voltage can vary greatly from +3 to +18 or so. In such cases, we will use the +15 volt symbol, as this will generally be available. When the +5 volt symbol is used, the voltage must be from a +5 supply.

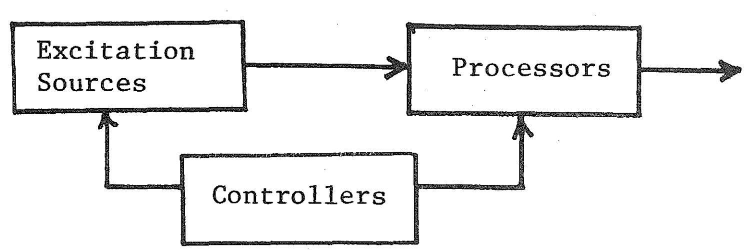
CHAPTER 1A

ELECTRONIC MUSIC SYSTEMS AND THEIR CHARACTERIZATION

CONTENTS:

INTRODUCTION

Most electronic music systems can be put into a very basic form consisting of three elements: Excitation sources, Processors, and Control devices. The interconnection is of the form shown below:

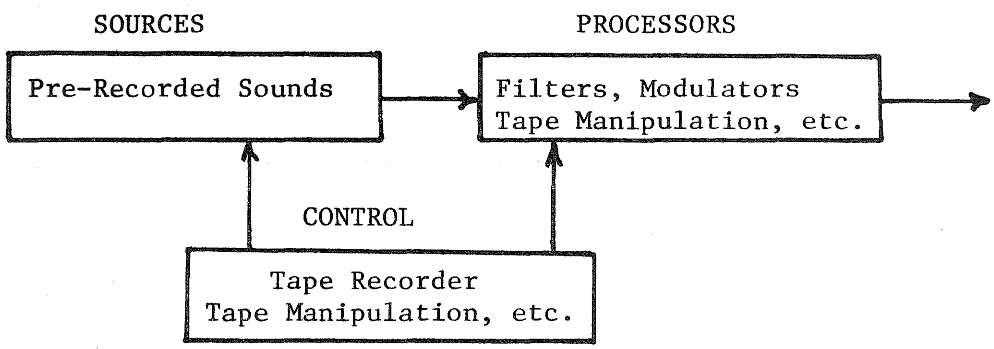


Typical excitation sources are oscillators of many forms, noise sources, and partially processed (e.g. recorded) sounds. Typical processors are filters, modulators, and controlled gain amplifiers. Typical controllers are keyboards, sequencers, translation devices (e.g., pitch to voltage converters), and computers.

EXAMPLE ELECTRONIC MUSIC SYSTEMS

MUSIQUE CONCRÈTE:

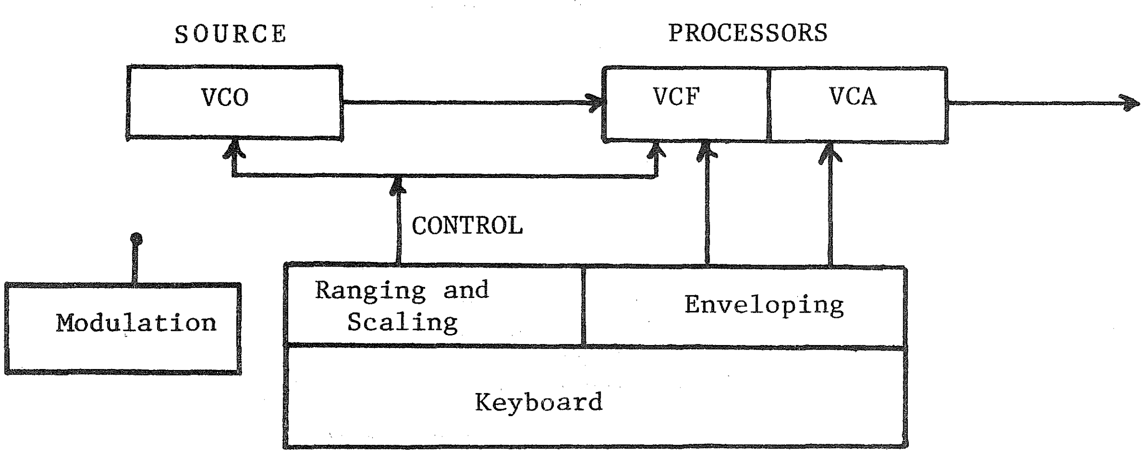
Musique Concrète involves manipulation of live sounds into a musical composition. The manipulation always involves a tape recorder to record the sounds and/or to process them. Musique concrète is the earliest form of electronic music, and one which is still used today by some serious composers along with newer methods, and as a teaching vehicle. We can fill in the blocks of the electronic music system as follows for musique concrète:



With the tape recorder, a live sound may be passed through a filter to change it and this can be recorded. The recorded sound can be further processed if desired. The tape can be chopped up and spliced in different ways. The tape speed and direction can be changed. Tape loops, delay lines, and any number of other techniques can be used.

VOLTAGE-CONTROLLED ELECTRONIC MUSIC SYNTHESIZERS:

The voltage-controlled electronic music synthesizer is probably the best known of all electronic music systems. These were developed by Robert Moog, Donald Buchla, and others in the late 1960's. The block diagram for a typical synthesizer is shown below:

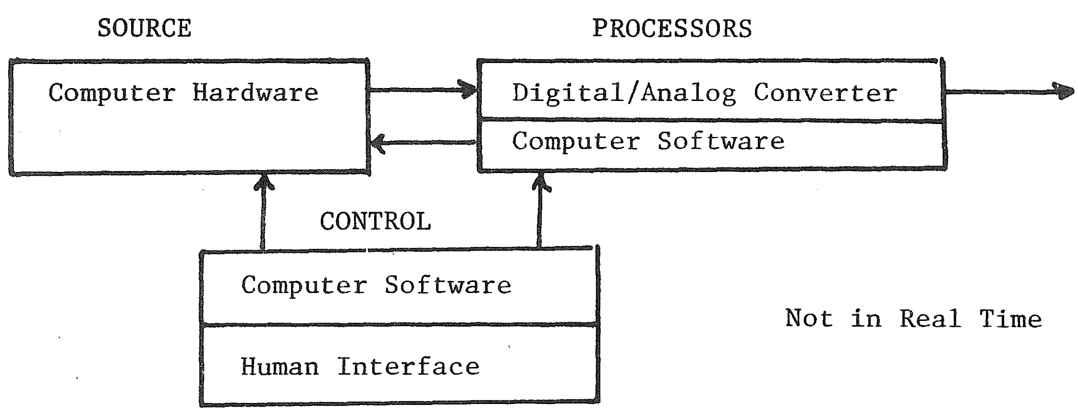


An important feature for the voltage-controlled synthesizer is that all the control lines from the controller (which is usually a keyboard) to the synthesizer modules are control voltages. The modules respond to these voltages to control pitch, amplitude, and tone color. The basic information from the keyboard (which key is down and when) can be expanded by setting circuits to respond to timing signals to give an overall set of control voltages that would be either to complex or to fast to obtain with manual control alone.

The keyboard controls the basic pitch level of the system as each note on the keyboard is pushed. In addition, timing signals from the keyboard produce envelopes. These envelopes are voltage contours which change through a full cycle each time a note appears. For example, one envelope will control the amplitude, defining a beginning and end to a tone. The voltage-controlled oscillator (VCO) produces a waveform rich in harmonics. The pitch of the VCO is controlled from a voltage that is proportional to the on the keyboard. The tone color is determined by the voltage-controlled filter (VCF) which usually has two control voltages: the keyboard voltage for pitch level, and an envelope to change the tone color as the note progresses. The VCF alters tone color by changing the harmonic content of the waveform from the VCO. Finally, a voltage-controlled amplifier controls the amplitude of the output signal. This amplitude is generally independent of pitch level and thus only an envelope is applied to control it. Various modulating signals may be inserted in different points to enrich the tone by adding additional harmonic and/or non-harmonic components.

COMPUTER POINT-BY-POINT SOUND SYNTHESIS:

Any sound waveform can be represented by the variation of its amplitude as time progresses. Sampling theory tells us that if the sound has frequency components up to a maximum value of f_{max}, then we can represent the sound by samples taken at a rate of $2f_{max}$ and no information will be lost. For audio waveforms which may have frequency components up to 15,000 Hz, a sampling rate of 30,000 Hz is enough. Thus, we can envision a digital computer that has been programmed to produce any desired sound. It first generates one number, then goes back to find what the next number should be and updates the output, and so on, 30,000 times per second. However, for all but the simplest waveforms, this calculation can not be done fast enough for the computer to operate in real time. Thus, the digital data must be stored during generation and played back later in real time once the computer has finished. A block diagram of such a system is shown below:

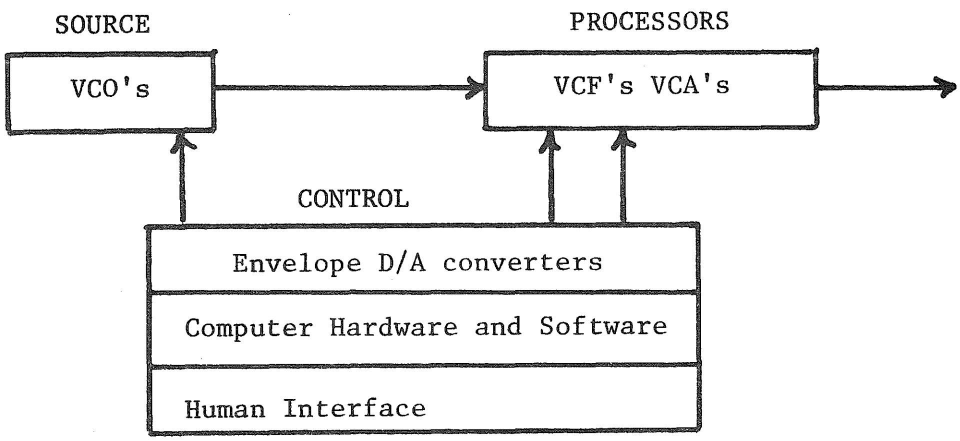


It is probably evident that the computer software (i.e., the way the computer is programmed and the way the user makes it do what he wants) is the key to the system. It would be nearly unworkable were it not for the fact that certain routines can be programmed in for repeated use. For example, it would be much easier to represent a sine wave of rising amplitude by its frequency, starting amplitude and time, and rate of change (four numbers) rather than to write a computer program to do this each time it is needed. A number of useful computer programs have been developed which give the user control over the major options with a minimum knowledge of computer programming.

A number of serious drawbacks remain with computer point-by-point methods: (1) The cost of time on a computer large enough to do the job can be prohibitive. (2) While some very specific sounds can be generated and often are by persons doing scientific type investigations, the library of sounds available to the composer is restricted unless he learns a good deal about computer programming. (3) The system works out of real time so there is a delay between the time the composer tries something and the time he finds out how it actually comes out. (4) While the system is based on sound principles of information theory as far as the mathematics goes, the ear does not need anywhere near the amount of information that is used. It is as though the bank sent you a statement every half hour.

COMPUTER CONTROL OF ANALOG SYNTHESIZERS:

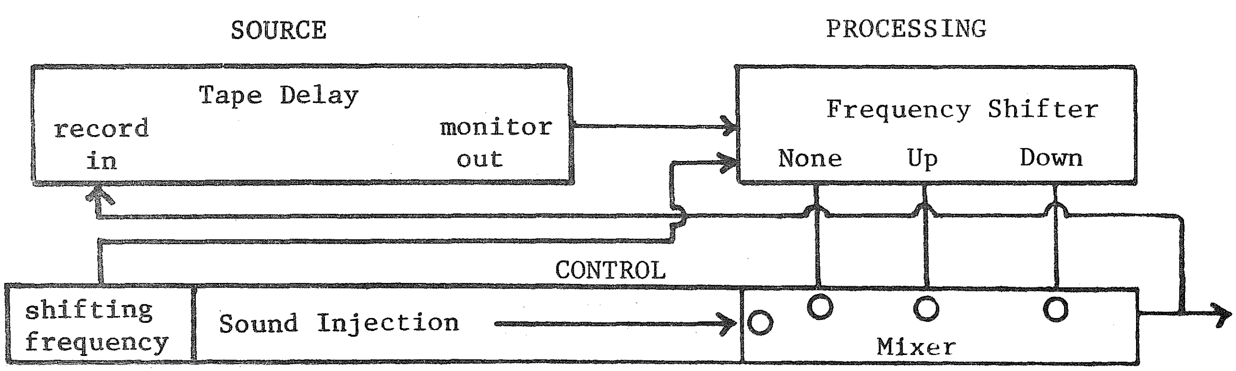
A computer can be used to supply a set of slowly varying analog control voltages rather than every point of the waveform to be generated. Even a very small computer may have the ability to keep up with this. The block diagram of such a system is shown below:



In this way, the computer serves as a multiple envelope generator where we have used the term envelope to refer also to voltages with discrete level changes such as are required for many types of pitch control. The human interface can be a standard musical keyboard, or a number of other controllers. The computer can operate in real time in response to timing commands from the human interface. Also, it is relatively simple to just have a record of the pitches and timing stored in the computer. Then the whole musical passage can be obtained at the push of a button. The computer can be thought of as a modular system where certain of the modules can be created as needed and in the numbers needed for a specific application. Since these are only software modules, they can be changed by just changing the computer program.

A TAPE DELAY FEEDBACK SYSTEM:

The block diagram of this system is shown below as an example of a non-standard system that can be rigged up for a specific application:



The tape delay line is formed from a tape recorder that has a record head and a playback head as well as the erase head. There is a delay as the tape passes between the record head and the playback. The signal from the playback head is fed to a frequency shifter (see chapter 6a) which is a means of changing the frequency of each component of an input spectrum by the same number of Hz. Control is obtained by forming a mixture of the original unshifted signal, the upshifted signal, and the downshifted signal, and feeding this back to the tape recorder input. This signal later emerges and is again processed by the frequency shifter and so on. Additional control can be obtained by injecting any desired sound into the tape recorder input through the mixer, and by controlling the shifting frequency.

This is the type of system that can be controlled without a lot of ability in the techniques of traditional music. For much of the time, the system does its own thing, but it is not random, produces many interesting sounds, and with practice it can be controlled quite well. Many unexpected sounds can be edited into other compositions.

THE CHARACTERIZATION OF THE ELEMENTS OF ELECTRONIC MUSIC SYSTEMS

EXCITATION SOURCES:

Excitation sources are usually characterized by their spectral content. For example, a sine wave output of an oscillator has a single spectral component, its frequency. A sawtooth output has all the harmonic overtones of the fundamental frequency in its spectrum. A noise source will have a continuous spectrum in general - all frequencies. It may also be possible to characterize the spectral evolution of the source as time progresses (if any). Such a progression may be found if the source has already undergone some processing, or if some processing is done on the source (e.g, dynamic depth FM resulting in a spectrum in which the amplitude of the components is changing).

PROCESSORS:

Processors are characterized by the parameter which they control. A filter is thus characterized by the way it alters the properties of the spectral components at its input. A low-pass filter passes the lower components and rejects the higher ones. Thus, there is a characteristic frequency of the filter (the transition region) that is of interest. A modulator is characterized by the property of the input that is changed. A VCA thus can produce amplitude modulation since a VCA controls amplitude. A pulse-width modulator changes the width of the pulse train at the output. Properties that would be of interest for a PWM system would be initial width, final width, rate of change, and directions of change in time.

CONTROLLERS:

Controllers are characterized by the human interface and the use of the controller output. A keyboard is manually controlled since it is the action of the players fingers that determine the output. If the output is a voltage, it can be used to control the pitch of a VCO or any other voltage-controllable parameter in the system. Other controllers are programmable - they are set up and activated. They do not depend on manual control except for activation. Such controllers are sequencers and computers. A third type of controller is the translation device. This device takes one sort of input which is under the control of the user and translates it into a control means that the system will understand. A pitch-to-voltage converter is such a device since it could be used to recover frequency information from the users favorite instrument (e.g., a violin) and use this to control a VCO. By use of such devices, the musician can use his own familiar instrument and expect that the subtle nuances he is able to achieve with it will be transferred in some form to the electrical device.

CHAPTER 1B

WAVEFORMS, ENVELOPES, MODULES, AND CONTROL

CONTENTS:

INTRODUCTION

The more or less standard "voltage-controlled electronic music synthesizer" which is known in many circles as just the "synthesizer" has been responsible for a large portion of the current interest in electronic music. Admittedly, much of the commercial success of the synthesizer has been "gimmicky", but this does not alter the fact that the synthesizer has also proven an extremely viable system for producing several types of music. Furthermore, understanding of the operation of voltage-controlled synthesizers will lead naturally to the understanding of other electronic music systems.

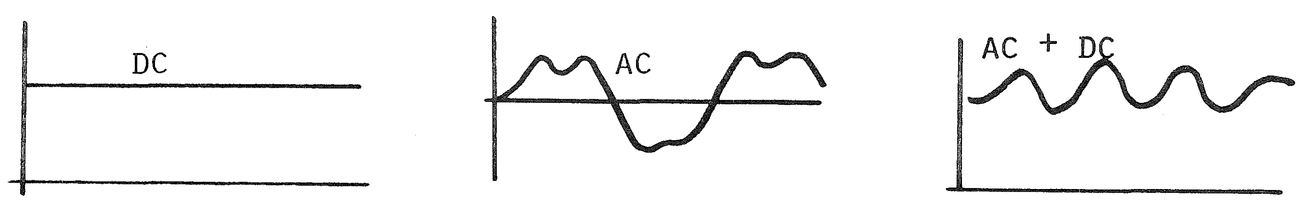
In this chapter, we first discuss the basic ideas of waveforms, and this is a feature that must be considered in all systems. Then, moving toward the synthesizer, we discuss the idea of an envelope. Next, we discuss the various types of modules that are most common, and finally we give a brief outline of the major means of control. The reader will find discussions of design details for the various modules in chapters 5a through 5j. Synthesis methods are discussed in chapters 2a, 2b, 2c, and 2d. Additional information on controllers can be found in chapter 2d and chapters 7a through 7c.

WAVEFORMS

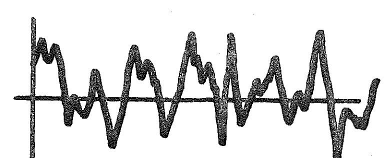
All electrical signals in electronic circuits have some sort of waveform. Direct current (DC) is the simplest, white noise is the most complex, in fact, it defies analysis directly. Probably the most interesting waveforms are the periodic, repeating waveforms.

To observe a waveform, we measure an instantaneous value of the voltage, and as time goes by, we make some sort of recording of these values and study this record. It should be made clear that when the record takes the form of a visual representation, e.g., is drawn on paper or displayed on the face of an oscilloscope, the physical quantity that is the "reference quantity" is <u>time, not space</u>. Time has been represented by space in such a case. So what? Well, time is an entirely different sort of quantity than space, and while some voltage waveforms actually represent space anyway, in music, we ultimately get down to the movement of the eardrum in time, and confusing the actual dimension involved will sometimes confuse what we are trying to do. A voltage varies in time, but we cannot "see" time, so that is why we make do with a space dimension, usually horizontally, left-to-right, to represent time, past-to-future. Keeping this "record" of instantaneous voltage is important in the following sense: If we measure the voltage only once, e.g., 3.2 volts at 5:00 PM, this tells us almost nothing about the waveform of which this "sample point" may be part. To determine the waveform, we need many measurments in time. Of course, at audio frequencies, these measurments must be made very rapidly, therefore automatically, as on an oscilloscope, which also displays the record. If you have ever watched the spot on an oscilloscope screen scan slowly across the face while moving up and down more rapidly, you can easily see how the left-to-right motion represents time. The same is true even when the spot moves too fast for your eye to see as anything but a continuous line.

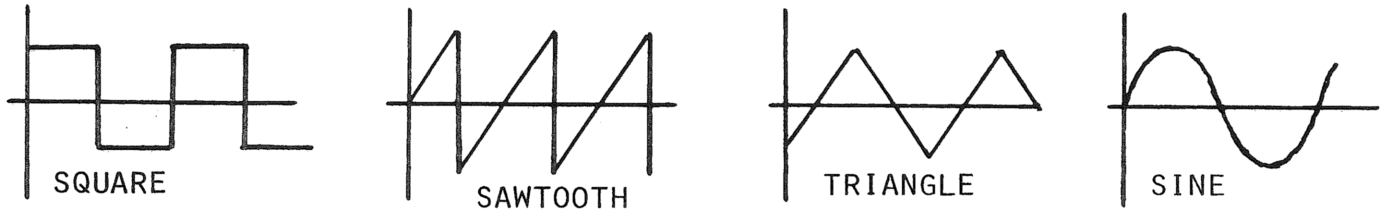
Direct current (DC) voltages cause electronic current of flow in only one direction. For something like a power supply, with hopefully a constant voltage, there is no variation in time, and thus nothing that the ear could hear (the eardrum must move in time) were this voltage converted to a sound wave. Alternating current (AC) voltages cause current to change direction from plus to minus and back rapidly, and since these can become audio sounds, it is mainly AC that we are interested in. Often voltage waveforms have both AC and DC "components":



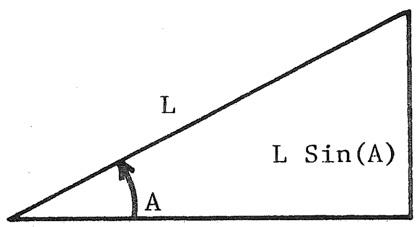
White noise is a waveform that appears randomly. One cannot predict when or where or how much voltage will appear. However, even though all previous records of a white noise waveform will not help in the prediction of the future waveform, the sound does have a characteristic sound, sometimes refered to as "static".


WHITE NOISE

Periodic waveforms are formed when a pattern is found in the record of the voltage. For example, a "square-wave" is actually reversing DC, and would at first appear as DC to an observer, but after a few cycles, the pattern is clear. On the other hand, the observer would immediately determine that a "sawtooth" waveform was not DC (because it is always changing), but still we must observe the whole cycle to determine if it is a sawtooth or perhaps part of a "triangular-wave".

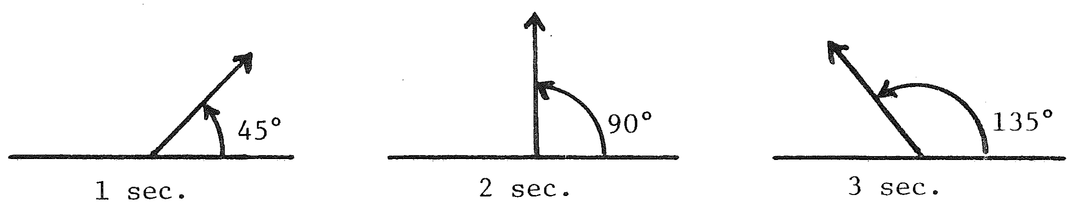


A "sine-wave" is the most basic periodic waveform, but certainly not the easiest to understand. Actually, it comes from trigonometry, where you may recall a triangle has the following properties:

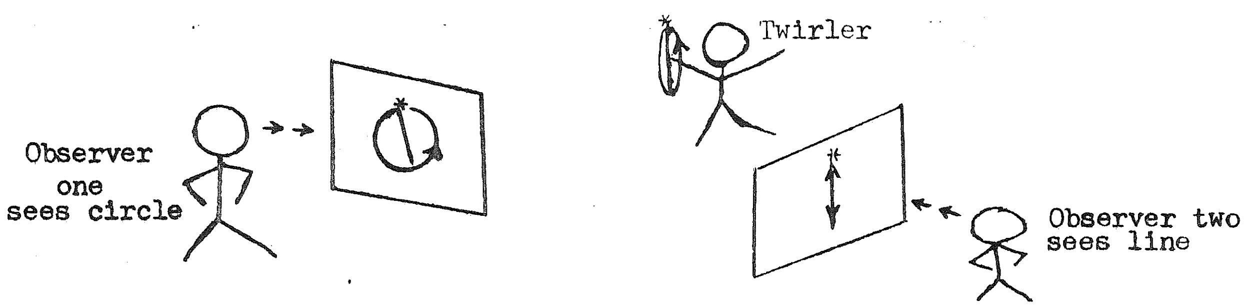


Where A is an angle, and the notation "Sin A" represents the "sine of A", etc. In essence, what we are saying is that if we put one end of a fishing pole of length L on the ground and aim it up at an angle A, the length of string to reach from the other end to the ground can be represented as L times the sine of A.

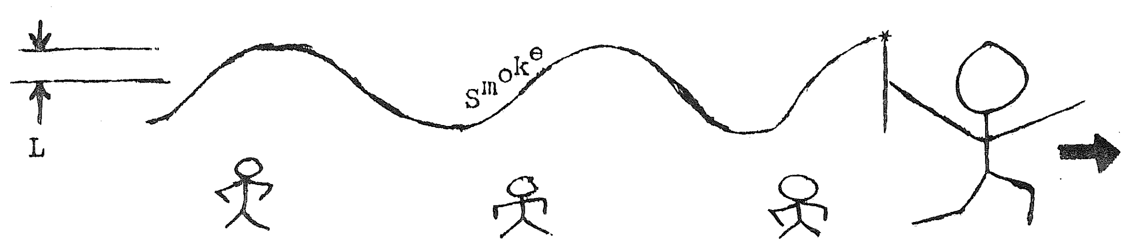
Now suppose we make the angle A depend "linearly" on time. In math this is to say that A is a "function of" time. For example, a rod rotating at a constant rotational speed might be aimed at $A=45^{o}$ at one second, 90^{o} at two seconds, 135^{o} at 3 seconds, etc.



Likewise, consider a baton twirler with a star on one end of the baton. As indicated below, observer (1) sees the star move in a circle, while observer (2) sees the star move up and down, and has no idea that the star is actually moving in a circle.



Now consider what the crowd (or a person walking beside the twirler) will see as the twirler moves down the street. We put a smoke marker on the star, so that the star leaves its own record. The result is a sine wave. Note that this is a left-to-right representation of time as discussed above.

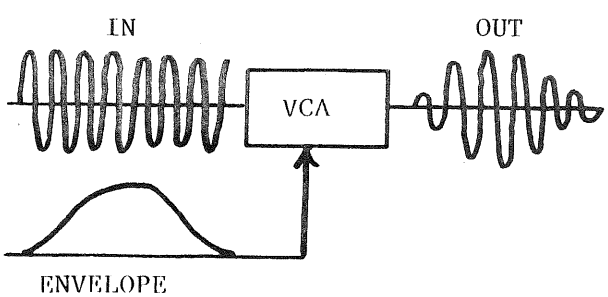


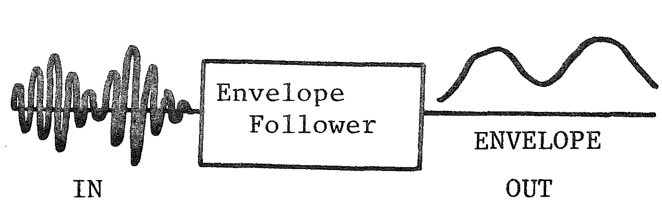
Leaving the twirler and going back to math, we can see why the smoke trail came out as a sine wave. The twirler spins the baton as a linear function of time: $A = \omega t$. [The Greek letter omega (ω) represents a constant speed of rotation that is measured in radians per second. One radian is $360°/2\pi$ or about 57°] The radian notation is the most natural mathematically, but the ordinary frequency in Hertz or the older name "cycles-per-second" is more familar. The relationship is $\omega = 2\pi f$ where π is the familiar constant 3.14159.... Thus we can write $A = 2\pi ft$, and since the twirler is seen side on, we see the sine, and therefore the waveform seen is $L \sin 2\pi ft$. Electrical voltages with this time dependence are sine waves.

Let's clean up the nomenclature a little. A periodic waveform can start at any one of its points, and proceed in time. Once back to its starting point (which will be at the starting voltage, although it may have passed through this voltage several times), we say it has completed one cycle. The time to do this is called the "period" The number of cycles completed in one second is called the "frequency", and you can consider this to be related to musical pitch. Frequencies are measured in "cycles-per-second", but someone has (gasp!) seen fit to recently change this clear and meaningful notation (c.p.s.) to "Hertz" abbreviated Hz. When we say one cycle, it is the singular of Hertz that we have in mind (whatever that is). Thus, one Hz is one c.p.s, etc. One thousand Hz is (or are?) refered to as one kilohertz (Khz). The human ear can hear sounds with frequencies corresponding to audio signals in the range from about 15 Hz to 20KHz. Frequencies below and above these (sub-sonic, and ultra-sonic) are useful for control signals and for modulation effects.

ENVELOPES

An envelope is a voltage waveform that varies in time, usually much slower that regular audio signals. These waveforms often repeat, but only at irregular intervals, often with different periods. For example, there could be an envelope appearing for each musical note in a composition. The magnitude of the voltage at any point of the envelope waveform controls or monitors a parameter of the sound. The most common example is the amplitude envelope as applied to a Voltage Controlled Amplifier (VCA) to give attack, sustain, and decay, but the reverse process of recovering the envelope from the signal is also useful, and is called envelope following. In either case, the envelope voltage is directly proportional to the peak height of the signal at the corresponding point, or as an "Upper Limit" on the peak at the points inbetween.





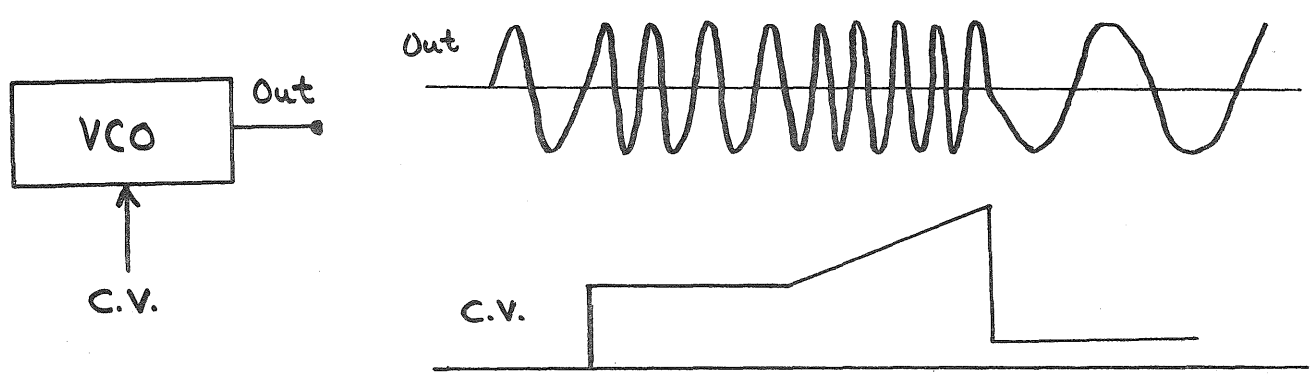
Among the important properties of amplitude envelopes are the "attack", the "steady state" or "sustain", and the "decay". The attack is the initial part of the envelope (e.g., how does the amplitude vary immediately after the key on a keyboard is depressed?) and the decay is the final part of the envelope, (e.g., after the key is released how long does it take for the sound level to die out?). Attack and decay are called the "transients", and familiar examples are the slow attack we hear as the amplitude builds up in a pipe organ, and the slow decay of a piano as a key is struck and held down. The transients are very important in making a musical sound interesting enough for the ear to bother to listen to it. Inbetween attack and decay is the steady state, a time during which the waveform magnitude is relatively constant. This would be present in an organ held tone, or a trumpet held tone, but not in a piano, where the decay begins immediately after the string if struck, i.e. attack and decay are directly connected. The steady state is probably only obtainable exactly in electronic instruments, since traditional instruments involve human beings, who can not hold the sound constant, but only approximate it. However, it is these subtle changes, in harmonic content, if not in amplitude, or slightly in pitch, that make the steady state at all interesting to the ear. Thus in electronic musical instruments, we must take special effort to make the steady state imperfect! This can be done with special control voltages of relatively low magnitude. Two familiar examples are the shallow amplitude variations at about 7 Hz, known as tremolo, a form of amplitude modulation, and similar variations of pitch, known as vibrato, a form of frequency modulation.

In addition, envelope waveforms can be used to control or monitor many other parameters of the sound, e.g. the harmonic content, i.e., the waveform of the output signal. Harmonic content can be controlled in two ways. First of all, the waveform from some source can be filtered, and the instantaneous characteristics of the filter are controlled by a harmonic envelope applied to a Voltage Controlled Filter (VCF). Thus the envelope controls the filter, and the filter controls the waveform as time passes. A second method is a so called "additive synthesis" method. In this case, a large number of harmonics are available, and the amount of each of these that we want in the final waveform is controlled by its own envelope. Here, the harmonic in question is applied to the input of a VCA, and the VCA is controlled by an envelope. The output is a mixture of such harmonics.

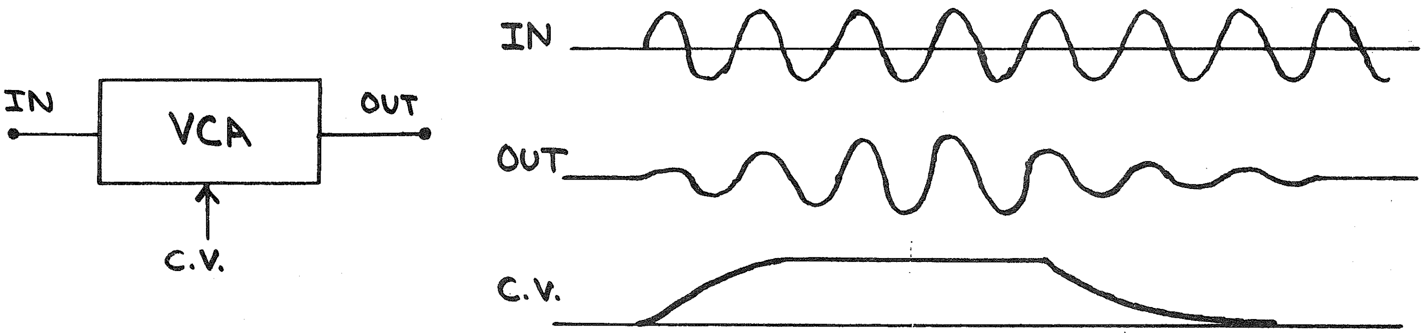
MODULES

Below are described briefly some of the more typical electronic music modules. A module strictly speaking is a functional block that is formed into an actual "black box" type of circuit. The user then selects whatever combination of modules he needs, places then in their proper positions, and interconnects them. This leads to the idea that the modules should be mobile - they can be set in different positions, different studios, etc. This type of packing can be very useful, but also very expensive. The actual circuitry and performance are independent of the mobility - only convenience will suffer from lack of mobility. Thus, we shall consider a circuit to form a module even if it is not mobile. In this case, the interconnections to the module have fixed panel positions and patch cords have to be brought to these positions.

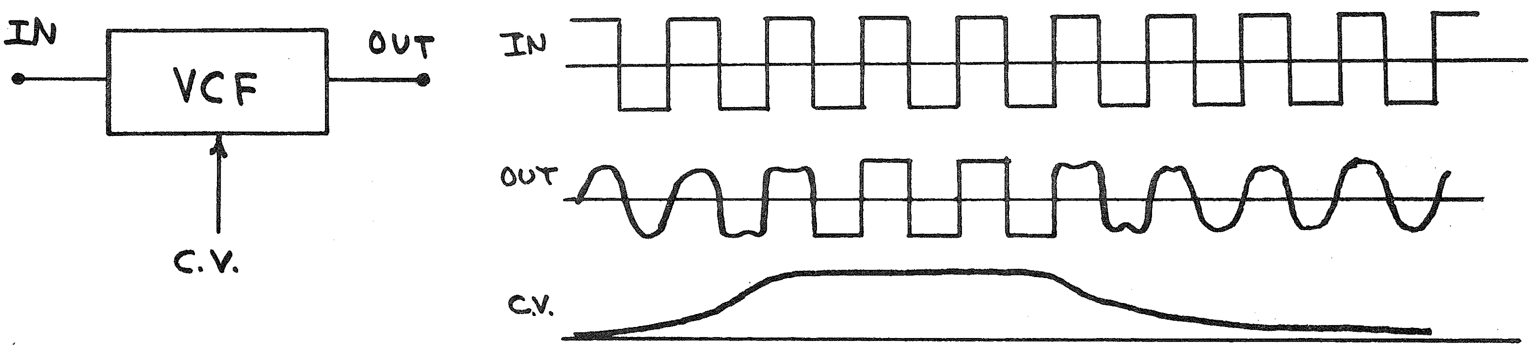
<u>VOLTAGE CONTROLLED OSCILLATORS</u>: A VCO is a waveform producing device. The frequency of the waveform depends on a voltage that controls it. This process is diagrammed below:



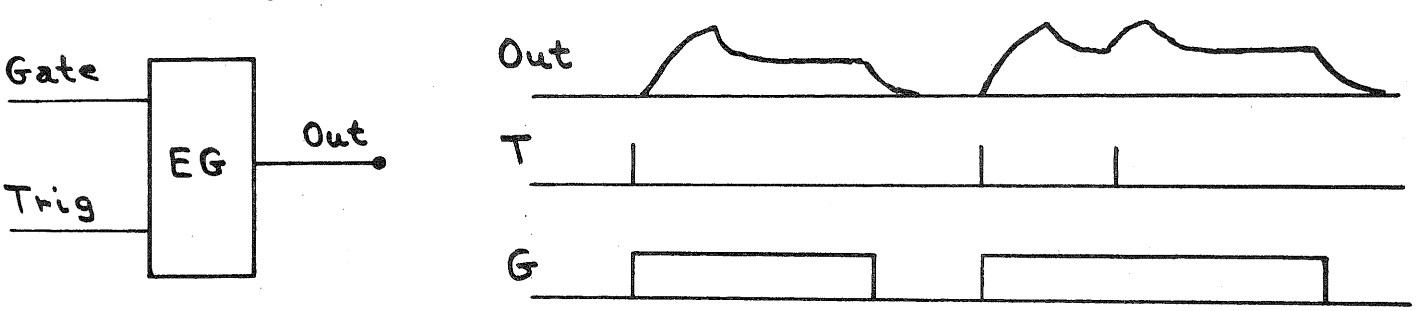
VOLTAGE-CONTROLLED AMPLIFIERS: The VCA is a waveform processing device. The amplitude of the waveform depends on a voltage that controls it. The process is diagrammed below:



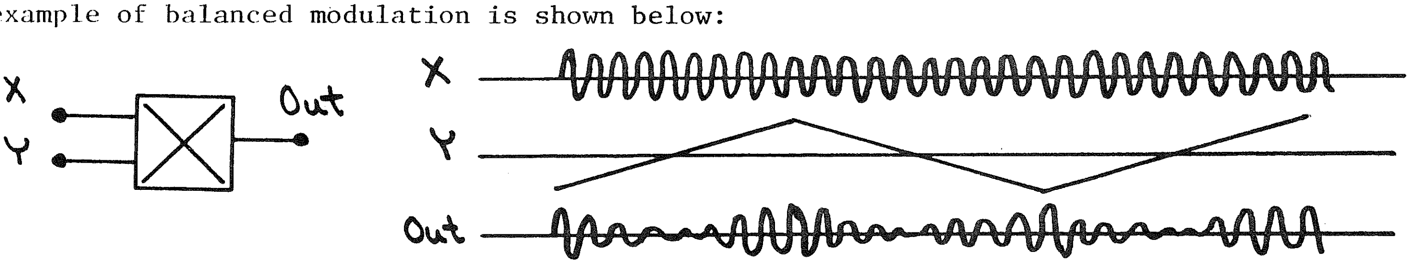
VOLTAGE-CONTROLLED FILTERS: The VCF is a waveform processing device. The harmonic content of the waveform is altered in response to a control voltage. In the example diagrammed below, we have assumed that the VCF is in a low-pass mode, and that the filter's cutoff frequency is controlled by the voltage. As the voltage rises, more high frequency components from the input square wave are allowed through. This alters the waveform from a square shape for high control voltages to a sine wave for low voltages where only the fundamental component is allowed through (see chapter 1d for more information on harmonics).



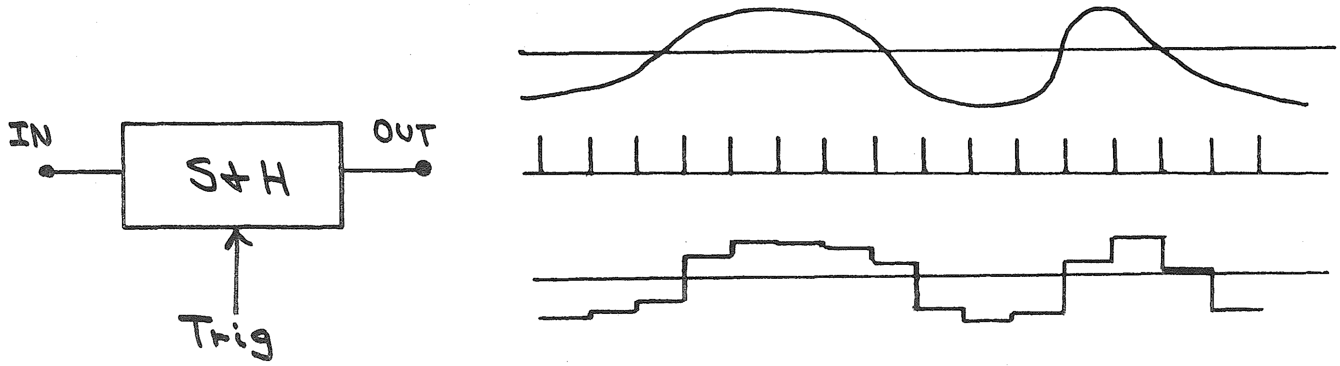
ENVELOPE GENERATOR: The envelope generator (also known as a transient generator or a contour generator) is an envelope producing device. The envelope is used as a control voltage, typically to control a VCA or a VCF. The EG responds to timing control signals, typically from a keyboard.



BALANCED MODULATOR: A balanced modulator (also known as a "ring modulator" or as a multiplier) multiplies two signals together. It can serve as a VCA as well as a modulator. An example of balanced modulation is shown below:



SAMPLE-AND-HOLD: The sample-and-hold is a waveform or envelope altering device. It chops up waveforms into discrete steps by catching an instantaneous value of the waveform when it is commanded to sample, and then holds the value until another sample command arrives.

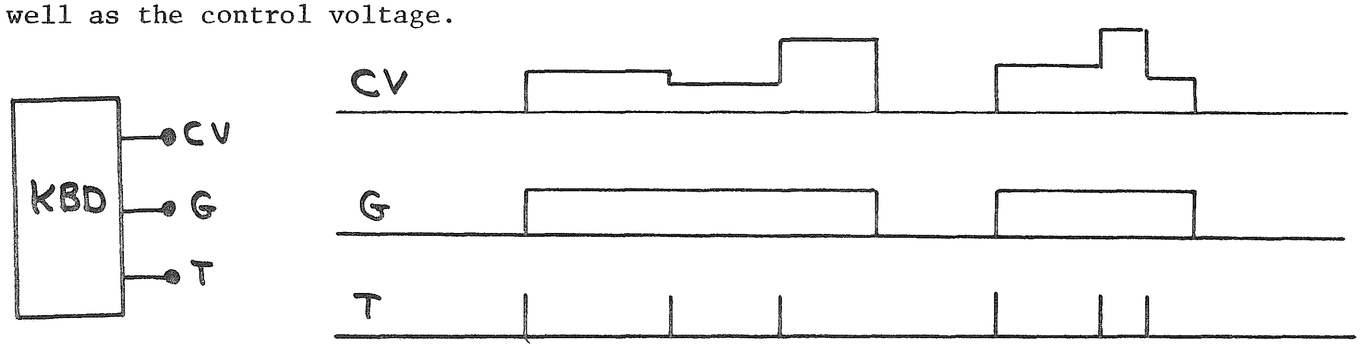


NOISE SOURCE: A noise source is a signal producing device. The signal is typically white noise (the same as the interstation "static" noise on an FM receiver). The noise is useful since its spectrum contains all frequencies and can be filtered for colored noise. The noise can also be sampled to give slower random signals.

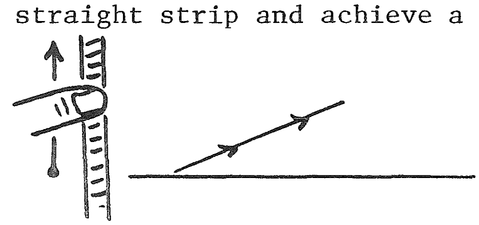
CONTROL

At some point some means to give the musician control over the various modules has to be provided. This applies to all electronic music systems. Below we will be discussing sources of control voltages for electronic music modules. It should be mentioned that a voltage is a control voltage by the way it is used, not by its source. Any voltage; waveform, envelope, or specific control source voltage may be a control voltage. A control voltage is one which is applied to a control input.

KEYBOARDS: Keyboards produce control voltages that vary discretely (in steps). Usually one voltage for each key. The actual voltage handling system involves some sort of interface that converts switch closures to control voltages, and may also provide timing signals such as gates (some key is down) and triggers (the key that is down is changing) as well as the control voltage.



CONTINUOUS CONTROLLERS: The keyboard provides control voltages with discrete steps. It is often useful to have continuous control of voltage. A simple potentiometer is such a continuous control, essentially just a voltage divider with an adjustable center point. Another form called a "ribbon controller" is also popular. With the ribbon controller, the performer can move his finger along a straight strip and achieve a continuous range of output voltages. An example output of such a continuous controller is shown at the right. Another controller known as a "joy stick" or X-Y controller produces two control voltages for different motions at 90° to each other.



SEQUENCERS: A sequencer is a control device that is programmed to provide a sequence of voltages upon command. Once the device is set, the sequence is available as needed. Sequencers thus can be thought of as producing a number of sequenced events where each event is represented by certain times and by certain voltages. A sequence might be: 1 sec. V_1 = 2 volts, V_2 = 3 volts; 2 seconds, V_1 = 1 volt, V_2 = 4 volts; 5 seconds, V_1 = 3.33 volts, V_2 = 2.56 volts, etc.

Sequencers can be programmed in a variety of ways. In some devices there is an individual control for each stage in the sequence. It is then possible to advance the sequence slowly while setting the correct voltages for each stage. In other sequencers, the sequence is read into a digital memory which stores the parameters. Such a memory can be entered with data from any standard data entry device or special interfaces can be made so that it is programmed by a standard musical keyboard.

TRANSLATIONAL DEVICES: A translational device takes some parameter in the real world that can be controlled by a musician (e.g., the output of his own instrument) and converts the output to one or more control signals. This is an information extraction problem. The usual information extracted is pitch, amplitude, and possibly tone color. These procedures are not particularly easy. The main problem is that the actual musical signal is very rich in information (that is why it is interesting to hear) and therefore it is difficult to presuppose what sort of information patterns will be found. Secondly, the information is often a blend. The pitch information and the harmonic information may vary together. As the pitch changes, so does the tone color. In fact, this is the usual case with conventional instruments where for example and instrument has a low range and a high range with a different tone color in each range.

CHAPTER 1c

BASIC MATHEMATICS OF MUSICAL ENGINEERING

CONTENTS:

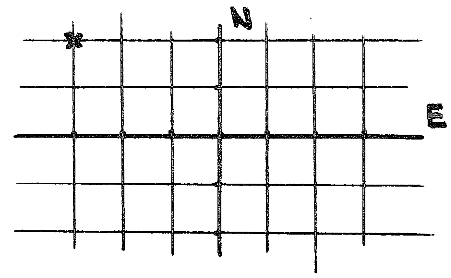
INTRODUCTION

As with all fields of engineering, mathematics plays an essential role in musical engineering. It is not just a tool, but is essential for the description of the devices that are used to produce music electronically. This chapter is not to be considered an introduction to mathematics. It jumps around from topic to topic and thus assumes that the reader is familiar with math at least through algebra. Chapter 1d covers some of the more advanced math that is useful for the electrical engineering aspects of musical engineering. The math given here is mainly application oriented. Each of the topics is chosen because there is a useful application that the musical engineer should be familiar with.

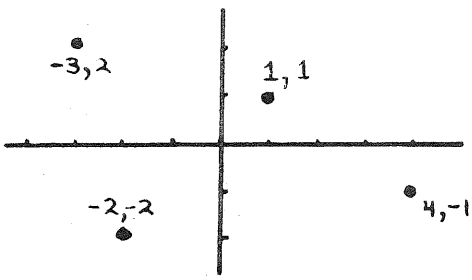
Certain topics such as exponential relationships are essential for the understanding of the nature of musical intervals. This leads to the consideration of the application of exponential circuits. Trigonometric identities are used as a means of examining spectra and as a means of determining the necessary hardware for certain functions. More on these will be found in chapter 2c.

GRAPHS, LINEAR AND NON-LINEAR RELATIONSHIPS

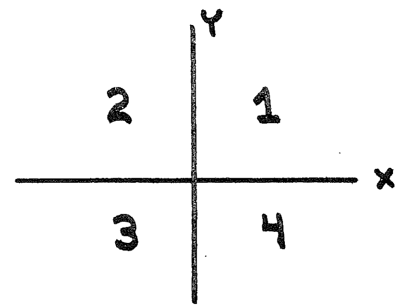
Nearly everyone is familiar with graphs and how to read them. The most familiar graph is the set of X-Y coordinates also called "Cartesian" coordinates In this system, a point on a graph is represented by its X-value and its Y-value. These values may be positive (greater than zero), negative (less than zero), or zero. Graphs are used to represent something that we are interested in that has two (or more at times) numerical aspects to it. Such a quantity is called a "vector" quantity. A familiar example is the directional system within a city. An address might be specified as two blocks North and three blocks West. We could represent this as two blocks North (the Y quantity, thus Y = +2) and three blocks negative East (the X quantity, thus X = -3). Thus we might give an address to a stranger as +2, -3. All this means nothing without a reference however. We would have to know that the direction refers to the corner you are standing on, or perhaps perhaps the center of the city. Thus, for a proper coordinate system there must be two axes (the X and the Y) and an origin (the intersection of the two axes - thus the point 0,0). The X axis thus represents all points where Y=0 and the Y axis represents all points where X=0. The axes divide the graph into four sections called quadrants. At the right we show a typical graph showing the location of the address in the city. Below on the left we show how points (vectors) are plotted on a graph by means of several examples. Finally, below on the right we show how the four quadrants are numbered.



GRAPH OF A CITY ADDRESS

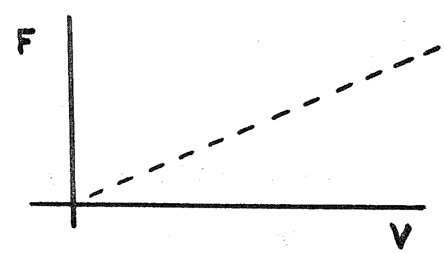


PLOTTING OF POINTS (VECTORS) ON GRAPH

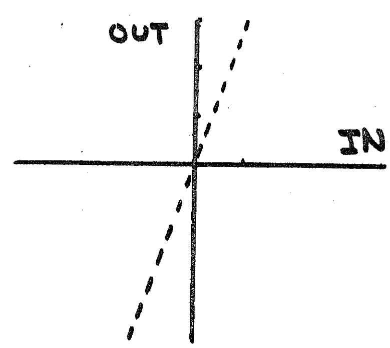


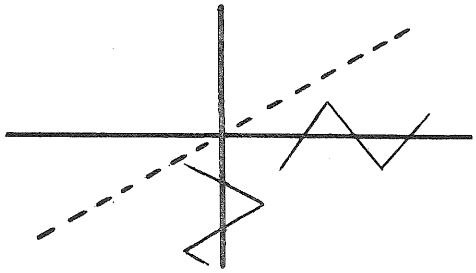
NUMBERING OF QUADRANTS

In electronic music we are often interested in two related quantities such as an input control voltage and an output frequency of a VCO. If we replace X by the control voltage and Y by the frequency, we might find a relationship as shown by the graph at the right. The relationship in this case is a straight line and is referred to as linear. Linearity is often a very desirable property of electrical systems. The linear relationship we have shown operates only in the first quadrant (positive voltages and positive frequency). It is possible to work in any combination of quadrants. For example, we might have an amplifier that amplifies voltage by a factor of 3 and this could work with both positive and negative voltages. We suppose first that all we want is linear amplification. This means that the relationship we need is shown by the graph at the right. We show below how this linear relationship relationship implies that an input voltage waveform (e.g., the triangle) has its waveform faithfully reproduced at the output. If the relationship between input and output is not linear, the waveform will be distorted. The result is non-linear distortion and will manifest itself by adding harmonics to the output that are not present in the input, or will change the relative proportion of the harmonics.

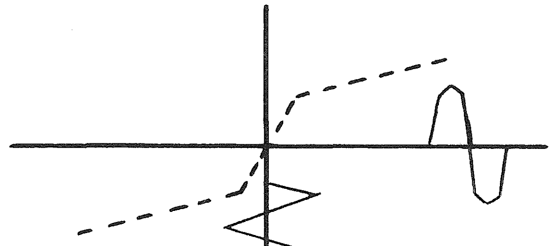


A LINEAR VCO

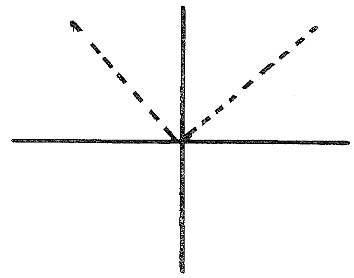




LINEAR AMPLIFIER



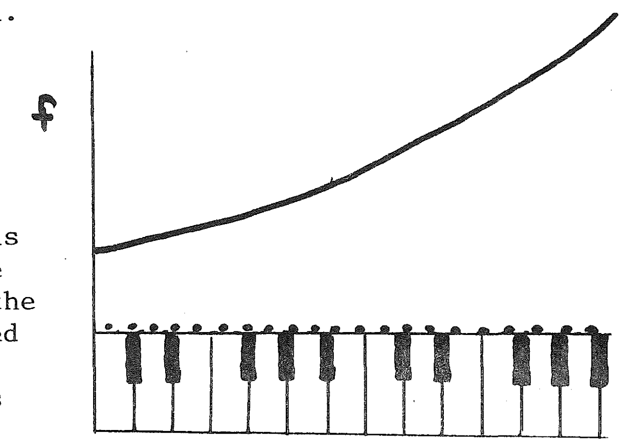
NON-LINEAR AMPLIFIER



ABSOLUTE VALUE AMPLIFIER

One should not get the impression that amplifier curves like the ones above only appear in the first and third quadrants. An inverting amplifier for example would have its curves in the second and fourth quadrants. If the output voltage had an offset, the curve would appear in the first and third quadrants over most of the range, but there would be a small section where the curve misses the origin and passes briefly through either the second or the fourth quadrant. Another circuit shown at the right is the "absolute value amplifier" which has its curves only in the first and second quadrants. This device gives only the numerical value of the input voltage but not the polarity.

While the linear VCO is useful in many applications, for music it is generally not the most useful. This is because in music, it is the interval between two notes that is of interest, not the exact number of Hertz difference. That is, if we are concerned with two notes with frequencies f_1 and f_2, it is the ratio (f_1/f_2) sometimes written f_1:f_2 that is significant, not the difference f_1-f_2. The most fundamental interval is the octave, the ratio 2/1.

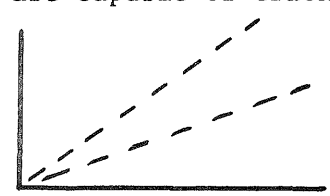


Suppose that we take a keyboard for an electronic musical instrument. Each key (black or white) has associated with it a switch of which one of the terminals comes out the top side. All the terminals are equally spaced as shown at the right. We assign to these terminals the values X = 1, 2, 3, ... and effectively place the keyboard along the X axis as shown. Along the Y axis we then plot the frequency that is expected from the key assuming that a standard "equally tempered" scale is desired by the musician. This results in a curve as shown. The curve is an exponential curve and is a consequence of the nature of traditional music and not something that was dreamed up for electronic music. We shall see that there is an advantage to using an exponential relationship when setting up a keyboard for a voltage-controlled system, but the reader should realize that the exponential relationship comes from the interval nature of music and is not just a matter of convenience for the engineer.

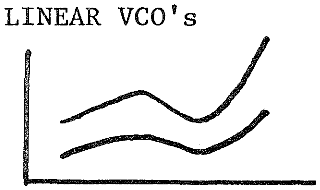
We noted above that the most fundamental interval is the octave, a ratio of 2/1. This implies that the second note depends on the first, every octave is just twice the frequency of the first. In fact, it is the nature of exponential relationships that they come about when the rate of change of the quantity being measured depends on the amount there. A familiar example is the discharging capacitor. The voltage on the capacitor depends on the charge stored in it. If we connect a resistor across the capacitor we know that the current that flows through it is equal to the capacitor voltage divided by the resistance. The current in turn is the rate of change of the stored charge. Thus the rate of change of charge (the current) depends on the charge stored. Suppose the capacitor is charged to 10 volts initially. You connect a resistor R and a current 10/R starts to flow. This current then lowers the voltage on the capacitor. After a certain time you find that the voltage is down to 5 volts. The current is now 5/R, half the original rate of change. After two time intervals, the voltage will be down to 2.5 volts and the rate of change has again halved.

Note that this is exactly the same as the way we would generate octaves given an upper frequency of say 10,000 Hz. We would take the change required for one octave down to be half the amount that is there. That is, we subtract 10,000/2 = 5000 Hz and get the octave below as 5000 Hz. Then we take half the remaining amount and subtract it to get 5000 - 2500 = 2500 Hz, and so on to 1250, 625, 312.5, 156.25, etc. Just as the capacitor never really discharges in theory, we never get down to zero Hertz.

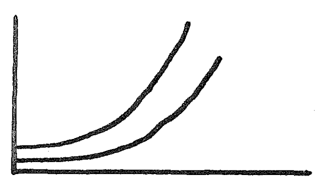
For electronic music it is most useful to have VCO's that are capable of tracking at a fixed ratio (interval). The simplest system that we can try is a pair of linear VCO's. We supply them with the same control voltage and they track up and down as is indicated at the right. However, it is not necessary that the curves be linear (or even anything nice) for them to track at equal intervals. We could consider any curved functions as long as the second VCO when connected to the same control voltage happens to be set for the interval above. This is shown by the second set of curves at the right. These do in fact track at equal intervals although the response to the voltage would be strange. We will of course want to look at another set of "nice" curves - the exponential functions. A set of exponential curves are shown in the third diagram at the right. Note that they have no zero in frequency at zero control voltage, but otherwise they are smooth increasing curves that do track at equal intervals. Thus, we will consider the linear and exponential VCO's in the following tracking experiment.



LINEAR VCO's



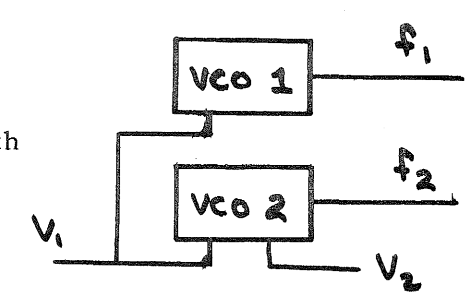
ANY OLD IDENTICAL VCO's



EXPONENTIAL VCO's

The setup is shown at the right. We have two VCO's. The two are supplied one control voltage V_1, and the second VCO receives a second voltage V_2. With $V_2 = 0$, suppose we want to have the VCO's track at an octave ratio starting with the octave 400 Hz/200 Hz for $V_1 = 1$ volt. We can see that the linear VCO's for this job are:

$$f_1 = 200 \cdot V_1 \qquad \text{and} \qquad f_2 = 400 \cdot V_1$$



VCO TEST SETUP

For exponential VCO's we need to put the control voltage in the exponent of the function. We will assume standard 1 volt/octave exponential VCO's which gives us for the exponential case:

$$f_1 = 100 \cdot 2^{V_1} \qquad \text{and} \qquad f_2 = 200 \cdot 2^{V_1}$$

Both these systems meet the requirement of tracking at an octave ratio and passing through the the 400/200 octave for $V_1 = 1$.

The next step in the experiment is to add one volt to the second VCO by changing V_2 from 0 to 1 volt. The output frequencies for the linear VCO's become:

$$f_1 = 200 \cdot V_1 = 200 \qquad \text{and} \qquad f_2 = 400 \cdot (V_1 + V_2) = 400 \cdot 2 = 800$$

And for the exponential VCO's:

$$f_1 = 100 \cdot 2^{V_1} = 200 \qquad \text{and} \qquad f_2 = 200 \cdot 2^{V_1+V_2} = 200 \cdot 2^2 = 800$$

So in both cases the ratio has been changed to two octaves at this point. The final step in the test is to now change V_1 to see if the tracking at two octaves is maintained. We get in the case of the linear VCO's for a 1 volt increase in V_1:

$$f_1 = 200 \cdot V_1 = 200 \cdot 2 = 400 \qquad \text{and} \qquad f_2 = 400 \cdot (V_1 + V_2) = 400 \cdot 3 = 1200$$

Thus the 2 octave tracking is not maintained. The ratio has dropped to an octave and a fifth (a 3/1 ratio). In the exponential case on the other hand:

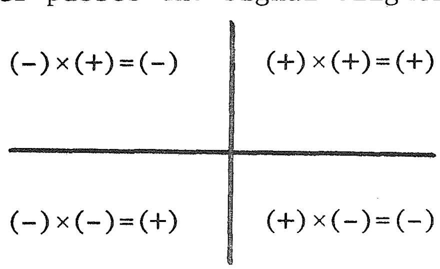
$$f_1 = 100 \cdot 2^{V_1} = 100 \cdot 2^2 = 400 \text{ and } f_2 = 200 \cdot 2^{V_1+V_2} = 200 \cdot 2^3 = 1600$$

so the exponential VCO's track at the new interval.

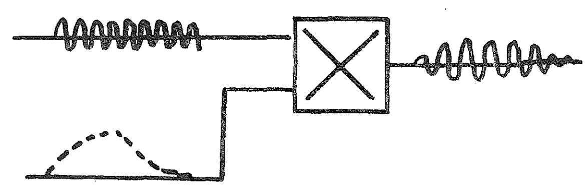
We are thus led to the conclusion that linear VCO's track only when they receive exactly the same control voltage. The interval that exponential VCO's track at can be changed by just changing one of the control voltages. This is one of the major advantages of exponential VCO's.

MULTIPLICATION IN GRAPH QUADRANTS

The X-Y type of graph can be used to understand different types of VCA's. We can first consider X and Y as numbers to be multiplied. Note that the sign of the product depends on the quadrant in which the point X,Y falls. An electrical circuit which will accept bipolar signals on both inputs and give the correct sign for the product is called a four quadrant multiplier. Such a circuit can be used as a VCA by letting the signal to be controlled by X and the envelope control signal by Y. Note in particular that the envelope must go precisely to zero for the off state. If it overshoots slightly the multiplier passes the signal slightly in the inverted sense:

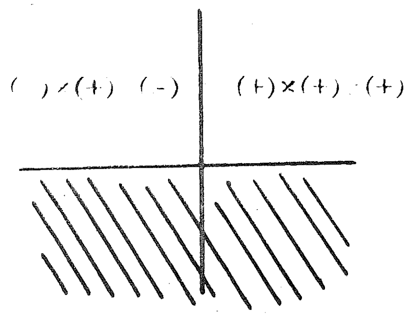


FOUR QUADRANT MULTIPLIER

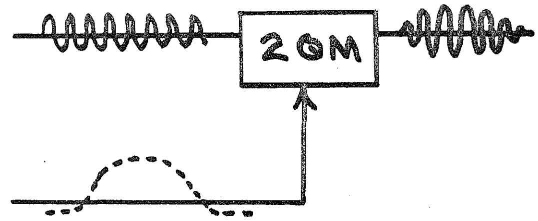


FOUR QUADRANT MULT. AS A VCA

A more restricted type of multiplier is the two quadrant multiplier. This device operates only in the first and second quadrant. X may be bipolar, but Y must be only positive in the sense that any negative signals will be treated the same as zero. We can now see two advantages of the 2 quadrant multiplier as a VCA. First, the zero need not be very precise, and secondly we can even bury the tails of the envelope in the negative Y part of the response so that envelopes will be truncated. This can be very handy when it is realized that for exponential envelopes. the voltage theoretically never goes to zero. Thus by truncating the envelope early a more abrupt and in many cases more natural decay is produced. Below we show the response of a two quadrant multiplier and its application as a VCA.



TWO QUADRANT MULTIPLIER



TWO QUADRANT MULT. AS A VCA

Thus, the 2 quadrant multiplier is in general a better VCA. The four quadrant multiplier is used as a balanced or "ring" modulator. Here is a good example where we can use mathematics to analyze an effect. If we are working with sine waves and if we use a circuit that multiplies voltages, can we use a math equation to tell us what comes out? Yes, and it really works. We look up the trig identity for the multiplication of two sine waves:

$$\mathrm{Sin}(x)\cdot\mathrm{Sin}(y) = (1/2)[\mathrm{Cos}(x-y) - \mathrm{Cos}(x+y)]$$

This tells us that multiplying the sine waves gives cosines that have frequencies that are the sum and difference of the input frequencies. We will have many occasions to use this result.

SCALING OF EXPONENTIALS AND PLACEMENT OF EXPONENTIAL CONVERTERS

We saw above the advantage of an exponential VCO. We used a response that was some constant times 2^V where V is the control voltage. Many readers are familiar with the more usual form of the exponential expression that would be:

$$e^V \quad \text{or } \exp(V) \qquad \text{where } e = 2.718\ldots. \text{ the base of the natural log.}$$

The relationship 2^V is convenient because it means that a change of one volt will double the frequency and hence we have the easy to remember one volt per octave. We are thus faced with the problem of converting a natural expression like Ae^{BV} to one like $C\cdot 2^V$. This is really quite simple, We just take the logs of both expressions.

$$Ae^{BV} = C\cdot 2^V \quad \rightarrow \quad \ell n(A) + BV\,\ell n(e) = \ell n(C) + V\,\ell n(2)$$

We can let $\ell n(A)$ and $\ell n(C)$ be lumped into one constant K, and since $\ell n(e) = 1$, we get:

$$BV = V\,\ell n(2) + K$$

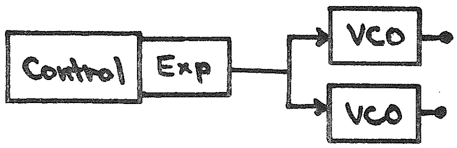
We can then again take exponentials to get: $e^{BV} = e^{\ell n(2)\cdot V}\,e^K$ and set e^K = another constant K'. Thus finally:

$$e^{BV} = K'e^{\ell n(2)\cdot V}$$

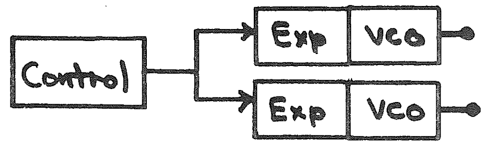
from which it can be seen that the scaling factor is $B = \ell n(2) \simeq 0.69$.

The significance of all this is that to go from any exponential relationship we may have handy (as in the physics of a semiconductor junction) to the one we want, all we need to do is scale the control voltage by a constant factor.

Next we have to inquire about the placement of exponential converters. Having decided to use them between the control voltage inputs and the actual device, we have to know if they belong on the input of one device or the output of the control device. It is simple to demonstrate that they must be placed on the inputs of each device to be controlled. Consider the two setups below:



Exptl. Converter on
Output of Controller



Exptl. Converter on
Each Input.

In the first setup, the two linear VCO's have no idea that the voltage they are receiving is from an exponential converter. Thus, if one of them is offset as in the demonstration on page 1c (5), they do not track anymore. The exponential converters on each input do the job correctly.

The case of one exponential converter can be thought of as a warping of the keyboard. This would be done by placing resistors of different values between the keys and then sending a constant current down the resistor string. Thus there is no real exponential converter except by the selection of resistor values. This may be satisfactory in non-critical applications where the intervals between any tracking VCO's and VCF's are set by panel controls and not be control voltages. On the other hand, it is not overly difficult to design quality exponential converters for the input of each device to be controlled. This gives the advantage that the VCO's and VCF's can be set to intervals by voltage offsetting, and the controller can be linear. In the case of exponential VCO's for example, keyboard string resistors can be identical valued, resulting in high precision without having to come up with a wide variety of precision values.

LOGARITHMIC RELATIONSHIPS

The exponential function can be seen as a sort of expansion. The inverse function is the log function and is a sort of compression. A number of useful relationships are conveniently expressed by log notation.

The "decibel" notation is a scale on which voltage or power rations are expressed in terms of linear increments. The basic decibel equation is:

$$db = 20 \log_{10} (E_1/E_2)$$

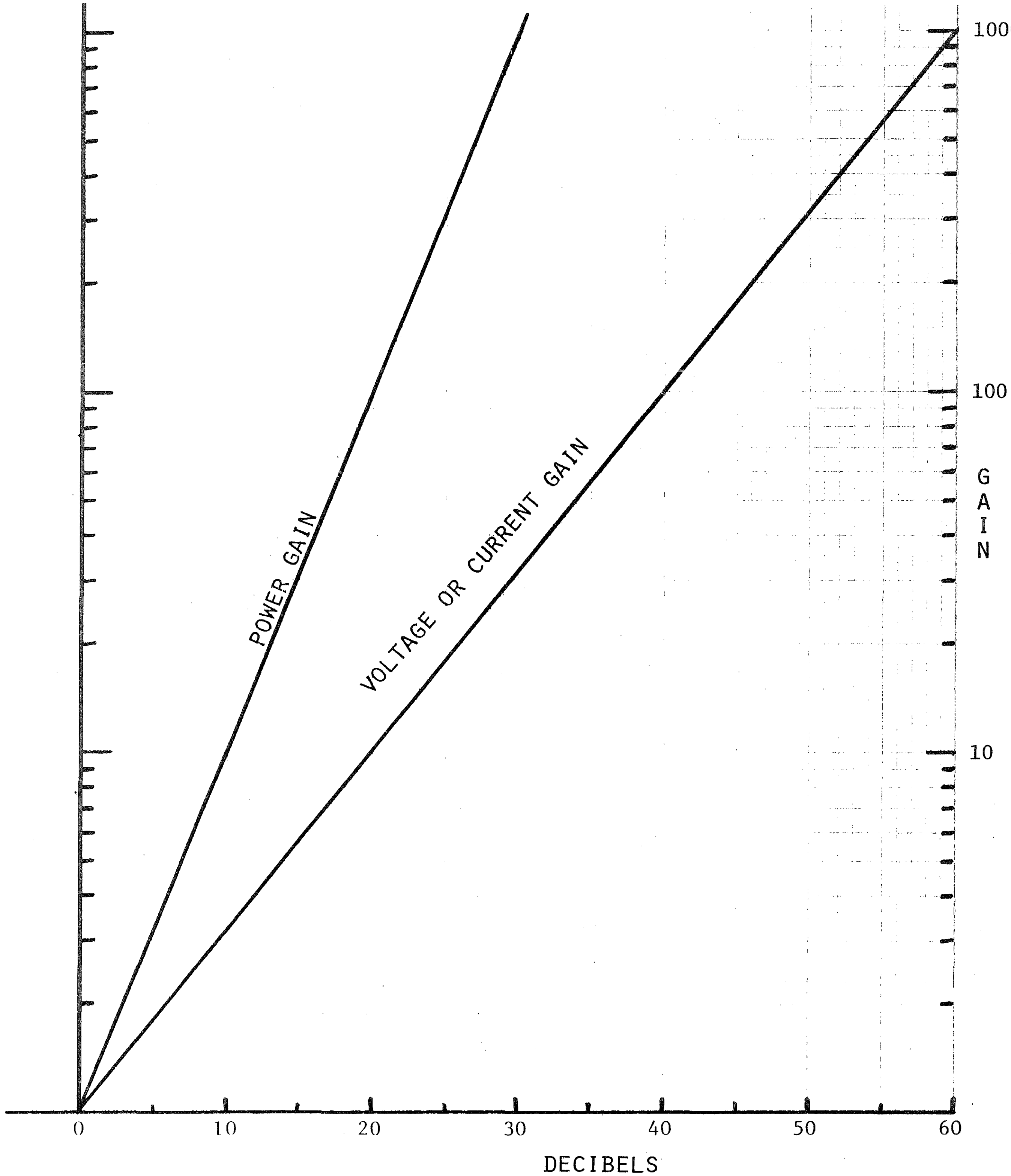
Where E_1 and E_2 are two voltages to be compared. If we compare two power levels, these represent the same as squared voltages so the db equation becomes:

$$db = 10 \log_{10} (P_1/P_2)$$

Common decibel ratios that are often remembered are:

1 db $\sim$ 10%	10 db $\sim$ 3:1
3 db down is about 0.7	20 db = 10:1
6 db = 2:1	40 db = 100:1
	60 db = 1000:1 etc.

From these, a good estimate of other db ratios is possible. For example, 46 db would be 40 db + 6 db (100:1 + 2:1) or a ratio of 200:1. For exact calculations, tables of logs can be used. The graph below should also prove useful:



Various types of log graph paper are available. Semi-log paper can be used for plotting exponential functions. This gives a straight line regardless of the exact exponential. Log-Log paper is useful for plotting frequency response as this allows a wide range of both the frequency and the amplitude to be plotted. In addition, a 6 db/octave rolloff appears as a straight line at 45° across the paper.

Chapter 1d

INTEGRAL METHODS OF ELECTRICAL ENGINEERING

CONTENTS:

INTRODUCTION

One of the principle passions of the electrical engineer is to take one mathematical function, multiply by a second, and integrate the combination over certain limits. The reason for this madness is ostensibly to simplify things or to put them in a more useful form. This is in fact so, even though it takes the engineering student some time to appreciate this and feel comfortable with it. The integrals take on such names as Fourier coefficients, Fourier Transforms, Laplace transforms, convolutions, autocorrelations, and the like. All these names at times tend to obscure the fact that some very similar things are being done under different names.

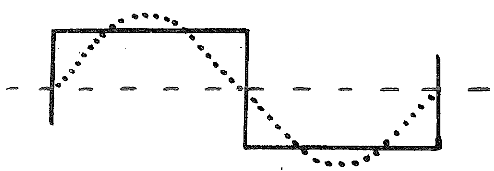
The transforms (Fourier and Laplace) are essentially alternative descriptions between time and frequency, both of which are descriptions used in music. Thus we have a chance to get two views of musical sounds by using transforms.

Here, we have not attempted to give a full description of transform and integral methods. Rather the methods are discussed briefly, often with reference to the same RC circuit example. The reader will get some feeling for these methods as a sort of "algebra" of electrical engineering.

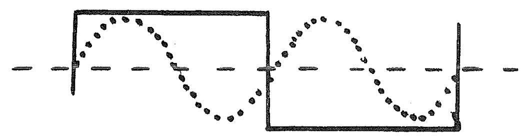
EYEBALLING THE FOURIER SERIES

All periodic waveforms can be broken down into a set of sinewaves called Fourier components. When we say sinewave we are actually referring to any "sinusoidal" waveform that may include cosines, or any phase relative to the actual sine. The mathematical method by which a waveform is analyzed for its Fourier components is called the Fourier Series (FS) expansion. The only sinewaves that appear in the FS are integer multiples of a certain sinewave called the fundamental. The actual FS requires the use of integral calculus, but there is a lot that can be done by just eyeballing the waveform. It is often possible to tell what harmonics are present just by observing the waveform.

For a start, consider the square wave which has somewhat of a clarinet sound. We look at the waveform and compare it with a sinewave of the same period. It is clear that the first harmonic fits the general form if we keep in mind that we will have to do something to sharpen up the corners later. We can next try the second harmonic:

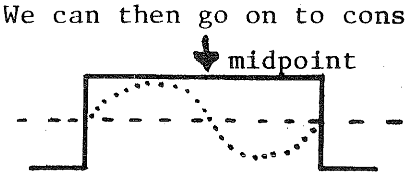


SQUARE AND FIRST HARMONIC

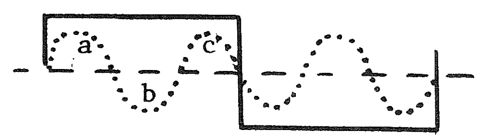


SQUARE AND SECOND HARMONIC

Note that the second harmonic appears exactly the same in the different halves of the square wave. Also, both halves of the square wave are symmetric about their midpoints (unlike the second harmonic). Whatever good the second harmonic might do in the first quarter of the square wave will thus be undone in the second quarter. Thus, the second harmonic is not included. We can then go on to consider the third harmonic:

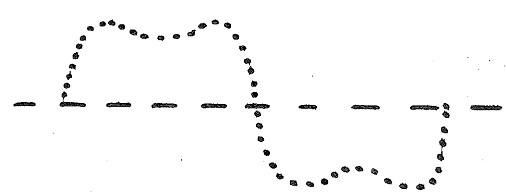


SECOND HARMONIC IN HALF SQUARE WAVE



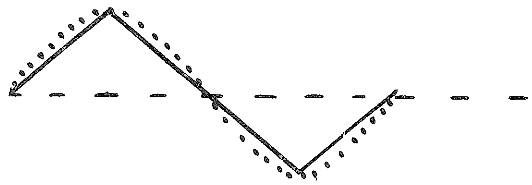
SQUARE WAVE AND THIRD HARMONIC

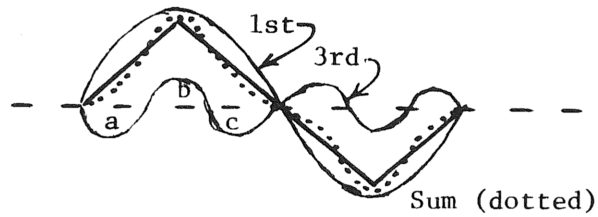
With the third harmonic we can see that the waveform is different in the different halves of the square wave. Furthermore, the bumps a and c will help fill up the corners while the bump b will help to flatten the center. Adding the first and third harmonics will give something like the waveform shown at the right. Note that it starts to take on the general shape of the square wave. If this is followed through it can be seen that the harmonics present are the 1st, 3rd, 5th, 7th, etc., but no even harmonics are included.



FIRST AND THIRD HARMONICS

Next we can consider the triangle wave. The same mode of reasoning that was used for the square will show that only odd harmonics are present. In fact, it can be seen that the first harmonic is already quite a good fit with the triangle. Adding in the third harmonic however starts to make things worse as we start to move in the direction of a square wave. However, it we subtract the third harmonic rather than add it we will improve the situation. This can be seen from the drawings on the top of the next page. The bumps a and c will start to flatten down the rise of the first harmonic that is too fast. The bump b will help add to the point of the triangle in the center. Continuing this reasoning we find that the triangle contains the 1st, 5th, 9th, etc. harmonics added and the 3rd, 7th, 11th, etc. harmonics subtracted. Note also that since the first harmonic and the triangle are such a good fit to start with, we expect far less total harmonic content in the triangle than the square.

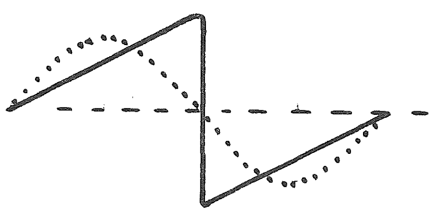




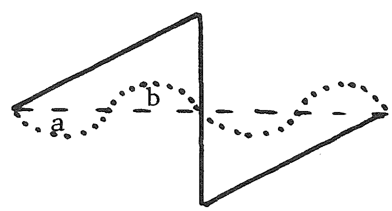
TRIANGLE AND FIRST HARMONIC

TRIANGLE, FIRST, THIRD, AND SUM

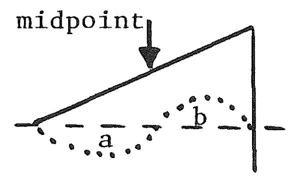
As a third example, consider the sawtooth wave. We can see that the first harmonic starts to fit in as long as we keep in mind we have to do something about the corners. We then look at the second harmonic and find that it is the same in different halves of the sawtooth, but the sawtooth is not symmetric about the midpoint of its halves. This means that the second harmonic will fit in. In fact, if we subtract the second harmonic we can see that the bump a will start to pull down the rise of the first harmonic which is too fast while the bump b will start to sharpen up the corner.



SAWTOOTH AND FIRST HARMONIC



SAWTOOTH AND SECOND HARMONIC



MIDPOINT OF HALF SAWTOOTH

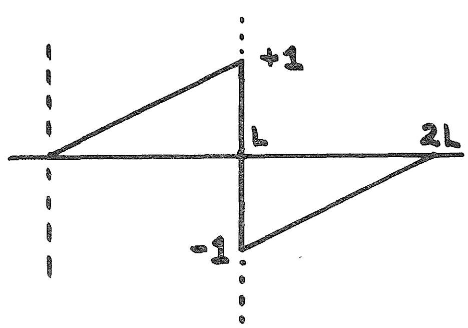
Continuing the process, it is not evident that the third harmonic is included until it is seen that the fourth harmonic may be subtracted so that the two work together to make the sawtooth better. We thus keep all harmonics for the sawtooth, adding the odd ones and subtracting the even ones.

With this eyeball method of Fourier analysis, a lot can be determined about the harmonic content of waveforms. However, the relative proportions in which the harmonics are mixed can not be determined accurately without mathematical methods.

MATHEMATICS OF THE FOURIER SERIES

We saw above that the sawtooth wave was formed from a series of sinewaves added in different proportions. We can represent this series by giving enough terms so that the pattern is evident, or by using the summation sign ($\sum$).

Thus, we can represent a sawtooth as:



$$f(x) = (2/\pi)\mathrm{Sin}(\pi x/L) - (1/\pi)\mathrm{Sin}(2\pi x/L) + (2/3\pi)\mathrm{Sin}(3\pi x/L) - \ldots\ldots$$

$$= 0.636\ \mathrm{Sin}\ (\pi x/L) - 0.318\ \mathrm{Sin}(2\pi x/L) + 0.212\ \mathrm{Sin}(3\pi x/L) - \ldots\ldots$$

$$= \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n}\ \mathrm{Sin}(n\pi x/L)$$

These are three forms that show the harmonic content of the sawtooth wave.

A general periodic function has a Fourier Series of the form:

$$f(x) = a_o/2 + \sum_{n=1}^{\infty} [a_n \,\mathrm{Cos}(n\pi x/L) + b_n \,\mathrm{Sin}(n\pi x/L)]$$

where:

$$a_n = (1/L) \int_0^{2L} f(x)\, \mathrm{Cos}(n\pi x/L)\, dx \qquad \text{and } b_n = (1/L) \int_0^{2L} f(x)\, \mathrm{Sin}(n\pi x/L)\, dx$$

The coefficients a_n and b_n are called the Fourier coefficients and represent the proportions in which the different harmonics are included. Note that each harmonic frequency has a sine term and a cosine term. The two together allow for the realization of an arbitrary phase for the overall harmonic.

For the FS, we have done the standard process: We took a function f(x), multiplied it by a second function Cos(nπx/L), and integrated the combination $\int$......dx over a full period of f(x). This process gives us a measure of the similarity between the two functions inside the integral sign. If the number is zero, the harmonic is missing from f(x). If it is large, the harmonic is present. If it is negative, the harmonic is present but of opposite phase. Thus, the integral can be thought of as a method of finding the similarity between functions in general and as such represents a sort of correlation coefficient.

To see how a coefficient is generated, we must first derive the proper expression for the function. In the case of the sawtooth, this is x/L for the first half (from zero to L) and (x/L)-2 for the second half (from L to 2L). Thus, we can find the first b_1 coefficient by two integrals:

$$b_1 = (1/L) \int_0^{2L} f(x)\, \mathrm{Sin}(\pi x/L)dx = (1/L) \int_0^{L} (x/L)\, \mathrm{Sin}(\pi x/L)dx + (1/L) \int_L^{2L} (x/L-2)\mathrm{Sin}(\pi x/L)dx$$

$$= (1/\pi^2) \int_0^{L} (\pi x/L)\mathrm{Sin}(\pi x/L)(\pi/L)dx + (1/\pi^2) \int_L^{2L} (\pi x/L)\mathrm{Sin}(\pi x/L)(\pi/L)dx$$

$$+ (-2/\pi) \int_L^{2L} \mathrm{Sin}(\pi x/L)(\pi/L)dx$$

Using tables for the integrals:

$$b_1 = 2/\pi$$

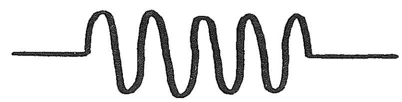
Other Fourier coefficients can be determined in the same way. It is well to be aware that there are several different ways of writing the expressions for the FS. One has to be careful to see how the basic periods are defined. Above we have used a period from 0 to 2L. Sometimes a period from -L to +L is used, or -1 to +1, etc. Another form that is popular is the complex form where instead of using separate a_n and b_n coefficients, the FS is determined from a complex exponential, resulting in a single coefficient c_n and a phase factor.

ANALYSIS OF NON-PERIODIC WAVEFORMS - - FOURIER TRANSFORMS

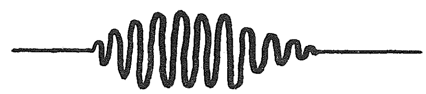
All musical waveforms are non-periodic. A periodic waveform begins at time minus infinity and ends at time equal plus infinity. Quite a concert - but not the sort of music we are used to. Thus we know that the FS is strictly not correct for musical applications and we have to determine where the limits are. These limits are determined by the absolute mathematics of the shorter waveforms, and perhaps more importantly by the limitations due to processing by the human ear.

A couple of limitations appear right away. For example, it was at one time thought that to produce the sound of a trumpet, it was only necessary to determine the overtones in the trumpet waveform and resynthesize. This procedure produces a sound not unlike the sound of a trumpet, but one which has some basic quality more akin to that of any other electronic sound.

A second hint that the FS is not the whole story is provided by the case of a short sinusoidal tone burst, or any tone that is suddenly turned up to full amplitude (as by simply closing a switch). One might produce a waveform of the general form shown at the right. Persons familiar with only a FS would likely argue that the only harmonic present is the frequency of the ungated sinewave. However, what is actually heard is a sharp click as the waveform is gated on. If the waveform has a duration exceeding about 100 ms, the click is followed by what seems to be a pure tone, and by another sharp click as the waveform is gated off. Often the click is blamed on the switch (e.g., the VCA) and is thought to be related to the thump that one hears when the control voltage to a VCA feeds through into the output. In fact, there is nothing wrong with the VCA. The click results from noise that is present mathematically in the gated waveform. The ear can be thought of as averaging over about the last 50 ms of sound it received. The noise in the spectrum results from sharp edges on the waveform, and thus is principally associated with the switching points. Thus the ear hears clicks since it preceives noise during the switching times, but after about 50 ms starts to forget it. Thus for long tones, the ear forgets the edges and hears only a pure tone until another transition occurs.

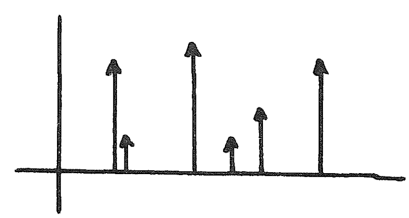


We also know that we can get rid of the clicks by rounding off the corners, causing the waveform to build up gradually and die down gradually as is suggested at the right. This can not really be understood on the basis of the FS. Yet we use the idea every time we apply an envelope to a VCA.

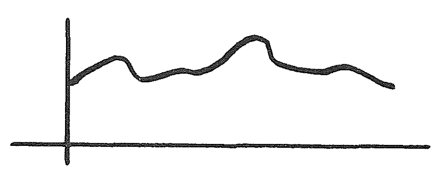


Another limitation of the FS is that we are restricted to a discrete set of sine waves that are all integer multiples of a common fundamental. We can and do build sounds with non-integer multiple sine waves, and these are musically useful if not essential. This is an extension of the principles of additive synthesis. Often times, such spectral components are the result of modulation processes.

At this point it is a good idea to say exactly what we are getting at when we refer to the spectrum of a sound. A periodic waveform that has a FS has what is known as a discrete spectrum. That is, all the frequency components appear in their own positions which are represented by discrete numbers. It is also possible to have a discrete spectrum where the spectral components are not integral multiples of a common fundamental, and which may be in no way related to one another. This could result from the arbitrary mixing of independent sine wave generators. An arbitrary discrete spectrum is shown at the right. A third type of spectrum is the continuous spectrum. The simplest spectrum of this type is the spectrum of white noise. No frequency is favored in this type of noise, so the spectrum is flat. Furthermore, all frequencies are possible so the spectrum is continuous. It is possible to have other types of continuous spectra as well. In fact, it is correct to say that all non-infinite waveforms in fact have a continuous spectrum even though for practical purposes they are sufficiently sharp to be considered discrete.



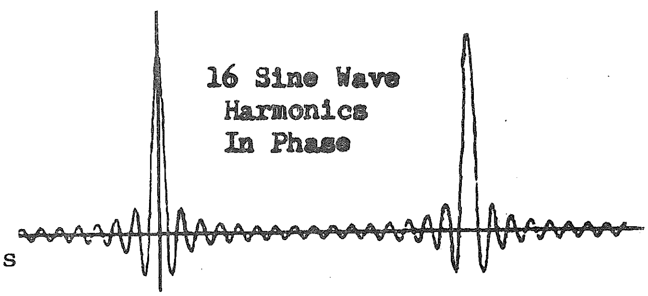
ARBITRARY DISCRETE SPECTRUM



ARBITRARY CONTINUOUS SPECTRUM

We can build up to the idea of a continuous spectrum by first considering the effect of of spectral components that are very close, and then consider the effect of having many spectral components. When two sinewaves of frequencies that are very close are mixed together, the result is known as beating. The two cancel and reinforce each other and thus build up localizations of energy in time. The beating is at a rate that depends on the difference frequency, and the pitch of the result is the average frequency. If we make the frequency difference very small, we can consider only one energy localization over the time span of interest. It is possible to think of any sine wave of variable amplitude to consist of two sine waves one of which has a slightly higher frequency than the actual one and the other a slightly lower frequency. In this sense, any waveform that is not steady state is seen to consist of a number of components and the actual observed waveform is a part of some "beat envelope." Again this makes the point that only a time invariant waveform has a pure spectrum.

Next we want to look at the sharpness of localizations. It would not be unreasonable to expect a musical note to be localized to a time of one second for example, and not reoccur for times on the order of hours. We can get a clue as to what is required by considering what happens when a large number of sine waves are used to build up waveforms. Chamberlin has shown [EN#25 (3)] that a sharp pulse is produced when 16 sinewave rmonics are added. The resulting waveform is shown at the right. What is actually happening is that one point on the waveform is selected and all the sinewaves are put in phase at that one point. A substantial amplitude results at that point, but at other points the sinewaves tend to average to zero. Since the actual waveforms that made this up are not part of a continuous spectrum, the overall energy localization repeats periodically.



Now, by combining the effects of close spacing and a large number of components, we can arrive at localizations that are both sharp and isolated. We can then consider the spectra of a large number of non-periodic waveforms as being composed in this way. The necessary mathematical tool is the Fourier Transform (FT) which can be regarded in some ways to be a continuous FS.

We should point out that the FS is in fact a transform. It transforms the description of a sound in the time domain (the waveform) to a description in the frequency domain (spectrum). The time domain description is in terms of the variable t while the frequency domain description will be in terms of the variable ω. To go from the time domain to the frequency, we want first to look at the complex form of the FS that we mentioned earlier.

$$f(t) = \sum_{n=-\infty}^{\infty} c_n \, e^{jn\omega_o t}$$

where:

$$c_n = (1/2L) \int_{-L}^{+L} f(t) \, e^{jn\omega_o t} \, dt$$

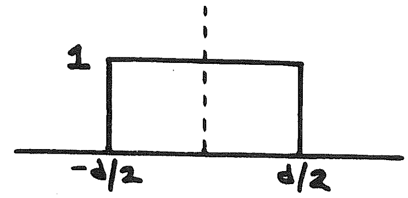
It is then a plausible step to go to the FT where the symbol F[ ] will indicate that the FT of the quantity inside the [] is to be taken. An alternative notation will be F(ω).

$$F[f(t)] = F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} \, dt$$

The inverse transform (analogous to Fourier synthesis form FS results) is possible. The inverse transform is denoted by $F^{-1}[\ ]$ or alternatively as just f(t). The inversion formula is:

$$F^{-1}[F(\omega)] = f(t) = (1/2\pi) \int_{-\infty}^{\infty} F(\omega)\, e^{j\omega t}\, d\omega$$

As an example, we can take the FT of a single pulse. The pulse is shown at the right. Denoting the pulse as p(t) we have:



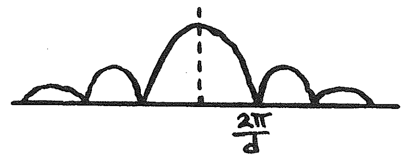
SINGLE PULSE

$$F\,[p(t)] = P(\omega) = \int_{-\infty}^{\infty} p(t)\, e^{-j\omega t}\, dt$$

$$= \int_{-d/2}^{d/2} e^{-j\omega t} dt = \frac{1}{-j\omega}\, e^{-j\omega t} \Big|_{-d/2}^{d/2}$$

$$= \frac{1}{j\omega}\, [e^{j\omega d/2} - e^{-j\omega d/2}]$$

$$= \frac{2}{\omega}\, \mathrm{Sin}(\omega d/2)$$

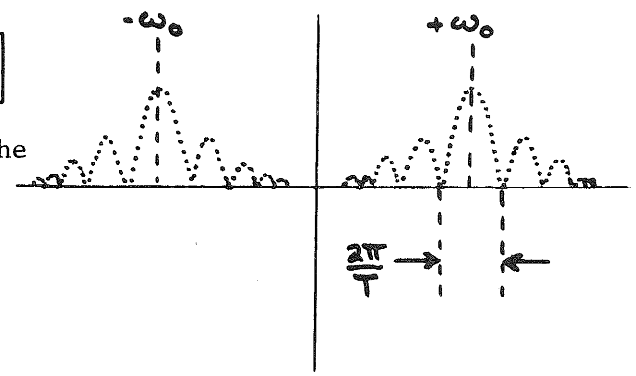


SPECTRUM OF SINGLE PULSE

The result shows that the single pulse has a continuous spectrum. Thus, we can expect an isolated pulse to result in a click.

As a second example, consider the sine wave tone burst. This has been discussed by E. Metzger [JASA 42 #4 (1967)]. For a sine wave Sin $\omega_o t$ which is gated off except when t is between -T and +T, the FT is

$$j\left[\frac{\mathrm{Sin}(\omega + \omega_o)T}{\omega + \omega_o} - \frac{\mathrm{Sin}(\omega - \omega_o)T}{\omega - \omega_o}\right]$$



A diagram of the spectrum magnitude is shown at the right. It is worth noting that the main lobe is of width $(2\pi/T)$. This can be considered as a measure of the frequency uncertainty $\Delta\omega$. The corresponding uncertainty of time is the duration of the burst $2T = \Delta t$. The product $\Delta\omega \cdot \Delta t = 4\pi$. This means that the more you confine the time, the greater the uncertainty in frequency and visa versa. This is a property of any two quantities related by the FT relationship, and is of the same origin as the famous "Uncertainty principle" of physics. For our purposes, what this implies is that short tone bursts may well have poorly defined pitch, or just appear as a click.

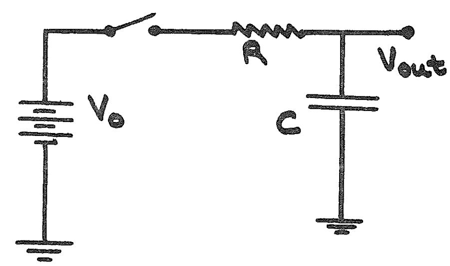
LAPLACE ANALYSIS

A more general form of the Fourier Transform (FT) is the Laplace Transform (LT). The FT related a time description to a frequency description and back through the inverse FT. The LT relates a time description to a complex frequency description and back through the inverse LT. The LT will be denoted by the operator L and the integral description is given by:

$$L[f(t)] = \int_{0}^{\infty} e^{-st} f(t)\, dt$$

where s is the "complex frequency." By complex we mean that it has a real and an imaginary part: $s = \sigma + j\omega$, where ω is the angular frequency. This means that the complex variable s exists mainly on a plain, called appropriately, the s-plane. One of the coordinate axes of the plane is the σ-axis, and the other is the $j\omega$-axis. The price we pay for increased generality is that the inverse LT is not represented by a simple integral formula except where σ=0 and the case reverts to the FT. The derivation of inverse LT's involves a contour integral in the complex plane. As in the case of integrals and many FT's, the practicing engineer usually resorts to tables when it comes time to get a LT or its inverse. They thus become a short-cut tool to solving many problems as we shall see. We are in fact very used to the s-plane since it is used in the analysis of filters. Yet, the LT equations seldom come in at all.

As an example of the use of a LT method, we will consider a simple circuit that is often used as an envelope generator - the RC series circuit shown at the right. We recognize this as a low-pass filter, an integrating network, and as an exponential ramp generator when a step waveform is applied. It is the application of the step waveform that we want to consider here. We will go through two methods of analysis as a comparison demonstration. The first method is the differential equation method. The second method will illustrate the LT method.



METHOD 1, THE DIFFERENTIAL EQUATION APPROACH

For a capacitor, the voltage is related to the charge by Q = CV and the current is the time rate of change of charge: I = dQ/dt. When the step waveform arrives (as when sumeone closes a switch), the voltage across the resistor (IR) plus the voltage on the capacitor (Q/C) must add to the applied voltage V_o. This gives the equation:

$$V_o = IR + V_c = R(dQ/dt) + Q/C$$

This is a differential equation: $dQ/dt + Q/RC = V_o/R$.

The solution to this type of differential equation is found in the following (outwardly fishy) way. You first throw out the forcing term V_o/R and solve for what is called a general solution. Next you find a "particular solution" which fits the forcing function. Then some mathematician somewhere was kind enough to prove that this is all you have to do. The complete solution is the sum of the two - and the final form is arrived at by applying initial conditions which for this problem are that the capacitor is discharged when the step voltage first arrives.

Setting V_o=0 we go for the general soultion (which will be seen to be the solution of the discharging RC network). This leaves:

$$dQ/dt = -\frac{1}{RC}Q$$

The variables can be put on separate sides of the equation:

$$\frac{dQ}{Q} = -\frac{1}{RC}\,dt$$

This is integrated on both sides to give

$$\ell n(Q) = (-1/RC)t + \text{Constant}$$

Taking the exponentials of both sides gives $Q = A\,e^{-t/RC}$ where the "Constant" has been incorporated in the new constant A. Next we look for a particular solution, and it is usually best to guess that this will have the same form as the forcing function, therefore we guess a constant B as the solution. This gives:

$$V_o = R\frac{dB}{dt} + \frac{B}{C} \qquad \text{and since } dB/dt = 0, \qquad B = V_oC$$

Adding the two solutions gives:

$$Q = V_oC + Ae^{-t/RC}$$

Since $Q = 0$ at $t=0$, we can plug these values in to give $0 = V_oC + A$, or $A = -V_oC$.

Thus we get the final answer:

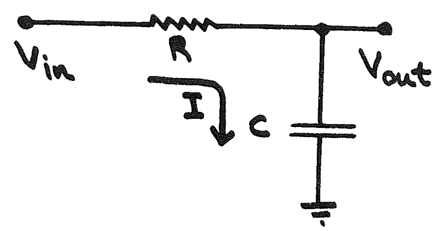
$$Q = V_oC[1 - e^{-t/RC}]$$

Or in terms of the output voltage: $V_{out} = V_o[1 - e^{-t/RC}]$ which is the rising exponential ramp we expected.

METHOD 2, THE LAPLACE TRANSFORM METHOD

The first step in the LT method is to derive the transfer function T(s) just as we would for a filter.

$$I = V_{in}/(R + 1/sC) = \frac{sCV_{in}}{1+sCR}$$

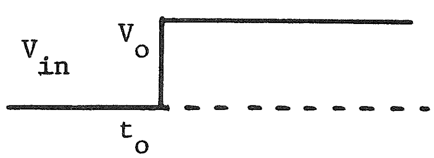


$$V_{out} = V_{in} - IR = V_{in} - \frac{sCRV_{in}}{1+sCR}$$

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

Above we replaced C by 1/sC and treat the capacitor exactly as we would a resistor. This allows for the frequency dependence and phase of the capacitive reactance without resorting to explicit notation.

If we were just working with a filter, we would be testing with sine waves and would be using $s = j\omega$ to determine the frequency response. Here we must represent the input by the LT of the time input. We thus consult tables to get the LT of a step function:



$$L[V_{in}] = \frac{V_o\, e^{-t_os}}{s} = V_o/s \text{ for } t_o=0$$

The Laplace transform of the output is now just the product of ithe LT of the input and the transfer function T(s), which already exists on the s-plane. Thus:

$$L[V_{out}] = \frac{V_o}{s(1 + sCR)}$$

To get the output, we have to take the inverse LT or 1/[s(1+sCR)]. We could probably find this by looking in tables, but it is often the case that it is necessary to reduce the expression to one that can be solved by better known LT's, the ones on the first page of the tables for example. The method used here is the expansion into partial fractions. We assume that numbers A and B exist such that:

$$\frac{1}{s(1 + sCR)} = \frac{A}{s} + \frac{B}{1+sCR} = \frac{A + AsCR + Bs}{s(1+sCR)}$$

Equating like powers of s in the numerators gives A=1, and ACR + B = 0, thus B = -CR. Thus:

$$\frac{1}{s(1 + sCR)} = \frac{1}{s} + \frac{-CR}{1 + sCR}$$

The LT of a sum is the sum of the LT's so we use:

$$L[1/s] = 1 \qquad \text{and} \qquad L[1/(s-a)] = e^{at}$$

This gives V_{out} as:

$$V_{out} = V_o[1 - e^{-t/RC}]$$

which is the same result we got for the differential equation method.

CONVOLUTION: TIME AND FREQUENCY

The integrals used above were used for transforms. They were of the basic form:

$\int f(t)\cdot K(t)\,dt$ where K(t) is an exponential (or sinusoidal) function called the Kernel of the transform, and f(t) is the function that is transformed.

We can also use functions that have approximately equivalent status under the integral sign. One such useful integral is the "Convolution Integral" which is formed below:

$$C(t) = \int_{-\infty}^{\infty} f(x)\, g(t-x)\, dx = f(t)*g(t)$$

where the $(*)$ is used as a short form of notation of the integral.

There are two important theorems concerning convolution that will be stated but not proved here. These basically describe how relationships between functions change when a transformation is made from the time domain to the frequency domain and visa varsa. We denote the FT's of f(t) and g(t) by F(ω) and G(ω) respectively. The two theorems are:

$$f(t)\cdot g(t) = (1/2\pi)\, F^{-1}[F(\omega)*G(\omega)] \qquad \text{FREQUENCY CONVOLUTION THEOREM}$$

and:

$$F(\omega)\cdot G(\omega) = F[f(t)*g(t)] \qquad \text{TIME CONVOLUTION THEOREM}$$

These theorems imply that multiplication in one domain leads to convolution in the other domain.

AUTOCORRELATION - A CONVOLUTION

A special type of convolution is the "autocorrelation" denoted by R:

$$R(\tau) = \int_{-\infty}^{\infty} f(t)\, f(t-\tau)\, dt$$

Thus the autocorrelation is a measure of the similarity of a function with a delayed version of itself. The autocorrelation is a convolution in time, so it is possible to relate this to a multiplication of functions in the frequency domain by using the time convolution theorem:

$$F[f(t)*f(t)] = F[R(\tau)] = F(\omega)\cdot F(\omega) = |F(\omega)|^2$$

The term $|F(\omega)|^2$ can be related to some sort of power spectrum as the spectral components are all squared. This can be related to the time waveform power by means of Parseval's Theorem:

$$(1/2\pi)\int_{-\infty}^{\infty}|F(\omega)|^2 d\omega = \int_{-\infty}^{\infty}|F(t)|^2 dt$$

Clearly if f(t) is some sort of voltage, then $f(t)^2$ is a power. Parseval's theorem then relates the total power integrated over time to the total power in the spectrum integrated over frequency. Thus, we arrive at an important FT pair:

> The Autocorrelation Function and the Power Spectrum form a Fourier Transform Pair.

IMPULSE RESPONSE

We are familiar with impulse response in several cases. If a short impulse is applied to a bandpass filter for example, it will "ring" producing a damped sinusoidal as energy is dissipated. We are also familiar with filter transfer functions:

$$E_{out}(s)/E_{in}(s) = T(s)$$

In the analysis that follows, we will follow tradition and replace T(s) by $H(\omega)$. This is just a change of name when we work with real frequencies. We could use $T(\omega)$ or $T(j\omega)$, but $H(\omega)$ is generally used. In the new notation, a transfer function can appear as:

$$H(\omega) = E_{out}(\omega)/E_{in}(\omega) \qquad \text{or:} \qquad E_{out}(\omega) = E_{in}(\omega)\cdot H(\omega)$$

We can again use the time convolution theorem to handle the multiplication $E_{in}(\omega)\cdot H(\omega)$.

We can next define the following FT's:

$$F^{-1}[E_{out}(\omega)] = e_{out}(t)$$

$$F^{-1}[E_{in}(\omega)] = e_{in}(t)$$

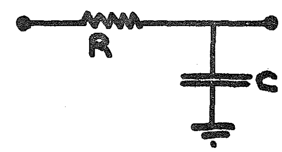
$$F^{-1}[H(\omega)] = h(t)$$

The function h(t) is in fact the impulse response of the system. We can also use the time convolution theorem to infer the following relationship:

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

Thus the time response is determined by convolving the input with the unit impulse response. This is reasonable. It says that the total response is the same is it would be if the input were broken into impulses and the response of the system to each of these impulses were added.

As an example, we will once more attack the RC series circuit excited by a unit step. We determined the transfer function to be:

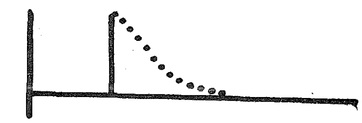


$$T(s) = 1/(1 + sCR)$$

thus $$H(\omega) = H(j\omega) = T(j\omega) = \frac{1}{1 + j\omega CR}$$

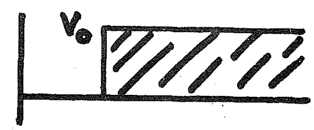
Consulting tables, we can determine h(t) as the inverse FT of $H(\omega)$.

$$h(t) = F^{-1}[H(\omega)] = (1/RC)\ e^{-t/RC}$$



This impulse response is shown at the right:

Now that the inpulse respones has been determined, the response of the RC to a step function is determined by convolving the step function with h(t). The step function is shown at the right. For practical purposes, this just changes the limits on the integral, since the step function is zero below t=0, and a constant V_o above.

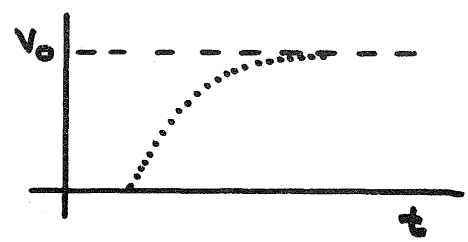


$$e_{out}(t) = \int_{-\infty}^{\infty} e_{in}(x)\ h(t-x)\ dx = \int_{o}^{t} (1/RC)\ V_o e^{-(t-x)/RC} dx$$

$$= V_o\ [e^{-(t-x)/RC}]_o^t$$

$$= V_o\ [1 - e^{-t/RC}]$$

Once, again, this is the same result, indicated at the right:



CHAPTER 1E

TOPICS FROM MUSICAL ACOUSTICS

CONTENTS:

INTRODUCTION

Musical Acoustics is the scientific study of musical sounds. Our interest in musical acoustics stems from our desire to produce viable musical instruments. Even most musicians who are looking for entirely new musical art forms admit to the success of traditional instruments. If we can learn why traditional (acoustic) musical instruments are successful, we will have at least one clue as to how electronic counterparts may be approached.

Many readers of this book will be engineers who play at least one musical instrument and they will be able to think back to how they learned to play. As with many learning processes, they probably were given the rules, learned them, and obeyed them. Since they later became engineers, it seems that at some point they had to know the reason why. Why are there 12 notes per octave? Why does the clarinet have that maze of keys and finger holes? Why does the dominant chord lead to the tonic?

Upon close examination, the engineer is probably next impressed by the precision and refinement of music, and shocked to learn how much is based on tradition - not on sound scientific principles. It then may seem to him that all this tradition has led to musical creatures that may merely be accidents of evolution - but somehow, it all does seem to work.

The next question is probably something like: "Yes, that works, but why not....?" What follows the "why not" may be provocative if not at times profound and useful. But how can the idea be tested? It isn't easy. The necessary musical devices must be prepared and turned over to musicians who learn to use them to realize music. Assuming that this can be done successfully, the experimenters may be up against an esthetic wall - Yes, but is it art?

The electronic technology has added greatly to analysis of traditional musical instruments and practices (including such "exotics" as spectral analysis by computer and acoustical holography). Also, existing synthesis equipment and computers add to the ease with which new ideas can be tested. For example, musical scales with an arbitrary number of notes per octave can be had at the turn of a dial with a voltage-controlled synthesizer. Working the same study with a piano would involve tedious tuning, a process taking at least a few hours. Also, computers can generate new sounds and structures that are too complex or too unfamiliar for musicians to attempt manually.

Most of the musical acoustic studies that have been written up will assure the engineer that there have been plenty of technical people who have been concerned with music and have studied it. He will likely find from these studies that evolution has accomplished what are in many cases the same as engineering compromises. It is interesting that many engineers will discard a tradition (e.g., a 12 tone scale) as sheer nonsense and then try to work it out right. Study will show them that the exact solution they assumed existed does not, and that within certain limits, the 12 tone scale is the best engineering compromise.

In the limited space in this book, we will be able to touch briefly on only a few points of musical acoustics. The reader is advised to consult additional references to gain the necessary insight to apply musical acoustics to musical engineering.

THE ACCURACY OF TRADITIONAL MUSIC

The exact precision of a musical piece and/or individual performance depends on the score, the instruments, and the number of players. We assume here that the performers are accomplished musicians so that ability plays no role in accuracy. We assume first that the only operative analyzer for the music is the human ear-brain, and that the human judges the performance to be correct barring possible difference on esthetic interpretation. We can then ask what happens when we bring in scientific instruments. Just how accurate are the pitches and rhythms relative to the score and relative to other performances? The answer is that in general they can be quite "bad" and still be judged as a good performance, perhaps as a better performance that a more accurate one.

What determines if a pitch is right or wrong? First of all consider the piano. Once tuned, the performer has no control over the exact pitches, so by definition, he is always 100% accurate. A single player in an orchestra also has little if any leeway in pitch once he has tuned to the rest of the orchestra and is drawn along by it. Things start to get more interesting when the player is turned loose as a soloist or even as a soloist with accompaniment, and where his instrument has continuous pitch capability (e.g., a violin of the voice). The control in this case can be influenced by the accompaniment if any, but to a large degree, artistic instincts take over to determine the right pitch. Some studies indicate that the written pitch may be missed by amounts exceeding a whole tone and yet sound right to even experienced listeners. Thus the element of musical context seems to play a role - the performer has a "story" to tell and does it with his own interpretation. If this were not so, surely we would not see much competition for recordings of the same piece by different artists.

Boomslitter & Creel have shown [J. Music Theory 7 #2 (1963)] that even over relatively short passages, opinions on the exact "correct" pitch for the same written pitch can vary due to local references. This sort of result is something for the engineer to consider since often a very strict set of pitches is used, and the ones that would be "correct" are not even available.

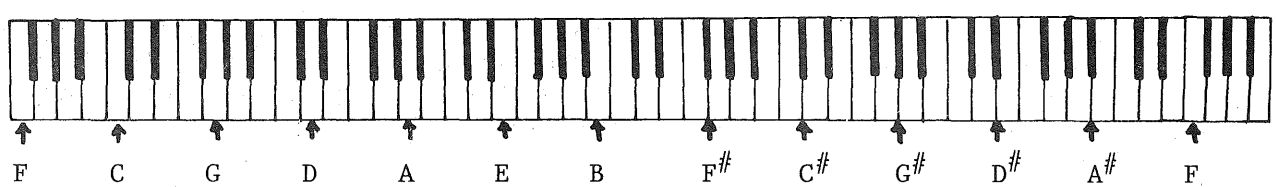
The situation on rhythm is much the same as pitch. Not only does tempo vary greatly, but the actual lengths of the notes in time may vary greatly from their written values. For example a dotted half note followed by a quarter may not be given the expected 3/4 and 1/4 time values. Instead a much larger lowest common denominator is required in many cases, e.g., 45/64 and 19/64. Again, an electronic sequencer will output rhythms with extreme precision (musically wrong?).

Probably the most important clue to an electronic sound for an average listener is the extreme percision of waveforms, stability and timing. This is not to imply that things would be better if things we made sloppy, as they just sound exactly sloppy in this case. Instead, a careful study of the implications of studies of musical accuracy should be made.

MUSICAL SCALES

It is the usual practice to divide the continuous range of pitch into discrete steps (called tones) with octave ratios (2:1) serving as the main markers. The choice of the number of tones per octave is not arbitrary since it is also the usual practice to play more than one pitch at one time, and therefore the subjective effects of various combinations must be considered. The subjective description is generally either consonant or dissonant. Dissonant intervals have long been associated with unpleasant sounds, but this is no longer true as modern music has made them sufficiently familiar to be accepted. Consonant intervals are generally associated with combinations for which the frequency ratio can be expressed by a small integer ratio. Dissonant intervals have high integer ratios or are irrational ratios. Regardless of where the individual sets the dividing line, the most useful scales are those which make possible both consonant and dissonant intervals. Thus, it is necessary to set some notes at small integer ratios.

The simplest non octave ratio is 3:2, the musical fifth. We could thus try to build a scale based on the fifth. We could start on musical F for example and build with fifths above it:



After 12 fifths, we have gone 7 octaves and arrived back at the starting note (F) and have generated a 12 tone scale. However, things are not quite right as 7 octaves is a ratio of $2^7 = 128$, while 12 fifths are $(3/2)^{12} \simeq 129.74632....$, so if we want perfect octaves, the fifths will have to be a little flat. The scale based on fifths was suggested by Pythagoreas (600 BC). It turns out that scales based on other simple ratios offer no real improvement, and the thing never comes out right.

It can be demonstrated that the 12 tone per octave scale can be arrived at on the basis of the best fit to the simplest integer rations. The procedure has been described by van Hoerner ["Universal Music" EN#35]. You first start with a prime number (5 in this case) and take all lower primes (i.e., 3 and 2 in this example). Next you consider all possible ratios of these numbers:

For prime 5:	5/4	4/5	5/3	3/5	5/2	2/5	5/1	1/5
For prime 3:	3/2	2/3	3/1	1/3				
For prime 2:	2/1	1/2						

Next, all ratios that are more than an octave are removed - leaving:

5/4	4/5	5/3	3/5
3/2	2/3		

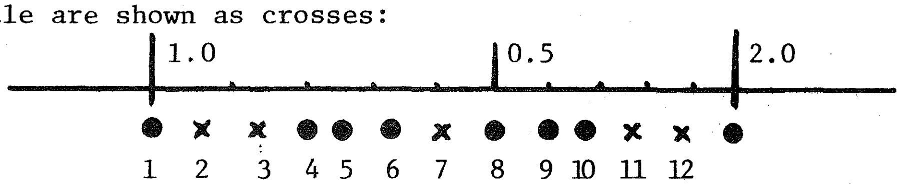
Then those ratios that are less than one are doubled (raised to one octave higher):

5/4	8/5	5/3	6/5
3/2	4/3		

These are all the possible simple integer ratios for prime numbers up through 5. These are all musical intervals named as follows:

Ratio	Name	Closest tone in C Scale
5/3 = 1.67	Major 6th	A
8/5 = 1.60	Minor 6th	G#
3/2 = 1.50	Perfect 5th	G
4/3 = 1.33	Perfect 4th	F
5/4 = 1.25	Major 3rd	E
6/5 = 1.20	Minor 3rd	D#

These are all well known consonant intervals. It is interesting to plot these ratios on a log scale (shown as dots below). Note that the spacing is pretty regular, and gaps in the scale are shown as crosses:



Thus we can have good approximations to all the small integer ratios, and if we fill in the gaps we can have a complete 12-tone scale. The various gaps can be filled by musical intervals from the table listed below. Most of these are relatively small integer ratios, but are large enough to supply dissonant intervals:

Gap Position	Ratio	Name	Comments
2	16/15 = 1.0667	Semitone	dentoed by s
3	10/9 = 1.1111	Minor tone	denoted by m
3	9/8 = 1.1250	Major tome	denoted by M
7	45/32 = 1.4063	Augmented 4th	45/32 = (9/8)·(5/4)
7	64/45 = 1.4222	Diminished 5th	64/45 = (4/3)·(16/15)
11	7/4 = 1.7500	Harmonic minor 7th	
11	16/9 = 1.7777	Grave minor 7th	
12	9/5 = 1.8000	Minor 7th	
12	15/8 = 1.8750	Major 7th	

With the gaps filed, we can choose a complete 12 tone scale. This is called a just tuned scale since all the ratios are integer ratios. We will show how a just major scale can be obtained. This will also serve to show how some of the intervals that were used to fill the gaps were arrived at, since to this point, they were given somewhat arbitrarily. Successive tones of the scale will be obtained by multiplying the lower one by either a semitone (s)-16/15, a minor tome (m)-10/9, or a major tone (M)-9/8.

Just Major Scale	Ratio	Step	Interval
C	1/1		Unison
D	1·(9/8) = 9/8	M	Major tone
E	(9/8)·(10/9)= 5/4	m	Major 3rd
F	(5/4)·(16/15)= 4/3	s	Perfect 4th
G	(4/3)·(9/8) = 3/2	M	Perfect 5th
A	(3/2)·(10/9)= 5/3	m	Major 6th
B	(5/3)·(9/8)= 15/8	M	Major 7th
C	(15/8)·(16/15)= 2	s	Octave

This is a just tuned major scale. It has one major drawback. Once an instrument is tuned for one key, the intervals will be different if it is used in another key. For example, the fifth from C to D is a perfect 5th with a ratio 3/2 = 1.50. If we try to play in the key of D, the fifth from D to A would have a ratio 40/27 = 1.481... and this is not a perfect fifth. The reason it is necessary to play in more than one key is that much of western music "modulates" from one key to another as the piece progresses. The compromise that has been arrived at is called the "equal tempered" tuning method where all intervals are spaced by an integer number of half steps. Thus all scales are equally good in any key, and by the same token, equally bad where errors occur.

When twelve tones are spaced so that the intervals are all equal, they are spaced at the 12th root of 2 apart. The half step is thus a ratio of 1:1.059... and the full scale becomes:

Equal Tempered Scale (chromatic version)	Ratio	Interval Name	Just Interval
*C	1.0000	Unison	1.0000
C#	1.0595	Half step	1.0667
*D	1.1225	Whole step	1.1250
D#	1.1892	Minor third	1.2000
*E	1.2599	Major third	1.2500
*F	1.3348	Fourth	1.3333
F#	1.4142	Diminished 5th	1.4063 or 1.4222
*G	1.4983	Fifth	1.5000
G#	1.5874	Minor 6th	1.6000
*A	1.6818	Sixth	1.6667
A#	1.7818	Minor 7th	1.8000
*B	1.8877	Major 7th	1.8750
*C	2.0000	Octave	2.0000

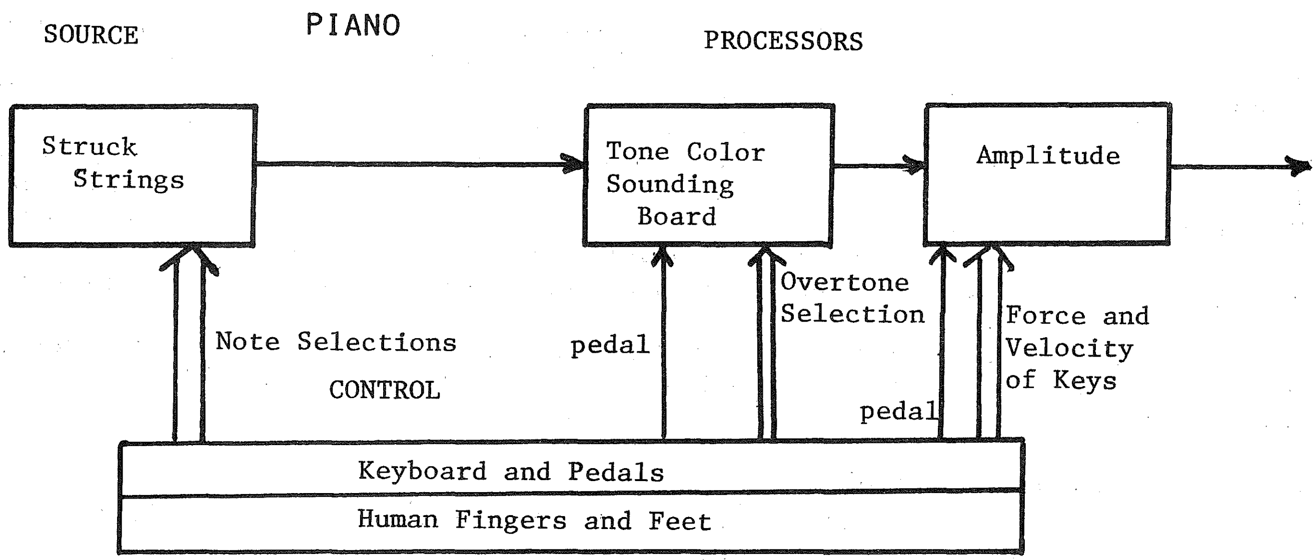
*Tones for equal tempered major scale

The 12-tone equal tempered scale is a natural scale for electronic music systems since a set of equal valued resistors can be used in a voltage divider to supply all the voltages required for an exponential VCO. A one volt/octave VCO for example gets voltages that are 1/12 volt apart. It is possible to obtain many other equal tempered scales as well by scaling the VCO response up or down. Scales of 19, 31, and 52 tones are of interest because they provide tones that are good approximations to many of the just intervals.

BLOCK ANALYSIS OF TRADITIONAL INSTRUMENTS

Below we will take a look at some traditional musical instruments in a block diagram type of approach similar to that we used in chapter 1a for electronic music systems. In general, these instruments consist of an excitation source, and some means of tone coloration - usually connected with a sound radiation mechanism. Traditional instruments have been designed mainly to produce melodic and harmonic structures. Thus they are oriented toward accurate pitch control. Dynamic control is usually fairly well advanced. It is in the area of tone color control and variation that traditional instruments may be found most wanting. None the less, the reader should bear in mind that there is a certain quality of traditional musical instruments that is often referred to as "realism" that is still somewhat of a mystery to the electronic instrument designer.

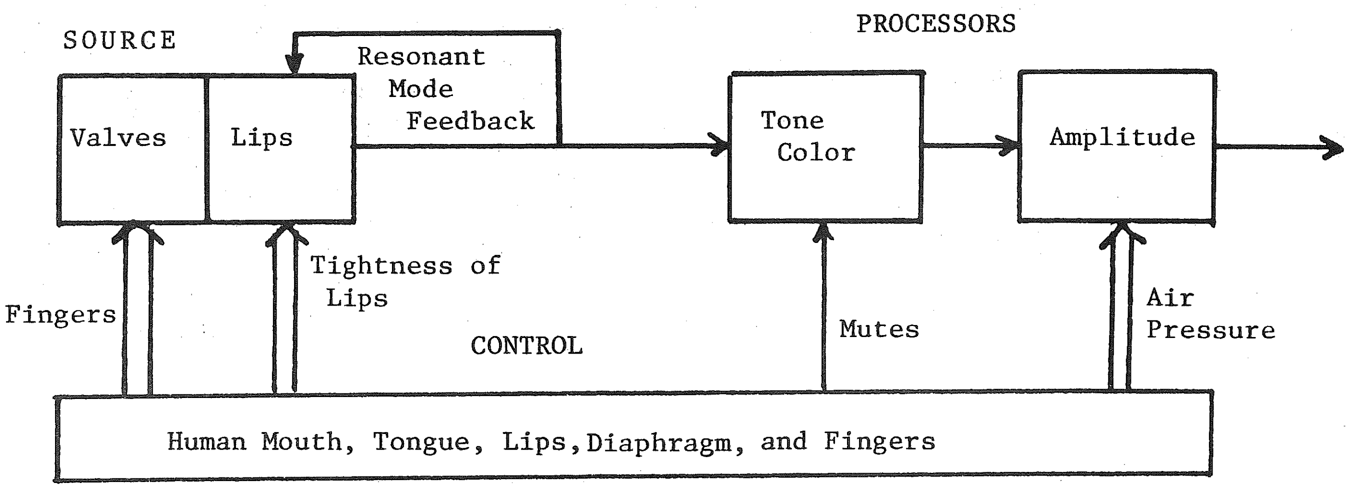
THE PIANO: A block diagram of a piano is shown below:



The excitation source results from the striking of the piano's strings. The performer has rather complete control over pitch and note selection through the use of the piano's keys. On the processing end, the performer has very good control over the amplitude (since the real name of the piano is "pianoforte",soft-loud). Additional amplitude control is provided by the pedals. However, the shape of the amplitude envelope is not variable - the user must use a sharp attack followed by an exponential-like decay. Slight changes of tone color are possible by using the pedals, but the main mechanism of tone color control is through the placement of overtones by means of multiple note selection. For the most part, the piano literature consists of works that exploit the polyphonic ability of the piano. A few modern composers have however used the piano to produce different tone colors (the proverbial "prepared piano") by adding things to the inside, or by striking the strings from the inside with various objects.

THE TRUMPET: The trumpet is quite similar to other brass instruments, and somewhat similar to all wind instruments. The block diagram of the trumpet is shown on the top of the next page. The excitation source consists of vibrating lips. The pitch is controlled by the tightness of the lips. However, as is indicated, there is a feedback mechanism that forces the lips to vibrate only in certain resonant modes. Additional vibrational modes are possible by pressing different valves which change the effective length of the trumpet tubing. A resonant condition is set up when a standing wave is set up in the tube such that there is an antinode at the lips. This antinode changes the

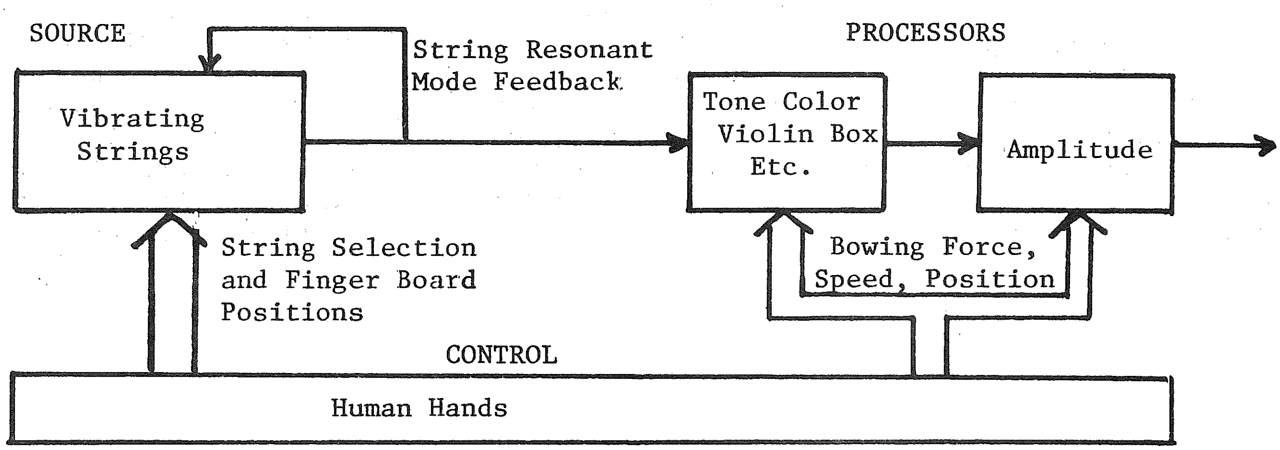
TRUMPET



pressure against the lips and allows them to open and close at a rate corresponding to the pitch. The player has reasonable control over the amplitude by controlling air pressure. However, only weak control over tone color is possible, mainly as a secondary effect of amplitude, or by the use of devices such as mutes. The main new feature in this block analysis is the feedback loop that forces the instrument to follow certain resonant modes. The instrument is monophonic.

<u>THE VIOLIN</u>: The block diagram of a violin, a typical stringed instrument is shown below:

VIOLIN



The excitation source is the vibrating string which can be driven by either a bowed action or a plucked action. Pitch control is achieved by placement of a finger on the string and pressing it to the fingerboard, effectively changing the length of the string. As with the wind instruments, there is a feedback loop that forces the string to vibrate only in certain modes at certain pitches. It takes a certain length of time for this feedback loop to be established and thus certain features appear in the attack phase that are not present in the steady state. It is sometimes supposed that a string is bowed by a stick-and-slip method in which the string moaves with the bow according to the coefficient of static friction. When the bow moves far enough the tension in the string causes it to snap free and back to a position determined by the coefficient of sliding friction. If this were true, the pitch would be determined by the velocity of the bow, not be the length of the string. What actually happens is that the vibrating string essentially kicks itself free of the traveling bow, much as the pressure antinode at the mouthpiece of the trumpet frees the sealed lips. This means that the range of bowing speed and pressure can vary through a substantial range as long as the two together are consistant within certain limits with the natural vibration of the string. In addition, the string can be bowed at different points along its length, and this adds additional possibilities for different sounds. Formant structure is added to the output by the various resonances of the violin box, it various component parts, and the air inside it. The upper frequency response curves may be very complex. Unlike the piano and the trumpet, the player can achieve a relatively wide range of tone color and amplitude due to the range of bowing parameters. In addition, some modern composers have used the violin in unique ways; for example, the fingers may be drawn along the strings to produce an unusual effect.

HEARING

There is really no way in this limited space to discuss much of what is known about human hearing and the problems that remain to be investigated. We give here only the briefest of overviews.

Somewhere between musical sounds that exist in the air and the final esthetic impression that exists in the human mind lies the hearing mechanism. It consists of at least the two ears, part of the nervous system, and part of the brain. Often we refer to all this as "the ear" as in the phrase: "the ear will not detect....."

The physical ear can be modeled as a sound receptor and detection mechanism. The detection mechanism is to be considered in at least some ways as a spectrum analyzer, and the time constant of the detection mechanism is something like 50 milliseconds.

The ear has a wide dynamic range for loudness (in excess of 100 db) and pitches of from 15Hz to about 20,000 Hz can be detected. The ear is much more sensitive to pitch variations (~0.1%) than it is to amplitude variations (~10%). Pitch detection is not based entirely on spectral information - neural processing is somehow involved in some form of pattern analysis. The ear identifies tone color partly on the basis of spectral analysis, but the actual identification of an instrument depends to a large degree on the transients in the sound, and not just the tone color. Surprisingly, the ear is relatively insensitive to the phase relationship of the components in a complex waveform.

CHAPTER 1F

BASIC PRINCIPLES OF MUSICAL ENGINEERING

CONTENTS:

INTRODUCTION

When the outline for this handbook was being prepared, it seemed appropriate to have a chapter on the basic principles of musical engineering. It seemed at that time that the exact formulation of these basic principles and their proper placement in this book would become clear as material was organized.

Actually, neither the principles or their proper placement in the handbook is crystal clear. What is clear is that (1) The principles cannot be presented easily without examples. (2) Many of the principles are only now starting to surface as methods are tried and evaluated, and (3) The setting up of the principles of musical engineering is greatly complicated by musical traditions, subjective evaluations, and esthetics.

It is perhaps these complications that make musical engineering relatively unique. While many enjoy the pure objectivity of other fields of engineering, with musical engineering we get to piece together knowns and unknowns, sure things and speculation, and ultimately use technology for new art forms.

A formal presentation of the principles of musical engineering is not possible at this time. In the actual presentaion below, we will have to settle for a few not quite random comments on how the principles seem to be shaping up. Many of these are based on personal experiences, lessons learned the hard way, and other such informal but effective methods.

BASIC GOALS AND PROCEDURES

THE END PRODUCT OF MUSICAL ENGINEERING IS MUSIC, NOT MUSIC SYNTHESIS SYSTEMS:

We must remember that the final end product of musical engineering efforts is actual music, not the music synthesis system. The most important point in this regard is that it really is the case that it is not the equipment, but the way it is used that is important. Some types of music do require more elaborate and more accurate systems than other types of music, but a $5000 synthesizer is no guarantee of good results. In fact, a complex and expensive system in the hands of the beginner may make it difficult for him to make any sound at all, let alone music. Many excellent pieces of music have been produced by the simplest types of equipment because the user could control it properly and used it in an artful manner.

In the same way, an excellent musical piece should not be credited to the synthesis system that produced it. The artist who produced the work may if he wishes acknowledge the utility of the system he used, and point out features that perhaps were essential for the realization of his music, but the synthesis system itself deserves no direct credit for the music it produces. Robert Moog was once giving an elementary talk on synthesizers to a group of physicists and played a tape of one of Walter Carlos' transcriptions of Bach. At the conclusion of the example, some of the audience began to applaud. This was uncomfortable as others wondered if Carlos or perhaps Bach would appear to take the bow. The misconception of those who were clapping was then made clear as someone shouted "Oscillator - Oscillator" in the manner in which a composer might be called to take a bow. As laughter died down, Moog shouted back "the next time I hear a string quartet I'm going to stand up and shout, 'Cat - Cat'." Unfortunately, the connection, let alone the point, seemed to be missed by most of the audience.

The production of music is a job for a musician, not an engineer. The necessity for musical form is something an engineer can appreciate, but it is the musician that must realize it. This is not to say that an engineer cannot produce music - just that if he does, it is the musician in him that does so, not the engineer.

THE MUSICAL ENGINEER

Musical engineers may come to the center ground from either the musical side or the engineering side. When a person decides he needs to work in musical engineering, it is necessary to evaluate his position to determine what he needs most. Obviously, the answer to this question depends on his background, but two general observations can be made by observing students who take up courses in electronic music. The musicians in the group seem to be facinated by the sonics of elementary synthesis techniques and produce musical pieces with some form, but with basically inferior technique as far as the exploitation of the sound synthesis potential is concerned. Engineering types on the other hand often produce sensational sonic effects with no sense of musical form. The musician needs to gain familiarity with the electronics so that he will feel free enough with it to explore new patches and methods. The engineer on the other hand must learn to stop experimenting and give some attention to the way sonics can be used to form musical pieces.

THE UTILITY OF GENERAL PROCEDURES

It is often the case in musical engineering that a general procedure is successful or unsuccessful due to its general nature and independent of the exact details of the procedure. In subtractive synthesis, the sweep of a VCF against a rich harmonic waveform is successful. The procedure is successful independent of the type of filter, the way it is swept, and the harmonic waveform against which it is swept. Procedures of additive synthesis often work with either sine waves or Walsh functions, the important thing being the general nature of the structures that can be produced. By the same token, a procedure that is tried and found to fail often cannot be patched up by some minor adjustment. This is perhaps because a successful musical sound must "entertain" the ear to a certain extent and whether or not it does this depends to a great degree on the basic structure of the system.

THE PRODUCTION OF SIMILAR SOUNDS

It often turns out that there are a number of ways of achieving the same basic result. For example, non-harmonic sounds can be produced by various methods of modulation. When this modulation gets heavy enough, it may be found that different modulation methods produce essentially similar sounds. This is important for two reasons. First it tells us that there is something essentially the same about the results of the methods. In this case, it is the complex structure of sidebands produced. The second point is that the designer is then faced with the decision of including or excluding a method from his system if the sounds produced by the method are similar to those that already exist.

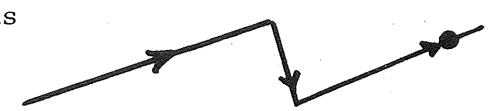
SYSTEM DESIGN PHILOSOPHY

ACCURACY AND STABILITY REQUIREMENTS

Accuracy and stability requirements depend on context. As previously mentioned in chapter 1e on musical acoustics, many traditional instruments are technically inaccurate although musically accurate.

Two examples from electronic music will serve to demonstrate the relation of accuracy to context. In chapter 5b on VCO design we will be discussing the design of accurate and stable exponential current sources. It turns out that once the VCO is scaled properly in its midrange, careful attention to accuracy is required only when there are two or more VCO's that have to track, or if the ends of the range have to be extended. Stability is required where there is a time reference. Once the first order temperature correction is made, drift can be detected only with a reference, when two VCO's drift apart, or when a VCO drifts with respect to a prerecorded section.

As a second example, consider the use of a sawtooth wave. For general electronic work, an engineer expects a sawtooth wave to consist of a linear ramp and have a very rapid reset to the base line. In musical engineering, there is generally no context for a sawtooth wave or any other waveform except possibly the sine wave. The closest thing to a sawtooth that appears naturally is probably the motion of a bowed string where the string moves with the bow and then snaps free and back down the bow. The string cannot snap back instantaneously, even relative to bow speed. Thus the fall is more like the waveform at the right. This waveform is rich in harmonics, although a few upper harmonics may be missing or weak as compared to the sawtooth. The point above is not to argue for a sawtooth with slow reset time, but to simply point out that for system design there is generally no musical reason to prefer one waveform over another. Thus, the engineer can do whatever is easiest and results in the best pitch accuracy and stability. If other engineering and musical factors are accounted for, there is no reason to insist on perfectly engineered waveshapes. The musician will likely fine there is enough variety in a standard set of waveforms like the saw, triangle, square, pulse, and sine. A programmable point-by-point waveform generator where each point on the waveform is set by a pot and some method is employed to smoothen between points is not to useful (although it may be very useful as a programmable envelope generator or a sequencer). The problem with the programmable point-by-point generator is that there is no complete and easy way to determine how the different levels in the waveform (in time) will effect the output sound that is perceived



mainly in terms of spectral content of the waveform. Furthermore, the programmable waveform generator is a lot of work to go to for what will probably be minimal improvement - the musician may find a standard waveform just as satisfactory. It should never be supposed that a programmable waveform generator will make possible all musical sounds unless some means of user control is provided that is fast enough and provides some sort of translation from the spectral description to the time domain description. [For a discussion of how various programmable waveform generators should be approached, see H. Chamberlin, "Comparison of Methods for Producing Arbitrary Waveforms and Envelopes," EN#23 (2)].

Probably the most useful way of categorizing waveforms is in terms of their total harmonic content (%) and their spectral density (all harmonics, missing even harmonics, etc.). In this scheme, the table below can be prepared:

% ↓ / Density →	Low	High
None	Sine	----
Low	Triangle	----
High	Square	Saw, Pulse

THE LACK OF LIMITS ON USEFUL RANGE

The musical engineer should not be fooled into thinking that he has only to deal with the audio range from 15 Hz to 20KHz. Lower frequency signals are of course needed for control signals. The question often comes up as to the actual range that is needed for any device used in electronic music. The answer is fairly simple - the range should be as wide as possible without exceeding the limits of direct or indirect perception by the ear.

Pitches can be perceived from 15 Hz to 20 KHz, so this is one range to consider for VCO's. However, there is no lower limits because the same VCO can be used for slow control signals. Higher limits on the upper side are also used sometimes as various forms of frequency division are often employed. This implies that the VCO range really cannot be made too wide. This means that exponential VCO's are not just a matter of convenience due to the interval nature of music, but for practical control, wide range is possible only with an exponential current source.

Wide range in amplitude can be achieved with transconductance control built into VCA's. This range is fortunately comparable to the full dynamic range of the ear. A wide range of tone color is largly a matter of making available all standard synthesis techniques and at the same time leaving the system flexible so that new methods can be tried.

Range should be expressly demonstrated to musicians and frequently reviewed. It is important that the musician realize that the devices do have wide range and that different "effects" result in different parts of the range due to the speed of change relative to time constants of the ear.

ALL COMPLEXITY OF A SYSTEM MUST BE MUSICALLY CONTROLLABLE

It is easy to design sound synthesis schemes with the possibility of extreme complexity. Digital computer point-by-point synthesis is a good example. It is also necessary, however, to provide control over the complexity of the system or it is all useless. De facto control is the minimum requirement - in most cases the control must be provided of a well defined, user oriented nature. If control of complexity is available, but not in a user oriented form, the excess complexity will be wasted and may even hinder the user. The digital point-by-point method demonstrates uncontrolled complexity. Complete control over extreme complexity is implied by the vary nature of the method, but it is not in a user oriented form. There is no translation from the sound in the composers mind to the point-by-point waveform that corresponds to it. To make the computer work at all, some form of program must be used. All possible complexity not covered by the program is wasted. What most of today's music programs give is roughly the equivalent of an analog synthesizer with a few advantages, but with the disadvantages of added expense (initial and continual operating) and inability to work in real time. An analog synthesizer on the other hand requires a certain amount of experimentation to get the desired sounds, but working time is relatively inexpensive and is in real time. The computer control of point-by-point synthesis is better suited for research situations where new structures and control means can be similated without having to build hardware.

As a basic guide to the degree of complexity that is needed, available, and controllable, the musical engineer can compare with examples and models. Useful material to help with this determination can be found by studying traditional instruments, successful electronic instruments, information theory, and psychoacoustics.

COMMUNICATIONS BETWEEN MUSICIANS AND ENGINEERS

It is essential that musicians and engineers learn to communicate. It is therefore desirable that they learn a good deal about each other's fields. Beyond this, communication requires the ability to ask the right question in the right way.

WHAT DOES THE MUSICIAN REALLY NEED?

Only the musician can answer this. But the engineer must ask the question in terms more specific than the question in the above title. For example, considering the range of the VCO, he might ask if the musician needs a range below 15 Hz. Since this can't be heard directly, the musician might assume he doesn't need it. Unless told, the musician may not realize that it is possible to use one oscillator to control another. Why should he know? Once shown that lower frequencies are useful as controls, he will be in a better position to know if he wants them. The engineer must continue to probe - "Would you use ....?", "Do you need....?", etc, and make sure the musician understands the implications of the devices being discussed. On the other hand, the engineer must be able to do at least a "first order" musical evaluation himself so that too much time is not wasted going over ground which the musician feels has already been covered or is of no musical importance.

Two schools of thought appear at this point: One says that things should be kept as general as possible. A second says that since there are usually several ways of doing the same general sort of thing, why make available more than one? This is really an individual decision that has to be made. In some cases the musician is only confused and hindered by multiple methods - for some types of music, only a small bag of tricks is required. Other musicians will need the elbow room of complete generality as a matter of psychological preparation for their work. The engineer must take considerable care to see that the musician understands the implications of such decisions on system operation.

Particular care must be taken to indicate that if a synthesizer is not prepatched, all but the very smallest have a potential for unique patches that may never have been tried by anyone. It is very easy for the musician to assume that there are certain ways of putting the system together, and that if other things were possible and useful, the engineer would tell him about them. The engineer should therefore stress the fact that there are many untried patches possible, and that control setting may further expand the potential. The musician must be told he is free to experiment.

THE MUSICAL ENGINEER MUST OFTEN DEMONSTRATE THE SONICS

Many composers conceive pieces in terms of sounds they know. For them, it is apparently less frustrating to work with known sounds than to have a sound in their head and no way of realizing it. There is nothing wrong with this and the engineer should not try to interfere once the work has been conceived. I once worked with a composer who had a section of the score marked "sawtooth wave". I knew of course that he really meant something like a sawtooth and would not want to work with something so simple. He wanted the sawtooth unaltered, and musically that is what he had to use. He knew the sound and wanted it that way.

When working with this type of composer, it is necessary to put the equipment through its paces as a demonstration. If this type of composer does experiment, his experimentation may tend to follow certain lines. He will learn new sounds from other musicians and engineers. It may be the case that his own inititive may have been stifled by unfamiliarity with the equipment or fear of doing the "wrong thing" when such wrong thing probably do not exist. Such composers will often ask the engineer how to make a certain type of sound and play a passage from some recorded work produced by someone else. In order to answer the question, the engineer himself must have done a lot of experimenting and careful listening. I once was asked how to make one type of sound and spent about an hour trying to achieve it with a VCF. It was simple pulse-width modulation which I did not recognize until some time later.

It is easy for the engineer to slip into a state where things are so familiar that the fundamental principle is forgotten - just as a native English speaker may use perfect grammer and yet retain no knowledge at all of the rules of grammer. Scott Wedge tells of the time he was demonstrating a synthesizer and was speaking of connecting modules with patch cords. Sometime during the discussion, confusion was apparent and he was struck with the realization that it was because modules have inputs and outputs and that inputs are different from outputs. After conveying this idea, the listeners were able to conceive of the idea of signal flow from one point to another, and understand what Scott was telling them. This means that the engineer must be clear in his own mind about what the fundamental principles are, and start at the beginning, and generally it would be a good idea to determine approximately what the musicians knowledge of electronics is. This can be done with a few "trial balloons" and watching the musicians faces, or be saying something to the effect of, "Stop me if you've heard this one before." This is a good idea because it avoids the embarassment of starting with your "this is a wire" speech and then being asked how you do second order temperature compensation of your exponential current sources!

Another thing to bear in mind is that musicians and engineers do have a common language - ordinary everyday speech. This can be most useful. For example if the musician wants something that sounds like a barking dog, but is afraid to ask, he might try to speak the engineer's language and ask for certain mixtures of noise and resonant filters, etc. Or he might try to explain the sound in suitable intellectual terms from a musical angle. If he just asks for a dog, he has a better chance of getting a dog.

CHAPTER 2A

SUBTRACTIVE SYNTHESIS

CONTENTS:

INTRODUCTION

The basic principle of sound synthesis that has been employed for the last ten years is that of "subtractive Synthesis." Simply stated, the method is to start with an electrical waveform that is rich in the features that we want to use for musical sounds, and break it down and use those features we want. This can be compared with the method of "additive synthesis" to be discussed in the next chapter. In additive synthesis, a waveform is built up from component parts.

As an example of subtractive synthesis, we might start with a waveform that is rich in harmonics and with a standard (i.e. largest ever used in the system) amplitude. We can then select certain components of the waveform by filtering, and select and control the amplitude by controlled attenuation.

An important feature of subtractive synthesis as it is used today is the principle of voltage control. That is, we make our filters and amplifiers (attenuators) vary in response to a voltage control input. The control voltages that can be used are typically grouped into three categories:

1) Control voltages from manual controllers such as keyboards which give the musician note-by-note control over the output.

2) Control envelopes produced by envelope generators. These are voltage contours that result when appropriate timing signals are recivied by the envelope generator. They go through one cycle and then wait for the timing signals to reoccur.

3) Periodic modulating voltages. Modulation will be discussed in Chapter 2c.

HARMONIC CONTENT OF WAVEFORMS

The usual waveforms available from VCO's include the sine, square, triangle. sawtooth, and variable pulse. Of these, the sine has no overtones, the square and triangle have only odd harmonics, and the sawtooth and pulse are the riches in harmonics. The harmonic content is obtained from the well known Fourier series analysis. It is perhaps surprising that even a small set of waveforms such as the above 5 will serve well for producing a wide variety of musical tones. Another source that can be used to advantage is the noise source (a wide-band spectrum containing a continuous set of frequencies.

Below are the Fourier series of the common waveforms. We have added to the list also the outputs of a half-wave and a full-wave rectifier with a sinewave input. These latter waveforms are added because they are easy to obtain and because unlike the other common waveforms, these two have only even harmonics:

SQUARE
$$f(x) = \frac{4}{\pi}\left\{\frac{\text{Sin}(x)}{1} + \frac{\text{Sin}(3x)}{3} + \frac{\text{Sin}(5x)}{5} + \cdots\right\}$$

TRIANGLE
$$f(x) = \frac{8}{\pi^2}\left\{\frac{\text{Sin}(x)}{1} + \frac{\text{Sin}(3x)}{9} + \frac{\text{Sin}(5x)}{25} + \cdots\right\}$$

SAWTOOTH
$$f(x) = \frac{2}{\pi}\left\{\frac{\text{Sin}(x)}{1} - \frac{\text{Sin}(2x)}{2} + \frac{\text{Sin}(3x)}{3} - \cdots\right\}$$

PULSE
$$f(x) = \frac{4}{\pi}\left\{\frac{\text{Sin}(a)\text{Cos}(x)}{1} - \frac{\text{Sin}(2a)\text{Cos}(2x)}{2} + \frac{\text{Sin}(3a)\text{Cos}(3x)}{3} - \frac{\text{Sin}(4a)\text{Cos}(4x)}{4} + \cdots\right\}$$

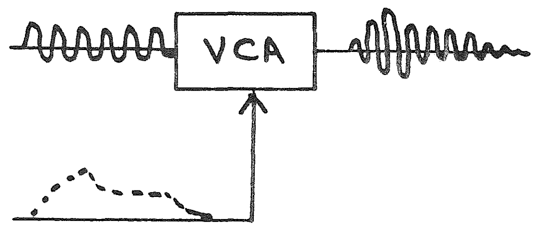
F.W.R. SINEWAVE
$$f(x) = \frac{2}{\pi} - \frac{4}{\pi}\left\{\frac{\text{Cos}(2x)}{1\cdot 3} + \frac{\text{Cos}(4x)}{3\cdot 5} + \frac{\text{Cos}(6x)}{5\cdot 7} + \cdots\right\}$$

H.W.R. SINEWAVE
$$f(x) = \frac{1}{\pi} + \frac{\text{Sin}(x)}{2} - \frac{2}{\pi}\left\{\frac{\text{Cos}(2x)}{1\cdot 3} + \frac{\text{Cos}(4x)}{3\cdot 5} + \frac{\text{Cos}(6x)}{5\cdot 7} + \cdots\right\}$$

The important things to remember are that the square and the triangle have the same harmonics, but the triangle has less total harmonic content (the square's harmonics fall off as $1/n$ while the triangle's harmonics fall off as $1/n^2$). The sawtooth has all harmonics and they fall off as $1/n$. The pulse has all harmonics which fall under a $\mathrm{Sin}(x)/x$ type of envelope that depends on the pulse width. In particular, for a pulse width of $1/m$, where m is an integer, every m^{th} harmonic is missing from the spectrum. The square is a special case of the pulse. Note also the DC components in the rectified sine waves - that's what rectifiers are for. While all the phases are set as either sines or cosines, the actual phase is relatively unimportant. The phase makes little difference as far as the sound is concerned although the waveshape will be changed.

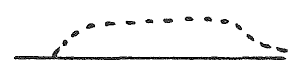
AMPLITUDE AND HARMONIC CONTENT ALTERATION

A voltage-controlled amplifier (VCA) is actually a voltage-controlled attenuator since it passes signals with a gain from zero to one in response to a control voltage. To control the amplitude of the sound we are synthesizing, an envelope is applied to the VCA control. The envelope starts from zero (defining the start of the tone) and ends back at zero (the end of the tone). During the time that the envelope is not zero, the amplitude control can take on a wide variety of shapes. The most popular envelope shape is the ADSR (attack, decay, sustain and release). A VCA controlled by an ADSR envelope is shown at the right. For convenience, exponential ramps are used for envelopes, as these are easily implemented as charging and discharging capacitors. It turns out that a wide variery of upslopes are acceptable as attacks. The AD part of the ADSR envelope is useful as a sharp percussive envelope. On the other hand, there seems to be a strong preference for the exponential decay as opposed to any other form. This is reasonable when you consider that many traditional instruments have something very close to exponential decay. Other types of envelopes that are often used are the AD (Attack-decay) and the AR (attack-release) which can be seen to be limiting cases of the ADSR.



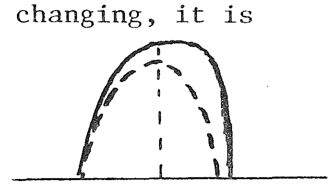
VCA WITH ADSR ENVELOPE

AD ENVELOPE



AR ENVELOPE

In addition to amplitude shaping, the tone color is also altered to some degree by the attack and decay transients. When the amplitude level is changing, it is easy to see that even a pure tone sinewave input will be altered. For example, a rising amplitude level would cause a sinusoidal waveshape to be altered slightly as indicated at the right. This is a general result - only tones with unaltered amplitude can have a spectrum given by the Fourier series.



The small change of tone color due to amplitude change is incidental to the amplitude control procedure. The main mechanism of tone color control is the voltage-controlled filter (VCF). The most popular VCF's are the 4-pole low-pass, 4-pole high-pass, and the state variable, which offers a 2-pole response with low-pass, high-pass, and band-pass outputs. Other types of filters can be used as well. Generally it is the filter's center frequency that is swept by the control voltage, but other parameters like filter Q and roll-off slope can be voltage-controlled.

In general, VCF's are controlled by ADSR envelopes since these envelopes are available. There is however, no real reason to prefer this type of shape for a filter control envelope. In many cases however, it may not be possible to tell the difference between the ADSR filter control and some similar shapes - the important thing seems to be the time constants of the envelope.

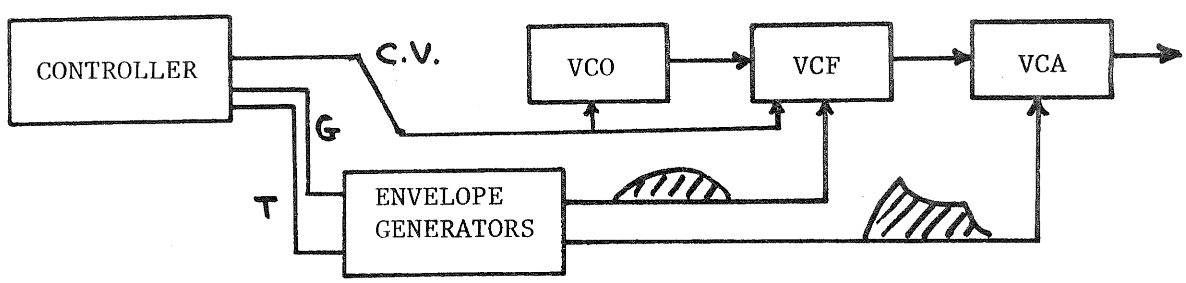
VCF's like all filters have a phase response as well as a frequency response. We often assume that a VCF simply alters the amplitude of the Fourier components of the waveform without making any other changes. It turns out that this is actually quite close to what the ear actually receives, since the ear is relatively (but not absolutely) insensitive to phase changes among the components of a complex wave. Even if we ignore phase for the moment, it is clear that the removal of some components by the filter will alter the amplitude of the output waveform. Thus the VCF is also to a degree an amplitude control device. However, for practical purposes, this amplitude variation is usually not important. First of all, the ear requires something like a 10% change in amplitude to even notice, and total harmonic content will often be only several times this value. Secondly, the apparent amplitude is often a function of the harmonic content. This can be demonstrated by comparing a sine wave and a sawtooth wave of the same amplitude and frequency. The sawtooth will seem much louder. The VCF becomes a noticeable amplitude control when it starts to cut through the fundamental frequency. In this case, it may act like a VCA for sine waves since it is altering only the fundamental component. Many synthesizer users may be fooled at some time by a filter frequency that is set so low that either nothing gets through, or there is only a short "blip" when the envelope reaches its peak.

Phase effects in filters become apparent in two cases. The first is when the waveform is displayed on an oscilloscope. By lowering the cutoff frequency of a low-pass filter with a square wave input for example, one might expect to see a waveform that shows only a limited number components remaining (e.g., the first and third). In fact, the resulting waveform will in general not resemble the theoretical square wave with higher components removed due to phase shift of the components in the filter. This phase shift is a function of frequency, and therefore each of the components is shifted by a different amount, and the waveshape is altered. It is only when the filter low-pass cutoff reaches the region between the fundamental and the 3rd harmonic that the fundamental sinewave is isolated and the expected waveshape is seen on the screen. To isolate individual harmonics, a sharp bandpass filter can be used. This is a common demonstration, but is often confused by "ringing" of the filter that is mistaken for actual harmonic content. Ringing can be seen since the frequency of the ringing varies with the filter frequency, not with the square wave input frequency. The demonstration of the presence of harmonics should be inferred from the presence of amplitude peaks in the filter output as the filter frequency is changed, not by the waveshape on the screen.

A second case where phase of the filter is important is where the filter's frequency is altered rapidly so that the output is subject to detectable phase modulation. This effect will be discussed in chapter 2c.

A BASIC SUBTRACTIVE SYNTHESIS PATCH

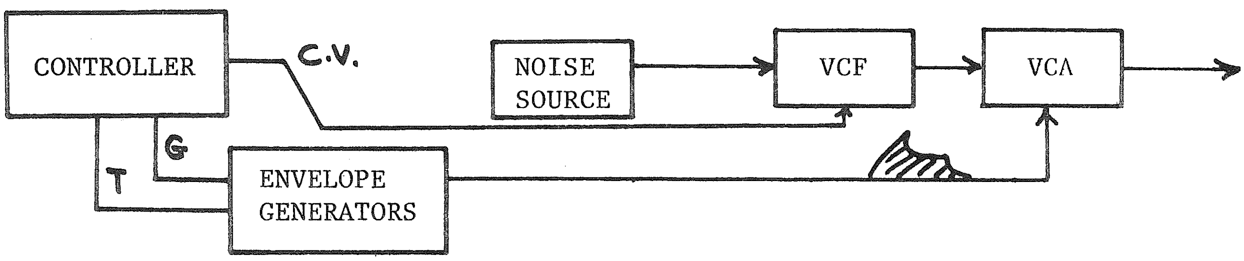
A very basic subtractive synthesis patch is shown below:



The VCO is controlled by the main control voltage from the controller and this same voltage also causes the filter to track. This means that in the absense of an envelope to the VCF, the waveform at the output of the VCF will have a constant shape as the voltage from the controller goes up and down. This results in a constant tone color. However, in addition to the main control voltage, the VCF is also controlled by an envelope, and this causes an alteration of the main tone color as a single note progresses. The envelope to the VCF is responsible for the familiar wwwoooooowww sound associated with much of the music produced on synthesizers today. The last module in the line is the VCA. It is controlled by the amplitude envelope and defines the beginning and end of the tone. Placement of the VCA and VCF processors is somewhat arbitrary. There is some difference because if the VCA is placed first, the transisent response of the filter is involved. Generally the VCA is put second as shown above because it shuts off the signal completely, including any noise in the output of the filter.

SUBTRACTIVE SYNTHESIS USING A NOISE SOURCE

Another useful patch replaces the VCO with a white noise source, sets the filter for a sharp Q as a bandpass (or sharp corner peaking for a low-pass) and removes the control envelope from the VCF:

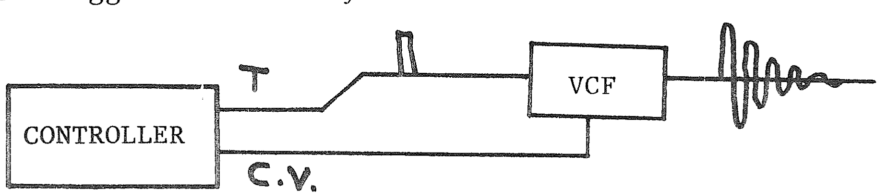


In this patch, a narrow band of white noise is allowed to pass through the filter according to the control voltage to the VCF. The process that results can be viewed in two ways. In a sense, the white noise is being filtered and frequency components in the range of the pass band of the filter are being allowed through. This results in a narrow band of frequencies with a moderate sense of pitch center, depending on the Q of the filter. It is also possible to think of the noise as exciting the filter into ringing. The filter thus responds to a series of random impulses. In either interpretation, the pitch center is easily made strong enough to carry a melody.

FILTER RINGING

This technique is similar to subtractive synthesis although strictly speaking it belongs to a class more properly termed resonant synthesis. It can be thought of as a special case of noise excitation where a single noise spike is used.

The high Q filter used above can be excited by a single impulse. The impulse can be the trigger from the keyboard as seen below:



The energy input to the filter by the impulse damps out causing the filter to ring. A VCA can be used to control amplitude, but the system can be used to provide its own envelope, as the ringing damps out exponentially. This results in a characteristic "raindrop" or popping sound.

CHAPTER 2B

ADDITIVE SYNTHESIS

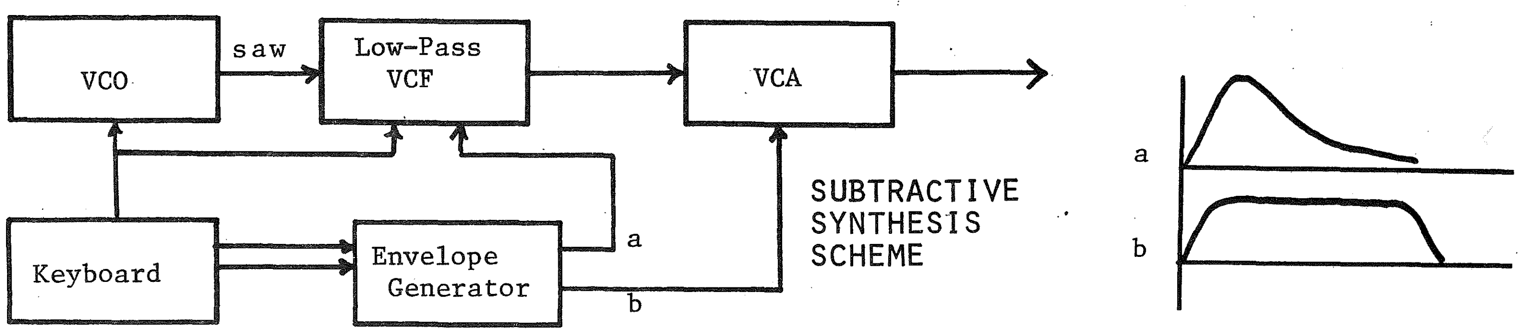
CONTENTS:

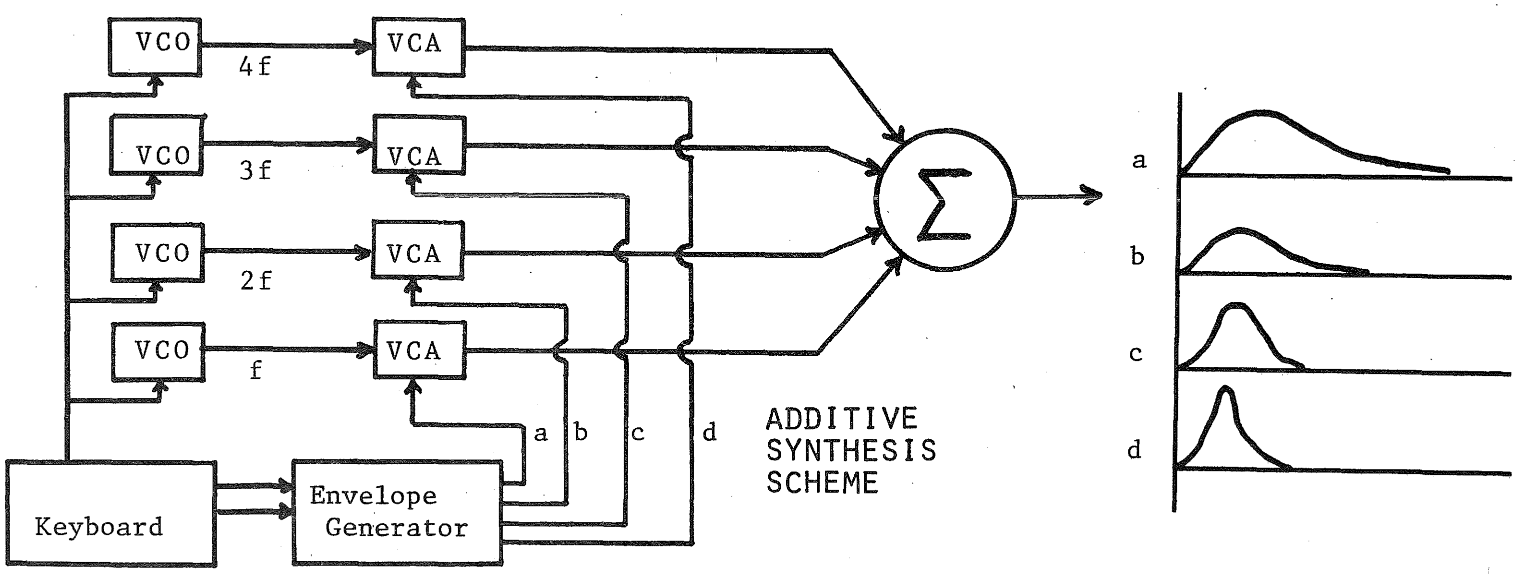
INTRODUCTION

It is conceptually simpler to obtain a desired result by assembling a necessary set of components rather than by starting with an excessively large set of components and discarding unwanted ones. Thus, additive synthesis is conceptually simpler than subtractive synthesis. We have considered subtractive synthesis first because it is more familiar and the hardware is often simpler. Additive synthesis is perhaps more general and is becoming more popular and practical.

COMPARISON OF ADDITIVE AND SUBTRACTIVE SYNTHESIS

The basic idea of additive synthesis is to first make available a large number of sources and group them into what is called a source bank. We can consider the source bank to consist of a large number of sinewaves, or any other group of sounds that proves to be useful and convenient. These sources are then mixed together in a time varying proportion to produce a time varying sound. In this way, harmonic structure can be made to vary with VCA's rather than with VCF's. It is possible to consider a roughly equivalent set of additive and subtractive synthesis methods. The two setups are shown at the top of the next page. The subtractive synthesis scheme is the standard one discussed in the last chapter. The additive synthesis scheme consists of a bank of VCO's each with its own VCA and envelope. Note that there is no overall VCA and envelope for amplitude control. This is not needed because when all the envelopes are low, the sum of the outputs is low. If all the envelopes are the same, there is no time dependent

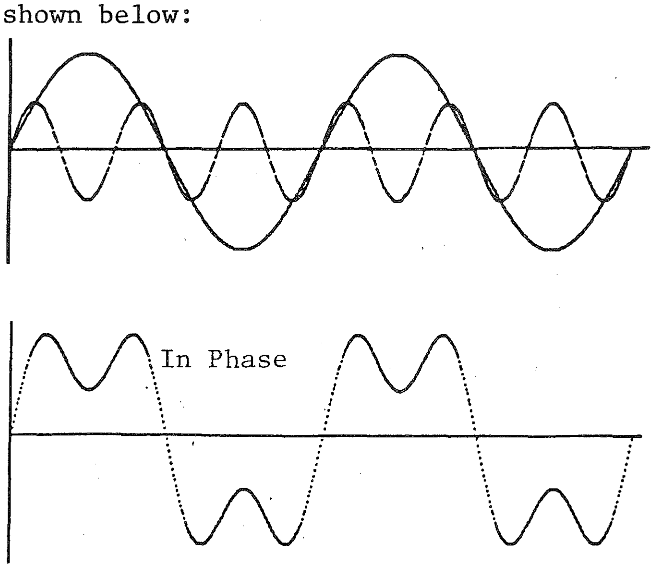


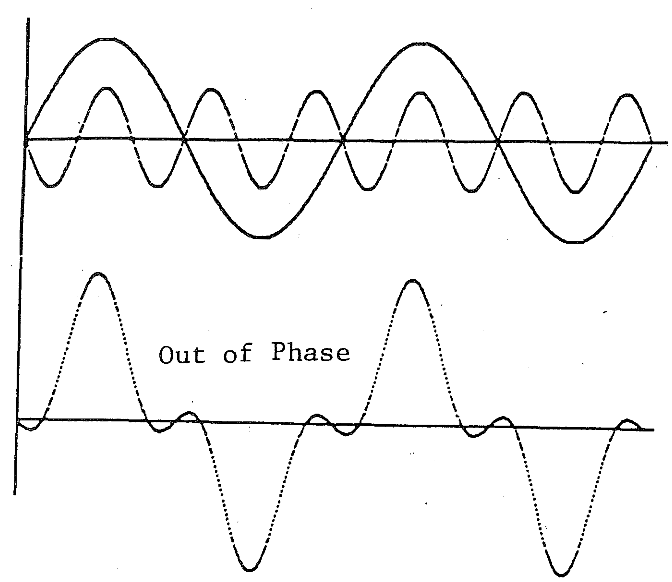


change of tone color. The two above techniques are roughly equivalent. The subtractive synthesis scheme would be preferred on the basis of simplicity in this case. Either system has a major problem in that the resulting sound is too exact - the harmonic relationship is too precise. In subtractive synthesis, additional non-harmonic tones can be added by modulation techniques. In additive synthesis, they can be added directly.

THE ROLE OF PHASE IN ADDITIVE SYNTHESIS

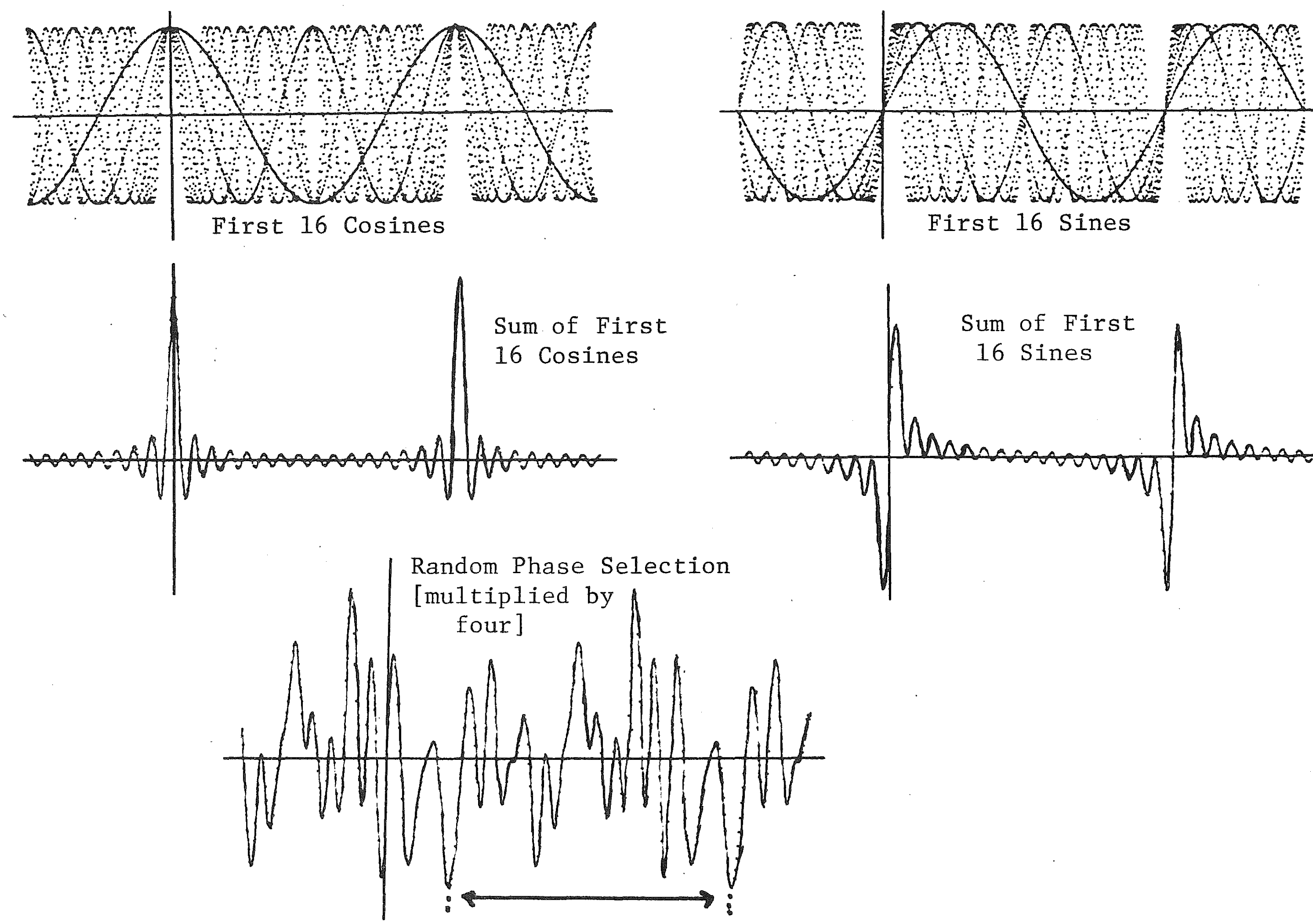
The role of phase in waveform generation was discussed by Hal Chamberlin in EN#25. The addition of the first and third harmonics can be done in phase or out of phase as shown below:





Even in this simple example, it is clear that the shift of phase has a large effect on the shape of the waveform. It is a surprising result that the ear is relatively insensitive to this change of phase. In cases where such a waveform is played into an open room so that much of the sound that reaches the ear is indirect, it is most difficult to tell the difference between the two. If played directly to the ears through earphones, it may be possible to tell a small difference. Obviously, this is of profound importance in the construction of source banks since we may be able to use only sine waves for example, and forget about the corresponding set of cosines. This result can be understood when we consider that the ear is in some sense a Fourier spectrum (power) analizer.

As a second example, Chamberlin showed the sums of the first 16 sines and cosines, as well as a random phase arrangement of the first 16 Fourier components, all of equal amplitude. The three sound alike in an open room, and nearly so with earphones. The waveforms are shown below:



ADDITION OF NON-HARMONICS

The source bank for additive synthesis may consist of any group of sounds, so it is possible to have any grouping of harmonics and non-harmonics, noise, or whatever. In general, the easy ways of producing a bank of harmonics locks the available harmonics at integer ratios as a result of direct waveshaping, or due to digital counting. Thus, special techniques may have to be employed to produce non-harmonic tones.

It is relatively easy with digital methods of producing harmonic source banks to add in non-integer dividers. A second technique which can be used is to have several source banks driven by different clocks. Waveforms can be selected from among the different banks. Finally, a number of modulation techniques can be applied to modulate the entire source bank. It will in general be necessary to provide some means of achieving non-harmonic tones if a wide variety of sounds are desired.

SOURCE BANKS

As mentioned above, there are no real rules about what can go into a source bank. Obvious choices are sine waves, but these are not easy to generate in a large set. One method that can be tried is to start with a single sinewave (cosine) and use multipliers and summers to realize identities of the form:

$$\mathrm{Cos}\ 2x = 2\ \mathrm{Cos}^2 x - 1$$

$$\mathrm{Cos}\ 3x = 4\ \mathrm{Cos}^3 x - 3\ \mathrm{Cos}\ x$$

$$\mathrm{Cos}\ 4x = 8\ \mathrm{Cos}^4 x - 8\ \mathrm{Cos}^2 x + 1$$

$$\mathrm{Cos}\ 5x = 16\ \mathrm{Cos}^5 x - 20\ \mathrm{Cos}^3 x + 5\ \mathrm{Cos}\ x$$

$$\mathrm{Cos}\ 6x = 32\ \mathrm{Cos}^6 x - 48\ \mathrm{Cos}^4 x + 18\ \mathrm{Cos}^2 x - 1$$

A similar set is possible for a bank of Sin functions, but this requires a Sin x and a Cos x basis. For complete generality, it is necessary to have both Sin and Cos functions in a Fourier synthesis scheme. However, as mentioned above, the phase of the harmonics is of secondary importance, so a single phase bank is a reasonable choice. The main problem with the above method is the difficulty of trimming the multipliers and summers. It is not practical to extend the scheme too far.

A digital sinewave generation method that gives 32 harmonics with nearly continuous phase selection has been described [H. Chamberlin, EN#39 and AES Preprint #1023 (E-4) May 1975]. It is likely that the method could be extended to give non-harmonic tones as well.

A second set of functions for a source bank that can be considered is a set of square waves or other rectangular functions formed by using digital dividers. A variety of waveforms can be tried. The easiest set is the square wave "subharmonics" that are formed by dividing down an oscillator. In such a case, the subharmonics are not in general harmonically related [see J. Ball, "The Digital Divider in Music Synthesis," EN#34 (8)]. There exists a simple set of rectangular functions called the "Walsh Functions" that have the property (like Fourier harmonics) of being a complete set in the mathematical sense. A report on the use of Walsh functions can be found in the reference papers in chapter 9d. Non-harmonic tones can be produced with the Walsh functions by the method shown in the reference. One remaining problem with the digital counting methods is that they are all locked to the single driving oscillator. This means that the "ensemble" effect is still absent. Ensemble effects result from minor variations as oscillators drift in and out of phase. For this reason, it is often useful to have several source banks driven by different oscillators.

It may be possible to do a limited amount of additive synthesis on a large synthesizer that is primarily intended for subtractive synthesis. If three or more VCA's and three or more envelope generators are available in addition to a mixer, any available oscillators can be used as a source bank for experiments in additive synthesis.

CHAPTER 2c

GENERALIZED MODULATION

CONTENTS:

INTRODUCTION:

A single electronic oscillator produces an inherently uninteresting sound*. The sound is too perfect and regular. Even when modes of user control are provided to give some variation, this may prove insufficient for extended musical tones which require subtle changes. To provide a variety of musical structures and timberal developments, it is necessary to alter or process the oscillator output in some manner. Generally, the exact parameters of the processing are made to vary in time, and this often leads to different regimes (and different terms for the sonic "effects" that result) according to the time constants of the change. This is because the ear which is the final receiver has its own time constants for signal detection. Of course, the mathematics of the processing does not change with a different time constant of the processing parameters. The electrical processing is objective. It is the subjective effect of the processed signal on the hearing mechanism that causes a different effect to be described.

*since the term "inherently uninteresting sound" is controversial, I should add that I have in mind waveforms of low or static information content that are not interesting to the ear as sounds. Any sound when properly used may be musically interesting.

A good example of what we are getting at is frequency modulation - the control of a VCO output frequency with a periodic waveform. At a modulation frequency of 0.1 Hz, the ear clearly follows the rise and fall of the pitch (a siren effect). At 10 Hz, the effect is close to a musical vibrato. At 1000 Hz, a complex, generally inharmonic spectrum results that is useful for bell-like and other percussive sounds. The mathematics is the same in all cases, at least to the point where the sound reaches the ear.

The FM example further points out the need to explore the full range of any control or modulating parameter for musical usefullness. The fact that many of these signal processing devices respond to control signals that are faster or more complex than those that can be produced with manual control leads to the need for voltage control of the processing parameters.

Processing of signals can be done in many ways and it is useful to have a general theory for these since there are many similarities. This is particularly true of what we may refer to as the general type of sounds that can be produced for different types of processing with similar control signals. For example, rapid and deep modulation by different processes tends to produce the same types of sound. We can group these signal processing techniques into three categories as follows:

1) Those which involve the interaction of two or more signals without one being considered the signal and the other the control. [e.g., linear mixing and balanced ("ring") modulation].

2) Those which result from the application of a control signal to the signal producing device itself. [i.e., types of frequency modulation].

3) Those which involve the processing of one signal by applying a control signal to some device external to the signal producing device. [e.g., AM, formant modulation, sampling, PWM].

We will refer to the above as "generalized modulations." One point should be made right away: while we often modulate, we seldom demodulate. The modulated signal is presented directly to the ear. This means that often lost information and/or extraneous information may not be at all important as far as the musical sounds are concerned, and the actual engineering of the modulator can be relaxed or simplified.

Much of what we do below relies on trig identities. It should be realized that these identities are correct and when we realize a trig operation electronically, the output will be correct to the accuracy of the electronics. Often we will find that one side of the identity gives us what we need to know about the spectrum while the other side tells us more about what the ear will actually hear. Both are useful. In many cases, we simply look at processing from the point of a single sinewave input, and handle complex waveforms by the principle of superposition. This is generally valid and useful, but it should be remembered that phase relationships between the components should be considered to see if they are important. Often they are not. Finally, we often choose sines or cosines somewhat at random. If you have worked much with trig identities, you know that these choices are ususally not important and everything has a way of working itself out. This is of course a matter of initial phase and is only important where there is a reference phase to consider.

We have included in this chapter several processes that are usually not considered as standard types of modulation. This has been done since analysis of these devices fits in nicely with modulations, and because the sounds produced may be similar to those produced by modulations.

ADDITION OF SIGNALS (LINEAR MIXING)

What happens when we add two sinewaves of different frequency but of the same amplitude? This is a common situation - for example when two standard 1000 ohm outputs are mixed by plugging both into a multiple. For this analysis, we have only to look at the trig identity

$$\text{Sin } A + \text{Sin } B = 2\cdot\text{Sin } \frac{A+B}{2} \text{ Cos } \frac{A-B}{2}$$

Note that the left side of the identity tells us all we need to know about the spectrum of the combined signal. Thus, no new spectral features are added by linear mixing. We look at the right side of the equation because it often tells us better what the ear will actually hear.

Suppose first that $A=\omega t$ and B is slightly smaller $(\omega - \delta)t$ where $\delta << \omega$. Then the right side of the trig identity becomes:

$$2\cdot\text{Sin}(\omega - \delta/2)t\cdot\text{Cos}(\delta/2)t$$

What the ear hears is a tone $\text{Sin}(\omega - \delta/2)t$, the average frequency, with its amplitude modulated by the term $\text{Cos}(\delta/2)t$. This is the phenomenon of "beating." Beating is usually described as an amplitude variation of the average frequency at a rate that depends on the difference frequency. The formula above clearly shows how the average frequency comes in $(\omega - \delta/2)$, but the multiplying term is $\delta/2$, half the difference, <u>not</u> δ. The apparent descrepency between the usual musical description and the mathematics is easily resolved by considering that you get two maxima in amplitude, one positive and one negative, for each cycle of $\text{Cos}(\delta/2)$. Since $\text{Cos}(\delta/2)$ is actually the "envelope" of the combined signal, there are two amplitude peaks for each cycle of the modulating term.

When the two sinewaves being added have a frequency difference in the range of 10 - 15 Hz, the beating sensation becomes so rapid that the ear can no longer follow the amplitude variations clearly. The sound in this case is described as having the quality of roughness. This roughness continues until the ratio of frequency reaches what is called a "critical band" at which time, the two tones are heard separately and the sensation is one of smoothness. This critical bandwidth varies with frequency and with the individual, but is on the order of a minor third (ratio 6:5). It should be mentioned that the two tones can be heard as separate tones below the critical bandwidth, but with roughness.

Above the critical bandwidth, the sensation is generally one of smoothness. However, as the difference frequency increases, several additional effects come in which cause the listener to hear other tones. These are caused by the non-linearity of the ear for large amplitude signals and by the ability of the ear to preceive pitch as an overall repetition rate as well as from spectral information. Also, for certain small regions, where the two tones are close to a small integer ratio, "second order beats" can be heard. The second order beats are quite different from first order beats: (1) There are no large amplitude variations (restricted to 10% or less). (2) The second order beat effect is binaural - the effect is still heard when the tones are applied to separate ears. This is not true of first order beats. The conclusion is that the first order beats are due to an ear mechanism; second order beats are due to an ear-brain mechanism. (3) If the second tone differs from an integer multiple by an amount ε, the beat rate goes as ε multiplied by the denominator of the fraction expressing the integer ratio. For example, for a mistuned fifth (close to 3/2, the beat rate is 2ε. For a mistuned octave (close to 2/1) the beat rate is ε. For a mistuned fourth (close to 4/3), it is 3ε.

Next we can look at the addition of two sinewaves of the same frequency but of different phase. This is really just a simple vector addition, and often a simple graphical construction is the easiest. The trig development however is as follows:

$$A_1 \mathrm{Cos}(\omega t + \phi_1) + A_2 \mathrm{Cos}(\omega t + \phi_2) = A_3 \mathrm{Cos}(\omega t + \phi_3)$$

where:

$$A_3 = \left\{(A_1 \mathrm{Cos}\ \phi_1 + A_2 \mathrm{Cos}\ \phi_2)^2 + (A_1 \mathrm{Sin}\ \phi_1 + A_2 \mathrm{Sin}\ \phi_2)^2\right\}^{1/2}$$

$$\phi_3 = \mathrm{Tan}^{-1} \frac{A_1 \mathrm{Sin}\ \phi_1 + A_2 \mathrm{Sin}\ \phi_2}{A_1 \mathrm{Cos}\ \phi_1 + A_2 \mathrm{Cos}\ \phi_2}$$

The most important point brought out by these formulas is that as long as the frequencies of the components are the same, the result is a "sinusoidal" waveform with a different amplitude and phase. Thus, even in cases where it is possible to hold the frequencies equal to avoid beating, the mixing of sinewaves gives nothing new for the ear to hear. The technique of mixing sinusoidals is useful where it is necessary to realize a certain relative phase by mixing two known phases, usually quadrature phases. For example, mixing two components that are 90° out of phase gives a 45° phase inbetween with an amplitude $\sqrt{2}$ times the original amplitude. The equations are also useful for generating a "three phase" signal (120° apart) from quadrature signals. These phases are often used to realize other trig identities where relative phase is important.

The most general case is where we assume that both frequencies and amplitude are different. It is not necessary in this case to consider different phases, as this only determines a starting point in time, and we can always shift this time axis. In this case, the trig development leads to:

$$A_1 \mathrm{Cos}\ \omega_1 t + A_2 \mathrm{Cos}\ \omega_2 t = A_3(t)\ \mathrm{Cos}[\omega_1 t + \phi(t)]$$

where:

$$A_3(t) = \left\{A_1^{\,2} + 2\ A_1 A_2 \mathrm{Cos}(\omega_1 - \omega_2)t + A_2^{\,2}\right\}^{1/2}$$

$$\phi(t) = \mathrm{Tan}^{-1} \frac{A_2 \mathrm{Sin}(\omega_2 - \omega_1)t}{A_1 + A_2 \mathrm{Cos}(\omega_2 - \omega_1)t}$$

The fact that the frequencies are different therefore (as expected) leads to a more dynamic waveform since both amplitude and phase of the output are functions of time. The waveform is not sinusoidal even though it is represented as a Cos term. This is due to the time changes in the amplitude and phase. In particular, the time rate of change of phase is effectively an instantaneous frequency:

$$\omega_{inst} = d\phi(t)/dt$$

The superposition is thus very similar to modulation since both phase and amplitude change at a rate given by the difference frequency $\omega_2 - \omega_1$.

Note that the simple addition of sinewaves considered in the first part of this discussion is a simple case of the above formulas. Also, we shall note later that the close connection between superposition of sinewaves and modulation is no accident as the modulation processes usually lead to a set of discrete sinewaves in the spectrum. These are referred to as sidebands.

We have considered above the mixing of sine waves. What happens if we mix two complex waveforms? Consider for example the sawtooth with Fourier series:

$$f(t) = 2 \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} \text{ Sin } n\omega t$$

$$= 2 \text{ Sin } \omega t - \text{Sin } 2\omega t + (2/3)\text{Sin } 3\omega t - \ldots\ldots$$

If a second sawtooth has frequency $\omega+\delta$, its Fourier series is:

$$f'(t) = 2 \text{ Sin}(\omega+\delta)t - \text{Sin}(2\omega+2\delta)t + (2/3)\text{Sin}(3\omega+3\delta)t - \ldots\ldots$$

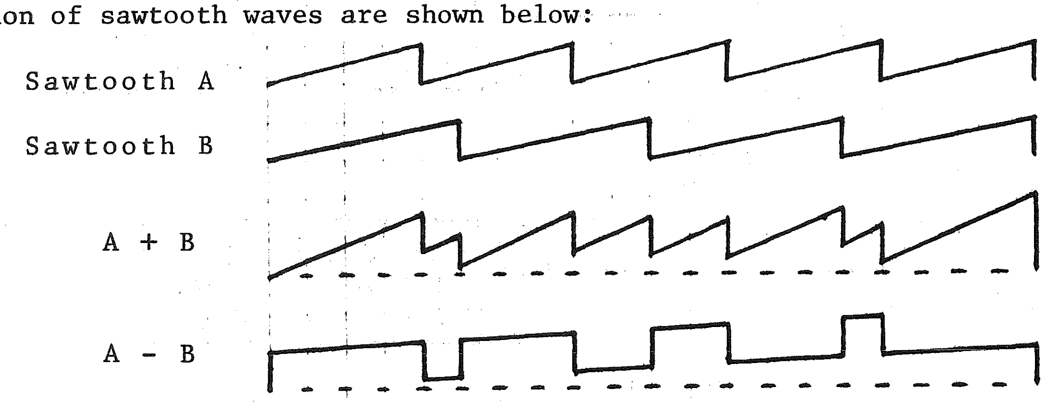
The sum of the two sawtooth waves is thus:

$$\begin{aligned} f(t) + f'(t) = & \; 2[\text{Sin } \omega t + \text{Sin}(\omega+\delta)t] \\ & -[\text{Sin } 2\omega t + \text{Sin } (2\omega+2\delta)t] \\ & +(2/3)[\text{Sin } 3\omega t + \text{Sin}(3\omega+3\delta)t] - \ldots\ldots \\ = & \; 2[2 \text{ Sin}(\omega - \delta/2)t \text{ Cos}(\delta/2)t] \\ & -[2 \text{ Sin}(2\omega-\delta)t \text{ Cos}(\delta)t] \\ & +(2/3)[2 \text{ Sin}(3\omega - 3\delta/2)t \text{ Cos}(3\delta/2)t] - \ldots\ldots \end{aligned}$$

Each of the overtones in one sawtooth wave beat against the corresponding overtone of the second. The beat rate is δ times the order of the harmonic.

In the case of beating sinewaves, we saw that the beat rate resulted in periodic amplitude variations. We saw that the beating resulted in annoying amplitude changes and that if special provisions are made to assure that beating does not occur, all that results anyway is another sinusiodal wave. Thus, the mixing of sinewaves is not a useful way of producing new sounds.

Thus, we can ask about the mixing of sawtooth waves as a useful method of producing new timbres. Experimentally, we note that the mixing results in the beating of overtones at different rates, and this does produce a richer timbre. Secondly, the beating does not produce annoying amplitude variations. It is also useful to mix a sawtooth with an inverted sawtooth. This results in a pulse waveform, and beating is equivalent to pulse width modulation. Addition and subtraction of sawtooth waves are shown below:



In the subtraction of sawtooth waves, it is often felt that amplitude variations will occur since it is possible for one sawtooth wave to completely cancel the other. It is true that a zero in the combined amplitude will result, but the important thing is not the zero, but the way the amplitude goes to zero. In the case of sinewaves, the amplitude goes to zero slowly, and builds up again slowly. In the case of the sawtooth, we saw that subtraction resulted in a pulse waveform. If this is examined carefully, it will be seen that the duty cycle of the pulse varies between 0 and 100%. The zeros in the amplitude of the combined sawtooth waves correspond to the 0 and 100% extremes of the PWM. The zero amplitude is therefore present only instantaneously and is not detected. It can be seen that this happens because the sawtooth has a sudden jump. In general, we can look at other waveforms to see if beating causes annoying amplitude variations. The table below summarizes the information on the waveforms:

WAVEFORM	HARMONICS	HARMONICS GO AS	SHARP TRANSITIONS?	DIFFERENT INVERTED EXCEPT PHASE?	ANNOYING BEAT AMPLITUDE CHANGES?
Sine	f	--	No	No	Yes
Triangle	f, 3f, 5f...	$1/n^2$	No	No	Yes
Square	f, 3f, 5f...	1/n	Yes	No	No
Sawtooth	f, 2f, 3f...	1/n	Yes	Yes	No

Thus we can see that the important factor for avoiding annoying amplitude variations during beating is to use a waveform that has sharp transitions (vertical sides).

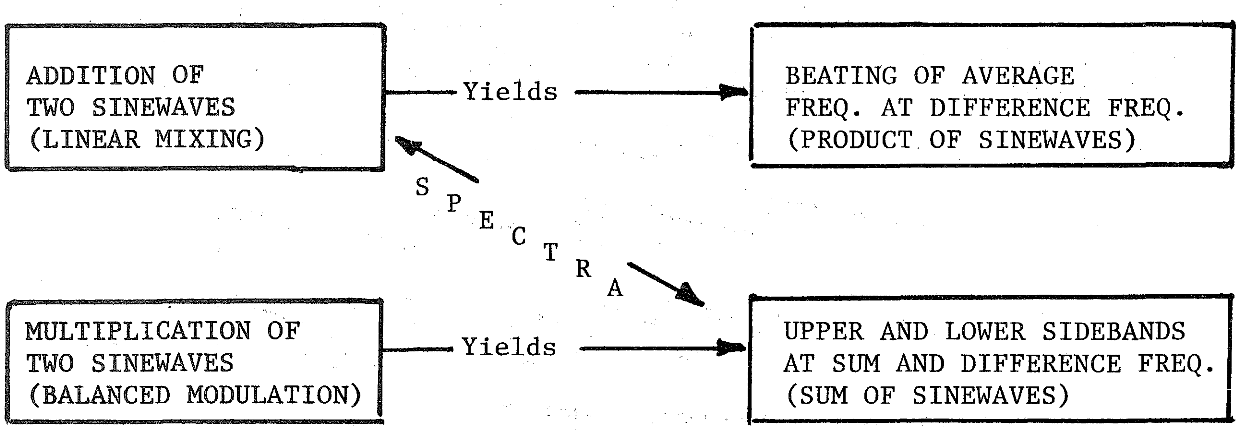
MULTIPLICATION OF SIGNALS (BALANCED OR "RING" MODULATION)

The basic equation of interest for multiplication of signals is basically just the inverse of the one used for addition of two sinewaves of the same amplitude:

$$\sin A \cdot \sin B = (1/2)[\cos(A-B) - \cos(A+B)]$$

That is, two sinewaves multiplied split into two "sidebands" at the sum and difference frequencies. Thus, we can see the following correspondence between addition and multiplication of sinewaves:

PROCESSING　　　　　　　　　　OUTPUT DESCRIPTION



Thus there is a strong correspondence between linear mixing and balanced modulation. In fact, given a stationary portion of the output of either device it is difficult to guess which device is being used. It is when one of the input frequencies is changed that it becomes possible to tell the difference. Consider for example two sinewaves of 400 Hz and 500 Hz. Linear mixing results in a 450 Hz output beating at 100 Hz. Multiplication results in sidebands of 100 Hz and 900 Hz. The linear combination could have been produced by the multiplication of a 450 Hz sinewave and a 50 Hz sinewave. Likewise, the multiplier output could have been simply the linear mix of a 100 Hz sinewave with a 900 Hz sinewave.

In balanced modulation, it makes no difference which of the two signals is considered the "signal" and which is the "carrier." It is the usual case that the higher frequency is considered the carrier, e.g.

$$\left.\begin{array}{ll}\text{Carrier} & 1000\text{ Hz}\\ \text{Signal} & 100\text{ Hz}\end{array}\right\}\ \text{Gives Sidebands}\rightarrow\left\{\begin{array}{l}1000-100 = 900\text{ Hz}\\ 1000+100 = 1100\text{ Hz}\end{array}\right.$$

But we can just as well use:

$$\left.\begin{array}{ll}\text{Carrier} & 100\text{ Hz}\\ \text{Signal} & 1000\text{ Hz}\end{array}\right\}\ \text{Gives Sidebands}\rightarrow\left\{\begin{array}{l}100-1000 = -900\text{ Hz}\\ 100+1000 = 1100\text{ Hz}\end{array}\right.$$

The -900 Hz frequency may seem artificial, but it is a simple matter to change it to a positive frequency since it is the Cos(A-B) term that gives this, and cosine is an even function. If the negative frequency appears in a Sine term, the phase is inverted according to Sin(-A) = -Sin(A).

FREQUENCY SHIFTING

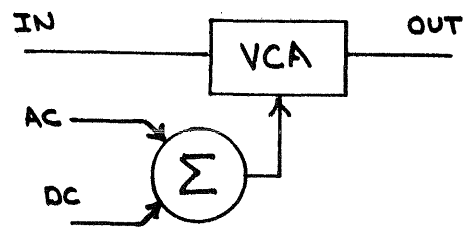
Balanced modulation is double sideband modulation - there is an upper and a lower sideband. This is a special case of amplitude modulation as we shall see, and in radio is known as suppressed carrier modulation. We can go one step further and suppress one of the sidebands. This is single sideband modulation, and when implemented for electronic music, it results in frequency shifting. Thus, a frequency shifter gives an upshift or a downshift to the components of an input signal.

Either a balanced modulator or a frequency shifter can handle a complex waveform by the superposition of sine waves. Thus, when a complex waveform is input, a whole set of sidebands results, and generally they appear in positions that are not integer multiples. The balanced modulator or the frequency shifter are therefore devices useful for producing non-harmonic tones from harmonic ones. Since the two input signals are often comparable in frequency, it is easy to have sidebands appearing at negative frequencies. These are reflected back to the positive spectrum, and a very complex spectrum can result. The non-harmonic tones are useful for percussive and bell-like sounds.

A balanced modulator is simply an analog multiplier and is easy to implement. A frequency shifter is more difficult (see chapter on frequency shifter design). Thus, the frequency shifter must be justified for some purpose other than the production of non-harmonic sounds, as a simpler device would serve for this. The frequency shifter is most useful when there is something to relate the shift to. This is the case when live familiar sounds are input. Since these sounds are known, it is possible to realize that they have been shifted in frequency at the output. Moreover, if the live sound is relatively harmonic in nature, the components will all be shifted by the same number of hertz (not by the same ratio) and this causes a non-harmonic image to be produced. In contrast to the balanced modulator, the frequency shifter provides only a single cluster of shifted components and is often easier to control. A second case where the shift can be related to something is when the shifter is placed in the feedback loop of a tape recorder used to realize tape echo. In this case, the echo shifts with each repetition. Frequency shifters are also potential frequency modulators. Frequency modulation is usually realized by feeding a control voltage to a VCO, but it can also be realized by feeding a control sinewave to a frequency shifter.

AMPLITUDE MODULATION

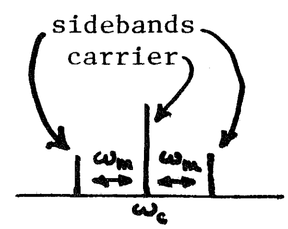
It is easiest to consider amplitude modulation (AM) in terms of the way it is generally produced in electronic music. The basic setup is diagrammed at the right. A periodic control signal is summed with a DC component and used to control the gain of a VCA. Basically this means that we start with an initial gain and then add a periodic variation that drives the gain above and below the initial level. Mathematically, we start with a "carrier" $\sin \omega_c t$, and initial gain of one, and a modulating term $m \sin \omega_m t$. For radio work, m is restricted to values between 0 and 1, (0 to 100% modulation), but for sound synthesis, larger values can be useful. The AM equation then becomes:



$$V(t) = (1 + m \sin \omega_m t)\cdot \sin \omega_c t$$

$$= \underbrace{\sin \omega_c t}_{\text{unsuppressed carrier}} + \underbrace{m \sin \omega_m t \cdot \sin \omega_c t}_{\text{balanced modulation (sidebands)}}$$

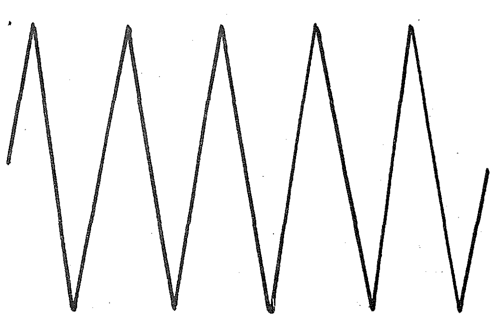
$$= \sin \omega_c t + (m/2)[\cos(\omega_c - \omega_m)t - \cos(\omega_c + \omega_m)t]$$

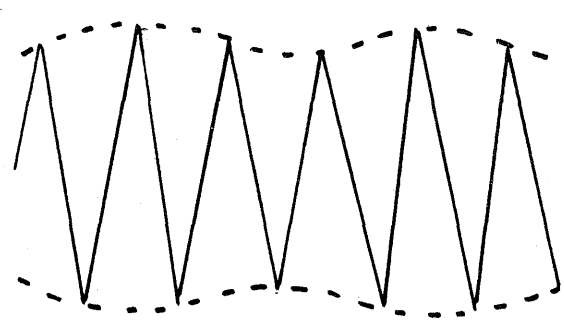


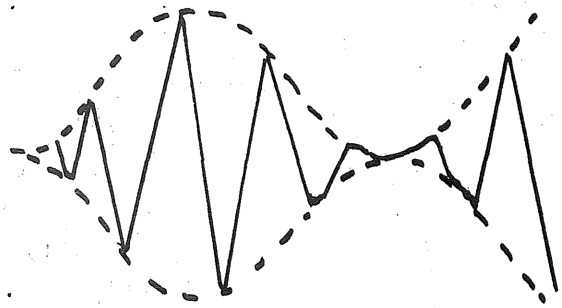
AM SPECTRUM

The factor "m" is the modulation index which determines the depth of modulation. The spacing of the sidebands is ω_m. It is clear that AM is just balanced modulation with part of the original carrier present. This carrier can be very important since it provides the original tonal center. Contrast this with balanced modulation which is just two sidebands of equal magnitude without a tonal center. Both AM and balanced modulation have their uses.

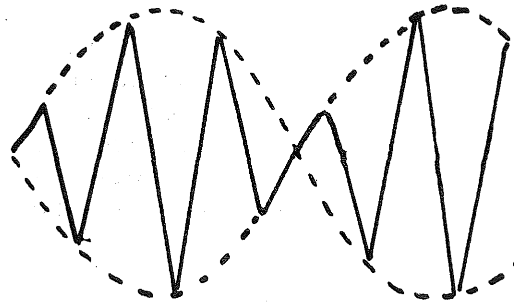
Modulation can often be identified by the pattern of the signal on the screen of an oscilloscope. Below we show several patterns that are typical.


UNMODULATED

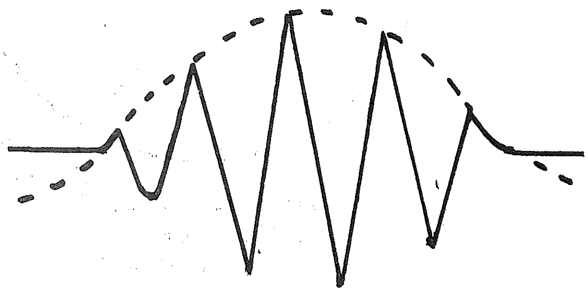

10% AM



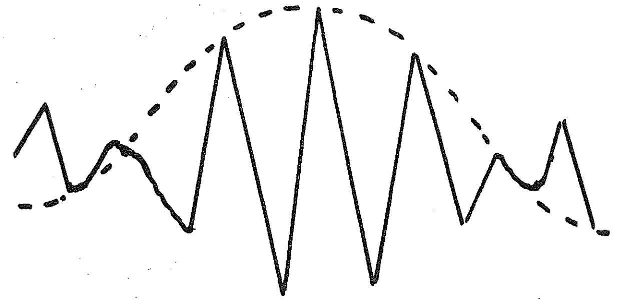
100% AM



BALANCED MODULATION
(or Beating)



OVERMODULATED AM WITH A
TWO QUADRANT MULTIPLIER

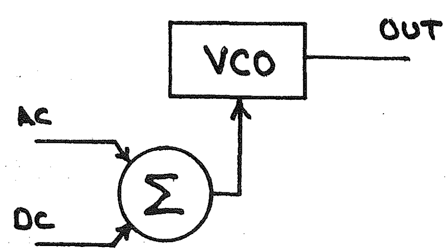


OVERMODULATED AM WITH A
FOUR QUADRANT MULTIPLIER

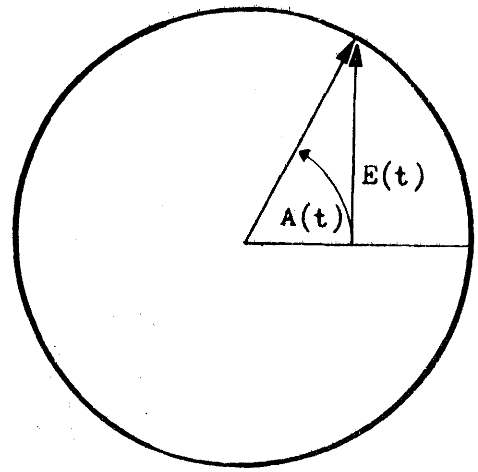
The 10% modulated signal is typical of the tremolo effect. Note also the difference in the envelope shape in the 100% AM and the balance modulated case. Both envelopes are sinusoidal in shape but the AM case has the sinusoid with DC offsets of 1/2, while the balanced modulation case has a single sinusoidal with no DC offset. The balanced modulation envelope is also the envelope that appears for "beating". This is in line with the equivalence of sinewave addition and balanced modulation described above. In the overmodulated AM cases with the two-quadrant multipler (normal VCA), the envelope dips below zero and is ignored for part of the waveform. The modulating signal is therefore equivalent to the part of the envelope above zero, and no longer sinusoidal. This means that the modulating signal is effectively one that has harmonics of the original frequency ω_m. Therefore, additional sidebands appear when the modulation increases beyond 100%, and this is in line with what we know from radio broadcasting, and is the reason that the modulation index is limited to one or less for radio. The AM equation that we started with does not apply because values of $(1 + m \text{ Sin } \omega_m t)$ that are less than one are repalced by zero. However, overmodulation with a four quadrant multiplier does not violate the original AM equation and no additional harmonics are produced in this case.

LINEAR FREQUENCY MODULATION

Frequency modulation (FM) is generally implemented as shown at the right. An initial DC voltage sets the "carrier" and an AC voltage is superimposed on it. We shall assume first that the VCO is linear (unlike most electronic music VCO's) to keep things as simple as possible. The exponential VCO case will be considered later as part of a review article.



The FM problem is essentially a phase modulation and we can consider the VCO output to be a rotating vector. The instantaneous angle of rotation of this vector is the phase. We shall denote this angle as A(t). The time rate of change of phase dA/dt is termed the instantaneous frequency, and is proportional to the control voltage input.



ROTATING VECTOR
MODEL FOR
PHASE MODULATION

$$V_{in} = V_o + V_m \operatorname{Cos}(2\pi f_m t)$$

$$F_{inst} = KV_{in} = KV_o + KV_m \operatorname{Cos}(2\pi f_m t)$$

$$= F_{CL} + \Delta F \operatorname{Cos}(2\pi f_m t) = (1/2\pi) \frac{dA(t)}{dt}$$

This can be integrated to give:

$$A(t) = 2\pi \int_o^t [F_{CL} + \Delta F \operatorname{Cos}(2\pi f_m t')]dt'$$

$$= 2\pi F_{CL} t + \frac{\Delta F}{f_m} \operatorname{Sin}(2\pi f_m t)$$

Thus the signal voltage is

$$E(t) = \operatorname{Sin}[2\pi F_{CL} t + \frac{\Delta F}{f_m} \operatorname{Sin}(2\pi f_m t)]$$ where $\Delta F/f_m = m$, the modulation index

This modulation equation can be converted to the spectral form. First, the identity

$$\operatorname{Sin}(x+y) = \operatorname{Sin}(x)\operatorname{Cos}(y) + \operatorname{Cos}(x)\operatorname{Sin}(y)$$

is applied to give:

$$E(t) = \{\operatorname{Sin}(2\pi F_{CL} t)\cdot\operatorname{Cos}[m \operatorname{Sin}(2\pi f_m t)] + \operatorname{Cos}(2\pi F_{CL} t)\cdot\operatorname{Sin}[m \operatorname{Sin}(2\pi f_m t)]\}$$

Next, it is necessary to bring in identities involving the "Bessel functions" in order to handle the cosine of a sine and sine of a sine terms. The identities are:

$$\operatorname{Sin}[m \operatorname{Sin}(x)] = 2[J_1(m)\operatorname{Sin}(x) + J_3(m)\operatorname{Sin}(3x) + J_5(m)\operatorname{Sin}(5x) + \dots]$$

$$\operatorname{Cos}[m \operatorname{Sin}(x)] = J_o(m) + 2[J_2(m)\operatorname{Cos}(2x) + J_4(m)\operatorname{Cos}(4x) + \dots]$$

This results in many terms of the type $\operatorname{Sin}(x)\cdot\operatorname{Cos}(y)$ which have to be multiplied out according to the identity:

$$\operatorname{Sin}(x)\cdot\operatorname{Cos}(y) = (1/2)[\operatorname{Sin}(x+y) + \operatorname{Sin}(x-y)]$$

or $$\operatorname{Cos}(x)\cdot\operatorname{Sin}(y) = (1/2)[\operatorname{Sin}(x+y) - \operatorname{Sin}(x-y)]$$

The final spectral equation becomes:

$$E(t) = J_o(m)\ \mathrm{Sin}(2\pi F_{CL} t)$$

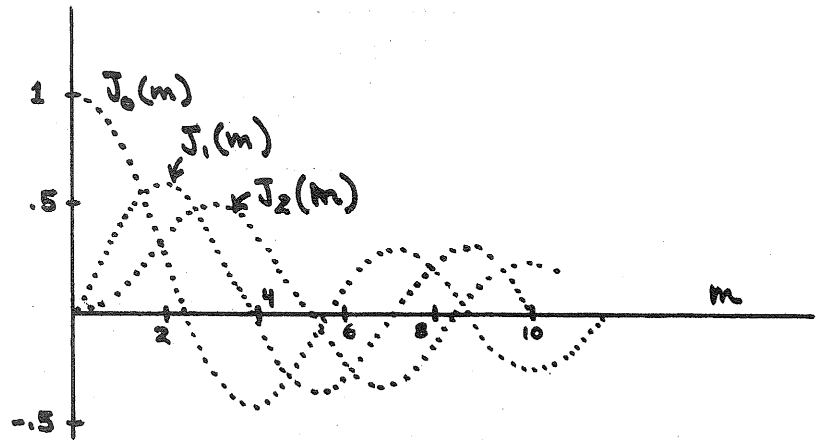
$$+ J_1(m)\{\mathrm{Sin}[2\pi(F_{CL}+f_m)]t - \mathrm{Sin}[2\pi(F_{CL}-f_m)]t\}$$

$$+ J_2(m)\{\mathrm{Sin}[2\pi(F_{CL}+2f_m)]t + \mathrm{Sin}[2\pi(F_{CL}-2f_m)]t\}$$

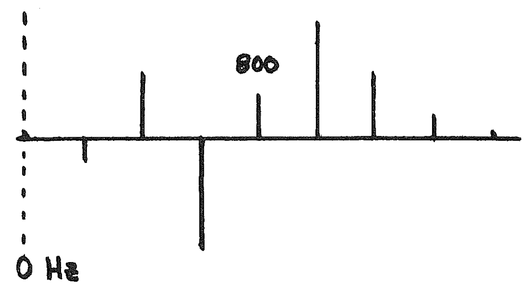
$$+ J_3(m)\{\mathrm{Sin}[2\pi(F_{CL}+3f_m)]t - \mathrm{Sin}[2\pi(F_{CL}-3f_m)]t\}$$

$$+ \ . \ . \ . \ . \ . \ . \ . \ .$$

The Bessel functions J_n are standard tabulated functions. [see reference tables in the back of this book.] The first few Bessel functions are plotted below:

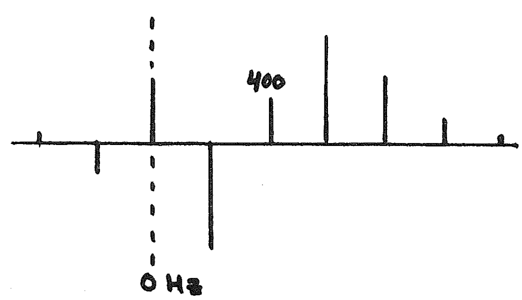


As an example, suppose the carrier (F_{CL}) is set at 800 Hz, and a modulating voltage is applied to drive the frequency up and down by 400 Hz at a rate of 200 Hz. What is the spectrum? The modulation index $m = \Delta F/f_m$ is 400 Hz/200 Hz =2. The Bessel functions for m=2 are: $J_o = .22$, $J_1 = .58$, $J_2 = .35$, $J_3 = .13$, $J_4 = .03$, and higher orders are insignificant. This gives a spectrum as indicated at the right. Note that this happens to be a harmonic spectrum - all the sidebands fall on multiples of 200 Hz. The resulting sound has a pitch of 200 Hz and overtones extending to 1600 Hz.



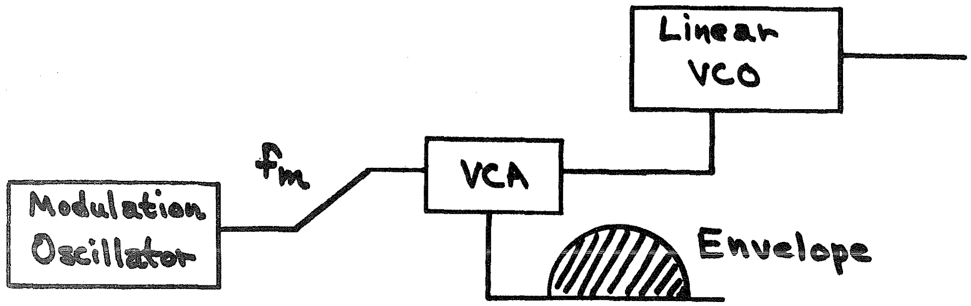
Note however that if we raise the carrier to 811 Hz for example, the sidebands follow unchanged except they are all 11 Hz higher. The spectrum in this case consists of components that are integer multiples of 1 Hz. This is too low to be heard as a pitch, and the resulting spectrum seems non-harmonic. The simple criterion for a harmonic spectrum is that F_{CL}/f_m should be a small integer ratio.

If we were to change the carrier to 400 Hz, we would have a spectrum as shown at the right. The important new feature here is that sidebands on the lower side fall below zero. What this actually means is that they will be "reflected" to positive positions. If the spectrum is harmonic, they will fall on the positions of unreflected lower sidebands (and possibly also the carrier and upper sidebands), and the resulting amplitude at that spectral position will depend on the actual phase



between the carrier and the modulating signal. In an analog system, it is difficult to set this phase exactly, so it is not too useful to consider the actual amplitudes in such cases. If the spectrum is not harmonic, the reflected sidebands do not fall on occupied positions in the positive spectrum, and there is no problem determining amplitudes.

Chowning [JAES 21 #7 (Sept. 1973)] described a method of dynamic spectrum generation where the modulation index was made to vary in time in response to a controlling envelope. Increasing the modulation index is a matter of increasing ΔF (increasing the amplitude of the modulating voltage). A setup using a linear VCO to do this is shown in the diagram below:



As the envelope rises, the modulation index increases, more sidebands appear on the flanks of the spectrum, and the tone quality becomes richer. When set for a non-harmonic spectrum, some very good bell sounds and other percussive sounds can result. When set for a harmonic spectrum, the modulation index controls the spectral evolution of the sound. Thus we can define a synthesized sound in terms of the modulation parameters and the envelope that controls the modulation depth. Chowning was able to describe these parameters for some traditional instrument immitations.

So far, we have been discussiong only sinewave modulation and sinewave outputs of VCO's in the modulation process. What happens when we try other waveforms? First of all, if we use a waveform other than a sine for the VCO output, we can expect to have a full set of sidebands about each component in the complex waveform. This is a simple matter of the superposition of sinewave oscillators. On the other hand, if we use a complex waveform for the modulating voltage, things are not so simple. To see why, we must first consider "multitone" modulation - modulation by more than one sinewave (not necessarily harmonically related). If for example the sinewaves have frequencies f_1 and f_2 then we get sidebands spaced at intervals of f_1 and sidebands spaced at f_2, but also sidebands spaced at f_1+f_2 and f_1-f_2. [see for example, M. Crosby, RCA Review 3 p. 103,(1938)]. The result for additional sinewaves compounds, and sidebands appear spaced at every possible combination of sinewave frequencies. Now, for harmonically related sidebands, the sum and difference terms will fall on existing harmonics. Thus the contribution for any one sideband may come from a combination of several sources. While no essentially new features will be added to the spectrum, the calculation of sideband amplitudes will be greatly complicated. Thus, this situation would generally be avoided. The interested reader can get a feeling for the calculation process from the review article that follows as an equivalent calculation is necessary for the calculation of the FM spectrum of an exponential VCO.

EXPONENTIAL FREQUENCY MODULATION

Since most of the VCO's that are used in electronic music are exponential, it is necessary to consider the FM process when applied to such a VCO. A full discussion of the necessary calculations is given in the review article reprinted from JAES 23 #3, April 1975. The main point that should be grasped is that if you want to use dynamic depth FM (as used by Chowning in the above description), it is necessary to have some sort of linear control added to the exponential VCO to keep things managable.

Reprinted from J. Aud. Eng. Soc. 23 No. 3, April 1975 with permission of the author

The Frequency Modulation Spectrum of an Exponential Voltage-Controlled Oscillator*

BERNARD A. HUTCHINS, JR.

Electronotes, Ithaca, N. Y. 14850

As the amplitude of a modulating voltage to an exponential VCO is increased, the pitch of the modulated signal rises, making dynamic depth FM unrealistic. The pitch rise is proportional to I_0, the first of the modified Bessel functions. Higher order terms involving additional modified Bessel functions can be used to compute the entire spectrum. Various methods of correcting for the pitch shift are possible, but the most useful solution to the dynamic depth FM problem with exponential VCO's is to add some form of auxiliary linear control.

INTRODUCTION: Exponential voltage-controlled oscillators (E-VCOs) are those which produce a frequency which is an exponential function of the control voltage. Ever since the advantages of E-VCOs in musical systems were pointed out [1], it has been realized, and at times noted [2] that the straightforward frequency modulation (FM) patch would not produce a spectrum that could be calculated from the well understood FM radio broadcasting theory. The radio equations apply to a linear voltage-controlled oscillator (L-VCO), not an E-VCO.

SIMPLE MODULATION PATCHES

However, imperfect understanding of the exact details of the E-VCO FM spectrum does not prevent use of the method. For example, the patch shown in Fig. 1 has proven useful for the production of clangorous sounds which have a tone quality that does not vary with pitch. Recently, Chowning [3] used a digital computer method to demonstrate an FM synthesis method where the depth of modulation changes dynamically as a tone progresses. The method employs the equivalent of an L-VCO. In the apparently straightforward realization of this dynamic depth FM method using standard analog music synthesizers (hence E-VCOs), one tries a patch such as the one in Fig. 2. The amplitude of the modulating signal is controlled by a voltage-controlled amplifier (VCA) that is in turn controlled by an approp-

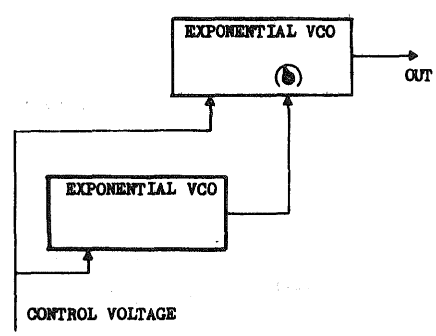


Fig. 1. Patch for clangorous sounds.

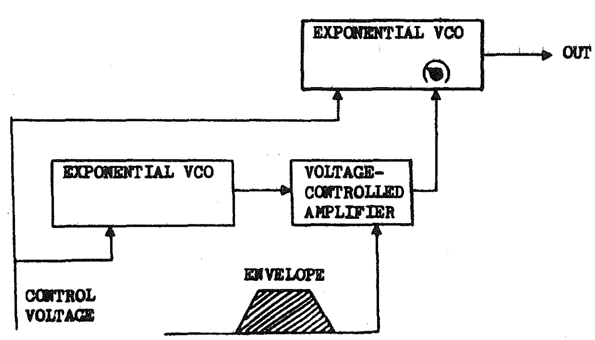


Fig. 2. Dynamic depth FM patch that produces pitch change.

* An earlier version of the paper was presented September 12, 1974, at the 49th Convention of the Audio Engineering Society, New York.

riate envelope. A problem immediately arises in that the E-VCO spectrum undergoes an overall pitch shift as modulation depth changes. While this can be a useful special effect, in general it is disconcerting.

CALCULATING THE E-VCO SPECTRUM

Attempts to use a dynamic depth FM method have made consideration of the E-VCO FM problem more important. A thorough analysis of the problem is useful for electronic music engineers as it both defines the problem and suggests solutions, and will also provide the engineer with a means of fielding questions from musicians who note the unusual behavior of their synthesizers.

The E-VCO FM problem is, like all FM problems, actually a phase modulation problem. This can be visualized as a rotating vector of constant magnitude, where the orientation angle of the vector is $A(t)$ as shown in Fig. 3. The signal voltage is then proportional to sin $A(t)$. The time rate of change of $A(t)$ is then proportional to what is termed the instantaneous frequency:

$$2\pi F_{\text{inst}} = dA(t)/dt. \tag{1}$$

In the L-VCO case, the instantaneous frequency is proportional to the control voltage $V(t)$:

$$F_{\text{inst}} = f_0 V(t). \tag{2}$$

For the standard E-VCO (1 volt per octave), the control voltage appears in the exponent, which has a base of two:

$$F_{\text{inst}} = f_0 2^{V(t)} = f_0\, e^{\ln 2 \cdot V(t)}. \tag{3}$$

It is convenient to start with a control voltage that varies as a cosine:

$$V(t) = V_0 + V_m \cos(2\pi f_m t) \tag{4}$$

where V_m is the amplitude of the modulating voltage and f_m is the modulating frequency. In the L-VCO case, this gives a constant-center frequency (the carrier) and a modulation depth ΔF.

$$\frac{1}{2\pi}\,\frac{dA(t)}{dt} = F_{\text{inst}} = f_0 V_0 + f_0 V_m \cos 2\pi f_m t$$

$$= F_{\text{CL}} + \Delta F \cos 2\pi f_m t. \tag{5}$$

This can be easily integrated to give the well-known FM radio equation:

$$E(t) = \sin A(t) = \sin\left[2\pi F_{\text{CL}} + \frac{\Delta F}{f_m}\sin 2\pi f_m t\right]t. \tag{6}$$

In the E-VCO case, cosine modulation gives:

$$F_{\text{inst}} = f_0 \exp\{\ln 2\cdot(V_0 + V_m \cos 2\pi f_m t)\}$$

$$= F_{\text{CE}} \exp\{\ln 2\cdot V_m \cos 2\pi f_m t\}. \tag{7}$$

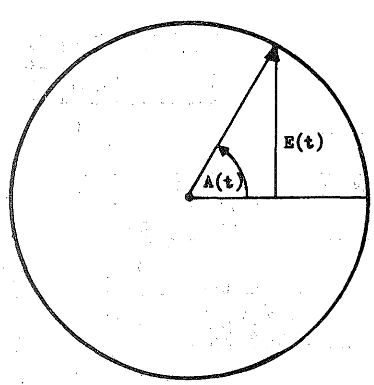


Fig. 3. Rotating vector model of phase modulation.

The exponential factor can be expanded according to the series [4]

$$e^{z\cos\theta} = I_0(z) + 2\sum_{k=1}^{\infty} I_k(z)\cos k\theta \tag{8}$$

where the Fourier coefficients are fortunately tabulated functions known as modified Bessel functions, or hyperbolic Bessel functions. The E-VCO expression for F_{inst} can then be integrated as in the L-VCO case to give $A(t)$, and $E(t)$ then becomes

$$E(t) = \sin\,[\,2\pi\; F_{\text{CE}}\, I_0\,(\ln 2\cdot V_m)t$$

$$+ \sum_{k=1}^{\infty} (2\cdot F_{\text{CE}}\,/\,k\cdot f_m)\, I_k\,(\ln 2\cdot V_m)\sin 2\pi\, k f_m\, t]. \tag{9}$$

The first five modified Bessel functions are shown in Fig. 4. The modified Bessel functions are related to the better known Bessel functions of the first kind J_n by the relation

$$I_n(z) = i^{-n} J_n(iz). \tag{10}$$

I_0 is very important as it will eventually be seen to determine the overall position of the FM spectrum. The higher order I_n terms are an indication of the number of terms that must be retained for a given degree of accuracy in the spectral calculations.

A form of the FM equation that is more useful than Eq. (6), or Eq. (9) in the E-VCO case, is the one that reveals the spectrum of the signal, as this is the one most easily related to the hearing process. Conversion of the modulation equation to a spectral equation is well known in the L-VCO case [5]. A modulation index m is defined as $\Delta F/f_m$, and a series of five identities, Eqs. (11)–(15), are applied to Eq. (6):

$$\sin(x+y) = \sin(x)\cdot\cos(y) + \cos(x)\cdot\sin(y) \tag{11}$$

$$\sin[m\,\sin(x)] = 2[J_1(m)\sin(x) + J_3(m)\sin(3x) + J_5(m)\sin(5x) + \cdots] \tag{12}$$

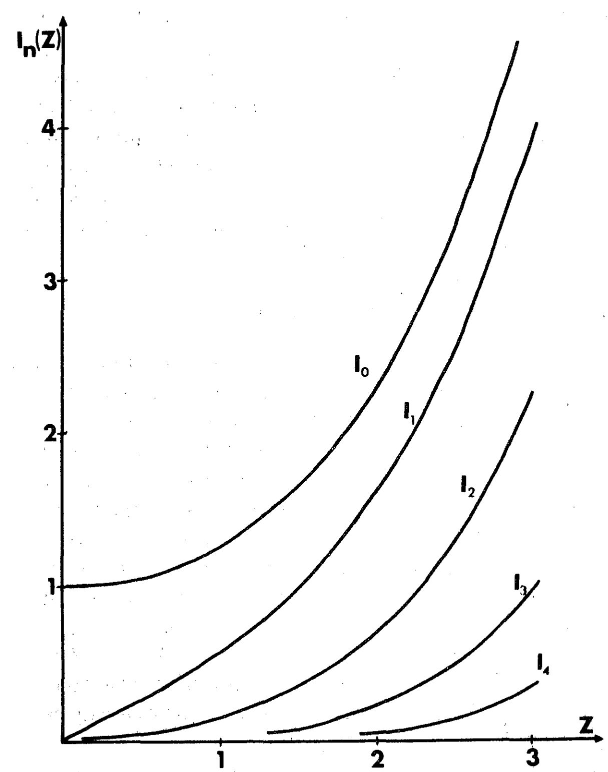


Fig. 4. First five modified Bessel functions.

$$\cos[m \sin(x)] = J_0(m) + 2\,[J_2(m)\cos(2x) + J_4(m)\cos(4x) + \cdots] \tag{13}$$

$$\sin(x)\cdot\cos(y) = \tfrac{1}{2}[\sin(x+y) + \sin(x-y)] \tag{14}$$

$$\cos(x)\cdot\sin(y) = \tfrac{1}{2}[\sin(x+y) - \sin(x-y)] \tag{15}$$

The spectral equation then becomes

$$\begin{aligned} E(t) = {} & J_0(m) \sin 2\pi F_{CL}\, t \\ & + J_1(m)\,[\sin 2\pi\,(F_{CL} + f_m)t \\ & - \sin 2\pi\,(F_{CL} - f_m)t] \\ & + J_2(m)\,[\sin 2\pi\,(F_{CL} + 2f_m)t \\ & + \sin 2\pi\,(F_{CL} - 2f_m)t] \\ & + \cdots . \end{aligned} \tag{16}$$

The J_n are Bessel functions of the first kind. The spectral equation thus shows that a series of sidebands are formed about the carrier, and since $J_{-n} = (-1)^n J_n$, the spectrum is symmetric about the carrier with respect to the absolute values of the amplitudes. Conservation of spectrum energy is demonstrated by the Bessel function identity

$$1 = J_0^2 + 2 \sum_{k=1}^{\infty} J_k^2 . \tag{17}$$

A typical spectrum is shown in Fig. 5 for the L-VCO case.

In the E-VCO case the calculations are greatly complicated by the additional terms. Consider first a limited case where I_2 is still negligible. The E-VCO modulation equation, Eq. (9), then becomes

$$\begin{aligned} E(t) = {} & \sin\,[2\pi F_{CE} I_0\,(\ln 2\cdot V_m)t \\ & + (2F_{CE}/f_m)\, I_1(\ln 2\cdot V_m) \sin 2\pi f_m t]. \end{aligned} \tag{18}$$

This has the same mathematical form as the L-VCO equation with

$$F_{CL} \rightarrow F_{CE} I_0\,(\ln 2\cdot V_m) \tag{19}$$

$$m \rightarrow (2F_{CE}/f_m)\, I_1(\ln 2\cdot V_m) \cong 0.69\, V_m\,(F_{CE}/f_m). \tag{20}$$

In this linear approximation, which is quite good for V_m less than half a volt, sideband positions and amplitudes about the carrier are calculated according to the L-VCO equations, and then placed about a carrier that has slid up the I_0 curve according to the magnitude of V_m.

It turns out that even when more terms are considered in the E-VCO equation, the entire spectrum still shifts along the I_0 curve as shown in Fig. 6. This replot of the I_0 function has additional labels to show the $\ln 2\cdot V_m$ axis, as well as the total frequency deviation and spectrum slide in units of semitones. It is the shift implied by this curve that is responsible for the (generally) annoying pitch variation that occurs as modulation depth changes while a tone progresses.

When three or more terms must be kept, calculations must go beyond the simple expansion about the L-VCO solution. For example, when I_4 can be neglected, but not I_3, the modulation equation becomes

$$\begin{aligned} E(t) = {} & \sin\,[\,2\pi F_{CE} I_0\,(\ln 2\cdot V_m)t + m_1 \sin 2\pi f_m t \\ & + m_2 \sin 4\pi f_m t + m_3 \sin 6\pi f_m t] \end{aligned} \tag{21}$$

where

$$m_1 = (2F_{CE}/f_m)\, I_1(\ln 2\cdot V_m) \tag{22a}$$

$$m_2 = (F_{CE}/f_m)\, I_2(\ln 2\cdot V_m) \tag{22b}$$

$$m_3 = (2F_{CE}/3f_m)\, I_3(\ln 2\cdot V_m). \tag{22c}$$

It is interesting that this equation has the same form that is obtained by considering the simultaneous modulation of a L-VCO with more than one sine wave. This can be understood by considering that the E-VCO expression for F_{inst} could be duplicated by applying an appropriate periodic waveform to the L-VCO. In the E-VCO case, the additional sine waves are all harmonically related, but the more general problem was considered as far back as 1938 [6]. The solution consists of the application of Eq. (11) and its corresponding identity

$$\cos(x+y) = \cos(x)\cdot\cos(y) + \sin(x)\cdot\sin(y) \tag{23}$$

to Eq. (21), considering the first term as x and the remaining terms as y, and repeating this process until all the terms are used up. The final result is [7]

$$\begin{aligned} E(t) = {} & \sum_{k_s=-\infty}^{\infty} \left[\prod_{s=1}^{S} J_{k_s}(m_s) \right] \\ & \times \sin\,[2\pi I_0\,(\ln 2\cdot V_m)F_{CE} + \sum_{s=1}^{S} 2\pi k_s f_m\,]\,t. \end{aligned} \tag{24}$$

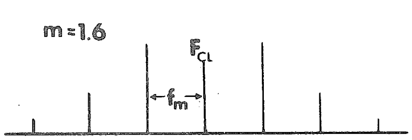


Fig. 5. A typical L-VCO spectrum.

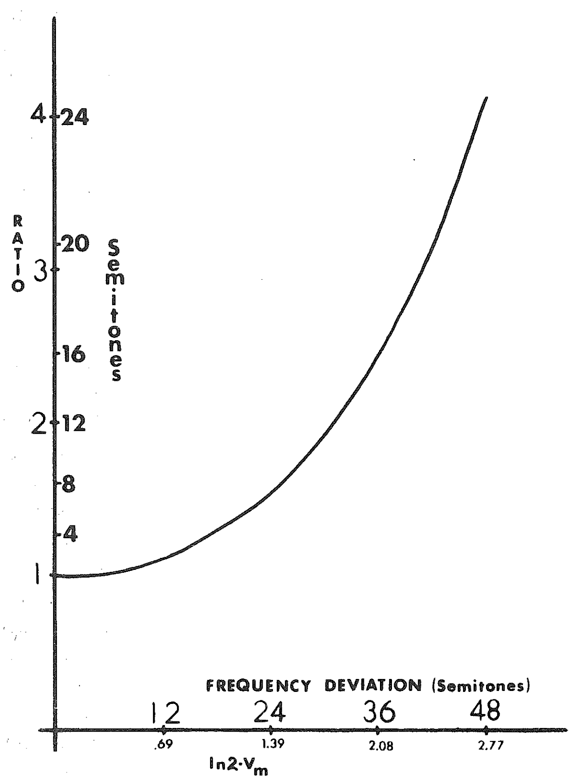


Fig. 6. Shift of spectrum with increasing modulation voltage V_m.

For the four-term case, this equation can be greatly reduced:

$$E(t) = \sum_{L=-U}^{U} \sum_{M=-U}^{U} \sum_{N=-U}^{U} \Big[J_L(m_1) J_M(m_2) J_N(m_3)$$

$$\times \sin 2\pi \,[I_0 (\ln 2 \cdot V_m) F_{CE}$$

$$+ Lf_m + 2Mf_m + 3Nf_m]\, t \Big]. \qquad (25)$$

In the above equation, U is the highest order L index which must be considered, corresponding to the smallest J_L which is still significant. To further simplify the calculations, consideration of the overall spectrum shift can be set aside and added in later. For computer calculations, the procedure is then as follows.

1) Select V_m and the ratio F_{CE}/f_m.

2) Determine all significant values of $I_n(\ln 2 \cdot V_m)$.

3) From the I_n values, calculate $m_1, m_2, m_3, \cdots$ and from these, calculate all significant values of $J_n(m_k)$.

4) Execute a nested "do loop" for each summation, three in the case of Eq. (25), and inside the innermost do loop

a) Determine the sideband in question: $L + 2M + 3N$;

b) Calculate $J_L(m_1)\, J_M(m_2)\, J_N(m_3)$ and add this to any previous contribution for the same sideband.

5) Reposition the sidebands. Space them at intervals of f_m about the carrier (zeroth sideband) shifted according to $I_0(\ln 2 \cdot V_m)$.

6) Any sideband that has a negative frequency should be changed to a positive frequency and be considered reflected back into the positive spectrum according to $\sin(-x) = -\sin(x)$.

EXPERIMENTAL VERIFICATION

These theoretical calculations can be verified with an experimental setup employing a spectrum analizer. A redrawn version of an experimental spectrogram is shown in Fig. 7. The predicted spacing of the sidebands at intervals of f_m and the predicted spectrum slide is observed. Also, the linearlike structure for $V_m = 0.5$ and the much more complicated pattern for $V_m = 1$ can be seen. In particular, there are more significant sidebands above the carrier than below.

Fig. 8 shows a plot of calculated and experimental sideband amplitudes for some 47 sidebands from six different E-VCO spectra. The agreement is within expected experimental and computational accuracy.

EXAMPLE SPECTRA

With this experimental verification completed, calculations can be viewed with more confidence. Fig. 9 shows a complete series of calculated amplitudes for the ratio $F_{CE}/f_m = 8$. When the ratio is lowered to 2.0, it is easier for significant sidebands to reach zero frequency and below. These sidebands are reflected back into the positive spectrum (and are experimentally observed as predicted). Such sidebands are very common in the L-VCO case, but are relatively rare in the E-VCO case for two reasons. First, the instantaneous frequency can never reach zero, and in general there are very few significant sidebands produced outside the range of F_{inst}. Second, as modulation depth increases and more and more sidebands start to appear at the extremes of the spectrum, the spectrum shifts up, pulling the lower ones away from zero. These lower sidebands often begin to become significant while in their reflected positions. In this case, as the spectrum shifts up, these sidebands are pulled down and around zero where they reemerge as normal lower sidebands. Sidebands behaving in this manner

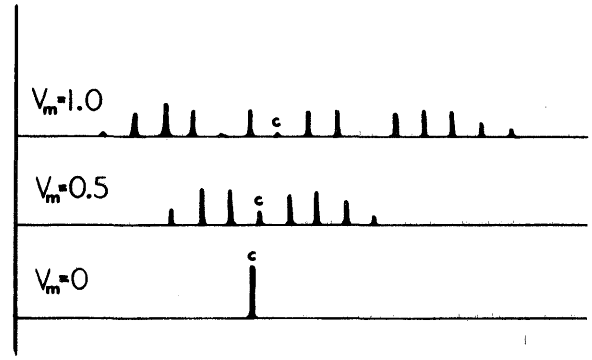


Fig. 7. Experimental spectrogram for $F_{CE}/f_m = 8$; $V_m = 0, 0.5,$ and 1.0

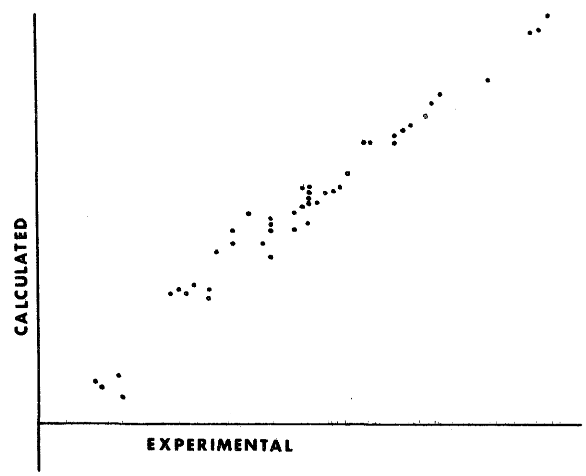


Fig. 8. Comparison of calculated and experimental sideband amplitudes.

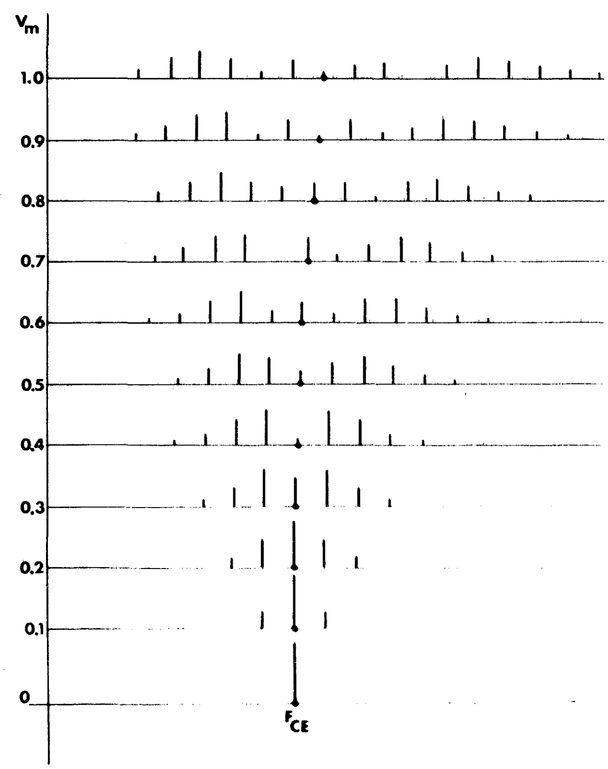


Fig. 9. Spectra for $F_{CE}/f_m = 8$.

can be seen in Fig. 10 for $F_{CE}/f_m = 2$, and in Fig. 11 for $F_{CE}/f_m = ½$.

In the process of calculation, the power in each sideband, which goes as the amplitude squared, is easy to compute and tabulate. A surprising result is noted: while there are more significant sidebands above the carrier, there is more power produced below. Both these results can perhaps be understood in terms of the graph of the E-VCO instantaneous frequency of Fig. 12. The instantaneous frequency goes much higher above the carrier than below, and therefore passes over more possible sideband positions on the high side, hence activating more of them. Although the carrier slides up with increased modulation depth (from the solid to the dotted line), the points at which the instantaneous frequency crosses at equal time intervals remain on the original carrier line. Therefore, the instantaneous frequency spends more time below the shifted carrier than above, and more power is distributed to the lower sidebands, even though there are fewer of them.

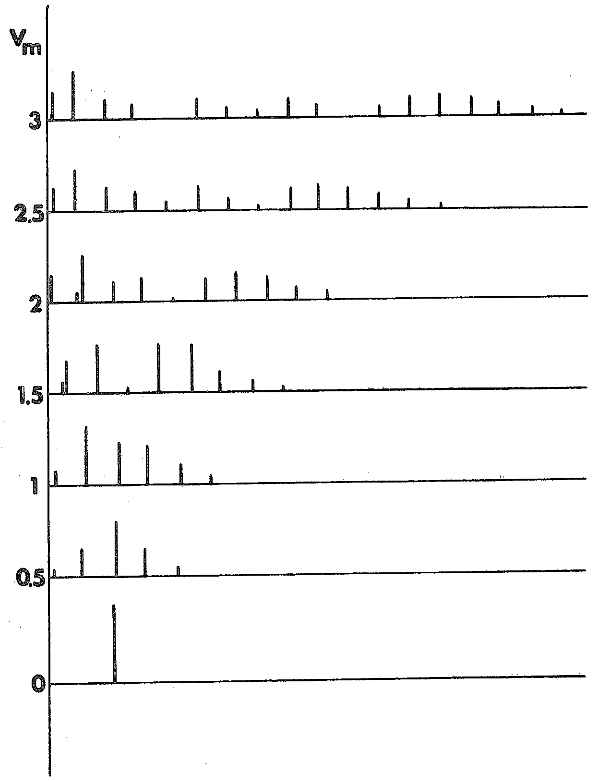


Fig. 10. Spectra for $F_{CE}/f_m = 2$.

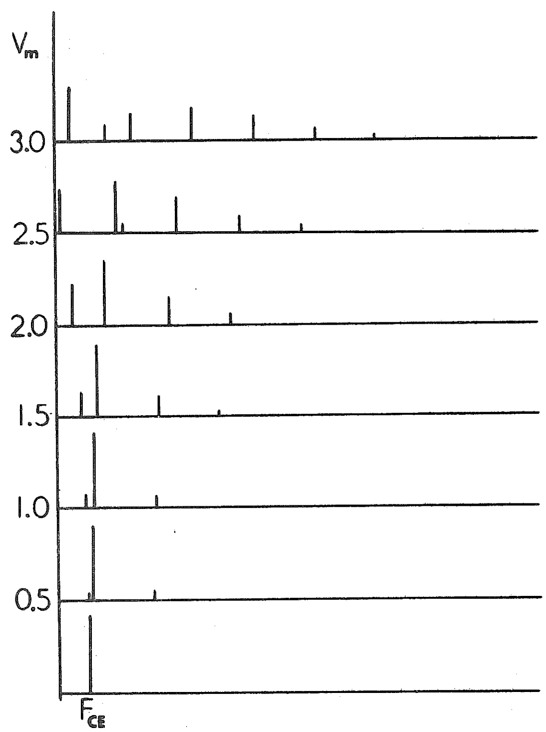


Fig. 11. Spectra for $F_{CE}/f_m = ½$.

Fig. 12 is also helpful in understanding the carrier shift along the I_0 curve. The areas between the F_{inst} curve and the shifted carrier are equal on both sides. This is equivalent to the dc level that would be required to produce the same F_{inst} curve with a L-VCO.

HARMONIC SPECTRA

An interesting application of the FM method is the production of harmonic spectra by allowing one of the sidebands to fall on zero frequency. In this case the carrier, all normal sidebands, and any significant reflected sidebands will fall on positions that are multiples of a common fundamental. In the E-VCO case, the condition for a harmonic spectrum is

$$F_{CE}/f_m = \frac{1}{I_0(\ln 2 \cdot V_m)} \cdot \frac{n_1}{n_2}$$

where n_1 and n_2 are integers. A typical harmonic spectrum is shown in Fig. 13. In the E-VCO case, unlike the L-VCO case, the condition for a harmonic spectrum is a function of V_m, making the condition impossible to maintain during dynamic depth modulation. A zero-frequency component in the harmonic spectrum actually just occurs as a dc weighting of the waveform. The actual amplitudes obtained for a harmonic spectrum depend on the relative phase of the original carrier and the modulating waveform, since reflected sidebands will have a phase that depends on these initial conditions. Control of this phase is difficult with ordinary synthesizers.

SUGGESTED SOLUTIONS

The theory presented has allowed the calculation of the E-VCO FM spectrum, and outlined the cause of the pitch shift problem encountered during dynamic depth FM. Several solutions to this problem are suggested.

A patch such as the one in Fig. 14 can be tried. Here the envelope controlling the modulation depth is inverted and used to pull the pitch back down as the modulation depth increases. When this inverted voltage is fed directly to the VCO, correction is

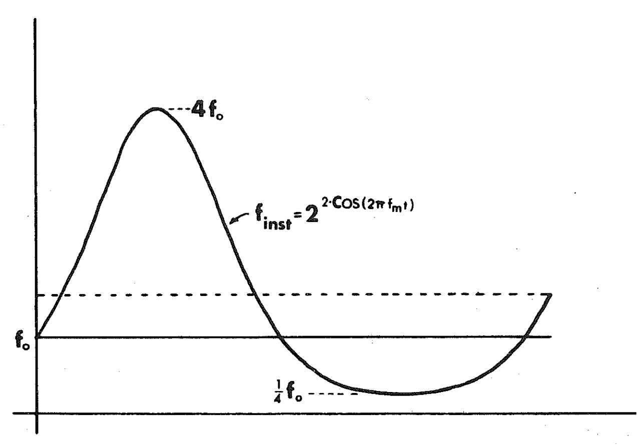


Fig. 12. Instantaneous frequency for one cycle of modulation.

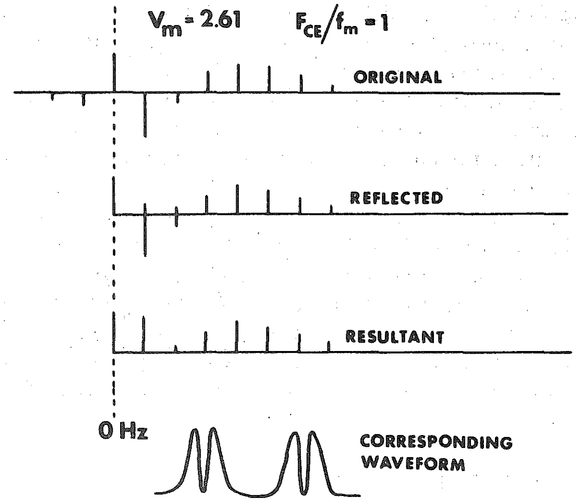


Fig. 13. Harmonic spectrum; $V_m = 2.61$, $I_0(\ln 2 \cdot V_m) = 2$, $F_{CE} / f_m = 1$, and corresponding waveform.

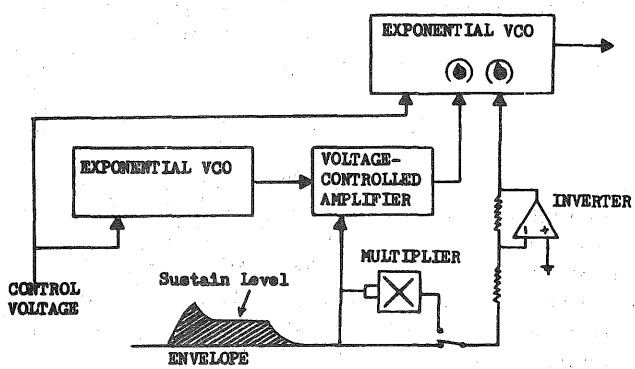


Fig. 14. Patch used to correct for pitch shift.

incomplete as the upward response to the envelope through the VCA is along the I_0 curve, while the direct path is the normal exponential. With the direct connection, complete correction is obtained only at one voltage level of the envelope, typically set at the sustain level. Incomplete correction during the attack phase often adds spectral features that enhance the musical quality of the tones. The same incomplete correction during the final decay phase, however, causes pitch variations that are generally bothersome. This can be partially remedied by using the same envelope for modulation depth control and amplitude control, and truncating the exponential tail early.

The correction can be completed to first order by squaring the envelope voltage before inverting and feeding back to the VCO. The squaring operation can usually be done with an available multiplier or "ring modulator" on the synthesizer. The normal exponential response is 2^V, so the response to V^2 goes as $1 + \ln 2 \cdot V^2 + \cdots$ or approximately $1 + 0.69\, V^2$. The $I_0(\ln 2 \cdot V_m)$ curve can be approximated by $1 + 0.119\, V_m^{\,2} + \cdots$, so by adjusting these levels, a degree of correction can be obtained that is quite satisfactory for small modulation depths. Since a correction voltage is fed to the VCO which normally sets F_{CE} according to initial control voltages before modulation is applied, the ratio F_{CE} / f_m varies dynamically as well, and this further complicates analysis.

A second and easier to use solution is to provide a linear control input to the E-VCO to supplement normal exponential controls. Typically, an E-VCO is basically a voltage-controlled exponential current source fed to a current-controlled ramp generator (capacitor with some form of discharge, or the current can be reversed). The exponential current source uses the exponential collector current to base-emitter voltage (proportional to control voltage) response of transistors [8]. If the standing current in these exponential current sources is linearly modulated, linear FM results. This modulation is before the exponential conversion, but is a collector current modulation, not a base-emitter voltage modulation. Thus the modulation remains linear at the output; only the limits of the current excursion depend on the base-emitter voltage. If the modulating frequency is tracking the original carrier, the total excursion for fixed modulation depth of the standing current results in a constant modulation index, since both ΔF and f_m are proportional to the control voltage. This constant modulation index linear FM gives a constant sideband structure relative to the carrier and is thus a constant timbre form of linear FM, and outwardly seems the most musically useful.

Another method is to simply add a linearly modulated current to the exponential current on the way to the current-controlled oscillator. This results in a constant frequency deviation independent of control voltage, and (approximately) constant bandwidth linear FM. This same feature can also be used to offset two otherwise tracking oscillators by a fixed number of Hertz so that the beat rate between them remains independent of frequency.

Both of the above methods will result in linear FM which does not exhibit a pitch shift during dynamic depth FM. A final solution that can be added externally to existing E-VCO's is a standard logarithmic amplifier [8]. Such an amplifier must offset standard bipolar signals so that they are always positive, take Log_2 of the resulting voltage, and feed this directly to a 1 volt per octave control input. This results in constant frequency deviation linear FM, but since it is before the exponential converter, it can not be used for a linear offset. A simple log amp that can be used with existing E-VCO's is shown in Fig. 15. Fig. 16 shows how the three linear FM methods discussed above are applied to a typical E-VCO.

CONCLUSIONS

While it is possible to calculate most of the features of the FM spectrum of an E-VCO, the complexity of the calculation process and the difficulty of setting analog controls to realize a set of conditions makes everyday application of these calculations impractical. However, an understanding of the theory does provide an understanding of the sounds realized by FM techniques on standard synthesizers, and suggests that some sort of linear control should be employed when attempting to use a dynamic depth FM synthesis method. Additional thought should be given to the possibility of reversing the process so that a given spectral evolution could be realized in terms of time-dependent modulation

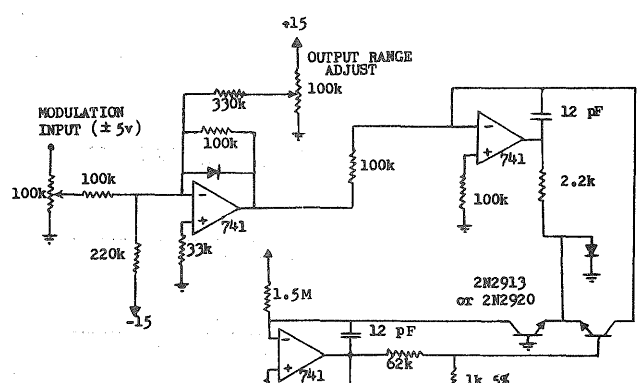


Fig. 15. Logarithmic amplifier for linear input control.

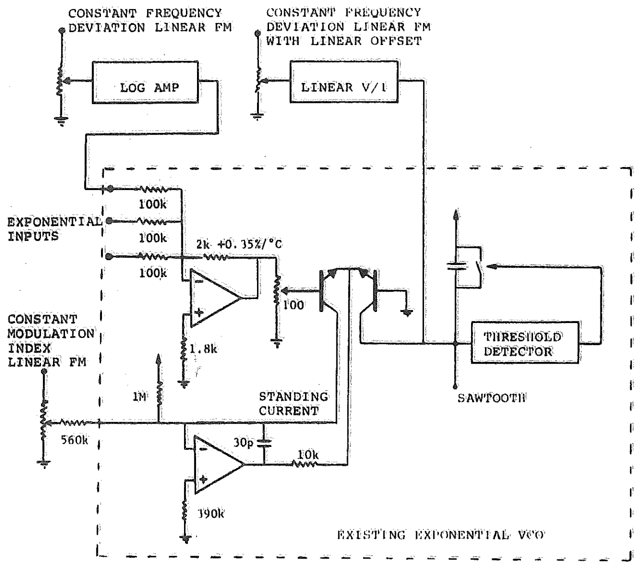


Fig. 16. Typical existing E-VCO illustrating basic methods of adding linear FM.

parameters. In this pursuit, the E-VCO with pitch pull-down correction may prove a more useful approach, as a variety of sideband distributions are possible, unlike the L-VCO which has only the one case. For general use of the dynamic depth method, simple linear controls seem the most useful.

ACKNOWLEDGMENT

The author is grateful to William Hemsath and Gordon Wilcox for valuable assistance, and for the use of equipment at the Experimental Psychology Interactive Computer Facility at Cornell University.

REFERENCES

[1] R. A. Moog, "Voltage-Controlled Electronic Music Modules," *J. Audio Eng. Soc.*, vol. 13, pp. 200-206 (July 1965).

[2] H. S. Howe, Jr., mentioned in review of Allen Strange "Electronic Music: Systems, Techniques, and Controls" in *Perspectives of New Music*, vol. 11, no. 2, p. 249 (Spring-Summer 1973).

[3] J. M. Chowning, "The Synthesis of Complex Audio Spectra by Means of Frequency Modulation," *J. Audio Eng. Soc.*, vol. 21, pp. 526-534 (Sept. 1973).

[4] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions* (Dover Publications, New York, 1965), p. 376, eq. 9.6.34.

[5] W. L. Everitt and G. E. Anners, *Communication Engineering* (McGraw-Hill, New York, 1956), p. 29.

[6] M. G. Crosby, "Carrier and Side-Frequency Relations with Multitone Frequency or Phase Modulation," *RCA Rev.* vol. 3, pp. 103-106 (July 1938).

[7] M. S. Corrington, "Variation of Bandwidth with Modulation Index in Frequency Modulation," *Proc. IRE*, vol. 35, pp. 1013-1020 (Oct. 1947).

[8] R. C. Dobkin, "Logarithmic Converters," National Semiconductor Corp., Santa Clara, Calif., Application Notes AN-30, Nov. 1969.

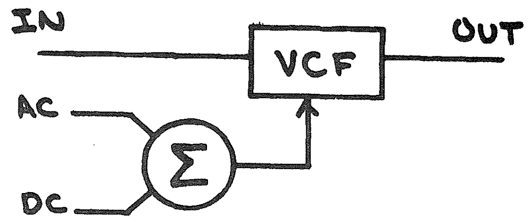
THE AUTHOR

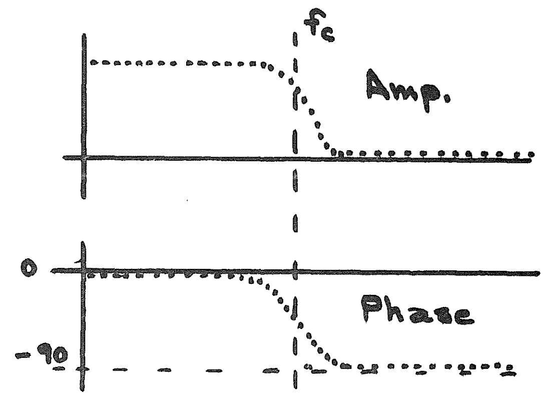
Bernard A. Hutchins, Jr. was born in Rochester, NY in 1945. He received a B.S. degree in Engineering Physics from Cornell University in 1967. After three years of military service, Mr. Hutchins returned to Cornell where he is now completing graduate work. He is also the editor and publisher of *Electronotes*, a newsletter devoted to electronic music and musical engineering.

2c (19)

FORMANT MODULATION

Formant modulation is the alteration of tone color by means of moving the position of a filter characteristic relative to the frequency of an input signal. In practice, the type of patch shown at the right can be used. We will assume that it is the characteristic frequency of the filter that is being altered by the AC component, although other parameters may be varied at times. The effect, particularly when used with a filter response that has an amplitude peak and a low modulation frequency, is known colloquially as a "Wow-Wow." What is actually happening can be quite complex mathematically. Consider for example a fairly simple filter response - the single pole low-pass. The amplitude and phase responses are sketched at the right. When f_c is swept by a voltage, both the amplitude and phase of any spectral component in the input waveform that lies fairly close to f_c will be modulated both in amplitude and in phase at the filter output. [Moog, AES Convention, Sept. 1974 reported measuring the phase change at as much as 1000°/octave when corner peaking of the filter was used]. In addition to the combined phase and amplitude modulation problem, calculations can be expected to be further complicated by (1) The fact that the response curves are not linear. This means that even for sinusoidal modulating signals more than two sidebands will be generated for the AM part of the overall modulation, and sideband amplitudes will be altered for the phase modulation part of the modulation. (2) If an expenential VCF is used, the sideband amplitudes will be further altered. However, the output pitch will not rise as it does in the case of an exponential VCO as modulation depth increases. The filter can not alter the input frequency except by phase modulation, an effect that is either too small (slow sweep) or too rapid (fast sweep and reversing) to be heard as a pitch shift.





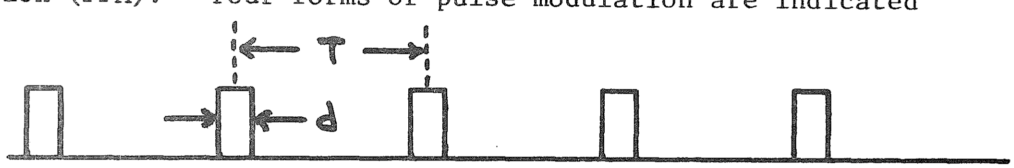
The total result can be expected to be a mathematically complex distribution of sidebands. However, only the computation of amplitudes is really complex. The sideband spacing is the normal modulation frequency, and the sidebands are centered about the original carrier. The subjective effect is in many ways similar to other phase modulations. With complex filter responses however, very rapid changes of phase and/or amplitude can produce some unique effects.

Many different parameters of a filter's response can be made voltage controlled and hence can be modulated. Voltage-controlled Q is a feature of many VCF's. This has the effect of controlling the sharpness of the resonance or of a peaked corner. Thus, the particular harmonic(s) closest to the characteristic frequency undergo amplitude modulation and generally simultaneous phase modulation.

PULSE MODULATION

There are numerous types of pulse modulation that can be used to produce complex spectra. Furthermore, most of these are easily implemented with simple circuitry or with existing synthesizer modules. The various types of pulse modulation we can consider include Pulse Amplitude Modulation (PAM), Pulse Width Modulation (PWM) of several forms, and Pulse Position Modulation (PPM). Four forms of pulse modulation are indicated below:

UNMODULATED PULSE TRAIN

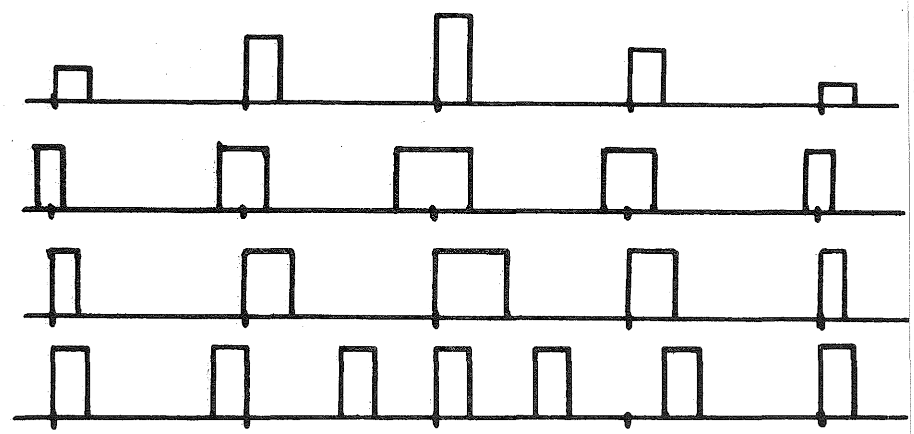


PULSE AMPLITUDE MODULATION

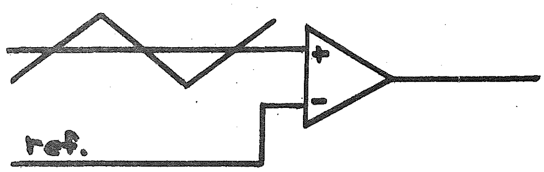
PULSE WIDTH MODULATION (Centered)

PULSE WIDTH MODULATION (Fixed Edge)

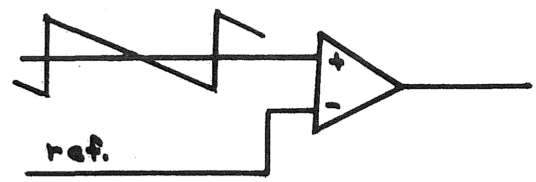
PULSE POSITION MODULATION



Pulse amplitude modulation can be implemented with a VCA, sample-and-hold, and a gating circuit. Pulse width modulation can be implemented with either a fixed center of a fixed edge by changing the basic type of driving waveform from something like a triangle to something like a sawtooth:



CENTERED PWM



FIXED EDGE PWM

Pulse position modulation is implemented by triggering a monostable from a driving waveform and varying the trigger point. For example, the monostable could be triggered when a driving sawtooth crossed a certain reference level. The reference level is then made to be the modulating signal.

Various types of pulse modulation have been studied [e.g., G. M. Russell, Modulation and Coding in Information Systems, Prentice-Hall 1962]. The basic starting point is the unmodulated pulse train represented by its Fourier series:

$$e(t) = \underbrace{Ad/T}_{\text{DC}} + \sum_{n=1}^{\infty} \underbrace{\frac{2A}{n\pi}\,\mathrm{Sin}[n\pi d/T]}_{\substack{\text{Fourier}\\ \text{Coefficient}}}\ \underbrace{\mathrm{Cos}(n\omega_o t)}_{\substack{\text{Fourier}\\ \text{Component}}} \qquad \omega_o = 2\pi/T$$

For the analysis of PAM, the term A is replaced by $A(t) = A(1 + m \mathrm{Cos}\ \omega_m t)$. With this plugged in, it is easy to see that the "1" gives back the original unmodulated pulse train. In addition, two terms are produced by the " $m \mathrm{Cos}\ \omega_m t$" term. The terms are:

$$\frac{Amd}{T}\ \mathrm{Cos}\ \omega_m t$$

and:

$$\sum_{n=1}^{\infty} \frac{2Am}{n\pi}\ \mathrm{Sin}(n\pi d/T)\ [\mathrm{Cos}\ \omega_m t \cdot \mathrm{Cos}(n\omega_o t)]$$

The first term is the modulating signal. The second term is the product of two cosines (balanced modulation) of the modulating signal and each of the Fourier components of the original unmodulated pulse train. These balanced modulation terms when combined with the Fourier components give a "carrier" for each Fourier component and two AM sidebands. The term in Cos $\omega_m t$ results from the DC term in the unmodulated pulse train. It can thus be seen that an alternative analysis could be based on the amplitude modulation of each of the Fourier components in the unmodulated pulse train. This is really no surprise.

When considering PWM, it is the term d that has to be replaced by a time varying term $d(t) = d + d_m \text{ Cos } \omega_m t$. This has assumed the centered form of PWM. Plugging this back into the unmodulated pulse train gives:

$$e(t) = Ad/T + (Ad_m/T) \text{ Cos } \omega_m t$$

$$+ \sum_{n=1}^{\infty} \frac{2A}{n\pi} [\text{Sin}(n\pi d/T + \frac{n\pi d_m \text{ Cos } \omega_m t}{T}) \text{ Cos}(n\omega_o t)]$$

The first term Ad/T is the average DC term while the second is a time varying DC term that varies with ω_m. This can be seen simply as the fact that the duty cycle changes with ω_m and this changes the DC weighting of any small time segment of the signal. The third term is the most interesting. It can be expanded using the identity for Sin(X+Y) and gives:

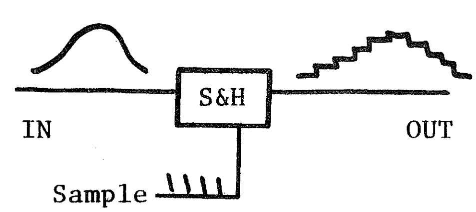
$$\sum_{n=1}^{\infty} \frac{2A}{n\pi} [\text{Sin}(n\pi d/T)\{\text{Cos}(\frac{n\pi d_m}{T} \text{ Cos } \omega_m t)\}] \text{ Cos}(n\omega_o t)$$
$$+ \sum_{n=1}^{\infty} \frac{2A}{n\pi} [\text{Cos}(n\pi d/T)\{\text{Sin}(\frac{n\pi d_m}{T} \text{ Cos } \omega_m t)\}] \text{ Cos}(n\omega_o t)$$

Reference to the earlier discussion of FM will show that the terms within the {} are Bessel series that can be expanded in terms of integer multiples of ω_m. The analysis closely parallels the phase modulation results. It soon becomes evident however that the "bookkeeping" becomes excessive here and that a thorough analysis would require a good deal of effort. Here we will just remark that the problem has been outlined and that it can be seen that analysis is similar to phase modulation. The one additional feature is the presence of a term at the modulating frequency which is not present in the general phase modualtion problem.

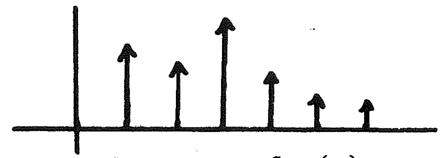
When we move on to fixed edge PWM, it can be seen that we first have to solve the PPM problem since we have to add a displacement to the term T: $T(t) = T + \tau \text{ Cos } \omega_m t$. The problem becomes very complex since the T term appears in the denominator of the pulse train expression. Russell has outlined a solution in an approximation that is not valid for the large modulation depth used in electronic music. Solution to the problem would seem to require a computer to do all the bookkeeping.

TIME SAMPLING

A system for time sampling a waveform is implemented as shown at the right. A sample-and-hold module is used to break a wavefrom into discrete steps. We will be concerned here with the case where both the input waveform and the sampling rate are periodic. We shall be interested in only the resulting spectrum of the output. We shall not be concerned with the recovery of any information.



We start with an input e(t) which we assume is periodic and thus has a spectrum of the general form shown at the right. The waveform e(t) is to be sampled at equal intervals of length T. We thus need a series of short pulses to act as triggers for the sampling unit. For a mathematical model, these pulses can be represented by a series of delta functions:

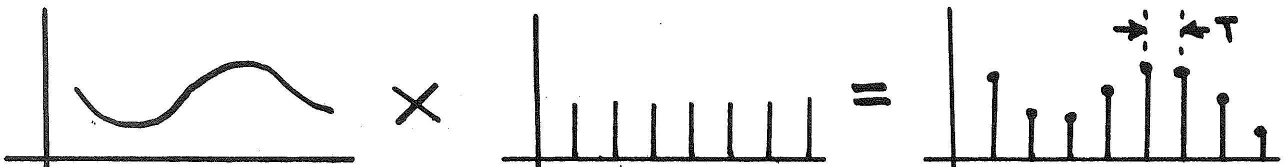


Spectrum of e(t)

$$s(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT)$$

Sampling s(t)

where the function $\delta(t)$ has a value only when t=0. We can diagram the series of sampling commands as shown at the right. We can then think of the sampling process as the multiplication of the input waveform e(t) by the sampling waveform s(t). We will eventually have to add to this a zeroth-order hold circuit to account for the holding action of the S&H. The sampling process as multiplication is shown below:



e(t) s(t) g(t)

where g(t) is the output of the sampler. It is a standard practice in electrical engineering to take Fourier transforms of waveforms to see how one process (multiplication) in the time domain behaves in the frequency domain. We can represent the Fourier transform operator as F and take the Fourier transforms of e(t) and s(t):

$$F[e(t)] = E(\omega) = \int_{-\infty}^{\infty} e(t)\, \exp(-j\omega t)dt$$

$$F[s(t)] = S(\omega) = \omega_o \sum_{n=-\infty}^{\infty} \delta(\omega - n\omega_o) \qquad \omega_o = 2\pi/T$$

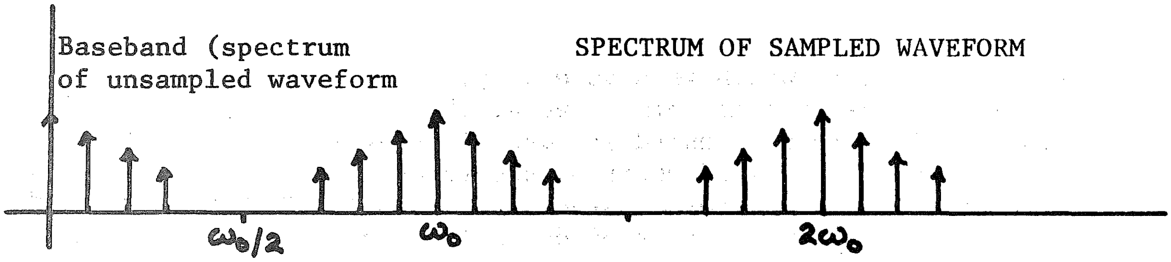
Since multiplication in the time domain becomes convolution in the frequency domain, the expression for the Fourier transform of g(t) is:

$$G(\omega) = F[g(t)] = (1/2\pi)\ [E(\omega)*S(\omega)]$$

where the (*) denotes the convolution. Convolution is an integral of the form:

$$[E(\omega)*S(\omega)] = \int_{\infty}^{\infty} E(x)S(\omega-x)\ dx$$

but it will serve our purposes here to simply note the nature of the spectrum $G(\omega)$ of g(t). The only additional fact we need is that the convolution of a delta function $\delta(t)$ with a function f(t) gives the function f(t). This can be easily seen by the fact that the convolution is essentially one function sliding by the other for each value of delay between them. Since the spectrum of the sampling function $S(\omega)$ is a series of delta functions, The delta functions in the convolution $E(\omega)*S(\omega)$ simply serve to spread out the original spectrum about each multiple of the sampling frequency. The result is indicated below:



Several important points can be noted from the drawing above. First of all, the basis of the sampling theorem can be seen. If the baseband contains frequencies only up to 1/2 ω_o, then there is no overlap from the spectrum spread about ω_o back down into the baseband. Therefore, a low-pass filter with cutoff at 1/2 ω_o will recover the information in the unsampled waveform. The second point is that while it is generally thought that the sampling frequency itself appears in the spectrum of the sampled waveform, this is only true if there is a DC component in the baseband itself. Finally, we should note that the spread of frequencies due to multiplication of waveforms is nothing really new. We saw the same thing when we studied balanced modulation. In fact, an alternative formulation of the balanced modulation results can come from the fact that time multiplication leads to frequency convolution.

The next thing we want to consider is the effect of the hold part of the circuit. Intuitively we can see that it will have an effect of reducing the high frequency content in the waveform. This is because it slows down sharp variations due to the holding action. Formally we have to look at the transfer function of the zeroth-order hold:

$$H(j\omega) = \frac{1 - e^{-j\omega/\omega_o}}{j\omega}$$

A series of trig. manipulations give the magnitude of this transfer function as:

$$|H(j\omega)| = (2/\omega)\ \mathrm{Sin}(\omega/2\omega_o)$$

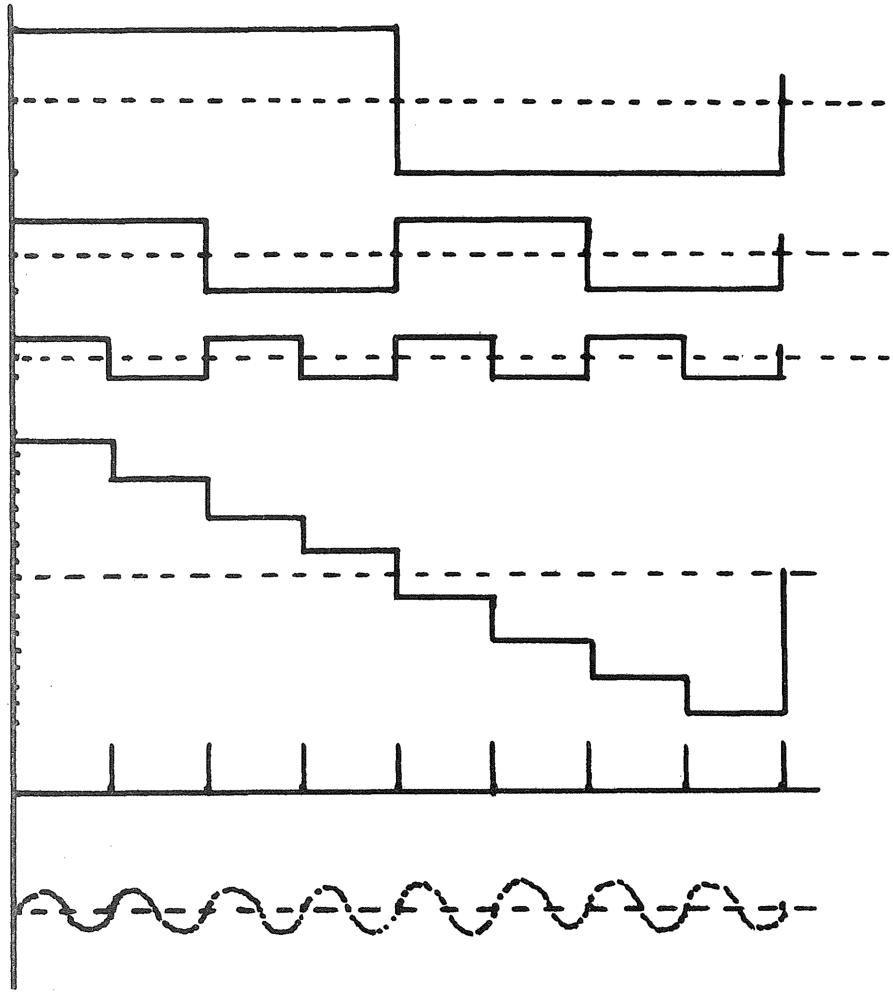
From this we see the result of the spectrum falling off as $1/\omega$. The other part of the magnitude function is a sinusoidal envelope varying as ω. This is the same as the envelope function of a pulse waveform.

We can thus summarize the results of time sampling as: (1) The duplication of the baseband spectrum about multiples of the sampling frequency. (2) This duplicated spectrum is multiplied by a decreasing sinusoidal envelope.

Note that there is no reason to obey the sampling theorem for the purposes of sound synthesis. The overlaps in the spectrum just give us a different sound.

In the generation of waveforms by a digital segment-by-segment approach, we often deal with what are called "reconstructed waveforms." These are of interest here because they are of the same form as certain sampled waveforms. A good example is the reconstructed staircase formed from square wave octaves as shown on the next page. We have shown only three square waves here, but it is probably already evident that the process will lead to a better and better approximation to the sawtooth waveform as the number of square wave octaves is increased. This can also be demonstrated by adding up the Fourier series of all the square wave components and seeing that this results in the Fourier series of the sawtooth waveform.

It should also be evident that the reconstructed staircase has no DC component since it is perfectly balanced, just as its square wave constituent square waves are. The reconstructed staircase thus does not contain the sampling frequency (8f) in its spectrum. This fact can also be demonstrated by summing the Fourier series of the first three square waves. The truncation of the sawtooth reconstruction at 4f leaves out the harmonics 8f,



16f, 24f, 32f, 40f,...etc. These are the harmonics that will be supplied by the higher square wave constituents. The square wave 8f will supply 8f, 24f, 40f,.... and the square wave 16f will supply 16f, 48f, 80f, ...etc. and so on until the entire sawtooth is completed (all harmonics).

It might have been supposed that since the "unsampled parent" waveform of the staircase is the sawtooth that the sampling frequency would appear in the staircase spectrum since (1) The sampling frequency and the sawtooth are "locked" at an integer ratio, and therefore (2) one of the harmonics of the sawtooth (the base band signal) would thus fall on the sampling frequency. The actual case is that it is not possible to violate the sampling theorem "after the fact," i.e., for reconstruction. The reconstructed waveform "assumes" that its parent was subjected to the sampling theorem. Another way to look at it is to observe that a sawtooth is not bandlimited since it contains all harmonics, and thus it is not possible to sample a sawtooth without loosing information. In a practical case, it is necessary to add a "guard filter" to the input to assure that no components above half the sampling system can get into the sampling device. An additional example, a reconstructed sine wave, is discussed in the wave-shaping section of chapter 5b.

CHAPTER 2D

CONTROL OF MUSICAL STRUCTURE

CONTENTS:

INTRODUCTION

When we refer to the structure of music and its control, we are referring to what might be termed the "note-to-note" variations. Here, it is necessary to keep a very open mind about what is to be termed a musical note however. In the first three chapters of this section, we have been concerned with the synthesis of sound by producing waveforms and altering their parameters in response to envelopes and other control devices. The control of structure will be concerned with the time ordering of these sounds. When done under the control of a musician, the imposed structure changes the sounds into music.

A simple and familiar example of this sort of control is the control of pitch and timing with an ordinary musical keyboard. Structure can be imposed by the musicians desire to produce a time sequence of pitches and his ability to do so by means of the keyboard. Pitch is not however the only element of the synthesized sounds that can be structured. Other factors are tone color, loudness, rhythm, and just about any element of traditional music. In addition, electronic devices make possible certain variations that are not possible in traditional music (e.g., 20 Hz of vibrato at a depth of one octave), and these may be controlled and structured. Probably the largest areas where the new technology adds elements to be structured are (1) The addition of new and rapidly changing tone colors. (2) Changes that are faster than is possible with manual control. (3) Simulated changes in the apparent spatial location of the sound. (4) Changes in dynamics that exceed the usual range. (5) Very high precision (exact electronic waveforms) and very inexact and unrepeatable results (use of random elements).

CONTROL - ELECTRONIC, PHYSICAL, AND PROGRAMMABLE

ELECTRONIC CONTROLS:

The most obvious means of control over electronic instruments is by means of the standard sort of electrical devices. One way to produce a control voltage is to simply use a potentiometer as a voltage divider. It is then possible for example to control the pitch of a VCO. The device used in this way is unsatisfactory for several reasons. First, all pitch changes are continuous and not discrete - at least to the limits of the human hands rapidity of motion. Secondly, the device only changes pitch, and another device is needed if we want the tone to change in amplitude. Thirdly, for the production of music in discrete steps, the musician must rely on feedback to adjust to the proper pitch or else have a very accurate dial control. This is not to say that such a controller is not useful for certain types of music. It can be a means of reaching pitches not on a discrete controller, and of "bending" pitches as a tone progresses.

Another type of electrical control is the switch. This device has only two states - off and on. In order to achieve a set of discrete pitches, a large number of switches can be used. In order to make them easy to operate, a push button type seems most useful (imagine a set of rotary switches as a musical controller for traditional music!). This leads of course to the keyboard idea, and of course, the standard organ or piano style keyboard is the most natural for musicians. This leads to some problems as every person who has ever demonstrated a synthesizer to an unsuspecting individual has no doubt discovered. They almost always try to play three or four notes at the same time. This is of course the logical assumption to draw from the use of a full keyboard. Then the proud synthesizer owner has to apologize for the fact that it only plays one note at a time. All early synthesizers were in fact monophonic devices - one note at a time. There is a logical reason for this. The synthesizer was intended as a solo voice. It provides a new tone color and has time and dynamic properties that exceed traditional instruments. Thus it is to be thought of more like a trumpet which is usually played one note at a time (some accomplished jazz musicians can get several notes out of various wind instruments). Thus the synthesizer is a monophonic instrument that just happens to have what looks like a polyphonic device. A number of polyphonic instruments are being developed which should fill the gaps many musicians feel.

Two other types of electrical controls have been used with synthesizers: the ribbon controller and the joy stick. A ribbon controller is a resistive strip that has a current flowing through it and which can be tapped at any point along its length to bring out the corresponding control voltage. In this sense, it can be thought of as either a linear pot or a continuous keyboard. The joy stick is a two domensional pot. Usually it is controlled by a lever arrangement so that a single rod can be moved up and down or from side to side. The joy stich has two outputs: an X-axis voltage and a Y-axis voltage.

PHYSICAL CONTROLS:

One of the problems with a controller made of switches is that it only has the two states, and the fact that a switch closes at a certain time tells us nothing about the physical manner in which it was closed. Was it pounded with a fist or timidly pushed? It of coures makes no difference as far as the electrical response of the system is concerned. However, musicians are used to having an instrument respond in a different manner when they alter the physical technique. Thus, the engineer is faced with the task of providing a means of sensing some of these physical parameters and then providing additional circuitry so that the detected parameter does something like the musician had in mind. For example, the musician might want a more rapid tap on the keys to give a stacotto effect. He might also want increased pressure to result in increased loudness. Sensing of velocity can be accomplished with two sets of contacts on each key. If these contacts make in sequence, the time between the switch closures is inversely proportional to the velocity of the key. It is also possible to sense the pressure with conductive foam or some similar material that changes resistance as pressure increases. Coming up with a means of recovering a physical parameter is the easy part of the job. It is a little more difficult to make a circuit that gives a change in response to the recovered physical parameter, and a lot more difficult getting this "tuned" so that it "feels" right to the musician. That is, knowing exactly what to do is the biggest problem.

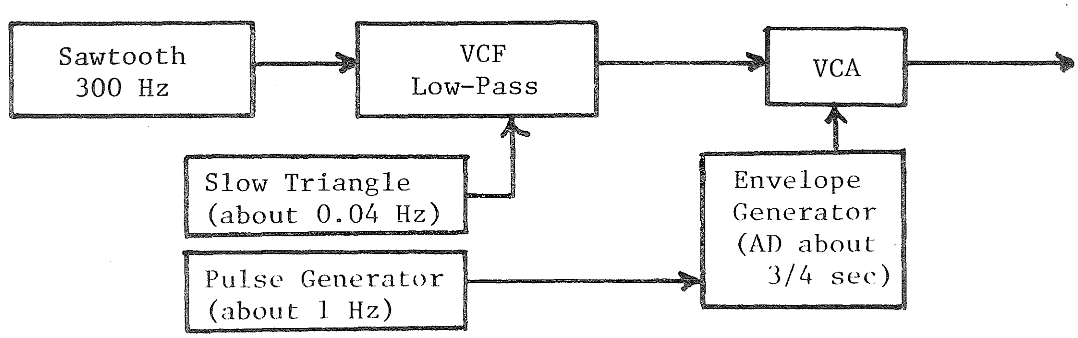
Another approach to the recovery of a physical parameter is to use the sound of a live musical instrument and then recover the useful information from that sound and use that to control the electronics. This is not an easy problem. The recovery of the amplitude information involves an envelope follower, and the exact time constants of the envelope follower can be set only when the pitch is known. Recovery of the pitch can be an even more difficult problem. This is mainly because the pitch is a subjective quantity. It is easy to recover the pitch of an electronic waveform - this is only a matter frequency to voltage conversion. Live sounds on the other hand tend to be much more irregular (that's what makes them more interesting to the ear) and thus it is harder to pick up the necessary patterns to recover the pitch. In some cases, it is possible to recover certain acoustical parameters directly from the vibrating structures of the original musical instrument (e.g., from the strings of a violin) and thus avoid complications due to the coloration of the acoustical signal by the sound radiating structures.

PROGRAMMED CONTROL:

Programmed control offers the composer new dimensions of speed and complexity. Musical structures that would be too difficult for even an orchestra made up of the world's most accomplished performers can be easily specified and will then appear automatically.

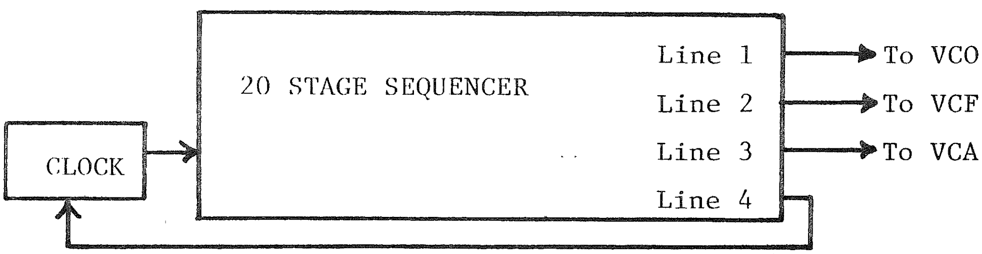
Programming permits the user to construct intricate rhythm and melody patterns and then alter them bit by bit if necessary. The rhythm can be programmed by setting a basic time division that is small enough that any desired note length can be represented to a desired degree of accuracy by an integral number of such small units. Melody patterns can be obtained by setting pitch values for each time segment (many adjacent segments will have the same value) and then recycling or reprogramming the sequence as necessary. An alternative memory scheme is to store a pitch value and then specify the number of basic time divisions that the value will last.

The patching systems used with most synthesizers are the first type of programming we can consider. The musician achieves control over the synthesized sound by setting up the patch and adjusting all the controls to whatever values he wants for initial values. This basically determines the sound itself, but there are also a wide variety of patches that control structure. The patch below illustrates two things: (1) It is a patch that controls structure independent of manual control. (2) It shows a structure change that does not involve any change of pitch.



This setup produces a change of tone color as notes change. The tone color is changed by the VCF controlled by the slow triangle wave. Notes are triggered by the pulse generator which controls a short AD type of envelope which in turn controls the output amplitude by means of the VCA. The exact sequence of notes out depends on many factors: (1) The exact waveforms selected. (2) The patch. (3) The initial settings of the controls. (4) The (chance) setting of initial phase between the VCF control voltage and the pulse generator.

A second means of programmed control is by means of the sequencer. A sequencer produces a series of musical events in a predetermined order. In the usual mode of operation, the user first sets all the parameters of the sound for each note, and then activates the sequence with a single control. There are many types of sequencers that are in use. One type is outlined below. It has n different states and m different outputs per state. Let's say n = 20 and m = 4. We can thus program 20 notes and have four control voltages available for each note. One of these control voltages will most likely control pitch, another might control a VCF and a third might control amplitude. When the sequencer is clocked into its first state, a set of voltages will be fed to the controlled modules and the first note will be produced. When the next clock arrives, the sequencer jumps into the second state and the second note is produced, and so on. The sequencer can be triggered by either an internal or an external clock. If the clock is made voltage controlled, the advance rate of the sequencer can be controlled by feeding back a voltage (e.g., line 4) from the sequencer. In this way, the duration between pulses can be easily programmed.



Some sequencers can be programmed by means of a standard keyboard. The musician simply plays in the sequence and it is stored in a memory. The information picked up can be just the pitches, the pitches and rhythm, or other possible combinations of information. In other devices, both pitch and rhythm can be programmed in from the keyboard and yet not have to be played in in real time - multiple key pushes serve to define longer notes.

Systems that employ a large or small computer have a wide variety of programming possibilities. Point-by-point sound synthesis on a large digital computer is totally programmed - like it or not. Smaller computers that drive analog modules are ideally suited for programming. The computer itself can take on the job of the sequencer. The sequencer then exists only in software and can be changed as needed. In addition, small computers with some form of outside memory (e.g., floppy disk) can serve to store vast amounts of material which can be read into the computer and "played" through the analog attachments. In this sense, the computer memory is the recording medium. In general applications, such a device is a useful tool for the composer since he can input his basic ideas and just change one parameter at a time until he gets the exact results he wants. Other material not suitable for his present composition can be stored in the computer memory for future use. If the computer has control over <u>all</u> the voltages in the analog synthesizer, then nothing has to be written down or reset. The composer simply has to record the address of the stored information.

STRUCTURE CONTROL BY TAPE RECORDER

A tape recorder performs the functions of recording sounds and also making them available in a form where time is converted to length. Also a strip of tape can have sounds recorded in parallel tracks so that relative times are locked together. This allows several forms of structure control.

1) Macro Tape Editing: This is more or less conventional tape editing used to organize major sections of a piece and adjust their timing.

2) Micro Tape Editing: In this operation, actual structures are created by the editing. Typical tape lengths for this type of editing would be from less than a second to several seconds, but the main criterion for micro as opposed to macro is that micro is essential to the actual production of the structure, not just a convenient way of presenting the structure.

3) Multitracking and Remixing: In this operation, up to 16 tracks can be recorded on a inch wide tape. Typically, one track is the sync track and serves as a reference while other parts are put on one at a time. This allows one musician to perform the functions of many and with only one synthesizer instead of many. Also, it is then possible to mix down the tracks and adjust the balance, perhaps selecting between two or more versions of the same part if desired. Four track quad tape recorders can also be used for this in small studios.

4) Tape Loops and Delay Lines: The tape can serve as a storage for sounds during performance and composition. If the tape is looped around, the sounds will be repeated for each loop cycle. In other cases, material can be recorded by one tape recorder and the tape can then be drawn across to a separate tape recorder for playback. This can be very useful for canon and fugue-like structures.

STRUCTURE CONTROL IN REAL TIME

Electronic music is often used in real time to supplement traditional instruments. In such cases, the synthesizer can serve as a solo instrument. Other electronic devices related to sequencers serve to produce rhythm backgrounds electronically. Also, electronic amplification and processing units have been used in real time for a long time. These processing devices inclued "fuzz boxes" (square wave chopping), "Wow-Wow" units (a VCF slowly modulated) and phasing simulators (phase shifters).

Some live performance groups use essentially all electronic instruments in concert situations. This provides some special problems - limiting the complexity of changes that can be made in the equipment due to limited time between the time it is finished for one section and the time it must be ready for the next section. Small prepatched synthesizers can be reset quite rapidly in many cases, but the prepatching imposes a limit on the variety of structures that can be produced. One solution is to have available enough equipment already set up and tuned. The performers just jump from one keyboard to another as the work proceeds. If time allows, setups can be revised by one performer wearing headphones while the others continue to play. However, working in real time will often impose limitations that will have a strong influence on the style of music that the group produces.

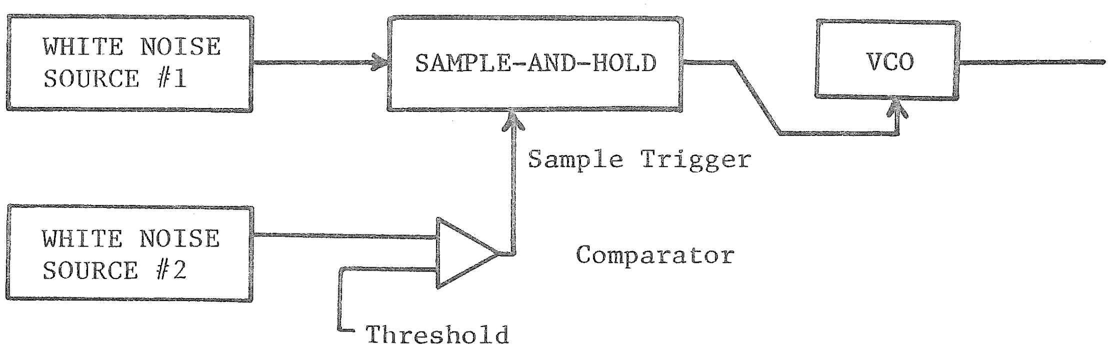
One way out of this is to speed the patching and resetting process. Here, digital computers and resetting circuits are useful. It would be easy for a musician to just preprogram all the parameters of his instruments for the different sections of the work he is presenting. Then, the patch could be advanced at the push of a button, or the performer could punch in a code for a patch and have it recalled from memory. The memory could contain a storehouse of his most used patches, or it could allow him to instantly store a patch he is using and recover it later. It should be pointed out that to do this sort of thing, all the parameters of the patch must be voltage controlled.

Thus, the synthesizers would have essentially no knobs. Each parameter would have a code and this would be given a voltage value for a given patch. It is also worth considering that the computer could replace much of the existing synthesizer - serving as a massive envelope generator. Real time systems where the computer serves to control analog modules, store structures, and program the patching are sure to become popular.

RELEASE OF CONTROL

In some cases, the musician will want to give up part or all control and let the system do its own thing. Control can be released by turning it over to a random source controller, or by letting a cycle complete itself without human control. It is often desirable in such cases to take a rigorous look at just how the structure is controlled since often the term "random" is used to describe processes that are in fact completely determined by initial conditions. The process is further complicated by the ability of the human mind to remember structures and recognize patterns. Often there may be little or no difference between the overall effect achieved with purely random control, pseudo random controls, or very long but repeating patterns.

We show below a system employing two white noise sources and a sample-and-hold unit. The structure produces a series of random pitches that change at random time intervals:



When the threshold level to the sample-and-hold is set fairly high, threshold crossings by the white noise are relatively rare and a fairly slow series of random pitches will result. When the threshold is lowered, the pitch changes become more rapid and get to a point where the ear tends to hear a texture rather than a series of pitches. That is, the ear tends to integrate the perception into a type of sound much as it does white noise itself. While the result is still random, the process results in the production of a sound that is basically repeatable. This is a case where a random process leads to something that seems basically non-random.

On the other hand, some processes that are not random may appear so. To demonstrate this, we can replace the white noise sources with mixtures of periodic waveforms. This will in general produce a pattern that seems random, and the same is true of certain "pseudo random" generators which produce very long but eventually repeating sequences.

Beyond these electrical examples, new music often has elements of chance that is built in for one reason or another. However, it is not always proper to contend that these are random. Jazz improvization or any improvization has its chance events, but is certainly not random. Likewise, a piece of music consisting of city street sounds has a structure that depends on chance, but is not random. You would probably hear a lot of sounds like cars, pedestrians, etc., but very few herds of elephants coming through. If it were random, an elephant would have the same priority as a garbage truck.

CHAPTER 2E

MISCELLANEOUS ELECTRONIC MUSIC TECHNIQUES

CONTENTS:

INTRODUCTION

There are numerous techniques of electronic music that do not fall under the subject headings of chapters 2a - 2d. The first group of these really involve tape recorders and have evolved from musique concrete. Other techniques employ other types of audio processing such as artificial reverberation. Finally, there are techniques that are closely related to additive or subtractive synthesis, but which require additional equipment or special setups.

TAPE MANIPULATIONS

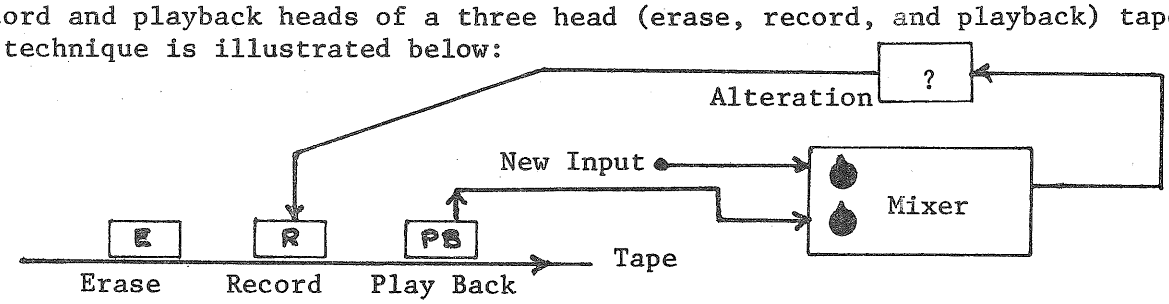
There are three basic types of tape manipulation that do not invlove delay effects: speed changes, direction changes, and tape editing (splicing). Speed changes have been used for novel demonstrations ever since multi-speed tape recorders have been available. Doubling (halving) the speed causes all pitches to be doubled (halved). In fact the effect is so familiar that it is interesting to speak at a normal rate into a frequency shifter while upshifting the frequency. Many listeners will associate a speed-up of the speaking rate even though none actually exists. In practice, these speed changes are used and the results rerecorded. In this way the process can be repeated over and over and speed ratios far in excess of those directly available from the tape recorder are obtainable. When the speed changes are in the upward direction, the original program material soon reaches the point where the original frequencies would be above the response range of the tape recorder electronics. The program would be lost except for the fact that the very lowest frequencies may still be audible at the top of the shifted recording, and elements of the musical structure may be changing rapidly enough to be perceived as low frequency pitches. Downward speed changes generally result in slow droans and low-frequency "grunts." However, certain high pitched sounds such as bird calls may be slowed to a degree where the entire song can be heard. At the original rate, the bird call may be too fast for the human ear and just sound like a shriek.

Direction changes can be very useful. However, this is only possible with a tape recorder that is intended to play in only one direction in the first place. Mono, two-track stereo, and four-track home recorders will do this, but the common four-track stereo recorder will not. To change the direction in time, the reels of the tape recorder are simply reversed, and the tape recorder is put into its normal playback mode. If this is done with a four-track stereo recorder, you simply get the program that is recorded "on the other side." Direction changes of the sounds of traditional instruments can greatly change the effect achieved since the attack and decay are effectively reversed (along with note order of course). A piano sound has a very sharp attack followed be slow decay. If this is reversed, the "attack" sounds more like an organ, and the "decay" seems entirely artificial.

Tape editing is the process of physically arranging the sounds by cutting the tape and resplicing it in various ways. This involves a certain amount of practice. The exact spot on the tape between selections and notes can be located (on some machines) by allowing the reels to move freely and turning the reels by hand while listening to the playback head. The tape that is over the playback head can then be marked with a grease pencil and pulled out to an editing block for splicing. The recommended editing system is the "Editall" splicing block with a sharp single edge razor blade. The razor blade can be demagnitized if necessary with a tape recorder type "head" demagnitizer. Good splicing requires both practice and patience but is worth the effort.

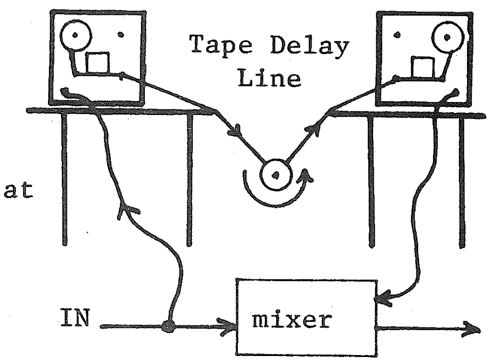
TAPE TECHNIQUES USING DELAY

The tape recorder is a method of delaying a sound. You can record something and play it back later. A number of interesting techniques involve the use of a carefully controlled delay. The shortest delay that can be used is the one that is obtained by the distance between the record and playback heads of a three head (erase, record, and playback) tape recorder. The technique is illustrated below:



When the input level from the playback head to the mixer is turned up, the material on the tape will be fed back with some delay resulting in the "tape echo" technique. The echo rate can be controlled to some degree with the tape speed, and the number of echos is controlled by the fraction fed back to the record head and is limited by the quality of the tape. When there is no input, the loop gain can be turned up to the point where the residual noise on the tape begins to build up. This results in a characteristic periodic swishing sound. In addition, some sort of processing device can be put in the loop to alter the sound on each "echo" and this results in a variety of special effects and structure progressions. An example of this can be found in chapter 1a, pages 4 and 5.

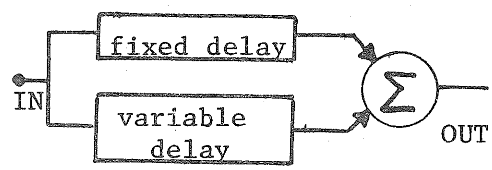
When two or more tape recorders can be used, longer but constant delays can be achieved by recording on one tape recorder, letting the tape pass to a second recorder and then playing it back. This is useful for "canon" types of structures when the timing is carefully done. As a practical matter, two problems must be considered when using this technique: First, the tape recorders may have "shut off" levers that will not be under the proper tension since the tape path is not complete. It may therefore be necessary to devise some method to hold these on. The second problem is that the tape recorders will not run at exactly the same speed. Thus, some arrangement must be made for taking up excess tape or making available extra tape. A simple method is shown at the right. An extra tape reel is used as shown and will move up or down as necessary. The reel should have a smooth edge and not have a threading slot as the slot might catch the tape as the reel revolves. If this take up method does not allow enough slack for the time that the delay line is used, it may be necessary to make up a new capstan for the recorder, or make speed adjustments if possible. Some tape recorders have capstans that can be changed for different speeds. With this type, it is fairly easy to make a replacement on a lathe. The outside diameter should be turned down to adjust for the necessary speed change. The inside should be bored with a boring bar, not drilled.



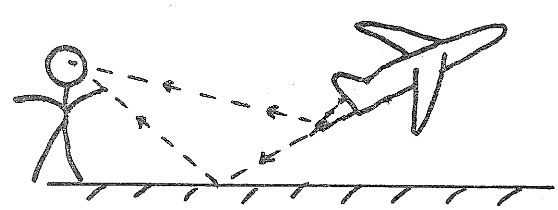
Another method of delay is to use a tape loop. This is a combination of the tape echo and tape delay line techniques. The tape is routed around and connected back to itself. Tape loops of from about 6 inches up to ten feet or longer are possible. The loop may be used in a record and playback mode, or just in the playback mode. Tape loops can be used to provide various repeating background figures.

PHASING (FLANGING) METHODS

The phasing (Flanging or "Jetsounds") method occurs when there is a small and varying delay between an original signal and the delayed version, and the two are mixed together. An idealized setup for producing this effect is indicated at the right. The best way to approach an actual phasing device is to use an analog delay line technique (Chapter 6c), but traditionally this has been done out of real time by using two (identical) tape recorders. With this method, the two tape recorders are set up in parallel, both recording, and both playing back through the playback (monitor) heads. The two monitor outputs are mixed together. Next, one of the recorders is slowed by simply touching the supply reel of the recorder. This causes the outputs to go in and out of phase with each other. As can be seen by the analysis in chapter 6c, the phasing effect is equivalent to a comb filter.



It appears that the term "jetsounds" is really no accident. Consider for example the situation shown at the right. The listener hears a direct sound from the jet and a delayed sound due to reflection. This results in a comb filter, and the notches in the comb vary as the jet moves.



For operation in real time, it is the usual practice to use simulators that use a phase shift rather than actual fixed delay. Analog delay lines will eventually make practical real time phasers. This is equivalent to the parallel tape recorders except there is a much shorter minimum delay possible - in the case of the tape recorders, the minimum delay is set by the fastest speed of the tape recorder and the spacing between the heads. This time is typically on the order of 100 ms, too much of a delay to be used in real time. With analog delay lines, the delay can be much shorter. Some users of phasing simulators will prefer them to true phasing. They may get used to the sound of the device they are using. Furthermore, the simulators may have non-linearities that cause harmonics to be generated. This will enrich the sound and may me missed if true phasing is used.

ARTIFICIAL REVERBERATION UNITS

Reverberation is due to multiple and very dense echos. In any live situation, much of the sound that is heard is indirect (perhaps 95% in some rooms). In an effort to add realism to electronic sounds, artific al reverberation can be added. The necessary echo density can be achieved with recursive structures using analog delay lines (see chapter 6c). At present, artificial reverb units are mainly mechanical, typically formed from two or more parallel springs that are driven on one end by a magnetic drive while the other end has a magnetic pickup to recover the vibrations and reflections along the spring.

In addition, this type of device may be used in excess of what would be expected in a live situation. In this way, electronic sounds can be given a spatial quality of distance in excess of what can be expected. This is a common use of artificial reverb units. Typically, a stereo unit can be used with its channels in series to greatly enhance the degree of processing. In addition, a sharp pulse can be added to a quiet line to give a shot-like output. Direct mechanical excitation is also possible. For example, a sharp tap on the unit will excite the springs for thunder-like effects.

ENSEMBLE EFFECTS

In traditional music, there are often many musicians playing the same line of the score. It is impossible for them to play exactly the same, but the general effect achieved is known as "ensemble" and is a very rich sound as compared to a single electronic oscillator for example. In electronic music, pure sounds and integer ratios are the general rule. Students of orchestration know that certain overtones can be enhanced by simply having a player play the overtone. However, this is never exactly in the "right" position. Thus, in electronic music, one approach to realism is to attempt to achieve artificial ensemble.

In large voltage-controlled synthesizers, VCO's can be used in parallel, all tuned to the same pitch initially. Small drifts and individual variations will cause beating effects that lead to ensemble effects. While many VCO's have a sync. option, using this will tend to destroy ensemble in most cases. Some VCO's have "soft sync" which is usually a phase-locked loop method. The PLL is always wandering slightly about the exact sync. position and this can also produce ensemble effects, depending on the exact system.

For polyphonic instruments, the "top-octave" generator is a popular means of getting all the required pitches from one oscillator. This however destroys all possibility of ensemble that might otherwise occur if different oscillators were used for different keys.

RESONANT SYNTHESIS

Many traditional instruments consist of resonant structures that are excited by some type of excitation. We can thus consider certain types of resonant filters for sound synthesis. One such method has been considered (filter ringing) in chapter 2a. In another approach, a whole bank of filters can be used as a sort of "sounding board" which enriches any sound that passes through it and which moves in relation to it. Consider for example that a periodic waveform will have the amplitude and phases of its harmonics altered by such a filter bank, and that these changes will be different for different frequencies of the periodic waveform. It is necessary however that the filters not be placed where they will enhance any one note or any group of overtones while leaving other notes with a generally attenuated set of harmonics. An empirical method of spacing filters has been described by Burhans [EN#40 (16)]. A Table of spacings is shown below. The filters are spaced at intervals of the fifth root of 2.1:

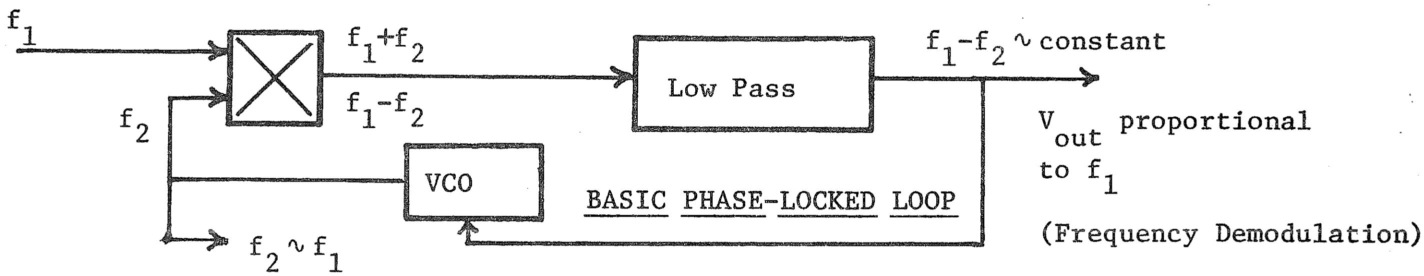
$\sqrt[10]{2.1}$ FREQUENCY SPACINGS

n	ratio	PEAK	VALLEY	PEAK	VALLEY	PEAK	VALLEY	PEAK	VALLEY
0	1.00	50		105		220.5		463	
1	1.08		54		113.4		238.1		500
2	1.16	58		121.8		255.8		537	
3	1.25		62.5		131.3		275		579
4	1.34	67		141		295		621	
5	1.45		72.5		152.2		319		671
6	1.56	78		164		344		722	
7	1.68		84		176.5		370		778
8	1.81	90.5		190		398		838	
9	1.95		97.5		205		430		903
(10)	2.10	(105)							
0	1.00	972.4		2042		4288		9005	
1	1.08		1050		2205		4631		9726
2	1.16	1128		2369		4974		10446	
3	1.25		1216		2553		5360		11257
4	1.34	1303		2736		5746		12067	
5	1.45		1410		2961		6218		13058
6	1.56	1517		3186		6690		14048	
7	1.68		1634		3431		7204		15129
8	1.81	1760		3696		7762		16300	
9	1.95		1896		3982		8362		17561
(10)	2.10							(18911)	

PHASE LOCKING

Phase locking can be used for a variety of purposes. A phase-locked loop (PLL) may be an integral part of some module, or it may be patched up from modules as needed. It should be pointed out right away hawever that the PLL is not directly useful as a pitch follower. A live signal is too complicated for a PLL to follow.

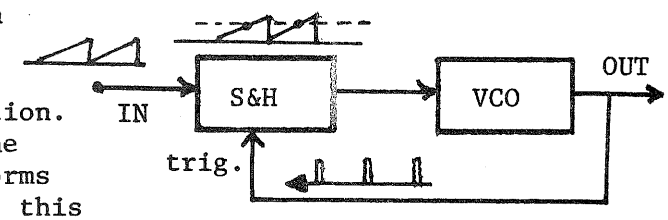
A basic PLL can be formed from a multiplier, a low-pass filter, and a VCO as shown below:



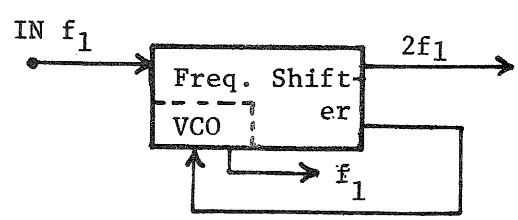
BASIC PHASE-LOCKED LOOP

This basic loop can be understood since the multiplier produces sum and difference frequencies. The difference frequency will be lower and will be passed through the low-pass more readily. This causes the VCO to wander around, but only moves that make the difference frequency less will be "rewarded". Eventually the difference frequency will go to zero (DC) and this will cause the loop to lock. However, it will still wander slightly about its stable point, and this can be useful.

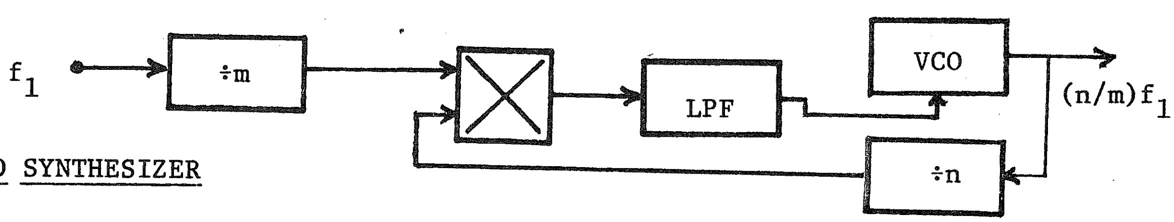
A PLL can also be implemented with a sample-and-hold as suggested by Bodley [EN#35 (4)]. The setup is shown at the right. It can be seen that this type of loop locks onto a portion of a smooth upslope. Points on the downslope are unstable as they move the VCO in the wrong direction. Note also that if a second voltage is added to the VCO control, the relative phase of the two waveforms (input and output) can be controlled. Therefore, this can be used as a phase modulator.



Another phase-locking method that does not involve the use of the low-pass filter uses a frequency shifter. Actually, this can be considered to be the ordinary PLL with a "perfect" low-pass filter. The patch is shown at the right. This type of loop is discussed in chapter 6a, pg 11. Note that when the loop locks, the upshift output of the freq. shifter is twice the input frequency while the VCO inside the freq. shifter runs at the input frequency.



Phase-locked loops have many applications. They may be patched up in electronic music systems to hold certain frequency ratios fairly constant. Another application that would probably involve a dedicated PLL is the frequency ratio synthesizer. It starts with the usual insertion of a ÷n circuit in the VCO feedback loop. If an additional ÷m is put on the input line, the ratio m/n is realized:



FREQUENCY RATIO SYNTHESIZER

CHAPTER 3A

BASIC APPLICATIONS OF:

OPERATIONAL AMPLIFIERS

CONTENTS:

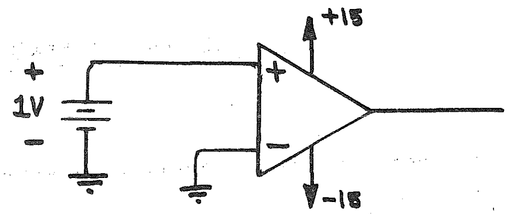
INTRODUCTION

Operational Amplifiers (usually just shortened to "op-amps") are the single most useful linear IC in general electronics use today. Designing with op-amps can be very easy in many cases; one can start with a set of "idealized" characteristics and worry about the limitations of real op-amps when it comes time to select the actual op-amp for the design or when the common op-amps do not give the desired performance.

We will be here first considering three ideal characteristics of op-amps one at a time and derive the basic applications procedures. We will then discuss real op-amps in regard to there performance limitations in slew rate and bias current. The discussion of current differencing amplifiers in chapter 3c will follow the same outline. Additional op-amp applications can be found in chapter 4a, where the circuits are treated as building blocks.

THE FIRST IDEAL CHARACTERISTIC - INFINITE GAIN

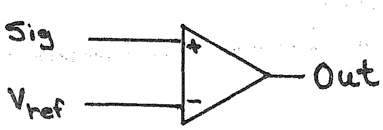
One of the ideal characteristics of an op-amp is infinite gain (or at least a very high gain). Suppose we say that the voltage gain is one million. This means that the differential input voltage to the op-amp is amplified by a factor of one million. The standard way of representing an op-amp is with a triangular symbol (Burr-Brown sometimes uses a curved back on the triangle to avoid confusion with other symbols, but the regular triangle is more common, and templates for triangles are readily available). One point of the triangle is the output. The "back" of the triangle has two inputs, an inverting input (-) and a non-inverting input (+). It is interesting that we are so used to amplifier configurations using negative feedback that we use the term "non-inverting" which seems to be a double negative. The differential input voltage is the voltage difference between the + and - inputs. Below is a circuit that uses the op-amp with a voltage gain of 1,000,000. We have here shown the power supply connections, but in general, we won't.

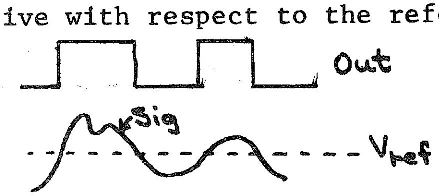


The differential input is 1 volt (the battery) and the gain is 1,000,000. Therefore, the output voltage should be 1,000,000 volts! Of course it is not; there is no way the output can exceed the limits of the supply voltage. The op-amp just does the best it can, which is, +15 volts.

We can see also that V_{in} can be as small as 15 volts/1,000,000 (15 microvolts) and we will still get a +15 volt output. Likewise, if a voltage input is more negative that -15 microvolts, the output becomes -15 volts. In practical cases, the value of the these minimum voltages that will cause the amplifier to drive its output to the supply limits cannot be calculated in this simple manner (due ot input "offsets"), but the point is that they are small. For all but a very small region around zero volts, the output will be at either +15 or -15. For this reason, the circuit above is often refered to as a zero crossing detector. An op-amp connected in this way is said to be operating "open-loop," and the circuit function performed is called a comparator action.

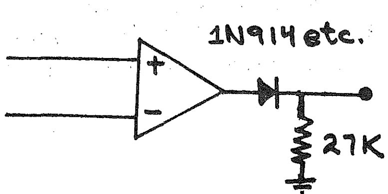
The basic comparator circuit is useful as a general signal level detector as well. In the general case, a reference voltage is placed on the - terminal, and the voltage to be level detected is put on the + input terminal. Whenever the signal reaches a level that is positive with respect to the reference voltage,





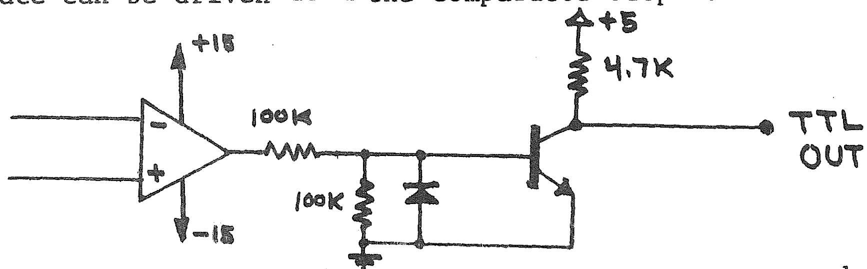
the output goes to +15, otherwise it is -15. If on the other hand, we consider the + terminal as the reference, and put a signal on the - input, the resulting output is reversed. As a practical manner, if you get it mixed up, just reverse the inputs.

Sometimes, you need an output that is some voltage (e.g., +15) for one state, and zero volts for the other state. This is a matter of blocking the -15 volts from the output. A signal diode will do this, and it is a good idea to load the



op-amp with a resistor in the range of 10k to 50k to prevent drift that could occur in some cases, depending on the circuit being driven.

When it is desired to interface this circuit with TTL logic, a standard interface can be driven from the comparator output. Since this interface is

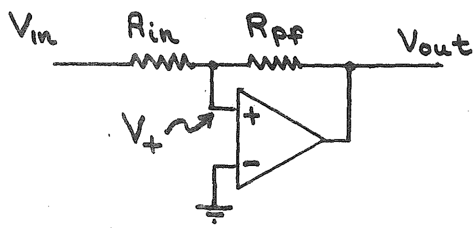


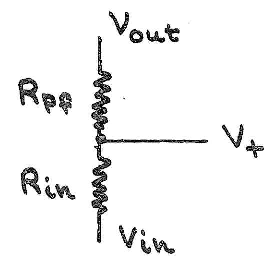
inverting, the inputs to the op-amp comparator are reversed from what they would be if the op-amp were used directly.

THE SECOND IDEAL CHARACTERISTIC - NO INPUT BIAS CURRENT

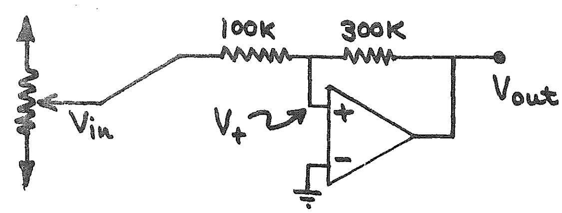
The next important applications come up when we consider the use of feedback - the return of part of the output signal to an input terminal. Here, we must use the second ideal characteristic of the op-amp: No current either enters or leaves the input terminals. That is, the inputs just respond to the voltages on them, but draw or supply no current, and hence do not alter the applied voltages in any way. In practice, we must sometimes allow for these currents, depending on their magnitude relative to the currents available (or charge stored).

Consider the following circuit:



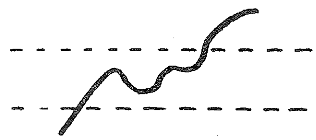


Since no current is drawn into the + input, the voltage V_+ depends only on V_{in} and V_{out}, and the ratio of R_{pf} to R_{in}. R_{pf} and R_{in} form a simple voltage divider. Now, ovserve that in general V_+ will not be zero, so the infinite gain of the op-amp will drive the output to either +15 or -15, just as in the comparator case. We are mainly interested in the transition points, i.e., the points where V_{in} will cause the output state to change. Consider the following example:



Assume first that the output is +15, then as long as V_{in} is positive, V_+ is also surely positive, since both ends of the voltage divider are positive. We then look at the case where V_{in} goes negative, and ask what negative value of V_{in} will cause V_+ to become zero. Simple algebra and Ohm's law show that this will be true when V_{in} reaches -5 volts. Slightly beyond -5 volts, V_+ starts to go negative, and the output then goes rapidly to -15 volts. Note that as it does this, a "bootstrapping" action takes place; the fact that it starts to change state causes it to change state even faster, adding a "snap" to the normal comparator action. Once at -15 volts output, a similar analysis shows that the output will not go to +15 again until the voltage V_{in} reaches +5 volts. The region between +5 and -5 volts is a dead zone as far as changes of state are concerned. This circuit has two level crossing points, and the one that is in effect depends on the state that the circuit happens to be in. The output that is actually found when the input is in the dead zone depends on which way the input voltage entered the dead zone, from the +5 side or the -5 side. A circuit of this type is said to have "hysteresis," which means the exact output depends on past history. The circuit is commonly known as a Schmitt trigger.

The Schmitt trigger is useful for several purposes. For example, when we are trying to level detect a signal with some unwanted noise present, we might get unwanted transitions due to the noise if we used a regular level detector. This

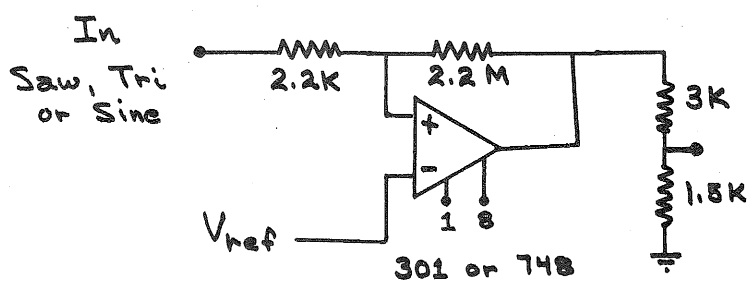


is important for example for a counting circuit where the extra transition would result in an extra counts.

The Schmitt trigger circuit with R_{in}= 100k and R_{pf}= 300k is also useful in VCO circuits where we want to generate a signal confined to the range from +5 volts to -5 volts. In this case, the Schmitt trigger operates almost exclusively in the dead zone; the transitions are triggered by excursions outside the dead zone and these transitions pull the signal back in.

Note that the percent of feedback is determined by R_{in} and R_{pf}. The magnitude of the dead zone about zero in either direction is just $R_{in}/R_{pf}\cdot 15$ volts. The maximum value of R_{in} is therefore R_{pf}, since $R_{in}=R_{pf}$ would mean that the dead zone would be the full supply voltage swing, and the Schmitt trigger would always stay in its first state. A type of oscillator circuit that is very common is the Integrator-Schmitt trigger combination. The peak to peak amplitude of the waveform from this oscillator is determined by the switching levels of the Schmitt trigger.

In many cases, a small amount of positive feedback, say 0.1% is useful when a waveform with a very fast rise time must be obtained. In this sort of Schmitt trigger, the output waveform is nearly the same as that from a level detector, since the feedback is so small. However, even that very small amount of positive feedback will sharpen up the edges of the pulses due to the "bootstrap" effect discussed above. A circuit often used here is the 748 or 301 type of op-amp that is used <u>without</u> any compensation.

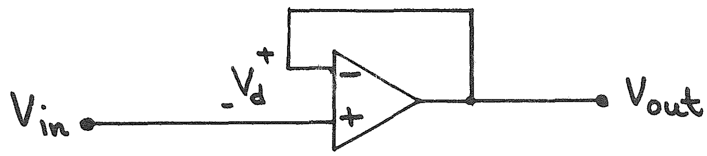


THE THIRD IDEAL CHARACTERISTIC - NO DIFFERENTIAL INPUT VOLTAGE AS LONG AS NEGATIVE FEEDBACK IS WORKING

Next we consider what is probably the most useful type of feedback in op-amp circuits - negative feedback. Negative feedback is feedback from the output to the inverting input. Negative feedback however can occur at the non-inverting input if there is an inversion somewhere in the feedback loop. Thus, the inverting or non-inverting inputs are not absolutes - they only treat signals in an opposite manner. The difference between positive, negative, or no feedback may be only a matter of a few degrees of phase angle.

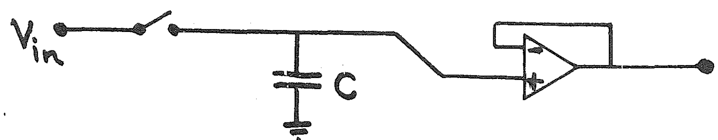
The second ideal characteristic still applies, and we shall see that the first ideal characteristic of infinite gain will, in the case of negative feedback, lead to a third.

The simplest form of negative feedback results when we simply connect the output back to the - input with a piece of wire. Consider the following circuit:



If V_d is positive as shown, then $V_{out} = A \cdot V_d$ will be forced lower, since the positive voltage appears on the inverting input. If on the other hand, V_d is negative, then the output will be forced up. The only possible stable point for the circuit is the case where $V_d = 0$. If the output should try to wander ever so slightly from a condition that makes $V_d = 0$, it will be forced back by the infinite gain of the ideal op-amp. Looking at the diagram above, we can see that $V_{out} = V_{in} + V_d$ and since $V_d = 0$, the output follows the input. The circuit is useful as a "follower" or "buffer" since the input impedence is very high in a practical case (infinite in the ideal case), and the output impedence is very low. This is useful when we want to distribute a voltage for use elsewhere without disturbing the source.

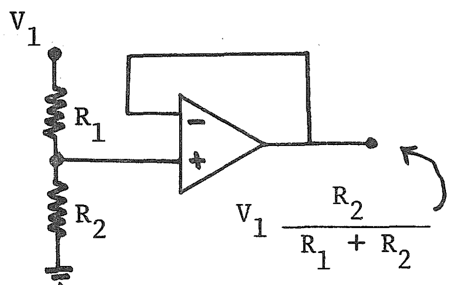
As an example, consider the use of the follower to monitor the voltage on a capacitor. When the switch opens, the last voltage on the capacitor is held, and the follower transfers this voltage to the output. The limit here is of course the actual bias current that flows into the op-amp terminals. The



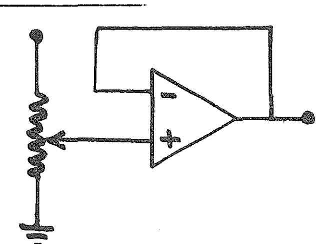
circuit is an elementary "sample-and-hold" device. In practical cases, the bias current for certain FET input op-amps can be as low as a few picoamps.

The follower will also leave a voltage divider undisturbed to the extent that to actual bias current is smaller than the standing current in the divider. Circuits that actually draw substantial current from the junction point will change the voltage as well. Several circuits that are basically voltage dividers are shown below:

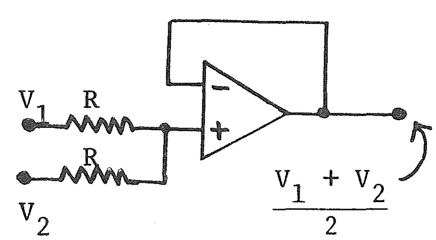
BUFFERED ATTENUATOR



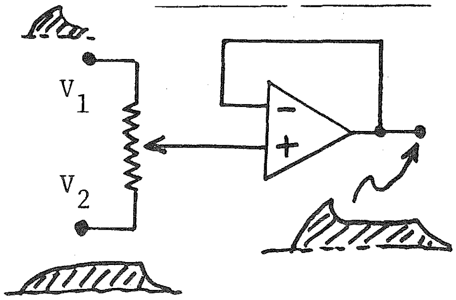
ADJUSTABLE BUFFERED ATTENUATOR



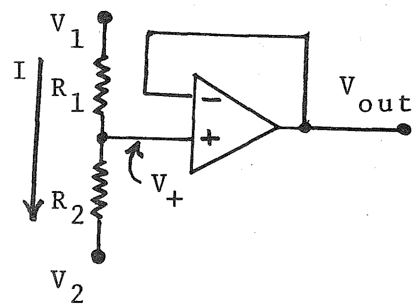
AVERAGER



WEIGHTED AVERAGER



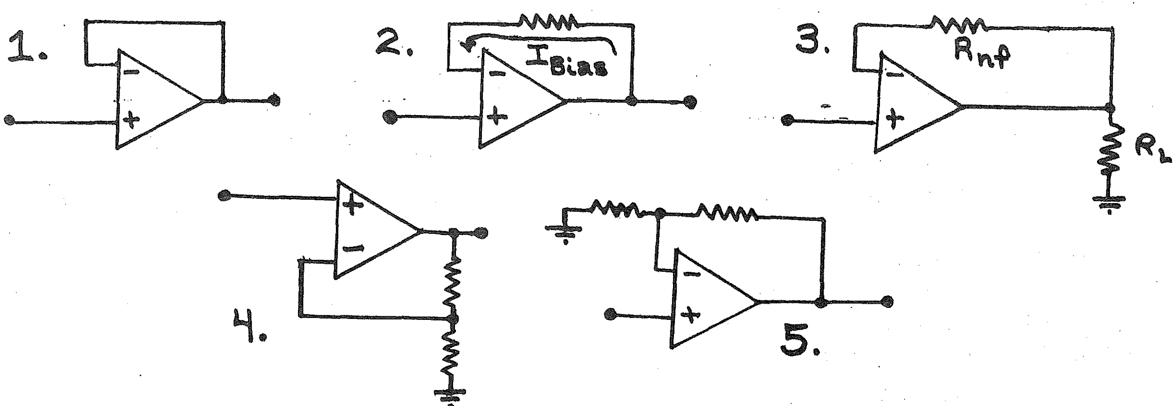
GENERAL ANALYSIS METHOD



$$I = \frac{V_1 - V_2}{R_1 + R_2}$$

$$V_{out} = V_+ = V_2 + I \cdot R_2$$
$$= V_2 + R_2(V_1 - V_2)/(R_1 + R_2)$$
$$= \frac{V_2 R_1 + V_1 R_2}{R_1 + R_2}$$

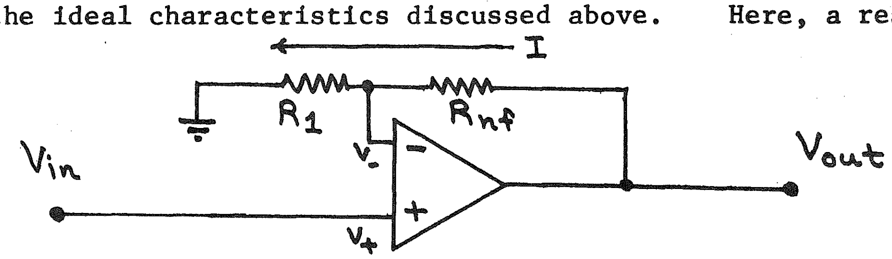
It should be pointed out that there are common ways of drawing certain op-amp configurations, but no real standards. One must always be careful to look carefully at the structure to see what is actually present. Of the five circuits below, the



first three are followers; the last two are non-inverting amplifiers (to be discussed below). The first circuit is a standard follower of the type we have discussed. The second circuit uses a resistor for feedback rather than just a wire. Note that for an ideal op-amp, it makes no difference whether this is a wire or a resistor, since no current flows in either case because the input will not accept or source any. In a practical case with a real op-amp, this resistor passes the amplifiers input bias current and this will cause a voltage drop to appear between the - input and the output. In some op-amps, this improves the performance, compensating for other non-zero offsets. The third circuit is a follower with a load resistor shown. This should not be confused with circuits of the type shown by the last two illustrations. These last two are non-inverting amplifiers. The circuit of the last two is exactly the same except for the way it is drawn.

THE NON-INVERTING AMPLIFIER

A non-inverting amplifier circuit is shown below. The circuit can be understood in terms of the ideal characteristics discussed above. Here, a resistor voltage



divider is used to feed part of the output signal back to the inverting input. The inverting input draws no current, so the voltage at the inverting input is determined by:

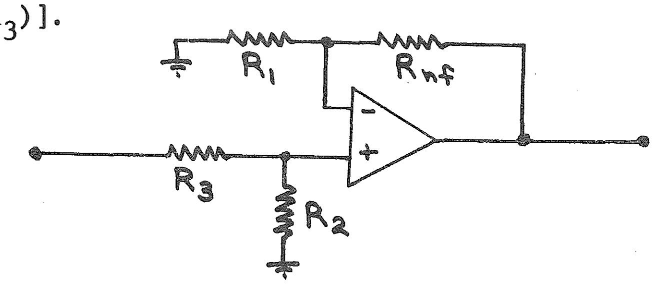
$$V_- = R_1 I = R_1 \, [V_{out}/(R_1+R_{nf})]$$

And since the differential input voltage is zero due to the negative feedback, $V_- = V_{in}$. Combining these two equations gives:

$$V_{out}/V_{in} = (R_1+R_{nf})/R_1 = 1 + R_{nf}/R_1$$

The circuit thus has a gain of one plus the ratio R_{nf}/R_1. The residual gain of one is the gain of the follower as can be seen by letting R_1 approach infinity in which case, the gain goes to one. The non-inverting amplifier does not change the phase of the signal. This is the meaning of "non-inverting," as compared to the inverting configurations that will be considered later, which change the phase by 180°. For audio signals, often times inversion or non-inversion is of little consequence. Control signals, on the other hand will be moving the wrong way if some provisions for handling inversion are not made.

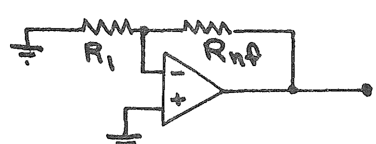
What happens if you need a gain less than one? First of all, to have an amplifier put out a signal with a level that is less than the input hardly seems like gain, or the proper function of an amplifier, but often times, the change of voltage level corresponding to a gain factor less than one is needed, and the term gain is still retained. If you do need a gain less than one, and can't handle the inversion (we will see that the gain of an inverting amplifier goes from zero on up), the best method is to use an attenuator (voltage divider) with a buffer as described above. One can of course add a voltage divider to the input of a non-inverting amplifier as shown below, in which case the gain of the structure is multiplied by the division ratio $([R_2/(R_2+R_3)]$.



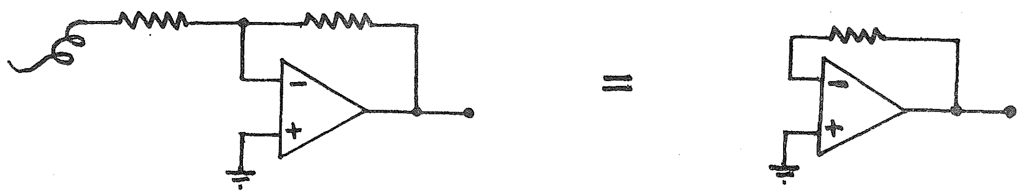
In general, there is no reason to use this structure, since the gain can always be set by just two resistors (R_1 and R_{nf} , or the two resistors on a voltage divider), but the structure may appear unintentionally. R_2 could be the output impedence of of a driving stage while R_3 is a load resistor added to the stage after setting the ratio R_1/R_{nf}. The overall gain would drop in this case. It is sometimes a useful trick to adjust a gain that is slightly too high by adding a load to the driving stage. Finally, compare this structure to the differential amplifier which we will be discussing later on. The grounded end of the resistor R_1 may be found in the case of the differential amplifier whenever the inverting input of the differential amplifier is held at zero.

THE INVERTING AMPLIFIER AND RELATED STRUCTURES:

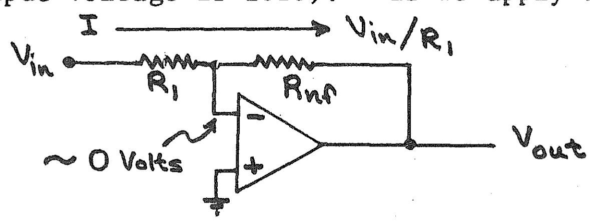
If we take the non-inverting amplifier, and ground the input, we get the (useless) structure shown below:



Since the input is zero, so is the output. Thus, both ends of both R_{nf} and R_1 are held at ground, and no current is flowing in either resistor. We can them lift the grounded end of R_1 off ground and let it float (which is the same as removing the resistor completely, but we leave one end connected for now). What we have now is a follower that is following ground potential.



Actually, we have arrived at the basis of a very important structure - the inverting amplifier. This is often analyzed in terms of what is called a "virtual ground," which means that the inverting input remains at ground potential, but no current passes between it and the circuit ground. A virtual ground is a consequence of the second ideal property (no current enters the inputs) and the third (the differential input voltage is zero). If we apply a voltage V_{in} to the



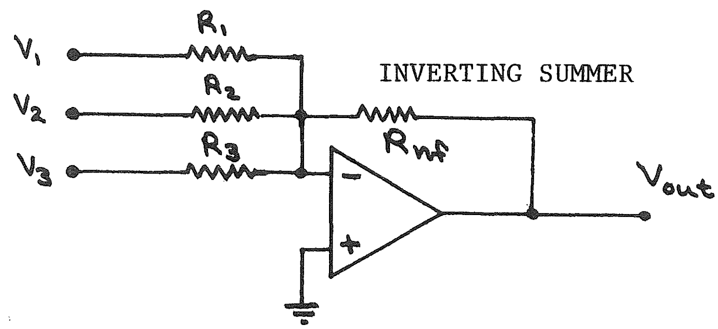
free end of R_1, a current V_{in}/R_1 flows in through R_1. The current can not enter the op-amp input, or go anywhere except out through the feedback resistor R_{nf}. Therefore, the output must go to a voltage $V_{out} = -I \cdot R_{nf}$ or:

$$V_{out}/V_{in} = -R_{nf}/R_1$$

The minus sign results from the fact that the current through R_1 is toward ground, while the current through R_{nf} is away from ground. This minus sign means that the amplifier has an 180° phase shift, and this is called an inverting amplifier. If we put an input of V_{in}= 1 volt, and R_1 is 10k while R_{nf} is 100k, we get an output of V_{out} = -10 volts.

What is the inverting amplifier good for? First of all, it is good for inverting, i.e., we might want to change a 5 volt signal to a -5 volt signal, and would therefore use an inverting amplifier with a gain of one. Secondly, since we don't have the residual gain of one, we can easily set gains less than one by adjusting the ratio R_{nf}/R_1. In audio circuits, the inverting amplifier may have certain bandwidth advantages for unity or relatively low gains since it is possible to use less compensation in these cases. One drawback is that the input impedence of the inverting amplifier is just R_1, even for an ideal op-amp, and this may be a little low depending on the source impedence. The non-inverting amplifier by comparison has a very high (ideally infinite) input impedence. However, by far the importance of the inverting structure appears when we consider multiple inputs. In this case, the inverting input becomes what is called a "summing node."

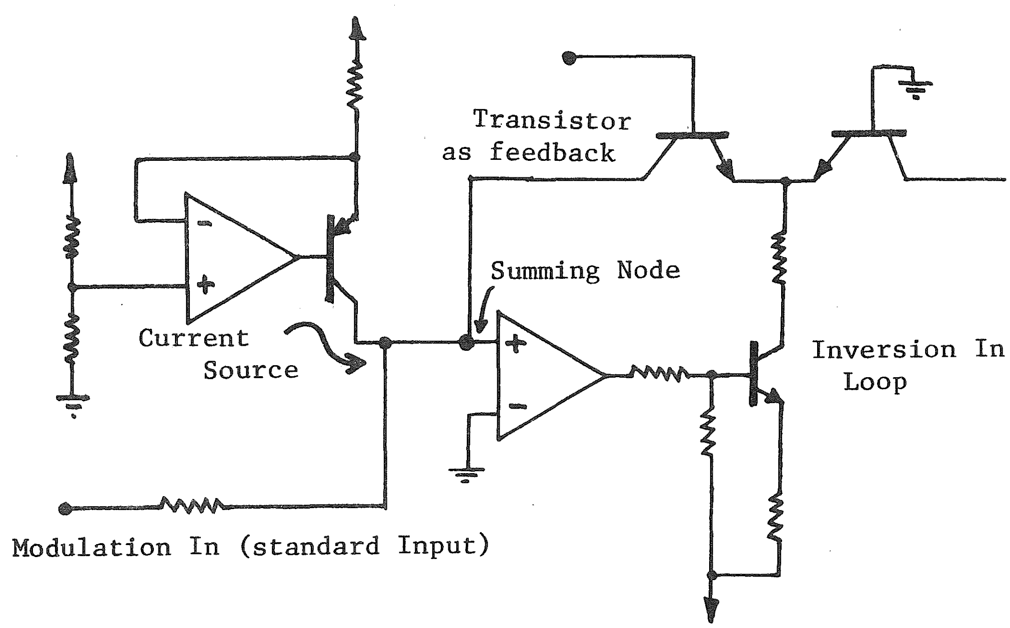
The "inverting summer" is formed by connecting a number of resistors to the inverting input of the op-amp inverting amplifier. It is possible to connect as many resistors to this point as you want. The currents that flow into (or out of) the summing node are algebraically summed. Finally, the output adjusts to whatever value is necessary to supply or remove a current through R_{nf} so that the net current in the summing node is zero. If the output is asked for a voltage that is in excess of the power supply limits, the summing node is no longer held at zero volts since negative feedback fails. An example of a summer is shown below:



If $R_1 = R_2 = R_3 = R_{nf}$, then $V_{out} = -[V_1 + V_2 + V_3]$, the simple negative sum of the voltages. In the more general case, the equation is:

$$-\frac{V_{out}}{R_{nf}} = \frac{V_1}{R_1} + \frac{V_2}{R_2} + \frac{V_3}{R_3} + \ldots\ldots$$

Two additional points about summing nodes should be brought out. First, they are current summing nodes, and it is not necessary that the current be supplied to them by means of dropping voltages across resistors. Current may arrive by a "current source", and you will often see transistor collectors, the output of CA3080's, and other devices connected directly to the nodes. Secondly, the return path that supplies the excess current need not be through a resistor or even direct from the output. It is not necessary that the output of the op-amp actually be putting out current corresponding to the excess amount. After all, the output is probably sending current to external circuits as well. What is necessary is that the output voltage causes the excess current to be supplied from somewhere. This overall feedback path may involve inversion as well, in which case negative feedback is fed to the non-inverting terminal, and the + input is thus the summing node. A circuit that is part of an exponential converter is shown below. It illustrates the points mentioned in this paragraph.

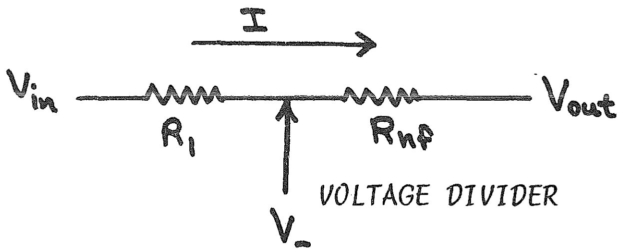


* * * * * * * * * * * * * * * * * *

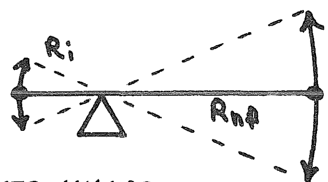
AN ALTERNATIVE ANALYSIS OF THE INVERTING AMPLIFIER

The concept of zero differential input voltage (and virtual ground) is correct and useful, but often seems like a swindle! Why should the input current be allowed to force its way around and out through feedback resistors? In fact, the concept of virtual ground is very easy to use, but another analysis perhaps gives a better idea as to what is actually going on. Since the inverting amplifier is so important, we will go through another analysis.

We start with the same amplifier structure, and will consider the input voltage to be positive. It is the usual practice to consider current to flow from positive to negative, even though electron current flows the other way. It makes no difference as long as we are consistent about this in a given circuit. We can also assume all input voltages to be positive so that current flows into the summing node. If we guess wrong, this simply means that the current will turn out to have a negative sign associated with it. As with the non-inverting



VOLTAGE DIVIDER



LEVER ANALOG

amplifier, we have only to look at the voltage divider comprised of R_1 and R_{nf}. This gives a voltage at the inverting input of:

$$V_- = V_{in} - (V_{in} - V_{out})\ R_1 / (R_1 + R_{nf})$$

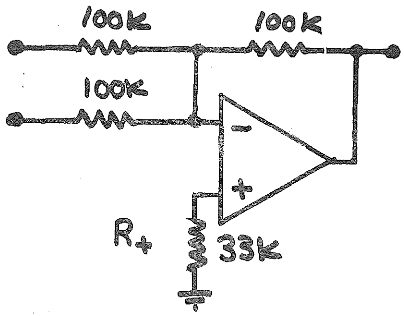
Next we consider what happens if V_+ is different from V_- . As in the case of the follower, we can see that the negative feedback causes the differential input voltage to go to zero. Since the + input is grounded, this means that V_- is zero. This leaves:

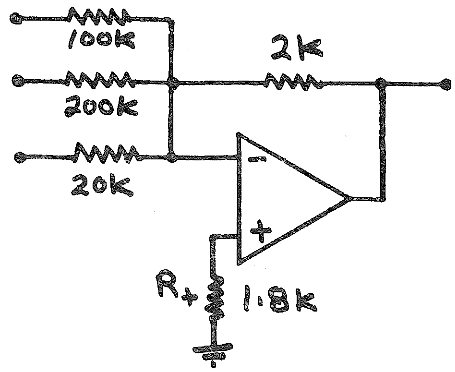
$$V_{out}/V_{in} = -R_{nf}/R_1$$

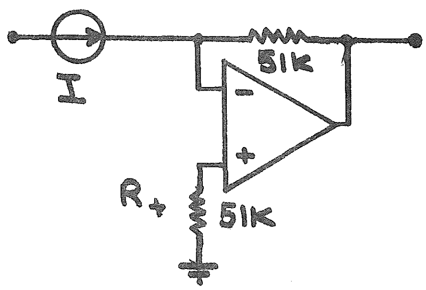
which is the same equation we had before for the gain of an inverting amplifier. Note in particular that this formulation has an exact analog in the equation for the circular motion of the end of a lever. The length of the lever beyond the point of rotation is proportional to the resistance value. This is an extremely useful analog to keep in mind.

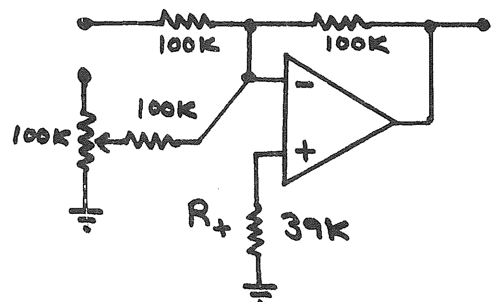
* * * * * * * * * * * * * * * * * *

When a high precision summer is needed, and in many if not most cases of general use, a resistor is placed between the + input and ground, rather than just grounding the pin directly. You should know that while this is sometimes very helpful, it is a relatively small effect. The purpose of the resistor is to compensate for the offset due to the current that actually flows into the other input. Rather than go into the theory, we will just say here that the value of this resistor is the same as the result one would get if the input resistors <u>and</u> R_{nf} were all put in parallel. However, bear in mind that since the actual current through this resistor is very small (zero for the ideal op-amp), it is not worth getting too fussy about this value. A value within 20% should be fine, even in cases where the actual input and feedback resistors are 1% types. Also, in a couple of cases, one can not be sure what input resistors are actually in use. If you have all the inputs connected to panel jacks, you may at times not use them all, and the other resistors will be effectively out of the circuit. Unused inputs could be shorted to ground with a shorting type of jack if this were necessary. Other input resistors might be connected to pot wipers, and here the effective resistance depends on the pot setting. You can't always win at this game, so you have to bear in mind that it probably does not make much difference. Calculation of this resistor, which we will call R_+ is shown for a couple of examples below:

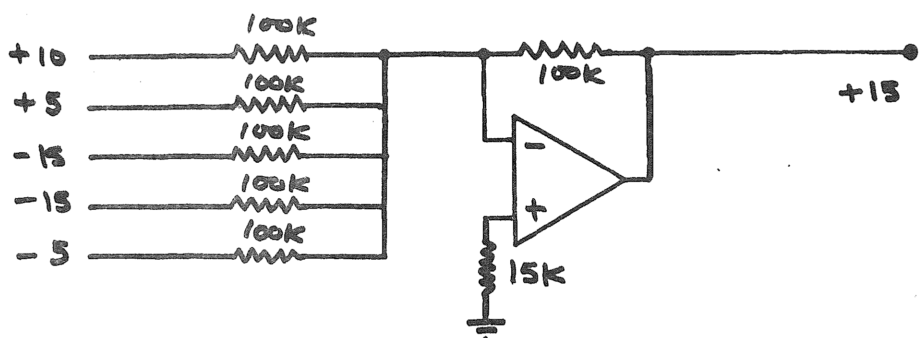






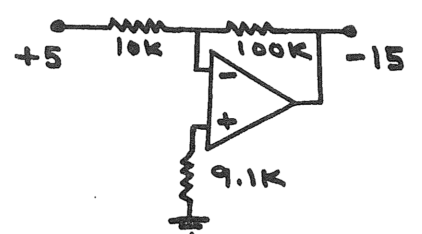


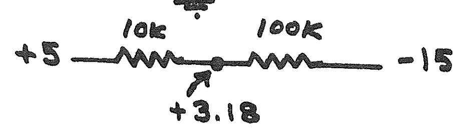
The essential point about negative feedback is that the differential input voltage is zero - as long as the output can adjust to do whatever is necessary to make it so. The output of course can not exceed its supply limits. Consider the circuit below:



The actual sum of the input voltages is +10 +5 -15 -15 -5 = -20. Thus the summer is asked for an output of +20 volts. It can't give more than +15.

A similar situation exists when the inverting amplifier is asked for too much gain. In the circuit at the right, the circuit only asks for a gain of 10, but the input is +5, so the output should be -50, but of course, it is only -15. We can see what has happened by considering the voltage divider comprised of R_1 and R_{nf}. As the voltage rises toward +5, the output falls toward -15, and once it gets there, it stops at -15. When the input actually gets to +5, the center of the voltage divider is at +3.18 volts. Thus the - input is at +3.18 volts while the + input is grounded. The op-amp has gone into the comparator mode. The controlled gain closed loop ($-R_{nf}/R_1$) has failed at -15 volts, and the open loop has appeared.

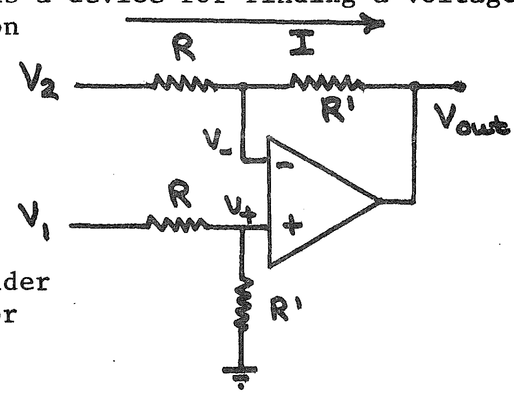




Another case where the differential input can be greater than zero with negative feedback is the case where the output can not keep up with changes of the input. This can occur when a very high frequency is applied to the input. For example, suppose a 100 kHz ±10 volt square wave is applied to the input of a unity gain inverter formed from an op-amp with a maximum slew rate of 0.5 volts/microsecond (e.g., the type 741). In the 5 microseconds between changes in the square wave, the output can only change by 2.5 volts, not the full 20 volts. It is thus making little triangle waves about zero volts while the input switches at +10 and -10. The - input is thus jumping between approximately +5 and -5 while the + input is grounded.

THE OP-AMP DIFFERENTIAL AMPLIFIER

The differential amplifier, or adder-subtractor, is a device for finding a voltage difference between two signals. The basic configuration is shown at the right. The analysis is fairly simple but is often confused. Start by observing that V_+ is obtained from V_1 by a simple voltage divider and the op-amp differential voltage is zero. Thus:



$$V_+ = V_1 \frac{R'}{R + R'} = V_-$$

It is then apparent that $I = (V_2 - V_-)/R$. Next, consider that the current I must flow out through the R' resistor from V_- to V_{out}. Thus:

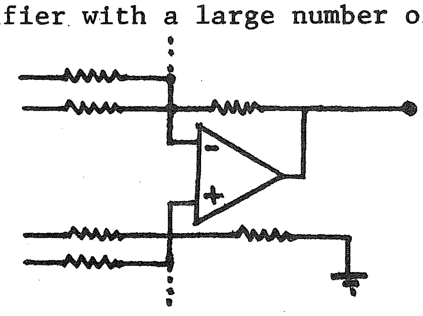
$$V_{out} = V_- - R'I = V_- - \frac{R'(V_2 - V_-)}{R} = \frac{R'}{R}(V_1 - V_2)$$

One common problem that occurs when the differential amplifier is applied is that the resistors R-R and R'-R' are not accurately enough matched. This is a particular problem when a high gain circuit (R'>>R) is used. What happens is that a "common mode" voltage causes the output to pin at one of the supply levels. A common mode voltage is one that is present on both inputs. For example, if you have two voltages of 10.0 and 10.2, and want to take the difference between them and multiply by 40, you expect 8 volts out. If however the resistors are not well matched, the common mode (10.1 volts) is amplified and may be more than enough to cause a serious error or to pin the output at one of the supply levels

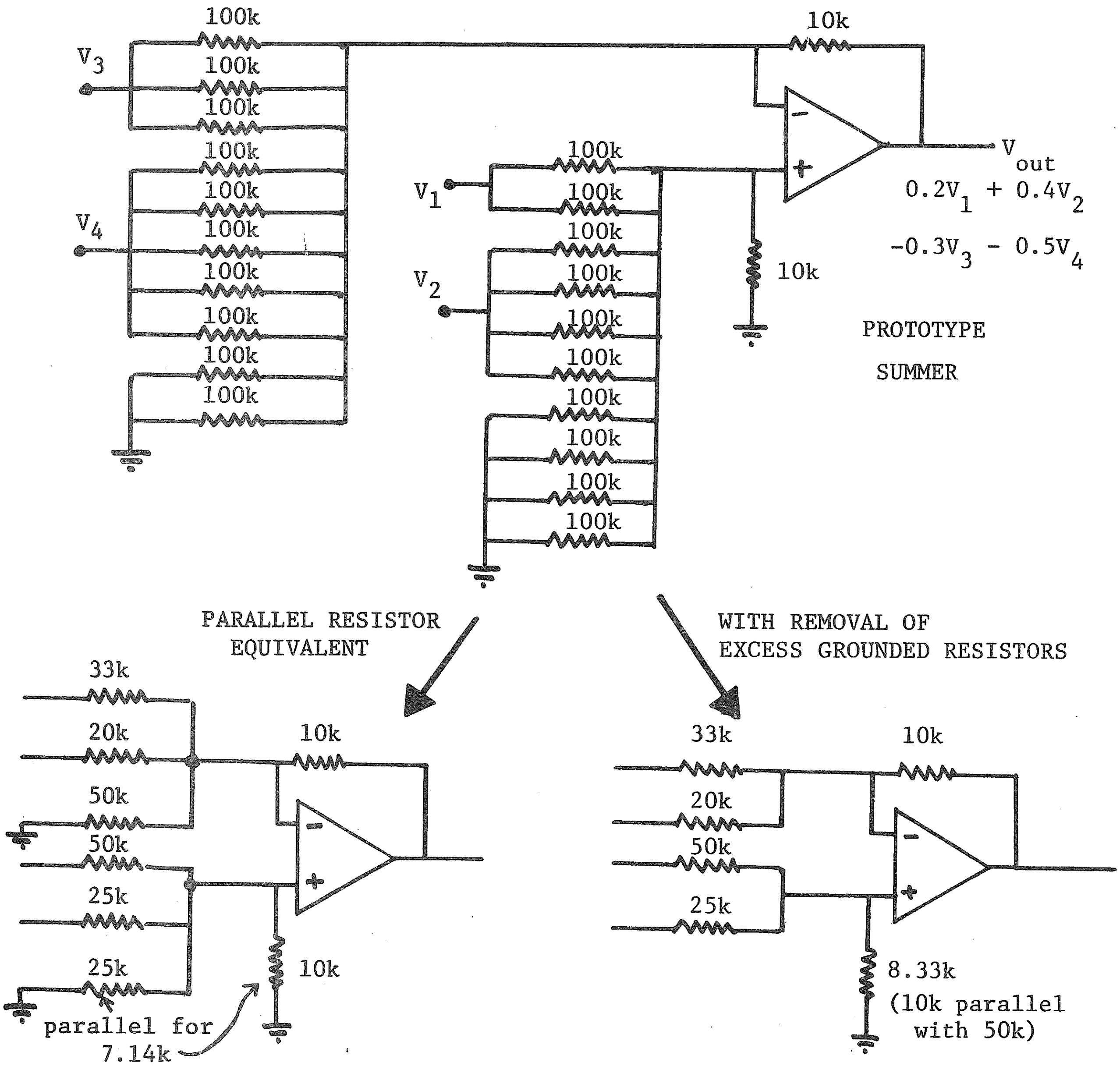
THE GENERALIZED ANALOG SUMMER

An extension of the differential amplifier is to allow for the formation of a generalized summer. That is, to allow for the summation of any number of voltages with either an inverted or non-inverted phase, and with arbitrary weightings. It is of course easy to perform this function with two op-amps, cascading one inverting summer into the second. In a single application for one piece of equipment, this is often the easiest thing to do. If a lot of units are to be made, the calculations may be worth while.

The first step is to expand the differential amplifier with a large number of additional inputs as shown at the right. The same number of inputs should be added to each side. Following the same analysis as for the differential amplifier, it can be shown that the device sums the voltages with equal weighting. Next, suppose the job is to sum $+0.2V_1 + 0.4V_2 - 0.3V_3 - 0.5V_4$. We select an input resistor $R = R'/10$ and use ten of these inputs on the + side and 10 on the - side.



We can then perform the specified summation by selecting two of the + inputs for V_1, 4 of the + inputs for V_2, three of the - inputs for V_3, and 5 of the - inputs for V_4, and ground all unused inputs. Then we can represent all the paralled resistors by their equivalent resistance. This gives the circuit on the left below. We could make the parallel combination of the 10k resistor and the 25k resistor, but a different simplification is possible. Note that we grounded four of the + inputs and two of the - inputs. Thus, two grounded resistors can be removed from each side. The final circuit is on the right below:



GENERALIZED ANALOG SUMMER CIRCUITS

In the above example, it was possible to obtain a simple network by removing the duplicate parallel resistors to ground. This method is straightforward and can be used to arbitrary accuracy by choosing enough inputs (on paper) to start with. Two points should be brought out though: (1) The unused inputs, if any in a given application, should be grounded, not left to float, and (2) In some cases, there will be more excess grounded resistors on the - input, and there will have to be a resistor to ground from both inputs. The whole procedure can be reduced to simple formulas to make things just a matter of calculation if this is desired. The procedure has been given by R. Kostanty in Electronics, Feb. 12, 1972 and by J. Fidler in Electronic Engineering, Nov. 1973. The basic steps are:

(1) Select a feedback resistor R'. Calculated input resistors R_i to give the desired gains on the inverting side (R'/R_i). Sum all the inverting gains to get K^- (a positive number).

(2) In the same manner and for the same value of R', select input resistors R_i for the non-inverting side to give the desired gains (R'/R_i). Sum these gains to get K^+ (which is also a positive number).

(3) Calculate a resistor value $R_x = R'/(1 + K^- - K^+)$. If this value is negative, connect it between the - input and ground, otherwise connect it between the + input and ground.

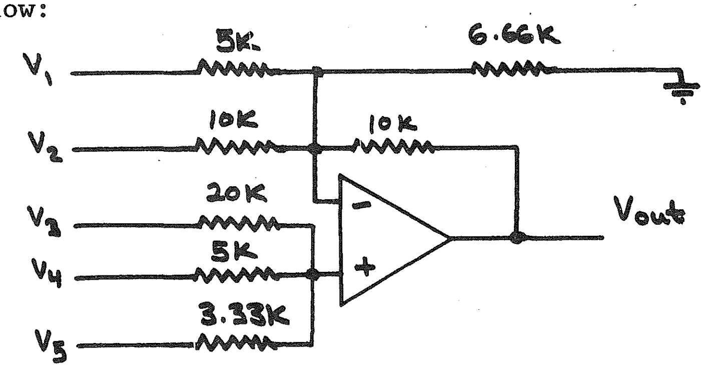
EXAMPLE: Find a summer to give $V_{out} = -2V_1 - V_2 + 0.5V_3 + 2V_4 + 3V_5$.

(1) $R' = 10k$. $R_1 = R'/2$. $R_2 = R'/1$. $K^- = 2 + 1 = 3$

(2) $R' = 10k$. $R_3 = R'/0.5$. $R_4 = R'/2$. $R_5 = R'/3$. $K^+ = 0.5 + 2 + 3 = 5.5$

(3) $R_x = 10k/(1 + 3 - 5.5) = -6.66k$. Connect a 6.66k resistor from the - input to ground.

The network is shown below:



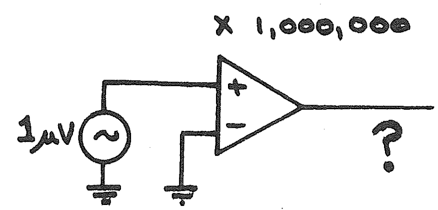
WORKING WITH REAL OP-AMPS

Working with real op-amps is almost always a question of selecting the right type of IC op-amp. The first rule when selecting a suitable op-amp is that to a first approximation, all op-amps are alike and to some degree approximate the ideal characteristics we have discussed above. Often, op-amps are compared to the type 741 which is an industry standard. It is a standard not in the sense that it is outstanding, but in the sense that you often hear such expressions as: faster than the 741, or draws less bias than the 741. The principle performance parameters of interest for electronic music are slew rate and input bias current. IC op-amps fall into two classes with respect to the way they are stabilized - internal compensation or external compensation. The internally compensated devices are essentially 5 terminal IC's - two terminals are + and - power supplies, the others are the two inputs and the output. Externally compensated op-amps have extra terminals that give access to the internal workings of the chip. Compensation is accomplished by adding external capacitors and resistors to these terminals. The advantage of internal compensation is simplicity; that of external compensation is flexibility.

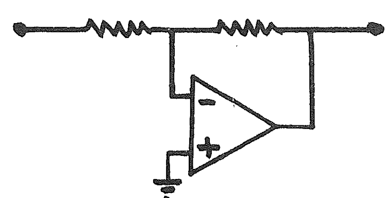
It should be mentioned at this point that one property of all real op-amps is that they must be supplied with a voltage power supply. In many cases, the connections for the power supply are not shown on schematics. In such cases it is usually the case that a standard +15 and -15 bipolar supply is implied. Op-amps usually do not have a pin to be grounded, although one or more pins may be grounded in a given application. While the ±15 bipolar supply is standard, it is not necessary in all cases. Some op-amps will run on as little as ±1.5 volts, an unbalanced supply (e.g., +12 and -6), or a unipolar supply (e.g., +9 with the negative pin connected to ground). The output voltage of course can never exceed the supply limits. One final note about the supply connections. It is good practice to bypass the supply lines by connecting a capacitor of about 0.1 mfd from the supply line to ground near the supply inputs to each amplifier. In many cases, this can be done for only every fourth or fifth amplifier. It is a good thing to do to avoid trouble and when trouble appears, it is a good thing to try to correct such things as high frequency noise and spikes that don't belong in the output.

COMPENSATION AND SLEW RATE

It is generally known that an op-amp (or any amplifier) must be properly compensated to prevent oscillation. This is really only true of feedback amplifiers, or perhaps is is safer to say: amplifiers where a portion of the output signal can reach the inputs either intentionally or unintentionally. It is sometimes assumed that oscillations and other instabilities are due to the high gain, but there must be feedback. The op-amp at the right might have an open loop gain of 1,000,000 at the frequency we are interested in, and we might want to amplify a signal with a 1 μvolt level. We expect a 1 volts signal out. A practical amplifier however will probably have offset, drift, and noise parameters such that the circuit will not even come close to working - the offset will probably cause it to act as a comparator. However, it will not oscillate since there is no feedback.

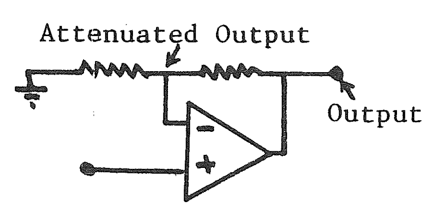


We have said that compensation is needed to prevent oscillations due to feedback. However, if you set up the simple amplifier shown at the right with negative feedback, you expect only negative feedback, and this will not cause oscillation. The simple analysis ignores the possibility of phase shift within the amplifier. There will always be some phase shift even though you look at the internal schematic of the op-amp and see no capacitors. The problem is that things are much too close inside the IC - significant stray capacitances are a necessary byproduct of small size. This is the reason that discrete components must often be used for some high speed circuits.

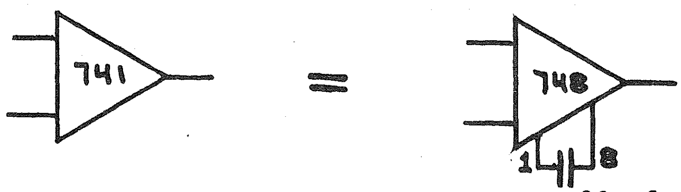


A stray capacitance from the input to the output (and numerous other capacitances inside) may make negative feedback into positive feedback. Phase shift is a function of frequency and can be expected to increase as frequency does. Thus at some frequency, the total phase shift can add to 180°, and if the gain at that point is 1 (0 db), the amplifier will start to oscillate. To determine the gain, we must look at the full loop signal path from the output back to the input. Any attenuation in this loop will help prevent oscillation. It is the puropse of compensation to reduce the gain so that when the phase shifts reach 180°, the gain is less than one.

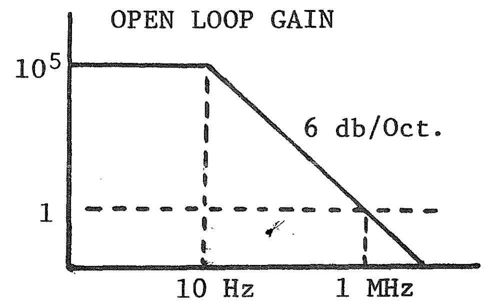
Although we commonly think of oscillation as being associated with high gain, it is the unity gain follower that requires the most compensation. This is because the output signal is fed directly back with no attenuation. If the amplifier is set up for higher gain, then the output signal is attenuated as it is fed back - this is how we get the desired gain with negative feedback. On the other hand, if we compensate for unity gain in all cases, we just slow things down.



We choose as a basis for comparison the 748/301 type of op-amp. The two types are fairly similar. To relate this to the popular type 741, we have only to connect a 30 pf capacitor across pins 1 and 8 of the 748, and we get a 741.



The 741 (or 748 with 30pf of compensation) is unity gain compensated. This means that if we look at a plot of the open loop gain as a function of frequency, it falls off at 6db/octave since the 30 pf capacitor forms a first order low-pass filter. The open loop gain is forced to 0 db (gain of 1) at a frequency of 1 MHz, well below the point at which the total phase shift could reach 180°. Thus the 741 is stable and in general requires no additional compensation. The only time where more compensation is required is where there is an additional amplifier (e.g., a transistor) in the feedback loop. In this case, the compensation can be added by adding an integrating capacitor between the - input and the output.



In many cases, we are mainly interested in handling and processing audio signals that have relatively large amplitudes. These may require little or no actual amplification. In such cases, the available open loop gain is of interest only to the extent that it improves or degrades the desirable effects of negative feedback (e.g., less distortion, lower output impedence). The main concern with such signals is: will the amplifier slew (change its output voltage) fast enough?

If we have a sine wave $V = A \sin \omega t$, the time rate of change of voltage is:

$$dV/dt = \omega A \cos \omega t$$

and the maximum rate of change of this slew rate is where:

$$d^2V/dt^2 = -\omega^2 A \sin \omega t = 0$$

This means that the maximum slew rate is at the point where $\sin \omega t = 0$, i.e., the zero crossings of $\sin \omega t$. The rate of change is just:

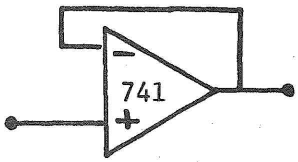
$$\left.\frac{dV}{dt}\right]_{\sin \omega t = 0} = \omega A$$

and since $\omega = 2\pi f$, the maximum slew rate required is:

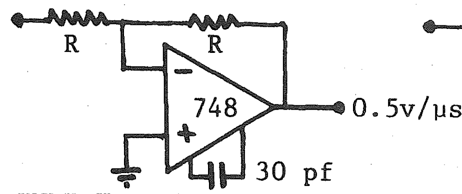
$$SR_{max} = 2\pi f_{max} A$$

The type 741 (748 with 30 pf) has a slew rate maximum of 0.5 volts/microsecond. For a signal with amplitude of 5 volts, this gives a f_{max} of 16 kHz, and only 8 kHz for a signal with 10 volts amplitude. For signals of 10 volt amplitude and frequencies up to 20 kHz, the required slew rate is 1.256 volts/microsecond. These figures apply only to sine waves. A square wave for example would in general have sides that are slew rate limited at any frequency. In general, if we try to exceed the maximum available slew rate, we start to turn waveforms into triangles, distortion will become a problem, and the output amplitude will drop. Thus, with the 741, we can expect problems if we try to go too close to the top of the audio range.

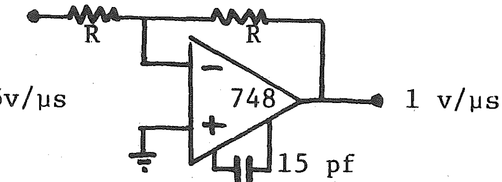
For unity gain operation, the 748 or 301 used as a unity gain inverter has a slew rate advantage if properly compensated. This can be seen from the example circuits below:



FOLLOWER



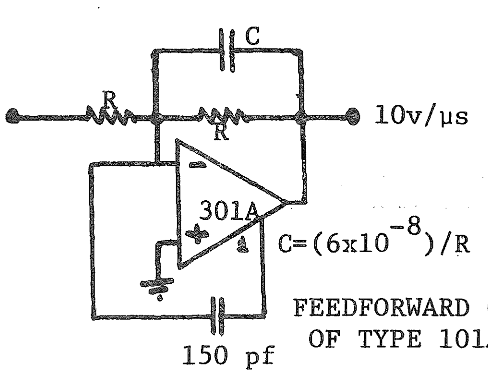
INVERTER WITH STANDARD COMP.



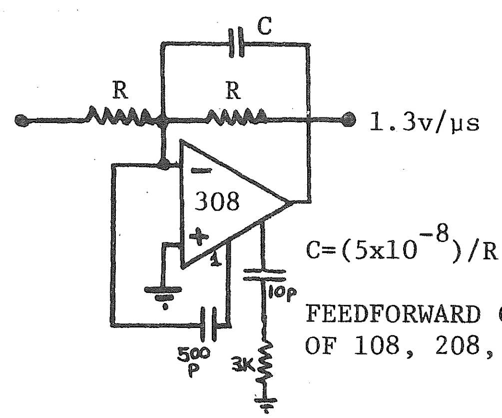
INVERTER WITH HALF THE COMPENSATION

In the case of the inverter, the voltage fed back is divided by two at the - input. This means that the signal that causes oscillation must have twice the amplitude it needs in the case of the unity gain follower. This allows half the usual compensation (15 pf instead of 30 pf) and doubles the slew rate to something like 1 volt/microsecond. For higher inverting gains, or for non-inverting amplifiers with a gain greater than one, the compensation capacitor can be cut back in inverse proportion to the attenuation factor. Thus, the gain of two non-inverting amplifier can be compensated with 15 pf. A gain of two inverter or a gain of three non-inverting amplifier can be compensated with 10 pf, etc. This can be continued until the gain reaches about 10 corresponding to a capacitance of about 3 pf. It is not practical to go much beyond this because either you can't get a capacitor small enough, or you already have it as stray capacitance. Also, it may not be wise to use less compensation than is required to get the slew rate up to the necessary rate as sharp input pulses may cause more overshoot and ringing. In some circuits, you can rig up a "gimmick" capacitor by attaching two lengths of insulated hookup wire to the pins for compensation and twisting them together until oscillation stops. For another method of getting the compensation capacitor down to increase slew rate, see the chapter on mixer design.

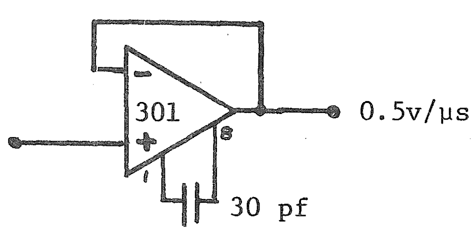
A number of circuits have been developed [R. Dobkin, Nat. Semiconductor Linear Briefs LB-2, LB-4 and LB-14] to increase slew rate by using a different type of compensation. LB-2 and LB-14 discuss feedforward compensation for the types 301A and the 308 respectively. LB-4 shows a method of extending the power bandwidth (essentially the paramether f_{max} used earlier) for the unity gain follower. Some circuits are shown below:



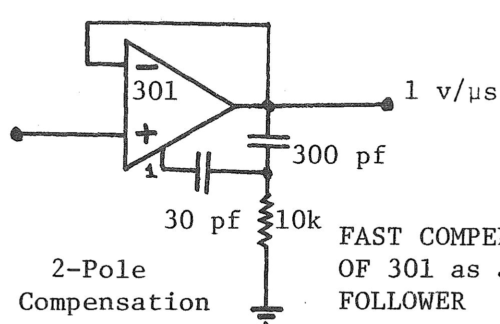
FEEDFORWARD COMPENSATION OF TYPE 101A, 201A, 301A



FEEDFORWARD COMPENSATION OF 108, 208, 308



STANDARD COMPENSATION OF 301 as UNITY GAIN FOLLOWER



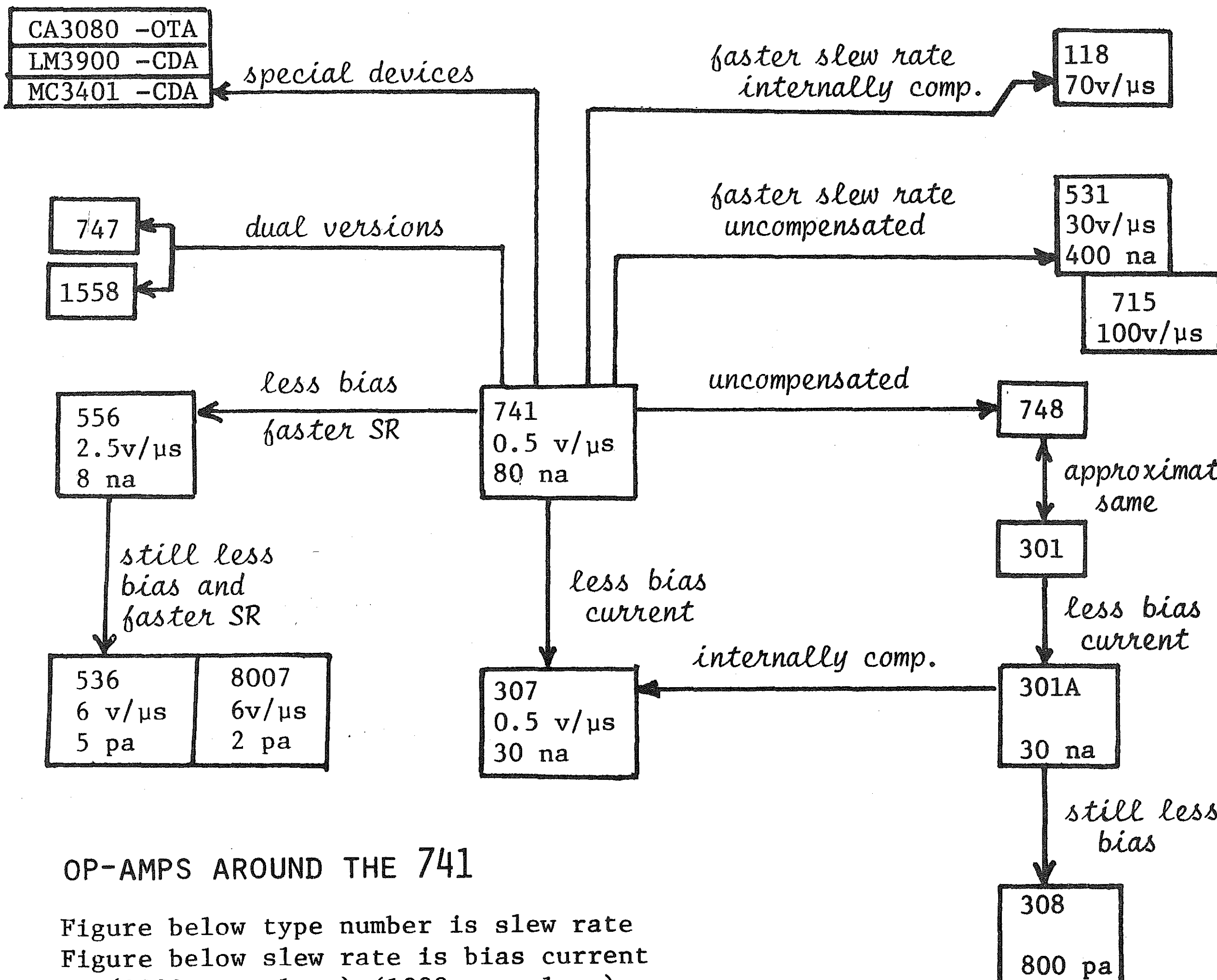
2-Pole Compensation

FAST COMPENSATION OF 301 as a FOLLOWER

INPUT BIAS CURRENT

Input bias current is the actual current the inputs draw. Whether or not this current is significant depends on what it is connected to. In general, it is not too improtant for audio signals, but is more important for control voltages. For example, when used as a follower on a capacitor in a sample-and-hold, the bias current (along with switch leakage) determines the rate of voltage droop during the hold. It is generally fairly easy to calculate what sort of effect the bias current will have on the circuit. The magnitude of the effect is also fairly easy to calculate. For example, the voltage on a voltage divider might be buffered by a follower. You can calculate the standing current in the divider and compare it with the magnitude of the bias current that the op-amp will steal. Once this is know, it has to be determined if the magnitude of the effect will significantly change the designed performance.

Below is a summary chart giving bias current and slew rate for a number of common IC op-amps using the type 741 in the center of the chart as a standard.



OP-AMPS AROUND THE 741

Figure below type number is slew rate
Figure below slew rate is bias current
(1000 pa = 1 na) (1000 na = 1 µa)
Slew Rates for uncompensated op-amps
depend on custom compensation of the device

CHAPTER 3B

BASIC APPLICATIONS OF:

OPERATIONAL TRANSCONDUCTANCE AMPLIFIERS

CONTENTS:

INTRODUCTION

The Operational Transconductance Amplifier (or its functional equivalent) is the single most important component for electronic music devices. Nearly every device requires a wide range controlled gain block. The OTA in common use is the RCA type CA3080. It has a differential input stage that is driven by an externally programmable bias current that controls the gain in a manner similar to the two quadrant transconductance multiplier (see chapter 5c on VCA design). There are no resistors on the chip - instead a series of transistor current mirrors are used. The output is a current rather than a voltage.

The fact that the OTA is current controlled is extremely important since it can be directly coupled to a wide range exponential current source. [It would be more convenient if the OTA were controlled by a current drawn out of it as it could be coupled directly to NPN exponential sources. The OTA requires a PNP exponential source or a current mirror. If a PNP source is used, an inverting stage is required to make control voltage response conventional - higher control voltages represent an increase in the controlled parameter.]

As a controlled gain stage, the OTA forms the basis of VCA's and VCF's. It can also be used as a signal switch for devices like sample-and-holds. With a saturated input, it can be used as a current switch useful for VCO's.

BASIC SETUP OF THE CA3080

The 3080, being a transconductance device has a transfer function $I_{out}/E_{in} = g_m$, where g_m is the transconductance. Since the 3080 has a standard differential two transistor input stage, the transconductance is given by [see chapter 5c]:

$$g_m = I_{ABC}\cdot(q/2k_BT)$$

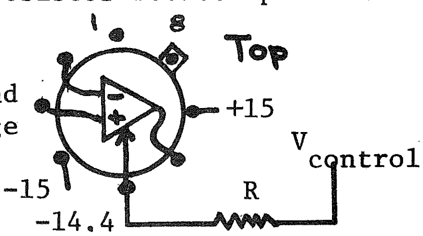
This gives the basic input-output relationship:

$$\boxed{\underset{\uparrow\atop\text{ma}}{I_{out}} = 19.2\ \underset{\uparrow\atop\text{ma}}{I_{ABC}}\cdot\underset{\uparrow\atop\text{volts}}{E_{in}}}$$

BASIC CA3080 EQUATION

I_{ABC} is the amplifier bias current and ranges from about 0.5 microamps to 0.5 ma*. The I_{ABC} can be supplied by a current source, or for a simple setup by a resistor between pin 5 and some voltage between -14.4 volts and +15 volts. Pin 5 of the OTA is approximately one diode drop above the negative supply. With such a setup, I_{ABC} is given by:

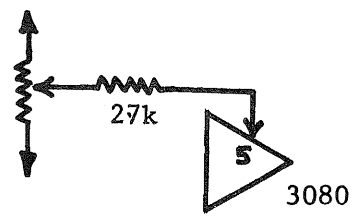
3080 Base and Pin 5 voltage



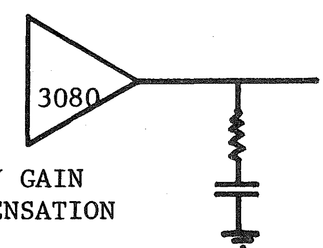
$$I_{ABC} = (V_{control} + 14.4)/R$$

This suggests a good test setup consisting of a 27k resistor connected between pin 5 and the wiper of a pot connected between +15 and -15 as indicated at the right.

TEST SETUP



The OTA is often used "open loop" and no compensation is required. When compensation is required, it has to be approached as a worse case problem, as the actual compensation depends on I_{ABC} and in most electronic music devices, I_{ABC} is a variable. The compensation for unity gain is shown at the right. It is also extremely fortunate that in at least two cases where a closed loop is used (sample and hold, voltage-controlled integrator) the compensation is a part of the circuit anyway.



UNITY GAIN COMPENSATION

TYPICAL GAIN CONTROLLED CIRCUIT

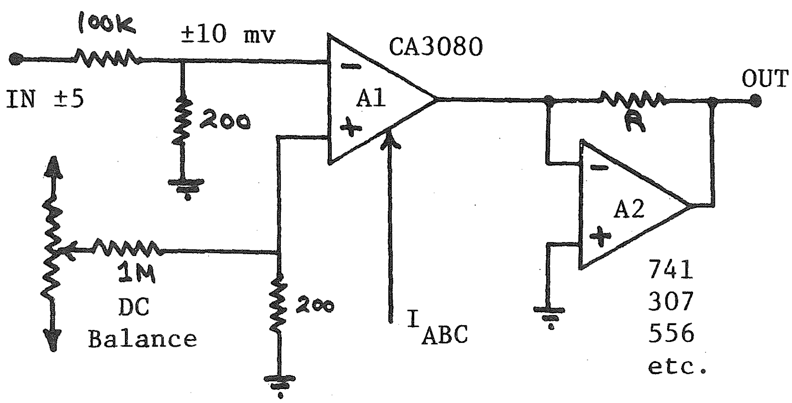
A typical gain controlled circuit is shown at the top of the next page. The circuit can be thought of as the basis of a VCA. The following is an outline of the design:

(1) The circuit is open loop (no feedback) so no compensation is required.

(2) We select input signals in the range of ±5 volts. The actual OTA input stage

---

* Note that the CA3080 can stand I_{ABC} values up to 2ma. There are at least two good reasons for not going this high: (1) The device looses linearity, and (2) It can start to warm up and the 3080 is famous for the ease with which it can be coaxed into thermal runaway. Considerable care should be used to make sure the current to pin 5 is not allowed to rise too high in a test setup. A direct short to ground of this pin, for example, will destroy the device.



BASIC CA3080 CONTROLLED GAIN CIRCUIT

is the ordinary two transistor differential amplifier. Thus, for linearity the inputs should be limited to 10 mV [see chapter 5c]. Therefore, a 500:1 voltage divider (100k:200Ω) is used on the input.

(3) If DC balance is critical, a balance stage as indicated should be used on the + input. If this is not necessary, the + input can be simply grounded.

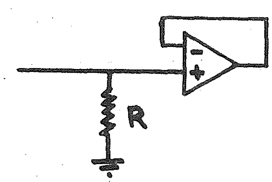
(4) At the maximum value of control current (I_{ABC}) that we choose (say 0.5 ma) we want the circuit to have unity gain. We thus calculate using the basic equation for the 3080:

$$I_{out}(\text{max}) = 19.2\cdot(0.5\text{ ma})\cdot(0.01\text{ volts})$$

$$= 0.096\text{ ma}$$

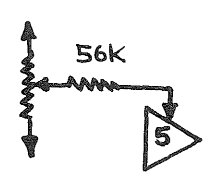
When the current to voltage converter A2 is attached, we want the maximum output current to drive the output voltage to ±5. This gives $R = 5/0.000096 \simeq 51k$

(5) The signal was supplied to the - input of A1 because the current to voltage converter A2 is also inverting. In open loop circuits, the only difference between the inputs to the OTA is the phase. [It may be helpful here to think of the OTA as an op-amp comparator with limited gain. Reversing the inputs changes the output phase by 180°, but the amplifier remains in the linear region.] There are two changes that will give an overall inversion to the output: (a) The + and - inputs can be reversed, and (b) The current to voltage converter at the right can be used.



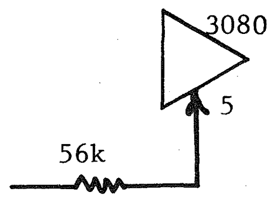
NON-INVERTING CURRENT TO VOLTAGE CONVERTER

(6) The control current can be supplied for a test setup by a 56k resistor connected to a pot wiper as suggested above. The 56k resistor across approximately 30 volts gives the max. current of about 0.5 ma for I_{ABC}.

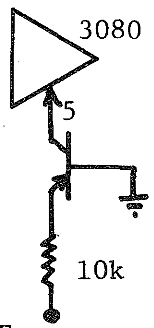


SIGNAL SWITCHING METHODS

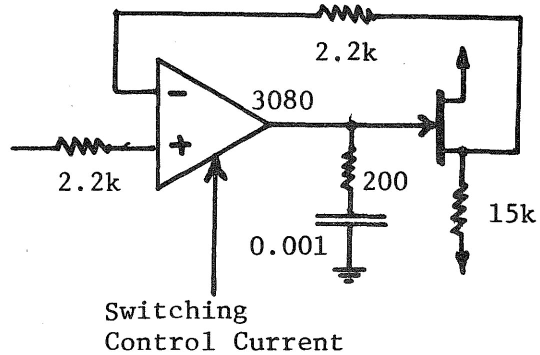
The OTA can be used in two "switching" modes. In the first mode, an audio signal at the input is switched by applying or removing the bias current. The simplest method of doing this is to have the current to the control pin be supplied by a resistor that is connected to a voltage that switches between the supply limits. A second method uses a transistor to drive pin 5 so that a 0 to +5 logic level can be used to control the switching. The CA3080 used in this way is an analog switch. Shown also below is the setup for a sample and hold circuit.



+15 = ON
-15 = OFF



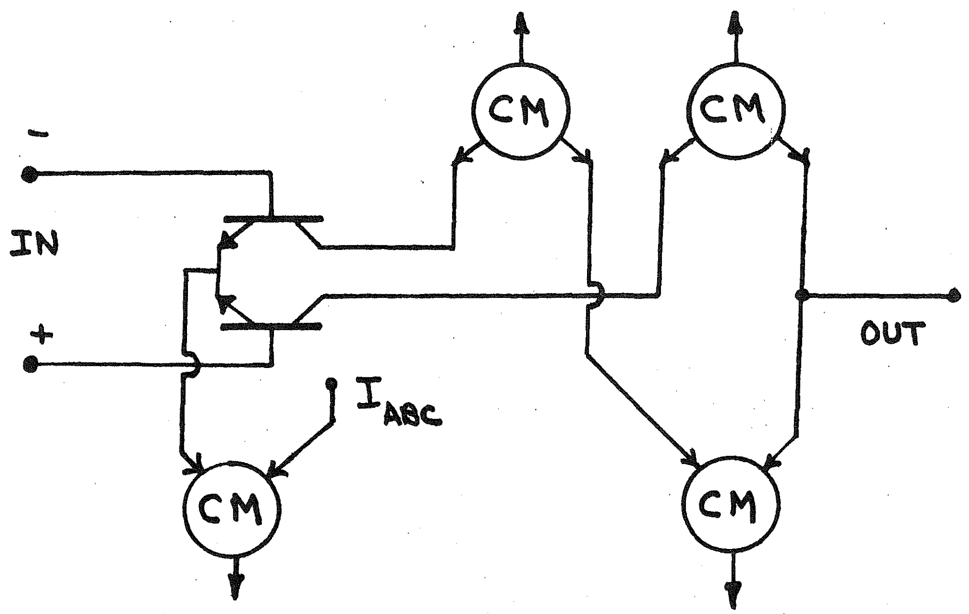
+5 = ON
0 = OFF



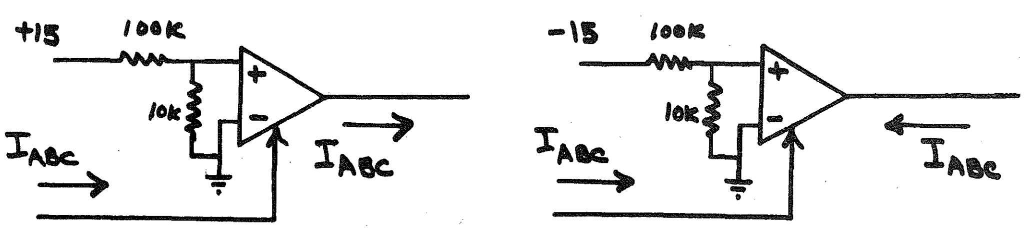
BASIC SAMPLE-AND-HOLD

CURRENT SWITCHING METHODS

The second switching mode is one where the input voltage differential is saturated. This has the effect of transferring the input bias current to the output in a direction determined by the sense of the input voltage. This can best be seen by considering the current mirror structure of the OTA as shown below. When the input voltage differential reaches about 40 mv or higher, most of the bias current passes through one of the transistors while very little passes through the other. The attached current mirrors thus either receive the bias current or no current at all. The OTA in this state is a voltage reversible current source useful for VCO circuits.



CURRENT MIRROR STRUCTURE
STRUCTURE OF OTA



TWO STATES OF CURRENT SWITCHING OTA

CHAPTER 3c

BASIC APPLICATIONS OF:

CURRENT DIFFERENCING AMPLIFIERS

CONTENTS:

INTRODUCTION

The Current Differencing Amplifier (CDA), also known as the Norton amplifier or Current Mode Amplifier is a relatively new device with the following useful properties: (1) It comes in a package of four amplifiers and is inexpensive. (2) It runs off a single power supply voltage ranging 4 to 36 volts (for the LM3900), or bipolar from ±2 to ±18.

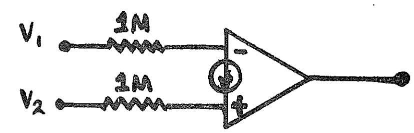
The basic applications considerations for the CDA parallel those of the ordinary op-amp. The main difference is that the inputs are currents instead of voltages. However, there is a wide area in common as far as general current summing techniques are concerned.

Basically, we find (1) The device is a high gain inverting amplifier and thus has an inverting input. The non-inverting input is obtained through use of a current mirror. (2) Inputs set at a voltage of one diode drop (about 0.6 volts). Resistors are used on the inputs to convert voltages to currents. (3) When used open loop, when the + current exceeds the - input, the output is high, otherwise it is low. (4) With negative feedback, the output goes to whatever voltage is necessary to equalize currents into the two inputs.

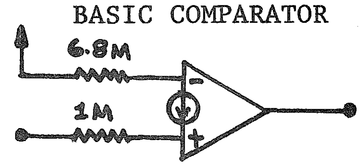
Due to the wide range of supply voltage, the device works nicely with CMOS IC's in many circuits. It is very useful for providing level detectors in such circuits. In fact, a number of logic circuits can be implemented with the CDA that would be difficult to find elsewhere. It can therefore be used in other logic families as well.

OPEN LOOP AND LOGIC CIRCUITS

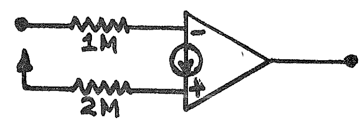
A basic CDA comparator circuit is shown at the right. Currents flow from V_1 and V_2 to the inputs which are at approximately 0.6 volts. If I(+) is greater than I(-) the output is high (positive supply) otherwise it is low (near ground). If negative voltages are to be compared, common mode biasing from the positive supply can be used and the negative voltages then steal current from the inputs. It is also possible to use unequal resistors on the inputs. For example, the circuit on the right has its reference level set by a larger (6.8 meg) resistor connected to the positive supply. When the current through the 1 meg resistor exceeds the reference current, the output will go high. The threshold is set for about 2.7 volts in this case. The third circuit at the right shows how the CDA can function as a logic inverter. The reference current is set into the + input in this case and is set at a value that should separate voltages about half way between logic levels.



BASIC COMPARATOR

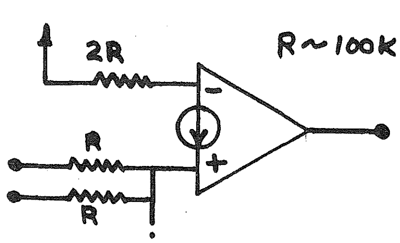


2.7 VOLT THRESHOLD

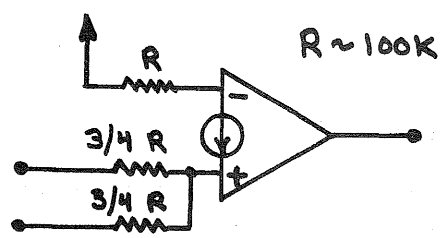


LOGIC INVERTER

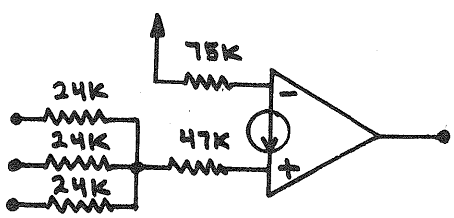
A simple extension of this gives a logic OR gate that can have a very large number of inputs. The AND gate is a little harder to implement except in the 2 and 3 input case. This is because for a very large number of inputs, the difference between all being high and all but one may be very small. For this reason, the diode network shown should be used for more than three inputs. It may be necessary for some logic inputs to add a second series diode to the + input. Either the OR or the AND gates can be made NOR or NAND by simply reversing the + and - inputs.



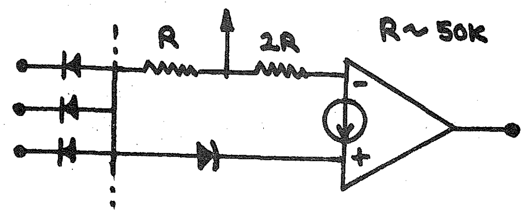
OR GATE USING CDA



2-INPUT AND USING CDA



3-INPUT AND USING CDA
(National Applications AN-72)

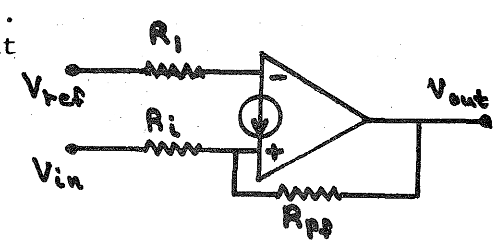


HIGH FAN-IN AND GATE
USING CDA

POSITIVE FEEDBACK

When a resistor is returned from the output back to the + input, a Schmitt trigger is formed in a manner similar to the op-amp Schmitt trigger. The basic circuit is shown at the right. Assume first that the output is low. Then the output will go high when:

$$V_{in} > V_{ref} \frac{R_i}{R_1}$$



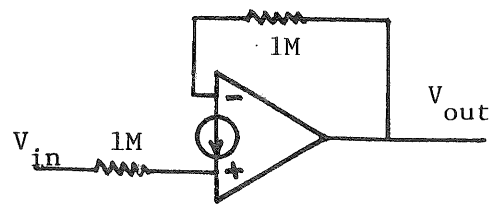
Once the output goes high, current supplied from the output back to the + input helps to hold the output high. The output will again fall when:

$$V_{in} < \frac{R_i}{R_1} V_{ref} - \frac{R_i}{R_{pf}} V_{out} \qquad \text{where } V_{out} \text{ is close to the supply voltage.}$$

The inputs can be reversed to give an inverting Schmitt trigger with a similar analysis.

NEGATIVE FEEDBACK

As was the case with the standard op-amp, we can next consider negative feedback. The simplest circuit is the follower shown at the right. Note that the output voltage is used to supply a current to the input through a resistor.



FOLLOWER

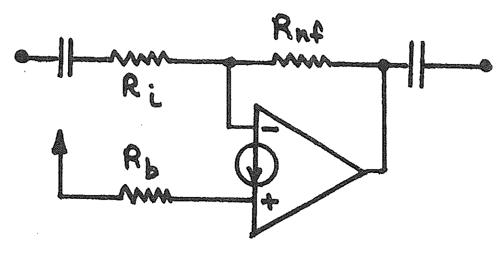
With the CDA, the ideal property is that the current into the - input will adjust to whatever value is necessary to equal the current into the + input as long as negative feedback is working. In the follower, the current into the + input is $(V_{in}-0.6)/R$ and the current into the - input must be the same. This gives:

$$\frac{(V_{in}-0.6)}{R} = \frac{(V_{out}-0.6)}{R} \qquad \text{or } V_{in} = V_{out}$$

Note however that V_{in} must be greater than about 0.6 volts for this application.

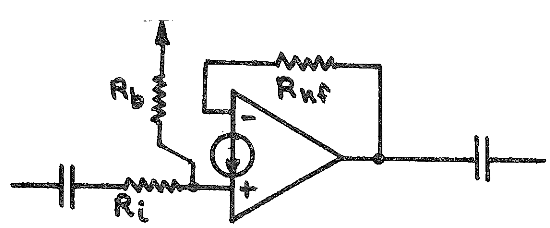
By simply changing the value of one of the resistors, a simple DC amplifier is formed. It is now apparent that with negative feedback the CDA performs a function similar to the op-amp summing node. The difference here is that the node seems to be split as the current sum applies to both nodes - both must receive the same current. The device makes more sense when it is realized that the current into the + node is drawn out of the - node by a current mirror. This is how the non-inverting input was formed.

A simple biasing scheme for an AC amplifier is shown at the right. R_{nf} has been set at $R_b/2$. A current V_s/R_b flows into the + input. Thus, the same current must flow into the - input. Setting the value of R_{nf} at half R_b means that the output is biased at about $V_s/2$. In the absence of an AC signal, the output just sets there. When an AC signal is applied, the current into the - input from R_i changes the DC condition. The output must change to compensate for this so that the total current is still V_s/R_b. This gives an AC gain of $(-R_{nf}/R_b)$, just what one would expect based on the op-amp inverter.



INVERTING AC AMPLIFIER

The non-inverting amplifier is obtained by applying the AC signal through a resistor to the + input instead of the - input. The circuit is shown at the right. The gain of the circuit is R_{nf}/R_i, so there is no residual gain of one as there is in the op-amp case.



NON-INVERTING AC AMPLIFIER

CHAPTER 3D

BASIC APPLICATIONS OF:

ANALOG MULTIPLIERS

CONTENTS:

INTRODUCTION:

Analog multipliers are devices which give the algebraic product of two input voltages. In musical engineering, multipliers find a direct application in the construction of balanced "ring" modulators. They are also sometimes used as the control device in a voltage-controlled module, but their dynamic range may be too limited for many applications. They often appear as a building block circuit in other structures.

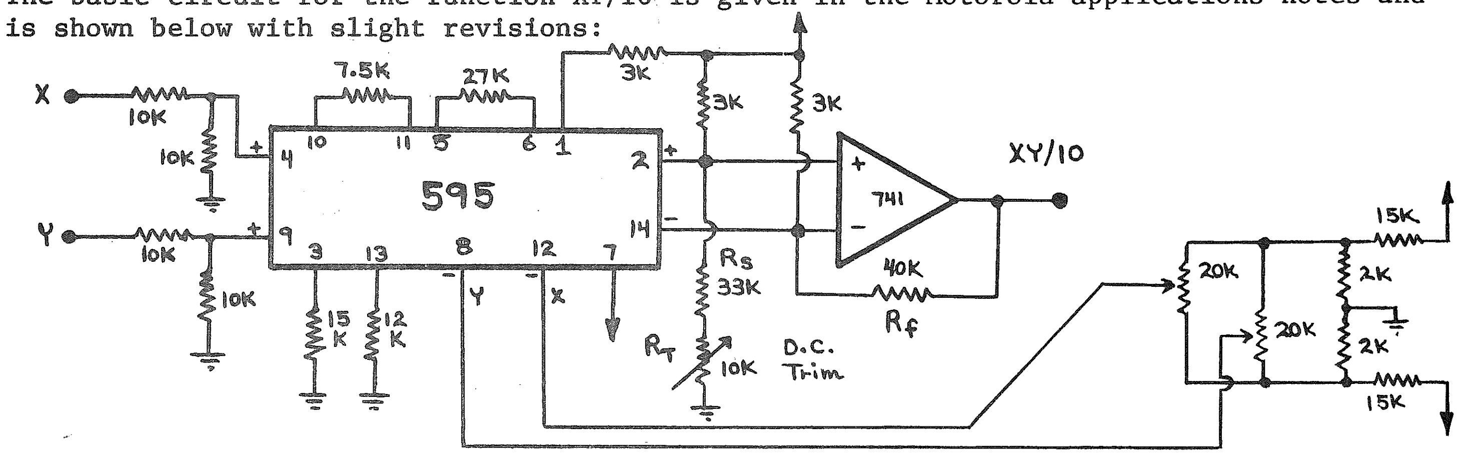
Integrated circuit versions of analog multipliers are now quite common, and can be found in a wide range of price and accuracy. A listing of some common types are given on the next page. We will deal here with the type 595 device which includes the MC1595 and the S5595, but the circuits are the same for the less accurate MC1495 and N5595 types. The 595 was the earliest device developed and is now a standard surplus item in the $1 to $2 range. It is generally accurate enough for most musical purposes. Other multiplier IC's involve fewer external components, but are more expensive. The 595 is quite easy to use once you become familiar with it.

IC ANALOG MULTIPLIERS

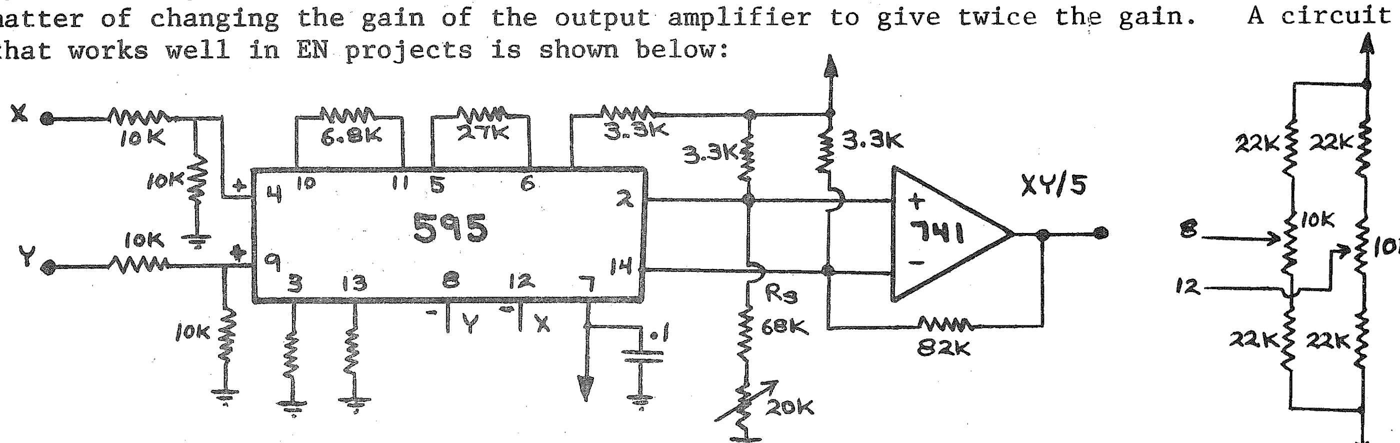
Manufacturer	Type Number	X Linearity Typical	X Linearity Max	Y Linearity Typical	Y Linearity Max.	Approx. Cost New	Approx. Cost Surplus
Motorola	MC1494	0.5%	1.3%	0.5%	1.3%		
Motorola	MC1594	0.3%	0.5%	0.3%	0.5%		
Motorola	MC1495	1.0%	2.0%	2.0%	4.0%	$5	$1.25
Motorola	MC1595	0.5%	1.0%	1.0%	2.0%	$6	$2.00
Signetics	N5595	1.0%	2.0%	2.0%	4.0%	$5	$1.25
Signetics	S5595	0.5%	1.0%	1.0%	2.0%	$6	$2.00
Intersil	8013C	0.8%		0.3%			
Analog Devices	AD531J	0.8%		0.3%		$12	
RCA	CA3091D	1.7%	3.0%	1.7%	3.0%	$6	

APPLICATIONS HINTS FOR THE TYPE 595

The basic application of the 595 as a four quadrant multiplier requires a level shifting op-amp, three or four trimmers, and misc. external resistors and bypass capacitors. The basic circuit for the function XY/10 is given in the Motorola applications notes and is shown below with slight revisions:



The function XY/10 is proper for a system with voltage levels that max. at 10 volts. For systems using a 5 volt level, the correct function should be XY/5. Basically, this is a matter of changing the gain of the output amplifier to give twice the gain. A circuit that works well in EN projects is shown below:



The following points about the type 595 should be noted:

(1) The device has 4 inputs and 2 outputs. The X inputs are + and -, the Y inputs are + and -, and a differential output is available. Commonly, the actual signal inputs are +X (pin 4) and +Y (pin 9) while the -X (pin 12) and -Y (pin 8) inputs are used for balancing. It is possible to interchange these functions to invert operation if required.

(2) The two outputs can be interchanged to invert the output (e.g. from XY/5 to -XY/5).

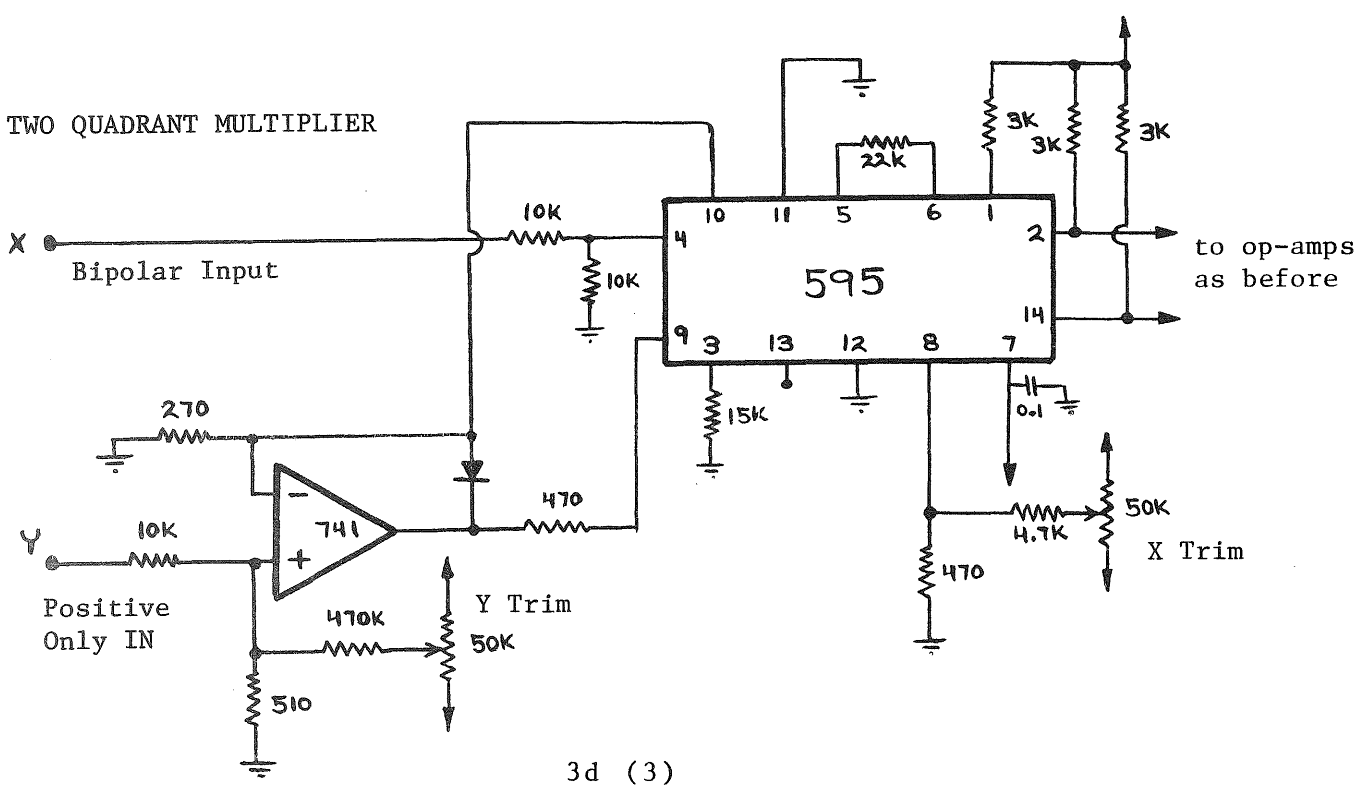
(3) Balancing can be done by either technique shown in the first two examples (by deriving what is equivalent to +5 and -5 supplies)or by other suitable techniques. The actual voltage swing needed for trimming is often only of the order of +2 to -2 volts. A third technique is shown in the example for a two quadrant multiplier below (see the connections to pin 8). The 470 ohm resistors are necessary even if balancing is not used and the other side of the resistor is grounded. This prevents RF instabilities that could occur if too low an impedence appeared at these pins. RF instability is also avoided by adding bypass capacitors to the -15 supply on pin 7. Bypassing of the +15 supply is also a good idea.

(4) As a quick test (useful for board testing), just ground the balancing pins. Apply +15 to both inputs (output goes to +15). Apply -15 to both inputs (output goes to +15). Apply +15 to one input, -15 to the other (output goes to -15, and with inputs reversed). Once balancing controls are connected, it is useful to balance the AC conditions first by applying a signal of about 5 volts at 1 KHz to one input while the other is grounded (or just open if the attenuator is on the input). Adjust the balance pot for min signal output. Repeat for the other input. Once done, the DC level control should be adjusted for zero volts out when the inputs are grounded. It may be necessary to change the value of the series resistor R_s in some cases to get this null. However, for best results keep the percentage of the range swept by the pot small so that an accurate DC zero can be obtained. A good way to check the zero is to set the scope for the most sensitive DC range and touch a ground wire to the output. If the trace deflects, the zero is not yet properly set.

TWO-QUADRANT MULTIPLIERS

Two-Quadrant multipliers have certain advantages in many applications. The two quadrant multiplier will be discussed in detail in Chapter 5c on VCA design. What we will look at here is the use of the 4-quadrant multiplier IC as a 2-quadrant multiplier as suggested by Jung (The IC Op-Amp Cookbook, page 262). The basic procedure is to disable one stage while adding a control to the current input to the remaining operative stage. The alterations are to pins 9, 10, 11, 12, and 13 as shown below:

TWO QUADRANT MULTIPLIER



CHAPTER 3E

BASIC APPLICATIONS OF:

IC TIMERS

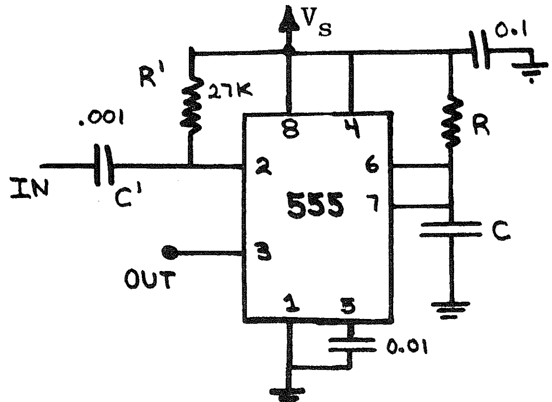
CONTENTS:

INTRODUCTION

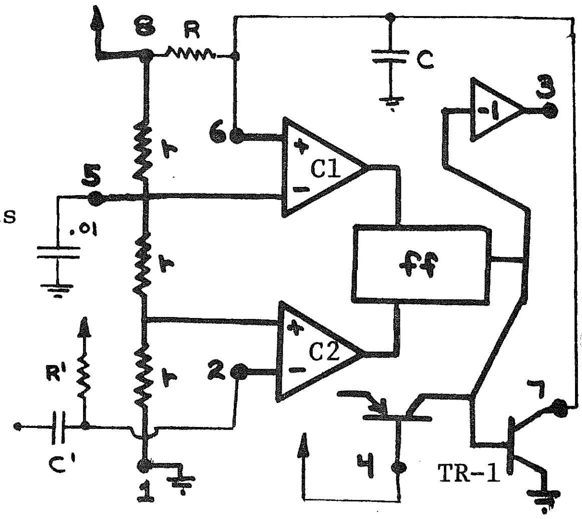
IC Timers are basically IC packages with several comparators, discharge transistors, a flip-flop, and possibly additional circuits. Various interconnections and external components can be used to provide monostable, astable, and special devices. The type 555 is a general purpose timer that is used in electronic music principally as a monostable (for timing delays, envelope delays, and sequencers). Most of the time, periodic waveforms for electronic music are obtained from other types of oscillator circuits, but there are cases where the IC dimer does an exact job that would be difficult with other devices. The 555 can be used as a monostable for pulse times from 1 microsecond up to a minute or longer. When shorter pulses are needed, the TTL type 74121 can be used to give pulses as short as 40 ns. To properly use the type 555, it is possible to either just capy an existing circuit, or to carefully consider the interconnection of the working elements. To use the 555 in unique applications, it is necessary to work with the inside functional blocks in the design.

MONOSTABLE (ONE SHOT) CIRCUITS

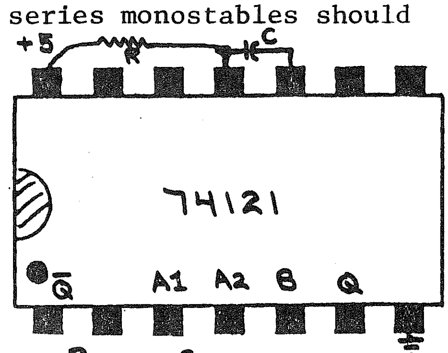
The 555 is very reliable as a monostable. The principle problems that the designer encounters in its application are: (1) Failure to provide the proper type and magnitude trigger signal, and (2) Failure to properly bypass the power supply line leaving the device open to noise triggering problems. The basic monostable circuit is shown at the right. The monostable is triggered by a short trigger pulse applied to pin 2. This pulse must be a voltage that starts near the supply level V_s, and falls below about 1/3 V_s for a period of time that is brief compared to the on time of the monostable. The differentiator network R'C' is used to provide the necessary trigger. The input side of C' should be a square wave or rectangular edged function that has a negative transition greater than 2/3 V_s. When the trigger arrives, pin 3 goes high and remains high until a time 1.1RC expires. It then falls until another trigger arrives. Any trigger that arrives while the output is high will be ignored. Note that a 0.1 mfd bypass capacitor has been added to the V_s supply line. This should not be left out. V_s can range from 5 to 18 volts. The capacitor on pin 5 is used to prevent noise problems. Pin 4 can be thought of as an "enable" pin, as this must be high to allow the output to go high. The monostable circuit can be made to have a voltage controlled on time by driving pins 6-7 with a voltage controlled current source instead of with resistor R.



To understand how the monostable works and to form a basis for considering other 555 circuits, it really is necessary to look at the internal elements as shown at the right. The internal elements are shown with heavy lines while the monostable interconnection is shown with a thinner line. Note that internal references are provided by the three series resistors labled "r". In the absense of warping by an external voltage on pin 5, there voltages are 1/3 V_s and 2/3 V_s. The stable state is when pin 3 is low, TR-1 is on, and pin 2 is high. A negative pulse triggers comparator C2, resetting the flip-flop. Pin 3 goes high, and TR-1 is shut off, allowing capacitor C to charge through R. When the voltage on C reaches 2/3 V_s, comparator C1 sets the flip-flop, pin 3 falls, and TR-1 discharges C. If it is desired to have the monostable retrigger, this can be accomplished by triggering pins 2 and 4 in parallel. This causes a glitch in the output which may or may not be a problem depending on the application.



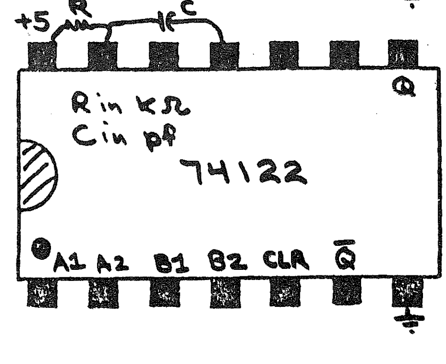
For short output times shorter than 1 microsecond, the TTL series monostables should be used. The setup of the type 74121 is shown at the right. The time is set by resistor R (which must be between 2k and 40k) and the capacitor C according to t = 0.69RC. Times up to 40 seconds are possible. To trigger on a positive going edge, ground A1 and A2 and apply the trigger to B. To trigger on a negative going edge, set A1 and B high and apply trigger to A2 (or reverse roles of A1 and A2).



For a short retriggerable monostable, the type 74122 can be used. This will give pulses of from 40 ns up. The time is determined from R and C according to:

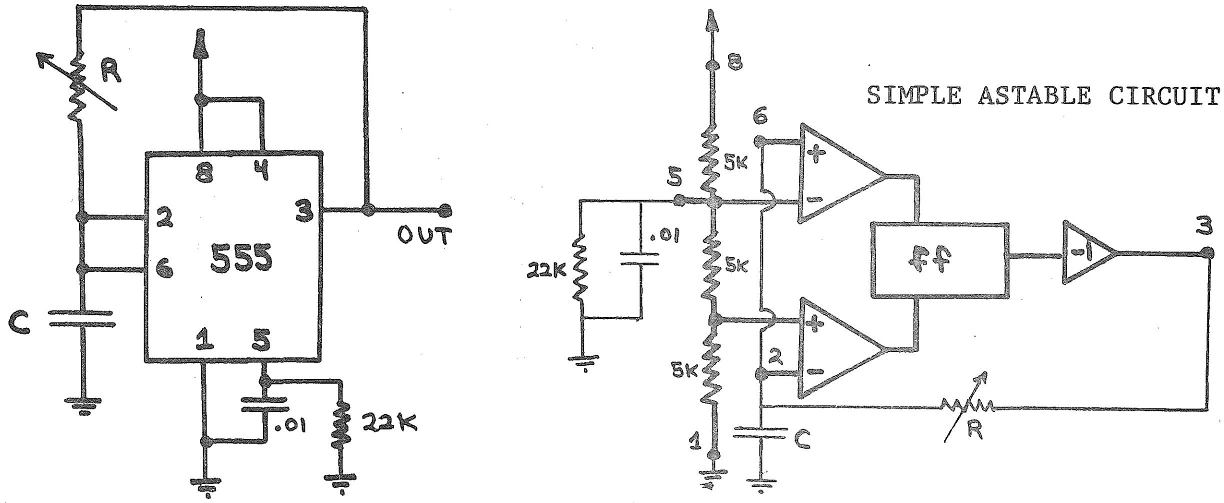
$$t = 0.32\ RC\ [1 + (0.7/R)]$$

For positive edge triggering, A1, A2, B2, and CLR should be high - apply trigger to B1. For negative edge triggering, A1, B1, B2, and CLR should be high - apply trigger to A2



ASTABLE (OSCILLATOR) CIRCUITS

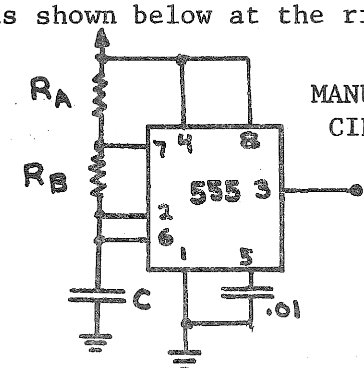
A simple astable circuit can be formed by resorting to the popular integrator - Schmitt type of oscillator. The Schmitt trigger can be comprised from the two comparators and the flip-flop. The integrator is formed with an external capacitor and a resistor feeding back the output. The basic circuit is shown below, along with its representation in terms of the functional units inside the 555.



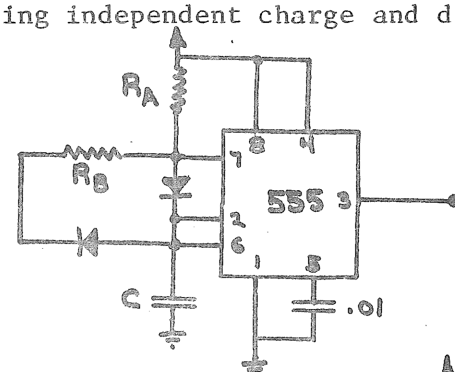
SIMPLE ASTABLE CIRCUIT

The capacitor voltage changes (exponentially, approximately triangularly) between 1/3 V_S and 2/3 V_S. The frequency is given by $f \simeq 0.78/RC$. Operation is quite simple. With the output high, the capacitor voltage rises until C1 triggers, causing the output to fall. The capacitor then discharges down until C2 triggers, causing the output to rise again. The 22k resistor connected to pin 5 tends to shift the reference levels slightly to give a more symmetric square wave.

The astable from the manufacturer's applications notes is shown below. This circuit has a duty cycle limited to less than 50%. The circuit is controlled by pin 7. When pin 7 is cut off, the capacitor charges through R_A and R_B. When pin 7 goes to ground (triggered by C1 reaching 2/3 V_S), discharge is through R_B only. The time for the capacitor voltage to change between the 1/3 and the 2/3 levels for a standard RC exponential decay is 0.685RC. This can be used to determine the frequency of oscillation as $1.46/[(R_A + 2R_B)C]$. M. Robbins has shown [Electronics, June 21, 1973] that the full duty cycle range can be achieved by setting independent charge and discharge paths as shown below at the right.

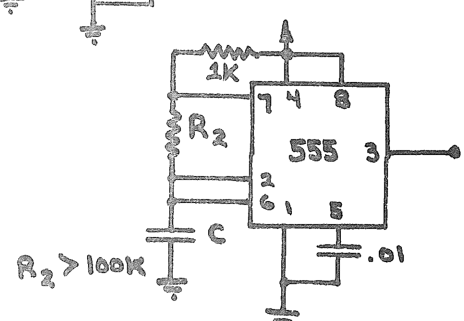


MANUFACTURER'S CIRCUIT

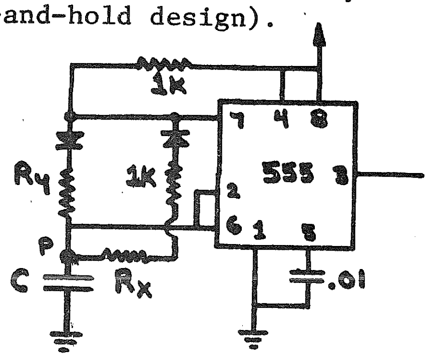


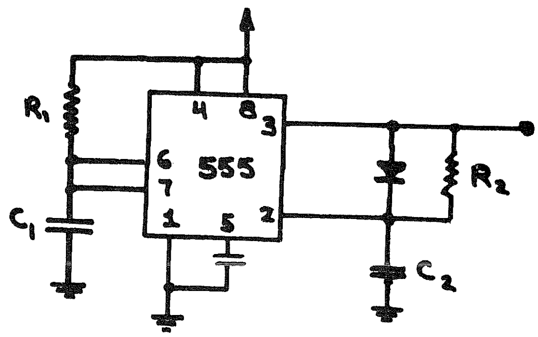
Jung has shown [PE Jan 1974] a simple square wave circuit that takes advantage of the current sinking of pin 7. The actual charge and discharge paths differ by less than 1% and the waveform is nerely square. The frequency is given by

$$f \simeq 1.43/R_2C$$



Two additional astable improvements have been described. A. Klinger [Electronics, Sept. 19, 1974] has shown a circuit that provides independent charge and discharge paths. The circuit is shown on the left below. R_x and R_y should both be much greater than 1k. If R_x and R_y are from a dual pot, the frequency of a square wave is controlled (same as Jung's circuit). If R_x and R_y are the segments of a single pot, the point p being the pot wiper, then the duty cycle is controlled independent of frequency. J. Carter, [EDN June 20, 1973] has described the circuit on the right below. This is an astable formed by self triggering a monostable. The charge and discharge times are independent. This circuit is very useful where you need both monostable and astable modes (see chapter 5g on sample-and-hold design).

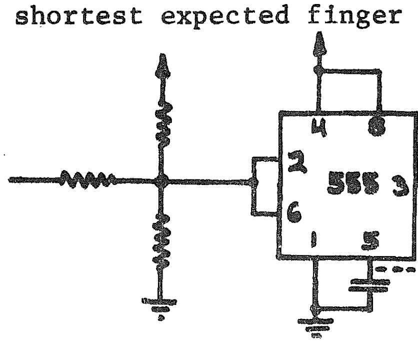




SPECIAL APPLICATIONS

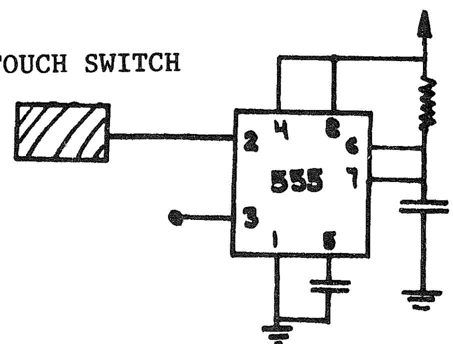
The 555 can be used as a Schmitt trigger by connecting the two comparators together and applying a weighted sum of the input, V_s, and ground as shown in the circuit below. If necessary, the voltage on pin 5 can be altered (e.g., by pulling it down to reduce the difference between the references) to suit the input.

The 555 can be the basis of a touch triggered switch as shown in the circuit below. The contact duration however should not exceed the time the output is high. Thus, certain applications will not work well if the output is high only for a time short compared with the shortest expected finger contact time.

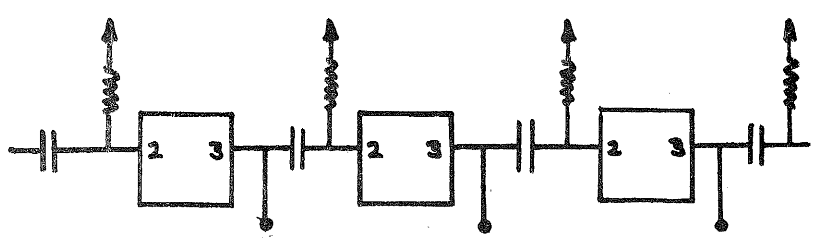


SCHMITT TRIGGER

TOUCH SWITCH



The 555 is ideal in certain sequencing devices as one 555 monostable can be used to trigger another down the line and so on. This could be for very rapid sequencing like one would use in digital circuits, or for slow sequencing on the order of seconds for the actual control of musical structure. The important thing here is to make sure all the power supply leads are well bypassed or else the switching may trigger more than one monostable down the line.



CHAPTER 3F

BASIC APPLICATIONS OF:

DIGITAL INTEGRATED CIRCUITS

CONTENTS:

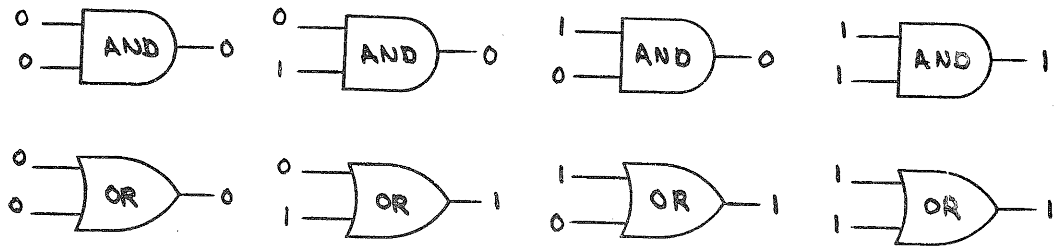
INTRODUCTION

Although the use of digital circuits in electronic music devices is increasing, and may eventually take over many or most of the functions now assigned to linear circuits, it will not be possible here to examine all the possibilities. Instead, we shall orient the present discussion toward applications that are really non-digital but use digital IC's. These are basically "one bit" digital circuits that interface functionally with analog circuits. We shall actually be doing three things here: (1) Examine the basic applications of the simplest IC digital circuits, (2) Indicate how some of the more complex IC's have been used in specific applications, and (3) Describe basic applications considerations for TTL, leaving CMOS for the next chapter.

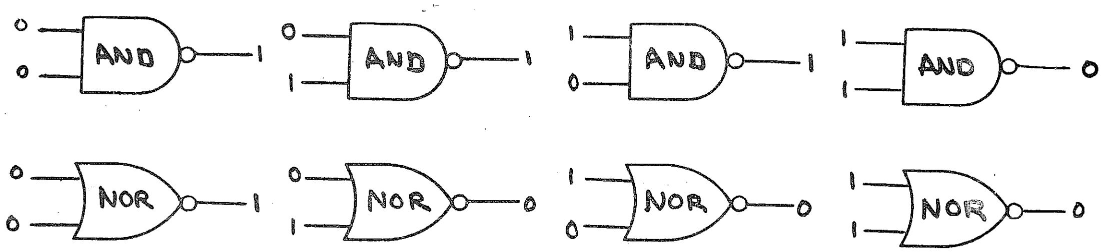
The reader interested in the development of digital structures for electronic music is advised to study the available literature of logic design, computer structures, and digital filters. A thorough study of the data sheets and applications literature on the larger MSI and LSI circuits should also facilitate applications and suggest many new uses.

APPLICATION OF LOGIC GATES

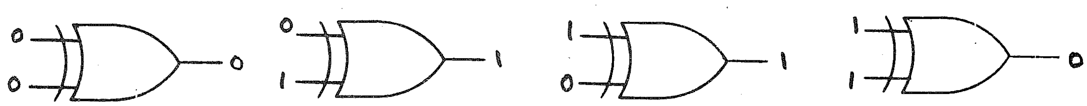
Logic gates can be used to make AND decisions or OR desisions - that's the original and obvious purpose. What we want to discuss here are the cases where special tricks can be used. What we say here applys to TTL, and to the 74C00 series of CMOS, and in general, any IC logic family that has the necessary logic gates. The basic gates are usually inverting, so they are called NAND and NOR. The AND and OR decisions are:



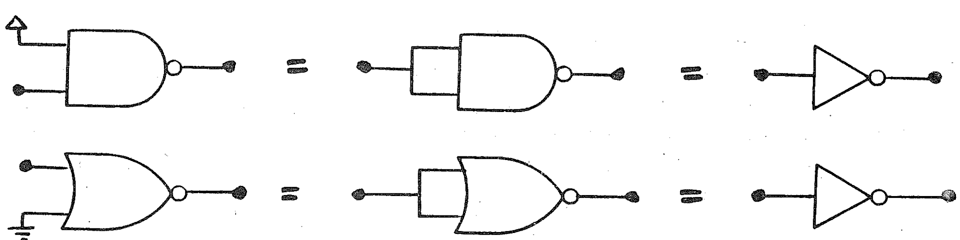
The NAND and NOR gates simply have an inverted output:



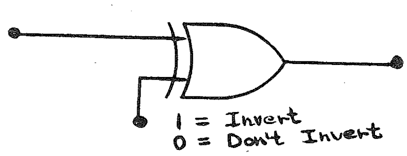
Another useful gate is the EXCLUSIVE OR gate which gives a decision for one and only one. However, by the same token, it can be thought of as giving one state if the two inputs agree, and the other state if the two disagree.



These gates commonly come four to the IC package, so often the trick is to keep the package count down by using what is available. In particular, the gates are often used as inverters. Often the inverters necessary are available as left over gates:

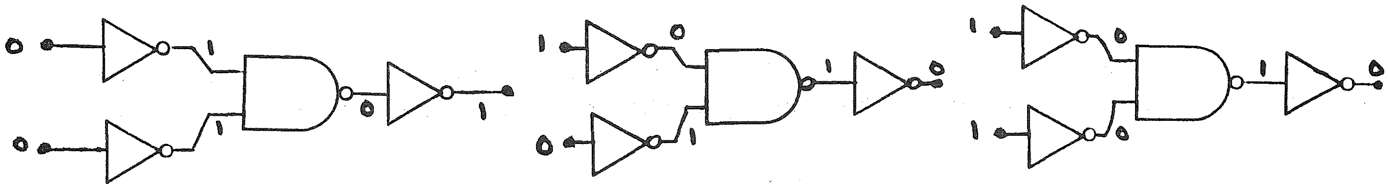


The EXCLUSIVE OR gate can also be used as an inverter. Also, depending on the state of one input, it can be programmed to invert or not invert a digital signal.

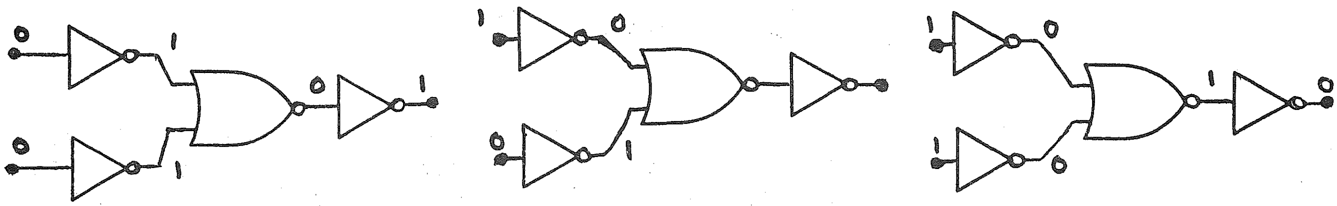


When one need an actual AND or OR gate, an inverter can always be added after the NAND or NOR gate, but often times, there are other ways. There is a basis principle of logic that says that AND and OR gates change function when one changes from negative to positive logic. In practical terms this comes down to something like: Here is an IC and this is the way it responds to input conditions - does it do the job I want done?

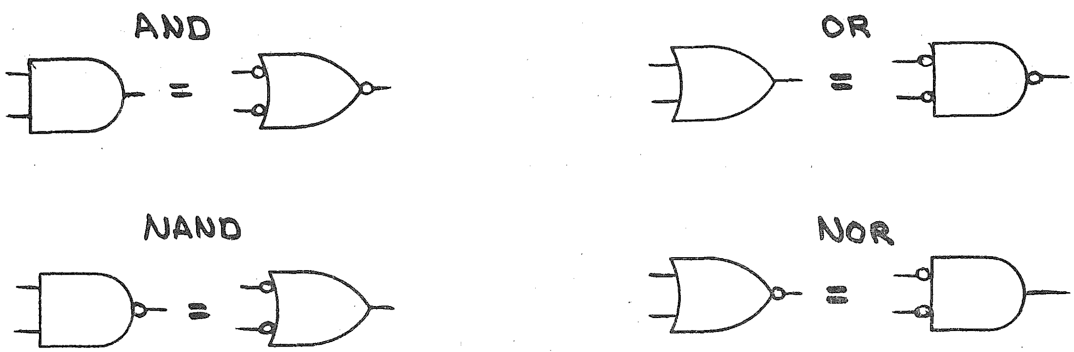
For example, suppose you need a NOR gate and have only a quad NAND. Use three of the NAND's as inverters, and connect up the following:



Checking out the three input conditions, you can see that this behaves as a NOR. without the final inverter, it is an OR gate. In the same manner, we can form a single NAND gate from a quad NOR. Leaving off the final inverter gives a AND.



The little circles on the outputs are used to signify that the output is inverted. The same circles (called bubbles) can be put on the inputs to signify that we have placed inverters there. With this notation, we have the following equivalent notations for AND, OR, NAND, and NOR.

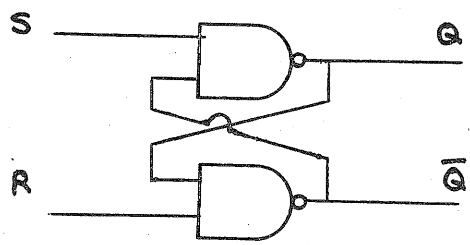


Up to this point, we have only found something that is useful for emergencies. Generally, there would be no point in using a full package when all that is needed is one gate. Where this sort of thing starts to get practical is in cases where what is needed is an AND or an OR (saving the output inverter) or where it is not necessary to put inverters on the inputs because the inverted versions of the states are already present in the circuit. If you have a flip-flop for example, you have both Q and $\overline{Q}$ available already.

If you need a single NAND and a single OR, then a quad NAND package will do. If you need a single NOR and an AND, then a single NOR package will do.

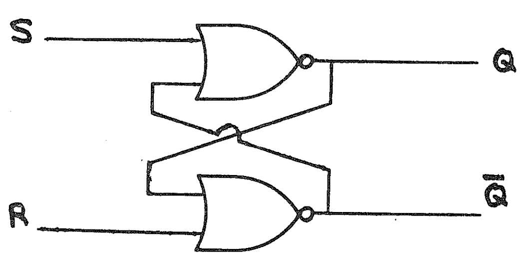
Next, as an example suppose you need an AND, and all you have is a left over NOR. That is you want to AND A and B together, but have only a NOR. You could bring in a quad NAND and use one section as an inverter on another. This would bring in another IC package and leave the extra NOR and two NAND's. The thing to look for here are $\overline{A}$ and $\overline{B}$ in the circuit somewhere. If you find them, you can feed them to the NOR gate and have an AND decision made. The saving could be very important for large production runs, but all this is secondary to the satisfaction you will probably feel for having beaten the system. The important thing to bear in mind is that the NAND and NOR gates can be made to do each other's jobs. This street runs both ways however. You should be aware that just because you see a NAND gate or a NOR gate, this does not mean that these are necessarily the actual logic decisions being made. Here, the notation with the bubbles on the inputs can be very helpful as the right sort of gate for the logic function being performed can be shown.

Two logic gates can be cross-coupled to form an elementary flip-flop called an R-S flip-flop or a latch. The configuration with two NAND gates is:



The two outputs Q and $\overline{Q}$ are complimentary (when one is in the 1 state, the other is in the 0 state). Both the inputs (R and S) are normally held at logic 1. Then, a logic 0 on S will set the flip-flop (Q=1, $\overline{Q}$=0) and it will remain in this state even if the logic 0 on S returns to a logic 1. The only way to change the state is to apply a logic 0 to R (reset) and the state of the outputs will become Q=0, $\overline{Q}$=1. It should be kept in mind that the two gates are still gates, and do not have any way of knowing they are connected as a flip-flop. Analysis of this type of circuit requires a few drawings and a belief that the gates still obey their own rules.

The corresponding R-S flip-flop with a pair of NOR gates is shown below. The circuit is essentially the same, except the inputs are held at logic 0 and the setting or resetting is accomplished by applying a logic 1 to S or R.



The R-S flip-flop is the basis of many envelope generator circuits. The flip-flop is set by a trigger signal, and an envelope is initiated. When the envelope reaches a certain stage, the flip-flop is automatically reset and the envelope finishes.

APPLICATIONS OF FLIP-FLOPS

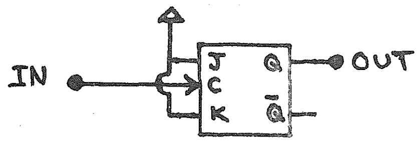
The R-S flip-flop formed from two logic gates as shown above is a simple latch type of circuit that "remembers" the last state change that it was asked for. With additional inputs, flip-flops can be designed to perform other memory and decision functions.

The J-K flip-flop (7473, 7476, 74107) has the usual Q and $\overline{Q}$ outputs. In addition it has two logical control inputs J and K, and a clock input. When a clock signal arrives (usually a negative going transition), the flip-flop either changes state or does not depending on the input conditions J and K:

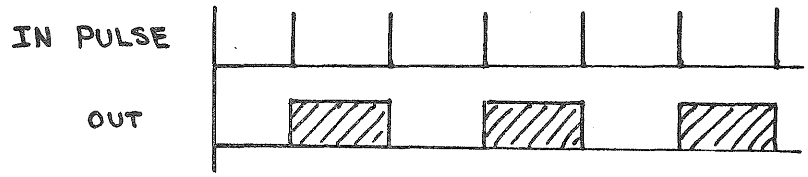
With J and K grounded, no change takes place
With J and K high, the state always changes ($Q \rightarrow \overline{Q}$, $\overline{Q} \rightarrow Q$).
If only J is high, Q goes high (or stays high if it is already high).
If only K is high, $\overline{Q}$ goes high (or stays high if it is already high).

Note that Q and $\overline{Q}$ always complement each other - they can not be in the same state ever.

With J and K high, the flip-flop divides an input square wave by two. Musically this is an octave division. An input pulse will be "squared up" by the same arrangement. Division by powers of two is a matter of cascading more flip-flops.



DIVISION BY TWO



DIVISION BY TWO SQUARING UP PULSE

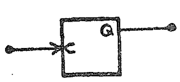
As indicated above, the basic divider circuit is the binary divider (BD) formed from a JK flip-flop. The BD divides by 2 and gives a symmetric square wave output for any input where triggers are periodic. Note that ÷4, ÷8, ÷16, etc is a simple matter of cascading ÷2 circuits.

Division by 3 (and higher odd numbers) can be done for a symmetric (S) or a non-symmetric (NS) output. Symmetric outputs for odd numbers 3 and higher require a symmetric square wave input. Division for these odd numbers is actually easiest with a walking ring counter (WRC) which gives a non-symmetric output. A method of symmetric division for any number has been described by D. A. Scott ["Divide-by-N Circuit Has 50/50 Duty Cycle," EDN July 5, 1973]. This method requires exclusive-OR gates in addition to flip-flops. For even numbers other than powers of two (6,10,12,14,18...etc.), it is usually best to factor out powers of two, implement the remaining odd number division by the easiest method, and then binary divide the output.

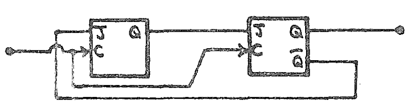
Below are shown a series of divider circuits for ÷2 to ÷16. These have been chosen mainly for symmetric outputs by the simplest method of those discussed above. A few non-symmetric dividers are included where these may be used to reduce IC package count.

The Following Codes Apply to the Divider Circuits: [Note: Unused J, K should be high]

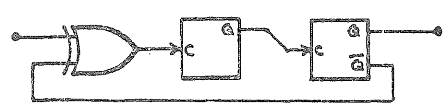
- BD -Binary Divider(s). The output is symmetric for periodic triggers to the last stage.
- NS÷3 -Non-Symmetric divide-by-3 (requires only two flip-flops). Used as a basis for several symmetric circuits (÷6 and ÷12).
- WROC -Walking Ring Odd Counter. Output is non-symmetric. Shown for ÷5 and ÷7; extension to ÷9 and ÷11, etc., is straightforward from these examples.
- EXOR -Exclusive-OR method from EDN reference above. Output is symmetric.



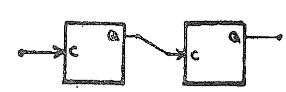
÷2 BD



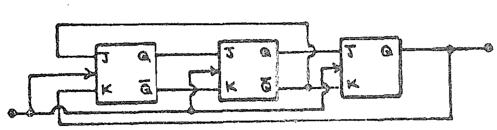
÷3 NS÷3



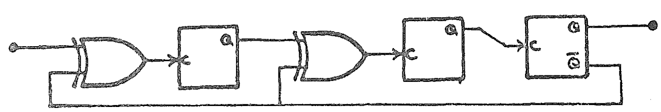
÷3 EXOR



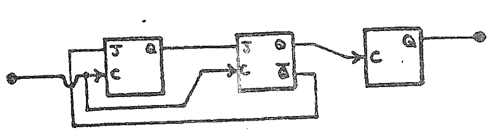
÷4 BD



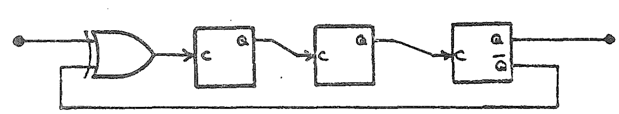
÷5 WROC



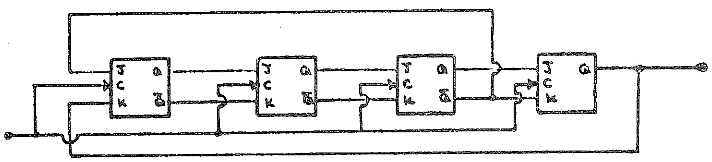
÷5 EXOR



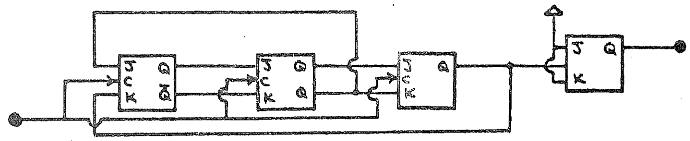
÷6 NS÷3 plus BD



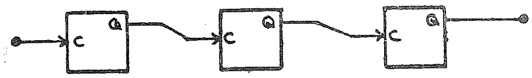
÷7 EXOR



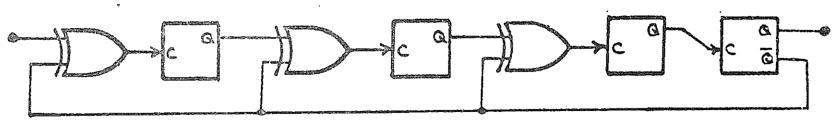
÷7 WROC



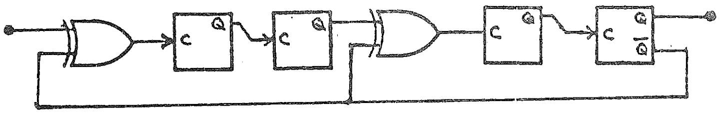
÷10 WROC plus BD



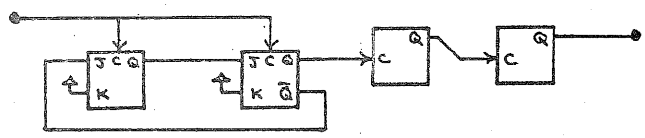
÷8 BD



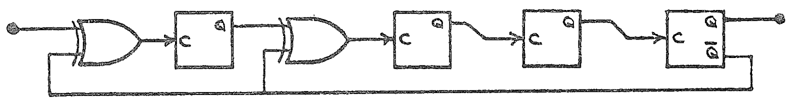
÷9 EXOR



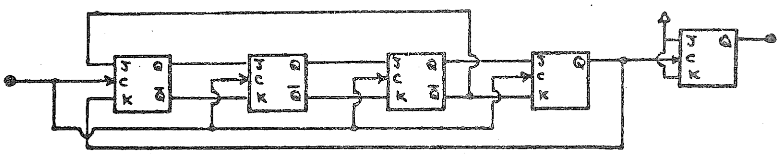
÷11 EXOR



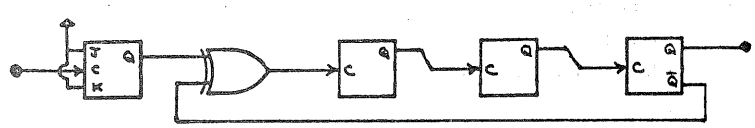
÷12 NS÷3 plus BD



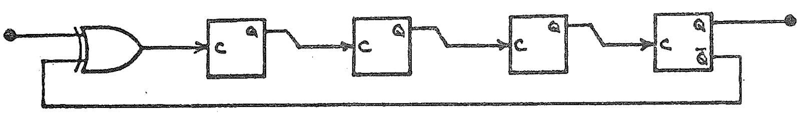
÷13 EXOR



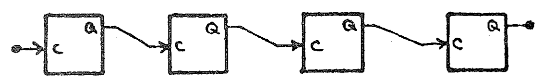
÷14 WROC plus BD



÷14 BD plus EXOR



÷15 EXOR



÷16 BD

<u>OTHER TYPES OF FLIP-FLOPS</u>

A D-Type flip-flop (TTL 7474, 74174, 74175) has the usual Q and $\overline{Q}$ outputs. The controls are a D input (D=Data) and a clock. When the positive edge of the clock arrives, the input on D is transferred to Q. Changes in D at times other than when the clock is going high are ignored.

A somewhat similar circuit is the latch (TTL 7475). The latch has Q and $\overline{Q}$ (usually) outputs, a D=Data input, and an enable control. With the enable high, data is transferred continuously to the Q output. When the enable goes low, the Q output retains the last D input it saw. It holds this until the enable again goes high.

SPECIAL APPLICATIONS OF MSI DEVICES

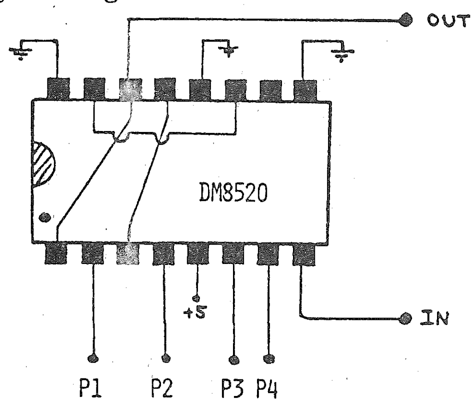
In addition to the standard digital applications for digital IC's, there are often special applications for the MSI devices (medium scale integration). Below we will be listing a few TTL devices and describe briefly useful applications:

<u>DIVIDERS</u>: In addition to the divider circuits using flip-flops, dividers can be formed from MSI devices.

The 7490 is a decade divider (÷10) which can be used as a ÷5 and a ÷2 by proper connection. The 7492 is a ÷12, which can be broken down into a ÷6 and a ÷2. The 7493 is a ÷16 counter which can be broken into a ÷8 and a ÷2. With these, certain dividers can be easily implemented with a single chip. For more divider circuits see the chapter on counters in R. Morris & J. Miller (Eds) Designing with TTL Integrated Circuits, Texas Instruments Inc., (1971), and the TTL Cookbook by Don Lancaster (Sams publishers).

UP-DOWN counters may also be used. These counters count either up or down depending on a control input to the chip. These include the 74190, 74192, and the 74193. These can be used for circuits like envelope generators that step up and then later step back down.

Programmable divide-by-n circuits can be used as programmable dividers for frequency division, or for keyboard circuit digital counting (usually using two such chips to expand the division to 8 bits). The TTL type 74193 can be used as a programmable divider. The type DM8520 from National is also popular. The setup of a single chip and the programming is indicated below. The output has one high pulse for n-1 low ones.



The Type DM8520
Programmable
Divide-by-N

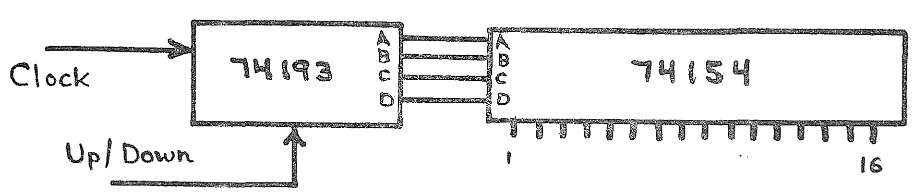
Programming of DM8520

÷by	P1	P2	P3	P4
2	1	1	1	0
3	1	1	0	0
4	1	0	0	0
5	0	0	0	1
6	0	0	1	0
7	0	1	0	0
8	1	0	0	1
9	0	0	1	1
10	0	1	1	0
11	1	1	0	1
12	1	0	1	0
13	0	1	0	1
14	1	0	1	1
15	0	1	1	1

The series of data distributers and multiplexers can be very useful. These include the following:

74150	16 line to 1 line selector
74151	8 line to 1 line selector
74152	8 line to 1 line selector
74153	Dual 4 line to 1 line selector (common address both sections)
74154	4 line to 16 line decoder
74155	Dual 2 line to 4 line decoder
74157	Quad 2 input data selector

A popular application is indicated below. The 74193 up/down counter is used to drive the 74154. The count address (A,B,C,D) from the 74193 is decoded by the 74154 into 16 different output lows. This gives the rough equivalent of a 16 stage bucket brigade counter. This has the added advantage that it can be made to go either direction.



The circuit can be used as an envelope generator for example, or as the basis of a sequencer. Additional applications of these selectors include keyboard scanners for digital keyboards.

Priority encoders such as the 74147 and 74148 have possible applications in keyboard circuits (EN#52). These can select a highest key when more than one is down.

Some packaged counter circuits like the 74164 8-bit parallel shift register can be useful for circuits that require access to the insides of the counter such a programmable pseudo-noise generators.

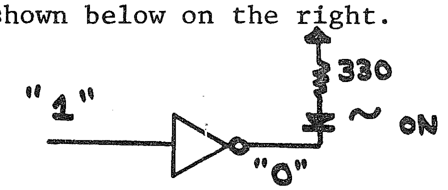
The types 7497 and 74167 are rate multipliers. A wide variety of useful applications of these circuits have been discussed by Stephen Wilson [EN#42 (5)].

DESIGNING WITH TTL LOGIC IC'S

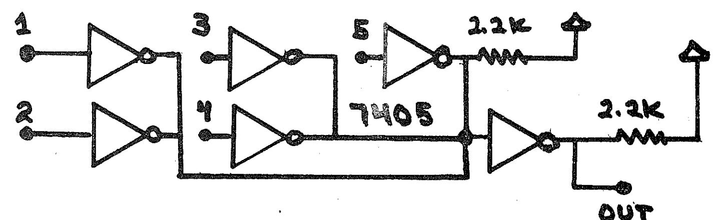
TTL is a current sinking type of ligic - the inputs must be pulled down by drawing approximately 2 ma from them. Thus, an unconnected TTL input is high, or will at least tend to stay high. However, if an input is unused and should be high, it should be tied to +5. If it is unused and should be low, it must be grounded and not left open.

TTL circuits generally have an "active pullup" known as the "totem pole output." This means that even though logic is handled through pull down, there is no need for a load resistor to allow the outputs to go high in the absense of pull down, unless the circuit is of the "open collector" type. There is generally no problem having a high output drive an attached input high as well. There is a limit on the total sink current that can be drawn however, and this determines the number of inputs that can be driven by any one output. This figure is called the "fan out" and is 10 for most TTL, but may be 30 for some "buffer" type of outputs.

Open collectors are used for driving something like a LED by sinking one of its terminals as shown below. It also makes possible certain "wired OR" circuits of the type shown below on the right.



OPEN COLLECTOR DRIVING LED



A WIRED 5 INPUT OR GATE

In general, two outputs should not be connected together, but this is possible with open collector structures like the wired OR.

It is a good idea (essential in many cases) to provide power supply bypassing capacitors for the TTL circuits. These should be distributed throughout the circuit. Usually values of 0.01 mfd every third gate package or one for every MSI package are recommended. These capacitors prevent heavy switching currents in one chip from being coupled through the power supply lines to another chip. These spikes could cause noise triggering if they are not "soaked up" by the bypass capacitors.

Chapter 3G

BASIC APPLICATIONS OF:

CMOS INTEGRATED CIRCUITS

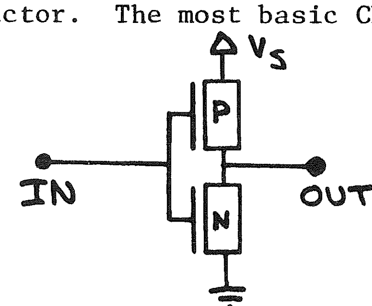
CONTENTS:

INTRODUCTION

CMOS IC's can be used in both digital and linear applications. The IC's have a number of properties that make them nearly ideal (wide power supply range, high input impedence, low power consumption, high noise immunity, and some unique devices such as the analog switch). The devices are available in the original 4000 series and a 74C00 series where devices are functionally equivalent to the popular 7400 series TTL. The IC's require a little more care than other IC's, but not as much as is often thought.

BASIC CMOS APPLICATIONS CONSIDERATIONS

CMOS stands for Complementary Metal Oxide Semiconductor. The most basic CMOS structure is the CMOS inverting amplifier which is diagrammed at the right. It can be considered to be two complementary FET's in series with a common gate. When the gate is low, the lower N-FET is off while the P-FET is on. The output is effectively a resistor of about 400 ohms to positive supply -- hence -hence it is high. With the input high, the P-FET is



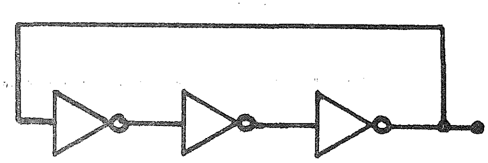
is off and the N-FET is on representing a resistance of about 400 ohms to ground - hence the output is low. The device is thus a logic inverter. The gate structure is an insulated gate and thus has a very high input impedence (10^{12} ohms) and is best modeled as a capacitance of about 5 pf. Input currents are only about 10 pa. When input voltage is close to half the supply, both the P-FET and the N-FET have about equal resistance and thus form a voltage divider that sets the output at half the supply voltage. This is the basis for the self biasing method for CMOS in the linear mode. When the input goes to either side of the supply half point, the output will favor one state. The region of uncertainty around the midpoint is quite small and thus CMOS is said to have high noise immunity. It should not be assumed however that the CMOS inverter behaves like a comparater with a reference at half the supply voltage. In many cases, the results are consistent with this assumption, but the actual case is that for a supply voltage V_s, noise of about 0.45 V_s will not propogate. This is just saying that an input that is close to the uncertain region will in general cause an output that is more removed from the uncertain region. A few stages into the system, the level switching between V_s and ground will have restored itself.

In logic circuits, one of the two FET's is off except when the device is actually changing state. Thus essentially no standby power is drawn (current drain is about 5 na). At audio frequencies and below, the on time is a very small percentage of the total time, and the average current is very small. The CMOS IC will run on any voltage between 3 and 18 volts.

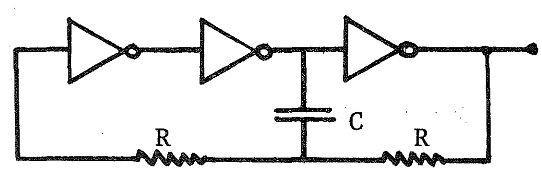
When used as logic devices, application is quite similar to other logical IC types. The main difference is that a much larger fan-out can be used due to the low input impedence. The main limitations on fan-out are due to the time required to charge the gate capacitances of the combined structure.

LINEAR CMOS APPLICATIONS

CMOS IC's can be used in a number of linear circuits and oscillators. Two oscillator circuits are shown below:



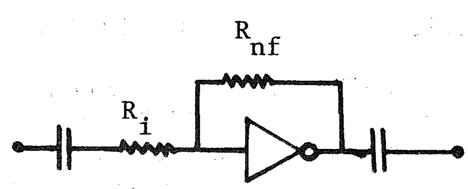
RING OSCILLATOR $f \sim 1/(n \cdot 2\times10^{-8})$



$f \simeq 0.559/RC$

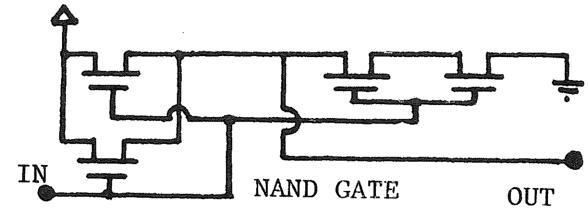
The first circuit is the ring type of oscillator that requires an odd number of gates (n). The design formula for the frequency is not especially accurate, but it does show that for practical numbers of gates, the frequency is in the MHz range. The second oscillator is more accurate and the frequency can be much lower.

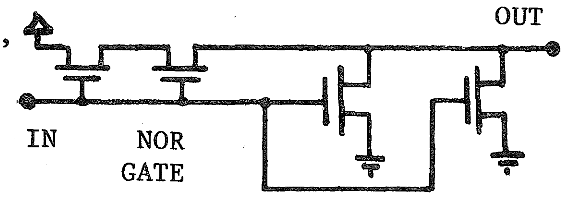
As suggested above, it is possible to use negative feedback to self-bias the CMOS inverter at (1/2) V_s. This is done by using a resistor (usually in the megohm range) R_{nf} as shown at the right. It is then possible to use an input resistor R_i and obtain an inverting gain of R_{nf}/R_i. The performance can be improved by using cascaded inverters instead of just the one. The total number of inverters must be odd. The input impedence is higher than the value of R_{nf}. Note that AC coupling is required in most cases.



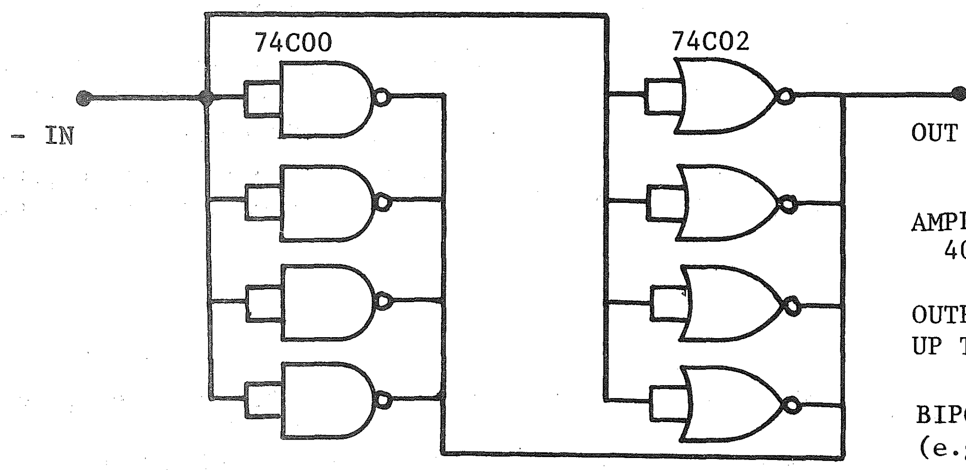
Another useful application of the CMOS IC's in linear circuits is to increase output drive current or input impedence. The FET inputs can be used to change an ordinary op-amp into a FET input op-amp [W. Jung, PE Aug. 1974]. However, the total power supply voltage to the op-amp must be reduced since the CMOS can only handle 18 volts total.

To see how the CMOS circuit can be used to boost output current, it is first useful to look at the way NAND gates and NOR gates are implemented using the CMOS structure. At the right it can be seen that when the inputs of a NAND gate are coupled, there is twice as much drive current from the + side of the supply as is possible from the inverter. This is because there are two channels for the current to flow through. If on the other hand the two inputs of a NOR gate are connected, there are two channels to ground available. Thus, compared to the simple inverter, the inverter formed from a NAND gate is a better source, and the inverter formed from the NOR gate is a better sink. This will become important when we consider the interfacing of CMOS with other logic systems.





In making a post amplifier from CMOS to increase current drive, we have seen that there is an advantage to using NAND gates for a source and NOR gates for a sink. This can be extended for bipolar operation in which case the NOR gate supplies current from a negative supply. By connecting a parallel combination of 4 NAND gates and 4 NOR gates, a post amplifier is formed as shown below that can supply currents up to 40 ma. The overall amplifier is inverting so this must be taken into account if the amplifier is used within a feedback loop (as will usually be the case).



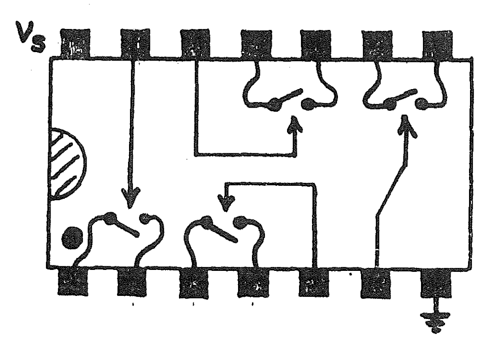
AMPLIFIER GIVING UP TO
40 ma DRIVE CURRENT

OUTPUT STAGE GIVING
UP TO 40 ma DRIVE CURRENT

BIPOLAR SUPPLY
(e.g. ±6)

THE ANALOG SWITCH

A unique device that is available in the CMOS line is the type 4016 quad analog switch. These switches will handle any voltage between the supply limits. When the switch control voltage is high (+ supply) the switch is on and the device becomes what is equivalent to a 400 ohm resistor. When the switch control is low (- supply or ground, whichever the case), the resistance is very high. The quad switch has a number of very useful applications since the designer can use what is actually a voltage controlled switch. The switch finds applications in sample-and-hold devices, envelope generators, and other control voltage circuits. The



analog switch can also be used as a control element for audio signals. The switch is controlled by a rectangular waveform of ultrasonic frequency and variable duty cycle. This changes the effective value of an attached component in proportion to the duty cycle.

CMOS HANDLING PRECAUTIONS

The following is a listing of handling precautions for CMOS IC's.

[1] Store CMOS in protective conducting foam or in styrofoam which has been first covered with aluminum foil. If Al foil is used over styrofoam, avoid pushing the leads into the same holes twice. Never use styrofoam without foil - store them in the open if no foil is available.

[2] Ground the soldering iron tip with a thin wire wrapped around it, or solder in CMOS by allowing solder to flow from a grounded circuit board trace (or any trace that has a DC path to ground, even in the megohm range) to the IC pin.

[3] Whenever possible, install the CMOS Ic's last. Make changes in the circuit with the power off. For major changes, remove the CMOS if this is convenient. When installing a CMOS IC, hold it on one hand and touch a ground (circuit ground) with the other hand before touching the IC leads to the mounting position.

[4] The inputs are protected by diodes that run to the supply levels. The input voltages should not exceed the supply limits or these diodes may conduct current and heat up the entire chip. When the power is off, both supply limits are zero. This means that a voltage source should not be connected to any input if the CMOS is not under power - unless there is a series resistor to limit current. The current in this case should never reach 50 ma, and should be limited to 10 ma or less. Thus a signal of 10 volts through a standard 1k output resistance from a synthesizer is right at the limit.

[5] Never let an input float. The input impedence is so high stray currents can charge the input. The input should be either grounded, connected to V_S, or connected to a used input. The selection is based on proper logic function and loading.

[6] CMOS should never get even warm when it is functioning properly. If it gets hot, shut down power immediately. Some early CMOS IC's had input protection that could turn into an SCR and burn itself up.

CMOS INTERFACING

The major interfacing interest is CMOS to TTL and TTL to CMOS - (CMOS on 5 volt supply).

TTL to CMOS: TTL will directly drive 74C00 CMOS. For 4000 series CMOS, it is a good idea to use a pull-up resistor on the TTL output (e.g., 10k from the TTL output to +5).

CMOS to TTL: Since TTL is a current sinking logic and TTL inputs float up to the high state naturally if not held down, there is no problem driving the high state. The question is will CMOS sink the necessary 2 ma from a TTL input. The answer is no. It is necessary to have two parallel CMOS outputs to sink TTL. Note however that this referrs to the CMOS inverter structure. Looking back to the discussion of linear applications above we saw that the NOR gate had twice the sinking ability of the inverter since it has two channels to ground. This means that the CMOS NOR gate can be used to drive TTL.

Chapter 3h

MISCELLANEOUS MUSICAL IC'S

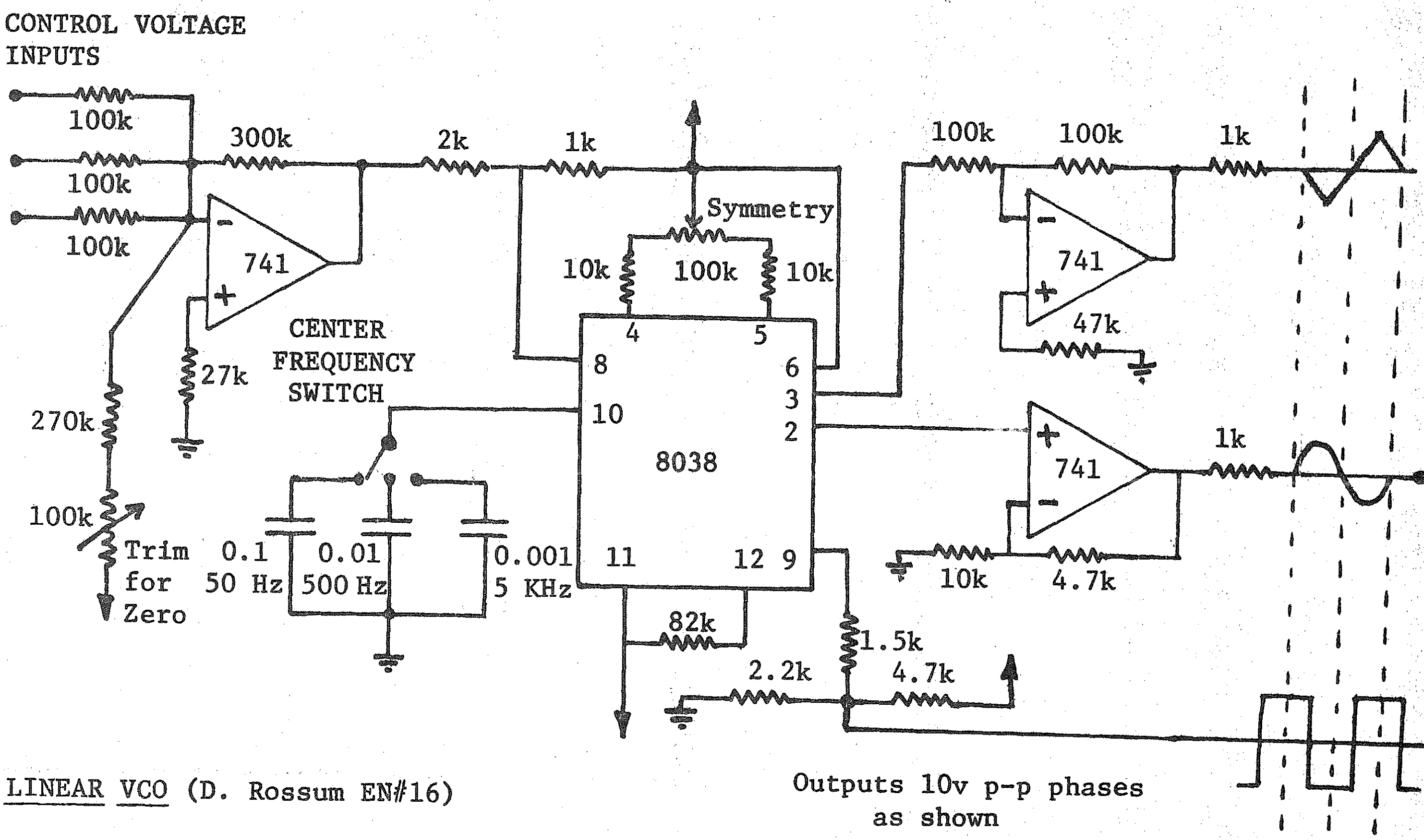
CONTENTS:

INTRODUCTION

There is a good deal of interest in IC's that are made with specific musical applications in mind (e.g., top octave generators), or which perform the general functions that are normally done by electronic music modules (e.g., a VCO chip). Some of these were intended for the consumer market for use in mass produced "electronic organ" type instruments. Of particular attraction to the musical engineer are those devices which might serve as an electronic music module or at least as a portion of it. As of this writing, no complete electronic music modules with satisfactory characteristics are readily available on a single chip. There are however many interesting and useful devices, a few of which are discussed below.

VCO AND FUNCTION GENERATOR CHIPS

A number of VCO chips (function generator chips) are available. All of these have a linear frequency/voltage response so are not directly suitable for the standard exponential keyboard VCO application. The earliest of these devices was the type 566 which still finds some applications. It provides square and triangular waveforms. A recent device is the Intersil 8038 which features triangle, square, and a sinewave output. The duty cycle of the rectangular waveform can be altered (along with the other waveforms) to provide a relatively sharp reset sawtooth. The sinewave is obtained by an on chip sinewave shaper which is driven by the triangle, and distortion can be trimmed to as low as 0.5%. Several attempts, with limited success, have been made to use this chip in electronic music VCO's by converting to an exponential response. It is probably most useful however to use it as a utility type of oscillator. Rossum (EN#16) has shown one such circuit for implementing voltage control and buffering the outputs. The circuit is shown below:



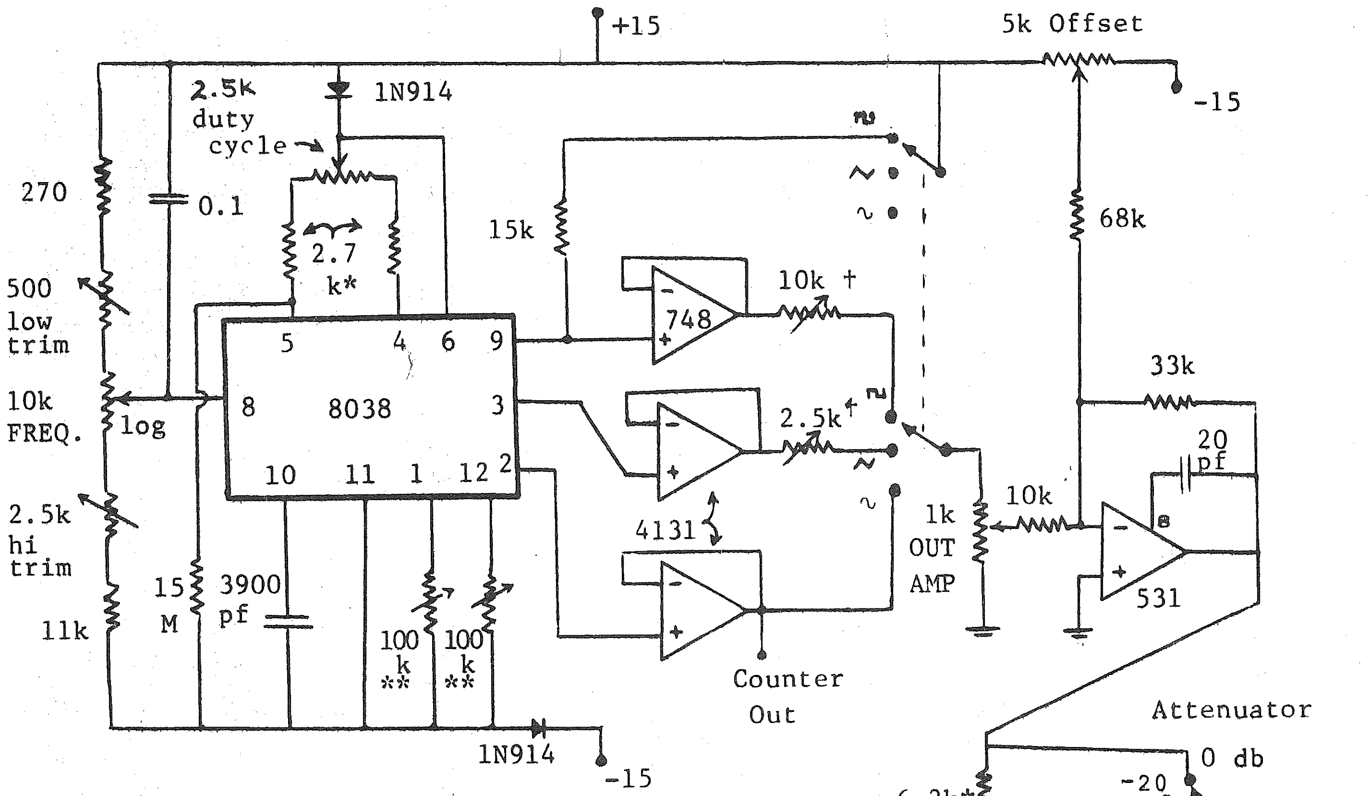
LINEAR VCO (D. Rossum EN#16)

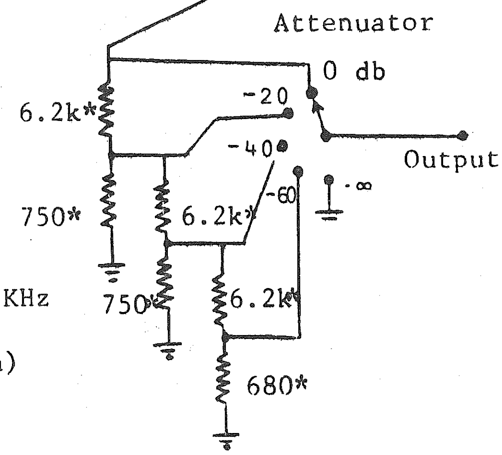
Outputs 10v p-p phases as shown

Anderton (EN#53) has described a circuit that uses the 8038 as a test bench oscillator. Note that this circuit is still voltage-controlled since the voltage on the pot wiper determines the frequency. The circuit also shows the use of a diode to effectively drop the positive supply voltage to obtain greater range as described in "Everything You Always Wanted to Know About the 8038," (Intersil, May 1973).

TOP OCTAVE GENERATORS

It is possible to start with a single oscillator operating in the low MHz range and employ suitable digital circuits (often divide-by-n counters) to obtain a set of frequencies in the audio range that approximate the tones of the equally tempered scale [see chapter 7b, pages 7 and 8 on "Digital Counting Interfaces," and the references listed there]. The need for such a set of tones is widespread enough to encourage a single chip implementation of the necessary circuits. One such chip is the Mostek MK5024 which operates on a single 15 volt supply and provides 13 tones (twelve tones plus an octave of the lowest one). The top note is a division of the input frequency by 239. The top octave generator can be used to generate a full keyboard of tones through the use of binary dividers which give the lower octaves. All the tones generated by this method are in effect locked together, so there is no "ensemble" effect created by playing more than one note. When a separate oscillator is used for each note (either in the top octave, or for all notes) it is possible to create richer blends due to slow beating between tones. In some applications however, the top octave generator is too much more efficient to be ruled out for lack of ensemble.



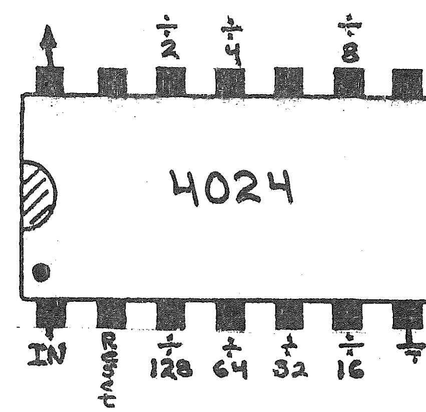


NOTES: * = 5% preferred
3900 pf should be polystyrene. Use large external cap. for subaudio signals
† amplitude adjustments - adjust triangle and square = sine amplitude
** adjust for min. sine wave distortion at 1 KHz

INEXPENSIVE FUNCTION GENERATOR (Craig Anderton)

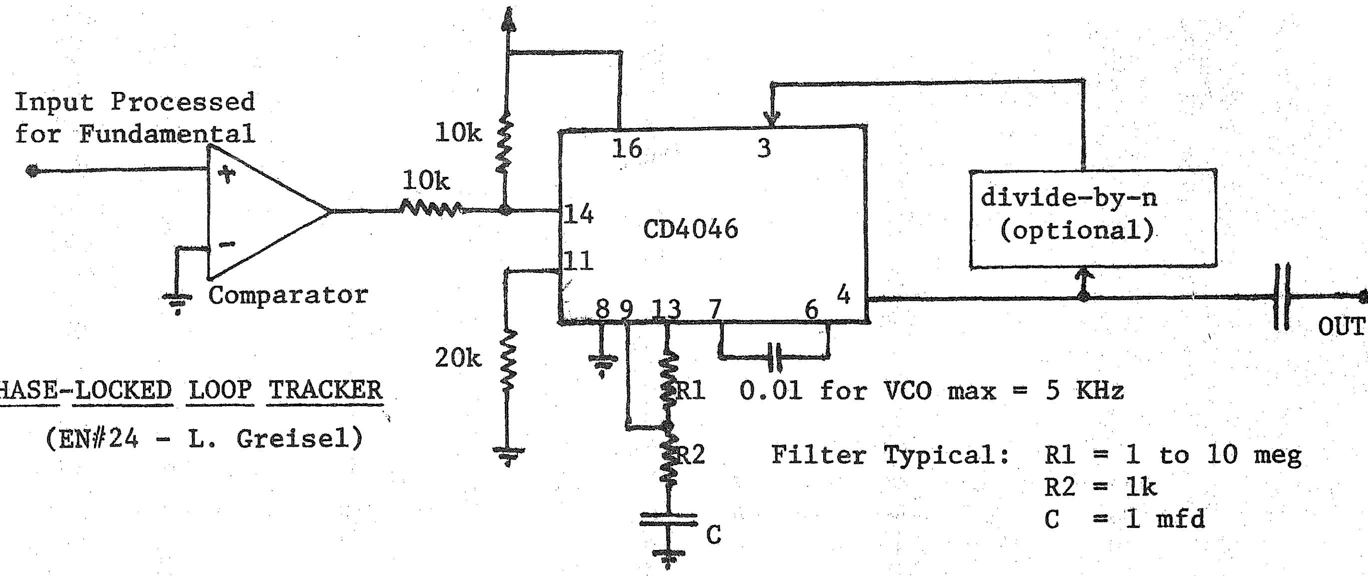
BINARY DIVIDERS

Square waves in octaves are generated by simply cascading flip-flops. These can be used to generate lower octaves for a keyboard for example. Such dividers can be implemented with single or dual JK flip-flops, but this in general requires a good deal of wiring (besides the interconnections between flip-flops, numerous J, K, preset, and clear inputs should be wired high). The type 4024 CMOS binary divider divides by 2, 4, 8, 16, 32, 64, and 128 on a single 14 pin chip.



PHASE-LOCKED LOOPS

Phase-locked loops are often used in frequency handling systems. The types 560 - 565 phase locked loop chips are simple and useful in limited frequency range applications. For general music handling, a wide range is required. The type 4046 phase-locked loop is useful in such applications as it discriminates frequency as well as phase, and can be used over a 1000:1 range. It cannot be used by itself as a pitch follower in a general system however. At least some input processing will be required even for a restricted pitch follower. One such tracking circuit has been described by Greisel (EN#24, pg. 4), and is shown below:



PHASE-LOCKED LOOP TRACKER

(EN#24 - L. Greisel)

The filter is the only part of the loop that needs to be optimized. The R1·C time constant determines the rate at which the loop slews from one input to the next. This permits the loop to serve as a glide circuit. The loop (like all phase-locked loops) does not lock completely on the input frequency, but rather wanders about the input. The rate at which it wanders is determined by the ratio R2 to R1. Smaller values of R2 cause a smooth wandering - larger values cause a more rapid jitter. Sample-and-hold can be obtained by inserting a switch between R1 and pin 13. When the switch is opened, the last frequency tracked will be held. The circuit has a wide capture range, but the comparator must be triggering on the fundamental. Thus, the success of the 4046 as a pitch tracker depends on input processing.

CHAPTER 4A

CIRCUITS USING IC AMPLIFIERS

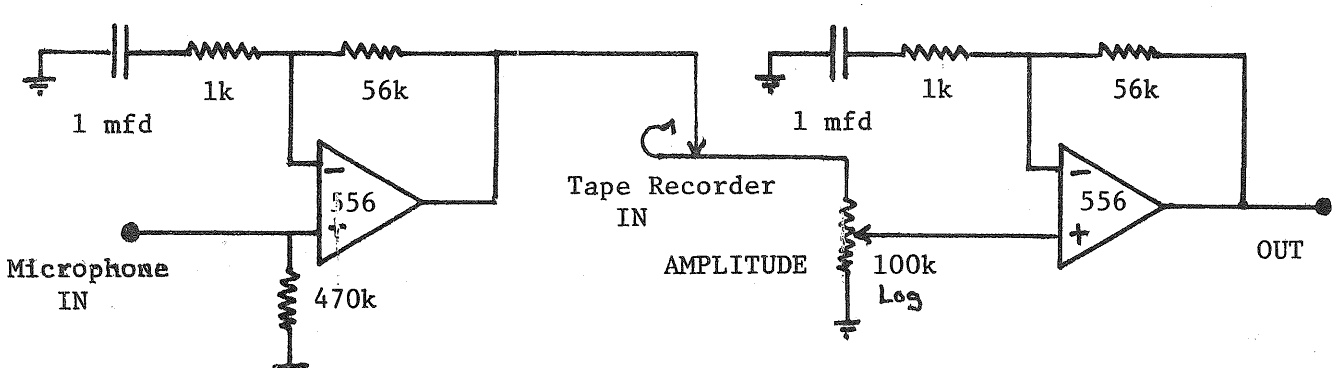
CONTENTS:

INTRODUCTION

There are numerous useful circuits that go somewhat beyond the basic applications of op-amps discussed in Chapter 3a, and yet are more general than specific electronic music modules to be discussed in chapters 5a through 7c. A few of these circuits that the reader should be aware of are given below. Additional applications can be found in the applications literature for op-amps, and books on op-amp applications such as Walter Jung's The IC Op-Amp Cookbook (H. Sams, 1974).

AUDIO CIRCUITS

Basically, most electronic music circuits are audio circuits - what we have in mind here are circuits for use with audio equipment in electronic music studios. We will illustrate the basic idea with an example, the microphone/tape recorder amplifier below:

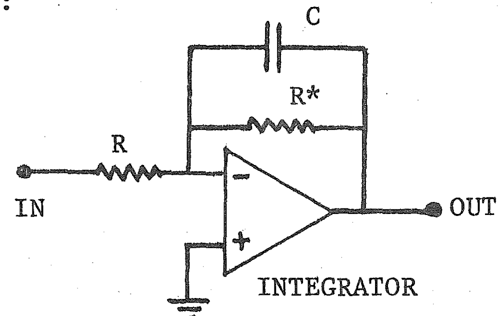


The circuit is useful when a signal from a microphone must be amplified for processing in synthesizer circuitry, or when a tape recorder level signal must be input. With the jack arrangement shown, plugging in a tape recorder level signal defeats the mic. input. The important thing to note about this circuit is the use of capacitors to return the AC signal to ground from the inverting input. This is a useful trick where high gain is used, as the DC gain is unity (a follower) and DC offsets for the most part can be ignored. The gain values are only suggested starting values. A synthesizer level signal may be plugged into the tape recorder input if the pot is used in its lower range. The circuit is a useful input amplifier for devices such as frequency shifters or pitch extractors where live signals must be input, but where it is also useful to be able to input test signals from a synthesizer or other higher level source.

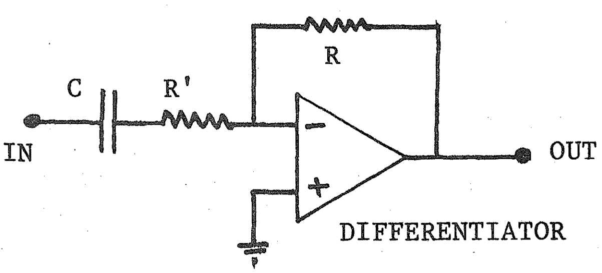
INTEGRATORS AND DIFFERENTIATORS

The main use of integrators in electronic music is for low-pass filter structures. The integrator is the basis of the popular state-variable filter. Differentiators find applications for detecting changes, as in sharp control voltage changes which occur when a note on a keyboard changes.

The basic integrator is an inverting amplifier with a capacitor in the feedback loop instead of a resistor. The basic differentiator is an inverting amplifier with a capacitor in the input to the summing node rather than a resistor. Often for practical circuits it is desirable to add a couple of resistors to limit the action of the circuits outside their intended regions of operation. The integrator for example would integrate its own offset voltage, and even if this were very small, the output would eventually ramp up. To prevent this, a resistor R* is added to bleed off excess charge on the capacitor. For practical purposes, this resistor can be much larger than the input resistor (several megohms for example). Often however, it is desirable to make it a smaller value. This turns the device into a single pole low-pass filter, but the region of interest may still be subject to a nearly ideal integrator. In state variable filters, the integrator is within a closed loop and this bleed-off resistor is not needed. The loop makes the necessary corrections. In a similar manner, a resistor R' may be added to the differentiator to limit the maximum gain of the device. Circuits for integrators and differentiators are shown below:



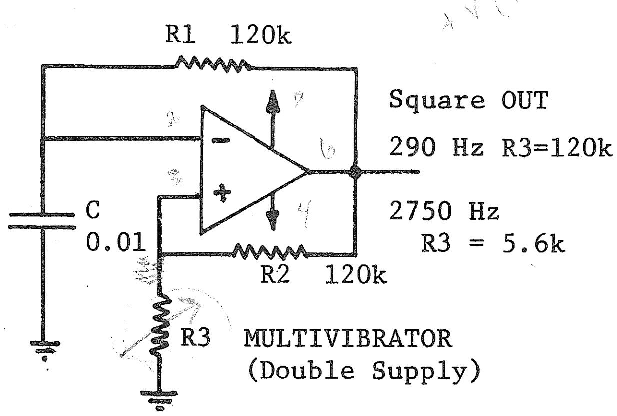
INTEGRATOR

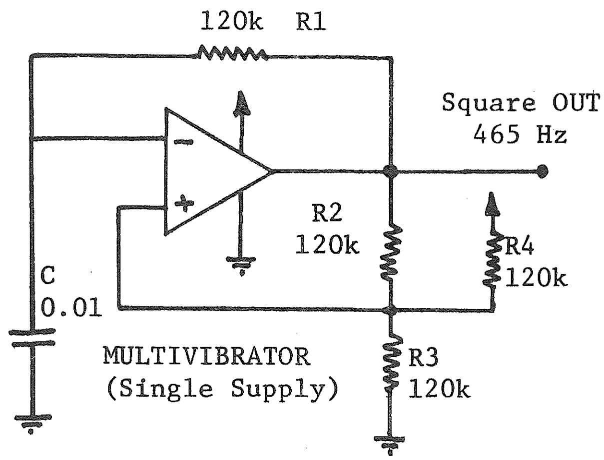


DIFFERENTIATOR

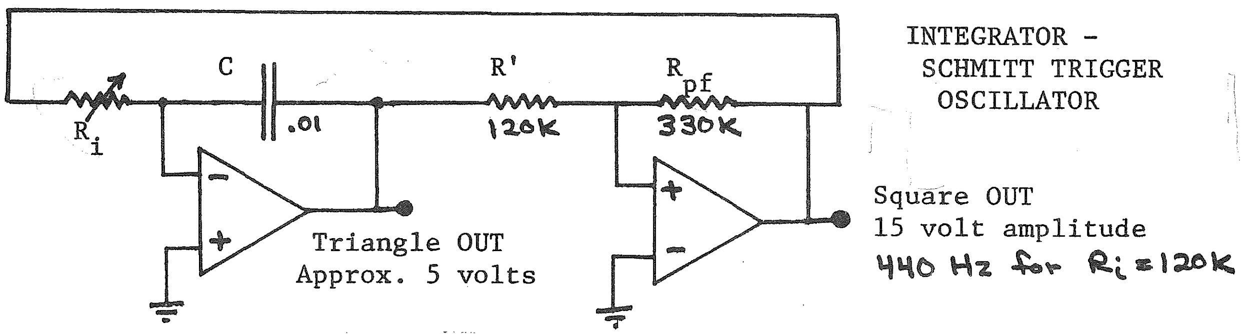
SIMPLE SIGNAL GENERATORS

Two circuits are popular for signal generators constructed from op-amps. The first is the multivibrator, and the second is the "Integrator - Schmitt-Trigger" type of triangle and square wave generator. The multivibrator uses a single op-amp as shown below. The output is a square wave that swings between the two supply volatges. The waveform at the - input is a triangle-like waveform but with curved slopes due to the exponential charging process. It is also possible by adding a single resistor to have a single supply generator working with an output between + supply and ground.





A triangle-square oscillator is formed with an integrator and a Schmitt trigger connected back-to-back. The frequency is adjusted by adjusting the integrating resistor R_i. The amplitude of the triangle can be adjusted by adjusting the resistor R_{pf} in the Schmitt trigger. In order to operate however, the value of R_{pf} must always be greater than R'. Frequency can also be adjusted by changing the capacitor, but the resistor is usually more convenient. If extended for very low frequencies (less than 1 Hz), it may be necessary to use ap-amps with low bias current requirements. If this is a problem, it will show up as a loss of waveform symmetry at the lower frequencies.

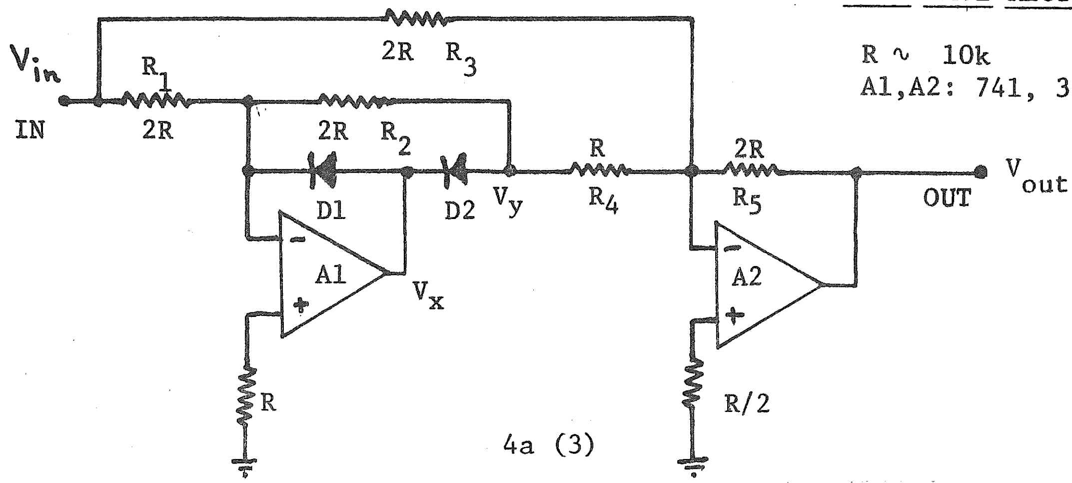


FULL-WAVE RECTIFIERS

A popular circuit for a full wave rectifier is shown below. When the input is positive, a current V_{in}/R_1 flows to the - input node (summing node) of A1. This current then flows out through R_2, setting V_y at $-V_{in}$, and through diode D2, setting V_x at V_y-V_d where V_d is the diode forward voltage drop (about 0.6 volts for a silicon diode). A2 is then seen to be summing $-(+V_{in}) -2(-V_{in}) = V_{in}$ through its inverting inputs R3 and R_4. When the input is negative, a current V_{in}/R_1 flows from the summing node of A1 to

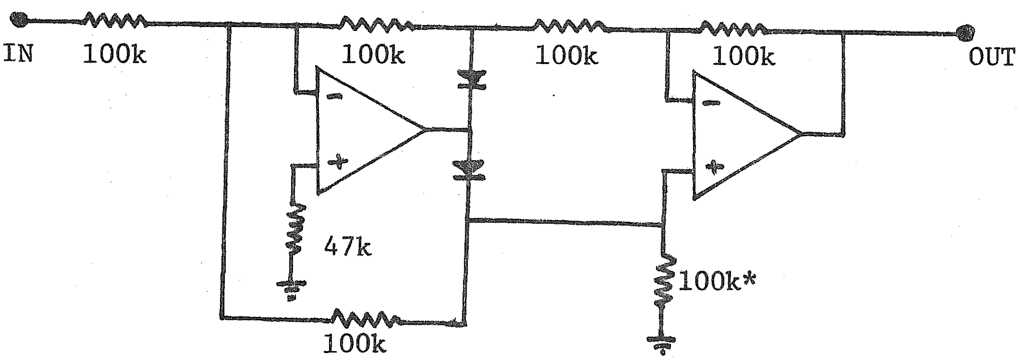
FULL WAVE RECTIFIER

R ~ 10k
A1,A2: 741, 307, Or 556



V_{in}. This current must be supplied to the summing node through D1. The only alternative paths for this current are through D2 and R_2 (for which the diode blocks the current), or from the summing node of A2 through R_4 and R_2 (but the summing node of A2 is a virtual ground just like the summing node of A1, so no current flows). By the same line of reasoning, no current flows through R_4, since V_x must be $+V_d$, and D2 blocks this voltage. No current flows between the virtual grounds through R_2 and R_4. Thus, all that is left is the inverter formed from A2, R_3 and R_5. Thus, the output is $-(-V_{in}) = V_{in}$.

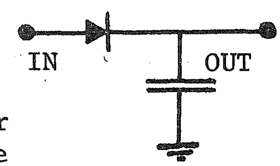
A second full-wave rectifier circuit that uses equal valued resistors (except for the 47k resistor which is not critical), was suggested by D. Rossum (EN#26, pg. 5). The circuit is shown below. The analysis is too much fun to give out (try it). Hints: The 100k resistor marked with a * is not essential to the analysis; assume the op-amps have zero differential input voltage, assume the right answer and work backward.



FULL-WAVE RECTIFIER

D. Rossum (EN#26)

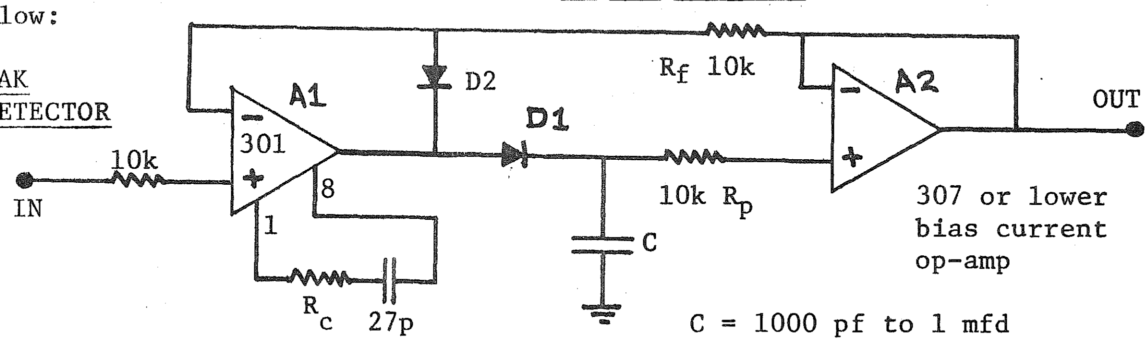
PEAK DETECTORS



The most basic peak detector circuit is just a diode and a capacitor connected as shown at the right. The capacitor charges to the input voltage (minus a diode voltage drop) and remains at that voltage when the input drops back. If a higher input voltage appears, the diode again conducts and charges the capacitor to the higher voltage. Thus the circuit output voltage is the highest (peak) voltage that the input has seen.

For a more paractical circuit, it is desirable to use an op-amp to compensate for the diode voltage drop, and buffer the capacitor voltage. A practical circuit has been described by W. Jung in his Op-Amp Cookbook. A typical circuit is shown below:

PEAK DETECTOR



The resistor R_p is a protection resistor for A2. It prevents dangerously high current levels that could flow during power shutdown. Resistor R_f balances the offset current of the + input of A1. Diode D2 speeds recovery. The actual peak detector is composed of D1 and C, and is placed inside the feedback loop of A1 and A2, thus correcting for the voltage drop across diode D1. Op-amp A1 is custom compensated to adjust for the extra pole in the response that results from the capacitor C and the output impedence of the op-amp. The resistor R_c is matched to the storage capacitor C as: $R_c = 2.7 \times 10^{12} \cdot C$.

CHAPTER 4B

ANALYSIS AND DESIGN OF ACTIVE FILTERS

CONTENTS:

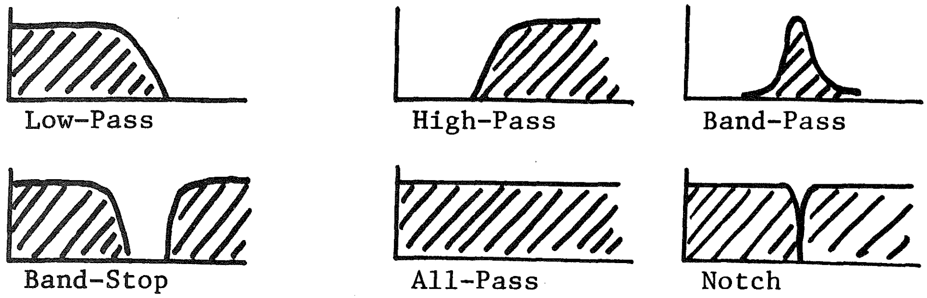
INTRODUCTION

The subject of active filters could fill several books this size and still just get started. Thus we are left here to consider the must useful approach to what has to be a brief discussion. The approach we are using here is to start with specific filter structures which you have probably seen, and to attempt to analyze them and examine the theories and procedures that lead to their design. This should enable the reader to attack other designs, or at least enable him to attack the relevant literature.

TERMINOLOGY OF ACTIVE FILTERING

There are many terms that are used to describe active filters and this may be confusing to the beginner and even to the professional engineer who has never had a complete course on active filters. We list below some of the terms and attempt to describe the most important aspects of the terms:

BASIC TYPE: Selection of basic type is dictated by the application. Basic types are: Low-Pass, High-Pass, Band-Pass, Band-Stop, All-Pass, and Notch. A rough sketch of the amplitude characteristics of each of these basic types is shown below. Note that the all-pass has a flat amplitude characteristic and is used for its phase properties:



CHARACTERISTIC: These are all the names of mathematicians: The most common are: Butterworth, Bessel, Chebyshev, Gaussian, and Butterworth-Thompson. Basically, this relates to a mathematical function that appears as the denominator of the "transfer function." This function in turn determines the finer features of the basic type.

FREQUENCY RESPONSE: The frequency response of the filter is the amplitude of the output as a function of frequency when the input is a sine wave of a constant amplitude. In the filter analysis, a complex frequency "s" is used for mathematical reasons. When it comes down to evaluating the actual frequency response, s is replaced by $j\omega$ where j is the square root of minus one. The transfer function $T(s)$ thus has a new variable $j\omega$ and the frequency response is $|T(j\omega)|^2 = T(j\omega) \cdot T(-j\omega)$. ω is the so called angular frequency and is related to the ordinary frequency by $\omega = 2\pi f$.

PHASE RESPONSE: All active filters shift phase as well as alter amplitude characteristics. The phase shift curve is a property of the filter. The total phase shift possible is 90° per pole. Some (all-pass) filters are used only for their phase properties (see chapter on frequency shifter design).

ORDER: The order of the filter relates to the number of RC sections in it, the number of poles, and the final rolloff of amplitude response. The order of the filter is the highest power of "s" that appears in the denominator (if the highest term is s^n, then the order of the filter is n). A "pole" of the filter occurs when the denominator takes on the value zero. Generally the denominator is a polynomial of order n, and it has n roots, hence n poles. The order of the filter usually indicates the final rolloff of high-pass and low-pass filters, and this is 6n db/octave. Generally, filters are built with second order sections. Each second order section has one op-amp, two capacitors, and two or more resistors.

NUMBER OF POLES: See order above.

CONFIGURATION: This relates to the way op-amps, resistors and capacitors are connected so as to realize a transfer function. In practice, there are relatively few useful configurations that can be used. Once a configuration is selected, and determined to be low-pass, high-pass, or whatever, the values of the R and C elements are chosen to realize the desired characteristic and frequency range.

TRANSFER FUNCTION: This is denoted by T(s) or H(s) (usually) and is the general output/input relationship for a configuration. It is determined by the simple application of circuit laws by representing all capacitors "C" by 1/sC, and doing calculations as though the 1/sC elements were resistors. The useful circuit laws are the usual:

1. The sum of the voltages around any closed loop = 0.
2. The sum of the currents into any node = 0.

The function E_{out}/E_{in} is determined as T(s). In the more useful configurations, T(s) will have a polynomial for the denominator. The basic type of filter is determined by the numerator. If it is $s^o = 1$, the filter is LP; if it is s^1, it is BP, and if it is s^2, it is HP. All-Pass filters have a denominator of the form $s^2 + bs + c$, or whatever form is necessary so that zeros appear in the numerator in the same way they do in the denominator.

It is now possible to see how the transfer function relates to the filter's frequency response. For example, consider the second order transfer function:

$$T(s) = \frac{s^n}{s^2 + bs + c} \qquad \text{for } n = 0, 1, 2$$

When n=0, the filter's response for small s (corresponding to low frequency) is just 1/c. For high frequencies, the response goes as $1/s^2$. This corresponds to a fall off of 12 db/octave. The response is thus low-pass, a constant at low frequencies, and falling off at high frequencies. When n=2, the low frequency response rises as the s^2 in the numerator. When the frequency is very high, the s^2 terms dominate the response which is then $s^2/s^2 = 1$, a constant. Thus, for n=2, the response is high-pass, rising at 12 db/octave to a constant value. For n=1, at low frequencies, the s^1 term dominates the low frequency response, rising at 6 db/octave. For high frequencies, the response falls of as $s^1/s^2 = 1/s$, a 6 db/octave fall off. This rise and fall results in a bandpass curve.

FREQUENCY SCALING: Once the characteristic has been set, the actual frequency of the filter characteristic can be raised (or lowered) by decreasing (increasing) the values of either all resistors or capacitors by an amount proportional to the frequency change desired.

IMPEDENCE SCALING: The frequency characteristic of the filter can be held in place while all resistors are increased by one factor as long as all capacitors are decreased by the same factor. This is commonly used to (1) adjust to a capacitor that is available. (2) change the input impedence level of the filter.

NORMALIZATION: This is a filter description where the characteristic frequency has been set to 1 radian. This usually sets some RC value to one. This often results in filters with 1 ohm resistors and 1 farad (!) capacitors. To obtain a practical design, the resistors are scaled up by some factor, and the capacitors are scaled down by the same factor. Normalization is used to simplify the mathematics.

The selection of a filter characteristic is determined by some requirement of the system to which it is being applied. No filter characteristic is ideal, and thus certain trade offs have to be made. The table below is perhaps oversimplified, but it does illustrate the types of trade offs that have to be made:

Characteristic	Major Advantage	Major Disadvantage
Gaussian (a)	Critical Damping (b)	Very Slow Rolloff
Butterworth	Maximally Flat Passband	Overshoots Step Response
Bessel	Constant delay (c) (linear phase)	Butterworth rolloff > Bessel rolloff > Gaussian Rolloff
Chebyshev	Very Sharp Rolloff (d)	Passband Amplitude Ripples
Butterworth-Thompson	Operates between the Butterworth and Bessel characteristics	

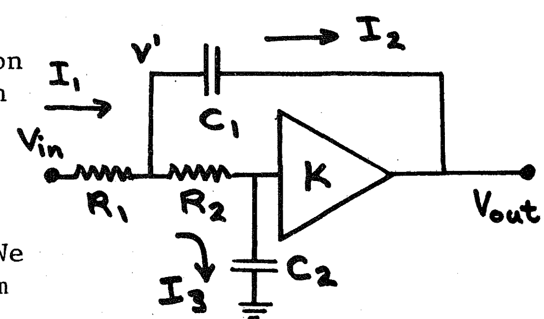
Notes: (a) (b) — The Gaussian filter is actually one or more cascaded simple RC filters with $T(s) = 1/(1 + sCR)$. Critical damping is observed in that these "filters" form simple AR envelope generators.

(c) — Since phase is linear with frequency, the time delay is constant. If the delay is not constant, a complex waveform will be distorted due to phase distortion. Bessel filters preserve waveform best.

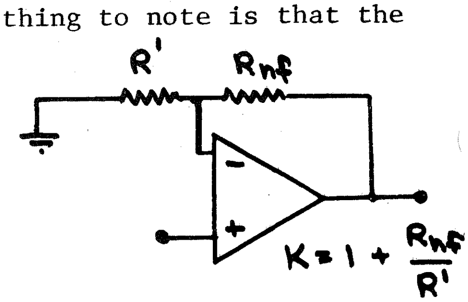
(d) — The very sharp rolloff occurs because the pass band ripples extend the corner of the response and it falls artificially fast. The corner rolloff may be much sharper than the final rolloff of 6n db/octave.

EXAMPLE FILTER ANALYSIS

As an example, consider the 2nd-order low-pass filter section shown at the right. First of all note that it looks like a low-pass filter. The R_2C_2 section is clearly a low-pass RC section, and C_1 is clearly in a position to reduce gain at high frequency. Later, we will want to show that it is a low-pass section by deriving the transfer function. Thus far, selection of the circuit at the right has determined the basic filter type, the order (2), and the configuration. We have still to determine the transfer function and from that select the characteristic.



The basic procedure in determining a transfer function is to first treat all the capacitors as though they were resistors of value $1/sC$, where s is what is called the complex frequency. Then use some network analysis. The first thing to note is that the amplifier marked "K" is just the ordinary type of noninverting amplifier as shown at the right. Thus the first step in the analysis is to realize that the input to the amplifier is simply V_{out}/K. We then set the unknown node voltage to V', and we can write equations for currents as follows:



$$I_1 = (V_{in} - V')/R_1 \qquad I_2 = (V' - V_{out})sC_1$$

$$I_3 = (V' - V_{out}/K)/R_2 = (V_{out}/K)sC_2$$

The I_3 equation can be solved for V':

$$V' = (V_{out}/K)(1 + sC_2R_2)$$

The node V' has only three connections, thus $I_1 = I_2 + I_3$. Plugging in I_1, I_2, and I_3 and solving for V_{out}/V_{in} gives the transfer function:

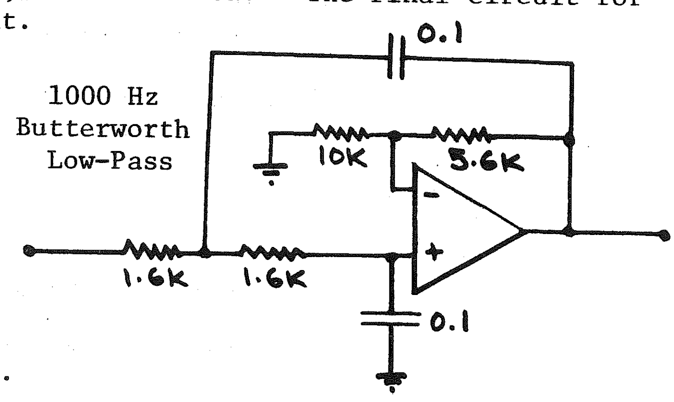
$$T(s) = \frac{V_{out}(s)}{V_{in}(s)} = \frac{K/(R_1R_2C_1C_2)}{s^2 + s\left[\frac{(1-K)}{C_2R_2} + \frac{1}{C_1R_1} + \frac{1}{C_1R_2}\right] + \frac{1}{R_1R_2C_1C_2}}$$

We have been careful here to indicate that when we write a transfer function to indicate that it refers to voltages at the inputs and outputs that are functions of s, not of t. The expression $T(s) = V_{out}(t)/V_{in}(t)$ makes no sense since the two quantities being equated are functions of different variables. What this means is that the voltages at the output and inputs are really the Fourier (or Laplace) transforms of the time domain voltages. The reason we so often think of the transfer function as a ratio of time varying functions is that it is always assumed that what we really mean is the study state amplitudes of sine waves.

A transfer function with only the zero power of s in the numerator, and a function of the from $s^2 + as + b$ in the denominator is a low-pass filter function. So far, nothing prevents us from selecting as many elements as we can to have whatever value is convenient. We can make a particular selection with $R_1 = R_2 = R$ and $C_1 = C_2 = C$. Thus makes T(s):

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

Now we have to make a choice of characteristic. If we choose a Butterworth, we can consult design tables and find that the denomonator should be $s^2 + \sqrt{2}\omega_o s + \omega_o^2$. From this we can see that ω_o must be identified with 1/R C which in turn means that [3 - K] must be set to 1.414, or K = 1.586. The meaning of K = 1.586 means that we just form the non-inverting amplifier with R' = 10k and R_{nf} = 5.86k (or probably the standard value 5.6k will be close enough in many applications). In the case of the Butterworth filter, ω_o is the frequency at which the amplitude falls to 0.707 (down 3 db) of its pass band value. This means that to set any frequency, we can start with a convenient value of R or C (and usually we choose C since it is easier to get a wider variety of values for R). For a cutoff frequency (3 db freq.) of 1000 Hz, $\omega_o = 2\pi f_o = 6283$. If we choose C = 0.1 mfd, then R = 1.6k. The final circuit for the 2 pole Butterworth is shown at the right.



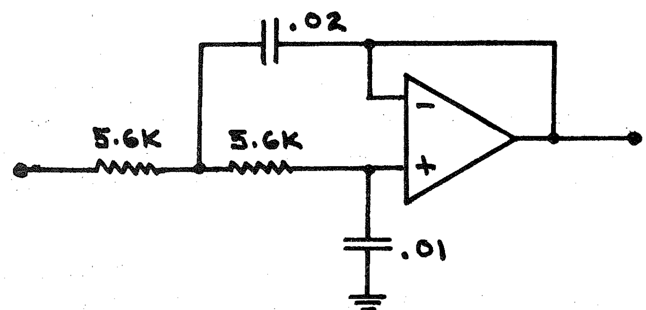
Note that the low frequency gain of the circuit is K, and not unity. The circuit does provide some gain in the pass band. In many cases, this is fine, but at times we can not allow this because if we have many such stages cascaded for higher order filters, the signal may clip against the supply levels. It is a simple matter to set K = 1 by just using a follower. With this new restriction, we will not be able to set the R and C values arbitrarily equal. The transfer function for K=1 becomes:

$$T(s) = \frac{1/R_1R_2C_1C_2}{s^2 + \frac{s}{C_1}[1/R_1 + 1/R_2] + 1/R_1R_2C_1C_2}$$

To make this take on the Butterworth characteristic, we can set $R_1 = R_2 = R$ and $C_1 = 2C_2 = 2C$. The transfer function then becomes:

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

Thus $\omega_o{}^2=1/2R^2C^2$ or $\omega_o = \sqrt{2}RC$. The (1/RC) coefficient of s is thus $\sqrt{2}\omega_o$, exactly the Butterworth case. It is indeed fortunate that the particular selection of $R_1 = R_2$ and $C_2 = 2C_1$ gives the right result. At worse, it is necessary to select three capac capacitors of the same value, use one of them for C_1, and the other two in parallel to give C_2. For example, select a cutoff of 2000 Hz. $\omega_o = 12566$. Select C_1 =0.01 mfd and C_2= 0.02 mfd for convenient values. This gives R = 5.6k. The circuit is shown below:



2000 Hz Butterworth
Low-Pass
Unity Pass-Band Gain

To form a higher order filter, it is necessary to obtain a transfer function with a higher order denominator. For example, a fourth order transfer function would have a form like the following for a low pass filter:

$$T(s) = \frac{1}{s^4 + bs^3 + cs^2 + ds + e}$$

This is difficult to do directly. What is usually done is to first factor the denominator into two second order polynomials.

$$T(s) = \frac{1}{(s^2 + fs + g)\cdot(s^2 + hs + i)}$$

Then, it is possible to realize two second order filter sections and cascade them. It is <u>not true</u> however that two cascaded second order Butterworth filters will give a fourth order - they will not. To obtain the fourth order Butterworth, it is necessary to work with two different second order filters, neither of which is a Butterworth, but which work together to give the higher order. Recall that the second order Butterworth polynomial (normalized to RC=1) is:

$$s^2 + 1.414\ s + 1$$

The factored fourth order Butterworth polynomial is:

$$(s^2 + 0.765\ s + 1)\cdot(s^2 + 1.848\ s + 1)$$

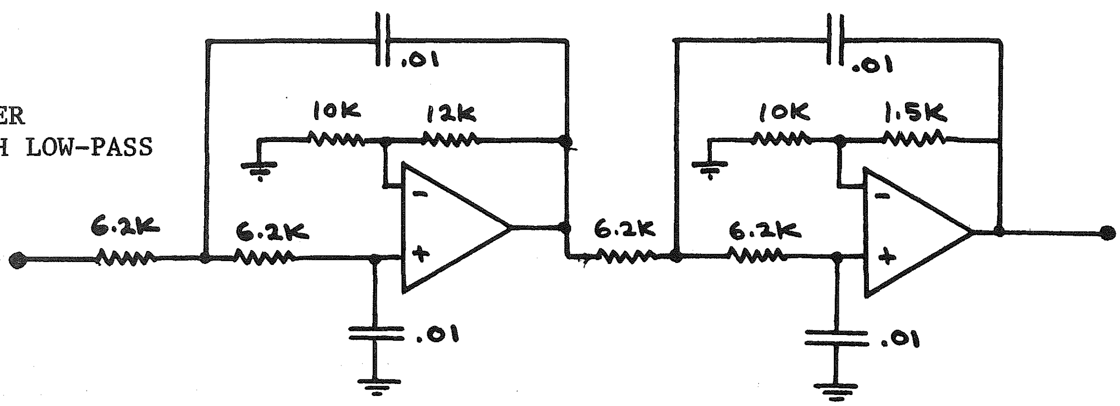
These denominators can be realized exactly as the first order polynomial was. We can choose the method using equal valued resistors and capacitors. With this, we can select the R and C values to suit the frequency we need, and select the gains K according to:

$$3 - K_1 = 0.765 \qquad K_1 = 2.235$$

$$3 - K_2 = 1.848 \qquad K_2 = 1.152$$

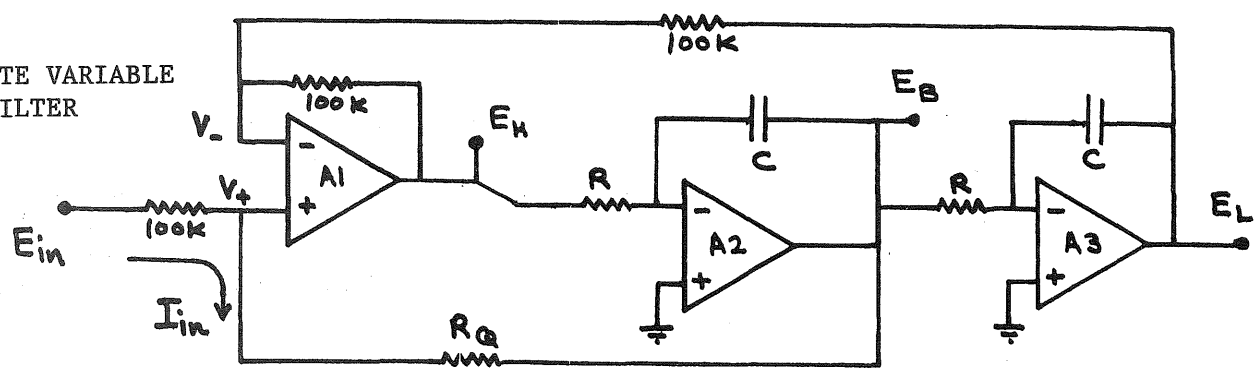
As an example of frequency scaling, suppose we want the filter's 3 db frequency to fall at 2580 Hz. We saw in the case of the 2nd order filter that the R and C values were 0.1 mfd and 1.6k for 1000 Hz. We can increase the frequency by decreasing the resistance. Since the frequency is to go up by a factor of 2.58, the resistance would divide by the same factor. This gives 1.6k/2.58 = 620 ohms. However, we might decide that the input impedence is too low with resistors this small. We can then do an impedence scaling. Here we change the resistors up by the same factor we change the capacitor values down (maintaining the RC product). Thus, we can select a capacitor of 0.01 mfd, and a resistor of 6.2k. The final circuit is shown below:

FOURTH ORDER
BUTTERWORTH LOW-PASS

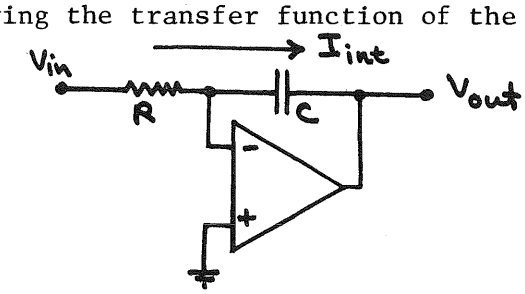


An extremely useful filter for electronic music is the "state variable" type. This filter is simple, reliable, and provides high-pass, low-pass, band-pass, and a notch output. A basic state variable filter is shown below.

STATE VARIABLE
FILTER



Note that the filter is just two integrators in a loop with a summing input network. We can simplify analysis greatly by first deriving the transfer function of the integrator as shown at the right. With an applied voltage V_{in}, a current I_{int} flows into the op-amp summing node through R. Thus



$$I_{int} = V_{in}/R$$

The same current must flow into the capacitor. Thus the output voltage must be:

$$I_{int} = V_{out}/(1/sC)$$

or: $$V_{out}/V_{in} = -1/sCR$$

This done, we can get rid of the node voltages E_B and E_L by relating them back to E_H as:

$$E_B = -E_H/sCR \qquad \text{and:} \qquad E_L = E_H/s^2C^2R^2$$

Next, we determine the current $I_{in} = (E_{in} - E_B)/(100k + R_Q)$

We then realize that the differential input voltage is zero for A1 due to negative feedback. Furthermore, the voltage V_- is the average of E_L and E_H. Finally, we can write V_+ from E_{in} and I_{in}. The proceeding three sentences can be summarized as:

$$\frac{E_L + E_H}{2} = V_- = V_+ = E_{in} - \frac{100k}{R_Q + 100k}(E_{in} - E_B)$$

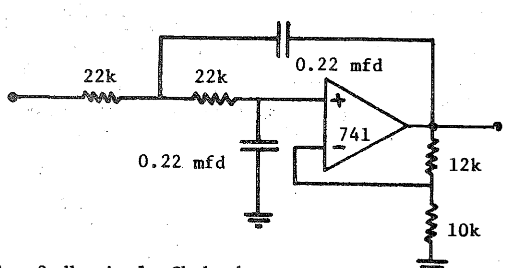
Next substitute for E_L and E_B the expressions derived from the integrator equation. Represent $100k/(R_Q+100k)$ as K, and solve for E_H/E_{in}. The result is:

$$E_H/E_{in} = (1-K)\frac{2s^2}{s^2 + \frac{2K}{RC}s + \frac{1}{R^2C^2}}$$

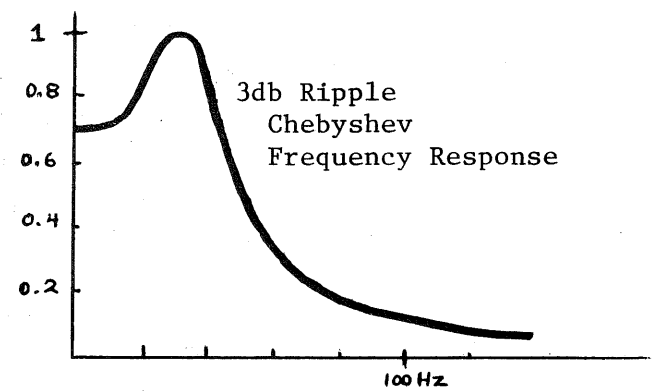
This is clearly a high-pass transfer function. The E_H output is the high-pass output of the filter.

It is then possible to see that the first integrator produces the output E_B which is a bandpass since multiplication by $-1/sCR$ removes a power of s from the numerator. Likewise, the second integrator removes another power of s from the numerator to give E_L as a low-pass output. Deriving the notch output is a matter of summing the low-pass and high-pass outputs. A careful study of the phase relationship shows that at the cutoff frequency, the two are 180° out of phase and when added, they cancel. The state variable filter can be set to give a Butterworth response by setting $2K = \sqrt{2}$. The control R_Q sets the Q of the filter by shaping the characteristic.

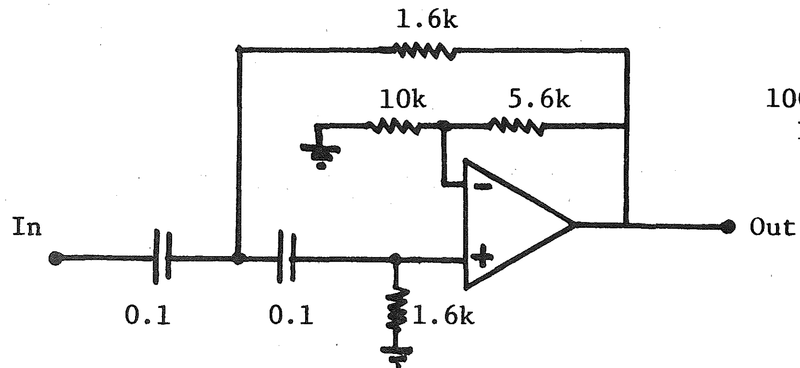
We should give one example of a filter having a different type of characteristic. For example, a 3db ripple 2nd order Chebyshev. Consulting filter tables we find that the appropriate denominator is: $s^2 + 0.767\ s + 0.841$. Thus if we want a 3db roll off point at 42 Hz, we design for a frequency of $0.841 \cdot 42$ Hz or approximately 35 Hz. Thus $1/RC = \omega_0 = 2\pi \cdot 35$ Hz. Setting C = 0.22 mfd for example, we get R at 22k. We can then use the variable K low-pass configuration and set 3-K = 0.767 giving K = 2.233. The filter circuit and its frequency characteristic are shown below:



Second-order 3 db ripple Chebyshev

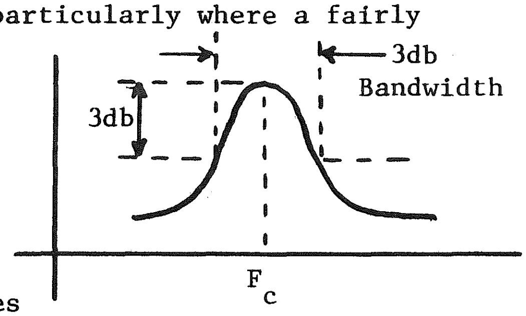


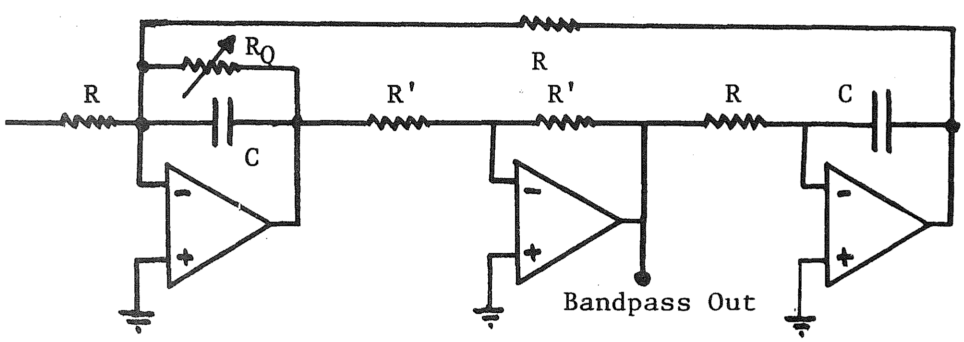
For the formation of high-pass filters, it is often possible to just change the positions of resistors and capacitors in the configuration. Very often. other basic types of filters are formed from a low-pass prototype. Below we have converted the second order Butterworth low-pass to a second order high-pass Butterworth.



1000 Hz High-Pass
Butterworth

The formation of band-pass filters that are of practical use can be more difficult. We have already discussed the use of the state variable filter as a band pass filter, and this is often a good choice, particularly where a fairly high Q is required. We shall define Q to be the center frequency (F_c) divided by the 3db bandwidth. This is illustrated by the diagram at the right. Note that this can be easily measured and provides a measure of the sharpness of the filter peak. A second type of multiple amplifier band-pass is the "Biquad" which is closely related to the state variable. Note that the Biquad does not have the regular state variable Q control but instead achieves damping by means of a resistor R_Q connected across the capacitor of one of the integrators. The Biquad is also capable of high Q (above 10 for example).





Biquad as
BANDPASS

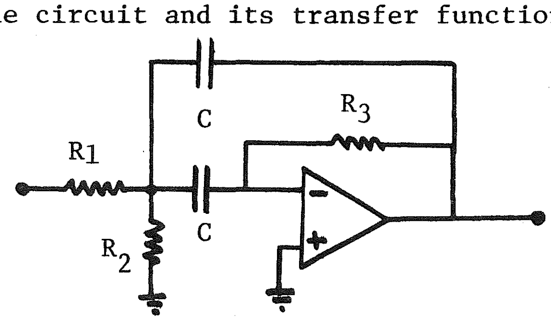
$F_c = 1/(2\pi RC)$

R_Q ranges from R to 100R

$R' \sim R$

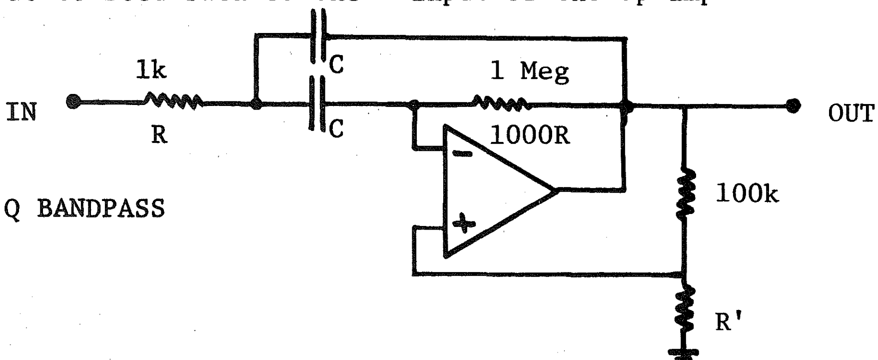
When it comes to single amplifier bandpass filters, it is most useful to go from the controlled gain type of configuration (amplifiers set for gain of K) and instead use the infinite gain ideal properties of op-amps. This configuration is known as the multiple feedback bandpass configuration. The circuit and its transfer function are given below:

$$T(s) = \frac{-s/CR_1}{s^2 + 2s/CR_3 + \frac{R_1 + R_2}{C^2 R_1 R_2 \; R_3}}$$



The constant term in the denominator is just ω_0 and the coefficient of s in the denominator is $\alpha\omega_0$ where $\alpha = 1/Q$. The circuit is suitable for low Q (less than 10).

When it is necessary to use higher values of Q, the state variable of biquad can be used. Another circuit that has been popular as a variation on the multiple feedback infinite gain which uses positive feedback [M.V. Mathews & J. Kohut, "Electronic Simulation of Violin Resonances," JASA 53 #6, June 1973, pg. 1620]. The circuit uses a voltage divider on the output to feed back to the + input of the op-amp:



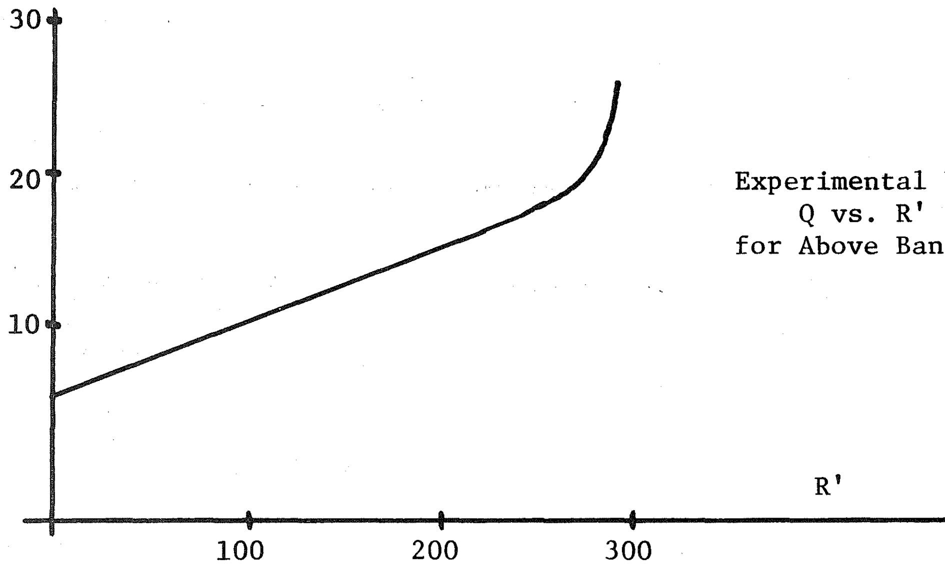
HIGH Q BANDPASS

The center frequency is determined by the RC time constant. An experimental formula for the the frequency is:

$$f_c = 1/(185 \cdot RC)$$

The Q of the filter is determined by the feedback ratio 100k/R' with R' being in the range of a few hundred ohms. An experimental plot of Q as a function of R' is shown below:

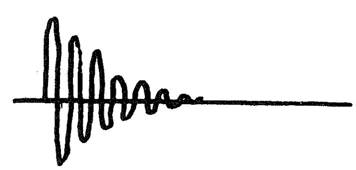
BANDPASS Q



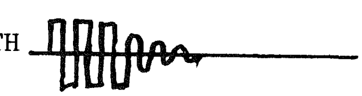
Experimental Plot of
Q vs. R'
for Above Bandpass Filter

High Q bandpass filters are often employed as tone generators by exciting them with a sharp pulse and letting the inserted energy damp out. The output is an exponentially decaying sine wave. The phenomenon is known as ringing. Oscillation is at the filter's center frequency, and the sound is remindful of raindrops or certain percussion devices. It is also possible to add a clipping network to the output which will limit the amplitude out providing a change of harmonic content. The relationship between the Q and the ringing time is neatly summarized by $T_r = Q/\pi F_c$ where T_r is the 1/e time constant - the time for the amplitude to go from 1 to 1/e ≈ 37%.

RINGING



RINGING WITH CLIPPING



CHAPTER 4c

CIRCUITS USING DISCRETE SEMICONDUCTORS

CONTENTS:

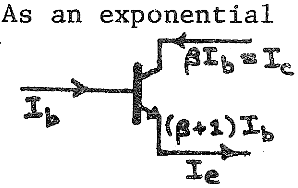
INTRODUCTION

In general the experimenter or individual builder of electronic music equipment need not resort to discrete components in their designs except in certain special cases such as those described below. At times, the commercial builder will be able to obtain a substantial saving by using less expensive discrete components in place of integrated circuits, sometimes with some sacrafice in performance.

At times, the discrete transistors we will be discussing may be best obtained from transistor arrays. This is true of the exponential and log circuits where the arrays offer an anvantage in temperature stability. On the other hand, in certain cases discrete components must be used to obtain the necessary speed. If a transistor array is used in such a case, the circuit may be slowed down by the capacitances of the close components, and the wider separation of the discretes was the main reason for expecting faster speed in the first place.

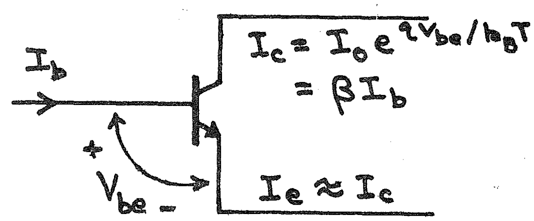
THE BASIC BIPOLAR JUNCTION TRANSISTOR (BJT)

The BJT is the "regular" type of transistor. It has three modes of operation that are of interest to us. These are: (1) As a current amplifier, (2) As an exponential (or log) current source, and (3) as a switch. The BJT is basically a current amplifier. The collector current is equal to the base current times a factor β or h_{fe}. For a good transistor, this β should be from 50 to several hundred, and fairly constant as

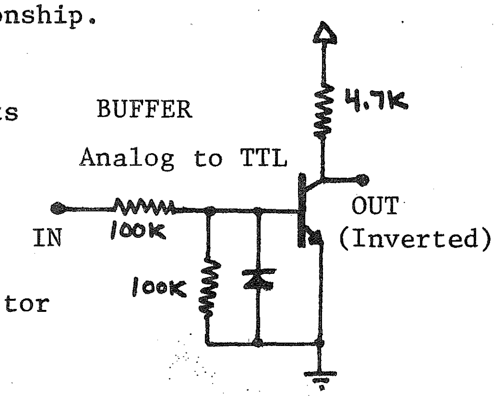


the current levels change. In the high-β approximation, $I_c \approx I_e$ and I_b can be considered negligible as long as the biasing circuit for the base has no problem supplying the current. This means that the standing currents in the biasing circuit should be substantially larger than the required base current.

The β-relationship between the base current and the collector current applies as long as the transistor is in the active region (not cutoff, or in saturation). When we move on to exponential circuits, the importance of this relationship is that for high-β transistors, we can ignore base currents. Then we can look at the relationship between the collector current and the base-to-emitter voltage (V_{be}). We find that this is an exponential relationship. Note that the β-relationship and the exponential relationship apply simultaneously - there is no conflict between the two, and the behavior of the transistors is not "programmed" for one relationship or the other. In regard to exponential or log circuits, it may be best to consider the results of applying currents to the collectors of the transistors. When driving the transistors from a current source, as with any current source, there has to be a voltage adjustment as the load changes (Ohm's law still works). This adjustment for the transistors is in terms of changing base-to-emitter voltage. Finally, we can consider that a small base current is actually flowing into the base to maintain the β-relationship.

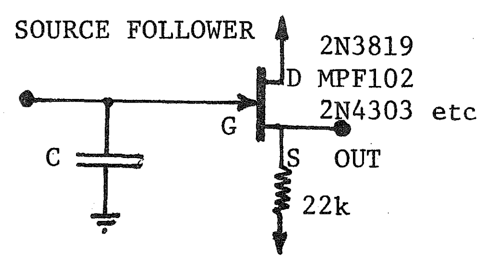


In the final mode of operation, the transistor acts as a switch. Typically, this is employed in logic circuits where only on and off states are involved. In particular, a single transistor and a few other components can be used to make a buffer between analog and digital circuits. In this case, enough base current is supplied to the base to put the transistor out of the active region, and close to or into the saturation region. In this region, the transistor collector is only a fraction of a volt above ground and is thus useful for a logic zero or for discharging capacitors to ground.



ELECTRONIC MUSIC APPLICATIONS FOR FET'S

There are three basic uses of FET's in electronic music devices: (1) As source followers for buffers, (2) As voltage-controlled switches, and (3) as triangle-to-sine waveshapers. There may also be applications as voltage-controlled resistors over limited ranges (see chapter 5c). A source follower to buffer the voltage on a capacitor is shown at the right (useful in sample-and-hold circuits and integrators). The circuit is shown with an N-FET but could equally well use a P-FET if it were turned upside down. The source follower will have a DC offset amounting to a fraction of a volt. In many circuits however, this buffer is inside a feedback loop and the offset is compensated by the loop. The FET can also be used as a switch. For example, a negative voltage on the gate will hold an N-FET off. If the gate voltage is allowed to float up, a voltage will pass between the source and the drain. The use of a FET as a sine shaper is discussed in chapter 5b.



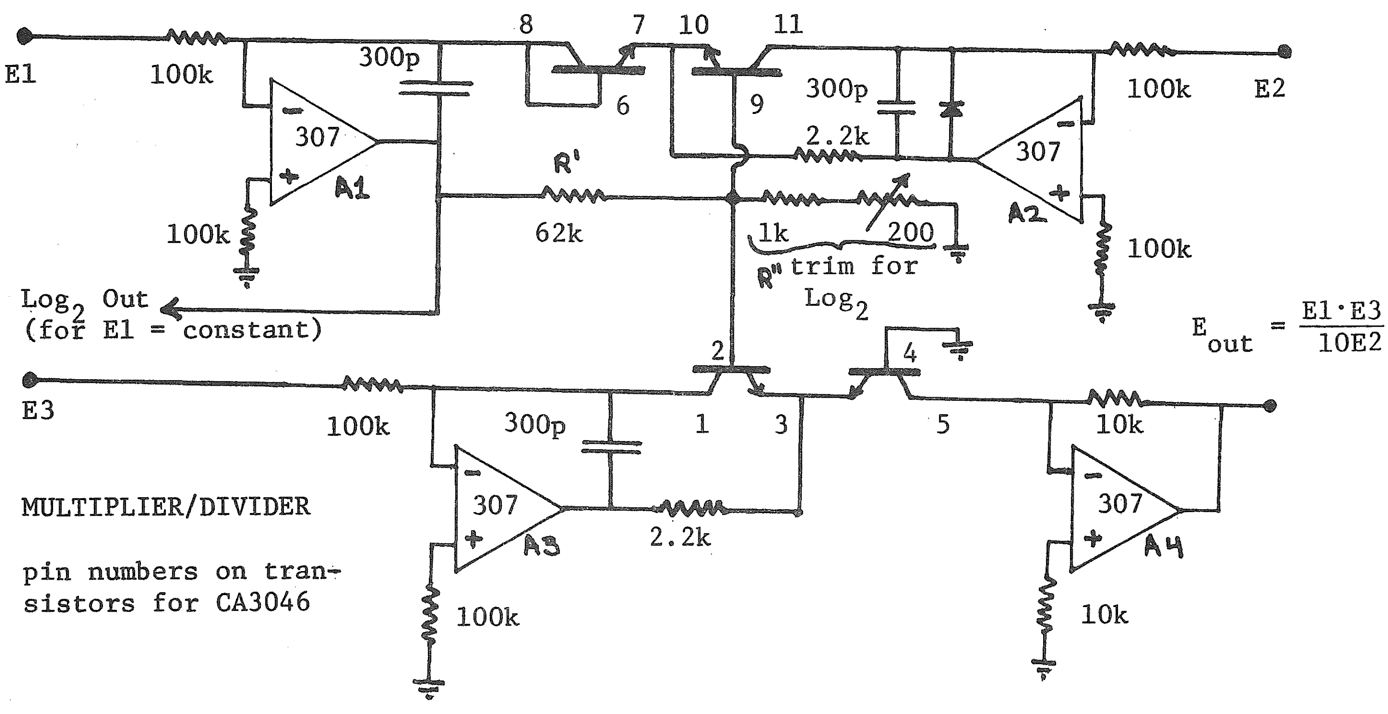
CIRCUITS EMPLOYING TRANSISTORS FOR THEIR LOGARITHMIC PROPERTIES

The collector current of a transistor is an exponential function of the base-to-emitter voltage:

$$I_c = I_o e^{qV_{be}/k_BT}$$

This relationship will be used often in chapter 5b where such transistors will be used to generate the necessary exponential control relationships for VCO's and VCF's. We also have need on occasion for circuits performing the inverse function (logarithmic) and it is usually the case that the circuit should take log to the base two to correspond to the standard 1 volt/octave exponential response. Such a log amplifier is shown in Fig. 15 of page 2C (18).

Another advantage of logarithms is that they change multiplication and division to addition and subtraction. Thus, log circuits are useful when analog multiplication and division are necessary. [Note that analog multiplication is often performed with the four-quadrant transconductance multiplier. By putting the multiplier in the feed-back loop of an op-amp, it is possible to divide - see the application notes on the type 595 multipliers. However, the linearity of such a circuit is not particularly good, and they are difficult to adjust]. The circuit below has a wide range and good linearity due to the use of log circuits. It has been adapted from National Semiconductor application notes AN31-19. The circuit multiplies and divides according to the inputs used. An attenuator consisting of R' and R" has been added to cause the circuit to opearte on a log to the base two level. This output can be brought out and used to drive exponential VCO's and VCF's for example. If only the linear divider is needed, the attenuator can be removed and the output of A1 can be fed directly to the bases of the transistors.

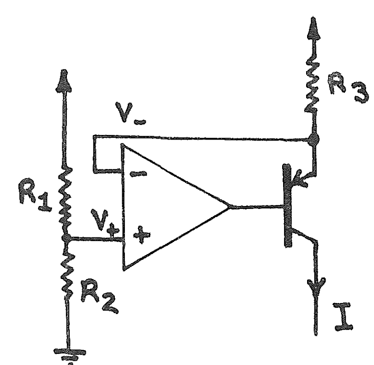


MULTIPLIER/DIVIDER

pin numbers on tran-sistors for CA3046

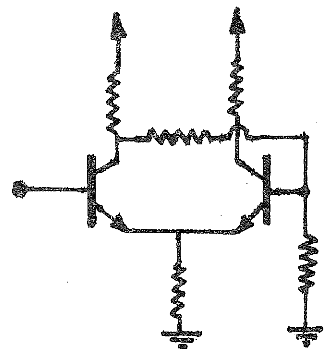
CURRENT SOURCES

A popular current source is shown at the right: The voltage $V_+ = (15 \cdot R2)/(R1+R2) = V_-$. Thus I_e for the transistor is $(15-V_-)/R3$. For high-β transistors, $I_c \approx I_e$. Thus, the current is regulated by the op-amp which adjusts the base voltage as is necessary. The current source will deliver the current $(15-V_-)/R3$ to any load that is at a potential from -15 up to within a volt or so of the voltage $V_+ \approx V_-$. Additional current sources and sinks can be found in National Semiconductor applications notes AN-20 and AN-31.

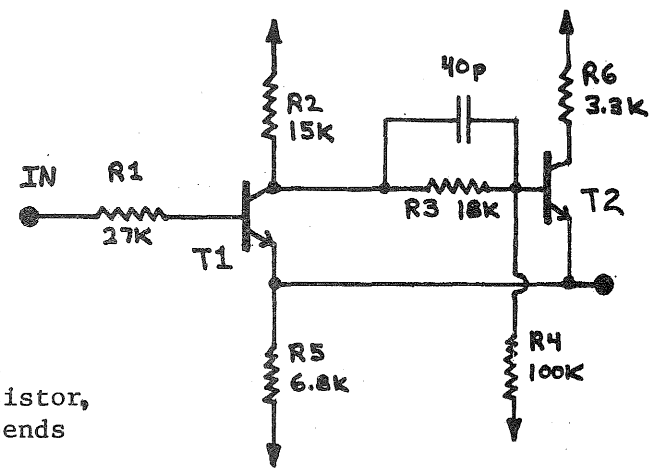


SCHMITT TRIGGERS WITH TRANSISTORS

VCO's often require a fast switch, and in the case of the integrator - Schmitt trigger type of oscillator, the two transistor discrete Schmitt trigger is hard to beat. The basic Schmitt trigger formed from two transistors is shown at the right where it has been drawn to show that it is basically a differential amplifier structure with positive feedback. The importance of the feedback is that it adds hysteresis to the output with respect to changes of input voltage.

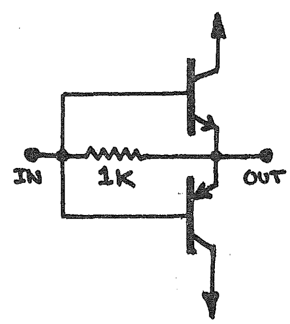


A Schmitt trigger that can be used to feed back ±5 volts from the tied emitters and which also has approximately the same ±5 volt trip points is formed from two transistors connected across a bipolar supply and is shown at the right. With T1 off, the base of T2 is biased at about +6 volts by R2, R3, R4, and the base current of T2. T2 is on and divides the ±15 supply across R6 and R5 so that the tied emitters are at about +5 volts. When the input voltage exceeds +5, T1 turns on and T2 is regeneratively shut off. This divides the ±15 supply across R2 and R5, and the tied emitters are at approximately -5 volts. The design example in chapter 5b uses this circuit. The 40pf capacitor is a "speed up" capacitor that is used to remove stored base charge in the transistor, thus speeding the switching. The exact value depends on the transistors used.



OP-AMP POWER OUTPUT STAGE

At times, it is necessary for an op-amp output to drive a load that requires more current than can be easily supplies (say more than 10 ma). In such cases, an emitter follower may help in many cases. For best results however, a power drive stage consisting of "Back-to-back" emitter followers as shown at the right should give the best results. The transistors should be a complimentary pair if possible.



CHAPTER 5A

MODULAR DESIGN - GENERAL CONSIDERATIONS
INPUTS AND OUTPUTS

CONTENTS:

INTRODUCTION:

An electronic music module is a self-contained (except for power supply) functional module for use in an electronic music synthesis scheme. Thus, a module is a part of a larger synthesis system. It can be modular in the sense that it can be pulled out of one physical position and plugged into another, or the modularity may be more conceptual in the "block diagram" sense. What we have in mind here is a unit which has a physical position (not necessarily movable) and which has accessable electrical terminals. These terminals are connected to the terminals of other modules through the process of "patching." By using a modular system as opposed to a "hard wired" system, the user is able to realize a wider variety of synthesis schemes.

A module consists of two parts: the electrical functional block (the "guts" of the module) and its interface with the rest of the world. Often this simply comes down to a circuit board on the one hand and the front panel fixtures on the other hand. This is the sort of picture we have in mind when we discuss a general module.

The module's functional section ("submodule" or circuit board) is the most difficult part of the design, and each different type requires a separate discussion. This is the purpose of the remaining chapters in Section 5 and Section 6. In this chapter, we want to discuss the interfacing procedures, as these are common to most types of modules.

The main interfacing mechanisms are: signal inputs, control inputs, timing inputs, panel controls, and outputs. These mechanisms in turn can be grouped according to the type of electrical structure and signal routing that is used:

1. Voltage Input Structures: These voltage inputs may be either control inputs or signal inputs; the only difference between signals and control is the use, not the input structure. Generally, an op-amp summing node is the voltage input structure. In addition, often times voltages from the panel controls enter the functional blocks through the voltage input structure. For example, the voltage input for the control voltage of a VCO often serves to handle externally applied voltages, and "Coarse" and "Fine" tuning voltages from panel pots.

2. Timing Signal Inputs: For timing signals, often the voltage level is of no importance. The important thing is that the voltage meets certain conditions, and that the time at which it meets the conditions is well defined. For example, we might want a "Gate" to appear when the signal voltage crosses the 0.5 volt level. This input structure generally requires a comparator.

3. Outputs: Outputting a voltage is largly a matter of making some voltage in the module available to the outside without disturbing the inside. This is of course, a buffer. In addition, it is often useful to have an output with certain "standard" properties, e.g., a standard 1k output impedence.

4. Individual Panel Controls: These are controls that are connected to the insides of the functional block directly, not through another input. An example might be a High/Low range switch.

We are thus left to consider shortly the details of the input structures and output structures. We shall see that there are logical and standard ways of doing these, but the actual combination of structures required in a given module will require individual attention. This is no small point, because the interface between modules is also the interface between the electronics and the musician, at least as far as the general type of synthesis process is concerned. Setting up this interface properly is one of the two main ways (the other being the controller) by which a musician makes the electronics make music. It is thus the interface that "shows" and this must be carefully considered. The main considerations are:

1. The exact combination of structures required: For example, a VCF needs a signal input, at least two control inputs, and one or more outputs as a minimum. A noise source on the other hand has no inputs at all (by its very nature) and the only interface structure other than an output or two would be some sort of control that controls the "character" of the noise. This requirement thus has a natural (what can be done), and a use (what does the specific user want to do in this specific case) aspect.

2. The panel layout: This involves the questions: What is a logical panel layout from the designers point of view? What is a logical layout from the musicians point of view? When these two layouts differ, is the musician better served by being forced to use the designer's choice?

3. <u>The Patching Method</u>: Plugs and jacks? Matrix switch? What?

These are all things that are important to musicians. Often the musician will not be in a position to appreciate the great difficulty you went to to develope a stable VCO. He will of course "appreciate" the "value" of a VCO that drifts badly. Suppose however that you do give him a good one; how is he to know that it is even possible to make a VCO that does drift? With the interface however, the musician will appreciate a set of controls and jacks that are conveniently organized, and which allow him to set up a desired patch without elaborate contrivances.

For the exact combination of interface structures required, it is suggested that the individual unit be considered carefully to see if all useful manipulations have been implemented. Secondly, all standard or expected patching systems should be considered. Thirdly, more structures than this minimum can be added as cost and panel space will allow.

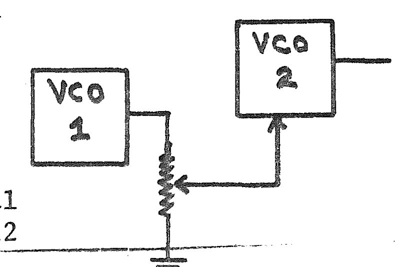
The question of panel layout is one of upmost importance. There are seldom any critical circuit layouts that dictate a certain panel layout, so the designer is in general free to choose the most logical setup. It is suggested that the designer choose a panel layout according to what he knows about the logical organization of the functional block. If properly done, this layout should be so logical that it would not be necessary for the designer to even label the jacks and knobs [of course he will do so for others!]. Using what seems like a logical layout to the designer who designed the actual "guts" of the thing has the advantage that the actual device and its capabilities will be more naturally represented, and this will be passed on to the user. None the less, panel layouts from other synthesizer modules should be studied, as there may be some sort of standard that should be used, or the designer may find that someone else's conception of the logic is better.

The actual patching system will depend a lot on the eventual use. For most purposes, standard 1/4" phone plugs and jacks are recommended. Many users of this book will be working on the expansion of existing systems, and these plugs and jacks are the most common for interconnections. Aside from that, plugs and jacks are the most conceptually simple patching system, and the most reliable system. Where panel space must be held to a minimum, smaller jacks or matrix switches can be used.

<u>VOLTAGE INPUT STRUCTURES</u>

The voltage inputting structure that we want to employ is the op-amp summing node. This structure is useful for both control inputs and signal inputs. The standard inputting resistor is 100k ohms.

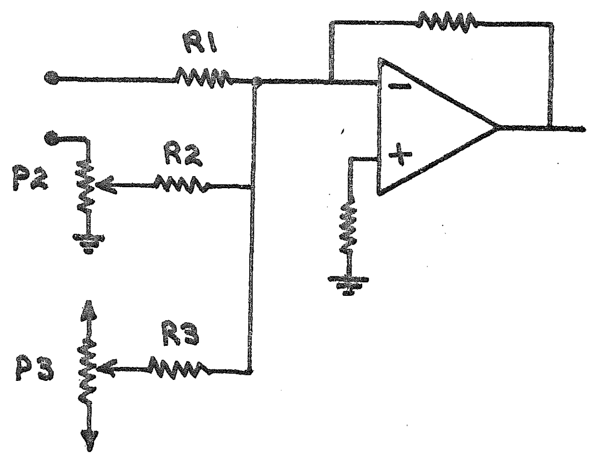
What we are actually considering is the transfer of one voltage to the input of a module. In many cases, we want this transfer to be incomplete - that is, we want to attenuate the voltage, and use only part of the available input. Thus, we must make some provision for attenuating the signal somewhere in the path. First of all, we want to make the case for the use of input attenuation. The device we will use to attenuate a voltage is almost always a variable resistor (a potentiometer). In the system shown at the right, assume that the output of VCO-1 is a ±5 volt sine wave (a peak-to-peak voltage of 10 volts). If we connect this to a standard VCO control input (to VCO-2), this would produce a swing of 10 octaves. Suppose we want only a small degree of frequency modulation. We need to reduce the signal from a 5 volt amplitude to a smaller value (e.g., 1/12



of a volt for a swing of ±1 semitone). We can simply insert a pot in the transfer line to form an adjustable voltage divider.

It is clear that this pot is a useful control, and we want to make it a permanent part of the system. The question then comes up as to where the pot should go. Should it be a part of the output of VCO-1, or part of the input of VCO-2. Well, we shall see that it should be part of the input structure of the driven module, not part of the output of the driving module. The basic reason for this is that the voltage level should be adjusted at the point of use. Suppose you want to use the output of VCO-1 to control two different VCO's. You want one to have a semitone swing of FM, and the second to have 3 semitones of swing. Once you have cut the voltage to 1/12 volt for the first, there is no way you are going to get the 3/12 volt for the second. Thus, the input attenuator allows each use to select whatever part of the maximum signal is required for the individual application, and still allows the voltage to be available at full value for other applications.

The summing node input structure has been studied in the chapter on the use of op-amps. The important consideration here is that the summing node is the - input of an op-amp that has negative feedback. The summing node remains at ground potential as long as the + input is grounded. Any excess current or current deficit into or out of the summing node is made up be the current through the feedback resistor. The overall device is an inverting, weighted summer.



We can now consider a typical input circuit as shown at the right. The input connected to R_1 is a standard 1 volt/octave control input. Generally, R_1 is 100k. This input would handle the main control voltage from the main controller of the system. For proper matching of such inputs as used on VCO's and VCF's, this type of resistor should be a 0.1% tolerance type - if there is more than one of this type on the module. Note that this input is used without attenuation. The input supplied through R_2 does have an attenuator on it. If the resistor R_2 is 100k, then the attenuator adjusts the response from 1 volt/octave down to zero. This is how microtonal scales could be generated. This type of input is also useful for inputting envelopes or any general control signal. This is the type we had in mind for input attenuation. Note that while R_2 could be 100k, it can be any value in a much wider range. For example, if it were 33k, the response would range from 3 volts/octave down. There would be less resolution, but a greater range. It should be realized however that these inputs have nothing to do with the overall range of the actual functional block, except for the fact that they may reduce it. The important thing as far as the functional block is the sum of the control voltages at the output of the op-amp. The functional block inside responds to this voltage. If the functional block only responds to a control sum from say -3 volts to +5 volts, then the range is limited to 8 octaves. If the control sum can only swing from -1 volts to +10 for example, then the range is cut to 6 octaves, and the control sum above +5 is wasted. The best way to determine the full value of the control sum is to connect up only one input to the module through a 100k resistor, and vary the input voltage from +15 to -15 and find out what the useful range is. The designer can then use the op-amp summation techniques to determine how to set the input resistors so as to keep the control sum in a useful range.

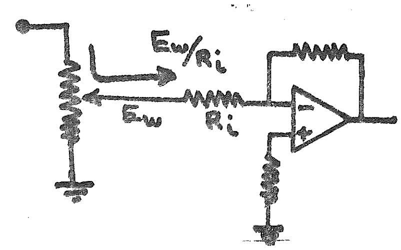
This leads to consideration of the third type of input resistor (R_3) which is a range control. For a coarse control, the value of R_3 should be selected so that it sweeps the full range of the control sum. It may be necessary to select a fixed resistor and connect it to either +15 or -15 to "center" this control. For a fine control, a value of R_3 much greater than 100k should be used. To sum this up, the following values are suggested:

Type of Input	Type of Control	Suggested Value
R_1	1 volt/octave standard	100k
R_2	Auxiliary input (e.g. envelopes)	33k to 100k
R_3	Coares	100k or lower
R_3	Fine	1.5 meg to 4.7 meg
R_3	Ultra Fine	15 meg to 22 meg

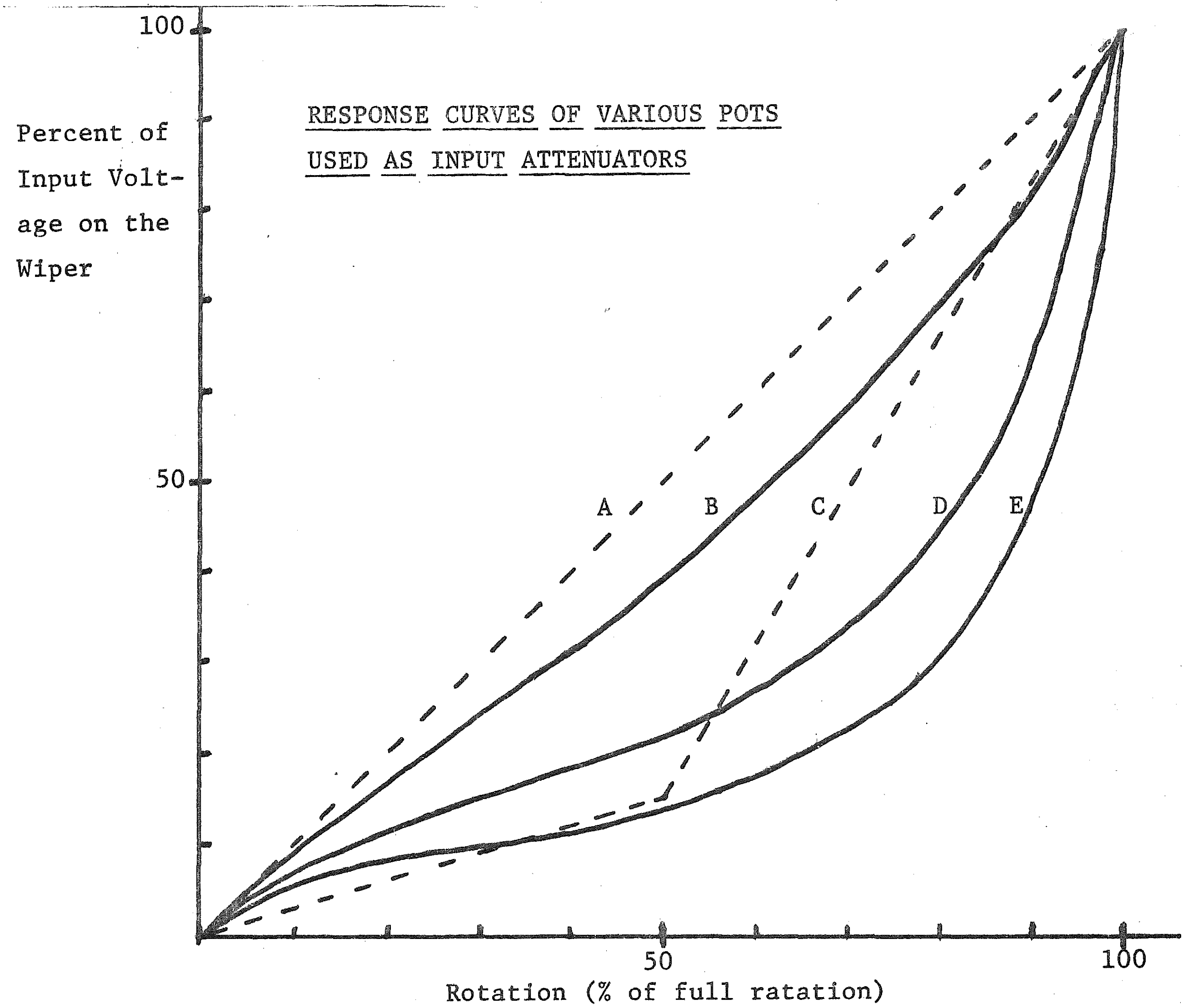
What we have said above was mainly directed at control inputs, but the summing node is also useful for signal inputs. Here, the R_2 type of input is generally used with R_2 = 100k, and set so that with the pot set to the top, the standard signal level enters the circuit with unity gain. The pot is thus strictly an attenuator. For the control inputs, for values of R_2 less than 100k, the summing node amplifier actually provides some gain (up to a gain of 3 for R_2=33k). This is useful for many control signals where the envelopes may be at a maximum level of 5 volts, and you may want more than 5 octaves of swing for a VCF for example.

It is now time to consider the pots. What should the value of the pots be? Should they be log or linear pots? First of all, consider that the pots are acting as voltage dividers in the type R_2 and R_3 input structures. Thus, it is natural to suppose that the pots should have a value much less than the attached R_2 or R_3 resistor. This is the standard consideration for a voltage divider. On the other hand, the voltage on the pot wiper has to go through the full range from one extreme to the other, so it is not necessarily required that the pots have low values. A log pot is one which has two linear sections. The sections are connected in the middle of the rotation. The log pot provides about 15% of its total resistance over the first half of the full rotation, and the remaining 85% on the second half. This type of pot is also called an audio pot. The reason for this latter name is that these are the types of controls that are used as "volume controls" and they are made that way since the amplitude response of the ear is logarithmic. Thus, there are certain applications where more resolution is required for small signals, and these are applications for log pots. None the less, the log pot still makes available all voltages throughout its range. Thus to a certain extent it can be said that the choice of pot is arbitrary. The considerations are the trade offs between input loading (for an R_2 type) and current consumed (for an R_3 type) on the one hand, and the degree of resolution and accuracy on the other hand.

We mentioned above that a pot is accurate as a voltage divider only when its value is much less than the resistor loading the wiper. The situation is illustrated at the right, where it is assumed that a 100k input resistance is attached to the wiper, and that the other side goes to a summing node (ground). The input structure is such that the voltage summed is always the wiper voltage, but in order to report the wiper voltage, some current must flow, and this steels some current from the voltage divider, and the



voltage on the wiper is different from what you would expect from the rotation of the pot shaft. A linear pot will not have a linear response. We show below calculated curves for five different cases. Curve A is the linear case, and would occur only where the wiper voltage is buffered, or where the pot is very small (10k or less). Curve B shows how the response changes when the pot is made 100k. Curve D shows the case for a 500k pot, and curve E shows a 1 meg pot. This warping of response is not as serious as one might expect - at least for the 100k case. Most all controls are set by ear, not by rotation position. The question is not one of accuracy, but of resolution. Note for example that a linear 500k pot has a curve that is more like a log pot (curve C). Curve C is the ideal log pot case. Higher resistance log pots would of coures show warping from curve C just as linear pots warp away from curve A. Note that curve E might be awkward in that it has a near linear region for very small rotation, a fairly flat region (10% to 50%), and then a very steep rise.



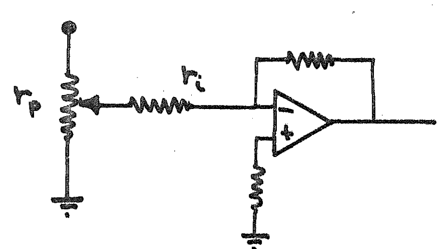
A.- Linear Curve: Applies to $r_i >> r_p$, or for the case where the wiper voltage is buffered.

B.- Linear 100k pot [r_i=100k, r_p= 100k]

C.- Log Pot for $r_i >> r_p$.

D.- Linear 500k pot [r_i=100k, r_p=500k]

E.- Linear 1 Meg pot [r_i=100k, r_p=1 Meg]

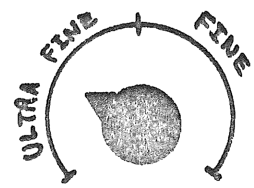


The following values and types of pots are suggested:

Type of Input	Type of Pot	Suggested Value†
Auxiliary Input (P_2)	Log*	100k
Coarse Control (P_3)	Linear	100k or 50k
Fine Control (P_3)	Linear of Log**	100k
Ultra Fine (P_3)	Linear	1 Meg or 100k
Audio (P_2)	Log	100k

* A log pot is selected for auxiliary control since resolution of small control voltages is often necessary, and many of the parameters that are controlled (amplitude and frequency) are subjectively logarithmic.

** A log pot is often a good choice for a fine control. The reason for this comes when you consider that before you adjust the coarse control, you usually center the fine control, or else you won't have range on either side to fine tune. With a log pot, you can set this initial position in the center of either of the two linear regions as shown at the right. In this way, you can have two different degrees of fine tuning that differ by about 10:1 in response. With careful selection of the R_2 type resistor to go with this, a very useful control is obtained.

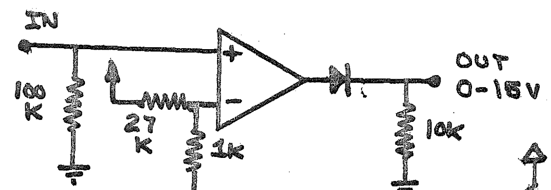


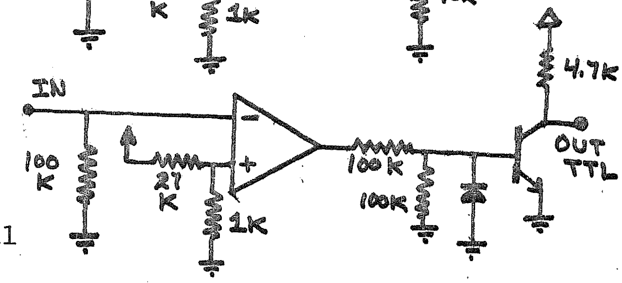
† Mostly 100k values are selected. Anything from 50k to 100k should be fine. While there is some warping of the rotation/wiper voltage curve with 100k, the main reason for prefering this value is that it causes only a 1% load on the standard output structures we will be using.

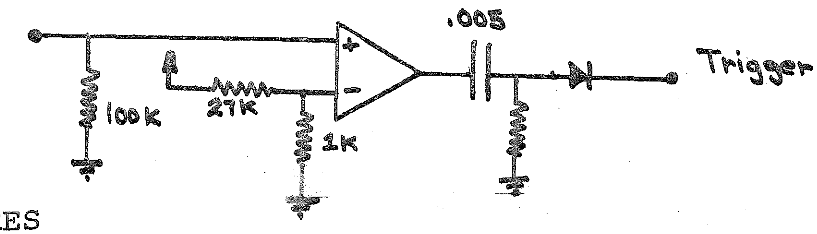
TIMING SIGNAL INPUT STRUCTURES

Timing signals are such voltages as gates and triggers which tell modules when to start and stop their functions. The most common of these signals are the gate and trigger signals from a keyboard that normally control one or more envelope generators. It often seems like these could be just permanent behind-the-panel connections, but there are cases where it is very useful to break into these lines. For example, the keyboard trigger can be brought out, fed to the input of a sharply tuned filter and made to "ring" the filter. On the other end of the line, we often want to trigger an envelope generator from some source other than the keyboard. Thus, we consider the inputting and outputting of timing signals. The outputting of a timing signal is the same as any other voltage output and will be considered with the other outputs.

Inputting an "ordinary" voltage for use as a timing signal is a little trickier. The essential thing about a timing voltage is that it represents some event that is defined in time. We are thus left with the problem of representing a time point by a certain point on a waveform that may be quite different from the standard timing signals (sharp pulses and rectangular waveforms). The obvious device for translating a small transition into a sharply defined one is the comparator. For example, we can define a gate as occuring whenever the input voltage exceeds 0.5 volts. A trigger might be defined whenever a waveform crosses the 0.5 volt level from the negative side. For this purpose, the ordinary op-amp comparator or "weak" Schmitt trigger is usually sufficient. In the examples that follow, we will choose 0.5 volts as a reference level, but other levels can be chosen. [Ground level is not chosen because too many signals have an off state very close to zero, and erratic triggering could occur.] The first example at the right shows a comparator used to define a gate whenever the input goes over 0.5 volts. The input impedence is 100k, and the output gate is either zero or about +13 volts. This is a useful gate level for CMOS circuits run between +15 and ground. If a TTL version of this is needed, it is possible to reverse the input terminals of the comparator, and employ the standard analog to TTL stage, which is inverting. This circuit is shown at the right. These same two circuits will work for inputting trigger signals if the pulse shape is established outside. In some cases, it is useful to obtain a trigger signal from a slowly varying input signal. In such a case, the comparator output can be differentiated to give a short pulse, and the unwanted transition can be blocked with a diode. The circuit below gives a positive trigger whenever the input crosses 0.5 volts from the negative side.





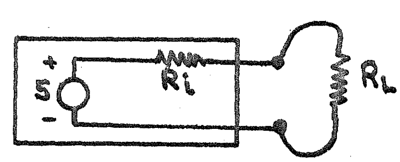


OUTPUT STRUCTURES

The standard output structure for electronic music modules is a voltage output with a 1000 ohm output impedence. This structure starts with a low output impedence voltage source such as the output of an op-amp, FET source follower, or emitter follower, and 1000 ohms of series resistance is added.

Since it is the output voltage that is used, we must ask how the module being driven will change the open circuit properties of the module that is the source. We have seen above that the standard input structure has an input impedence on the order of 50k to 100k. Thus, we can see that the load can be expected to be a few percent of teh output impedence.

First it is necessary to take a look at the meaning of output impedence. Suppose you have a black box with two output terminals and you measure the output voltage with a device that has a very high input impedence (scope or FET input voltmeter for example). You find the voltage is 5 volts, and you then connect a 5k resistor across the output terminals and expect a current of 5/5000 amps (1 ma) to flow. Actually, you will find a lower value of current flowing. This is because the voltage source has some impedence or its own. It is the usual practice to consider this to be a resistance R_i in series with an (Imaginary) perfect voltage source. This internal resistance can be very important in many practical problems. A familiar example is the ordinary flashlight battery which may measure a full 1.5 volts, and yet fail completely to light a bulb due to large increases of internal resistance as the cell ages. Suppose the black box we are considering is a perfect 5 volt source with a 1k series resistor R_i as shown at the right.



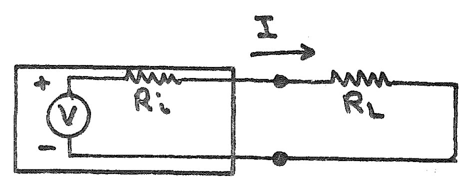
Two things become apparent when we look at the circuit as a series resistance problem and just apply Ohm's law. The current is actually 0.83 ma and the voltage on the output terminals drops to 4.17 volts, down 17% from the open circuit case. This is called loading. What happens when we consider other loads?

When we put a 100k resistor on the output terminals, we get a current of 0.0495 ma, for an output voltage that has dropped to 4.95 volts, or about a 1% drop. This is not too serious in most cases. Another interesting case is the load of 1k. In this case the voltage drops to 2.5 volts, exactly half the open circuit voltage. This is a common way of measuring output impedence - you just load down the circuit until the voltage drops to half its unloaded value. The load resistor and the output impedence then have the same value. It is nearly as easy, and perhaps easier on the circuitry, to just load it down by 10 or 20 percent, and use a standard series resistance calculation using Ohm's law.

As a variation, consider a vary high output impedence (say 10 Megohms) and use the device to drive a series of loads from zero ohms to 100k. For zero ohms, the current that flows is 0.5 amps, and the voltage is zero (short circuit). For a load of 100k, the current is 0.495 amps, and the voltage is still very small compared to 5 volts (0.0495 volts). Thus, the current is very nearly constant for a wide range of loads, and the voltage is very low. This is the simplest example of a current source.

Thus, a low output impedence (relative to expected loads) is necessary for a voltage source. A high output impedence (again relative to expected loads) is necessary for a current source. The case inbetween, where the two are of similar magnitude should be avoided.

Since we are interested in inputting voltages for signals and controls, we have to be able to output voltages, and thus the voltage source (low output impedence) is used. Note that it is not necessary (nor is it desirable) to have input and output impedences matched. It is necessary to match impedences only when power is to be transferred, or for preventing reflections on transmission lines. We are transferring voltage, and we want to do it from a low output impedence to a high input impedence. This mismatch is desirable, but note that the other one (the current source) isn't. Transfer of power is the middle case. It is perhaps instructive here to show that the transfer of power is most efficient when impedences are matched. To do this, consider the circuit at the right. The power developed in the load is:



$$P_L = I^2 R_L = \frac{V^2}{(R_L + R_i)^2} R_L$$

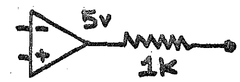
This expression can be differentiated with respect to R_L to give:

$$\frac{dP_L}{dR_L} = \frac{V^2(R_i^2 + 2R_iR_L + R_L^2) - V^2R_L(2R_i + 2R_L)}{(R_i^2 + 2R_iR_L + R_L^2)^2}$$

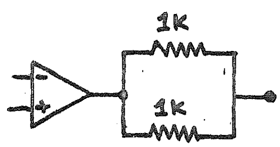
Setting this equal to zero to get the minimum gives: $R_i = R_L$.

Getting back now to the low output impedence, it might seem like to very low (few ohms) output impedences of op-amps, emitter followers, or source followers would be the best we could do. However, there are several advantages to adding a 1000 ohm series resistance, and few drawbacks. For one thing, it protects both the output and the inputs by limiting current in the event that something goes wrong. Op-amp outputs can usually stand at least momentary shorts, but it is not good practice to take advantage of this. It is virtually impossible to plug in any plug and not short the live contact to ground in the process. The 1k output resistor holds this in check. We shall see shortly that the standard output also allows some degree of mixing of signals. The errors that are caused by this internal resistance are as we have shown above, typically a few percent.

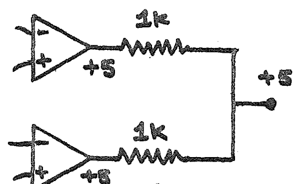
To show how output mixing can be achieved, consider the typical output stage (a) shown at the right. The output voltage is 5 volts, and the output impedence is 1k. If we now parallel the 1k output with a second 1k resistor as shown in (b) below, nothing changes, except the output impedence drops to 500 ohms. Nothing additional happens if we drive the second resistor with its own op-amp as in (c) below, as long as both voltages are the same. The mixing effect comes about when the voltages are different. A current now flows between the two op-amp outputs. Since the two resistors are the same value, and form a voltage divider, the voltage at the output is the algebraic average of the two op-amp outputs. Thus we have a simple 1:1 mixer (d).



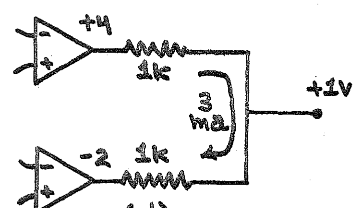
(a)



(b)



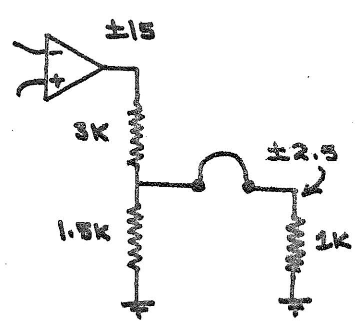
(c)



(d)

If three or more jacks are shorted, the average of all signals shorted is obtained. The main drawbacks of this method are: 1) The output impedence of the combination of n inputs drops to 1k/n. 2) The mixing is always at equal levels: 50%-50%, 33%-33%-33%, etc, and you can't get something like 70%-30%.

Often, we have signals that are at voltages that are higher than the level at which we want to output them. For example, when waveshaping a smooth signal into a pulse, a comparator is usually used, and this output swings between ±15 and we may want the output to be only ±5. To reduce this, we select the appropritae resistance ratio (2:1 in this case), and select actual resistances so that if the two resistors were placed in parallel, the resulting resistance would be 1000 ohms. For our example, we find that this requires that the resistors be 3k and 1.5k. The circuit is shown at the right. To demonstrate that the output impedence of this voltage divider is in fact 1000 ohms, we have shown it with a 1k load. The output voltage with a 1k load is 2.5 volts, which is half the open circuit voltage of 5 volts, and as we demonstrated above, this is the condition for the output impedence being equal to the load.



The 1k output resistors (or the equivalent voltage dividers) are generally used on all front panel output jacks. One exception to this is the main control voltage from the main controller of the system. This control voltage is standardized to be applied to a 100k summing node for a one-volt/octave response. Some sort of output jack for this main control voltage is usually necessary for interfacing with other systems even where the main distribution of this voltage to modules is by switches. If a 1k resistor were used here, it would cause the effective input impedence to the summing node to be 101k, a 1% error which would be very serious for pitch control.

CHAPTER 5B

VOLTAGE-CONTROLLED OSCILLATOR DESIGN

CONTENTS:

INTRODUCTORY NOTES:

This chapter is quite important, and is fairly difficult. This chapter and the three that follow it contain the majority of the standard techniques that are used in the more or less standard electronic music modules: VCO's, VCA's, VCF's, and Envelope Generators. The actual introduction to this chapter that follows is three pages long, and is actually an overview of the full chapter. Many readers may wish to read only the introduction carefully, skim the rest of the chapter, and return to the exact details of this chapter after going over Chapters 5c, 5d, and 5e. It will be found that many of the ideas will reappear elsewhere, and it is perhaps a good idea to have the proper overview before digging into the major part of this chapter.

INTRODUCTION:

Electronic music VCO design presents one of the toughest problems for musical engineers. The reason for this is that the human ear is extremely sensitive to changes in pitch, and many compositional processes (such as multi-tracking) further bring out errors in pitch relationships.

It should be realized that there is a world of difference between the requirements of a simple system with one VCO, and the requirements of a professional system with several (or many) VCO's. If the VCO is being designed for a simple system, a relaxed design will suffice providing it has reasonably accurate exponential response. On the other hand, a professional system requires that VCO's maintain accuracy with respect to each other, and with respect to time. In professional systems, the VCO's are often used in parallel, and must track accurately to maintain a desired tone color. Also, professional recordings are often made by "multi-tracking." This means that one part of a composition is recorded on one tape track, the tape is backed up, and the second part is superimposed on the first. It is therefore important that the oscillators not drift during the time it takes to put down one part and the time when the second track is superimposed. The same is important for live performance when it is necessary for a synthesizer to stay tuned to other live instruments. Drift of an oscillator is caused by the variation of one of the electrical parameters of the system. This is largely a matter of changes in a components electrical characteristics with temperature.

I would rate the important factors of VCO design in the following order:

1. Exponential response (frequency is an exponential function of the control voltage).
2. Thermal stability. (Low drift)
3. Accuracy of tuning. (precise exponential response).
4. Wide range. (full range voltage swept, not switched).

Other important additional features are:

1. Standard input and output structures.
2. Control inputs with variable attenuation.
3. Gating and syncing features.
4. Different types of frequency modulation inputs.

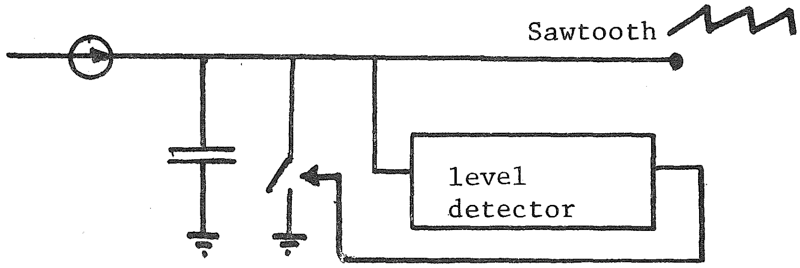
Less important points are:

1. A wide variety of waveshapes.
2. Highly precise waveshapes.

The basic design procedure for a VCO is as follows:

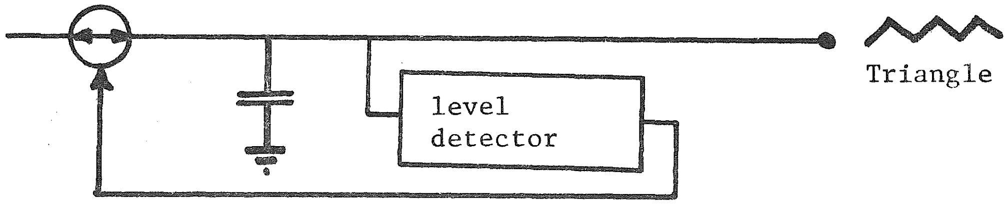
1. Selection of the basic oscillator circuit (current controlled).
2. Selection of an exponential current source with control inputs.
3. Connecting the oscillator and exponential current source and compensating either or both for proper tracking, accuracy, and stability.
4. Waveshaping the basic waveform into a larger set.

THE BASIC OSCILLATOR: In selecting a basic oscillator circuit, the musical engineer will generally select either a sawtooth or a triangle generator. A sawtooth generator is outlined below:



When the voltage reaches a preset level, the level detector closes the switch rapidly discharging the capacitor.

Alternatively, the current source can be periodically reversed (in response to the level detector) to give a triangle:



The level detector and reset switch (or current reversal) must be as fast as possible. Any reset time delay here will cause the oscillator to go flat on the high end, and the amplitude level may overshoot.

The triangle or sawtooth are generally selected because they are linear ramps, and maintain a linear relationship between control current and frequency. This is because the linear ramps (dv/dt is linear with current) spend all their time changing at a constant rate, and the limits on the waveform voltage are set by the level detector. Furthermore, either of these waveforms can be easily wave-shaped into the other common basic waveforms. It would be most difficult to start with a square wave for example and obtain other waveforms with a constant amplitude level. Some oscillators do in fact start with a sine wave output, but this is generally used in a configuration that allows for a quadrature (both sine and cosine) output, and not as a general purpose VCO.

THE EXPONENTIAL CURRENT SOURCE: The only exponential current source in general use is based on the fact that the ordinary transistor has an exponential collector current relative to linear changes of base-to-emitter voltage. [digital exponential generators offer an attractive alternative for the future]. This provides a simple solution, and a very wide range of the control parameter as compared to what could be obtained with voltage control. Control ranges of seven or more decades are possible with this sort of current control. If the same seven decade range were to be controlled with voltage, and the upper voltage were set at 10 volts, the low end control voltage would have to be 1 μV. This would present a substantial problem to say the least. The transistor does this control function naturally.

The transistor current control has a major drawback in the area where it does the most damage - it is extremely sensitive to temperature. Left uncompensated, an exponential current source formed from a single transistor could easily shift the VCO frequency several octaves as the circuit warmed from an outside winter temperature to room temperature. Thus, the engineer must use matched pairs of transistors to zero out as much of the temperature dependence as possible, and then use a temperature compensating gain stage (temperature compensating resistor) or a thermostating method.

CONNECTING THE BASIC OSCILLATOR AND CURRENT SOURCE: Once a basic oscillator and current source are selected, the two must be made to work properly together. This is generally no problem over the middle range of operation. On the low end, leakage current will determine the lowest frequency. The bottom of the exponential range is determined by the transistor's saturation current. It may be necessary to adjust the VCO timing capacitor. On the high end, two factors combine to make the VCO go flat (lower than the expected pitch). The factors are the reset error discussed above, and the base-to-emitter "bulk" resistance of the exponential converting transistor.

High end compensation techniques for the exponential converter can be employed to compensate for the bulk resistance. Special techniques are possible to compensate for reset error, or in many cases it is possible to just overcompensate the exponential stage to take care of the reset error too. In the latter case, a certain amount of matching of breakpoints in the response curves is necessary.

WAVESHAPING: The final step is waveshaping into a more or less standard set of waveforms that may include: sawtooth, inverse sawtooth, triangle, square, pulse, and sine. All of the required waveshaping methods are fairly simple and sufficiently accurate for musical purposes. One possible exception is the sine. There is no simple exact method or doing this conversion, and what is generally used is a method of rounding the triangle, and this generally gives at least 1% harmonic distortion. This may not be at all objectionable, but it is audible. A purer sinewave can be obtained by following the rounded triangle with a tracking low-pass VCF.

Fortunately, precise waveshape is usually not required as either the ear can not hear the difference, or there is simply no reason to prefer the harmonic content of a perfect sawtooth for example, to a sawtooth with some imperfect features. A possible exception to this rule comes about when the VCO is used to provide low frequency control signals, in which case waveform faults may become audible as they appear in the controlled parameter. However, this depends on the individual case, and no general rule can be made.

DESIGN OF THE BASIC OSCILLATOR:

The basic oscillator consists of a capacitor, a current source, a level detector, and a resetting mechanism. The waveform that is used is the voltage on the capacitor which is charged by the current source. The linearity of the frequency with current is implied by:

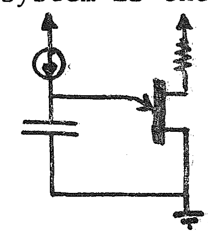
$$I = dq/dt = C\,dv/dt, \quad \text{or } dv/dt = I/C \quad \text{volts per second.}$$

The voltage swing of one cycle of the oscillator depends on the waveform. It is the peak-to-peak voltage for a sawtooth waveform and twice the peak-to-peak for a triangle. If we denote the voltage swing for one cycle as V_{1cycle}, then the frequency becomes (neglecting reset time):

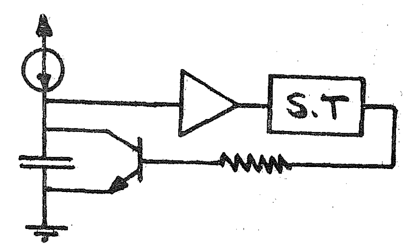
$$f = \frac{dv/dt}{V_{1cycle}}, \quad \text{where the units are:} \quad \frac{\text{volts/sec}}{\text{volts/cycle}} = \frac{\text{cycles}}{\text{sec}} = \text{Hz}$$

With these equations, the full frequency range can be calculated from the full current range. If this range is not satisfactory, either the current range or the capacitor can be changed. Note however that the decision is not arbitrary. First, since we are employing exponential current sources, we prefer to operate in a current range where the collector currents have the best exponential accuracy. Secondly, in general we would like to keep the capacitor small ($\sim$1000 pf) in the case of a sawtooth generator, since we have to discharge it rapidly. Keeping the capacitor small assures that the resistance of the discharge switch will not unduly slow the reset process. In the triangle oscillator, since the current is reversed, the small capacitor size is not so important, and the current range of the exponential source can be optimized. In either case, the capacitor should be polyethylene or polystyrene for the best results, and the voltage on the capacitor should be well buffered with a FET or a FET input op-amp. The bias current drawn by the buffer must be very small compared with the lowest exponential current to be used.

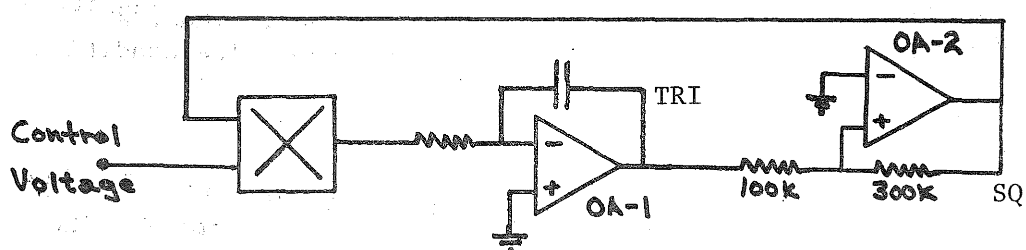
The rest of the circuit is a level detector of some sort connected to a reset switch or current reversing switch. A simple system is the standard relaxation oscillator formed from a UJT which serves both as a level detector and a switch. The basic circuit is shown at the right. The main drawback of this simple system is that the frequency range is limited due to the leakage current on the low end, and the fact that it will not discharge completely if the current is too large.



A good low leakage switch is the ordinary BJT switching transistor. This can be combined with a Schmitt trigger (and possibly a monostable) to discharge the capacitor. The Schmitt trigger should be either a high speed logic type, or one formed from discrete transistors. A block diagram is indicated at the right. Op-amp Schmitt triggers are not satisfactory since they are much too slow at the higher frequencies, Other types of switches include various types of switching FET's, and CMOS analog switches. Note that usually along with the sawtooth wave, a pulse waveform is available in the circuit and can be brought out.



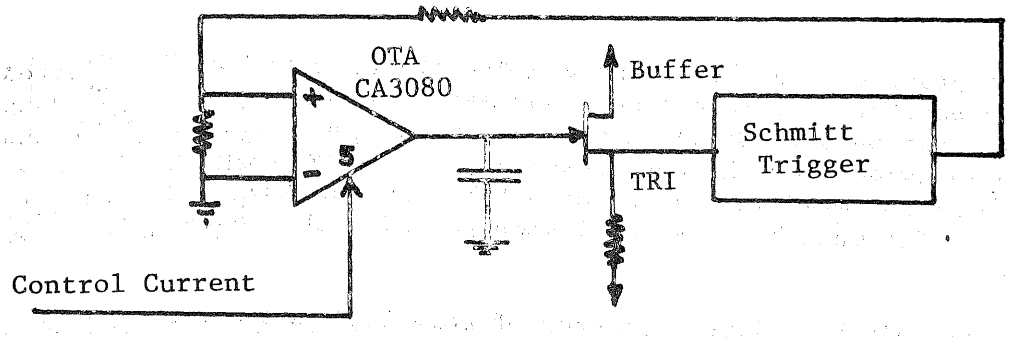
The basic idea of the triangle generator is illustrated below for a circuit which demonstrates the principle, but which is not satisfactory for wide range VCO's due to the limited dynamic range of the multiplier:



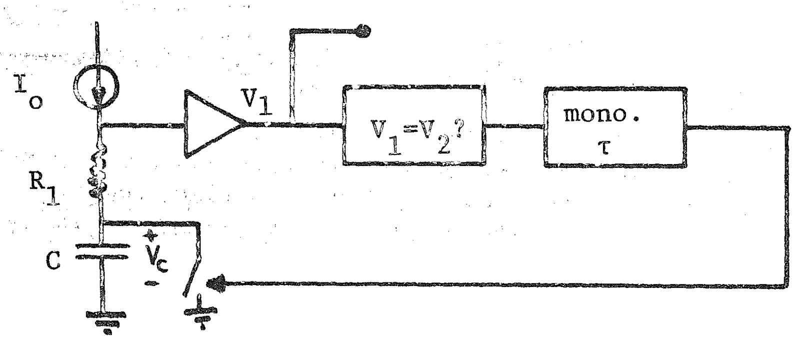
Triangle-Square Wave VCO

Here, the current is reversed by the Schmitt trigger (OA-2) and the magnitude of the current into integrator (OA-1) is controlled by the multiplier.

Since we are working in general with exponential current sources, what we need is a device that selects either the current itself, or a current of equal magnitude but opposite direction. The OTA is useful for this purpose. If the input voltage to the OTA is greater than about 100 mv, the input stage is saturated, and the output current will be of the same magnitude as the bias current, and the direction will depend on the polarity of the differential input voltage. [to understand this, consult the next chapter on VCA's to see how a differential input stage saturates, and apply this to the current mirror structure of the OTA]. The basic current controlled triangle oscillator is shown below:



As we mentioned, a reset error will cause the VCO to go flat on the high end. The first approach to avoiding this problem is to make the discharge time as short as possible. This involves a fast switch, a small capacitor, and a low resistance switch in the case of the sawtooth oscillator. A second approach to reset error correction has been described by Franco [Sergio Franco, Hardware Design of a Real-Time Musical System, U. of Ill. Dept. of Computer Science report UIUCDCS-R-74-677, Oct. 1974]. The basic idea is to swamp any actual reset error with a much larger "dead time" and then compensate for the dead time. The principle is illustrated by the circuit below:



A series resistor R_1 has been inserted in the current line. Instead of buffering the capacitor voltage, it is the voltage across the resistor-capacitor series connection that is monitored and fed to the level detector. Thus, a voltage $R_1 I_o$ has been effectively added to the capacitor voltage. The condition for the start of the discharge cycle is thus:

$$V_1 = V_c + I_o R_1 = V_2$$

At the trigger point, assume the capacitor has been charging for a time T. The capacitor voltage is therefore:

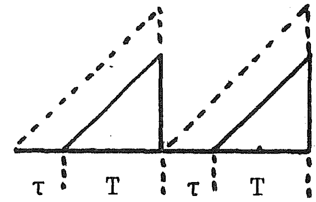
$$V_c = I_o T/C$$

After the trigger point is reached, the capacitor is discharged and held at zero for a time τ. The time for a full cycle is therefore $T + \tau$ and the frequency is given by:

$$1/f = T + \tau = V_c C/I_o + \tau = V_2 C/I_o - R_1 C + \tau$$

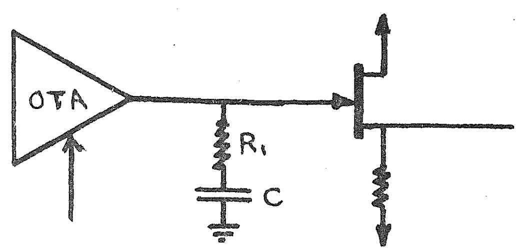
If we now make $R_1 C = \tau$, then the frequency is independent of τ, and depends only on I_o (C and the trigger level V_2 being fixed).

The time τ must be shorter than the period of the highest frequency we need. Also, note that some alteration of the sawtooth occurs as indicated at the right. This is because the sawtooth is held down for a time τ. This may or may not be serious depending on the appliaction. If the VCO is being used mainly for its pulse output (the output of the monostable), the distortion of the sawtooth is of no importance. It is also difficult to argue that the distortion of sawtooth shape is very important if it is held to some small value, perhaps 10% or so. While the distorted sawtooth has a different tone color, it may be just as useful as the standard sawtooth. There is a corresponding decrease in amplitude that comes with the distortion of the waveshape, but if this is not large, the ear can not detect it. Where this distortion would come in is where the sawtooth were ultimately shaped into a sine.



Above, we have assumed that the reset time τ has been set artificially long. This was to swamp any variation in the actual minimum reset time that occurs in the diacharge process. What happens if we try to apply this method to correct for an existing small reset error. First of all, τ may not be constant enough. Secondly, the reset error may be small enough that it can be corrected for incidentally to the correction for the bulk resistance error that occurs in the exponential converter (to be discussed). Otherwise, there is no reason why the correction scheme won't work. We can therefore consider extending Franco's method to the triangle oscillator as well as the sawtooth oscillator.

First, we ask how the reset error in the triangle generator manifests itself. If there is a time delay in reversing the current, the amplitude of the triangle overshoots, and when it does reverse, it has further to go back down. It is perhaps most useful here to just consider what happens if we insert the series resistor as in the sawtooth case:



If the delay in reversing the current is t', then the amplitude overshoot is:

$$I_o t'/C$$

With R_1 added to the circuit, the trigger points are effectively lowered by an amount $R_1 I_o$, and if t' is set to $R_1 C$ as was done in the sawtooth case, the trigger points have been effectively lowered by $R_1 I_o = I_o t'/C$, which is the same as the amplitude overshoot. Here, we are simply dealing with the existing delay, and nothing artificially long has been added. There is no distortion of the triangle wave shape, although its amplitude is decreased.

SYNCING OF OSCILLATORS

It is of course useful to have oscillators track each other as accurately as possible, and this is the purpose of careful design. However, there is a limit to the accuracy that can be achieved. For exact tracking, some form of syncing is necessary. Before discussing syncing methods, it should be pointed out that while it should be possible to have good or exact tracking in a synthesis system, it is not necessarily so that the more useful sounds are achieved with better tracking. Some very interesting sounds are achieved with imperfect tracking. Thus, accurate tracking is something that we should have when and if we choose to use it, but is not to be considered something that always leads to better sounds. Consider for example, that locking oscillators at small integer ratios may just produce additional harmonics of the master oscillator. This same thing could be done by just using a different waveform. Thus, syncing as a useful tool begins when we consider oscillators synced at frequencies between the harmonics, e.g., in the first octave above the master oscillator as would be done when producing chords.

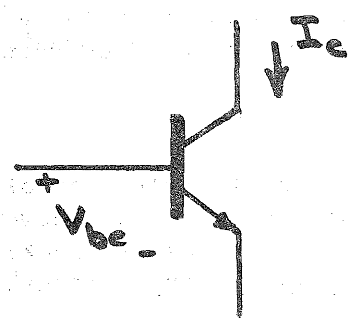
Syncing methods can be divided into two classes. The first class includes those methods that provide trigger signals which force the oscillators that are being controlled to alter their normal response. The second class is the class of phase locking devices. These methods are inherently inexact - they are always searching for the correct frequency, wondering around it according to the parameters of the phase locked loop.

In the first class, syncing with triggers can be achieved by designating one oscillator as the master (usually the lowest), and using its reset trigger to reset all the "slave" oscillators. This is called "Hard Sync." A second method is "Soft Sync," which is achieved by having a common sync terminal on all oscillators. No oscillator rules. What happens is that when any oscillator resets, it will reset all oscillators that are within a certain fraction of their normal reset points. Thus, certain oscillators will reset a little early. Generally, the factor is about 5%. This is accomplished by feeding all reset pulses to a common bus, attenuating the pulses to about 5%, and coupling this so that it drops the reference level of the oscillators. A third method belonging to this first class is the gated oscillator method. Here, we work with longer rectangular signals, not just short pulses. These longer signals gate the oscillator on and off. Thus, for example, when one oscillator has a square wave output going high, it may turn on another oscillator. The two are thus synced at the point where the square wave rises.

In the second class, phase locking can be achieved either as a part of the VCO design, or as an externally patched loop. This method is based on standard phase locking principles, and operates with the master-slave principle. The slave oscillator is forced to track the master by means of a correction voltage around the loop.

EXPONENTIAL CURRENT STAGES

The basic exponential converter stage (exponential current stage) is based on the fact that in an ordinary transistor, the collector current is an exponential function of the base-emitter voltage:



$$I_c = \alpha I_{ES}[e^{+qV_{be}/k_B T} - 1]$$

$$\simeq \alpha I_{ES} e^{+qV_{be}/k_B T}$$

for $V_{BE} >> k_B T/q$

where: α is forward current gain (≈ 1)

I_{ES} is the emitter saturation current (10^{-12} to 10^{-14} amps)

e is the base of the natural log: 2.71828.....

q is the charge of a single electron: 1.60219×10^{-19} Coulombs

V_{be} is the base to emitter voltage (in mV)

k_B is Boltzmann's constant (8.6167×10^{-5} electron-volts/°K)

T is the temperature in °K (Absolute or Kelvin temperature) 273 + °C

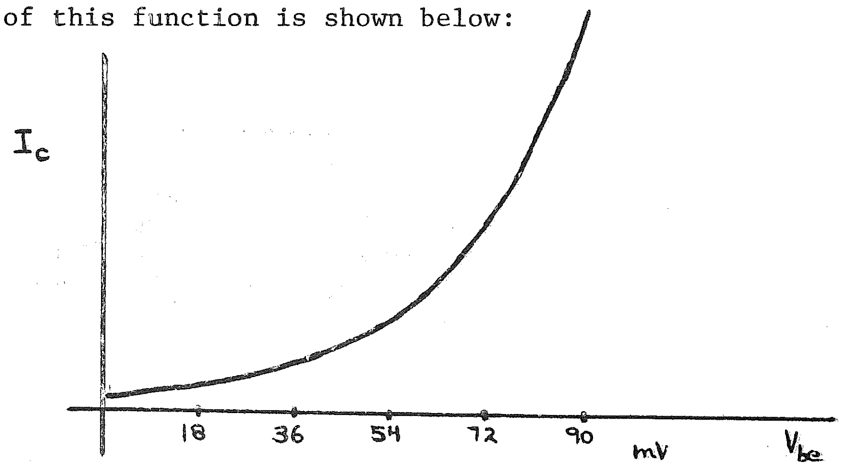
$k_B T/q$ is approximately 26 mV at room temperature (300°K)

The last entry ($k_B T/q \approx 26$ mV) makes the essential exponential equation:

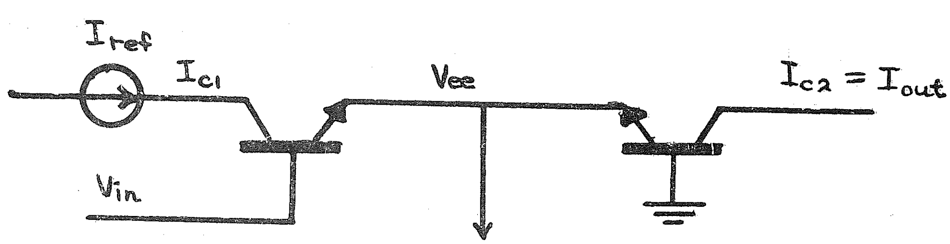
$$\boxed{I_c = \alpha I_{ES} e^{+V_{be}/26}}$$

V_{be} is in mV

A plot of this function is shown below:



The single transistor exponential converter is unsatisfactory in practical circuits due to the temperature dependence of the I_{ES} term which doubles with each 10°C change of temperature. Two methods of improving this are available. The first is simply holding the temperature of the transistor constant, but this is not particularly easy, and is generally used only in conjunction with the second method. The second method is to use a matched pair of transistors, which should be on the same semiconductor chip so that temperature tracking is optimal. The method is illustrated below:



The equations for the two transistors are:

$$I_{c_1} = \alpha I_{ES} e^{(V_{in}-V_{ee})/26} \qquad\qquad I_{c_2} = \alpha I_{ES} e^{-V_{ee}/26}$$

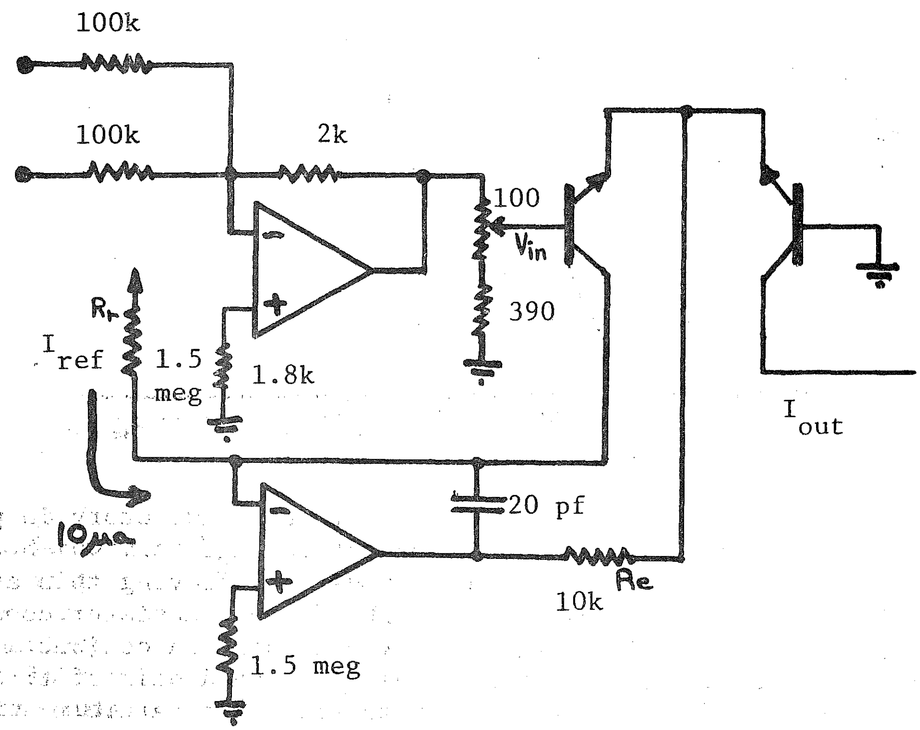
Combining these two equations gives:

$$I_{c_1} = I_{ref} = I_{c_2}\, e^{V_{ee}/26}$$

or: $$I_{out} = I_{c_2} = I_{ref} e^{-V_{in}/26}$$

Thus, for the matched pair, the I_{ES} term cancels out.

For a practical circuit, it is necessary to add a unit to scale and range of the input voltage, the constant I_{ref} current, and a means of collecting the excess current at the emitter junctions. This can be done with two op-amps as shown below:



The VCO's we will be working with have a one volt per octave response. Thus, the equation for the frequency is:

$$f = f_o 2^V = f_o e^{\ln 2 \cdot V}$$

The exponential function that the transistor gives naturally is $e^{V_{in}/26}$, so we have to do some scaling. The easiest way to see what we need is to observe that for a change of voltage that we will call v', we need a 2:1 change in output current:

$$\frac{e^{(V_{in}+v')/26}}{e^{V_{in}/26}} = 2 = e^{\ln 2}$$

This gives $V_{in} + v' = V_{in} + \ln 2 \times 26$ or $v' \approx 18$

This means that we will have to scale down our standard 1 volt/octave inputs to give 18 mV/octave at the base of the transistor that is doing the exponential conversion. Thus, the input voltage will have to be scaled down by something in excess of 50 to one. In the example current source above, this is accomplished by first attenuating the 100k inputs with a 2k resistor in the feedback loop of the op-amp. Next, a trimmer was inserted in a voltage divider to give the exact response required. Since the op-amp output can not exceed the limits of -15 to +15, V_{in} cannot exceed about ±270 mV. Thus the $\exp(V_{in}/26)$ term ranges from about exp(-10) to exp(+10) or from about 0.0000454 to about 22,000. The useful exponential range of the transistor is for currents from about 1 pa to 1 ma, a range of 10^6, so we have more control voltage swing than we really need for normal exponential converters. It remains therefore to set the value of the reference current for the proper range. When $V_{in} = 0$, the output current is the same as the reference current. We can thus expect to set this current in about the middle of the useful range (which would be at 1 μa). However, with the current source shown (op-amp summing node) we can see that 1 μa is a pretty small current unless the op-amp has a very low bias current. Thus it is common to select a higher value for the reference current from 10 μa to 100 μa which corresponds to values of R_r from 1.5 meg to 150k.

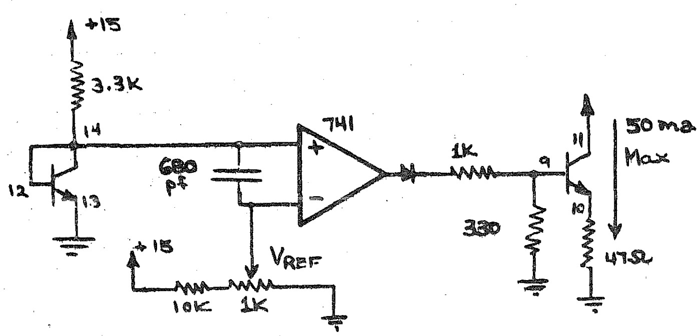
This explains the function of all the resistors except R_e. Observe that the output of OA-2 runs at a negative voltage to collect the excess current from the tied emitters. At the highest frequencies, I_{out} will be much greater than I_{ref}. This means that the current through R_e will be essentially the current I_{out}. The current through R_e is limited by the lowest voltage (-15) that the output of OA-2 can reach. Thus, R_e sets a maximum limit on I_{out}, restricting it to less than $15/R_e$. This is a useful way of limiting the current if it must be done, and not having to worry about how high the control sum (output of OA-1) can possibly go.

At this point, we should point out that while the I_{ES} temperature dependence has been cancelled out, there is still a temperature dependence in the output which we can see by looking at the complete form of the exponential expression:

$$I_c = I_{ref} e^{qV_{in}/k_B T}$$

The 1/T term in the exponent remains. If we are thermostating the transistor pair, this is under control, otherwise, we must compensate for this for the very best performance. Since our circuits operate around room temperature (300°K) the 1/T term is about one part in 300. The exponent thus changes by one part in 300 per °C change in temperature. To compensate for this, we can make V_{in} change by one part in 300 per °C as well. This is done with a temperature compensating resistor which changes its resistance upward by one part in 300 per °C, and placing this resistor in the feedback loop of the op-amp input summer. The usual resistor used here is the type Q81 from Tel Labs of Manchester, NH. The required temperature coefficient is +0.33%/°C, and these are usually found as what is nearly equivalent, 3500 ppm/°C. The resistor is placed in good thermal contact with the matched pair of transistors.

Thermostating of the transistor pair is a usable technique, but it does require a good deal of extra current from the power supply. This is because the semiconductor material must be held at a temperature well above room temperature, since there is no simple electrical means of cooling. Matched pairs of transistors can be obtained with their own internal heaters (Fairchild 726) or the designer can make his own using a transistor array chip with at least 4 transistors. Two are used as the matched pair, one as a temperature sensor, and one as a heater. The basic idea is shown by the circuit below:



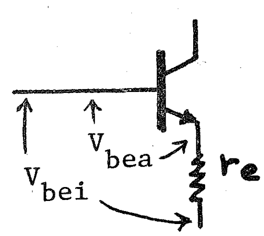
Pin numbers for CA3046

References, EN#33, EN#52

While this does work, it does require power, and does produce some heat that could find its way into other components and cause them to drift. To summarize the situation on temperature compensation, the I_{ES} dependence is removed by using a matched pair, thermostated or not makes little difference. The 1/T dependence of the exponent is compensated by a temperature compensating resistor or with a temperature controlled matched pair.

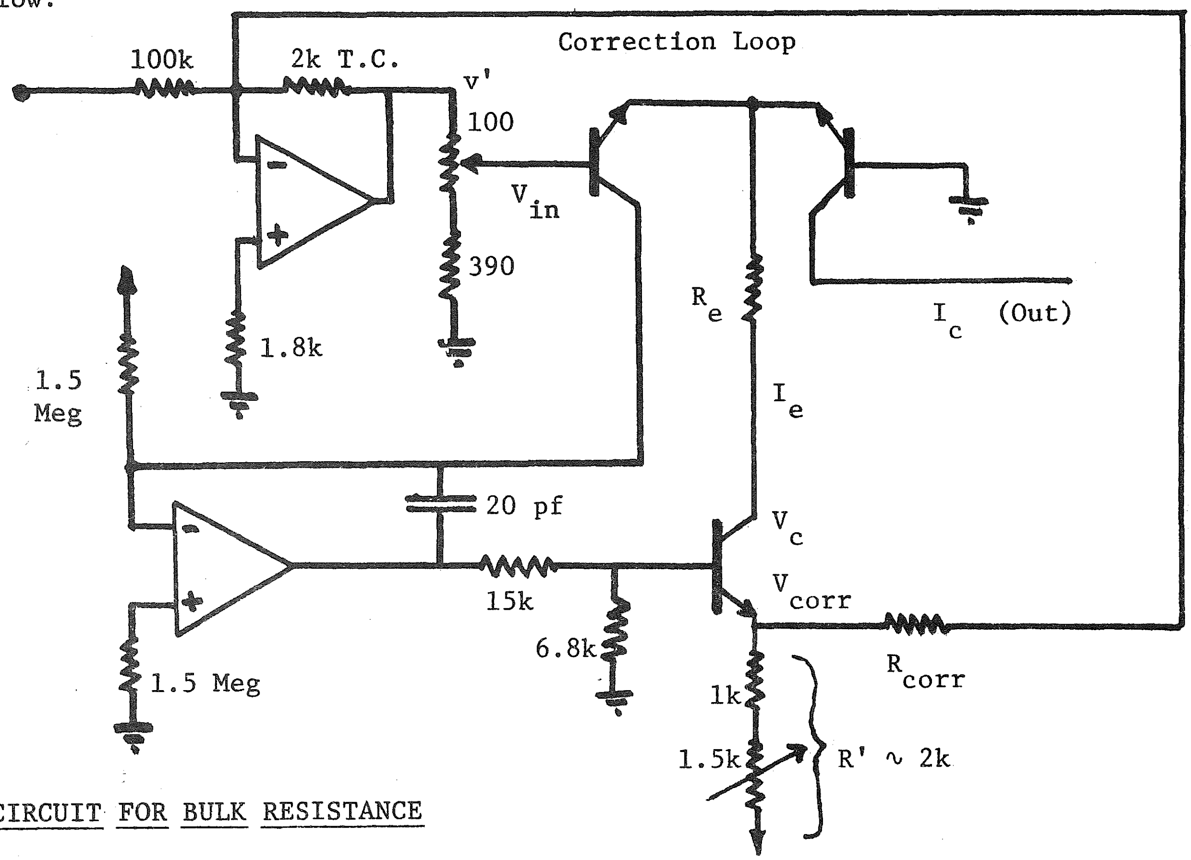
The exponential current stage above was a current sink, it pulled current into it. At times we need an exponential current source. This is true when we are driving the type CA3080 OTA as a control element. It is unfortunate that the CA3080 requires control current input to it rather than pulled out of it. To see why this is, consider that we have to use a PNP pair for a current source. This means that we need a rising input voltage for more current, not a falling one as in the case of the NPN pair. Thus, the inverting input summer is not enough if we want rising voltages to represent rising frequencies. What must be done in this case is to use an inverter stage, or a current mirror on the exponential output. This type of circuit will be demonstrated as the example design in this chapter. Note that while this extra inverter is a drawback in the VCO case, it can prove useful in VCF design as it provides a means of inverting the filters response to a control envelope.

We mentioned that it is necessary to compensate the exponential converter for the bulk resistance of the base-emitter junction. To understand this, consider the transistor model at the right where we show this bulk resistance as an equivalent series resistor r_e. This means that the actual transistor equation is:



$$I_c = \alpha I_{ES} e^{q(V_{be} - r_e I_c)/k_B T}$$

This bulk resistance proves a problem only at high currents (high frequency) as otherwise the $r_e I_c$ term is very small. To correct for this, we have to develope a voltage proportional to I_c and feed this back. A suitable circuit is shown below:



CIRCUIT FOR BULK RESISTANCE

CORRECTION

Analysis of the circuit goes as follows: The bulk resistance is generally on the order of 10 ohms. For certain "log conformance" transistors, it may be as low as 1 ohm, but we will use the larger value. The circuit is similar to the earlier design, except the emitter line is split into three parts: the voltage across R_e, the voltage across the extra transistor, and the voltage across R'. In normal operation, The two currents I_{ref} and I_c add together and pass through both R_e and R'. The correction voltage V_{corr} is thus $I_e R'$, V_c is approximately $-I_e R_e - 0.6$, and V_b is approximately $V_{corr} + 0.6$, held there by the op-amp. When the current I_c becomes large, then the voltage error due to bulk resistance is:

$$I_c r_e$$

and the correction voltage is: $V_{corr} = R'I_c \simeq 2k \cdot I_c$

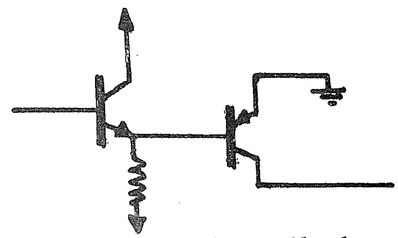
We must feed a portion of V_{corr} back through the input so that it cancels the error term. Thus:

$$2k \cdot I_c \cdot \frac{2k}{R_{corr}} = I_c r_e = 10 \cdot I_c \qquad R_{corr} >> 2k \quad V_{in} \sim v'$$

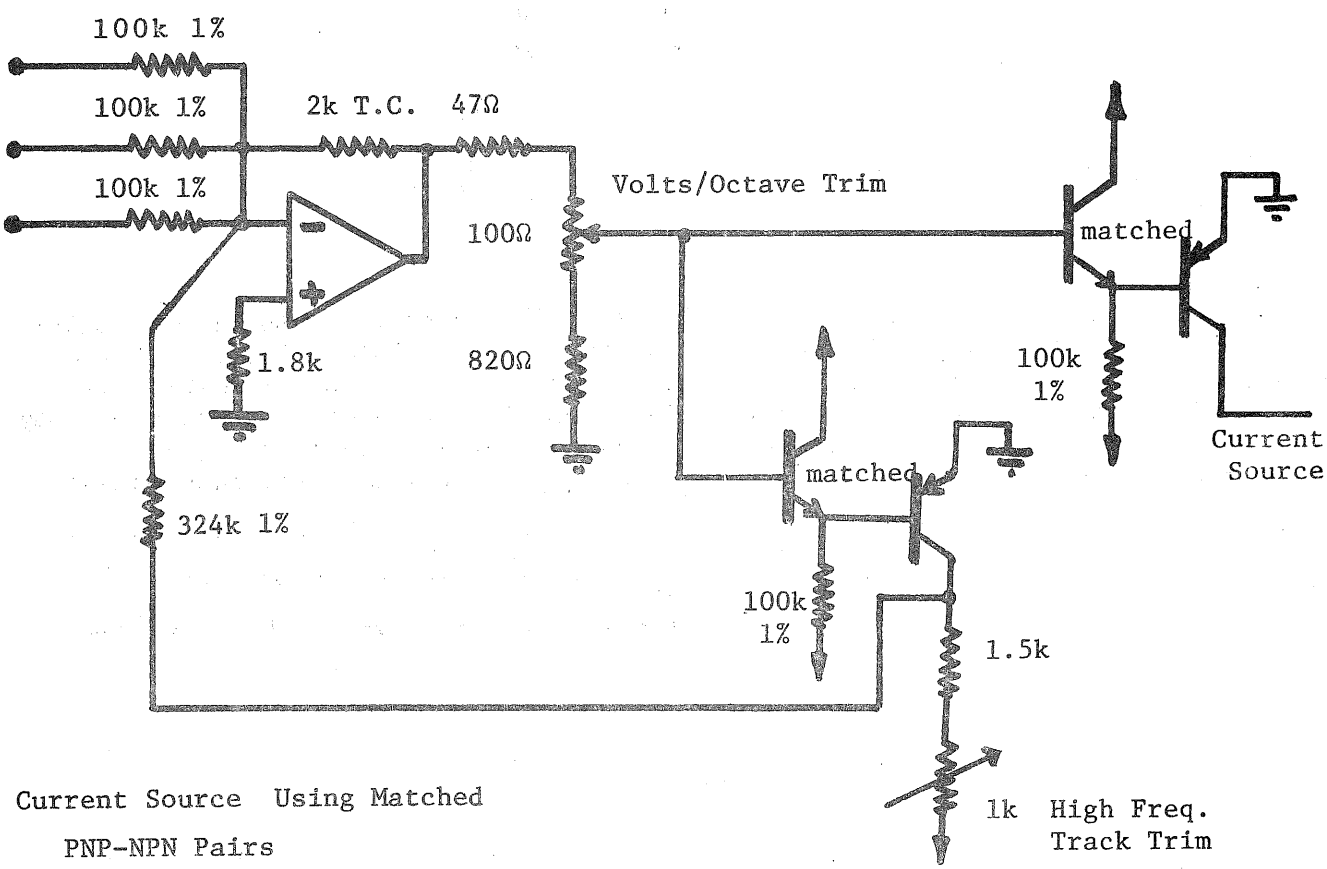
from which we find that $R_{corr} \simeq 400k$. This would be the value of resistance to start with.

We noted above that this bulk resistance error is a high end error like the reset timing error. Where reset timing error is fairly small, it is often possible to just overcompensate the exponential converter for bulk resistance error, and get both at the same time. In the event that the timing error must receive individual care, it may be possible to just set a "ballpark" value for the resistor in the current line, and do the trimming with the bulk resistance correction trimmer.

A technique of exponential conversion employing a complementary pair of transistors was described by T. Mikulic in EN#37 (2). The device is basically an emitter follower of one type driving the exponential converting transistor of the other type. The basic circuit is shown at the right. By using a NPN transistor as an emitter follower, and a PNP transistor as the converting transistor, a current source suitable for driving OTA's is achieved. If the pair are matched for I_{ES}, the temperature dependence of I_{ES}



is cancelled, and the 1/300°K term is compensated as in the circuits described above. The bulk resistance term is compensated with a parallel matched pair which supply a correction voltage. A full circuit is shown below:

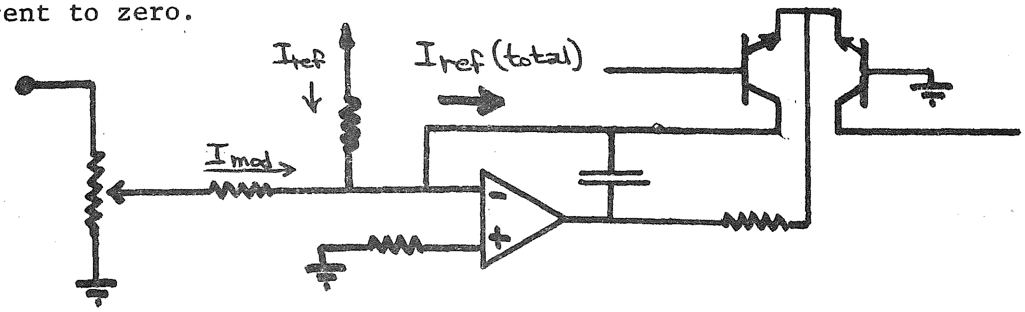


Current Source Using Matched
PNP-NPN Pairs

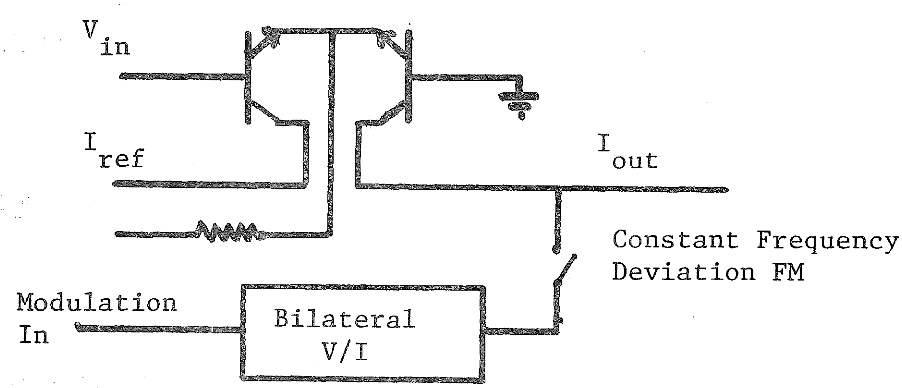
One disadvantage of the circuit is due to the difficulty of obtaining the necessary matched PNP-NPN pairs. It is not possible to match different type pairs to the same precision that can be achieved with monolithic matched pairs.

LINEAR CONTROL INPUTS:

The use of linear frequency modulation for the production of dynamically changing spectra is an essential technique. As has been shown in the chapter on generalized modulations, this is most easily achieved with linear control which are used to supplement the exponential controls. The easiest linear control to implement is the one that results in a constant modulation index. This is achieved by a linear modulation of the reference current to the exponential converter. When setting up such a control, it is only necessary to realize that the reference current is achieved as a current flowing between a supply voltage and ground (a summing node). A second input to this summing node can add or subtract from the reference current. Note also that this sort of input can also be used to gate the oscillator by reducing the reference current to zero.



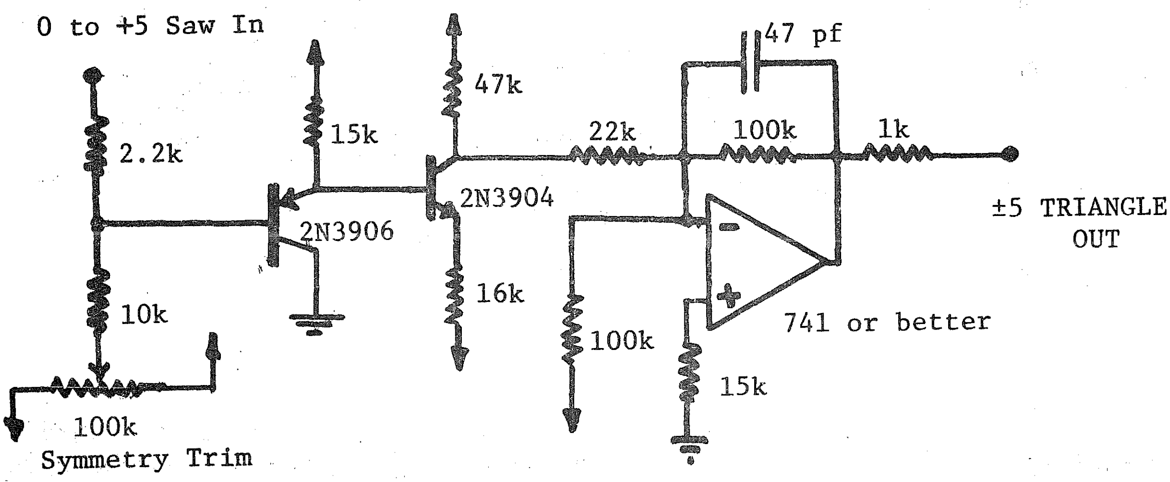
A second type of linear control is acheived by adding a current source to the current line at the output of the exponential source. This control must be carefully done. The result is linear FM with a constant frequency devation. This is also useful for achieving a constant offset between two otherwise tracking VCO's. If two exponential oscillators are offset by 10 Hz at one frequency, the beat frequency will be 10 Hz. One octave higher, the beat frequency will be 20 Hz. Since beating provides useful choral effects, it is useful to have a constant beat rate. The linear control achieves this. Note however, that this implies that one of the "tracking" VCO's is <u>not</u> exponential. The linear offset is a residual number of Hertz that destroys the exponential relationship. Of course, this was being used to advantage in achieving a constant beat rate, but it does point out an important requirement of this sort of linear control - when it is not needed, it should really be off. Thus if this is implemented, there should be an actual switch or some method to insure that there is no leakage current, otherwise, the exponential tracking may be thrown off enough to be annoying.



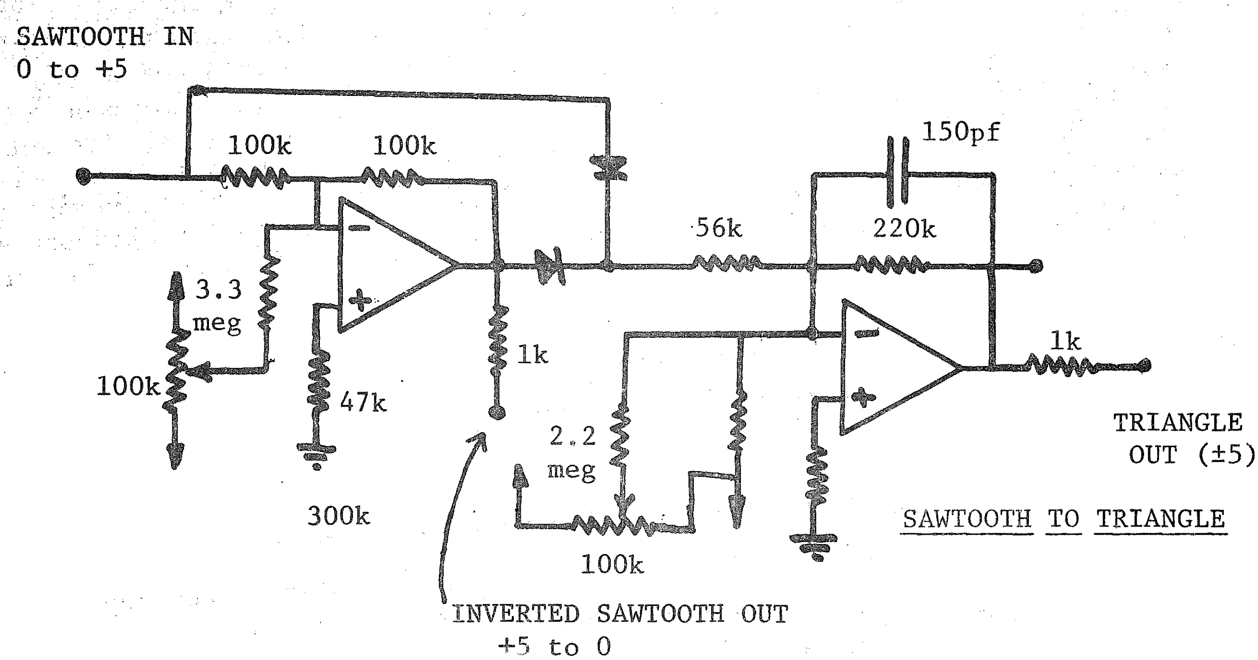
WAVSHAPING CIRCUITRY

SAWTOOTH TO TRIANGLE:

One basic method of converting a sawtooth wave to a triangle was devised by Moog (JAES 1965). It consists of allowing a transistor to either operate as an inverter or a follower. The break point is the point at which the transistor saturates. A circuit for converting a zero to +5 volt sawtooth to a ±5 triangle is shown below. The capacitor is used to remove a small glitch in the waveform that occurs when the transistor goes into saturation:



A second method of converting a sawtooth to a triangle is to first derive the inverted sawtooth, and then employ a circuit that selects the larger of the original or the inverted version. A circuit for doing this is shown below. This method also has a glitch when the diodes switch, and the capacitor removes this.

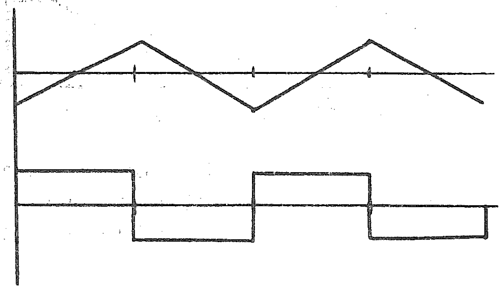


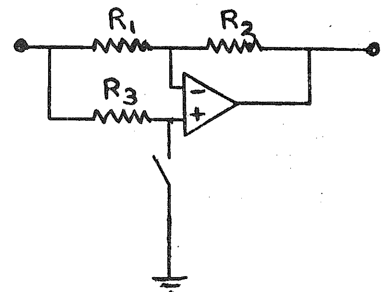
SAWTOOTH TO TRIANGLE

Note that of all the standard waveforms, only the inverted sawtooth makes much sense. This is useful as a control signal. Except for the pulse, the inversion of other waveforms is just a 180° phase shift.

TRIANGLE TO SAWTOOTH

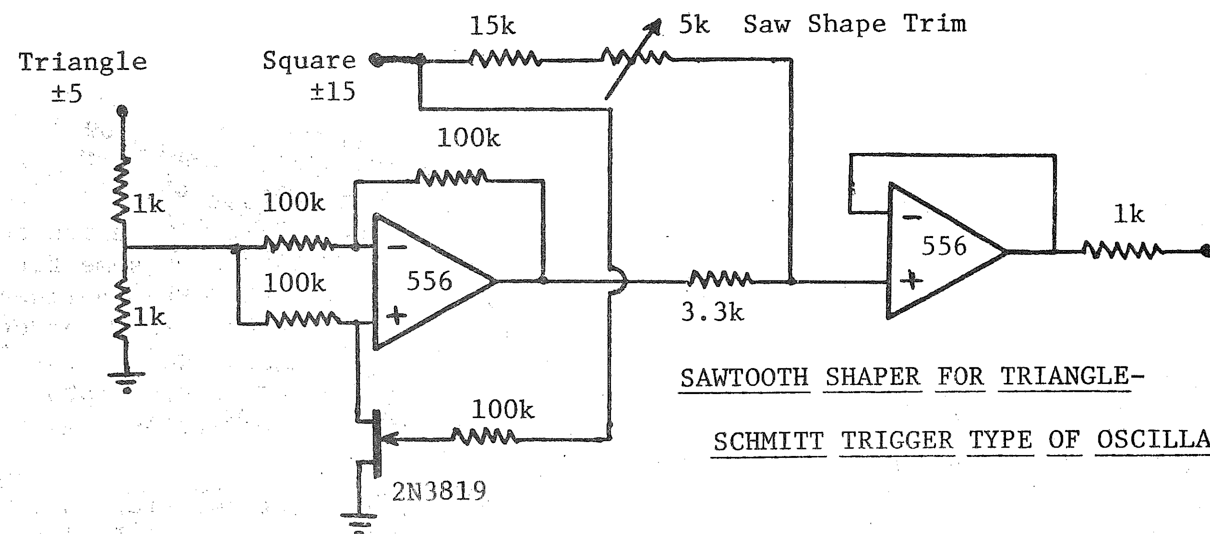
The integrator - Schmitt-trigger type of triangle oscillator also produces a square wave that is in phase with the triangle, and this makes possible a sawtooth conversion method based on a switched inverter-non-inverter.





With the switched closed, R_3 is just a load resistor on the driving stage, and the circuit is an inverter. With the switch open, no current flows through R_3 so the + input of the op-amp is as the input potential. Since the differential input voltage of the op-amp must be zero, the - input is also at the input potential. Thus no current flows through R_1, and no current is available for R_2 either. Thus the output is at the input potential, and the circuit is a follower.

If the circuit is switched by the square wave from the Schmitt trigger, every other half on the triangle wave is inverted, and a double frequency sawtooth results. If part of the square wave is added back in to the double frequency sawtooth, the original frequency is recovered. The basic idea is shown below. In a specific case, some of the phases may be different, but the scheme is similar. It should be realized however that there is a slight "glitch" in the final sawtooth at the midpoint of the ramp. This is due to the switching time of the switch, which is typically a FET switch or an analog switch. The glitch width is independent of frequency and is inaudible. It will however cause an extra count on a frequency meter.

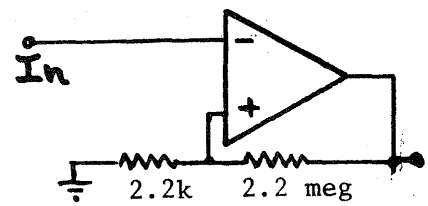


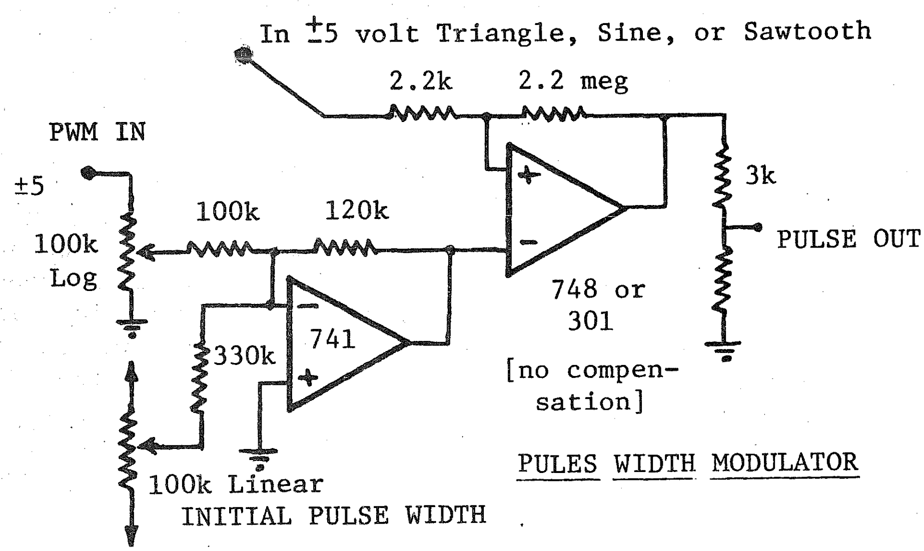
SAWTOOTH SHAPER FOR TRIANGLE-SCHMITT TRIGGER TYPE OF OSCILLATOR

SQUARE WAVES AND VARIABLE WIDTH PULSES FROM OTHER WAVEFORMS

A basic triangle generator has a square wave available directly. The second waveform in a sawtooth generator is the short reset pulse. In general however, we want to be able to derive rectangular waveforms from other smooth waveforms. The basic circuit here is the comparator or weak Schmitt trigger.

The circuit can be formed from an uncompensated op-amp such as the 748 or the 301. The basic circuit is shown at the right. To derive a square wave, it is only necessary to connect the input of the Schmitt trigger to a balanced sine, triangle, or sawtooth. To derive a pulse of variable width, the reference level is made variable. It is also easy to make the reference level respond to an externally applied voltage, and this permits pulse width modulation (PWM). It should be realized however that while this does permit PWM for any smoothly varying waveform, the variation in pulse width with time will depend on the driving waveform as well as the reference waveform. A simple PWM circuit is shown below:



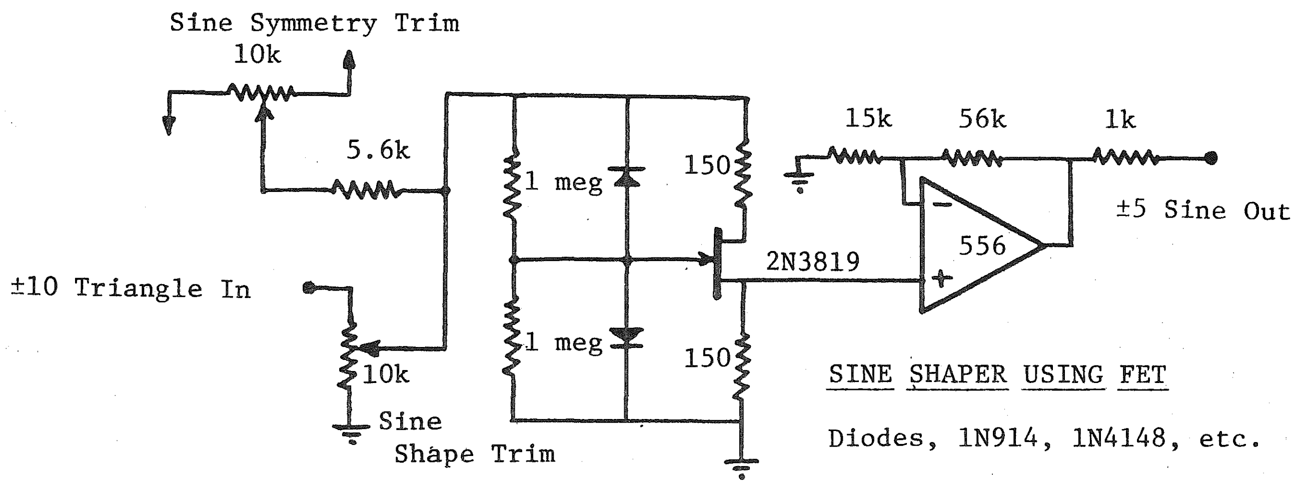


PULES WIDTH MODULATOR

OBTAINING A SINE WAVE

As mentioned above, there is no exact method of obtaining a sine wave from a triangle, and obtaining a sine from any other waveform would be even more difficult. Several methods are available for shaping a triangle into an approximate sine. First, there are a variety of diode waveshaping circuits which form the sine by a series of linear sections. This is a general method, but generally it is too combersome for the possible accuracy. What is more useful is to select a device that is non-linear and use the right portion of the curve to "distort" the triangle in the direction of the sine. In simple terms, we want to round the triangle. One useful device for this purpose is the FET. The exact details of the method are given by Middlebrook and Richer [Electronics, Mar. 8, 1965]. A practical circuit for converting a ±10 volt triangle to a ±5 volt sine is shown on the top of the next page:

Another simple device for shaping a triangle into a sine is the two transistor differential amplifier. In general, this method gives about the same purity sine wave as the FET. It has one advantage (only one trim has to be made) and one possible disadvantage (it does not look best when it sounds best!). In the next chapter of VCA design, it is shown that the two transistor differential amplifier is linear only for small input signals on the order of 10 mv. Beyond 40 mv, the curves start to bend over. Thus, if we drive (i.e., overdrive) the differential amplifier with a triangle, the output is bent toward a sine.



SINE SHAPER USING FET

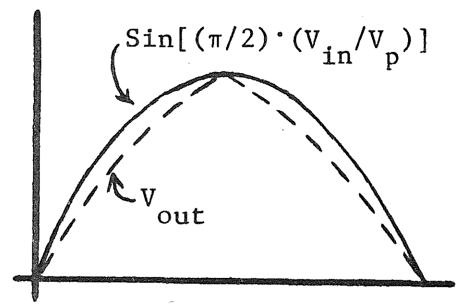
Diodes, 1N914, 1N4148, etc.

From the equations for the two transistor differential amplifier, it can be shown that if the peak of the triangle is V_p, the amplitude of the approximating sine is:

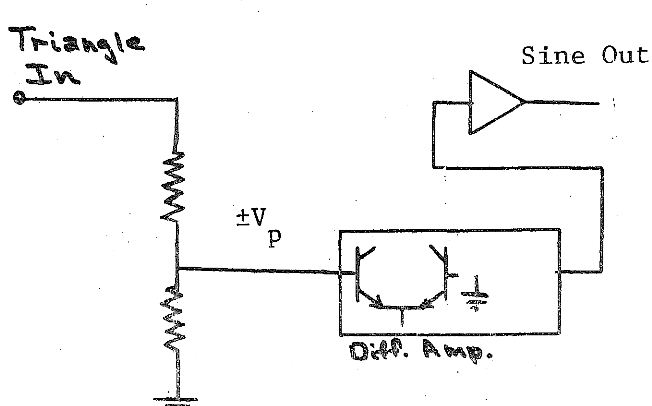
$$A = 2\left[-0.5 + \frac{1}{1 + \mathrm{Exp}(-V_p/26)}\right]$$

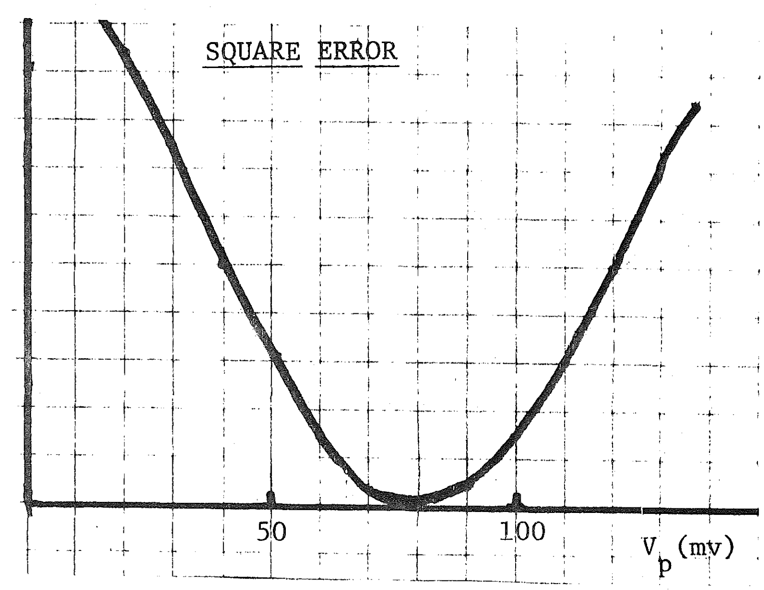
Thus, the output sine approximation is:

$$V_{out} = \frac{2}{A}\left[-0.5 + \frac{1}{1 + \mathrm{Exp}(-V_{in}/26)}\right]$$

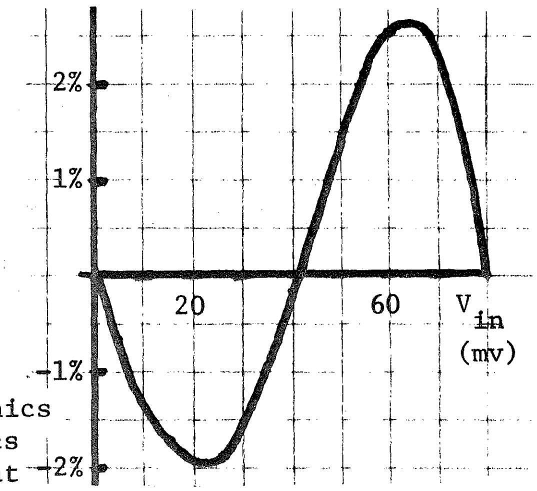


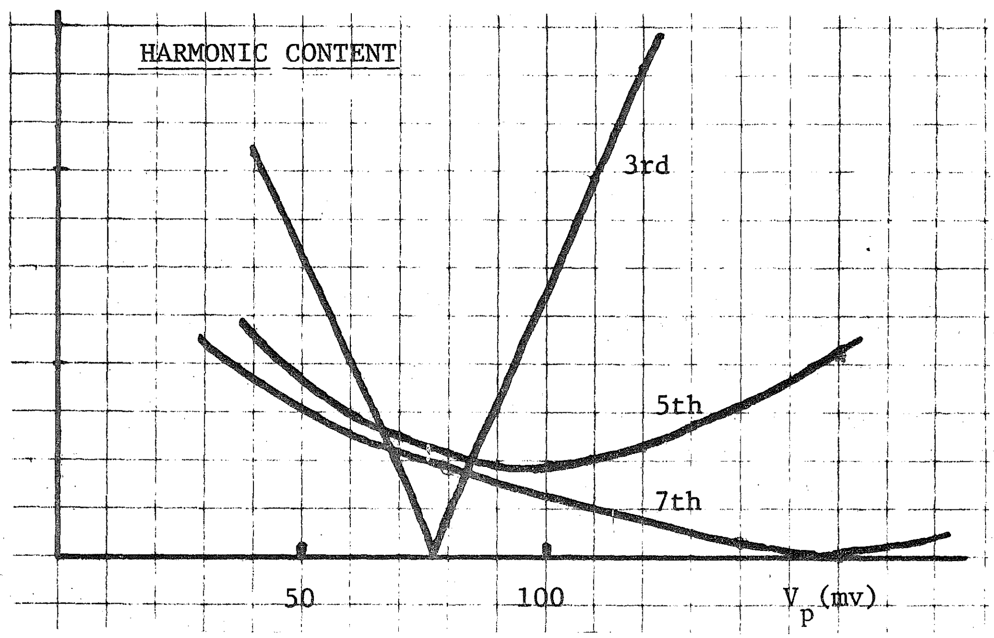
It thus remains to find the best fit of V_{out} to the sine as a function of V_p. As V_p is increased, the output will go from a triangle to the best rounded approximation to a sine, and round on through toward a square. It is useful to consider the minimum square error as a measure of the best fit. The plot of square error as a function of V_p is shown below:





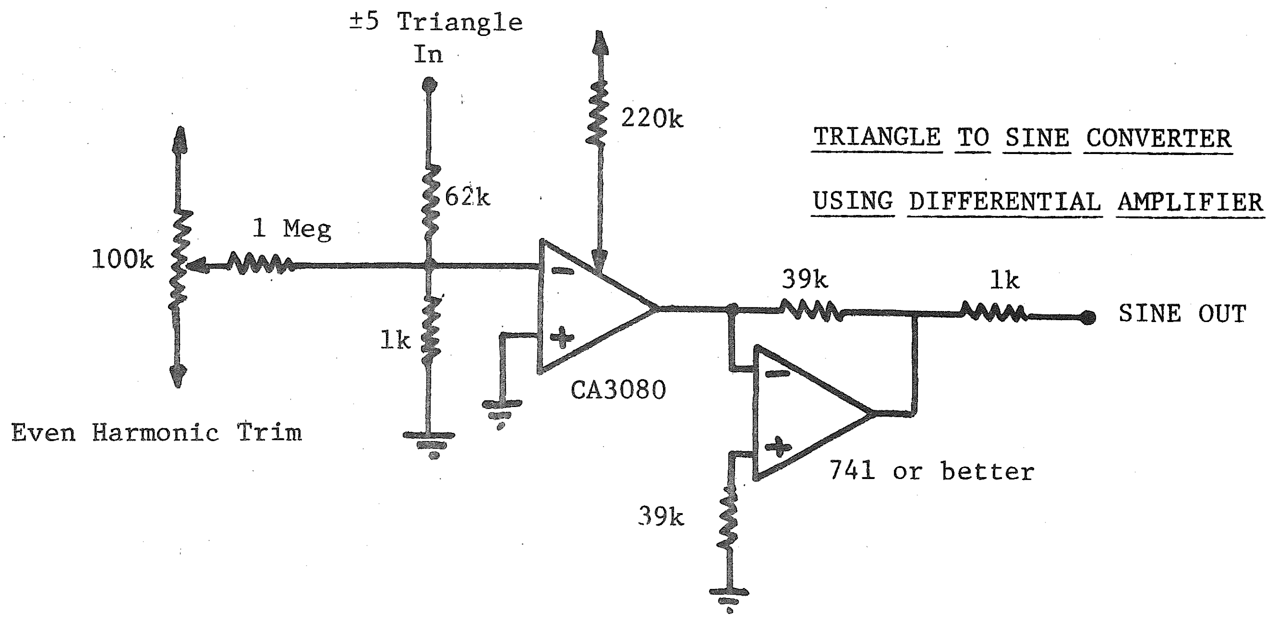
The minimum square error occurs for V_p = 80 mv. A plot of the actual error for 80 mv peak is shown at the right as a function of V_{in}. This shows that for the best least square fit, the maximum absolute error between the sinewave and its approximation is less than 3%. It is also useful to look at the harmonic content of the resulting waveform to see where the various harmonics go through a minimum. By symmetry, there are no even harmonics in the output. In practice, we will set V_p to give the best fit, and the one trim that we will have to make is for a balanced waveform to keep even harmonics from coming out. A plot of the 3rd, 5th, and 7th harmonics are shown below. Note that the 3rd harmonic goes to zero at about 77 mv, and the 5th harmonic goes through a minimum at about the same point or a little higher. Thus, we see strong evidence for prefering a peak input of 80 mv. Listening tests bear this out.





When V_p is actually set for 80 mv, one disadvantage does appear that is really of no importance, but which may have to be considered for commercial designs. The sine wave still has visable points on it at 80 mv, and it is not until about 100 mv input that the waveform appears well rounded (it is actually over rounded at this point, and more distortion results). The FET sinewave shaper on the other hand looks okay when it is set for minimum distortion. Bear in mind however, that neither method is perfect.

For a practical circuit, we can use the OTA since this has a differential input structure (line all op-amps) and its total gain can be controlled (unlike all op-amps). A practical circuit that sets the peak triangle input at about 80 mv is shown at the top of the next page. In this circuit, only one adjustment is required. The circuit using the FET actually has two trimmers of which one (sine shape) is actually a trimmer to adjust to the specific device used. The OTA on the other hand uses the set of "Universal" diff amp curves, and the performance should be device independent.

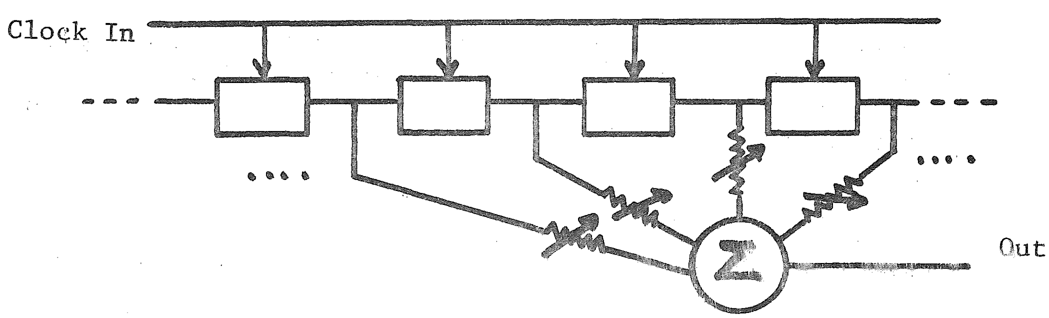


TRIANGLE TO SINE CONVERTER

USING DIFFERENTIAL AMPLIFIER

DIGITAL WAVESHAPING

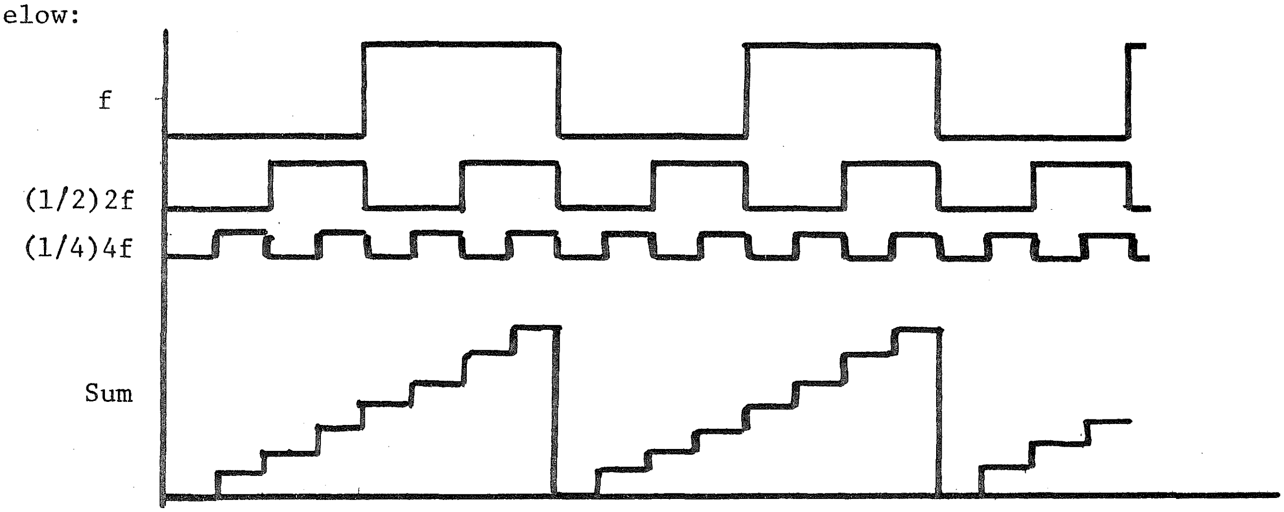
When it is necessary to generate arbitrary waveforms (and we must be careful to determine if it really is necessary), the digital method is probably the easiest. The basic idea is as follows: You start with a frequency that is N times the repetition rate of the waveform you are trying to generate. You then let this trigger a digital counter that has N states. For each state of the counter, a voltage is put out that corresponds to the approximate value of the waveform at that time. This results in a stepped approximation to the desired waveform. The stepped waveform can be smoothened with a filter or by a piecewise linear method if this is necessary. The simplest type of counter for this purpose is the bucket-brigade type of digital counter. This counter passes a logical "one" down the line with each clock pulse, and at the end, the pulse returns to the first counter stage and the process repeats. Note that this circuit, and other digital waveshaping circuits are structurally similar to sequencer circuits, and often times, the same circuitry can be used for both, a fact that should not be overlooked. For an actual voltage output, a weighting and summing circuit is connected to the counter. This is an analog method of forming the proper shape.



Another method of digital waveshaping is to use some sort of memory or look-up table to give an approximate value of the waveform as a digital word. For example, for a given segment of a sine wave cycle, a value for the sine wave could be read from a digital memory, and this could be output through a standard digital to analog converter. Note also, that in such a case, it is often possible to store in the memory fewer values than will be present in the output cycle, since it is often possible to use the symmetry of the waveform to advantage.

General programable waveform generators must be considered for the given application, and there is not too much that can be said ahead of time about them. Instead, we will take a look at a few stepped waveforms to see their properties. It should be realized that when we use stepped waveforms, we are really working with sampled data systems, and the mathematics that applies to the sampled data systems applies here too.

A special stepped waveform that is popular in electronic organ circuits is the "staircase" wave. The staircase wave is generated by inputting sharp pulses into an integrator, or it can be generated from octave square waves as is indicated below:



The staircase is a stepped approximation to a sawtooth wave as can be easily seen.

The digital generation of sine waves has received considerable attention. One surprising result is that for a given number of segments N, the proper level selected for each segment is the value of the sine wave as it would be sampled at the center of each segment. This minimizes the mean square error. For N segments in time, the clocking frequency of the digital counter is N times the frequency. The spectrum of the reconstructed sine wave contains the harmonics: f, Nf-f, Nf+f, 2Nf-f, 2Nf+f, 3Nf-f, 3Nf+f, etc. The total harmonic distortion due to the segmentation in time is given approximately by:

$$D_{ht} = \frac{\pi}{\sqrt{3}\ N}$$

The above distortion figure assumes infinite precision for the amplitude levels. One case where this is certainly not true is where the amplitude level is represented by a digital word. If there are N' possible digital levels, then the distortion due to segmentation in amplitude is:

$$D_{ha} = \frac{1}{\sqrt{6}\ N'}$$

For example, for eight segments in time, the first formula gives about 23% distortion.

When more accuracy is required in a sine wave "look-up" system, it is possible to use trig identities to reduce the momory storage requirements. The useful trig identity is:

$$\operatorname{Sin}(x+y) = \operatorname{Sin} x \ \operatorname{Cos} y + \operatorname{Cos} x \ \operatorname{Sin} y$$

Here, x can be the coarse value of the angle, and y can be a small correction. It is thus possible to have a table for x of the ordinary one quadrant type, and a table for small values of Sin y and Cos y .

Another possibility for the digital generation of waveforms is to use a set of orthogonal functions and sum them in a series just as sinewaves are summed to form complex waveforms. A set of orthogonal functions that are all rectangular is available in the "Walsh functions," and a discussion of this method is found in this book in the section an additive synthesis. With the Walsh functions it is often possible to use fewer summing resistors. The digital part of the circuit has approximately the same hardware complexity as a point by point method. The calculations are harder.

Before ending our discussion of the digital waveshaping methods, we should say a few things about the requirements of the VCO that drives the counter. It obviously must run at a much higher frequency than normal, and we already know that the full audio range starts to put VCO designs to the test even when they only have to run at the original frequency. Thus, it would be a considerable problem if ordinary VCO's were required to run at 20 times the normal frequency or higher. The percent error of the upper frequency would be transfered down by the digital counter to the same percent error down below, and this would be worse than the same VCO working directly at the lower frequency. Thus, if a VCO is to be used for a digital waveshaper, it is good to use a different design. What saves the day here is that it is not necessary to maintain a waveform for the digital waveshaper, all you need is a pulse to trigger the counter, and the VCO can be doing all sorts of tricks of its own and it won't matter as long as the pulse arrives on time. Probably the most useful trick here is to swamp the actual reset time with a much longer one, and then compensate for the fixed reset time as described earlier.

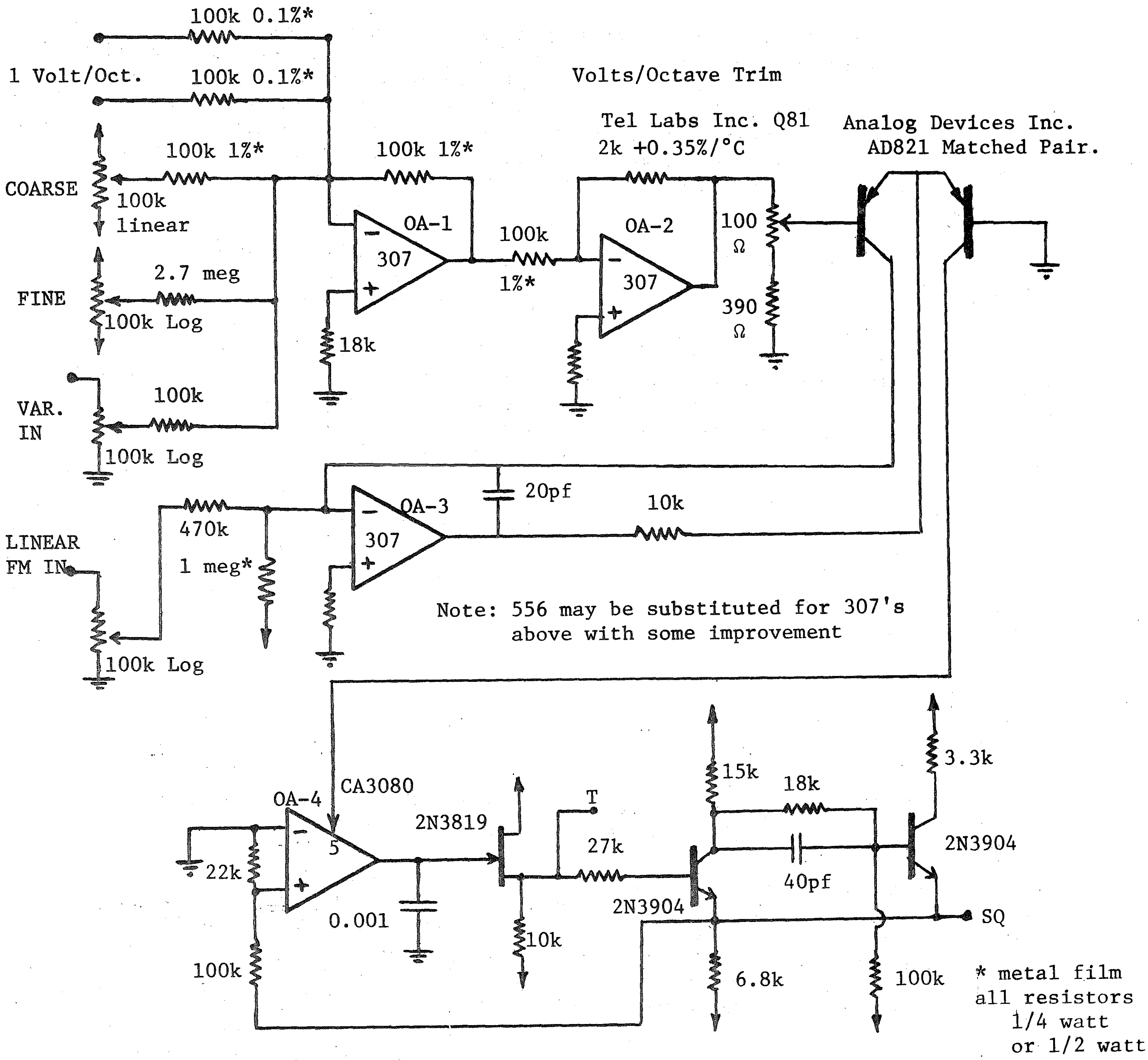
VCO DESIGN EXAMPLE

The VCO design example is based on the circuit in EN#46, which has proven a reliable circuit. Most of the structures in the design have been studied in detail in this chapter. A few specifics will be mentioned.

Since we are using the CA3080 in the basic oscillator, we have to use a matched PNP pair in the exponential converter. This makes necessary the inverter stage (OA-2). We have also added a (*) to some of the resistors to indicate that they should be of the metal film type for stability. The lowest frequency is a small fraction of 1 Hz. The highest possible frequency is obtained by considering the current limiting of R_e. The maximum current through R_e is 15/10k = 1.5 ma, and this is approximately the max. current into pin 5 of the CA3080, and the max. current out. The triangle output has a full cycle swing of 20 volts (-5 to +5, and +5 back to -5). Thus:

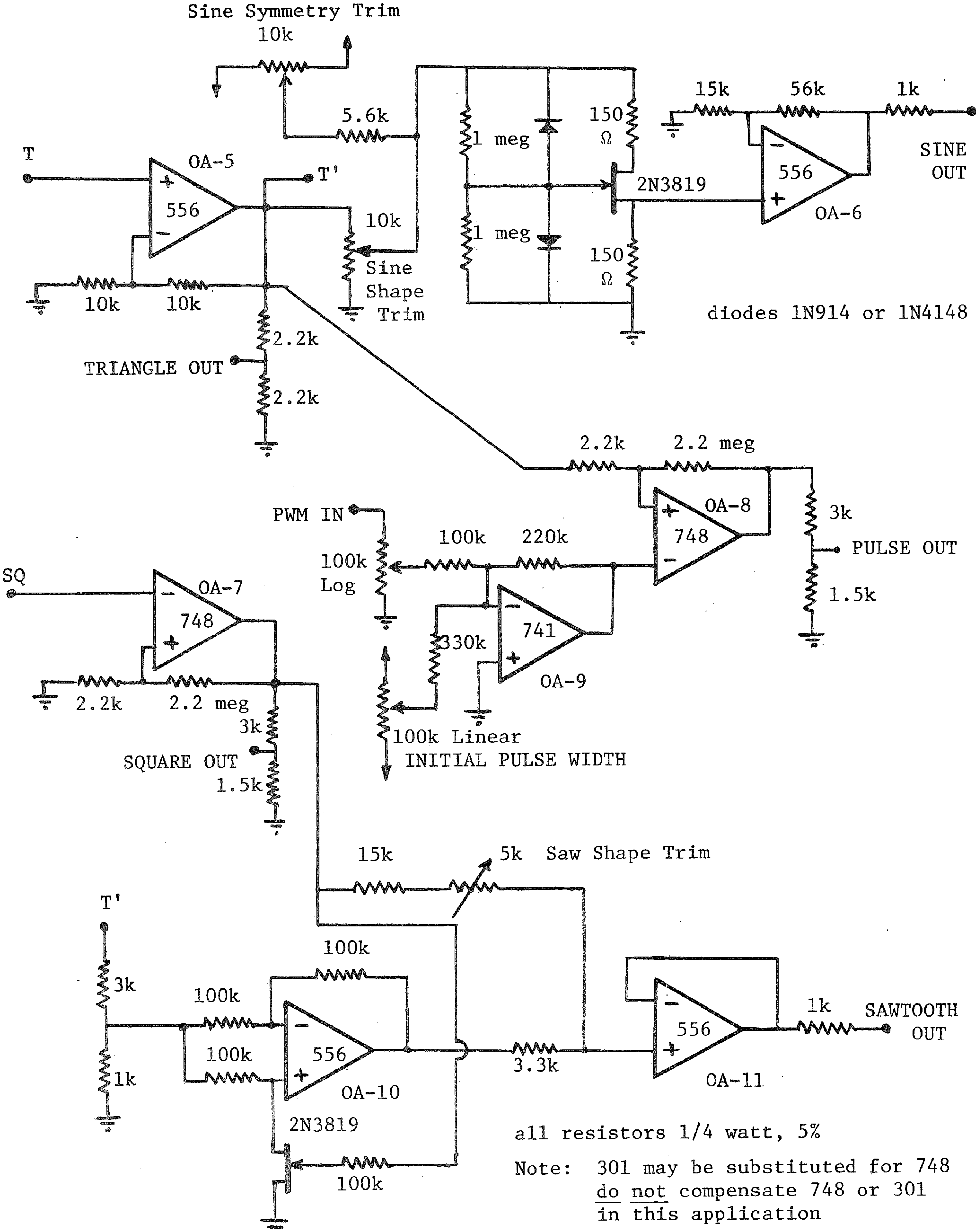
$$f_{max} = \frac{1}{\Delta t} = \frac{1.5\ \text{ma}}{0.001\text{mf} \cdot 20\text{volts}} = 75\ \text{kHz}$$

The Schmitt trigger selected for the design example is the two transistor discrete version discussed in EN#46. A linear FM control input has been added (EN#49). There are two outputs from the basic oscillator, (T) triangle, and (SQ) square.



VCO DESIGN EXAMPLE, EXPONENTIAL CURRENT SOURCE AND BASIC OSCILLATOR

The waveshaping circuits have all been discussed. The first step in the process is to buffer and amplify the triangle (OA-5), and then divide the output by tow to give a ±5 triangle out. The ±10 volt triangle (T') drives the FET sine shaper. It is necessary to have at least a ±7 volt triangle for this sine shaper. The PWM circuit is the same as in the discussion except the gain of the summer for the reference voltage has been increased since the driving waveform is ±10 volts. OA-7 serves to "buffer" the Schmitt trigger square wave, since any current drawn from the two transistor Schmitt trigger could upset the levels, and certainly a 1k output here would upset things if the 1k resistor were connected to the tied emitters. The sawtooth circuitry is the same as in the waveshaping discussion except it is being driven by T', the ± 10 volt triangle, so there is a 4:1 attenuator on the T' input.



all resistors 1/4 watt, 5%

Note: 301 may be substituted for 748
do not compensate 748 or 301
in this application

VCO DESIGN EXAMPLE, WAVESHAPING CIRCUITRY

Chapter 5c

VOLTAGE-CONTROLLED AMPLIFIER DESIGN

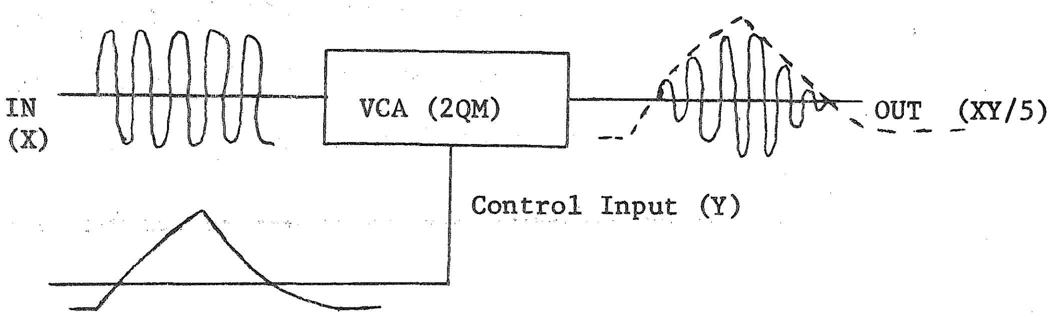
CONTENTS:

INTRODUCTION:

A voltage-controlled amplifier (VCA) is usually a two-quadrant multiplier. The VCA thus has one input that can handle voltages of either polarity with respect to ground, and a second input that handles only one polarity (positive). The two-quadrant multiplier is adjusted so that the gain of the circuit is one when the unipolar input is at its maximum value. For example, if it is used in a system where the maximum signal levels are ±5 volts, we could take the bipolar input to be X, and the unipolar input as Y. The X input can then have any value from -5 to +5. The Y input can receive any voltage from -5 to +5, but it treats all voltages from -5 to zero just as it does zero volts. The Y input therefore responds to voltages in the range of zero to +5. When the X input is full ±5 volt signal, we want the same signal at the output when the Y input is +5. This means that the response of the multiplier should be $X \cdot Y/5$. For a system with standard signal levels of ±10, the multiplier should give a response of $X \cdot Y/10$. These same responses are the correct ones for any four-quadrant multipliers in the system, and one often considers the use of a four-quadrant multiplier as a VCA. In many cases, the four-quadrant multiplier (4QM) can be used, but for a wide dynamic range, the two-quadrant multiplier (2QM) has a great advantage.

A typical 4QM has a dynamic range of 40-60 db. The ear has a dynamic range well in excess of 100 db. Thus it is reasonable to require that a VCA used for amplitude control have a range of at least 80 db. The better quality 2QM's have this necessary dynamic range (80-100 db). The main reason why the 2QM has a larger dynamic range is that it is simpler. Often, a 4QM is composed of two 2QM's that work together. This requires critical balance of components to work right. On the other hand, most synthesis systems have at least one 4QM that serves as a balanced or "ring" modulator, and this can often be used as a VCA in cases where the full dynamic range of the 2QM is not needed. This might occur for example in a case where a mudulating signal is enveloped in. It may well be that the circuit being modulated will not even respond to the modulating signal when it is down by 60 db, or if it does, the ear may not be sensitive enough to detect the modulation.

A second reason for preferring a 2QM over a 4QM for a VCA used for amplitude control is that the control signal need not be nearly as precise in its low range. In particular, for complete shutoff (as far as the ear is concerned) the control signal need be only zero or a little negative - a precise zero is not required. Also, amplitude envelopes in common use have an exponential decay. This is reasonable when you consider that many musical instruments employ vibrating structures that also have (at least approximately) an exponential decay. However, the exponential decay to zero volts sometimes seems to hang on a bit too long to sound musically right. In such cases, the expoenntial decay envolope can be made to decay to a slightly negative voltage. Feeding this to a 2QM causes an earlier cutoff. This sort of "truncated exponential" decay would not be possible with a 4QM.



Truncated Exponential Envelope

DESIGN OF TWO QUADRANT MULTIPLIERS

The most popular 2QM design in use today for VCA's is the one based on the principle of "variable transconductance" as applied to the differential amplifier. The term "transconductance" is a combination of transfer function (the output divided by the input) and conductance, which is reciprocal resistance (I/E). Thus, transconductance is the output current change relative to the input voltage. For generality, transconductance is taken as the ratio of differentials (dI/dE) and is often denoted g_m. The units of transconductance are reciprocal ohms or mhos (ohms spelled backward), but often we just think of transconductance in terms of a ratio of amps to volts.

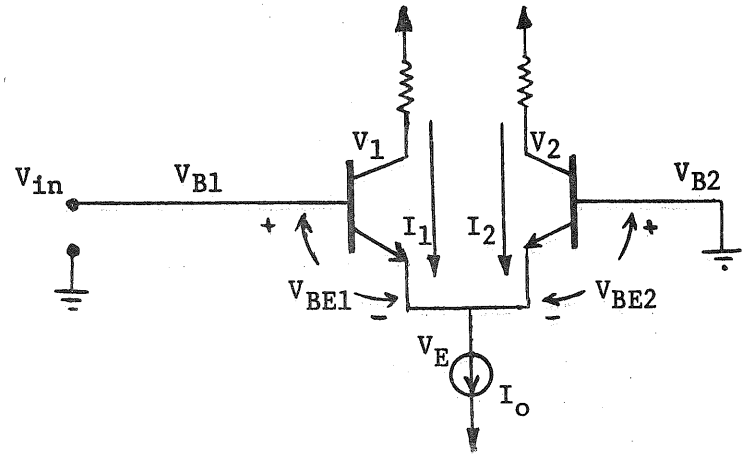
We noted in the discussion of exponential converters that the transistor obeys the relationship:

$$I_C = I'(e^{qV_{BE}/k_BT} - 1)$$

where I' is some reference current in the range of 10^{-12} to 10^{-14} amps, q is the charge of the electron, and k_BT is about 1/40 of an electron-volt at room temperature. The -1 can be neglected for forward bias. The term q/k_BT is just a voltage of 1/40 volts (about 26 mV) and is often denoted as V_T. With these simplifications. the relationship between collector current I_C and base-to-emitter voltage V_{BE} is:

$$I_C = I' e^{V_{BE}/V_T}$$

We can next consider the simple two transistor differential amplifier that is driven by a current source I_o connected to the tied emitters.



We will assume that the transistors are a matched pair. Clearly the two emitter currents must add to I_o, and if we apply the exponential equation and the fact that the emitter and collector currents are nearly equal (high β transistors), we get:

$$I_o = I_1 + I_2 = I' e^{V_{BE1}/V_T} + I' e^{V_{BE2}/V_T} \qquad \text{Eqn. 3}$$

This can be solved for I' as:

$$I' = I_o/[\, e^{V_{BE1}/V_T} + e^{V_{BE2}/V_T}] \qquad \text{Eqn. 4}$$

and this can be substituted back into Eqn. 3 to give:

$$I_1 = I_o/[\, 1 + e^{V_{BE2}-V_{BE1}/V_T}] \qquad \text{Eqn. 5a}$$

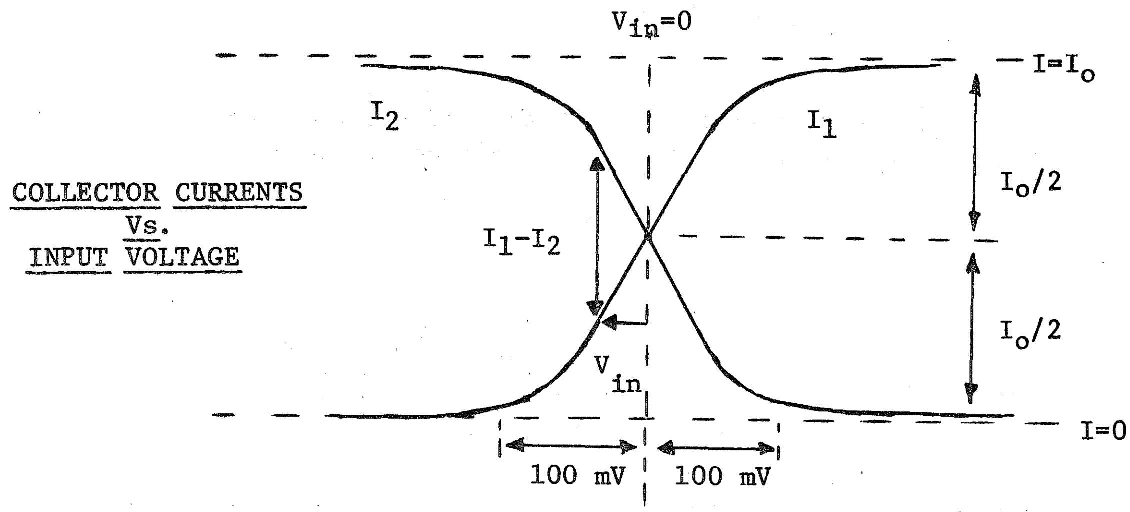
$$I_2 = I_o/[\, 1 + e^{V_{BE1}-V_{BE2}/V_T}] \qquad \text{Eqn. 5b}$$

Next note that $V_{BE2} - V_{BE1} = V_{B2} - V_{B1} = -V_{B1} = -V_{in}$ since $V_{E1} = V_{E2}$ and $V_{B2} = 0$.

$$I_1 = I_o/[1 + e^{-V_{in}/V_T}] \qquad \text{Eqn. 6a}$$

$$I_2 = I_o/[1 + e^{+V_{in}/V_T}] \qquad \text{Eqn. 6b}$$

The limiting and special cases for these two equations are when V_{in} is large and positive which makes $I_1 \approx I_o$, $I_2 \approx 0$, when V_{in} is large and negative, which makes $I_1 \approx 0$, $I_2 \approx I_o$, and when $V_{in} = 0$,in which case $I_1 = I_2 = I_o/2$. The two curves are plotted below:



From these curves, we can note several observations. First of all, if V_{in} exceeds about 100 mV, there is little further change in the collector currents. This means that the amplifier saturates at about 100 mV. Further, about the central point (V_{in} = 0), the curves are approximately linear. We will want to look at the non-linearity as a function of the magnitude of the input voltage. The actual output from the differential amplifier will result from the differential current $I_1 - I_2$. Note that the shape of the curves is independent of I_o, but as I_o increases, the magnitude of $I_2 - I_1$ for a fixed V_{in} increases. This is what we are looking for: a gain that depends on an external voltage or current.

Getting back now to the transconductance, we are now in a position to make the calculation of g_m. The differential current is simply twice the value we have calculated for either I_1 or I_2. We can thus write:

$$I_1 - I_2 = 2I_o/[1 + e^{+V_{in}/V_T}] \qquad \text{Eqn. 7}$$

This can be differentiated with respect to V_{in} by the usual rules for differentiating a quotient to give:

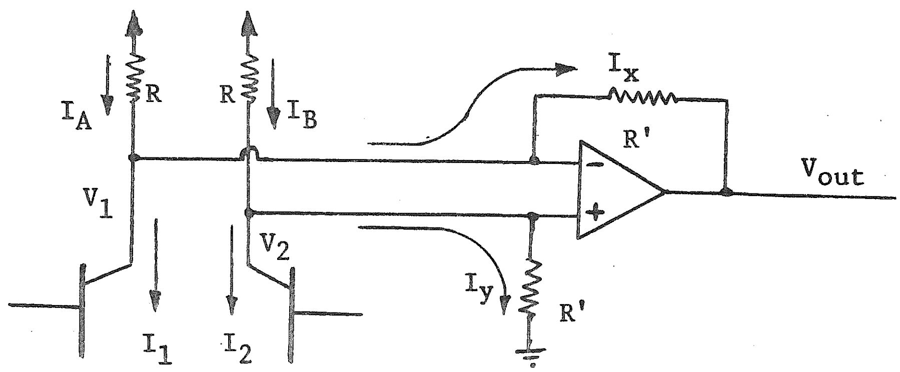
$$\frac{d(I_1 - I_2)}{dV_{in}} = \frac{2I_o(e^{+V_{in}/V_T})(1/V_T)}{1 + 2e^{+V_{in}/V_T} + e^{+2V_{in}/V_T}} = g_m \qquad \text{Eqn. 8}$$

This can be evaluated at the center point (V_{in} = 0) to give:

$$g_m = I_o/2V_T = I_oq/2k_BT \qquad \text{Eqn. 9}$$

To get some idea of the non-linearity of this tramsconductance, we can evaluate g_m at other values of V_{in} simply by plugging the values of V_{in} into Eqn. 8 and seeing how far this value of g_m deviates from the center value given by Eqn. 9. For V_{in} = 10 mV, the change in g_m is about 4% so we can see that we have to keep input signals on the base of the transistors below this value to avoid distortion. This is why there is usually an input attenuator (voltage divider) on the inputs. It is of course desirable to keep signal levels high in other circuits to have a favorable signal to noise ratio. In the differential amplifier however, we must limit the voltages to prevent distortion (the non-linearity is due to the exponential relation, Eqn. 2). However, even 10 mV is a very small signal, and we don't want to go much below that or the signal to noise ratio on the base of the transistor is too high. We will see later that there are ways of "predistorting" the signal input voltage to linearize the transfer curves, but many practical applications can be met simply by limiting the input voltage to 5 to 10 mV.

It is necessary to get the signal out of the differential amplifier as a single ended (referenced to zero volts) output signal. The differential current must be converted to a voltage difference. A differential amplifier formed from an op-amp could be employed, but this requires high input impedence to avoid loading the collector resistors, and matched resistors on the op-amp circuit to obtain a high common mode rejection. In fact, it should be pointed out that to this point, the collector load resistors have not been involved in the discussion or the calculations - they could have been left out. Along with the matched collector resistors, we would have 3 matched pairs of resistors if we tried the op-amp differential amplifier. A better solution is to use the currents I_1 and I_2 to rob current from the op-amp input nodes with both the input resistors connected to the positive supply voltage. Or, looked at another way, the collector load resistors become part of the op-amp differential amplifier. In either case, the thing looks a little fishy, and has to be carefully analyzed to show that it actually does work. The circuit is shown below:



Analysis goes as follows: $I_A = I_x + I_1$ and $I_B = I_y + I_2$ by just summing currents. Next, observe that since the differential input voltage of the op-amp must be zero, $V_1 = V_2$ and this tells us that $I_A + I_B$. This tells us that:

$$I_x - I_y = I_1 - I_2$$

Also, $V_1 = V_2 = I_y R'$, and $V_{out} = V_1 - I_x R'$ so:

$$V_{out} = R'(I_y - I_x) = R'(I_1 - I_2) \qquad \text{Eqn. 10}$$

So the output of the op-amp is proportional to the current difference.

We can now look at the voltage gain of the whole circuit. For small signals and a constant current I_o, g_m is close to its V_{in}=0 value of $I_o/2V_T$. We can then replace the differentials by small excursions and write:

$$V_{out} = R' \ (I_1 - I_2) = R' g_m \ V_{in}$$

Where $d(I_1-I_2)$ and dV_{in} in Eqn. 8 have been replaced by (I_1-I_2) and V_{in}. The voltage gain of the amplifier is thus:

$$G = V_{out}/\ V_{in} = g_m R' = I_o R' q/2k_B T \qquad \text{Eqn. 11}$$

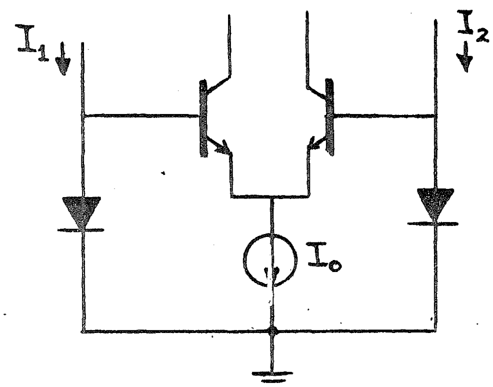
Note in particular that the gain is proportional to I_o, and inversely proportional to the absolute temperature T.

We noted above that the non linearity for a peak input voltage of 10 mV was on the order of 4%. It should not be assumed however that the same figure carries over to the case of harmonic distortion. The harmonic distortion in this case is only about 1%, a much smaller figure.

At this point it is necessary to consider the application intended for the 2QM as this will determine where we go from here. For general purpose VCA's (60-80 db dynamic range), what has been done above is sufficient, and all that needs to be added is the controlled current source. For special purpose VCA's or for the very best quality VCA's, it may be necessary to improve the linearity to increase the dynamic range without increasing distortion (80-100 db dynamic range). This can be done by pre-distorting the input voltage so that it compensates for the non linearity of the curves for the differential amplifier. This will allow the input of signals of 50 mV peak with no increase in distortion. Since the noise level is fixed, the dynamic range is improved along with the signal/noise ratio.

It is generally not necessary to compensate for the 1/T term in the g_m term if the 2QM is used for controlling audio amplitude. The ear can not detect small changes in amplitude any where near as well as it can small changes in pitch. In certain applications, the 2QM might be used as a control element for VCO's or VCF's, and the 1/T term could prove troublesome. Interestingly enough, the same method that is used to predistort the input signal to linearize the 2QM happens to correct for temperature drift as well. [Some times, things do work out right.]

The circuit method for linearizing the multiplier and correcting for temperature is known as the "Gilbert Multiplier" after the original developer. This is simply a matter of placing a diode across the input and feeding the input in as a current source. That this will work can be seen from the diode junction equation:



$$I_D = I_{ES}\, e^{qV_D/k_BT}$$

where V_D is the junction voltage, and I_D is the current. (The other symbols are the same as we have been using for transistors.) This can also be written as:

$$V_D = \frac{k_BT}{q}\ \ell n\ (I_D/I_{ES})$$

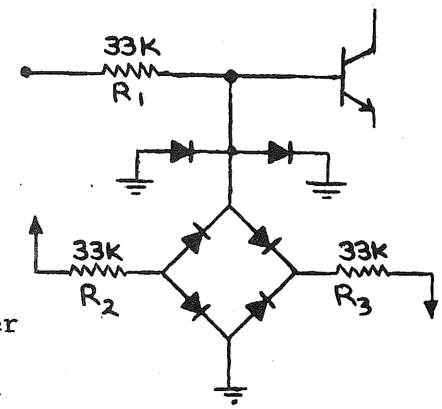
In practical circuits (and monolithic circuits) the diodes will be diode connected transistors, and they are assumed to be matched. The full analysis of the Gilbert multiplier is available elsewhere [e.g., the original: B. Gilbert, "A Precise Four-Quadrant Multiplier with Sub-Nanosecond Response" IEEE J. Solid State Circuits, Dec. 1968, or the Non Linear Circuits Handbook (Analog Devices), pg. 217]. It will suffice here to point out that if we plug the expression for V_D back into the equations for collector currents, we will get terms of the form:

$$e^{(k_BT/q)\ \ell n\ (I_D/I_{ES})\ (q/k_BT)}$$

and it is seen that both the exponential terms and the temperature terms will go out.

The necessary combination of semiconductor elements is not presently available as a single chip, so it is necessary in VCO design to make provisions for an imperfect match. The necessary elements are available as 4QM IC's, but these would have to be converted to 2QM's (see chapter on multipliers). Generally, excellent results can be obtained even with unmatched devices, as long as adjacent pairs are matched.

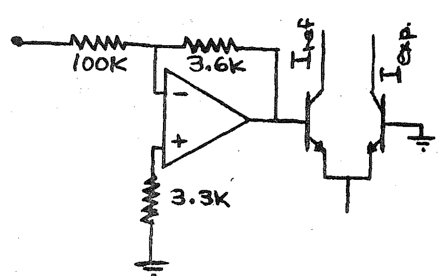
Jung has described a useful presistortion technique [AES Preprint 987 (M-1), Sept. 1974]. The predistortion is obtained from a matched diode array. The diode array replaces the smaller of the two resistors in the input attenuation. A typical circuit is shown at the right. The diodes generate the log function as described above. The diode bridge has an equivalent resistance on the order of 100 ohms, so the 33k resistor (R_1) serves as a current source for the input. With this predistorter, input levels of 50 mV can be used with a total harmonic distortion on the order of 0.25%, as compated to 1% for 10 mV input without the predistorter, and about 5% for 50 mV in without the predistorter. The diode array is available as the type CA3019. This predistorter can be used with either a discrete (matched pair) version of the differential amplifier as described above, or with VCO's fromed with the CA3080 OTA.



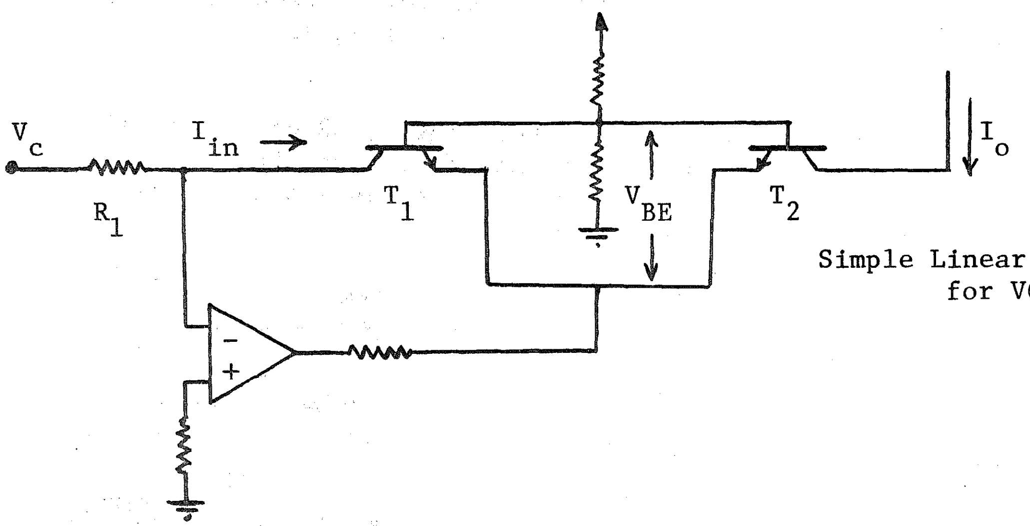
For best results, the resistors R_1, R_2, and R_3 should be matched to 1%.

CONTROLLED CURRENT SOURCES FOR VCA'S

We will discuss first exponential current stages for VCA's since we have studied these extensively in the last chapter. Basically, little is changed here except we need not be so precise, and it is not necessary to make the 1/300°K compensation. The exponential response is usually used as a second type of control input, the linear one being primary. It is of course used where sharper response is desired. The only thing to consider here is the scale factor. We have something like 60 db of dynamic range to divide among the maximum control range. If we take this as a 5 volt signal, then we arrive at 12 db/volt. Since 12 db is nearly a ratio of 4:1, we look for a scale factor that will give a 4:1 change for a 1 volt change of control voltage. We saw in the last chapter that a 2:1 (octave) change was an 18 mV change as far as the converting transistor was concerned. Thus, a 4:1 change is represented by a change of 36 mV at the base of the converting transistor. The input circuit using 100k to a summing node would thus have 3.6k in the feedback circuit as indicated at the right. Note that this is all that is really needed. There is no real need for a trimmer on the scale adjust, and the 3.6k resistor need not be temperature compensating.



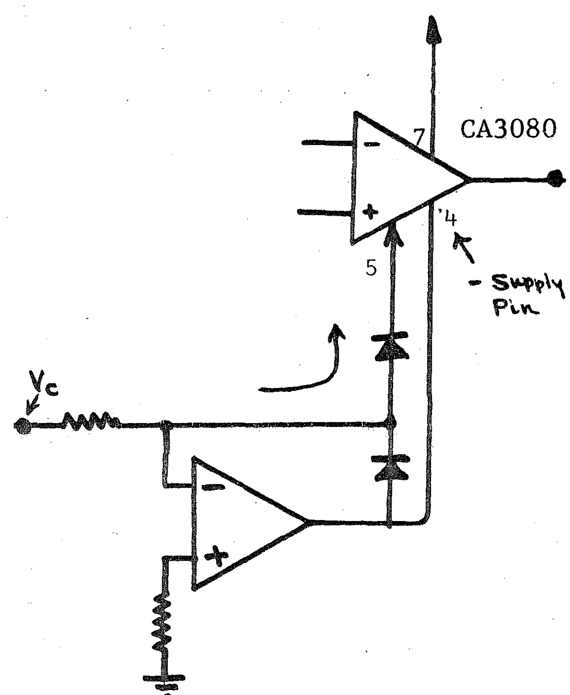
For linear control, the current must be provided by a linear voltage/current converter. Also, we want in general a circuit that increases its output as the control voltage increases. Depending on the required direction of I_o and the structure of the V/I converter, the device may function as a current mirror. A basic structure is shown on the top of the next page.



Simple Linear Current Source for VCA

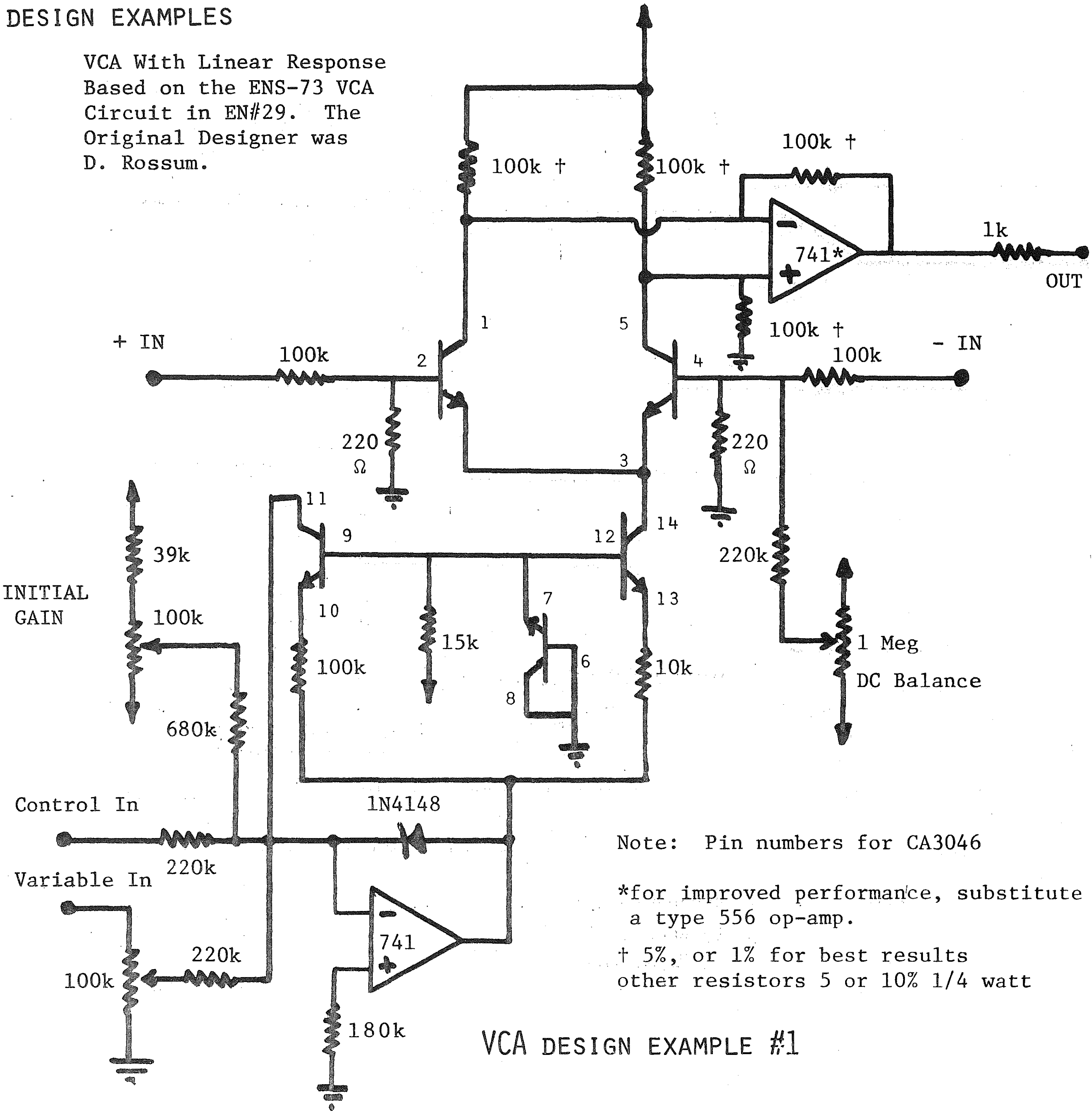
The (-) input of the op-amp is a summing node, so the input current is V_c/R_1. The same current is thus the input current to the collector of T_1. The two transistors have their bases and emitters tied together, so the V_{BE} of both must be the same. Thus, for a matched pair, $I_o = I_{in}$, and a type of current mirror has been formed. At first sight, the structure resembles an exponential converter (as do differential amplifiers in some drawings), but since there is no spread in V_{BE}, it is not. The circuit is from the Signetics application notes on the type 531 op-amp, and the full diagram of a 2QM can be found there. Other refinements that may be found in this type of current source include input protection with a diode, provisions for maintaining a current ratio instead of the mirror, and often the bases are held slightly negative with one or two diode drops to keep the transistors out of saturation. These refinements can be found in the design examples.

A popular type of current source is the op-amp summing node. With this, current is said to flow to a "virtual ground." However, as pointed out in the chapter on op-amps, it is not that the current forces its way in, but that the output of the op-amp draws the current out so that the summing node remains at virtual ground. Thus, the output must be connected to the far side of the load somehow in order to use the summing node as a current source. We have used this principle in exponential converters where the load is a transistor collector and the current is collected through an emitter resistor. Jung has shown [JAES, 23 #3, April 1975 same as AES Preprint 987] that the CA3080 can be driven by a summing node current source by connecting the op-amp output to the negative supply terminal of the CA3080 instead of connecting this pin to the usual -15.



DESIGN EXAMPLES

VCA With Linear Response
Based on the ENS-73 VCA
Circuit in EN#29. The
Original Designer was
D. Rossum.



Note: Pin numbers for CA3046

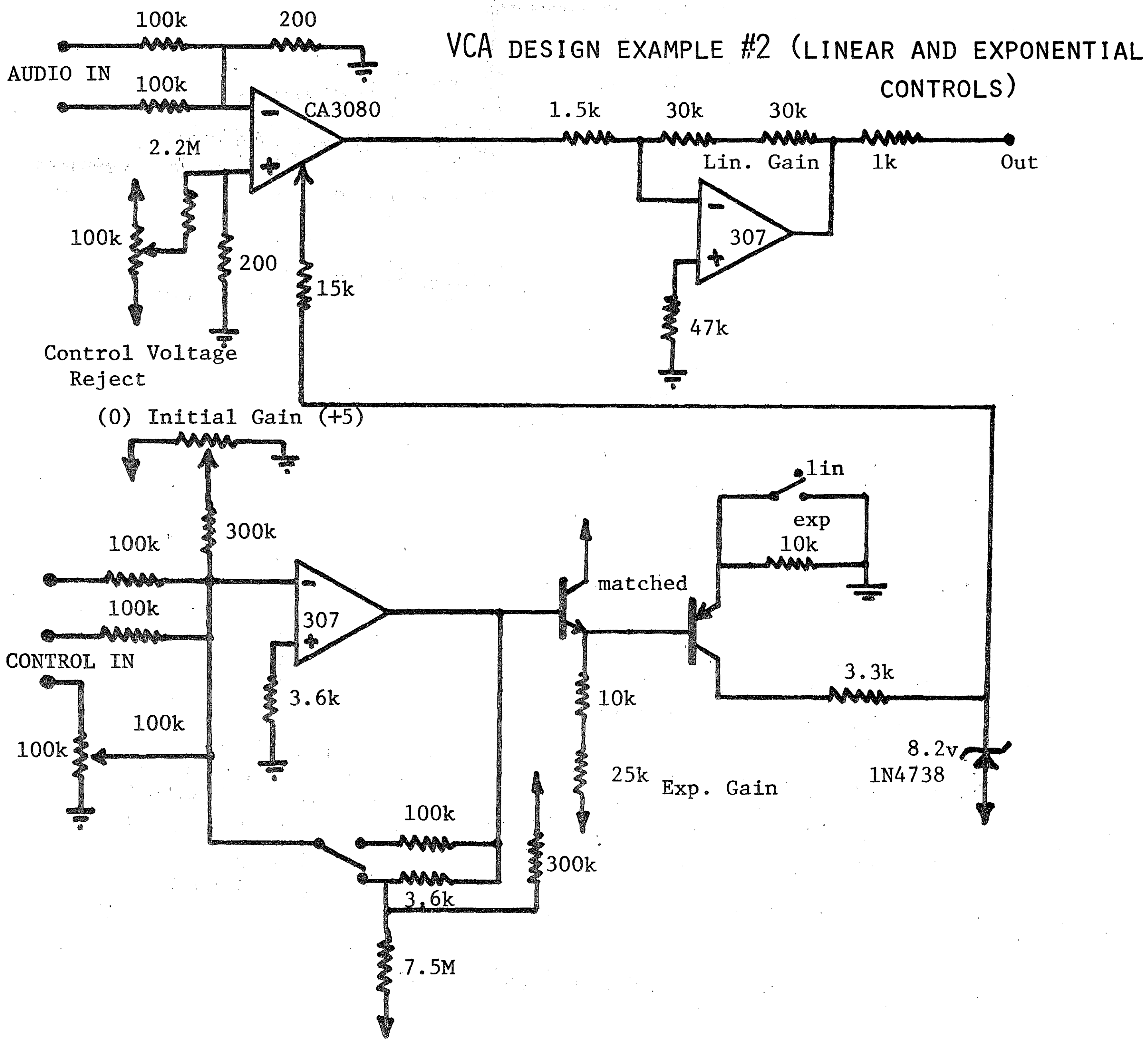
*for improved performance, substitute a type 556 op-amp.

† 5%, or 1% for best results
other resistors 5 or 10% 1/4 watt

VCA DESIGN EXAMPLE #1

The VCA in the first design example includes the basic two quadrant multiplier that was discussed in detail above, and the linear current source with the refinements mentioned. The circuit has proven very reliable in several applications. Control voltages can sum to +5 to give unity gain. The DC balance trimmer can be adjusted by applying an audio signal to the control input and adjusting for minimum control signal feedthrouth in the output. The 1N4148 diode protects the control input from a voltage reversal. The diode connected transistor (pins 6,7, & 8) is used to drop the base voltage slightly negative. Some designers use a second series diode here, and this can be added if desired. The diode bridge structure could be used as an input attenuator if desired, but for a general purpose VCA, this is not necessary.

VCA DESIGN EXAMPLE #2 (LINEAR AND EXPONENTIAL CONTROLS)



The second design example is from EN#34, a circuit designed by T. Mikulic. This circuit may be switched for either linear or exponential response. The VCA has a dynamic range up to 100 db. The linear response is (control voltage/5) and the exponential response is 12db/volt. When the sum of the control voltages is +5, the gain is 1 in either control mode. To adjust, start with the linear mode. Set "initial gain" to "+5" position. Adjust lin. gain so that input and outputs are equal. Set to exp. mode and adjust exp gain for the same level of output. Finally, remove the input signal from the audio inputs and connect to a control input. Adjust the control voltage reject trimmer for minimum feedthrough.

Chapter 5D

VOLTAGE-CONTROLLED FILTER DESIGN

CONTENTS:

INTRODUCTION:

The VCF is probably the most important part of the standard synthesizer - at least to the extent that we wish to provide new and unique sounds. A great deal of variety is possible in VCF's (more than is used). The basic design philosophy for VCF's is to first select a filter design that determines certain parameters of the sound that passes through it, and secondly to try to realize the filter with voltage-controlled elements. For the first step, the whole field of active filter technology is available. For the second step, the designer must select a control element that will work in the filter configuration. Control devices in common use are:

1. The dynamic resistance of semiconductor junctions
2. FET's as voltage-controlled resistors
3. Multipliers (two and four quadrant types)
4. OTA's in various configurations
5. Analog switches with ultrasonic switching rate and variable duty cycle.

The design of a complete VCF thus falls into the following steps:

1. Selection of an active filter design
2. Applying voltage-controlled elements to control the filter
3. Addition of the proper voltage-control input section to control all elements from a single exponential source.

Of the three steps, we will be mainly dealing with only the second. The first step goes back to the chapter dealing with active filters, and in any case, the basic active filter selected is often the one that lends itself most easily to the voltage control devices we have in mind. This in part (along with the musical usefullness) accounts for the popularity of the state variable filter designs and the 4-pole low-pass and high pass designs. The third step is very similar to the VCO case as far as the actual input structure and exponential current source is concerned. The only new feature here is the the need to control simultaneously several control elements from the same exponential source. We are thus left to take a careful look at control elements for step two of the design. The design examples cover the design of a state variable filter and the four pole low-pass (as the EN reprint). This reprint from EN#41 contains some extensive information on the overall design process and active filter analysis techniques as well.

CONTROL ELEMENTS FOR VCF'S

A filter structure based on the dynamic resistance of semiconductor junctions was developed by Moog [A Voltage-Controlled Low-Pass High-Pass Filter for Audio Signal Processing, AES Prperint #413, 1965]. The basic principle is that the dynamic resistance is proportional to the standing current in the junction. In the study of VCO's and VCA's we frequently used the semiconductor junction equation

$$I_c = I'e^{qV_{BE}/k_BT}$$

This equation can be easily differentiated to give:

$$dI_c/dV_{BE} = I'e^{qV_{BE}/k_BT}(q/k_BT) = \frac{qI_c}{k_BT}$$

The "dynamic resistance" is just the reciprocal of the derivitive (from Ohm's Law, R=V/I and dynamic resistance = dV/dI).

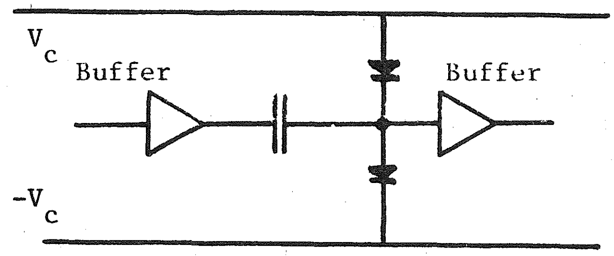
$$R_{dynamic} = (\text{constant}) \cdot \frac{1}{I_c}$$

The dynamic resistance is thus inversely proportional to the standing current and over a range comparable to the range of exponential accuracy of the junction.

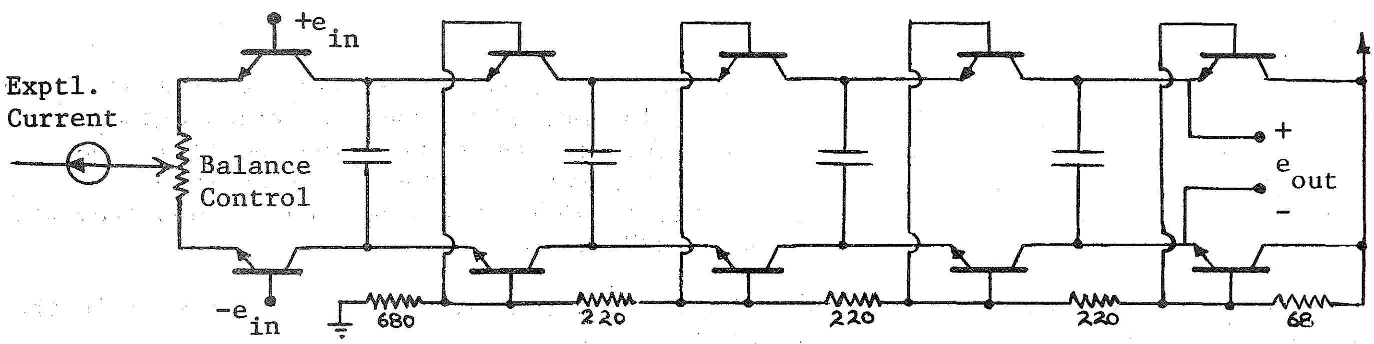
A few words on the significance of dynamic resistance and its utalization are in order. We see that by setting the standing current, we set the junction at a certain point on the exponential curve. For small excursions about this point, the variation of current with voltage is approximately linear, and looks like a resistance with slope dV/dI. Thus, a small AC component will see the effective resistance dV/dI. Changing the standing current thus gives a different slope on the V-I curve, and hence a different dynamic resistance.

Since we must work with small excursions in the presence of large standing values, we must work with a balanced driving signal and recover the signal by a differential means. Low-pass and high-pass filter structures can be easily implemented in this way as long as matched junctions are used and a certain amount of trimming is made.

In the realization of the high-pass filter as described by Moog, the diodes that serve as the dynamic resistors also do their own exponential conversion. This is a relatively simple matter of providing a balanced control signal across the diodes as shown for one "RC" section at the right. For matched diodes, the center point is at zero volts in the absense of an applied signal. The R section is actually the dynamic resistance of the two diodes if they were in parallel.



The "Moog Ladder" structure for the Low-Pass section is indicated below:

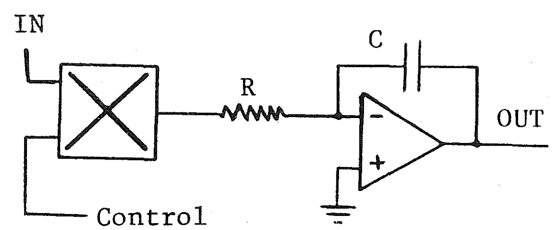


The current from the exponential converter is divided between the two sides of the ladder and passes through all the transistors on one side. The two transistors on the left side modulate the currents when a signal to be filtered is applied. The signal is removed from the far end of the ladder by differentially removing the signal from the emitters of the transistors on the right.

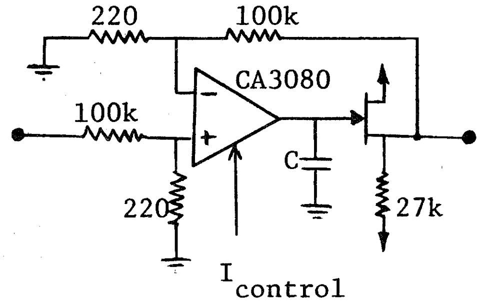
A simple low-pass, band-pass, high-pass structure using dynamic resistance of diodes was given by Steiner in Electronic Design #25, Dec. 6, 1974.

A second type of control element is the FET. The basic idea is to control the source-to-drain resistance with the gate voltage. Filters realized with this technique are generally limited by the linearity of the voltage-resistance curves of FET's, and by limited range. For this reason, they are unsuitable as general purpose electronic music filters, but may be useful as building block circuits in other surutures. Circuits of this type may be found in: N. Doyle, "FET and Op-Amp," Radio-Electronics July 1970; V. Georgiou, "Voltage-Tuned Filter Varies Center Frequency Lineraly," Electronics, Nov. 6, 1972, and D. Malham, "Voltage-Controlled High-Pass Filter," Wireless World, July 1972.

A third type of control element is the analog multiplier. This is applied by using the multiplier to effectively control a time constant of the filter as is indicated by the implementation of a voltage-controlled integrator at the right. Depending on the multiplier used, the linearity can be quite good, but the dynamic range is seldom better than 100:1 at best. To be really useful, the range must be increased with a two-quadrant multiplier of the OTA as described below.



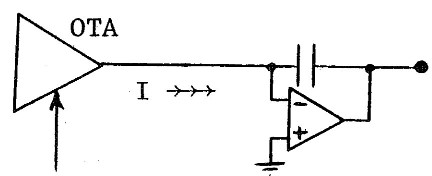
The realization of a voltage-controlled filter by placing a reactive element in the feedback loop of a voltage-controlled amplifier was suggested by Moog [JAES 13 #3, pg. 200 July 1965]. The amplifier can be a two quadrant multiplier, or an OTA if the current output is to be used. When we say "reactive element" it actually means a capacitor since inductors are out for active filters at audio frequencies. The OTA is ideal for this application since it has good linearity and a wide range. The realization of a voltage controlled low-pass section is a simple matter of dumping the OTA's current output into a capacitor and then buffering the capacitor voltage, and feeding it back. This effectively places the capacitor in the feedback path. The circuit is a voltage controlled RC section. The reprint article that is part of this chapter will show how this can be expanded into a more complex four pole low-pass filter.



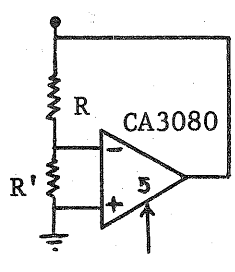
If the feedback connection is removed from the above section, a voltage controlled integrator results. The circuit is unstable however, and must be used in a closed loop system such as a state variable filter to make it useful. Such a filter is covered in the second design example in this chapter.

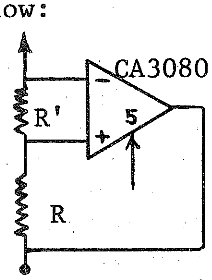
In other filter structures, the OTA may serve as a voltage-controlled resistance or conductance. The ability of the OTA to do so is implied by the term transconductance. Basically, the OTA puts out a current proportional to the voltage applied to it and scaled by the bias current that controls it. However, the OTA is not by itself a voltage-controlled resistor since it is not bilateral - the current only flows from one "end" in response to a voltage at the other "end".

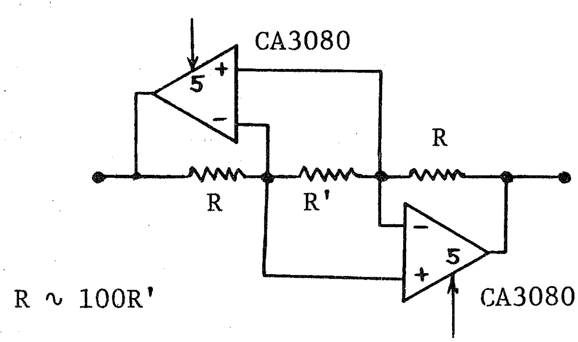
Note first that the OTA effectively serves as a voltage-controlled resistor in the integrator circuit above. This can perhaps be made clearer by considering how the OTA can be used to drive an operational integrator as shown at the right. The op-amp summing node receives the current output of the OTA.



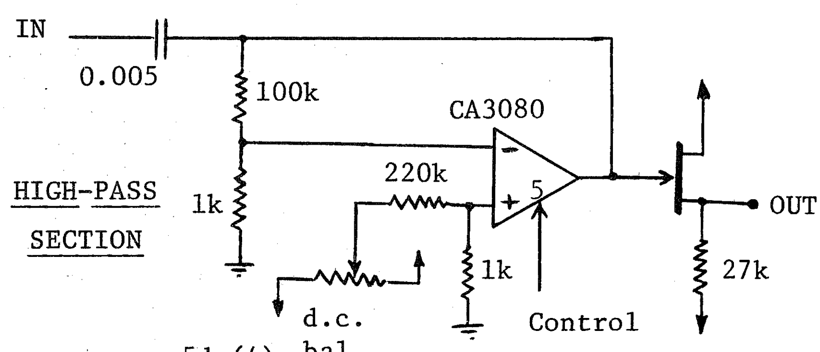
Wilcox (EN#40, pg.7) has shown how the OTA can be used as a voltage-controlled resistor referenced to ground, power supply, or back to back as a floating resistor. The circuits are shown below:



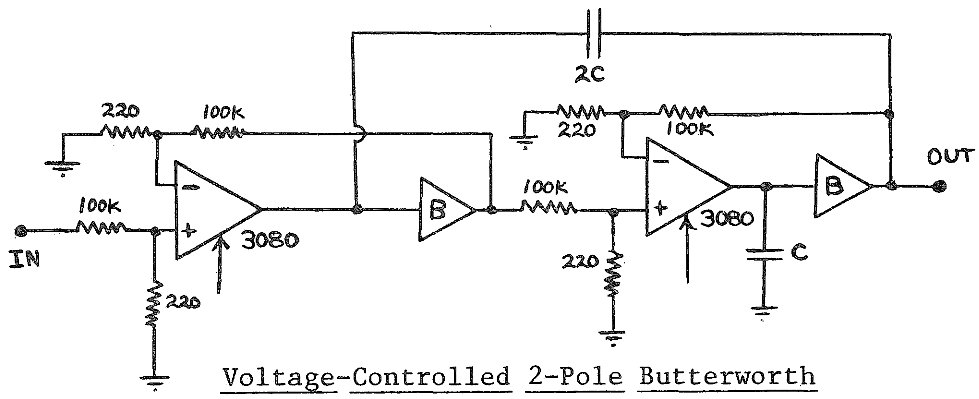




The voltage-controlled resistor can be used in several filter structures, for example the high-pass section shown at the right. The section is equivalent to a single high-pass RC section, and gives 6 db/octave roll off (roll-up in this case). Sections may be cascaded for a sharper cutoff slope.

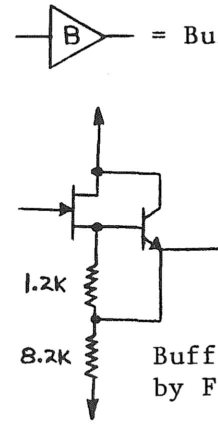


In general, the OTA can be inserted as a VCR in many filters to control them. A good example is the two pole low pass Butterworth (using capacitors C and 2C) as described in the chapter on active filters. The VCF version has been implemented by Franco [Hardware Design of a Real-Time Musical System, U. of Ill. Computer Science Report UIUCDCS-R-74-677 (1974) pg. 51]. A simplified diagram of this implementation is shown below:



Voltage-Controlled 2-Pole Butterworth

(after Franco)



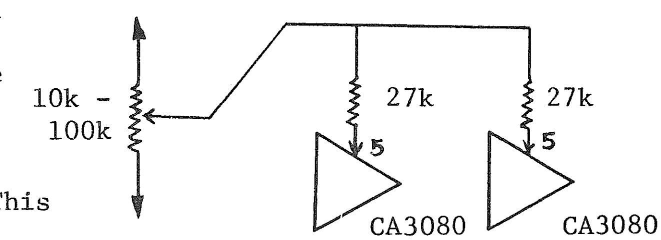
Buffer used by Franco

The final control element we will be discussing is the analog switch. The basic idea is to turn the switch off and on very rapidly (relative to audio frequencies) and vary to duty cycle. This has the effect of controlling the value of a series resistor or capacitor in proportion to the duty cycle. For electronic music work, it would probably be necessary to make the duty cycle (pulse width) an exponential function of control voltage. For examples of the analog switching method, see: T. Vestergaard, "A Regulated Low-Pass Filter Based on Pulse Width Modulation," AES Preprint 923 (L-3) 1973, or B. Wilkinson & T. Marshall, "International Music Synthesizer," Electronics Today International, June 1974. Control range on these devices is on the order of 100:1.

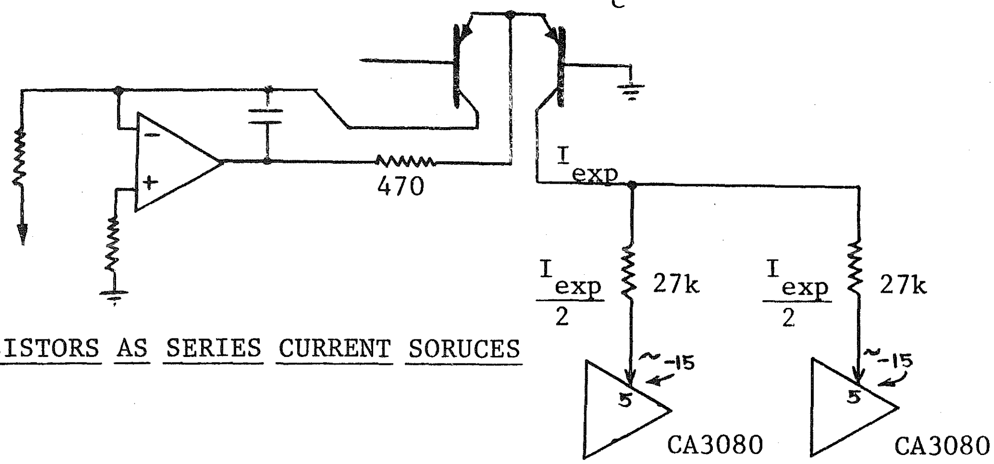
ADAPTING CURRENT SOURCES FOR VCF'S

For electronic music work, the VCF's should generally have the same response as the VCO's in the system, 1 volt/octave. Thus, the exponential converters used are basically the same, and the only additional design problem is the distribution of the control current when it must service more than one control device.

We can start by examining the usual test setup for the OTA circuit - feeding a current from the wiper of a pot connected between ±15 through a 27k resistor to pin 5 of the CA3080. We can thus control several CA3080's in parallel as shown at the right. First of all, we will assume that the CA3080's are matched. This may not be true, and in cases where it is necessary, CA3080's can be matched by setting up a simple gain test circuit. This is mainly important for four-pole filter designs since each stage acts independently. In state variable designs, the response goes as the product of the gains and the two work together. Once we are satisfied with the match, we can set up the test circuit above and by varying the pot setting, put the CA3080 through its useful range. Note that pin 5 of the CA3080's remains very close to negative supply at all times. This means that pin 5 is a very low impedence compared to 27k and the resistors act essentially as current sources.



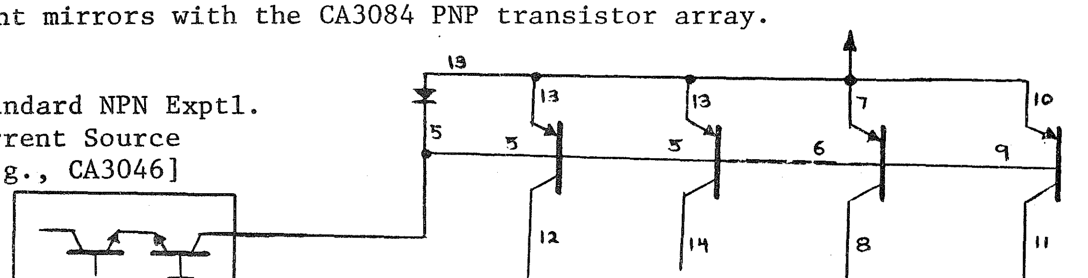
We want in general to drive the CA3080's with the current from an exponential converter. We can either divide the current among the CA3080's, or use a multiple controlled current source to supply the exponential current to CA3080's in a parallel arrangement. If we consider a typical PNP exponential converter as outlined below, we see that the current actually originates from the output of the op-amp which supplies current to the emitters of the PNP pair. This must be a potential that is positive, and recall that the pin's 5 of the CA3080's are near -15. Thus, the 27k resistors are still acting as current sources, and since the collector supplies a current I_c to two (virtually) identical



USE OF RESISTORS AS SERIES CURRENT SORUCES

circuits, the current is divided equally between the CA3080's. This is a particularly simple mehtod that is useful in a state variable design.

For more precise results, the OTA's can be driven form a multiple current mirror by a method suggested by Sergio Franco. In this way, we can use an NPN exponential converter (which as we discussed in the VCO chapter, has certain advantages), and realize the current mirrors with the CA3084 PNP transistor array.

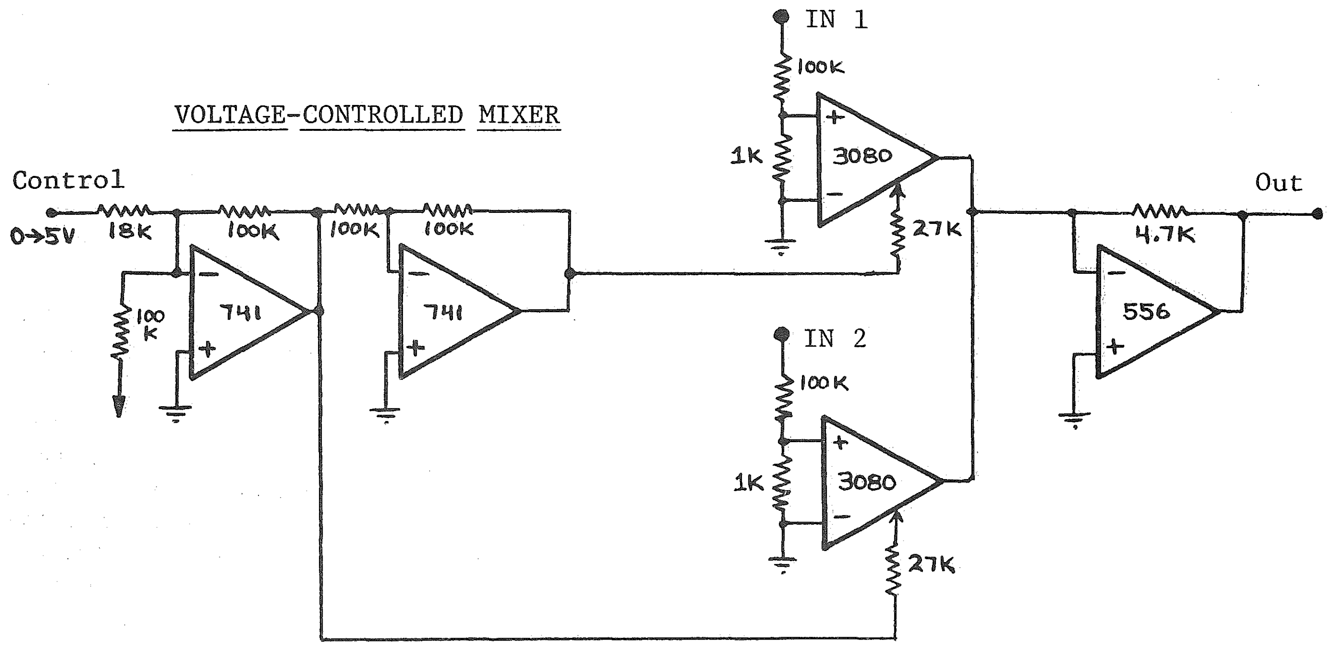


Pin Numbers for CA3084

VOLTAGE-CONTROL OF OTHER FILTER CHARACTERISTICS

Characteristics other than the filter characteristic frequency may lend themselves to voltage control. This may be a simple matter of making some resistor(s) in the filter voltage-controlled resistors. A good example of this is voltage-controlled filter Q. In the state variable filter for example, the filter Q is controlled by a single feedback resistor to a summing node. This can easily be voltage-controlled by using an OTA to deliver a current directly to the op-amp summing node. This implementation is shown in the second design example.

Another interesting means of controlling the characteristic is to employ a voltage-controlled mixer to combine outputs from different filters or filter sections. The mixer may be a part of the actual filter, or it may be an external device. However, it is necessary to realize that when the outputs of two filters are mixed (summed or subtracted), the output is not the resulting sum of the frequency responses in the areas where the two responses overlap. This is a simple matter of considering the phase of the two outputs. Consider for example that if the responses of the low and high-pass sections of a state variable filter were summed, only a small notch would result; it is the fact that the phases cancel that gives a sharp notch in the center. Thus, we must either do a careful phase analysis or simply do an experiemnt to see how the total response curve changes with the mixture. The voltage-controlled mixer section outlined below responds to an envelope varying from 0 to +5 volts. The outputs of the two op-amps vary from +15 to -15 for the first, and -15 to +15 for the second. Thus, the control currents to the two OTA's also vary with one rising while the other falls. The outputs of the OTA's are connected together and fed to a current-to-voltage converter, the final op-amp. Thus, the mixture varies with the envelope voltage:



DESIGN EXAMPLE #1; REPRINT: A FOUR-POLE VOLTAGE-CONTROLLED NETWORK; ANALYSIS, DESIGN, AND APPLICATION AS A LOW-PASS FILTER AND A QUADRATURE VCO: by Bernie Hutchins [reprinted from EN#41]

The op-amp inverter circuit shown in Fig. 1 should be a familiar device by now. It inverts the polarity of the input, and scales it by a factor of R_f/R_1. With the input grounded, the output remains very close to ground as well, but what happens when we replace the resistor R_f with a capacitor C as in Fig. 2? Initially, current V_1/R_1 which flows to the summing node (the - input junction point) would normally go through R_f. Since R_f is replaced by C, the output current must flow into C, thus charging it. The output of the op-amp will change as is necessary to maintain the differential input voltage at zero (i.e., the summing node at virtual ground). The output thus goes from zero toward -15, but since this is the supply voltage, it can go no further than -15. At this point, the feedback quits, and the differential input voltage starts to go positive giving the op-amp a second good reason for remaining at -15 volts output. Finally, the - input will reach the input voltage $+V_1$, the capacitor is charged to a total voltage of $15 + V_1$, and the input current stops.

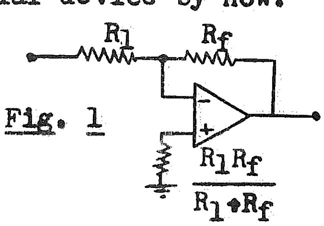


Fig. 1

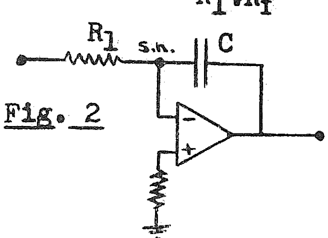


Fig. 2

You might expect that if you connected the input to ground, the output would remain at ground. In the case of a real op-amp, this is not true because the differential input voltage and/or the bias current to the - input will cause the capacitor to slowly charge, as these values will never be zero. The circuit in Fig. 3 will illustrate this nicely.

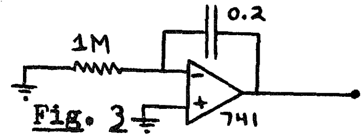
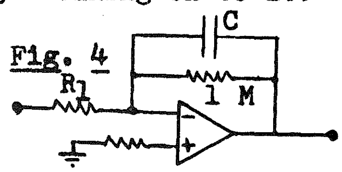


Fig. 3

The circuit with the capacitor in place of R_f is of course the usual op-amp integrator circuit, and we have been using this as a filter, so we are most interested in its AC characteristics. So why all the worry about the DC characteristics above? Simply because even small DC errors may cause these integrators to charge up against the supply or close to it, causing clipping and/or complete cutoff.

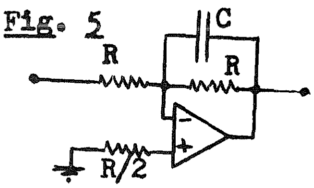
The traditional means of stabilizing the DC characteristics of the integrator is to effectively make a leaky capacitor to drain off any charge leaking on to it. This can be done with the circuit in Fig. 4.

Fig. 4



We can next consider the circuit of Fig. 5 where we use both a capacitor and a resistor in the feedback circuit. At DC and very low frequencies, the capacitor's reactance is much more than the resistance R, and the magnitude of the gain at low frequencies is thus just 1, as the capacitor is effectively out of the circuit. At very high frequencies on the other hand, the capacitor's reactance will be much less than R, and the resistor is effectively out of the circuit. The gain is therefore $X_c/R = 1/2\pi fRC$, and the thing starts to look like a filter as the gain (transfer function) depends on frequency. Since the response falls off as 1/f, we can see that for a one octave change, the gain drops by $\frac{1}{2}$. This is, in decibles, $20 \log_{10} \frac{1}{2} = -6$ db/octave. Since gain decreases with frequency, it is a low-pass filter. For reasons which we will consider later, it is called a single pole filter.

Fig. 5



We have seen very often in EN that integrators have been used to form filters. This is the two-pole design using two integrators known as the state variable, the universal active filter, the biquad, and by a variety of other names. In this circuit, DC stability is achieved through the use of the inverter. Since we had good results making voltage controlled filters by this means using voltage controlled integrator sections (with multipliers, or with CA3080 OTA's), we want to consider using these voltage controlled integrator sections for higher order filters. First, we have to take a look at some filter theory as it applies to the integrator.

We must first become familiar with the "s-plane", which is the "map" on which filters live. The quantity "s" is a complex frequency. We normally think of frequency as existing on a scale from zero to ∞ along one axis, On the s-plane, frequency has both a real, and an imaginary (in the mathematical sense) aspect. In fact, the useful physical properties of the filter will be derived from a frequency response above the s-plane while following a path defined by the <u>imaginary</u> axis. This is a good example of the fact that there is really nothing very imaginary about the so called imaginary numbers. Their main use in filter theory is to keep track of phase information automatically. In order to appreciate why we want to look at filter response above a plane, rather than just along the one line of interest, consider the following: Suppose you are to travel along a straight line on a map, and there are no features marked along this line. To get an estimate of what your journey may be like, you would look for features near your line of travel. For example, a mountain peak near your path will most likely indicate some high ground will be encountered along your path. On the s-plane, the straight line path you inspect is the positive imaginary axis, the elevation of the path is the filter's frequency response, and the mountains are the "poles" of the filter.

Consider for example the simple integrator if Fig. 2 as a filter. The gain of the integrator is $-1/2\pi fCR$. In filter terminology, this gain is called a "Transfer Function" which we will denote by T, and we will write this in terms of the complex frequency s as $T(s) = -1/sCR$. Note that s is generally written as $\sigma + j\omega$ where σ (sigma) is the real part, and ω (omega) is the imaginary part, and $j = \sqrt{-1}$.

The most important question to ask about T(s) is: where does it blow up. That is, where does it become infinite, or equivalently, where does the denomenator become zero. For the integrator, T(s) becomes infinite when $s \rightarrow 0$. Thus, if we plot T(s) over the s-plane, we get a steep mountain around $s = 0$ as shown in Fig. 6. The mountain is called a pole. The positions of the poles of a filter are important in the sense that they greatly influence the contour of T(s) over the entire s-plane. In particular, we examine the "height" of T(s) above the $j\omega$ axis but pole position is also important in determining the stability of a filter. If the poles are located on the $j\omega$ axis or if they move into the right half of the s-plane (positive values of σ), instability (oscillations) may occur. This is the case of a filter looking so hard for a certain signal that it makes its own! Active filters are those which use an amplifier stage to position the poles in the desired positions on the left half of the s-plane. Books on active fulters often indicate the pole positions by a cross mark (×) rather than by the mountains we have used.

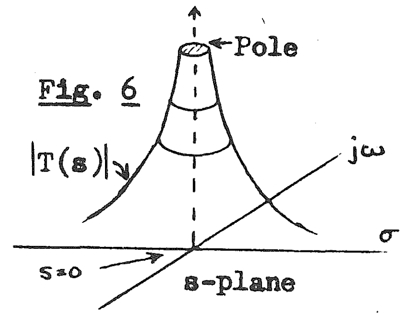


Fig. 6

Note that the mountain gets steeper and steeper as you get closer and closer to $s = 0$. It becomes a thin rod (which may be the origin of the name "pole"). In our practical circuit however, the pole has a flat plateau as shown due to the finite voltage of the power supply. It is also to be observed that this pole might be a deep hole if we are looking at T(s), but we generally look at $|T(s)|$, which is in the case of $T(j\omega)$: $[T(j\omega)\cdot T(-j\omega)]^{\frac{1}{2}}$. At $s = 0$, the response becomes infinite, or we might say that the circuit "oscillates" at zero frequency (DC), and this is what we observed when we saw that the circuit of Fig. 2 responded all the way to the power supply voltage.

Traveling along the $j\omega$ axis, we see the response falling off as 1/f which we have identified with a 6 db/octave drop. This can be drawn as in Fig. 7, but usually is drawn on a log-log plot as in Fig. 8, where the 6 db/octave response becomes a slope of -1. The log-log plot is usually used for filter frequency response.

Fig. 7

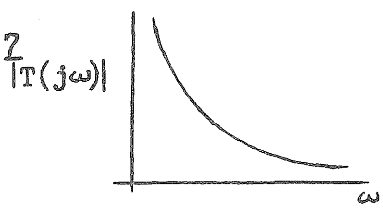
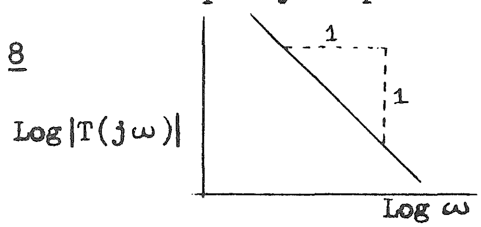


Fig. 8



Suppose for the moment that we have somehow avoided the DC stability problems, how close is the op-amp circuit of Fig. 2 to an ideal integrator, and does it look anything like what we would consider a useful filter? These characteristics are displ yed in Fig. 9. The interested reader is directed to a separate reference[1] for a discussion of the exact features of the op-amp integrator curve. Note the similarity of this curve to the "open-loop gain curves" of certain internally compensated op-amps such as the 741. This is no coincidence, as such op-amps are actually compensated with a single pole.

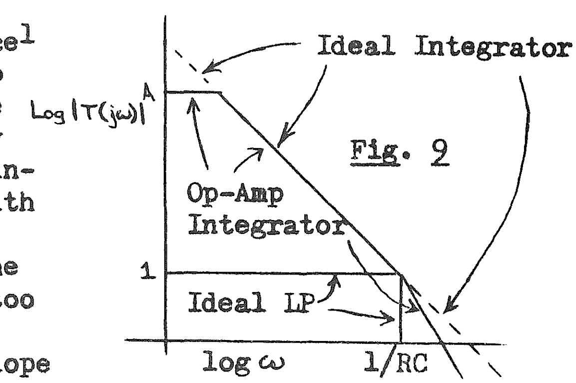


Fig. 9

The op-amp integrator doesn't look much like the ideal low-pass filter as you can see: T(s) is way too high at the low frequency end, and doesn't fall off vertically as the ideal low-pass does, but with a slope of -1. Let's look at a means of improving the fall off first.

If we cascade two or more stages like Fig. 2, we get the product of the transfer functions of the individual stages. For two stages this becomes $T(s) = 1/s^2R^2C^2$ and T(s) falls off as $1/s^2$ (12 db/octave). There are now two poles of T(s), both of which are at $s = 0$, making this what is called a "second-order pole". Likewise, we can cascade three integrators for 18 db/octave (3-pole), and four integrators for 24 db/octave (4-pole).

We have already seen how to bring the low frequency gain down - use the circuit of Fig. 5. The transfer function for this is:

$$T(s) = \frac{-\text{Feedback}}{\text{Input R}} = \frac{-\text{Parallel R and C}}{R} = \frac{\frac{-R/sC}{R + 1/sC}}{R} = \frac{-1}{1 + sRC}$$

This has a pole where the denominator of $T(s)$ becomes zero: when $s = -1/RC$. Thus, we have moved the pole from $s = 0$ to $s = -1/RC$, thus pushing it back down the negative sigma axis. This is equivalent to moving the mountain back from the $j\omega$ axis. Fig. 10 shows this, and Fig. 11 shows a log-log plot of the response along the $j\omega$ axis.

Fig. 10

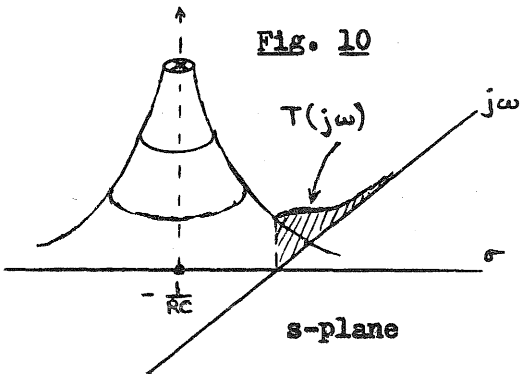
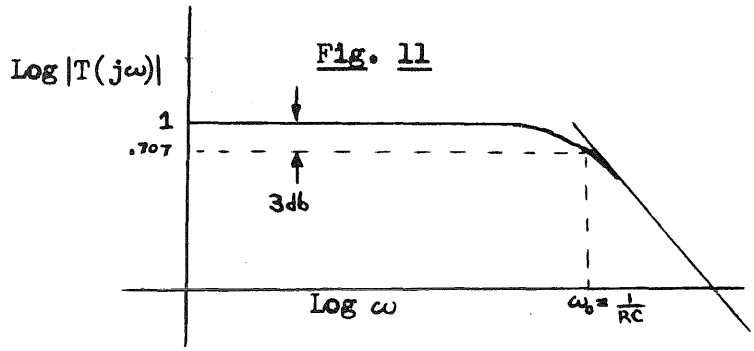


Fig. 11



At $s = 0$, $|T(s)| = 1$, and from that point on, it begins to fall, slowly at first, and finally at 6 db/octave. We can now sharpen this drop off as we suggested before by cascading stages, adding more poles. Fig. 12 shows experimental data taken on 1, 2, 3, and 4 stages (1, 2, 3, and 4 poles at $s = -1/RC$).

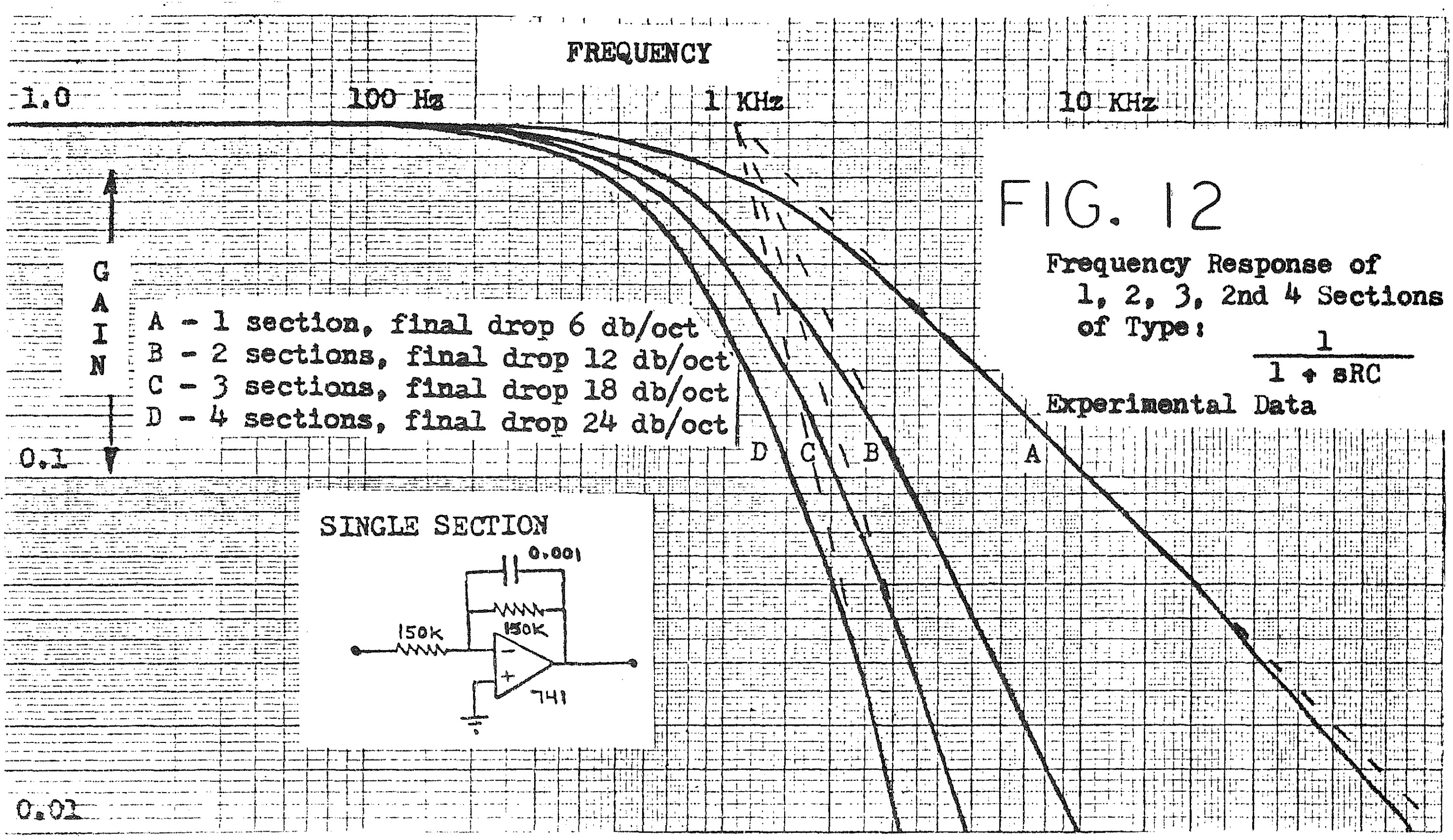


FIG. 12

Frequency Response of 1, 2, 3, 2nd 4 Sections of Type: $\frac{1}{1 + sRC}$

Experimental Data

These are fairly good approximations to the ideal low-pass filter as we go to higher and higher order, but it should be pointed out that this is not the best that can be done. The best approximation (while keeping the pass band flat) is the so called Butterworth response or maximally flat response. The single stage used in Fig. 12 happens to be a single pole Butterworth filter. Cascading two first-order stages increases the number of poles, and makes the filter second-order, but <u>does not</u>

preserve the Butterworth characteristics in the overall response. Two first order stages give:

$$T(s) = (1/1+sRC)(1/1+sRC) = 1/(1 + 2sRC + s^2R^2C^2)$$

$$= \frac{1}{\frac{s^2}{\omega_0^2} + 2\frac{s}{\omega_0} + 1} \quad \text{where } \omega_0 = 1/RC$$

which can be compared with the second order Butterworth response of:

$$T(s) = \frac{1}{\frac{s^2}{\omega_0^2} + \sqrt{2}\frac{s}{\omega_0} + 1}$$

To get a true Butterworth response, the stages must be designed to work together to give the correct transfer function.[2] A comparison of a 4-pole Butterworth and the corresponding 4-pole, 4-sections damped integrator response is shown in Fig. 13. The Butterworth has a sharper corner. Fig. 14 shows the poles for the cascaded sections (all four poles piled up) while Fig. 15 shows the poles for the Butterworth.

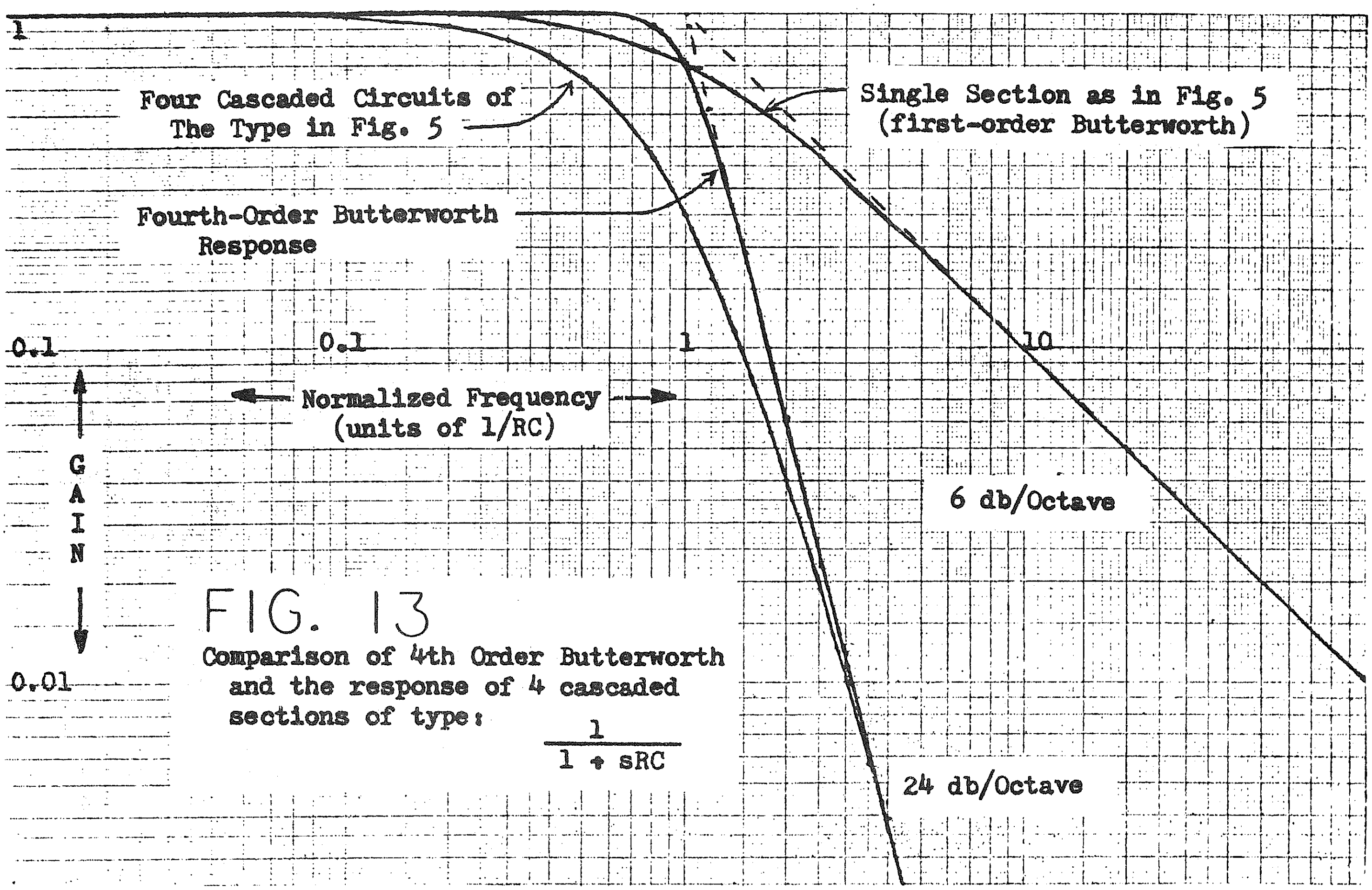


FIG. 13
Comparison of 4th Order Butterworth and the response of 4 cascaded sections of type: $\frac{1}{1 + sRC}$

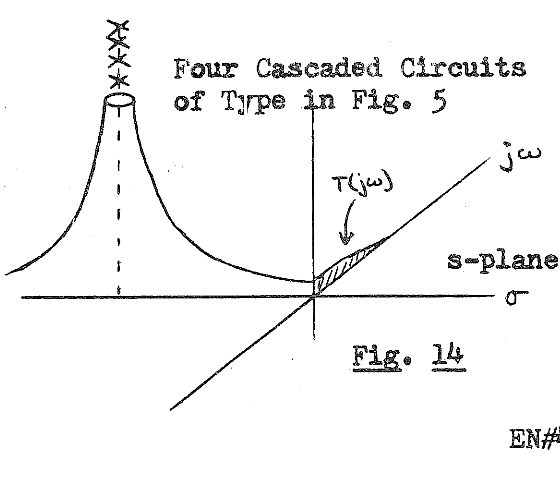


Fig. 14

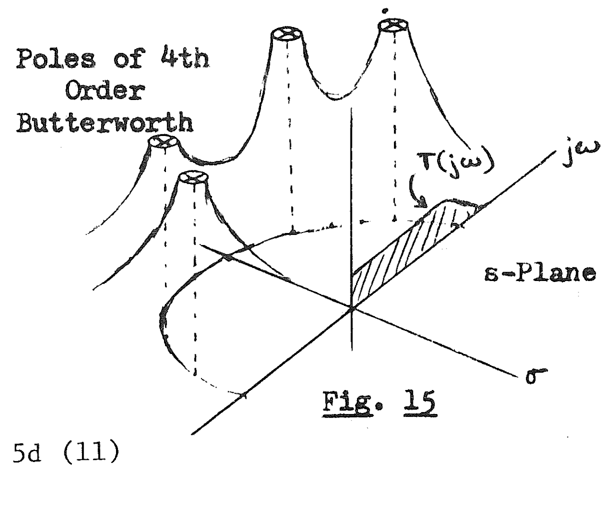


Fig. 15

Moving the poles around toward the $j\omega$ axis and away from the σ axis causes higher ground to appear at the middle and corner of the response curve, thus making the pass band flatter, and the corner sharper. In any event, we will select the 4-pole cascaded sections, as this is easier to do in a practical circuit. The difference between using four cascaded sections of type $1/(1 + sRC)$ and the Butterworth may not be important musically, especially when we consider that the filter is often used with feedback to produce corner peaking. A successful electronic music VCF[3] does in fact use the cascaded section response.

Adding Feedback:

It is a usual practice in filter design (called the leapfrog approach) to consider feedback from different stages. In the present case, we want to feed the output of the last stage back to the input. Here, we must consider not only the magnitude of signal fed back, but the phase as well. The phase response of the single section is shown in Fig. 16. At low frequency, the phase of four sections adds up to 0°. At very high frequencies, the phase adds up to 90° + 90° + 90° + 90° = 360° = 0°. At $\omega_0 = 1/RC$, the phase is 45° + 45° + 45° + 45° = 180°. So by adding an inverter to the output, and feeding it back to the input, we can enhance frequencies around 1/RC while further rejecting lower and higher frequencies. The result is corner peaking and is illustrated in Fig. 17. A control pot can be added to the feedback to control the sharpness of this peak. This control then serves as the "Q" control for the filter.

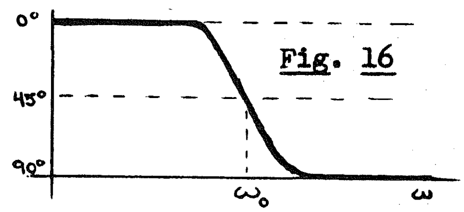


Fig. 16

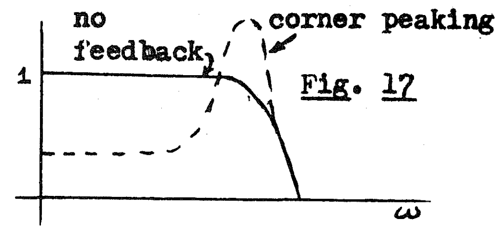


Fig. 17

Making it Oscillate:

When we put a sine wave into the filter and adjust the frequency to 1/RC, each stage decreases the amplitude by $1/\sqrt{2} = .707$. After four stages, this will be down to ¼ the input. If we now replace the input sine wave with the output amplified by a factor of four, the filter will oscillate. In practice, we are inverting the signal first, either by a separate op-amp or using the opposite input if available, and changing the feedback resistor to ¼ the normal input value. For the experimental network of Fig. 12, an inverter is added to the output, and a resistor of about 37k replaces the 150k input resistor. The frequency of oscillation is 1/RC, and the successive stages produce 45° phase shifts, so we can take signals from the second and fourth stages for example to get a quadrature oscillater. The fourth stage is at ½ the amplitude of the second, but this overall filtering is useful for the following reason. If the feedback is too small, oscillation will die out. Making the feedback larger will cause the first stage to clip at the power supply slightly. We want to run this first stage at a slight clip to keep oscillation stable, and thus the second stage will filter out this clip and give a signal level of about 10.6 volts. The fourth stage will have a very pure sine wave at a level of about 5.3 volts. In practice, all that we need to do to make the filter oscillate is to increase the Q control to the point where oscillation just starts.

The Complete Voltage-Controlled Network:

Voltage controlling a four stage network is now a simple matter of using voltage controlled integrator sections and feeding back each stage for unity gain at DC. This stage was first introduced by Terry Mikulic in EN#33 (5) and the modification for unity DC gain is shown in Fig. 18. We will use the exponential converter and summing stage from the VCF in EN#37 (6). The complete filter is shown in Fig. 19. The total range as drawn is 1 Hz to 22 KHz as either a filter or an oscillator. If you need more low range and are using the network mainly as a quadrature oscillator, you may want to change the capacitors from 150 pf to 1000 pf.

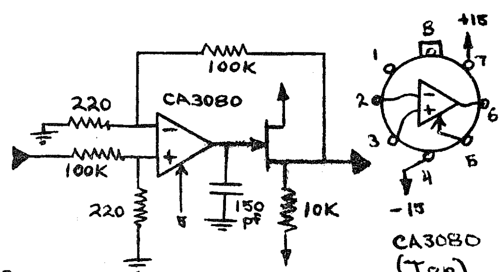


Fig. 18

In Fig. 19, the filter output can serve to take out one of the quadrature signals when used as an oscillator, as the 5.3 volt level should be close enough to the 5 volt levels of the ENS-73. The other quadrature signal from the second stage can be reduced to a 5 volt level with 1000 ohms output impedence by using a voltage divider of two 2k resistors as shown. If you desire to use the quadrature oscillator as a circular location modulator, you may want to obtain the inversions of the quadrature signals as well. A set of four op-amps can be used for this as shown in Fig. 20. This could be simplified a little, but this circuit with four op-amps allows accurate settings of signal levels. For more information on location modulation, see EN#33 and section 8a below.

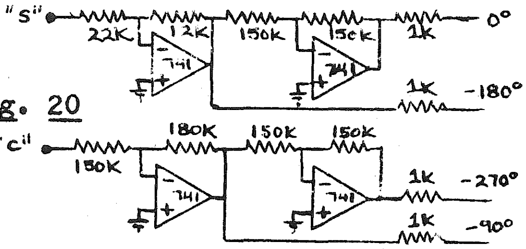


Fig. 20

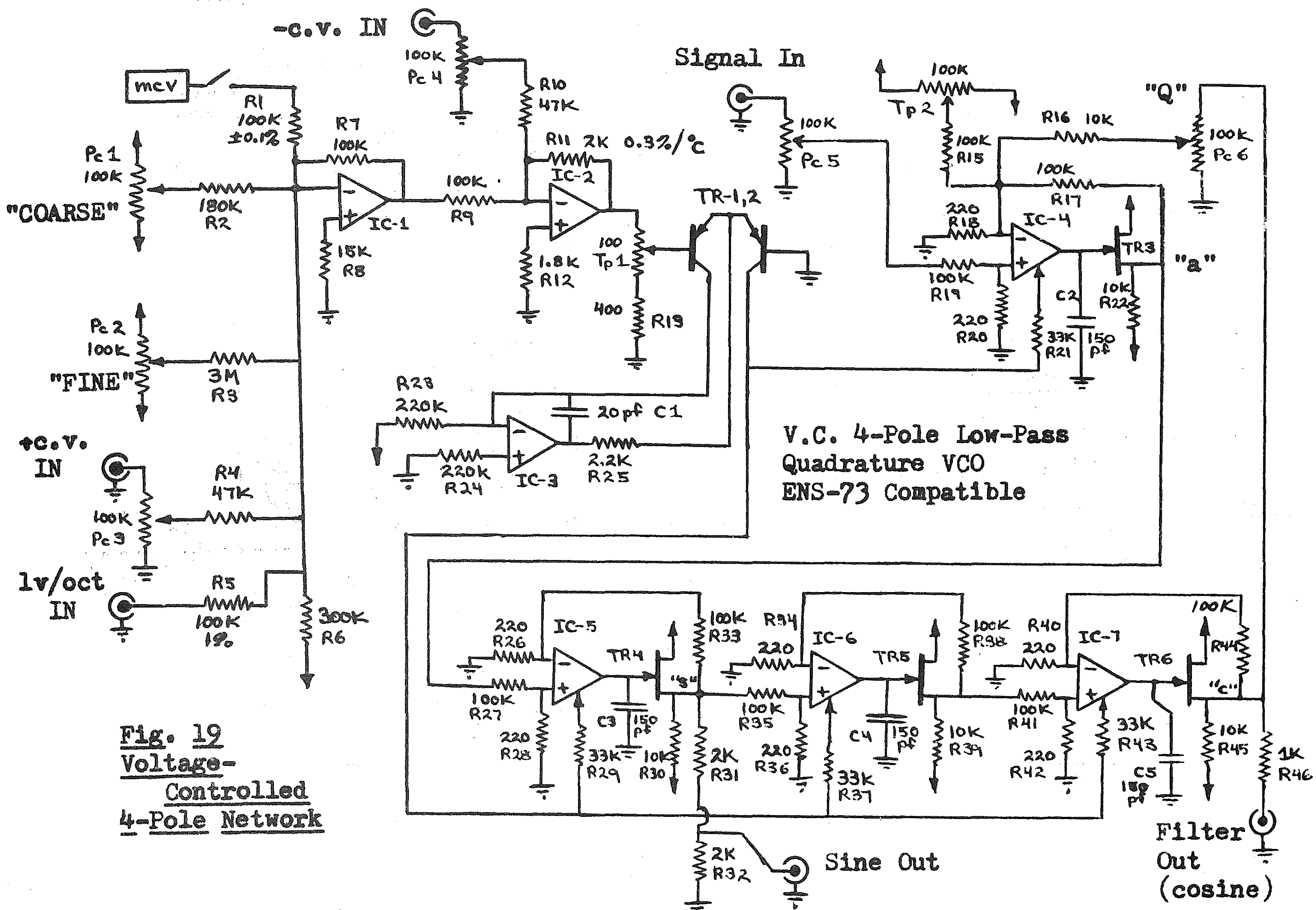


Fig. 19
Voltage-
Controlled
4-Pole Network

IC-1, 2, 3: 307 or 741
IC-4, 5, 6, 7: CA3080
TR-1,2: Matched Pair AD821
TR3, 4, 5, 6: 2N3819 FET's

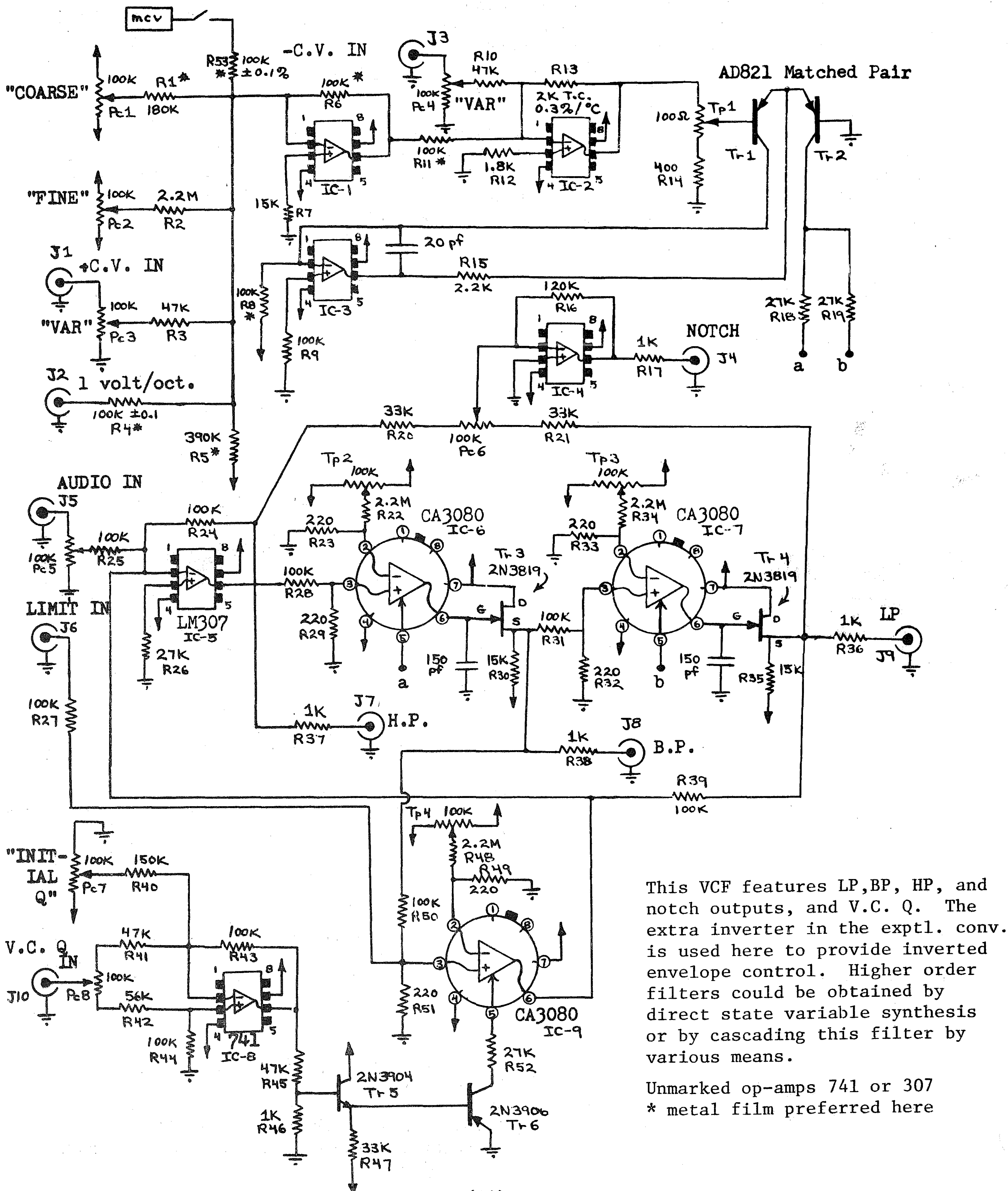
Adjust Tp1 for 1 volt/octave
Adjust Tp2 for symmetric clip at "a" during oscillation at low frequency.

The 2k 0.3% temp. comp. resistor and TR-1,2 should be in close thermal contact.

REFERENCES:

1) J. Graeme, G. Tobey & L. Huelsman, Operational Amplifiers Design and Applications Mc-Graw Hill (Burr Brown) 1971 (page 213)
2) Any good book on active filters, or use the short overview in Millman & Halkias, Integrated Electronics, Section 16-6 (Mc-Graw Hill 1972)
3) R.A. Moog, A Voltage-Controlled Low-Pass High-Pass Filter for Audio Signal Processing, AES Preprint 413 (Oct. 1965)

DESIGN EXAMPLE #2, VOLTAGE-CONTROLLED STATE VARIABLE FILTER (from EN#37, based on original design of T. Mikulic, EN#33 & EN#34)



This VCF features LP, BP, HP, and notch outputs, and V.C. Q. The extra inverter in the exptl. conv. is used here to provide inverted envelope control. Higher order filters could be obtained by direct state variable synthesis or by cascading this filter by various means.

Unmarked op-amps 741 or 307
* metal film preferred here

CHAPTER 5E

ENVELOPE GENERATOR DESIGN

CONTENTS:

INTRODUCTION

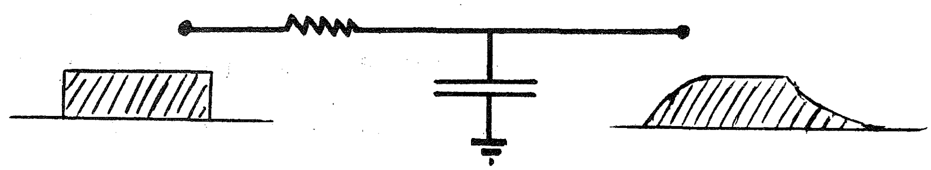
An envelope generator is a circuit that generates a control voltage contour in response to one or more timing signals. The envelope is in general non-periodic, but repeats at irregular intervals. It produces a voltage that is generally zero before timing signals arrive, and left alone, will return to zero sometime after the timing signals disappear. In a sense, it sets the time limits on a musical event, and varies some musically important parameter during the event. An envelope generator output is periodic only if the timing signals arrive periodically. Typically, the timing signals will arrive from a keyboard in response to the players fingers, and correspond to the musical notes the player is trying to produce. Envelope generators may also be triggered by timing signals from an electronic generator (usually a low frequency oscillator) or some type of sequencing circuit.

Timing signals are termed either "gates" or "triggers." A trigger is a very short voltage pulse that serves as a timing marker (e.g., a voltage level changed at time t=5 seconds, and here is a trigger reporting that fact). It may initiate some envelope action. A gate is a longer rectangular waveform of constant voltage and with sharp onsets and terminations. A gate denotes a condition, and the onset and stoping times of the condition (e.g., a key is down on the keyboard at time t=5, so here is a voltage level denoting it. At time t=8, the key is lifted, and the gate disappears.)

ATTACK-RELEASE (AR) ENVELOPE GENERATORS

A gate is also the simplest form of an envelope. The problem with using a gate directly is that the rising and falling edges are too sharp. If the envelope controls some parameter of the sound (e.g., amplitude or pitch) and this parameter changes too fast, the ear hears an annoying click rather than an interesting transition. This click is due to the wide bandwidth continuous spectrum (noise) that results when one examines the Fourier transform of the resultant waveform associated with the sharp edges. Under certain conditions, the ear accepts these clicks, and at other times they are wrong depending on context and musical conditioning of the listener.

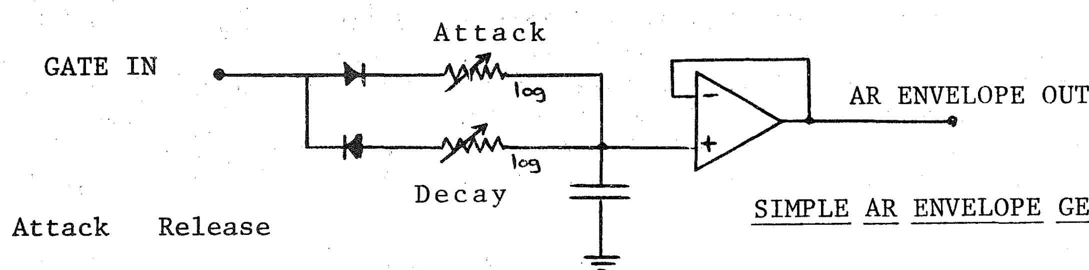
To slow down these changes, we have to take the sharp edges off - remove the high frequency Fourier components. This calls for a low-pass filter. While it is instructive to consider how the filter shapes the gate by removing the high frequency components, it is the time domain analysis (step response) rather than the frequency domain analysis that is usually used here. The problem is the familiar one of applying a voltage to an RC series circuit, letting it charge, and then removing the voltage and letting it discharge.



The result is the familiar exponential rise and fall. It is then a simple matter to add a diode so that the rise and fall time constants are different, and to make the resistors variable so that a wide range of attack and decay times can be obtained. The time constants are found by multiplying together R(in ohms) and C (in farads) to give time in seconds. If R is 1 megohm and C is 2 microfarads, then the time comstant is:

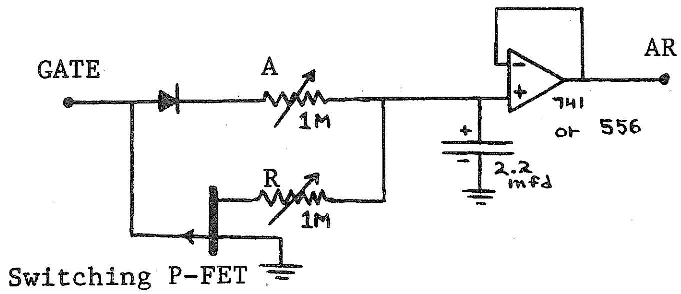
$$T = RC = 1 \times 10^{6} \times 2 \times 10^{-6} = 2 \text{ seconds}$$

This means that it will take 2 seconds for the voltage level to change by $1 - e^{-1}$ or about 63% of the total voltage. After 4 seconds, the voltage will have changed by 63% <u>of the remaining voltage</u>, and so on until the voltage change per unit of time constant is very small. The resulting circuit is an envelope generator commonly referred to as the AR (attack-release) type:



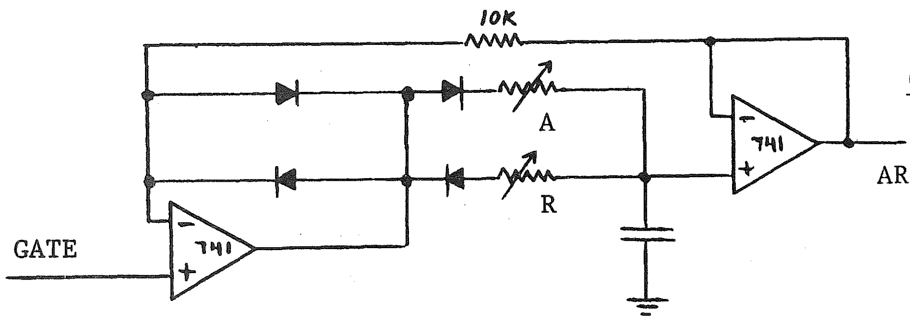
SIMPLE AR ENVELOPE GENERATOR

It should be noted that the AR generated formulated in this way is accurate only to the limitations of the voltage drops (about 0.6 volts) across the diodes. It is more of a problem with the decay than attack, since it is only necessary to add 0.6 volts to the gate level to get the full attack. For the decay section, a FET switch or other analog switch can be used to provide a true ground reference level, or a compensation scheme can be used:



MORE ACCURATE GROUND REFERENCE USING SWITCHING FET

(D. Rossum, EN#22)



COMPENSATION OF AR FOR DIODE FORWARD VOLTAGE

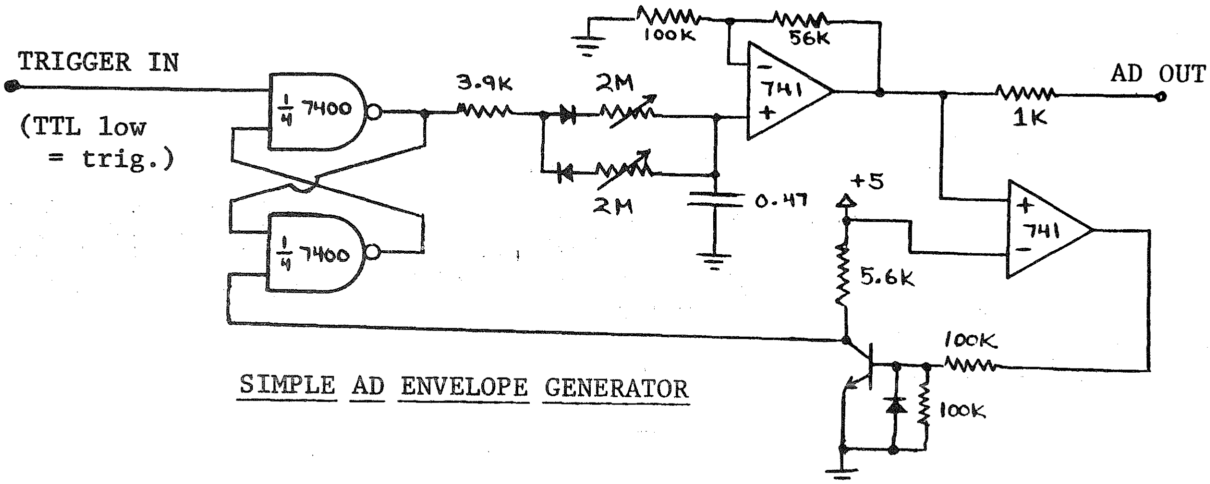
(C. Hovey, EN#31)

Some comments should be made on the relationship between the time constants of the envelope generators and the actual attack and decay times. First note that since the envelopes are exponential, they are never mathematically completed, and as we have seen, they are often truncated. Also, the ordinary "RC" time constant is only a 63% change, and we expect the effect of envelope control will be sharper than this in most cases. However, for an important case, amplitude control, we find that the sensitivity of the ear is not especially sharp. A good figure to work with is a 1 db change in amplitude which is just at the level the ear can detect. This corresponds to a change of about 12%. Thus, we can expect the ear will not detect any change in amplitude once the amplitude rises to within 12% of its final value. This corresponds to about two RC time constants. Thus if we have values of R=2 meg and C=1 mfd, we can expect attack times of up to 4 seconds. Decay is a different story however since the final value of decay is the lowest level the ear can detect. Starting with the maximum amplitude of the envelope, after one RC time constant has expired, the amplitude will be down to about 1/3. This goes on until the tone is no longer audible. In general we can expect this to go on for many RC time constants due to the wide dynamic range of the ear. The limiting factor here is actually the VCA or course, not the ear, so we are working with something like a 60 db range. Decay by 60 db would take about six RC time constants. In actual operation, the decay portion of the envelope is often truncated so that the exponential decay comes to an abrupt end short of the dynamic range lower limit of the VCA.

We should also note that in general the pots used in envelope generators should be logarithmic. This is quite simply because very short envelope times are useful, and a few milliseconds on the short end are much more important than they are on an envelope time ranging into seconds. Likewise, the sustain level control is actually an amplitude control, and should be logarithmic.

ATTACK-DECAY (AD) ENVELOPE GENERATORS

A second type of generator is the AD (Attack-Decay) type. This envelope is also exponential for both the rise and the fall, as is the AR type, but there is no sustain level. The envelope rises to a peak value, and then immediately goes into a decay condition. The envelope is initiated by a trigger signal, or by the rising edge of a gate signal if a trigger is not available. Once initiated, the envelope is automatic and does not depend on any condition remaining or disappearing to determine its shape. The overall duration of the AD envelope depends only on the time constants set. The AD envelope generator can be formed from an RS flip-flop (two two-input logic gates cross coupled):



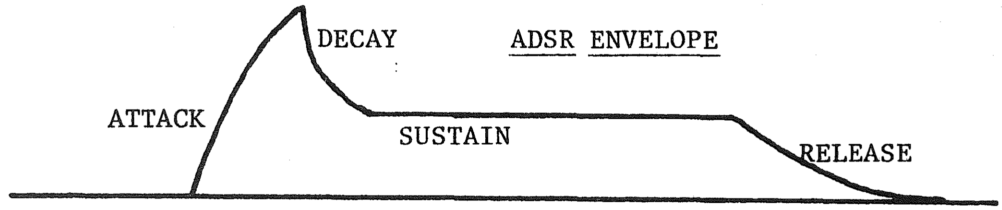
SIMPLE AD ENVELOPE GENERATOR

Note one important point about the AD generator: The decay is a standard exponential, but the attack is not. A rising exponential never reaches the peak, so we cannot set the condition for starting decay to be the envelope reaching the peak, as it will never get there. What we can do is one of three things that are roughly equivalent. [1] Set the reset point somewhere below the voltage toward which the attack is charging. [2] Let the attack charge toward a higher voltage than the envelope you need (e.g., toward 6 volts for a 5 volt envelope). [3] Amplify the voltage slightly before it goes to the threshold detector. In either of these cases, the voltage is trying to charge toward some level that is above the actual envelope output. In an extreme case, where this voltage is much higher than the envelope peak, the attack curve will be only a small part of the exponential rise, hence more linear. Also, the apparent time constant will be shortened. On the other extreme, if this voltage is only slightly higher than the envelope, the AD envelope will be very flat on top and timing may be subject to drift, and different units will behave differently due to component tolerances. As a general rule, the charging voltage can be set about 20% above the envelope voltage for a satisfactory envelope, and consistent results. Fortunately, the ear is not very particular about the exact shape of the attack portion of the curve. When reset occurs, the capacitor has the peak voltage and decays toward zero from that voltage. It has no way of knowing how this voltage was related to the attack curve, and is just the regular exponential decay.

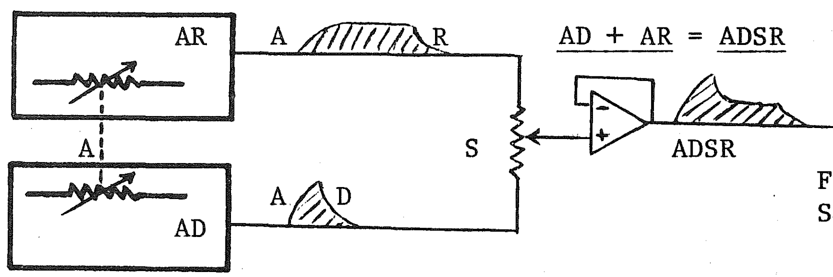
ATTACK-DECAY-SUSTAIN-RELEASE (ADSR) ENVELOPE GENERATORS

Perhaps the most popular type of envelope generator is the ADSR type. (Attack-Decay-Sustain-Release). This envelope has three phases and four control settings. During the first phase, the envelope rises exponentially toward the charging voltage. What we said above about the AD attack also applies here. After reaching a peak voltage, the generator goes into the second phase: decay to a sustain level above zero. In the final stage, the generator decays from the sustain level to zero. Typically, the three times (attack, decay, and release) can be set with pots, and the sustain level can be set between the peak voltage (making it an AR generator) and zero (making it an AD generator). Typically, a sustain level of about 1/3 the peak is used.

At times, the decay portion is called "initial decay" and the release portion is called "final decay." In some designs, the initial and final decay times are the same.



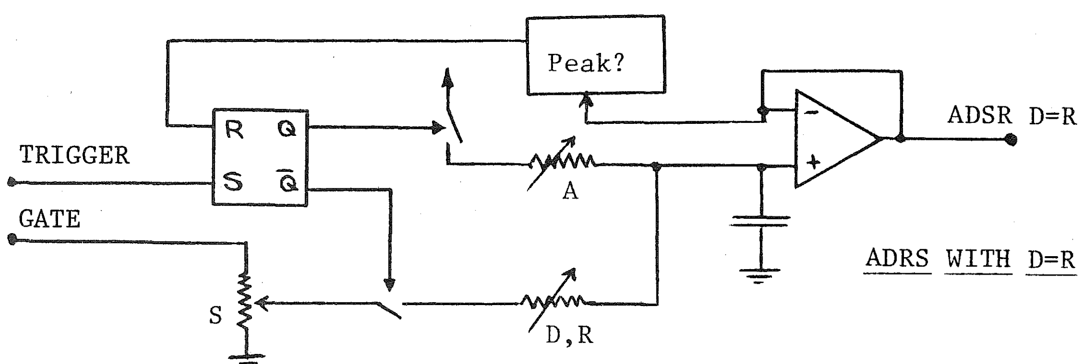
We noted above that the ADSR envelope generator can (by setting the sustain level) be made to function as an AR or an AD envelope generator. This suggests that the ADSR envelope generator can be formed from the sum of an AD and an AR. The ADSR output is obtained by mixing the two, and the control that sets this ratio is the sustain level control. The attack times of both the AD and AR sections must be the same, so a dual pot must be used.



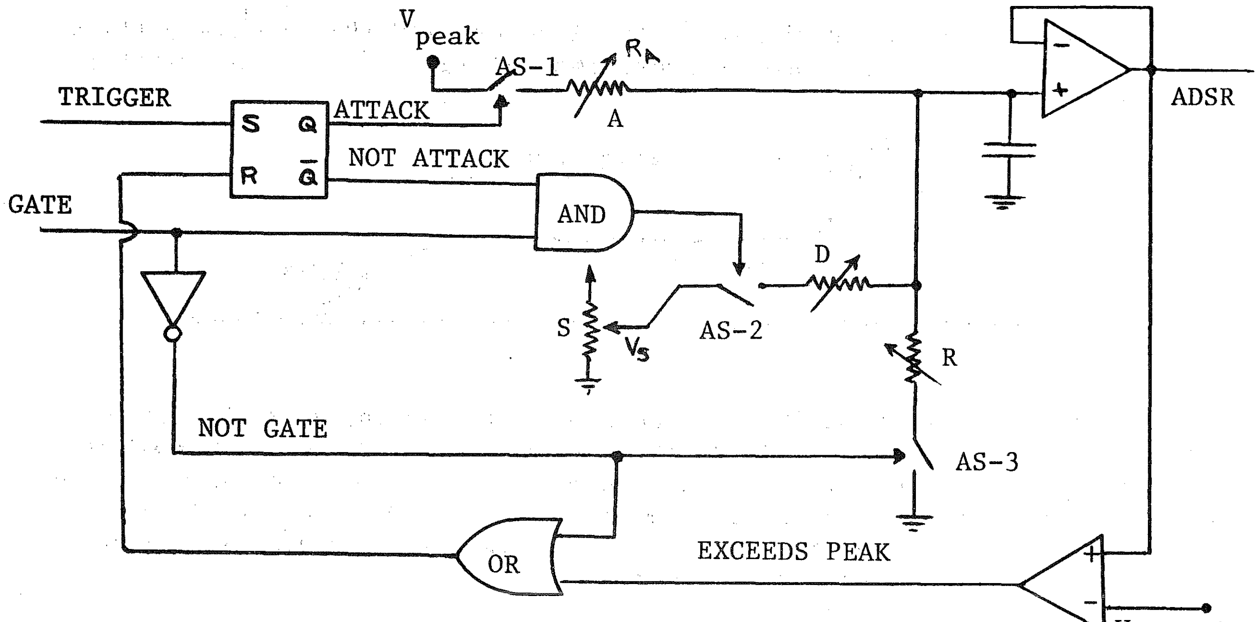
For Full Circuit
See EN#45 (15)

If the trigger signal is not available, the leading edge of the gate is differentiated (and inverted if necessary) to give a trigger. The obvious limitation of doing this is that the generator is not retriggerable. When the trigger is available, a new AD type peak can be made to appear even if the gate does not disappear and come back. In keyboard playing, this would be the case when the player moves from one note to the other without lifting a key inbetween. If there is no trigger, there will be no new attack in this case; only the pitch will change. A circuit that senses the change of control voltage from the keyboard and reports this as a trigger will cause a new attack phase.

The ADSR circuit with initial and final decay times equal is relatively simple, as it only involves an attack phase, and a decay to a level that is proportional to the gate (sustain level). When the gate goes to zero, the sustain level drops in proportion (i.e., it becomes zero as well) and the former sustain level decays to zero.



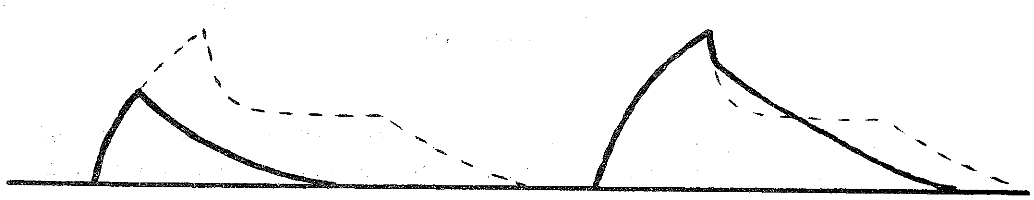
The full ADSR generator can also be formed by using digital logic and analog switching to examine the gate and trigger conditions and respond accordingly.



FULL ADSR LOGIC SCHEME (Reference D. Rossum, EN#22)

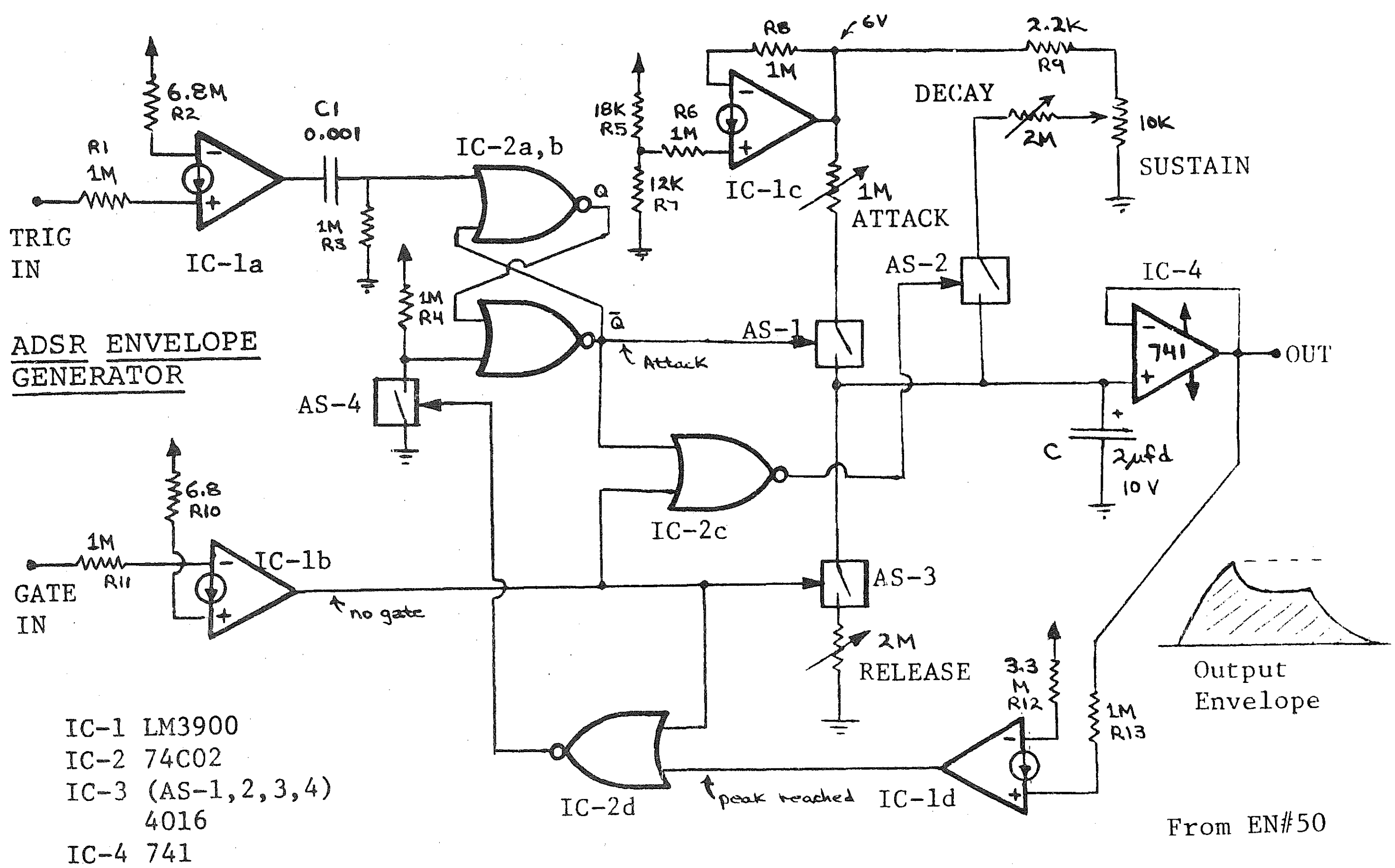
In the normal mode of operation, a gate and trigger arrive at the same time. The R-S flip-flop is switched to Q=1, and the analog switch (AS-1) is turned on. The capacitor begins to charge through R_A forming the attack portion of the envelope, while all other analog switches remain off. When the capacitor (as buffered by the op-amp) voltage reaches a preset value below the charging voltage, the comparator switches, and this output is transferred through the OR gate ending the attack phase. Q goes to zero, and $\overline{Q}$ goes to the one level. Assuming the gate remains high, the signal from $\overline{Q}$ is coupled through the AND gate to turn on AS-2. The capacitor voltage then starts to decay toward the sustain level voltage V_S. The comparator therefore switches back to its original state, but this does not retrigger the R-S flip-flop - it just remains where it was with Q=0. The capacitor voltage approaches V_S and remains there until the gate disappears. When the gate goes to zero, the AND gate is cut off, opening AS-2, and the NO GATE signal then closes AS-3, and the capacitor voltage decays toward zero.

The above is the normal mode of operation. Other modes are those in which the gate disappears earlier. We can examine this in the different envelope phases. If the gate disappears during the attack phase, the NO GATE signal resets the R-S flip-flop to Q=0, turning off AS-1. Since there is no gate, AS-2 does not come on, since the $\overline{Q}$=1 is blocked by the GATE=0 at the AND gate. The third switch, AS-3 is turned on by the NO GATE, so the capacitor voltage decays toward zero. The second shortened mode is the one where the gate disappear during the second phase. The second phase is the D to S phase, and it should be realized that theoretically, this phase (like the R phase) never really completes itself, since decaying exponentials are involved. For this reason, theoretically, this shortened phase differs in no way from the normal mode. It is just a matter of degree, and what the ear can detect - the electronics is the same. What happens electronically is that the envelope starts to decay toward zero before it gets (for practical purposes) to the sustain level. These modes are diagrammed below:



DESIGN EXAMPLE - ADSR ENVELOPE GENERATOR

An example of a practical ADSR design is shown below. The design example also illustrates certain principles of logic design that can be used to reduce the number of IC's in the circuit. The R-S flip-flop can be formed from two two-input logic gates cross-coupled. We could choose either NAND or NOR gates for this purpose. Since both an AND type decision and an OR type decision must be made, we would like to apply the De Morgan theorem to use the same gate type. In particular, the idea of using a single quad gate package is attractive. It turns out that the NOR package is better on several counts. First of all, since the triggers we are assuming are short positive pulses, it makes sense to use the R-S flip-flop formed from two NOR gates. This means that the attack phase can start as soon as the trigger rises. If we were to use two NAND gates, we would need to either use an extra inverter, or differentiate the trigger to obtain a negative going transition from the power supply level. We then use one of the remaining NOR gates for the OR, planning to follow it with an inverter. The remaining NOR is used as an AND gate by feeding the inputs the inverted versions of the two conditions we want to AND (De Morgan's theorem). Thus since we want to AND the conditions "NOT ATTACK" and "GATE", we NOR the conditions "ATTACK" and "NO GATE." Fortunately, the ATTACK was the condition we had in the first place (in any case, the RS flip-flop makes both available), and the NO GATE is the condition we need to control the third analog switch. The NO GATE signal is obtained from the GATE input buffer. We would at this point seem to be one inverter short, but this can be obtained from the analog switch, since these come four to an IC package anyway and the fourth one would go to waste if not used for something.



If we use an input buffer on the trigger input, then there is nothing that prevents us from considering the R-S flip-flop formed from two NAND gates, as the buffer can invert the positive going trigger to the negative going spike (logic

"1" going briefly to zero and back). We would then be using a NAND gate plus an inverter for the AND. To use the remaining NAND as an OR, we would have to have available "GATE" and "NOT ABOVE PEAK" (the comparator output). Since we also have to use the NO GATE condition, we are an inverter short. Finally, to make matters worse, we see that the OR output is not what is needed to reset the R-S flip-flop - we need a NOR gate, hence one more inverter. The NOR gate package has a clear advantage here. In a general case, the situation must be carefully considered. If the analog switches happened to be turned on by a logic zero, it would be a different story.

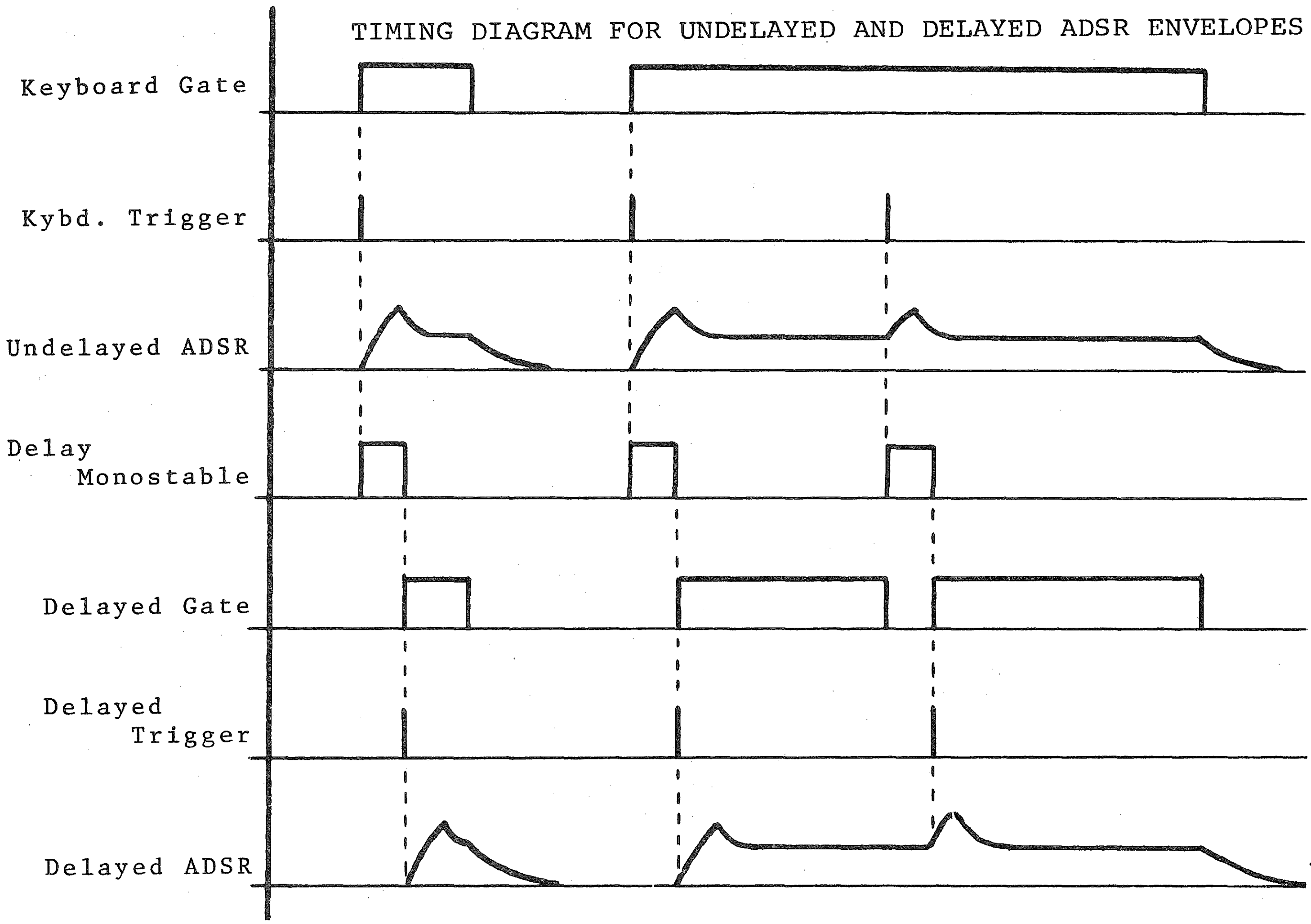
ENVELOPE DELAY UNITS

It is often desirable to add an initial delay section to any of the envelope generators we have discussed above. First of all, we must examine where a delay makes sense, and how this relates to the timing signals, and the envelopes we eventually want.

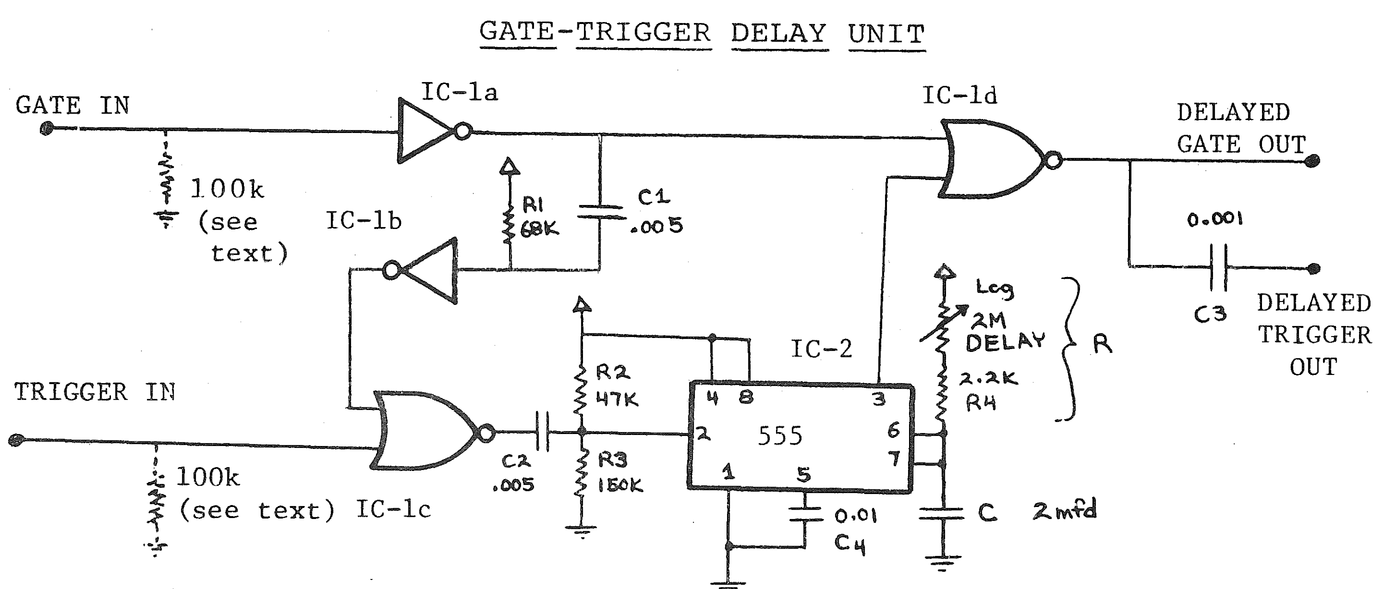
First, it makes sense to delay an AD type of envelope. For example, the AD envelope responds to a trigger, and the trigger comes when a player pushes a key on a keyboard. If instead, we let the trigger from the keyboard turn on a monostable, and then when the output of the monostable falls, we trigger the AD generator, we have delayed it. The user can push a key on the keyboard, go out to lunch, and come back later to hear his note if he wishes. More practically, he may set a sequence of envelopes to follow one another after just pushing a key once.

Contrast this with the idea of delaying a gate. It is possible and makes sense to delay the onset of the gate. We can consider delaying an AR envelope by simply using a monostable to block the gate for a period of time. The user would push a key, and after a delay, the gate appears. But what happens if he lifts the key before the delay time expires? What should happen? The only thing we can consider at this point is to not have anything happen. If a gate were to appear when the delay time expired, when would it end? A gate is extended in time, and must have starting and stopping points. The starting time can be delayed, but we must then either use a "monostable gate" (one that has a preset duration that starts when the delay ends - just as an AD envelope does), or the ending time of the real gate must be delayed by the same amount, and then the ending time can depend on the player.

What is often used as a base for a delayed envelope generator is the ordinary ADSR generator with gate and trigger controls, and a "gate and trigger delay unit," which may well be an integral part of the ADSR envelope generator circuit. Generally, we are talking about a timing control system where a trigger arrives whenever the gate does, and where a trigger may arrive to restart the "AD" part of the generator without the gate disappearing at all. It is then best to consider the trigger as the start of the delay, but for practical purposes, the delay should be triggered by whichever occurs first - the gate rise or the trigger. If the gate happened to arrive slightly before the trigger, the AR part of the ADSR envelope could rise slightly, because the monostable that blocks the gate has not yet risen. This would give a "blip", delay, ADSR response, where the "blip" is unwanted. Thus, we have the monostable delay triggered by either the gate rise or the trigger. When the delay falls, it is then necessary to recover the gate and the trigger. One way of doing this is to AND the "NO DELAY" and the "GATE" condition. Thus, when the delay time expires, and the original gate is still on, a delayed gate rises, and the leading edge of this becomes the trigger. Suppose that the gate still remains. The ADSR envelope then goes through A,D, and S. If the original trigger input receives another pulse at this time, the monostable delay is activated, and the gate disappears at the input of the ADSR. The ADSR then goes from S to R, and when the delay expires, it goes through A,D, and S again. Finally, when the original gate disappears, the ADSR goes into a final R.



DESIGN EXAMPLE - DELAY UNIT FOR ENVELOPE GENERATOR



IC-1a,b,c,d = 1/4 7402 or 74C02 (two sections used as inverters)   ↑ = +5v

A full description of the envelope delay unit design example can be found in EN#51. The basic implementation follows the discussion above. The circuit can be coupled directly to the ADSR design example. Both the gate and trigger levels should be 5 volts. It is possible to buffer these inputs with type 3900 quad amplifiers as was done in the ADSR design example. If this is done, and the delay unit is made a permanent part of the ADSR envelope generator, then the buffers on the ADSR unit are not needed, and these can be used to drive the delay unit.

SPECIAL TOPICS IN ENVELOPE GENERATION

LINEAR RAMPS: Although we have been using exponential waveforms exclusively (with certain truncations), and there is some reason to prefer exponential decays based on musical acoustics, linear ramps can be employed in the designs by simply replacing the charging resistors with current sources.

TIME SUSTAIN: In some designs, a sustain section of the envelope that is a preset time can be used. Basically, this is a matter of making the gate available from a monostable triggered by the trigger. The same design that was used for the delay unit can be employed for this by a simple switching arrangement if the delay is not used in this mode.

PROGRAMMABLE WAVEFORM GENERATORS AS ENVELOPE GENERATORS: If a programmable waveform generator is slowed down and made to cycle through its waveform only once and then reset, it may be a useful programmable envelope generator. In fact, this is a preferred application of this sort of device. Basically, the ADSR envelope has been used to the exclusion of all others, probably because it produces amplitude control that is very similar to traditional musical sounds. A programmable waveform generator thus makes possible the examination of other types of amplitude control. More importantly, it is possible to examine the effect of different envelope contours on other parameters of the sound where ADSR envelopes have been used because nothing else was available, and where there is less reason to prefer this type of envelope shape based on musical acoustics.

THE MINICOMPUTER AS A MASSIVE ENVELOPE GENERATOR: Minicomputers are becoming very common and are being used for music synthesis. These small computers are not fast enough for direct digital synthesis (point by point), but can easily generate many control waveforms since these vary slowly relative to the computer speed. Thus, we can think of the minicomputer as a multiple envelope generator, and generalize our concept of envelope somewhat to include any control voltage (i.e., "regular" envelopes, gates for switches, and stepped control voltages for discrete changes of pitch for example). It is easy for the computer to run through a series of routines (say 10 to 100 of them) and come up with a number which is output through the D/A for each routine. This voltage can be stored in a sample and hold corresponding to the envelope. The computer then goes ahead with the next routine and updates the next envelope, and so on. Eventually it gets back to the beginning in plenty of time (relative to the time constants of the change at the ear) to update it without causing a serious discontinuity. Thus the minicomputer, properly programmed can be a massive envelope generator. It can be programmed from a normal sort of musical keyboard for note-to-note changes. In addition to the computer, one D/A is needed, one sample-and-hold for each envelope, and the proper sampling sequencer. This can be connected to control traditional synthesizer voltage-controlled modules. The hardware is probably much simpler than the software in this case.

CHAPTER 5F

BALANCED "RING" MODULATOR DESIGN

CONTENTS:

INTRODUCTION

Sinced a balanced modulator is basically just an analog multiplier, much of the design is done when an accurate 4-quadrant multiplier (4QM) is designed and balanced (see chapter 3d). What remains to do is add circuitry external to the actual multiplier to accomplish the following:

(1) Input and output signals with standard levels and impedences.

(2) Allow for AC coupling as needed.

(3) Permit the use of the balanced modulator module as an amplitude modulator, frequency doubler, and a VCA as needed.

Of these, the first is basically just what has been done with other modules. Input attenuation is achieved with pots, and the actual input is the standard op-amp summing node. The output is taken through a 1k resistor from the output of the op-amp used in the multiplier level shifter.

Assuming we have gone to a lot of trouble to balance the 4QM, we don't want to upset this balance by introducing a signal with a DC component. Often the outputs of VCO's and signal processing devices have a small DC offset that is usually no problem, or they may develope a drift. The direct connection of such a signal to the inputs will effectively unbalance the multiplier and cause the frequencies on the other input to appear in the output. In evaluating "carrier rejection" in balanced modulators, it is necessary to make sure there is in fact no DC component in the signals or the modulator may be blamed for doing exactly what it should. One solution is to make provisions for AC coupling of signals.

It is generally desirable to have any module have the capability of performing more than one function if this is possible. By adding an intentional DC offset, the device is capable of amplitude modulation. It is however necessary to do this in such a way that the DC balance of the 4QM can be easily restored without exacting testing. Thus, a one polarity DC voltage is used as the offset. With the inputs tied together, the device is capable of frequency doubling, and if this is done a lot, it may be desirable to include a switch for shorting the inputs. Finally, a balanced modulator is often used as a 4QM to perform VCA functions where a wide dynamic range is not too important. For this reason, DC coupling of the inputs should be available.

ADAPTING THE ANALOG MULTIPLIER

When implementing input structures, it becomes obvious that since we have several input structures in parallel, we will drive the inputs from op-amp summers. This causes some changes in the basic multiplier. (1) Since the op-amp summers are inverting, we will input through pins 8 and 12 instead of pins 4 and 9. (2) We can do the AC balance trimmer settings on the op-amps. This means that pins 4 and 9 can be grounded (through 470 ohm resistors for RF stability). (3) Likewise, since the impedence of the op-amp outputs is low, 470 ohm resistors are inserted between the outputs and pins 8 and 12.

We can implement AC coupling using capacitors as a switched option on the inputs. The factor to consider in selecting this capacitor is the low frequency response. We thus ask: what is the lowest frequency we would <u>like</u> to input to the balanced modulator? The first response is: DC, but this is ruled out by the need for AC coupling. We therefor look into the nature of balanced modulation at low frequency. We saw in the chapter on generalized modulation that it was the same as "beating" and very similar to 100% AM. Thus, note the following points:

(a) If a low frequency (below 15 Hz) is fed through by a DC component in the other input, it is inaudible anyway.

(b) If a high frequency is fed through by a DC offset in the other input (which, since we are considering low frequencies here, is of low frequency) then the carrier fed through will be between two already close sidebands. If the low frequency is 7 Hz or lower, beating will be heard in the output and this is an amplitude variation of the average frequency, which is the exact position of the signal leaked through. Thus, we might expect very little effect due to the leakage.

(c) Most effects achieved with the input of a low frequency can be achieved with a similar method using DC coupling.

Thus, we might conclude that AC coupling down to about 7 Hz might be reasonable. The input resistance is something like 100k, so if we allow the capacitive reactance to go as high as 5k (5%), then the capacitance should be:

$$C = 1/[2\pi f \cdot X_c] \simeq 1/[6 \cdot 7 \cdot 5000] \simeq 5 \text{ mfd}$$

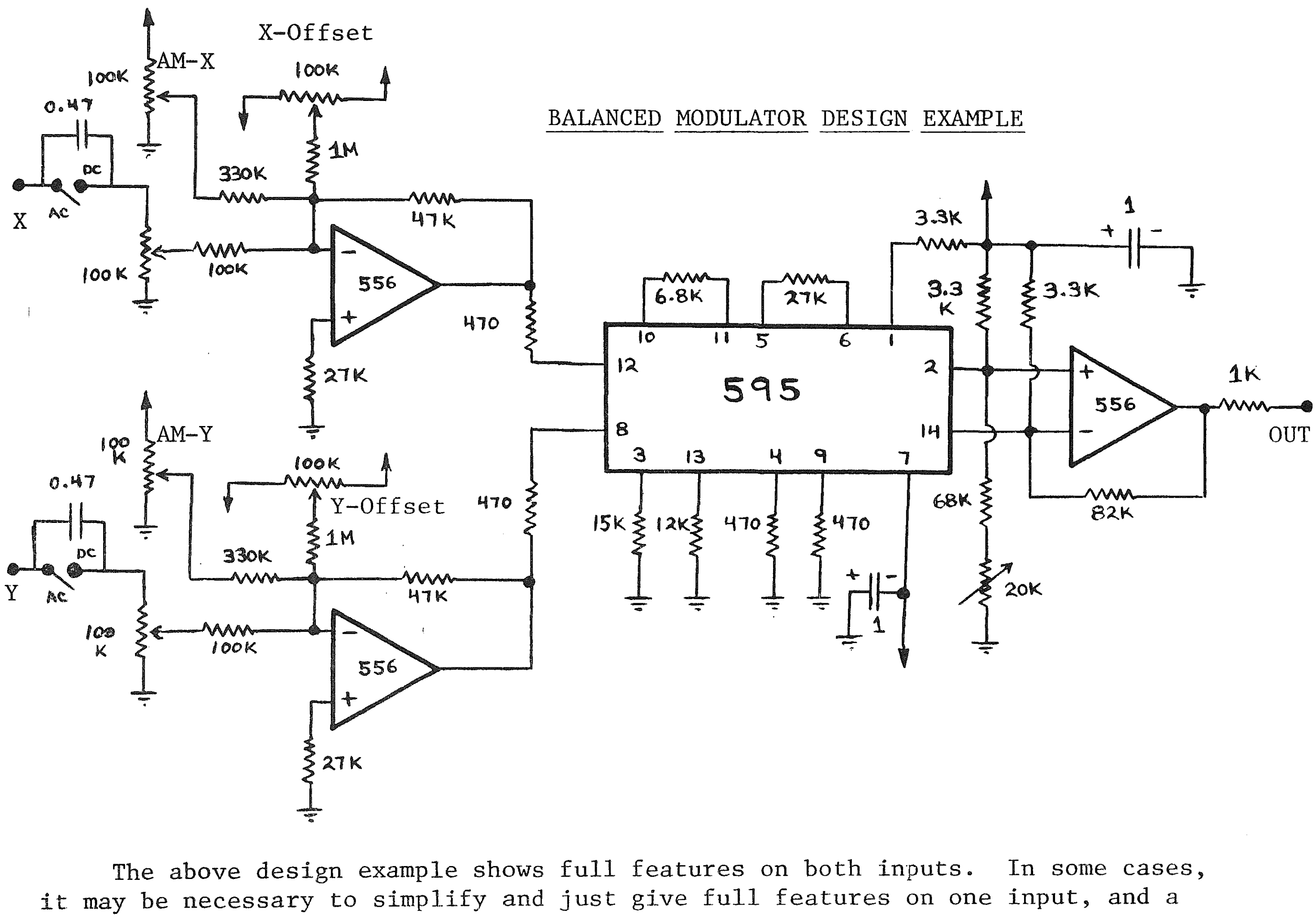
This 5 mfd capacitor can be achieved as two series 10 mfd electrolytics back to back. In many cases, a smaller capacitor will be enough and can be found as a standard dielectric type (e.g., a 0.47 mfd capacitor).

In order to obtain AM effects, a combination of AC and DC must be fed to one input while AC is fed to the other. If a DC term is fed to both inputs, both the original frequencies along with the sum and difference frequencies will appear in the output. In setting up these DC inputs, it is best to just use one polarity; that is a pot connected between + and ground for example. This makes it easy to restore balance by just returning the control to the low position. If the pot were connected between + and -, it would have to be carefully balanced after each use. In the design we will show, we use a +15 to ground for one DC offset (+5 due to input resistor), and -15 to ground for the other. This gives some choice. An important often overlooked function of such a device is its ability to act as an "in line" attenuator. With the suggested design, this can be a controlled inverting attenuator as well. For example, if you have the balanced modulator free, and need an inverter, just set the Y offset to the full - position and insert the signal to be inverted into the X input.

DESIGN EXAMPLE

The specific features of this design example have been discussed above. The balancing technique is to first return the AM controls to zero. Next, insert a ±5 volt 1000 Hz signal (or so) into the X input (AC coupling with the gain all the way up). Adjust the Y-Offset for minimum signal at the output. Repeat for the Y input. Repeat both steps until no improvement is noted. Balance the DC with the 20k pot, adjusting the 68k resistor up or down if necessary.

BALANCED MODULATOR DESIGN EXAMPLE



The above design example shows full features on both inputs. In some cases, it may be necessary to simplify and just give full features on one input, and a simple input on the other.

CHAPTER 5G

SAMPLE-AND-HOLD DESIGN

CONTENTS

INTRODUCTION

A Sample-and-Hold (S&H) designed as a synthesizer module is useful for many purposes. It can be used to chop up any waveform (signal, noise, envelope) into steps. The S&H module is different in several ways from S&H circuits used in other parts of synthesizers. For example, there is often a S&H circuit in a keyboard interface. This S&H must be accurate and have minimal droop rate, but it need not be particularly fast. Other S&H circuits in synthesizer devices may have to be very fast, but they may not need to hold a voltage very long. The S&H module is an intermediate case, and may have some additional features for programming the sampling according to the user's needs. Typical specifications for a S&H module might be:

1) Sampling rates from several hundred hertz down to 1 Hz or lower.

2) Acceptable droop rate (less than 1%) at the longest sampling interval.

3) The ability to slew from one signal amplitude maximum to the maximum of opposite polarity in one sample command.

4) Provisions for external sampling commands.

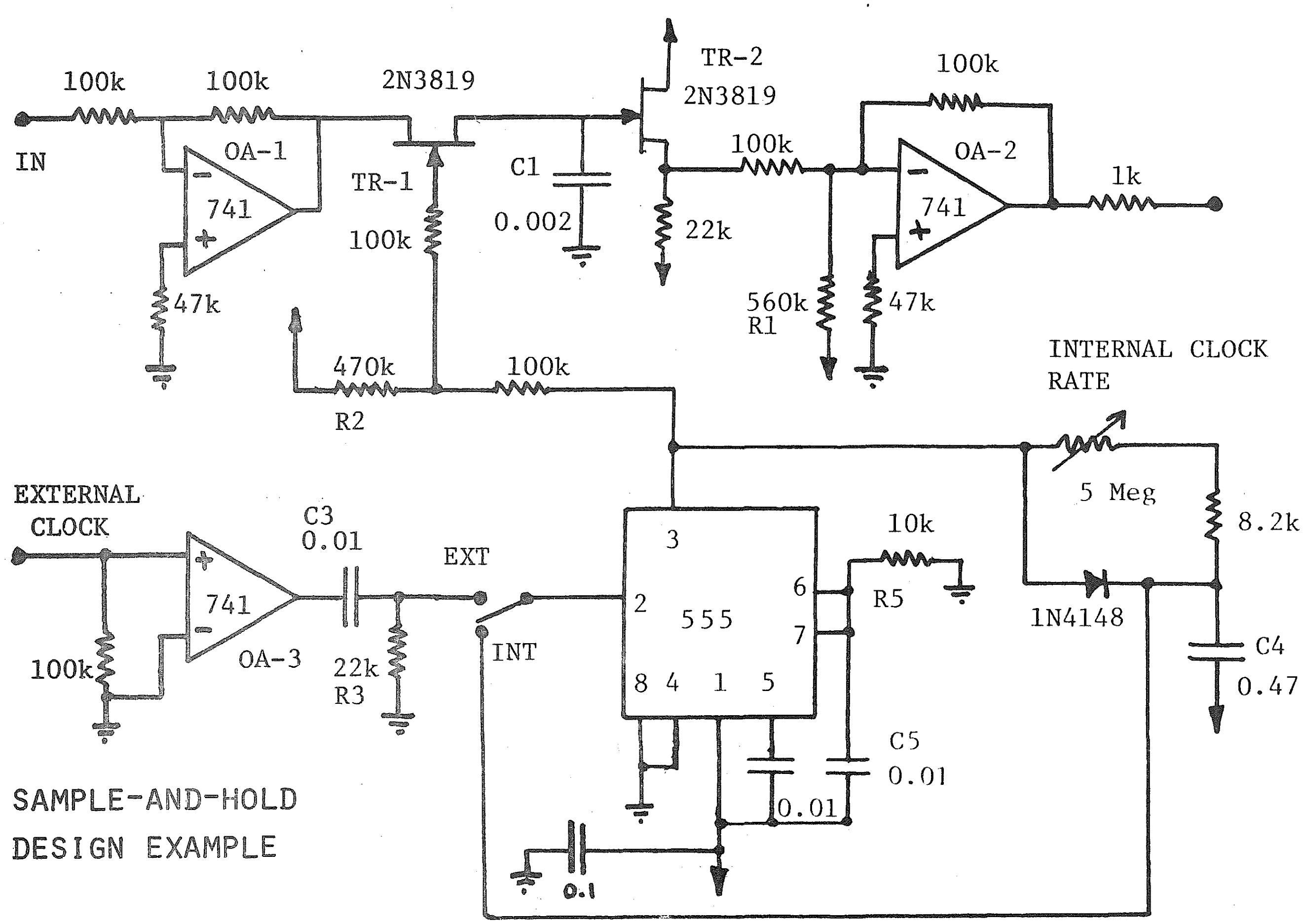
5) Extra features such as slew rate limiting.

The S&H unit is basically an analog switch of some type followed by a holding capacitor and a buffer to slow the discharge of the capacitor. In addition, there is some sort of sampling command that must tell the switch when to sample. The analog switches that are available are the FET, the CMOS analog switch (type 4016) and the CA3080 OTA. The CA3080 circuit [see RCA Application Note ICAN-6668] is a good choice in many cases, but it can not fill all needs due to limited current drive. The type 4016 CMOS quad analog switch makes a simple and cheap S&H for some applications. In the design example, we will use the FET switch.

SAMPLE-AND-HOLD DESIGN EXAMPLE

In the design example, TR-1 serves as the sampling switch. Since it is an N-FET, it is held off by a negative voltage applied to the gate (negative voltages repell the negative electrons which are the charge carriers). The FET switch is turned on when the gate voltage is allowed to float up to the input voltage from OA-1. C1 serves as the holding capacitor, and the voltage on C1 is buffered by source follower TR-2. This source follower is necessary to prevent excessive drain of current from the capacitor. However, the FET buffer causes a substantial offset voltage which is excessive in this case. [In many cases, the FET is inside a feedback loop and the offset is not important. In other cases, the offset is nulled out later on.] It is convenient to correct for this offset using the inverting summer (OA-2) and the op-amp is also useful for lowering the output impedence of the device as a whole. To compensate for the inversion, the input buffer (OA-1) is made an inverter as well. The input impedence is thus the standard 100k. It can be seen therefore that the actual voltage is stored as the negative of the input. Resistor R1 serves to compensate for the offset of the source follower. If a different FET is used, this value may have to be changed some.

The heart of the timing section is the type 555 timer. Note that this is essentially the standard monostable configuration for a 100 μsec. pulse out (C5•R5). The 555 is powered between ground and -15 instead of the usual +15 to ground. This makes no difference to the 555, but the pin 3 out is thus normally held down around -15, and rises to give a 100 μsec. pulse rising toward zero. The resistor R2 serves to pull this output up (as seen by the gate of TR-1) and this allows the FET switch to be clocked with a gate voltage in the proper range.



SAMPLE-AND-HOLD
DESIGN EXAMPLE

OA-2 serves as a comparator to allow for external clocking by any signal that crosses zero. The negative transition from the output of the comparator is coupled through differentiator C3-R3 to trigger the 555 for one sampling command out. In the internal mode, the pulse out of pin 3 charges C4 through the diode. Then this capacitor voltage discharges back toward pin 3 when the pulse falls (to -15) through the series 8.2k resistor and the 5 Meg RATE pot. When this level falls below the trigger level for pin 2 (about -10 volts in this case), another pulse is triggered.

The specs for the design example are:

Input signal range:	+5 to -5 volts (higher voltages inaccurate)
Internal Clocking Rate:	1/2 Hz to 240 Hz
External Clocking Rate:	up to 8 kHz
Droop Rate:	1%/sec.

If slower sampling times can be used (less than 100 μsec.) then R5 can be increased to increase the sampling time, and C1 can be increased in proportion to give a slower droop rate.

ADDITIONAL FEATURES

A number of special features can be added to S&H modules. One interesting device is to add slew rate limiting. This is simply a matter of inserting a resistance between TR-1 and C1 to slow the charging rate. This has the effect of limiting the maximum change that is possible for each sampling command. Only frequency ratios (input freq. to sampling freq.) that are close to integers will give amplitudes that reach the true signal maxima. With a noise input, this means that successive samples will be correlated to some degree and relatively large amplitudes will become rare. This has the effect of changing the probability distribution of the sampled noise. As slew limiting increases, the time averaged amplitude will be greatly decreased. It is possible to use a dual pot; one section determines the slew while the other increases the gain of the output amplifier as slew limiting increases. This tends to keep the maximum amplitudes in the same range while the probability distribution changes.

Another device is the cascading S&H module. In this device, when a new sample command arrives the stored sample is first transferred to a second S&H before a new voltage is loaded in the first S&H. This can be extended to additional S&H units. Delayed sequence effects can result, and a number of analog delay line devices can be used in an elementary form, often with excellent musical results. Implementing such a multiple S&H unit is just a matter of sequencing individual S&H units with a device that cycles for each S&H in response to one input trigger. The delayed outputs appear on separate panel jacks. No additional controls are necessary.

CHAPTER 5H

NOISE AND RANDOM SOURCE DESIGN

CONTENTS

INTRODUCTION

Noise sources and random signal sources can be employed for many puropses in electronic music for a wide variety of effects. Such sources may have a variety of spectral and statistical properties, and may be pseudo-random (not actually random because it repeats, by short segments seem random). The concept of randomness is difficult to begin with and the processing of this type of signals by the human mind further complicates analysis. Thus, we will take the approach here of simply describing the devices that produce the signals without an extensive discussion of the properties of the signals. The devices used are mainly the so called "white noise" generators, and pseudo-random sequence generators. To a large degree, the exact nature of the output depends on the processing of the device output by electronic circuits external to the actual noise source.

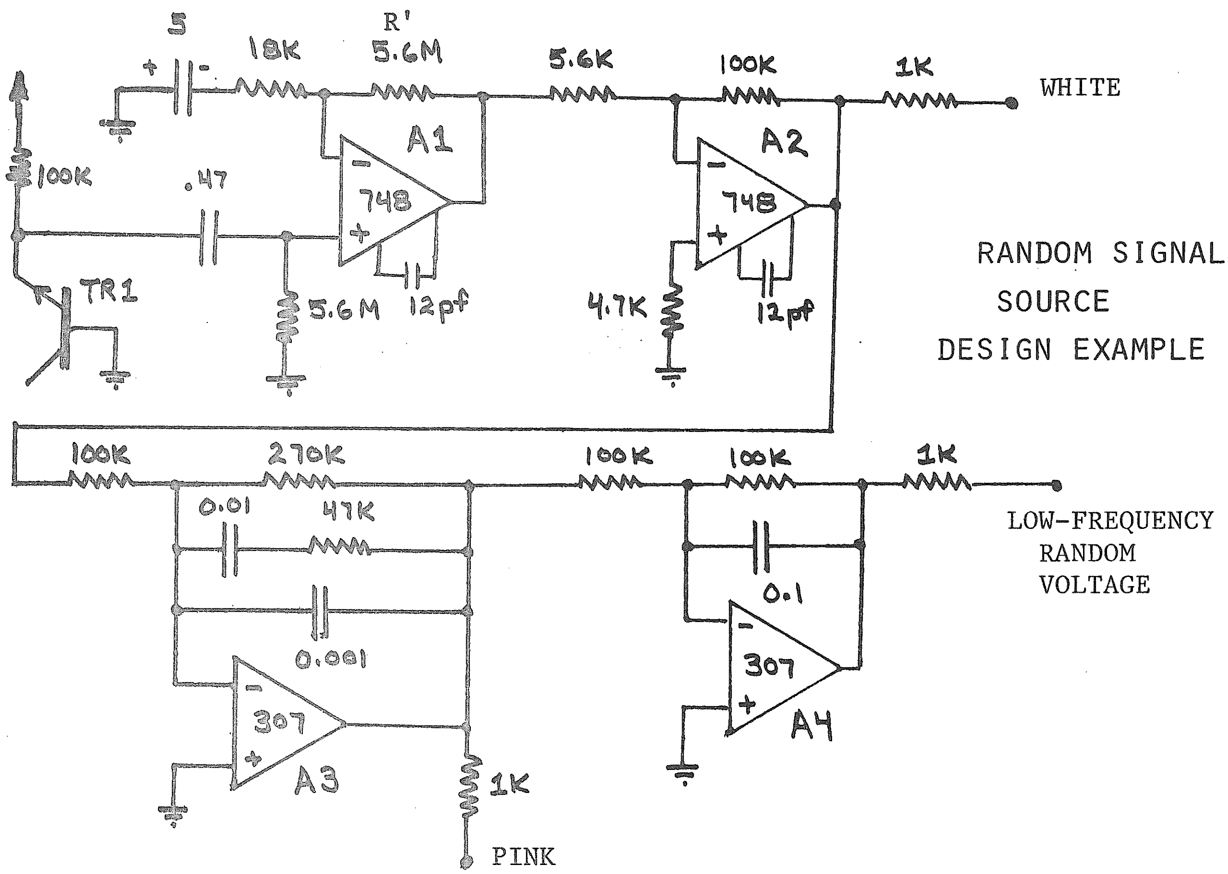
NOISE SOURCES USING SEMICONDUCTOR JUNCTIONS

A back biased semiconductor junction produces a good approximation to true white noise over most of the audio bandwidth. [We will consider white noise to be a signal with no periodic components and a time averaged spectrum that is flat. It is sometimes specified that white noise have a statistical amplitude distribution that is Gaussian - a bell shaped curve with zero as the most probable amplitude.] The amplitude of the junction noise voltage is fairly small and needs considerable amplification to reach synthesizer signal lavels. An amplifier with good high frequency response is required for a good "crisp" white noise sound.

Since white noise has a flat spectrum, it can be filtered to select or reject certain portions. In many cases, this is done by filters external to the noise source by techniques that are standard synthesizer patches. It is possible however to add certain filters inside the noise module to give different outputs. White noise has a flat spectrum, hence equal power per unit bandwidth. [Note that this is the type of spectrum one gets from thermal noise and "shot" noise where the output voltage squared is proportional to the bandwidth.] Another type of noise is "pink noise" which gives equal power per octave. The sound is not so "crisp" since there is less high frequency energy. Finally, it is useful to heavily filter the white noise to extract a low-frequency portion of the noise for use as a random control signal. The example design has this type of three output processing.

DESIGN EXAMPLE, WHITE, PINK, AND LOW-FREQUENCY RANDOM VOLTAGE SOURCE

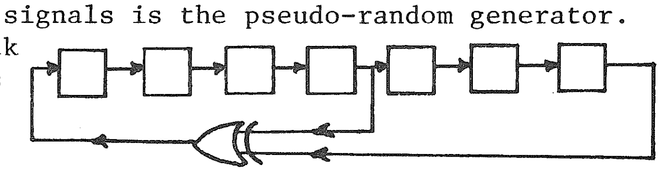
This design is based on the noise and random source designed by T. Mikulic in EN#34. The source of the noise is the back biased junction of TR-1. The white noise from TR-1 is amplified by A1 and A2 to give a 5 volt amplitude at the output of A2. If this gives too much amplification, R' can be cut back to 4.7 M or 3.3 M. A3 forms a pink noise filter rolling off at approximately 3 db/octave. A4 forms a single-pole low-pass filter that gives the low-frequency random voltage out. The output of A4 averages about 8 zero crossings per second. The overall performance level depends a lot on just how noisy the actual transistor happens to be. The low-frequency output is also effected by the burst noise of A1. It may take the circuit 10-20 seconds to establish its DC level after power is applied and the outputs may be pinned at the supply level until this DC level is established.



RANDOM SIGNAL SOURCE DESIGN EXAMPLE

PSEUDO-RANDOM SEQUENCE GENERATORS

A second approach to the generation of random signals is the pseudo-random generator. Pseudo-random sequences can be generated by feedback shift registers. An example of such a generator is shown at the right. The feedback is obtained by exclusive-ORing the outputs of certain stages and



feeding this to the input. The flip-flops are devices that pass their input state to the output when clock pulses arrive. If the feedback is properly chosen, a maximal length pseudo-random sequence results.

Since each filp-flop has two possible states, an n-stage shift register has 2^n possible states total. Of these, the all-zero state is not allowed since this feeds back a zero to the input with all zeros inside, thus stalling the shift register in an all-zero state. The 7 stage shift register above thus has $2^7-1 = 127$ possible states. Short sequences (say 3 to 6 stages) are useful for generating random tone sequences [see D. Lancaster, "Psych-Tone," PE Feb. 1971] and for musical timbre generation [see R. Burhans, JAES 20 #3, April 1972]. Longer sequences (say 12 - 34 stages) are useful for generating pseudo-noise of various types.

The problem of selecting feedback conditions so that the shift register cycles through all 2^n-1 states is not trivial, but the necessary data is available*. Testing for sequence length is not difficult. One sets up an n-input AND gate to detect the "all ones" state on the shift register. When the "all ones" state arrives, a gate rises and passes clock pulses to a counter. When the "all ones" state again appears, the gate falls, and the counter reading is the sequence length.

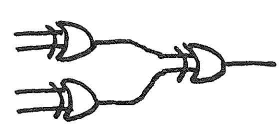
DESIGN OUTLINE FOR A PSEUDO-RANDOM SEQUENCE GENERATOR

Suppose we decide to make a pseudo-random white noise generator for audio use. We know from sampling theory that for a 15 kHz bandwidth, we need at least 30,000 samples per second. [Depending on how the output is taken, it may be necessary to run at higher rates.] If we further stipulate that the noise should run at least 5 minutes before repeating, then the sequence should be at least 300 sec × 30,000 samples/sec or 9,000,000 long. This can be met by a sequence $2^{24}-1 = 16,777,215$. To get this, we start with a 24 stage shift register. We then consult a reference (Peterson & Weldon) and find for order 24 (page 490) the first entry is 100000207F. This is an octal representation of a "primitive irreducible polynomial" from which we can get the feedback condition. First we change from octal to binary:

1	0	0	0	0	0	2	0	7
0 0 1	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 1 0	0 0 0	1 1 1
	1 2 3	4 5 6	7 8 9	10 11 12	13 14 15	16 17 18	19 20 21	22 23 24

Shift Register Feedback Conditions

Note there are n=24 binary digits to the right of the first "1" in the sequence. The remaining "1's" are in positions 17, 22, 23, and 24. This tells us that the feedback should be an exclusive-OR of stages 17, 22, 23, and 24. The necessary 4-input exclusive-OR gate is made by cascading two-input exclusive-OR gates as indicated at the right.



*See for example D. Lancaster, "Understanding Pseudo-Random Circuits," Radio-Electronics April 1975. The theory can be found in W. Peterson & E. Weldon, Error Correcting Codes, MIT Press (1972); feedback conditions can be derived from the tables in Appendix C. J. Peatman, The Design of Digital Systems, McGraw-Hill (1972) gives design information for registers up to 20 stages using JK flip-flops; see Appendix A-3, pg 415. A table for sequences from 2 to 34 stages derived from Peterson & Weldon by the process shown in the example in this chapter is presented at the end of this chapter.

There are various ways of getting a usable noise signal out once the sequence is generated. First the sequence could be used directly. An integrated (low-passed) version of the sequence could also be used. Standard D/A techniques could also be adapted to give a weighted output of the various stages. This could be set to give various amplitude distributions to the output and at an average amplitude level that is independent of clocking frequency.

DESIGN DATA FOR PSEUDO-RANDOM SEQUENCE GENERATION

n	2^n-1	Feedback from Stages:
2	3	1, 2
3	7	2, 3
4	15	3, 4
5	31	3, 5
6	63	5, 6
7	127	6, 7 [4, 7]
8	255	2, 3, 5, 8 (5, 6, 7, 8)
9	511	5, 9
10	1,023	7, 10
11	2,047	9, 11,
12	4,095	6, 8, 11, 12 [3, 9, 10, 12]
13	8,191	9, 10, 12, 13 (9, 11, 12, 13)
14	16,383	4, 8, 13, 14
15	32,767	14, 15
16	65,535	4, 13, 15, 16
17	131,071	14, 17
18	262,143	11, 18
19	524,287	14, 17, 18, 19
20	1,048,575	17, 20
21	2,097,151	19, 21
22	4,194,303	21, 22
23	8,388,607	18, 23
24	16,777,215	17, 22, 23, 24 (*19, 24)
25	33,554,431	22, 25
26	67,108,863	20, 24, 25, 26 (*21, 26)
27	134,217,727	22, 25, 26, 27 (*19, 27)
28	268,435,455	25, 28
29	536,870,911	27, 29
30	1,073,741,823	7, 28, 29, 30 (*23, 30)
31	2,147,483,647	28, 31
32	4,294,967,295	10, 30, 31, 32
33	8,589,934,591	20, 33
34	17,179,869,184	7, 32, 33, 34

NOTES: See note at bottom of page 5h (3). Data in () are from Lancaster. Data in [] indicates alternative feedback - many more are possible as well. The (*) indicates feedback suggested by Lancaster for non-maximum sequences (less than 1% short of maximum however) that require only one EX-OR gate.

CHAPTER 5I

MIXER AND MULTIPLE DESIGN

CONTENTS

INTRODUCTION

Mixers and multiples in electronic music serve to aid patching as well as performing their mixing (summing) function. Multiples are just jacks tied together to serve as a common voltage point. Their design is more human engineering than electrical. Mixers for electronic music synthesizers are basically op-amp summers, and the design is concerned with human engineering of input and output structures. These mixers differ from the usual audio mixers in the following ways:

1) They generally work with high level signals on the order of several volts rather than low level audio signals.

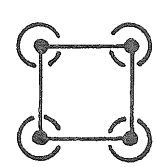
2) The mixers are often used to handle slowly varying and DC voltages, not just audio frequencies. This implies the need for DC coupling.

3) The mixers are often used for purposes other than mixing (e.g., in-line attenuation, inverting, signal distribution).

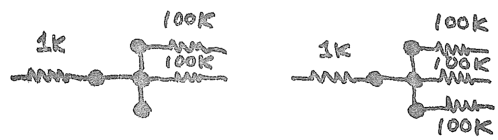
Fortunately, the design of such a mixer is basically just a matter of employing an op-amp summer

DESIGN AND PLACEMENT OF MULTIPLES

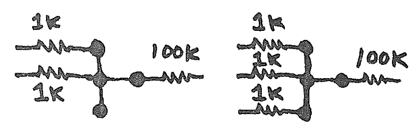
A multiple is nothing more than several jacks connected together to form a mixing point or a distribution point. The most popular multiple is the 4-way as shown at the right. One 4-way multiple for every 20 panel jacks is a reasonable estimate to start with. The 4-way can serve as a voltage distribution point if the distribution is from a low impedence output to several high impedence inputs. It can also serve as a mixer (equal level mix) for up to three standard (1k) outputs to a high impedence input. It can also be used to mix two outputs and distribute them to two inputs. The



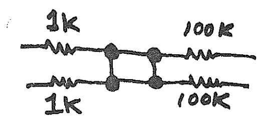
various combinations are shown below:



DISTRIBUTE

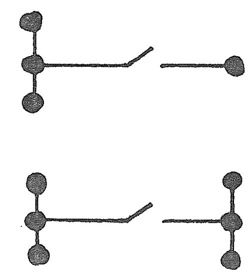


MIX



MIX AND DISTRIBUTE

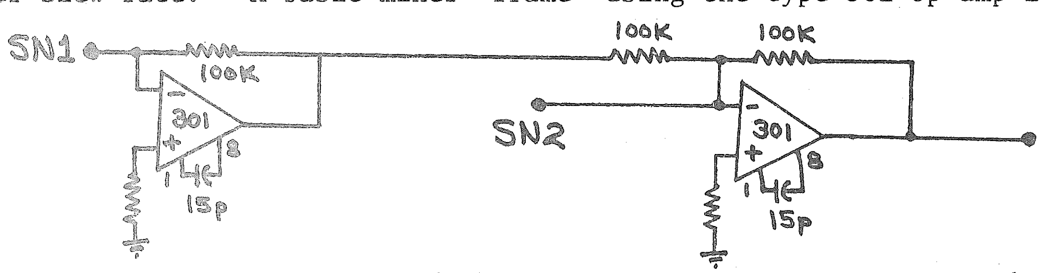
Another use of a multiple is on an input (usually a control input) where two or three jacks may be connected together so that the voltage to the module can be routed to another location without unplugging the first and going through another multiple (saving one patch cord since one of the connections is permanently wired behind the panel). Another useful type of multiple is the switched 3-1 or the switched 3-3 shown at the right. The switch can be used to increase the fan-out of the multiple, but more importantly, the switch can serve as an "in-line" switch where such a switch is handy in a given patch but is not required in a general setup. Switched 2-2 multiples are a good choice in place of 4-way multiples if a small switch is used so that little additional panel space is required. Often, all the jacks of a multiple are located close together on the panel. It is not necessary that this be so in all cases. In fact, a strong case can be made for a multiple where inputs are widely separated and serve as sort of "trunk lines" to route signals around behind the panel. Switches in the line can also be useful here.



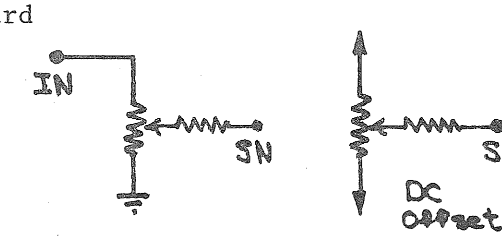
Some switches have a switching arrangement so that the switch is activated when a plug is inserted. These normally closed jacks are useful for a "patch over hard wire" system where where a given patch is in effect until it is defeated by plugging in an overpatch. Such systems can lead to confusion unless carefully studied however. Some interesting multip' s are possible with jacks having normally open switches where plugging on a plug expands th size of the multiple. You then have to skip a jack to start a new multiple.

DESIGN OF MIXERS

The basic mixer is the op-amp summing node structure. Since this is an inverting structure, it is generally useful to follow this with another inverting stage. This makes possible either an inverted or a non-inverted output. It will be shown that there is some advantage to using an externally compenstated op-amp here since it can be custom compensated for a faster slew rate. A basic mixer "frame" using the type 301 op-amp is shown below:

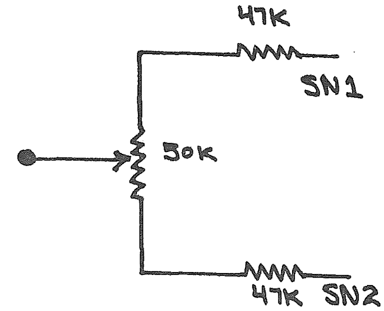


Before considering the compensation of the op-amps, it is necessary to select input structures. The most basic is of course the standard attenuating pot. This can be used to feed a signal into either of the summing nodes. If fed into the second summing node the two outputs are not simple inversions as they are when the second node is left as is. Offest voltages can be fed in with a pot connected across +15 and -15, and this can be very useful when handling control voltages. However, the

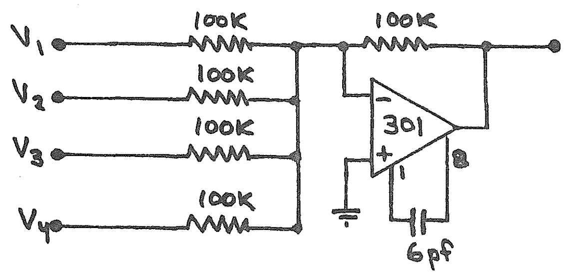


DC offset should generally be removed with a switch to make sure it is really off when not wanted. A DC balance control may be put in as a trimmer to zero the DC level of the overall mixer if this is desired.

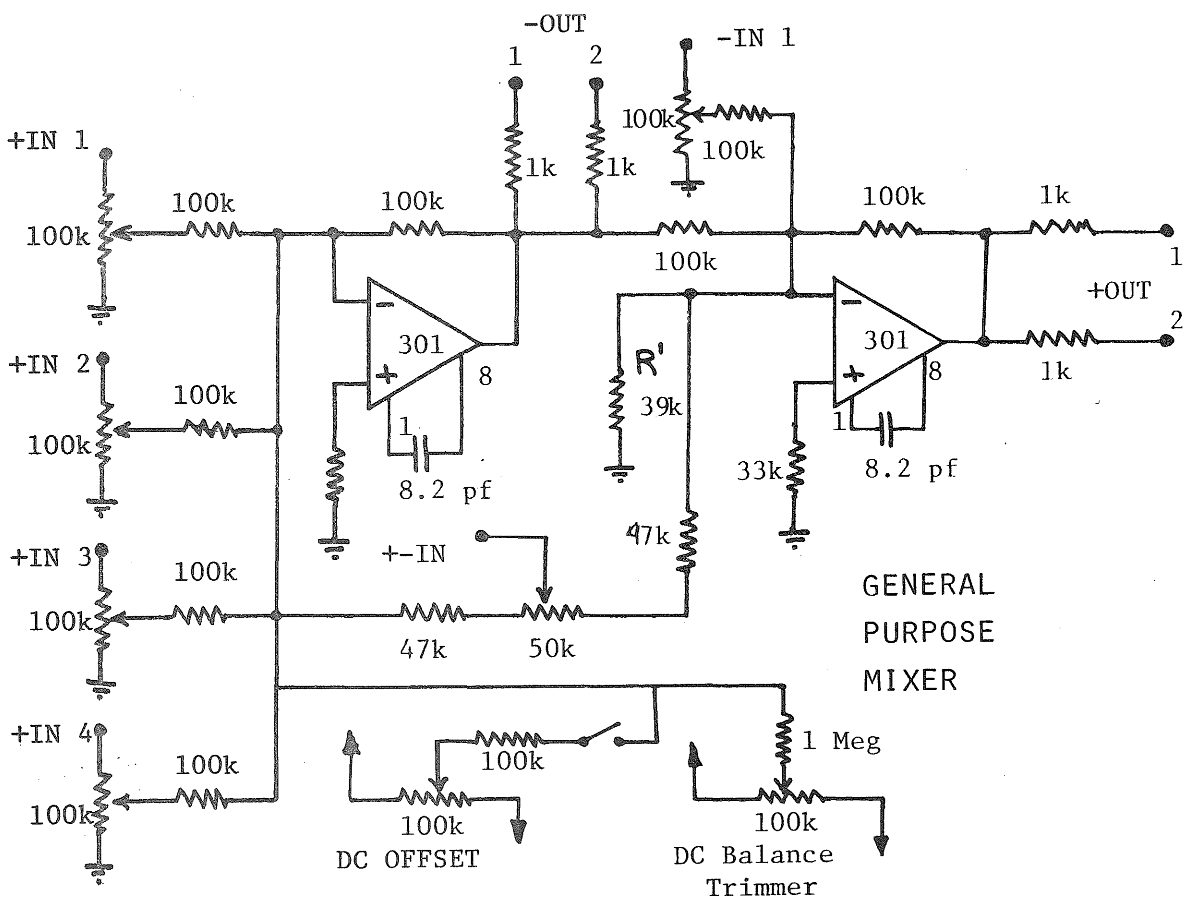
Another interesting connection is the input to both summing nodes off opposite ends of the pot, inputting the signal on the wiper. The device as shown at the right gives the equivalent of a full signal into 100k when the pot is in one extreme, and the full inverted signal with the wiper in the other extreme. With the wiper centered, both currents to the summing nodes balance and cancel out. Thus, it is possible to control the signal from its full value to its inverted version with one control.



Once all the input structures have been selected, the op-amp compensation can be determined and adjusted if necessary to obtain the necessary slew rate for the mixer. Note first that we are using unity gain circuitry with an inverting configuration. As was pointed out in the chapter on op-amps, this means that we can cut the compensation compensation in half due to the attenuation of the feedback stage (2:1) so we can make the capacitor 15 pf for just a start. Jung has further pointed out [Electronic Design 21 #20, Sept 27, 1973] that in the case of an inverting summer with unity gain for each input, all the inputs act as though they were in parallel as far as attenuation of the output back to the - input is concerned. For example, if we have the circuit shown at the right, the compensation can be cut to 1/5 of its unity gain value (thus from 30 pf to 6 pf). This improves the slew rate to over 2 volts/μsec from the value of 0.5 volts/μsec for a 30 pf capacitor. This may well seem like getting something for nothing since this implies that even when we are not using an actual summer, we can just simulate extra inputs with a resistor from the summing node to ground. This is in fact true and can be used to increase the slew rate, but there is a price to pay. In the above example (or its simulated case with one 100k input and a 33k resistor from the - input to ground) the bandwidth is reduced from its unity gain value to a value appropriate for a gain of 5 (14db). This may not seem to important since there is apparently a lot of extra bandwidth still above the audio range, but the important thing is that this 14db is subtracted from the loop gain of the amplifier and with this reduction goes a reduction of the performance parameters that benefit from excess loop gain. In short, it does not pay to push this thing too far - set the compensation for the maximum slew rate needed and not much higher. This reduction in performance can perhaps be understood by considering that the properties of the op-amp in this configuration depend on the - input being a virtual ground; it is at ground potential, but no current flows to it. As we add more resistors from the - input to ground (particularly resistors with values low relative to the feedback resistor as is suggested to get the compensation capacitor down), the - input starts to look more like a real ground, not a virtual one.



We will not attempt here to give a design example of the type given in earlier chapters. This is because mixer design tends to be oriented toward individual needs, and here we will be showing a design that illustrates all the points brought out above. Thus, the example given here is one that you will probably want to cut down. The mixer has 4 non-inverting inputs, a DC balance and a DC level control. It has one inverting input and both inverting and non-inverting outputs. One dual polarity input control is also illustrated. Note that the pot in this case must really be 50k - you cannot scale all the resistors in this case to 100k. We have shown type 301 op-amps in this design, and the compensation (8.2 pf) is smaller than the nominal inverting amplifier case of 15 pf. This is done for A1 due to the increased feedback attenuation due to multiple inputs. Note that we must assume that the inputs may be unused with the pots set to the top, so these are 200k to ground, not 100k. The parallel combination of inputs is thus something on the order of 40k, so the attenuation is 40k/140k or about 0.29, so this says that the capacitor can be about 30·0.29 = 8.6 pf (thus we choose the standard value of 8.2 pf). On the second op-amp, there are fewer multiple inputs so to be sure the total equivalent parallel resistance is smaller than 40k, we add a 39k resistor R' to ground. This allows the same compensation for A2. A similar resistor can be added to A1 if the number of inputs is reduced in a given application. An op-amp such as the 556 could be used to obtain higher slew rate without these compensation tricks. In such a case, R' would not be used, and the compensation capacitors would not be present. With the circuit shown using the 301's, the slew rate is in excess of 1.5 volts/μsec., which is more than enough for 20 kHz ±10 volts sinewave inputs which require about 1.2 volts/μsec.



CHAPTER 5J

POWER SUPPLY DESIGN

CONTENTS:

INTRODUCTION

Power supplies often seem like an imposition - most circuits need one, and they can be both bulky and expensive. Furthermore, once one is built, you still have to build something else to get something that does something interesting. A small device that you may want to carry around needs its own supply, and even if you have a central supply back in your lab with amps to spare, it won't help on the road. At some point, the user has to decide how he wants to physically locate supplies: one central supply or a supply in each piece of equipment.

The basic procedure for power supply design is to first build an unregulated supply and then add a regulator. In general, we will find that it is quite easy to do both. The unregulated supply involves the selection of a transformer, a rectifier, and a filter capacitor. Integrated circuit voltage regulators are then applied to obtain the proper voltage from the unregulated supply. Usually, a few external capacitors are added to stabilize the regulator, and finally various protective devices such as current limiting and overvoltage protection can be added if these features are desired and are not available as part of the normal function of the IC regulator. The first step in the design is to determine the power requirements of the circuits to be powered.

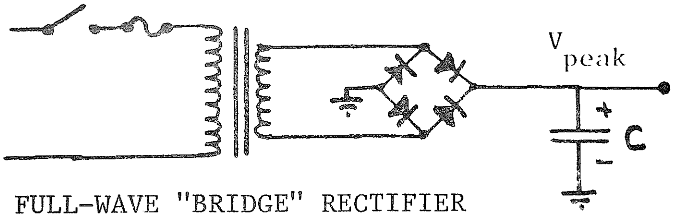
DETERMINING POWER REQUIREMENTS

The basic supply requirements are ±15 volts bipolar for op-amps and other linear IC's, and +5 volts for TTL and other digital circuits. As CMOS digital IC's come into wider use the +5 supply may not be needed, as CMOS can run on voltages from +3 to +18. Other supply voltages may be called for in some designs and must be considered carefully. Four points in this regard should be noted: (1) Power supplies for op-amps are sometimes indicated as ±12 or ±13, but most op-amps will work on ±15. (2) Common positive voltages other than +15 and +5 are +9, +12, +18, and +24. Of these, +12 and +18 devices will often run on +15. The +12 and +9 voltages can also be supplied from +15 by using a Zener diode to drop the excess voltage. (3) A bipolar supply of +12 and -6 was used at one time but is now uncommon. (4) Most digital IC's will run on +5 volts (TTL, DTL, and CMOS). RTL will also work on +5, but it is a good idea to drop the +5 to about 3.8 volts by two series silicon diodes.

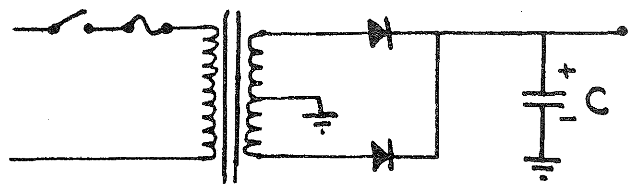
Determining the required current is more difficult. Thus, it is necessary to make some rough estimates and then make sure there is plenty of extra available. Often times you will be adding on other circuits later anyway. Also, designing a regulator for more current than is usually drawn will assure that it will run cooler. With some devices, it is better to limit the current at just slightly more than the device requires when functioning properly. Any excess current will shut down the regulator. In most cases however, it in more convenient to just use a supply with current to spare. When estimating current needs, the following rules of thumb can be used: (1) Allow 2 ma for each op-amp and add to this the current corresponding to +15 volts through the resistance connected to the output of the op-amp (if any). For example, if there is a 4.7k load resistor on the op-amp, when the output is at either +15 or -15, about 3 ma will be delivered to the load. Add to this the 2 ma for the op-amp itself, and you get a total requirement of 5 ma. (2) For TTL, estimate 10 ma for each package of gates, and up to 60 ma for a more complicated TTL package. Consult data books or the TTL Cookbook if you need accurate figures. For slow speed (including audio), CMOS requires very little current and need not be considered. When it gets right down to selecting a regulator, often times the choice is between a low current version (around 100 ma) and a high current version (around 1 amp). Careful consideration of the current requirements is therefore necessary only when the rough estimates come close to one of these limits.

THE BASIC UNREGULATED SUPPLY

As a first step, a transformer can be selected. The transformer should have a standard AC primary and the secondary voltage should be selected to be the same or slightly higher than the output voltage of the supply you are trying to design (e.g., an 18 volt transformer will work for a 15v supply). The current rating of the secondary should be at least as high as the peak current of the supply. The transformer may also be one of twice the voltage with a center tap (e.g., a 12.6V transformer center tapped for a 5V supply). In fact, there are a couple of advantages to this as we shall see. We then proceed to set up an appropriate rectifier circuit using diodes. It is best here to use a full-wave rectifier, not a half-wave, as this will mean that we can use less filtering. There are two choices of full-wave rectifiers, depending on whether we have a center tapped transformer or not.



FULL-WAVE "BRIDGE" RECTIFIER



CENTER TAPPED FULL-WAVE RECTIFIER

Given the choice, the center tapped transformer method is preferred for several reasons: (1) It uses only two diodes (which is not to say the two diodes saved will save you the possible expense of the center tapped transformer). (2) Since there is a voltage drop across any real diode, there is only one such drop for the center tapped transformer, two for the bridge rectifier case. (3) Since only half the transformer and only one diode are used at any one time, the current ratings of these two can be halved. (4) The setup is easily expanded to give the corresponding negative voltage for bipolar supplies as we shall see.

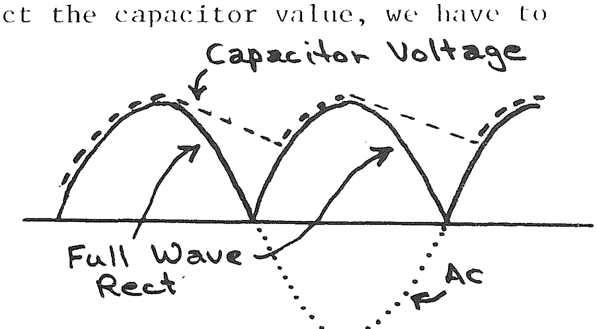
What is the value of V_{peak}? The first thing to realize is that the voltage rating of the transformer is the RMS value, so the actual AC voltage amplitude out is $\sqrt{2} = 1.414...$ times the RMS value. For example, a 6.3 volt filament transformer has an output amplitude of 8.9 volts. This voltage is applied across a diode to the capacitor C. The diode drop is often thought to be 0.6 volts for a silicon diode, but since the diode conducts heavily, it may effectively be higher. A good safe value is twice the standard or 1.2 volts. Next we observe that the transformer voltage charges the capacitor through the diode, and the capacitor voltage follows the transformer voltage minus 1.2 volts as it rises. When the transformer voltage starts to fall, the diode is back biased and capacitor holds the highest positive value it had. Thus, we see that the peak voltage is:

$$\boxed{\begin{array}{l} V_{peak} \simeq 1.4(V_{RMS}/2) - 1.2 \\ V_{peak} \simeq 1.4(V_{RMS}) - 2.4 \end{array}} \quad \begin{array}{l} \text{[center tapped transformer]} \\ \text{[full wave bridge setup]} \end{array}$$

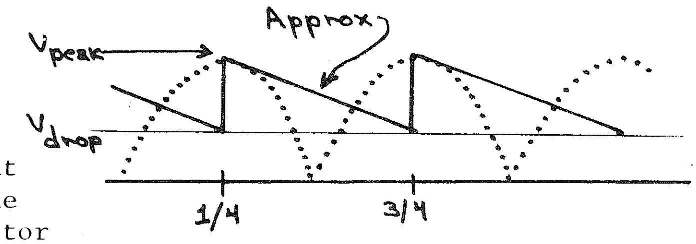
where V_{RMS} is the advertized transformer voltage, and 1.4 is taken as a close enough approximation to the square root of two.

What diodes should be used? In most cases, a diode rated at one amp 100 volts will do (or a higher rating on one or both values). In the case of the center tapped transformer setup, 1/2 amp averages through each diode for a 1 amp supply. The reverse voltage rating of the diode (i.e., the advertized voltage rating) should be at least twice the value of V_{peak}. Note that the maximum voltage across the diode occurs when it is holding back V_{peak} and the transformer voltage is at its negative maximum, and this adds up to about twice V_{peak}. For the highest supply voltages we are considering here (15 volts), the maximum allowable value of V_{peak} is determined by the maximum allowable input voltage to the regulator. In most cases, this value for an IC regulator is 30 - 35 volts. Thus, the diodes should be rated at at least 70 volts, and 100 volts is better. The diodes should be mounted with short leads, but slightly away from any mounting surface. The short leads should conduct heat away from the diodes to a large metal tab on a PC board if possible as this will aid in cooling.

How should the capacitor C be determined? First of all, the voltage rating should of course be greater than V_{peak}. Secondly, the capacitor should have a value as large as is necessary, but not much larger - contrary to the design of unregulated supplies where you usually make it as large as possible. In the case of a regulator, making the capacitor too large will just cause the regulator to run hotter and will not improve anything. To understand how to select the capacitor value, we have to consider how the transformer and diodes charge the capacitor, and what the capacitor voltage looks like when a current is being drawn. We will find that it looks something like the diagram at the right. The transformer voltage first charges the capacitor up during the first quarter of the AC cycle, and then drops out, leaving the capacitor to supply any current to the load. During the third



quarter of the AC cycle, the transformer again charges the capacitor up to V_{peak}. Next recall that the charge on a capacitor is related to the capacitance and the voltage by $Q = C \cdot V$, and current is the time rate of change of charge; $I = dQ/dt = C(dV/dt)$. This tells us that a constant current will discharge the capacitor along a straight line. We can now approximate the actual discharge as shown at the right, bearing in mind that we are playing it safe since the capacitor will actually start recharging before the actual 3/4 point. In IC regulators, there is a minimum input voltage called the dropout voltage below which the output voltage of the regulator can not be maintained. The capacitor voltage must therefore remain above the dropout voltage at all times. The dropout voltage is typically 2 to 3 volts above the regulator's rated output voltage. We can now develope a simple formula for the capacitor size.



The linear discharge section in the diagram above runs from V_{peak} to V_{drop} in a time of 1/120 of a second. Thus we can set a value for the maximum allowable ripple voltage on the capacitor: $V_{ra} = V_{peak} - V_{drop}$. This value V_{ra} is taken to be ΔV while the time of 1/120 of a second is taken to be ΔT. Since we are using an approximation that is a linear discharge, we can replace the differentials in $I = C(dV/dt)$ by ΔV and ΔT and arrive at the equation solved for C as:

$$\boxed{C = \frac{I}{120 \cdot V_{ra}}}$$

where: I is the maximum current to be drawn

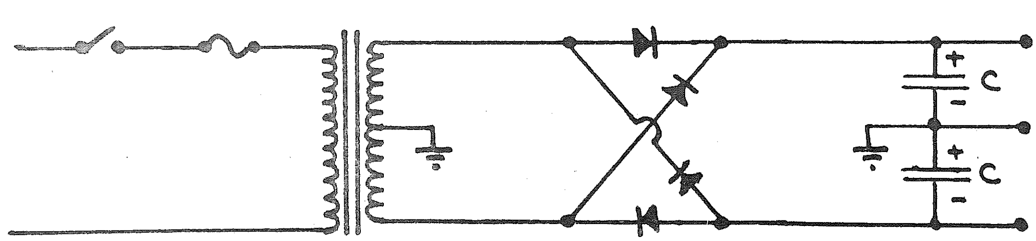
$V_{ra} \approx V_{peak} - (V_{out} + 3)$

V_{peak} is given on page 5j (3)

EXAMPLES: Suppose you want to make a 1 amp 5 volt supply using the LM309 or a 7800 series regulator and a 12.6 volt filament transformer that is center tapped. The peak voltage on the capacitor is $(6.3 \cdot 1.414) - 1.2$ or 7.7 volts. The 5 volt regulator needs 2 volts above the output or 7 volts minimum. This gives an allowable ripple of only 0.7 volts. For a one amp output, the capacitor is $C = 1/(120 \cdot 0.7)$ or about 12,000 mfd. This is a rather large capacitor, but the supply is useful for several reasons: (1) It is often used as a test bench supply and the current required is well under the maximum. (2) The design was on the conservative side and there may be more peak voltage available, and (3) Having such a low peak voltage means that the supply runs with very low heat dissipation and runs very cool.

In this case, we might have considered the full wave bridge. This would have given a peak voltage of $(12.6 \cdot 1.414) - 2.4$ or about 15.4 volts. The allowable ripple is then $15.4 - 7 = 8.4$ volts. Thus, $C = 1/(120 \cdot 8.4)$ or about 1000 mfd. Here, the capacitor is much smaller, but the regulator will run hotter.

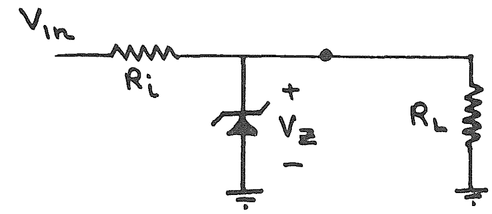
For a bipolar supply, the setup below is used:



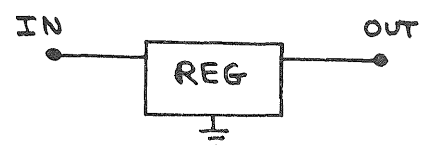
The following points about the bipolar supply should be noted: (1) If we remove the ground connections, this is just a full wave bridge rectifier. (2) It is also really a double sided center tapped transformer setup as well. Two diodes and one capacitor have been added to use the negative transformer voltages as well. We have added just what we would have used in the first place if we wanted a negative supply. (3) The current rating of the transformer must be the same as the output current rating of either polarity, not half the output current rating as would be the case if single polarity supply were used alone. Other calculations remain the same as though the supplies were separate.

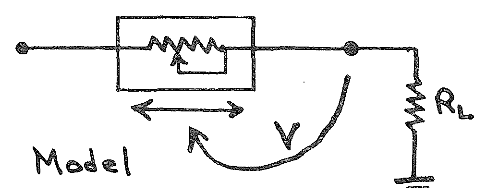
BASIC CONSIDERATIONS FOR REGULATORS

The simplest sort of regulator that can be used is the Zener doide regulator. This can often be used to derive a lower voltage from a standard supply when some odd voltage value is needed. The basic Zener diode regulator is shown at the right. The Zener diode can be thought of as drawing whatever current is necessary so that the excess voltage from the primary source is dropped to the Zener voltage across R_i. If part of the dropping current is drawn by the load, the Zener draws less. Note that the Zener must be prepared to dissipate the full power $V_z \cdot I$ where V_z is the Zener diode voltage, and I is the current through R_i, which is $(V_{in}-V_z)/R_i$. If the primary voltage source is a regulated supply, there is no need for further regulation, and the Zener diode can be used in series to drop the voltage. It will then dissipate only $V_z \cdot I_L$, where I_L is the actual load current being drawn.



Most IC regulators are either three terminal devices or three terminal devices with extra terminals for special functions. The three terminals are quite simply: input, output, and ground. The basic idea is to apply an input voltage which is above the dropout of the regulator, and below the maximum allowable input voltage. The output is then the rated output of the regulator. A simple model for the three terminal regulator is that of a series resistor which has a value determined by the output voltage. Note an important fact that is implied by this model: the regulator must dissipate the excess power that is not delivered by the load. On the other hand, there is no substantial standby current drawn, so for low current requirements, the regulator has very little power to dissipate. In this sense, the IC regulator is more like the series Zener regulator than the shunt regulator. For low currents, the input ripple is small, and the power dissipated by the regulator is essentially $(V_{in}-V_{out}) \cdot I_L$. However, we discussed earlier the fact that the actual input voltage in a proper design has a waveform something like the one shown at the right. This means that the average power dissipated is more like one-half of the larger value. In any case, except for currents that are on the order of 10% of the rated value, the IC regulator should have proper heat sinking. It will be found that most positive regulators are fairly easy to heat sink since the mounting surface is electrically common with ground. Thus, they can be mounted in contact with chassis ground.

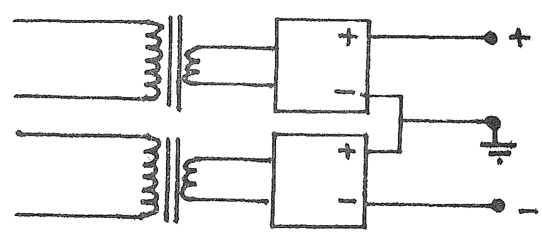




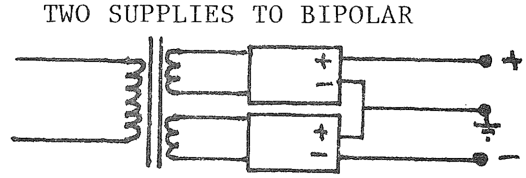
Regulators for negative voltages on the other hand generally do not have the mounting surface common to ground, and it is necessary to provide electrical insulation for the regulator, or to electrically insulate the heat sink. The IC regulators are often able to sense their own temperature and shut themselves down if they get too hot. In operation, it is possible to cautiously place a finger on the regulator to see if it is getting too hot. If you can hold your finger on it, it is probably okay.

IC regulators need some sort of compensation capacitors to keep them from having stability problems. In general, it is necessary to look at the literature on the regulator you are using to find out what type of capacitor is used, how big, and where it is placed. What seems to be most common is a capacitor on the input terminal to ground to be used if the regulator is mounted an "appreciable distance" from the capacitors in the unregulated supply. Since the term "appreciable distance" can refer to anything, you had probably better put them in to be safe. In the regulator circuits that we will be showing, these are generally put in. These are in parallel with the filter capacitors that follow the rectifier, and it is often confusing to beginners to see something like a 0.1 mfd capacitor in parallel with a 5000 mfd capacitor on a schematic diagram. The important thing is the physical placement of the smaller capacitor. These smaller capacitors are on the regulator terminals, and it is important that they are in fact mounted on the terminals and go to a solid ground with short leads. While there capacitors may not seem necessary when the power supply is first tested, they may be needed when it is put on the actual load. The same general thing can be said about capacitors on the output. These are often specified as not being necessary for stability but do improve the transient response. In general, these are a good idea. Some regulators have an upper limit specified for this capacitor; the idea being that the capacitor may draw too much current when first turned on and put the regulator right into current limit. However, many designs are used successfully with much larger capacitors, and in any case, there is often a fairly large accumulated capacitance on the output due to proper supply bypassing down the line. Many of the stabilizing capacitors are specified as being Tantalum electrolytics because they have a power factor more favorable for suppressing RF instabilities. Tantalum capacitors are expensive and somewhat hard to get. However, they should be used where they are specified so it is a good idea to get a few surplus capacitors to have on hand for power supplies. If you don't have the tantalum, try a normal electrolytic in parallel with a 0.1 mfd ceramic capacitor.

IC regulators come in three types: positive regulators, negative regulators, and tracking regulators (positive and negative). These types can be used with external pass transistors to increase their current output. A bipolar supply can be made in several ways. The simplest way conceptually is to make two identical supplies, ground the negative terminal of the first, and the positive terminal of the second. In this case, it is necessary that the two supplies have no electrical points in common until the output stage. This means that the two supplies either have to have their own transformers, or separate windings on the same transformer. A center tapped transformer will not work. The second method is to use the center tapped transformer to form a bipolar unregulated supply, and then to use a positive regulator for one polarity, and a negative regulator for the other side. Alternatively, the unregulated bipolar supply can be input into a tracking regulator.



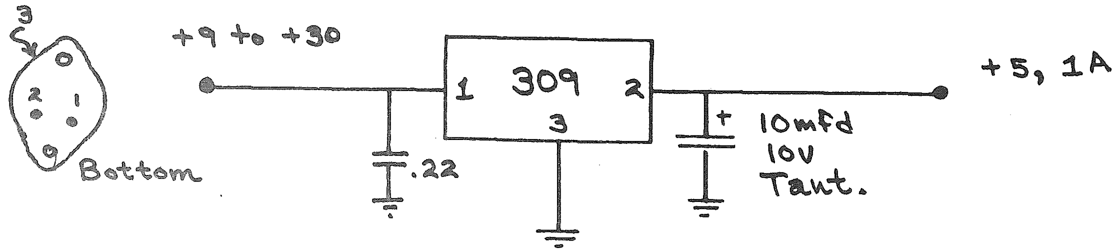
TWO SUPPLIES TO BIPOLAR

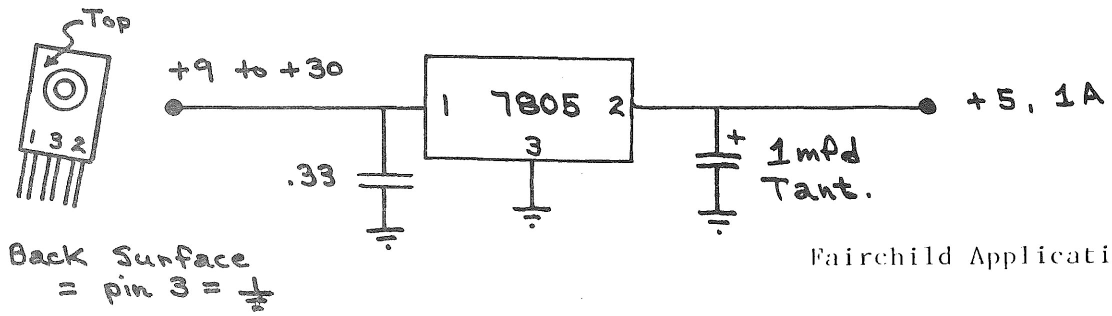


Upmost care should be used to connect regulators carefully. The pin connections can be confusing. If you test the unregulated supply first (as you should), be sure to discharge the filter capacitors before connecting the regulator to avoid costly errors.

INTEGRATED CIRCUIT REGULATOR CIRCUITS

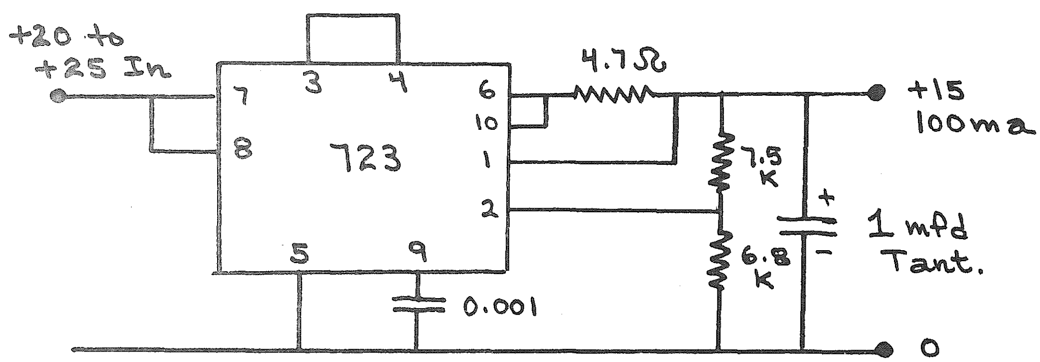
5 VOLT SUPPLIES: Five volt supplies are fairly simple to construct using either the LM309K regulator or the type 7805. The circuits are shown below. If only about 100 ma or less is needed, the LM309H can be used in place of the LM309K.



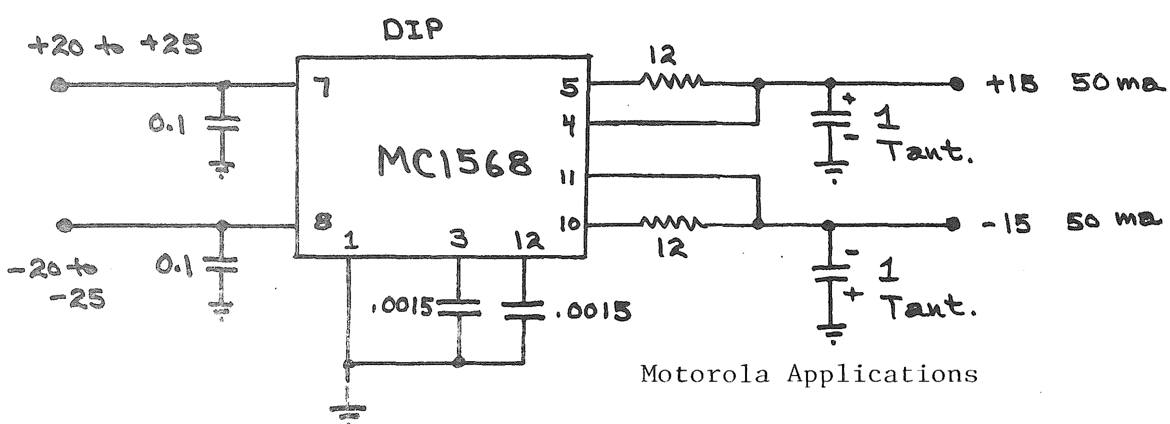


Fairchild Applications

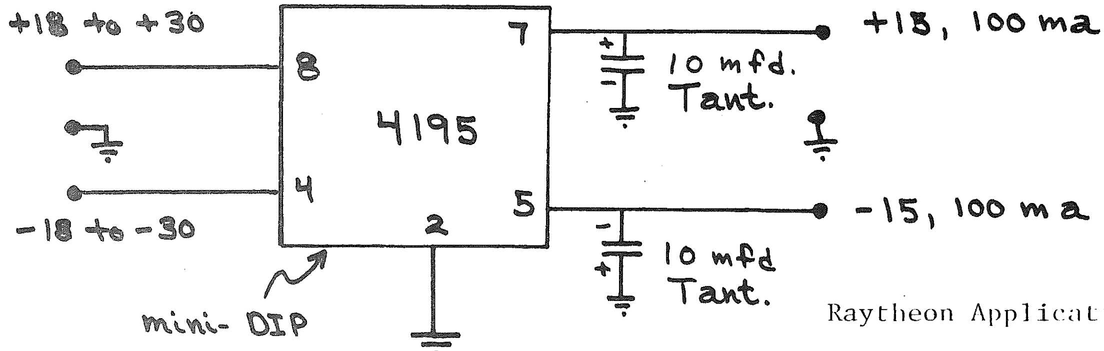
±15 VOLT, LOW CURRENT SUPPLIES: The type 723 regulator has been a favorite for some time. A basic +15 volt regulator is shown below good to 100 ma. Two of these can be used in the manner described on page 5j (6) to give a ±15 volt supply.



A tracking regulator can make a simple bipolar supply for low current requirements. Shown below are circuits using the MC1568 for a 50 ma supply and one using the Raytheon 4195 for a 100 ma supply.



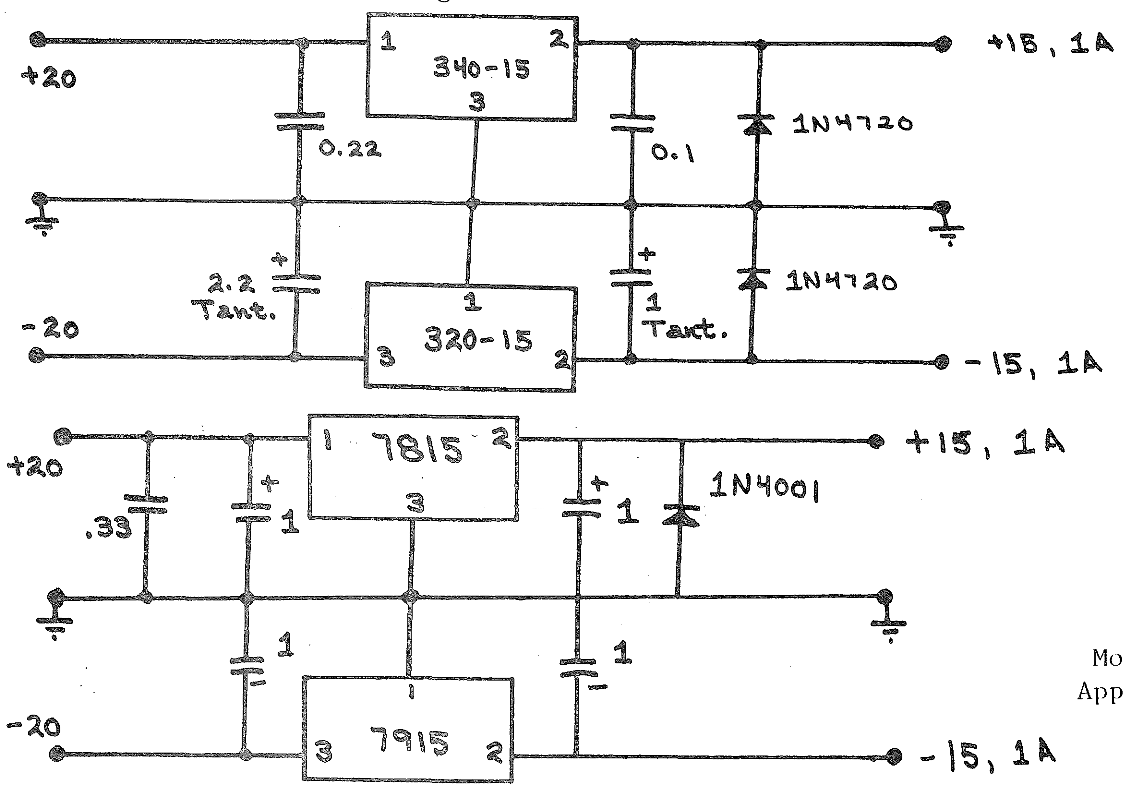
Motorola Applications



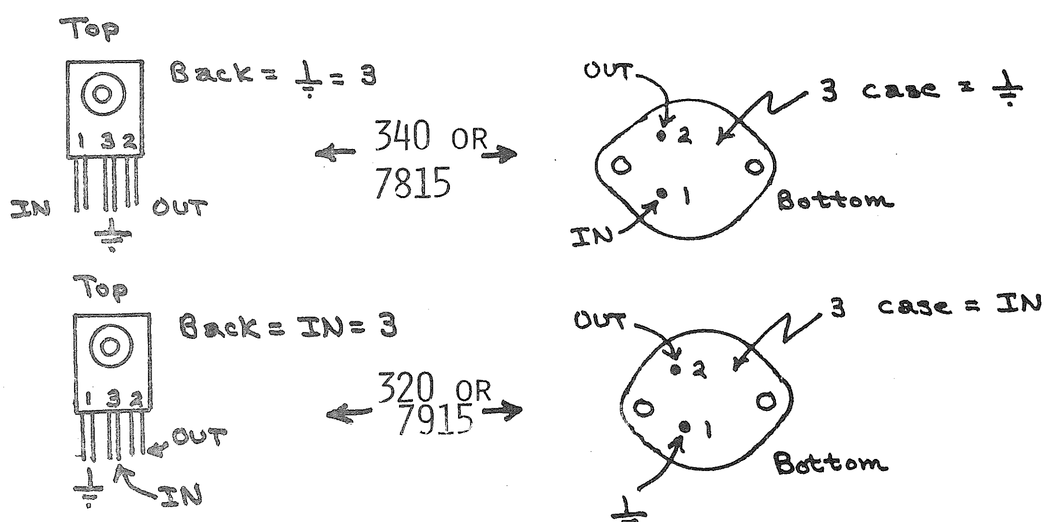
Raytheon Applications

±15 VOLT, HIGH CURRENT REGULATORS:

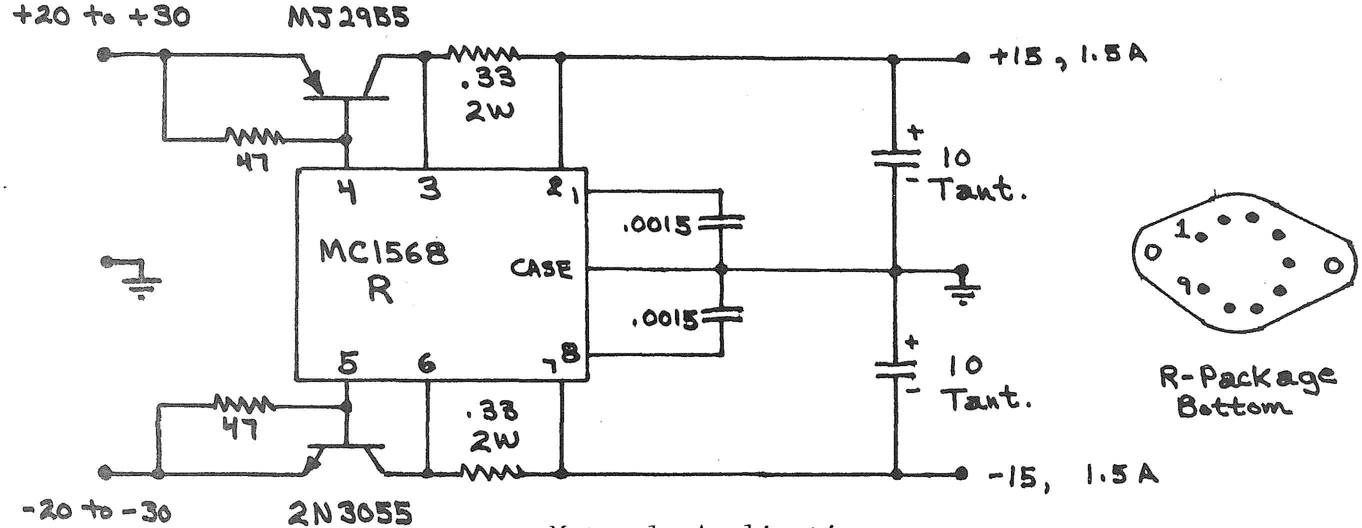
When more current is needed, pass transistors can be added to the tracking regulators or three terminal positive and negative regulators can be used. The basic choices for the three terminal regulators are the type 340 and 320 combination, or the type 7815 and 7915 combination. The regulator circuits are shown below.



Motorola Applications



For higher currents, pass transistors can be added to the three terminal regulators, or to the tracking regulators. A circuit good to 1.5 amps is shown below. A similar circuit for a 2.5 amp supply can be found in the Raytheon applications notes on the type 4195 regulator.



Motorola Applications

PROTECTIVE CIRCUITS

Circuits using IC regulators are often protected fairly well by the regulators themselves. The following comments should prove useful to those who need additional protection.

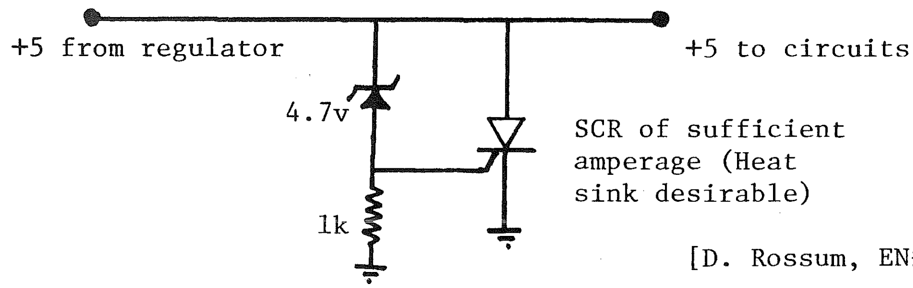
FUSE RATING: The overall supply should be fused at the AC power line. The value of the fuse is determined by power drawn, not current drawn from the supply. For example, a bipolar 15 volts supply at one amp delivers up to 30 watts, but the total power must also include the power dissipated in the regulator. If the V_{peak} value used is 22 volts for example, the power is 44 watts. Allowing for transformer and diode losses, probably about 60 watts is drawn from the AC line. This corresponds to about 1/2 amp, so the proper fuse value would be 3/4 amp or 1 amp of the slow-blow type.

CURRENT LIMITING: Current limiting is the maximum current that the regulator will pass before shutting down. This may be a built in feature of some regulators. Tracking regulators and some others have a current sense terminal. The output current is passed through a small value resistor (less than 1 ohm in many cases). When the voltage drop across this resistor approaches 0.6 volts (or lower if the chip is hot) a blocking transistor is activated to limit the current. The current sensing resistor is thus determined by something like $0.6/I_{max}$.

FOLDBACK CURRENT LIMITING: This is similar to current limiting, but here when the limit is reached (or if a short circuit reduces output voltage to zero), the current is cut back to less than the value it would deliver under a normal load.

THERMAL SHUTDOWN: Most of the newer regulators have the ability to sense their own temperature and shut down if they get too hot. The currents corresponding to this shutdown will depend on the heat sinking of the regulator.

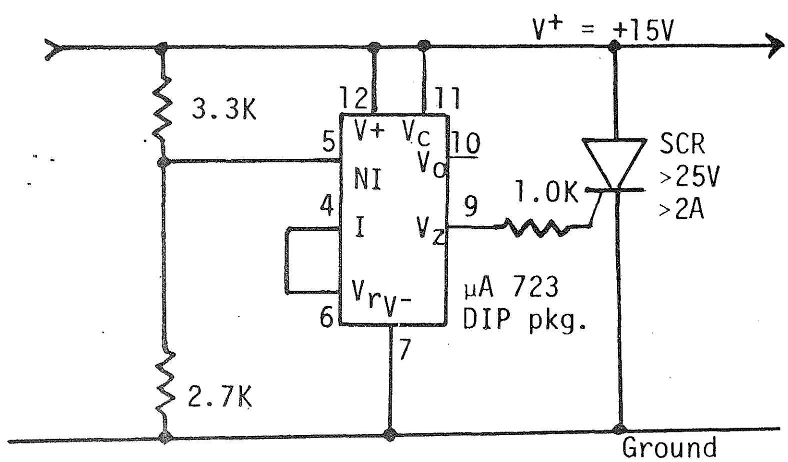
CROWBARS: The above features are mainly to protect the regulators, make them blowout proof, and indirectly to protect the attached circuit as well. A crowbar is a device added to the output to protect the attached circuitry from overvoltage. As soon as an overvoltage appears, the load current is shorted to ground by a SCR, holding the voltage to a very low value and either blowing the fuse or putting the regulator into a shutdown mode. Overvoltage could occur if the regulator failed and the unregulated voltage reached the output line. A far greater danger occurs when a 5 volt line is in the same circuit with a +15 volt line. With any amount of breadboarding at all, it is on' a matter of time before the experimenter will short the two accidently and possibly destroy any TTL devices in the circuit. Thus, it certainly makes sense to add a crowbar to the 5 volt line. A suitable crowbar circuit is shown below and is recommended for any bench supply and for other supplies connected to expensive equipment. For a bench supply, it is very useful to have LED indicators to show when a supply is not putting out voltage as this may indicate that the crowbar has fired.



[D. Rossum, EN#35]

The crowbar circuit can be tested by touching +15 to +5 with no other circuitry attached to the supply. The SCR should fire and remain on even when the +15 is removed. To turn it off, turn the power off. It is also a good idea to see if the crowbar is triggered by any transients such as switches in the equipment or other power line switches nearby. If it is, it may be necessary to increase the Zener diode to 5.2 volts and/or add a bypass capacitor (start with 0.1 mfd) to the 5 volt line at the junction of this line and the Zener diode.

If the attached equipment is very valuable, it is perhaps a good idea to take the following steps in designing and installing a supply. (1) Use only factory first regulators, no' surplus devices. (2) Stay well away from the maximum value for V_{peak} and draw less curren than the rated amount. (3) On first installation, add diodes to the supply lines at the input of the equipment to prevent damage from accidental supply reversal. This is important if the supply is often attached and detached. In a final installation, the diodes can be removed. (4) A crowbar can be added to the 15 volt lines as well as the +5. A crowbar of this type is shown below:



To protect the -15 volt supply line, connect the +15 side of the circuit to ground, and the ground side to the -15 supply line.

[D. Rossum, EN#45]

Chapter 6a

FREQUENCY SHIFTER DESIGN

CONTENTS:

INTRODUCTION:

The terms "frequency shifter" and "spectrum shifter" are used for the same devices. To be exact, it would probably be better here to call them spectrum shifters, because they do shift the entire spectrum, be it a complex set of spectra components or just a single frequency.

It should be realized that a frequency shifter or a spectrum shifter are not devices that transpose a frequency up or down by a preset musical interval. Rather, they shift each spectral component by the same number of Hertz. This will in general mean a drastic change in tone color. For example, if a sawtooth wave is shifted up by 50 Hz, and its frequency to start with is 200 Hz, then the components in the original sawtooth:

200 Hz, 400 Hz, 600 Hz, 800 Hz, ...

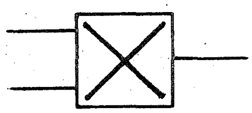
will be shifted to:

250 Hz, 450 Hz, 650 Hz, 850 Hz, ...

This means that the components are shifted from their positions as integer harmonics to non-harmonic ratios. This is useful for producing some different tone colors, but there are easier ways than spectral shifting (e.g., balanced modulation) to produce this general type of sound. The value of a spectrum shifter lies more in the processing of real (live) sounds with which the listener can easily relate, since he knows how the live sound sounds unaltered. There are also numerous other applications where the spectrum shifter is a part of another more general device or patch.

METHODS OF FREQUENCY SHIFTING:

The Double Hetrodyning Method: Frequency shifting is the exact same process as radio single-sideband modulation. The difference in the audio case lies mainly in the fact that the two signal frequencies are on the same order of magnitude. One method of frequency shifting is very similar to radio single sideband. In the radio case, the program material and the RF signal interact in what is called a "mixer," and this results in the sum and difference frequencies being produced. Thus, we know that the use of the term "mixer" in the radio case is different from the electronic music case, since our mixers are linear, and we do not want one to produce sum and difference frequencies. What is really being considered is some sort of non-linear element, and it is the square term in the polynomial representation of the non-linear curve that is useful. For electronic music, we simply use an analog multiplier directly. In fact, the analog multiplier (balanced modulator) is the basis of most frequency shifting schemes. We can take a look at the multiplier and its equations for reference in the discussion to follow:

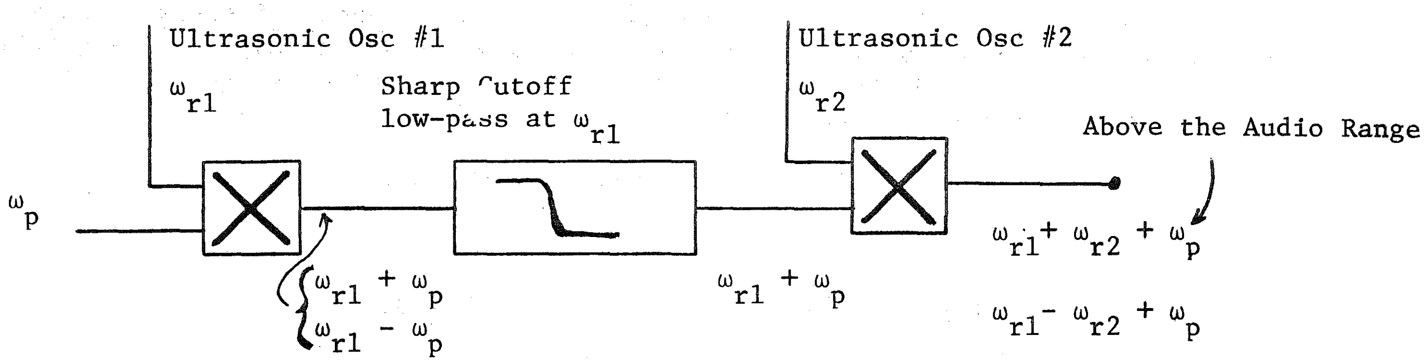


$$\operatorname{Sin} x \cdot \operatorname{Sin} y = (1/2)[\operatorname{Cos}(x-y) - \operatorname{Cos}(x+y)]$$

$$\operatorname{Cos} x \cdot \operatorname{Cos} y = (1/2)[\operatorname{Cos}(x+y) + \operatorname{Cos}(x-y)]$$

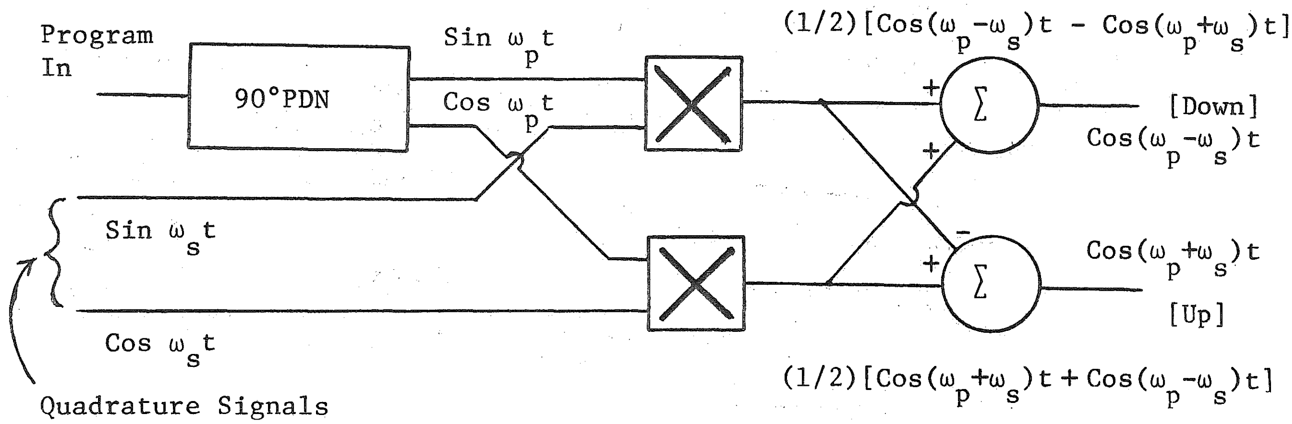
$$\operatorname{Sin} x \cdot \operatorname{Cos} y = (1/2)[\operatorname{Sin}(x+y) + \operatorname{Sin}(x-y)]$$

In the double hetrodyning method, the signal to be shifted is multiplied with a ultrasonic signal. This produces sum and difference frequencies that are both ultrasonic. Next, one of the sidebands is filtered out. Finally, the remaining sideband is multiplied with another ultrasonic oscillator, and this produces two more sidebands, one which is shifted way up higher into the ultrasonic, and the other may well end up in the audible range. The general setup is shown below. For simplicity, only the frequencies are shown on the diagram, all the waveforms are assumed to be sines:



Note that the frequency shift can be either up or down, depending on the frequency difference $(\omega_{r1}-\omega_{r2})$. A major limitation of the method is that the lower limits on the shift are determined by the sharpness of the low-pass filter and the degree of unwanted sideband feedthrough that can be tolerated.

<u>The Phase Shift Method</u>: The phase shift method is based on the use of two multipliers, quadrature signals, and two summers. The scheme is shown below:



An important part of the design is the 90° Phase Difference Network (90°PDN) and a good deal of discussion will be given to these networks. This will be shown in the reprint article that follows, and additional discussion will follow that. What we want to discuss first is the method of providing the quadrature shifting signals (Sin $\omega_s t$ and Cos $\omega_s t$).

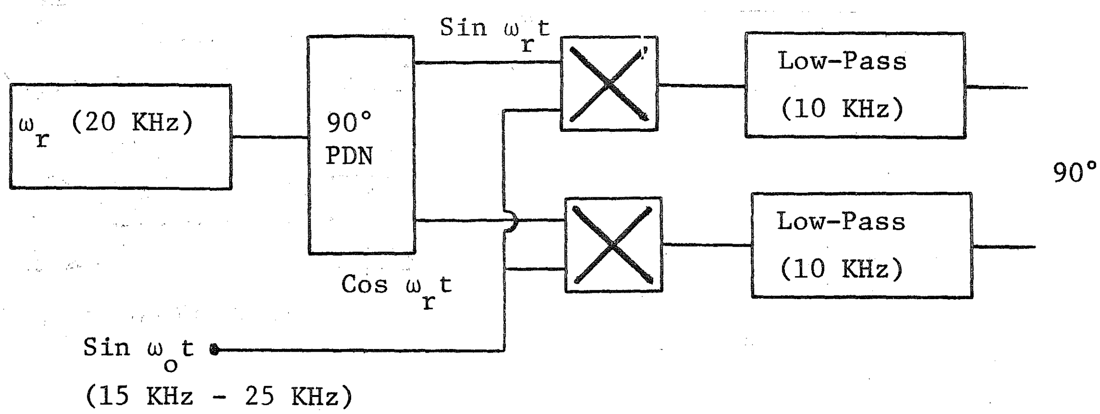
<u>Methods of Providing a Quadrature Signal</u>: The various methods of providing the quadrature signal that we want to discuss are:

1. A quadrature oscillator.
2. A sinewave oscillator with a 90°PDN.
3. The hetrodyning method.

There are several methods of designing a quadrature oscillator. A method based on a voltage-controlled four-pole network is discussed in the chapter on VCF design. This is a useful approach as it can be used as a VC Low-pass filter when not used as a quadrature oscillator. The frequency shift is just the range of the quadrature oscillator, and can be made very small with proper design of the oscillator.

In the second method, a 90°PDN is used to provide the the quadrature signals. The range of useful operation is defined by the overlap of the oscillator range and the bandwidth of the 90°PDN. Extension of this to very low frequencies is difficult due to the size of the capacitors in the phase shifter. In favor of this method, we can see that any available sinewave oscillator can be used, and inputs other than sinewaves can be used for the shifting signal resulting in complex shifting patterns.

A third method has been described by Bode and Moog ["A High-Accuracy Frequency Shifter for Professional Audio Applications," JAES 20 No. 6, (1972)]. This method allows the shifting of frequency through zero shift. [Any of the above methods permit the shifting of the downshifted signal through zero frequency and back up again.] This frequency shifter can thus serve as a frequency modulator by shifting the input signal about its original frequency. Also, shifts very close to zero can be obtained without going near to the limits of the shifting oscillator since zero shift is just a frequency in the middle of the shifting oscillator's range. The basic scheme is shown below:



The actual device used by Bode and Moog had a low-pass filter on the reference oscillator, and a squelch circuit that passed the reference signal only when there was an input to the program inputs of the basic scheme. Two further points should be made: [1] The 90°PDN for this application needs only one pole, and need not be a wide band network, since it works at only one frequency (ω_r). [2] The two low-pass filters at $\omega_r/2$ should be identical, otherwise they may create a phase difference that messes up the phase properties created by the multipliers.

In the above scheme: ω_r is a reference frequency on the order of 20 KHz.

ω_o is a frequency input as a sine wave in the range of 15 KHz to 25 KHz.

The two equations for the multipliers are thus:

$$\mathrm{Sin}(\omega_r t)\cdot\mathrm{Sin}(\omega_o t) = (1/2)[\mathrm{Cos}(\omega_r-\omega_o)t - \mathrm{Cos}(\omega_r+\omega_o)t]$$

$$\mathrm{Cos}(\omega_r t)\cdot\mathrm{Sin}(\omega_o t) = (1/2)[\mathrm{Sin}(\omega_r+\omega_o)t - \mathrm{Sin}(\omega_r-\omega_o)t]$$

The frequency $\omega_r+\omega_o$ is thus in the range of 35 KHz to 45 KHz. The frequency $\omega_r-\omega_o$ is thus in the range of 5 KHz to -5 KHz. We shall call $\omega_r-\omega_o$ by ω_s and shall consider the significance of a negative value of ω_s shortly. First, we note that with the 10 KHz low-pass filters, the two outputs of the filters are (since the $\omega_r+\omega_o$ term is blocked):

$$\left.\begin{array}{r} \mathrm{Cos}\ \omega_s t \\ -\mathrm{Sin}\ \omega_s t \end{array}\right\} 90°$$

When $\omega_s < 0$, the outputs of the filters become:

$$\left.\begin{array}{r} \mathrm{Cos}(-\omega_s t) = \mathrm{Cos}(\omega_s t) \\ -\mathrm{Sin}(-\omega_s t) = \mathrm{Sin}(\omega_s t) \end{array}\right\} 90°$$

Thus, when we return to the original phase shift scheme, and apply this quadrature signal, the multiplier that gets the two sines has the equation:

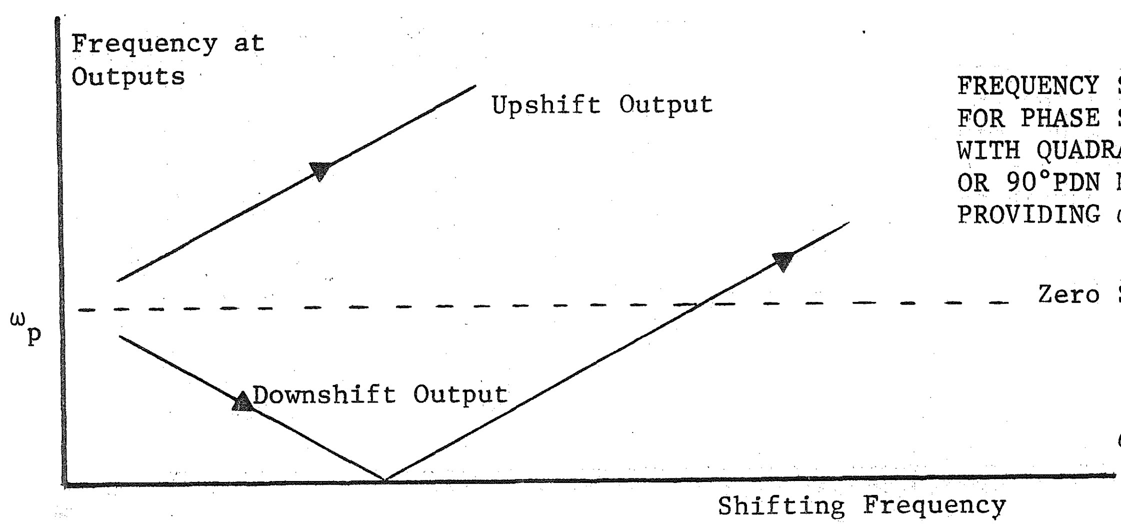
$$\mathrm{Sin}\ \omega_p t\cdot(-\mathrm{Sin}\ \omega_s t) \text{ in the case } \omega_s > 0$$

and

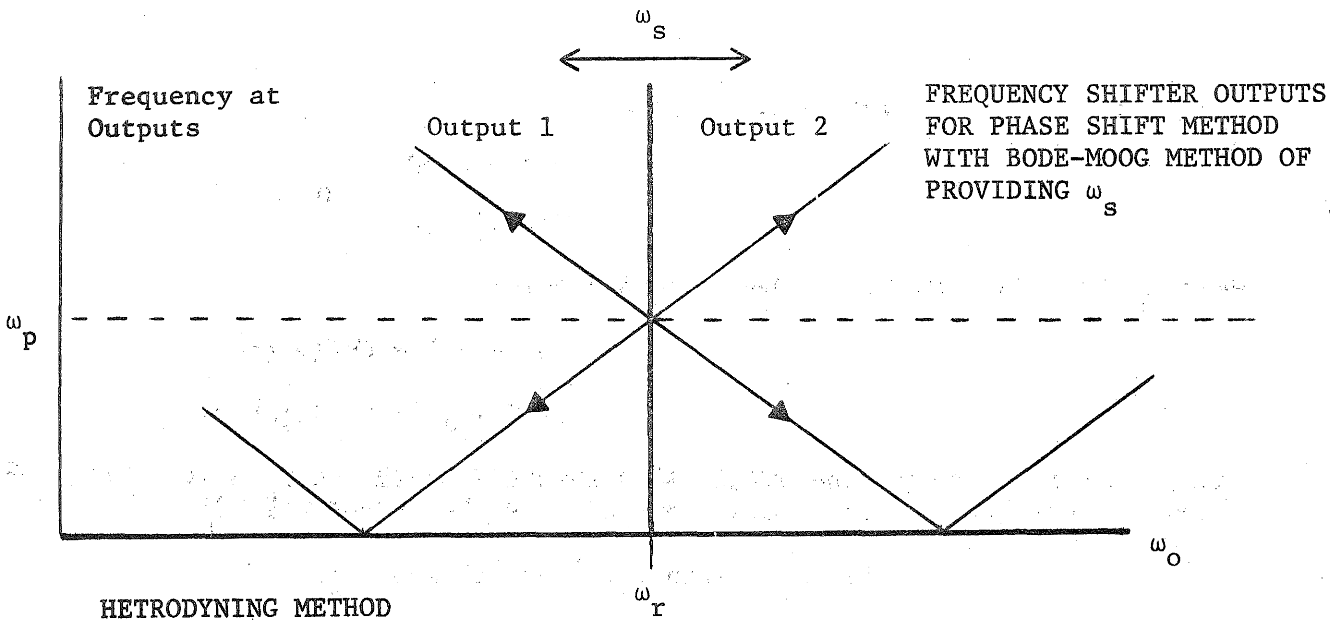
$$\mathrm{Sin}\ \omega_p t\cdot(\mathrm{Sin}\ \omega_s t) \text{ in the case } \omega_s < 0$$

The net effect of the negative values of ω_s is to change the polarity at the output of one of the multipliers in the basic phase shift scheme, and this changes the functions of the two summers so that the upshift and the downshift appear at the outputs of different summers. This means that looking at the output of any one summer as ω_s goes through zero, the output changes from an upshift to a downshift or visa versa. The overall shifting result for either the quadrature oscillator (or 90°PDN) method of providing the quadrature signal, or the hetrodyning method is illustrated below. The principle differences to note are:

[1] The hetrodyning method can easily reach zero shift.
[2] The shifting can go either direction, and thus the same output can shift through zero shift.



FREQUENCY SHIFTER OUTPUTS FOR PHASE SHIFT METHOD WITH QUADRATURE OSCILLATOR OR 90°PDN METHODS OF PROVIDING ω_s



Reprint: "Design of 90° Phase Difference Networks and Application to Frequency Shifter Design," by Bernie Hutchins, reprinted from EN#43

Most people who built the "Poor Man's Frequency Shifter" in EN#15 found the biggest problem to be the errors in the quadrature oscillator which was unstable. With the development of the new quadrature oscillator in EN#41, it is appropriate to look at the frequency shifter again. This time, we will find it is the 90° phase difference network (90°PDN) that must be improved.

It is important first of all to understand that a 90°PDN is not a 90° phase shifter, as 90° phase shifters are restricted to a single frequency of operation and we would like to operate over a wide bandwidth. The only phase shifter that gives the same phase shift over all frequencies is the trivial case of the 180° phase shifter, i.e., the inverter. The 90°PDN is a device which (for an input frequency component f in a bandwidth F_L to F_H) gives two outputs such that the phase shift of one will be ϕ, and the phase shift of the other will be $\phi + 90° \pm \delta$, where δ is an error term whose maximum value is ϵ. Typically, ϕ will be several hundred degrees or more, and ϵ will be from 1° to 5°. The potential bandwidth $(F_H - F_L)$ and accuracy (small ϵ) are traded off against each other, and if either or both must be improved, the 90°PDN must have more poles.

We can now immediately jump into the fire by starting with the inverter circuit and ask if there is any way to combine the 0° signal (input) and the 180° signal (output) to get an intermediate phase shift. At the same time, we don't want any network attenuation that depends on frequency (we don't want any filtering). Combining the 0° and 180° signals with resistors won't work as we would just get signal cancellation and no new phase shifts. Inductors are out, as they always are at audio frequencies, due to size and weight. Thus we try something with capacitors such as shown in Fig. 1. At low frequencies, the capacitor is effectively out of the circuit, and the 0° signal is fed through R to the buffer to give 0° phase shift out. At high frequencies, the capacitor is a low reactance path, and the resistor is effectively out of the circuit. What about inbetween frequencies? Well, recall that the current into a capacitor always leads the voltage across it by 90°. Also, since this is a series RC circuit, the current in both R and C must be the same. Furthermore, the current through a resistor is always in phase with the voltage

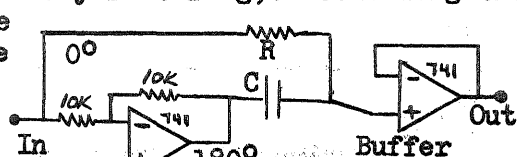


Fig. 1

across it. At this point, the so called "Phasor" diagram becomes useful. The phasor diagram is one in which the magnitudes of voltages, currents, and impedences are indicated by the length of a vector, and relative phases are indicated by angular orientation. For the RC series circuit, the phasor diagram is shown in Fig. 2. Also, since this is a series circuit, the two voltages V_C and V_R must add (vectorially) to the applied voltage, which in Fig. 1 is $2V_{in}$ (connected between $+V_{in}$ and $-V_{in}$). This is shown in Fig. 3, and in Fig. 4, we flip the phasor diagram around a little to show the corresponding situation.

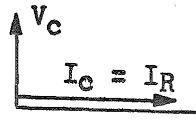


Fig. 2

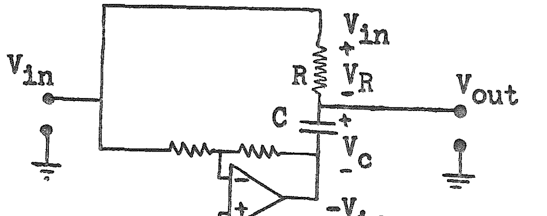


Fig. 3

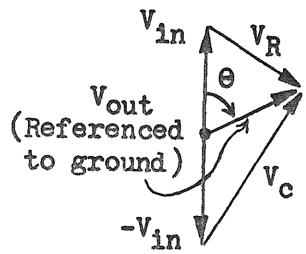
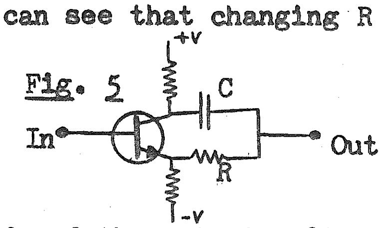


Fig. 4

Angle between V_R and V_C is always 90°. The angle between V_{in} and V_{out} is θ

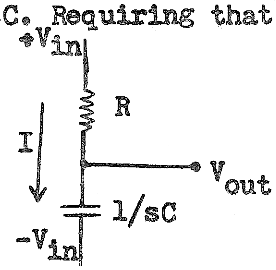
V_{out} thus follows the arc of a circle as frequency changes, due to the change in V_C. Thus, the phase shift depends on frequency, or you can see that changing R or C will change the phase shift for a fixed frequency. This basic circuit is often refered to as the "classical" phase shift circuit. Another popular version of this same circuit is shown in Fig. 5, where a single high gain transistor serves to provide a buffered input, and to give both an inverted and a non-inverted signal.

Fig. 5



In addition, the phasor diagram shows that the magnitude of the output voltage is a constant, a fact that you can easily see by applying a little geometry and requiring that the angle between V_R and V_C be 90°.

The s-plane analysis (See EN#41, Sect. 3) is also useful here, and is often used with this type of network. We just treat the capacitive reactance as a "resistor" of value 1/sC. Requiring that the current in the series elements be the same, we get:



$$I = \frac{2\ V_{in}}{R + 1/sC} = \frac{2\ V_{in}\ sC}{1 + sCR}$$

$$V_{out} = V_{in} - IR = V_{in}(1-sCR)/(1+sCR)$$

$$T(s) = V_{out}/V_{in} = (1 - sCR)/(1 + sCR)$$

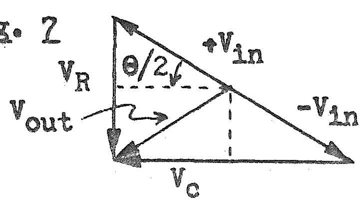
The pole of T(s) occurs at $s = -1/RC$, as the denomenator goes to zero for this value. At the pole, the capacitive reactance has magnitude $1/sC = R$. Thus, the magnitudes of V_R and V_C are the same, and we can see from the phasor diagram that the angle of V_{out} becomes 90°. Thus, the section has a phase shift of 90° when the input frequency is the pole frequency, i.e., when $2\pi f = 1/RC$.

A network with a transfer function of the form $(1 - sCR)/(1 + sCR)$, or the product of two or more such terms, is called an "All Pass" network. This is because of its transfer function along the positive imaginary axis (the frequency response) which is:

$$T(j\omega) = \left[(1-j\omega CR)\cdot(1+j\omega CR)/(1+j\omega CR)\cdot(1-j\omega CR)\right]^{\frac{1}{2}} \equiv 1$$

The transfer function is thus unity, independent of frequency, and passes all frequencies without attenuation. Were it not for the fact that we are interested in the phase properties of the network, a simple piece of wire would have sufficed to this point!

Fig. 7



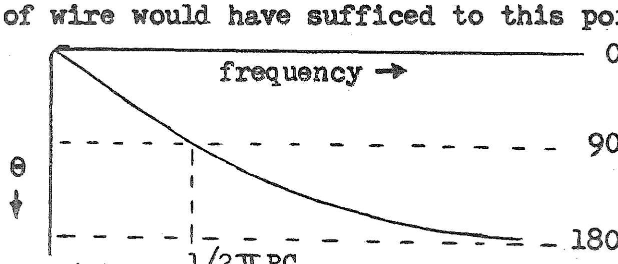


Fig. 8

We have argued that the phase changes from 0° at the lower frequencies to 180° at higher frequencies, but what about the general solution? Again we resort to the phasor diagram. From Fig. 7, we can see that Tan $(\theta/2) = V_R/V_C$, which in turn gives:

$$\theta = 2 \text{ Arctan } (IR)/(I\cdot\frac{1}{sC}) = 2 \text{ Arctan } (sCR)$$

thus the phase varies as the Arctan as indicated in Fig. 8.

Next, we can consider obtaining additional phase shifts beyond 180° by cascading additional sections. We might get something by cascading two sections (Fig. 9) that would look like Fig. 10, but if we are clever enough we may be able to get one phase roll-off to join with the next to give a relatively smooth shift (Fig. 11).

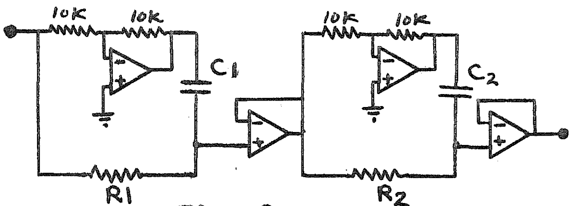


Fig. 9

Fig. 10

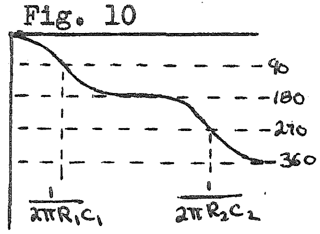
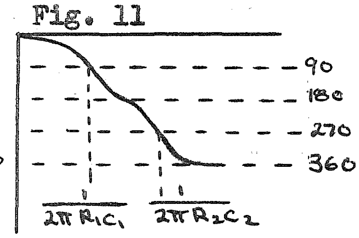


Fig. 11



This tells us how to make a single network (composed of cascaded sections) have a smooth phase change. If we add a second network, like the first, but with poles in displaced positions, we can make this lag the other by any angle, which may be 90°. Such a situation is shown in Fig. 12, where the two networks are called A and B. The diagram is exaggerated to show the meaning of the various terms we used above. The appropriateness of the terminology "Phase Difference Network" should be clear.

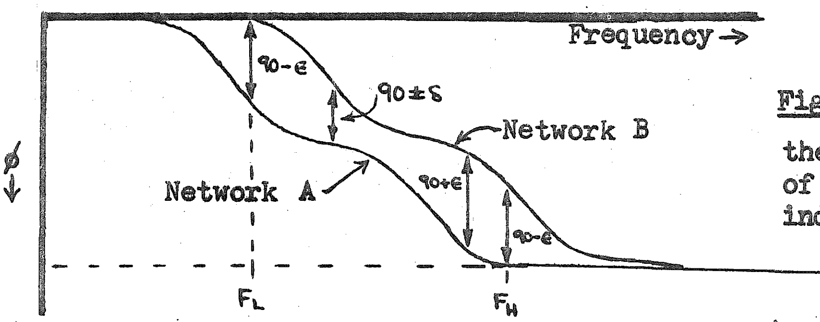


Fig. 12

the angle ϕ is the sum of the θ angles of the individual sections

The procedure is then to determine the pole positions so that the phase difference function ripples about 90° in a prescribed bandwidth. The theory here is quite involved and is neither necessary for understanding and successful design, nor particularly illuminating, so unless you have a fancy for such things as Cauer functions, elliptic functions, conformal transformations, and the like, you will probably prefer simple design tables. The full theory is available elsewhere[1,2], and we will make use of the design procedure of Weaver[3] to determine pole locations. Another method of determining pole locations is given by Albersheim and Shirley[4]. The specific network we shall use as a design example is perhaps more complex than is necessary for this application, as it was originally designed and built for a test of the method rather than for the actual frequency shifter. Alternative design tables are given in Appendix A for those who feel that relaxed specifications will suffice.

DESIGN EXAMPLE: Suppose we want to cover a band from 15 Hz to 15 kHz with a maximum phase error (assuming zero component tolerance) of 0.2°. When we build with real components, we will of course obtain more error. For this ratio of F_H/F_L (1000:1) a design graph[5] tells us that we need a total of 12 poles, 6 for Network A and 6 for B. We follow steps 1 to 4 of Weaver's method to arrive at pole positions, and this is all we need. These normalized poles are given in Table 1 below, and are in terms of F_L, so we multiply them by F_L (15 Hz) to get the actual pole frequencies. Next, since the poles of the network we want to use are at $s = -1/RC$, we take $2\pi f = 1/RC$ to solve for the RC time constants of the pole frequencies. Finally, select a convenient value for C (i.e., what's on the shelf) and solve for the corresponding R. Since this R must supply bias current to the op-amp buffer, it is well to keep $R\cdot I_{bias} \ll$ usual voltage levels handled. Values under 100k should be fine in most cases.

TABLE 1: Parameters of 12 Pole 90° Phase Difference Network

$F_L = 15$ $F_H = 15{,}000$ Phase Error approximately 0.2°

	Normalized Pole	Pole Freq.	RC	Chosen C (mfd)	R (k ohms)
Network A:	.3609	5.41	2.94×10^{-2}	.47	62.6
	2.7412	41.1	3.87×10^{-3}	.47	8.24
	11.1573	167	9.51×10^{-4}	.047	20.3
	44.7581	671	2.37×10^{-4}	.0082	28.9
	179.6242	2694	5.91×10^{-5}	.0082	7.20
	798.4578	11977	1.33×10^{-5}	.0082	1.62
Network B:	1.2524	18.8	8.47×10^{-3}	.47	18.0
	5.5671	83.5	1.91×10^{-3}	.047	40.6
	22.3423	355	4.75×10^{-4}	.0082	57.9
	89.6271	1344	1.18×10^{-4}	.0082	14.4
	364.7914	5472	2.91×10^{-5}	.0082	3.55
	2770.1114	41552	3.83×10^{-6}	.0082	.467

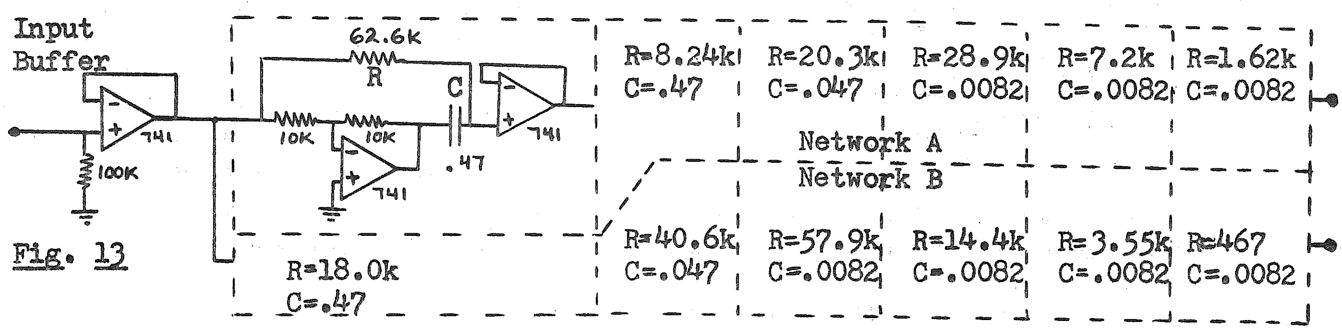
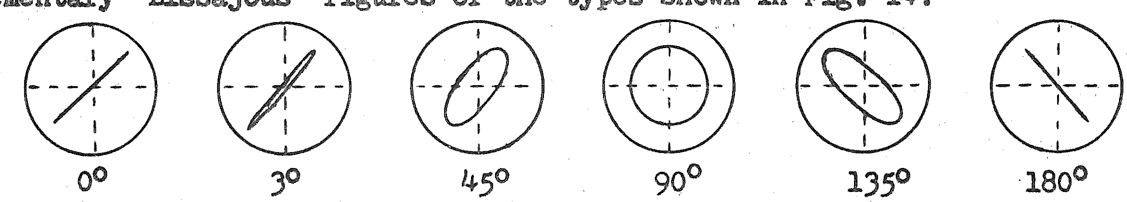


Fig. 13

FINE TUNING: When first assembling the design example, the closest 5% resistors can be used, with fine tuning to come later. As you probably know, an oscilloscope can be used to display phase angles by applying one signal to the X input, and the other signal of the same frequency, but different phase, to the Y input. This method forms elementary "Lissajous" figures of the types shown in Fig. 14:

Fig. 14



While the circular pattern (90°) is a good rapid check for basic operation of the overall network (Fig. 15), the eye is not accurate enough to judge minor variations in the circle. For fine tuning, it is better to use the quadrature oscillator (QO) and the fact that the phase shift of any section is 90° at its pole frequency. This allows the use of a straight line pattern (0°) as a tuning indicator by use of the hookup of Fig. 16.

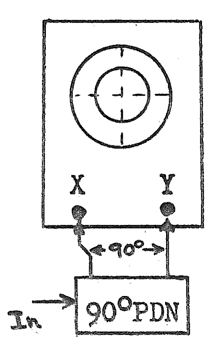


Fig 15

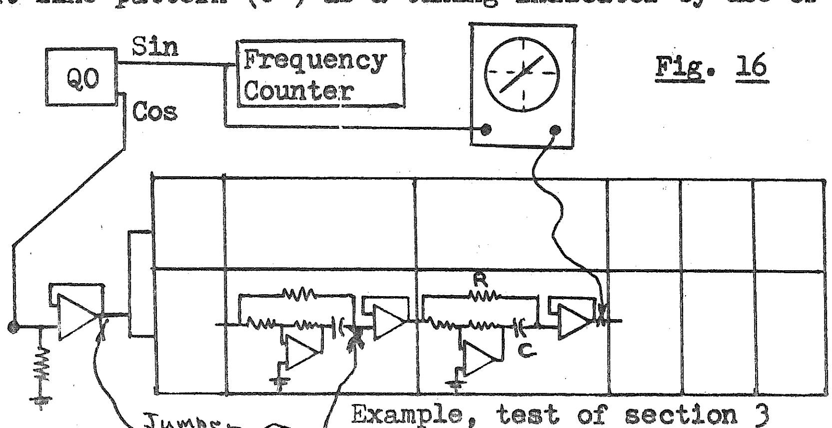


Fig. 16

Example, test of section 3 of the B Network

The sine signal from the QO is fed to the 90°PDN. The signal from the input buffer is then jumpered to the input of the output buffer of the stage proceeding the one under test. The QO is then set to the pole frequency and the R value is adjusted for the in-phase (0°) straight line pattern. If you have a shoe box full of excess trim pots, you may want to use them for making this adjustment, but usually, given a little time and patience, an extra series or parallel resistor will be the most that you will need. Keep this in mind when designing a circuit board.

THE FREQUENCY SHIFTER: Now, if you have the QO and the 90°PDN operational, you need only two balanced modulators (multipliers) and a couple more op-amps to complete the frequency shifter. We have carried multiplier circuits many times in the past, but for the sake of completeness, we will show the multiplier circuit used below:

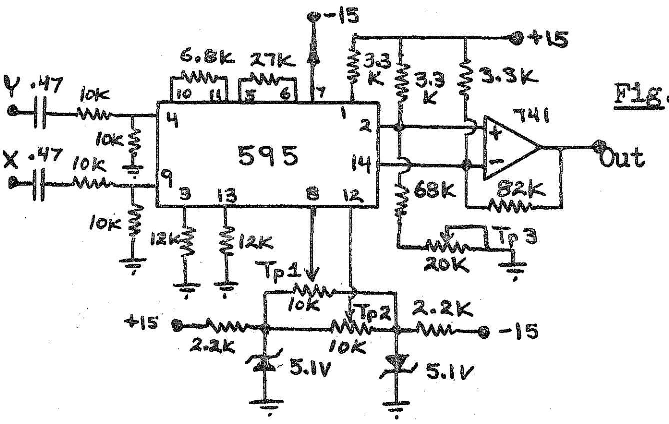


Fig. 17

Allignment
Connect signal to Y and adjust Tp-1 for no signal out. Remove from Y.
Connect signal to X and adjust Tp-2 for no signal out. Remove from X.
Adjust Tp-3 for a zero volt DC level.

Note the use of AC coupling on the inputs in this case. This is a good idea with all balanced modulator inputs which handle only AC, as any DC componet will cause the carrier to feed through.

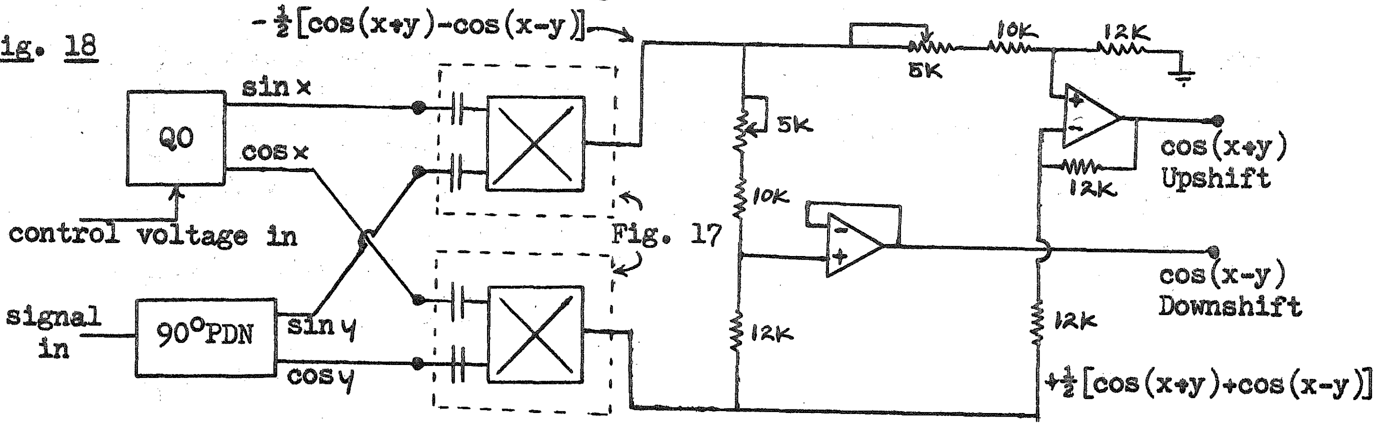
The frequency shifter design is based on one part of the more sophisticated design of Bode and Moog [6] using the trig identities (balanced modulator equations):

$$\cos x \cdot \cos y = \tfrac{1}{2}[\cos(x+y) + \cos(x-y)]$$

$$\sin x \cdot \sin y = -\tfrac{1}{2}[\cos(x+y) - \cos(x-y)]$$

These are applied such that sin x and cos x are the outputs of the QO, and sin y and cos y are the outputs of <u>any</u> Fourier component of the signal applied to the 90°PDN. The two sines are applied to one multiplier, and the two cosines to the second multiplier. The sum and difference of these multiplier outputs then give downshifted and upshifted signals respectively. Reversing the multiplier connections (i.e., $\sin x \cdot \cos y$ and $\cos x \cdot \sin y$) reverses the upshift and downshift outputs, but in either case, both an upshift and a downshift are available simultaneously. The complete frequency shifter is shown in Fig. 18:

Fig. 18



USING THE FREQUENCY SHIFTER: It is important that we see claarly what the frequency shifter actually does. First of all, consider the test setup where we put a 1000 Hz sine wave into the input and adjust the QO signal to 300 Hz. The signal at the upshift output will be a 1300 Hz sinewave, while the signal at the downshift output will be a 700 Hz sinewave. Increasing the QO signal frequency will cause the upshift to continue up, while the downshift will go down to zero and through into another upshift, trailing the original upshift by twice the input frequency (see Fig. 19). If a complex waveform is applied to the input, the waveform will not be preserved for two reasons. First, the 90°PDN will provide a different phase shift for each of the components of the complex wave (phase distortion) such that the waveform will be altered at the outputs. Secondly, each of the components output from the 90°PDN will be shifted by the same number of Hz (the QO frequency), so they will no longer be simple overtones of a single fundamental. The overall result is a non-harmonic tone color which is useful, but in many ways similar to those produced by various forms of modulation.

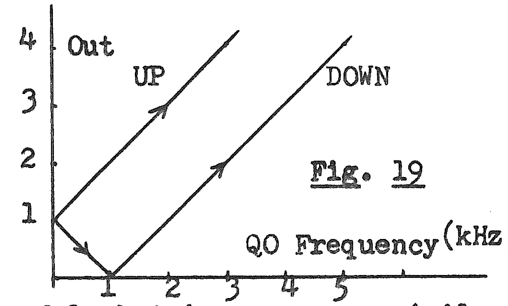


Fig. 19

The shifter is very effective with the human voice, live instruments, and other live sounds. Since these are complex waveforms, tone color can be drastically altered by frequency shift processing. This can be used in at least two ways. First, the shifting frequency can be set on the same order or lower than the program material input. In this case, the basic motion of the program material is preserved, but with new and often surprising sonorities. In the second case, we can set the shifting frequency much higher than the frequencies present in the program. Both outputs are upshifts in this case. This produces a relatively narrow band of frequencies in the region of the the shifting frequency. This may be interesting in itself, but the entire spectrun can be transposed down by tape manipulation and/or reshifted down for additional useful processing.

Finally, we can treat the frequency shifter as an electronic music module with two inputs and 3 outputs as shown in Fig. 20. One can fit this into the synthesis system in several ways. For example, the downshift output can be connected to the control voltage input as in Fig. 21. A little thought (considering the usual PLL scheme) will show that this is a phase locked loop. For a sine wave input, the VCO out tracks the input while the upshift output provides a frequency doubled track. The PLL will lock on to complex waveforms as well, providing a variety of complex outputs. The capture and lock ranges of the PLL are adjusted by the input attenuator on the control voltage input.

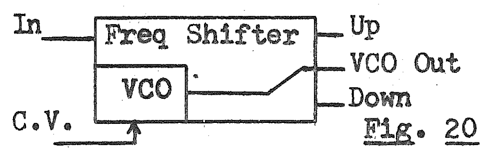


Fig. 20

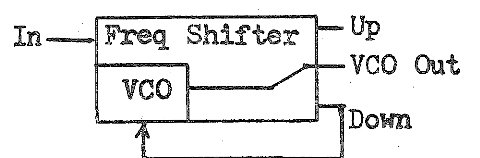


Fig. 21

APPENDIX A: Alternative 90°PDN Design Parameters: Below are listed some alternative design information for 90°PDN's with various bandwidth ratios, number of poles, and maximum phase error. Low values of phase error were chosen so as to permit less requirement of precision passive components, as in the design example in the text. Design is accomplished in the same way it was done in the text example. The question comes up as to how much phase error is allowable. A possible guide is the design graph of Bedrosian[5] which refers to single sideband generation (which is actually what we are doing). This shows the rejection of the unwanted sideband to be approximately: Rejection(db) $= 20 \log_{10}(100/\epsilon)$

TABLE 2: Alternative Phase Difference Networks

Ratio F_L/F_H	Number of Poles	Phase Error	Normalized Poles of Network (Multiply by F_L for actual pole frequencies)				
1:10	4	1°	Network A:	0.4787	5.2967		
			Network B:	1.8879	20.8875		
1:100	6	2°	Network A:	0.5206	6.0332	47.0664	
			Network B:	2.1245	16.5747	192.0597	
1:100	8	0.3°	Network A:	0.3835	3.1753	14.6036	72.6981
			Network B:	1.3755	6.8476	31.4923	260.7026

1:250	8	0.7°	Network A:	0.4468	4.2706	24.4061	147.8213		
			Network B:	1.6912	10.2433	58.5395	599.4564		
1:1000	8	2°	Network A:	0.5541	6.6265	53.2463	438.4678		
			Network B:	2.2806	18.7806	150.9075	1804.4704		
1:1000	10	0.7°	Network A:	0.4367	3.9201	20.8429	110.3799	620.6588	
			Network B:	1.6111	9.0596	47.9778	255.0930	2289.6428	
1:3000	10	1.1°	Network A:	0.5249	5.2142	34.0870	226.1747	1498.6177	
			Network B:	2.0018	13.2640	88.0098	575.3431	5714.7537	
1:3000	12	0.4°	Network A:	0.4333	3.5298	16.7786	81.3251	390.2776	1950.4137
			Network B:	1.5381	7.6868	36.8889	178.7991	849.8957	6922.3937
1:10,000	12	1°	Network A:	0.5684	4.7228	26.0677	156.9082	916.5275	4962.4934
			Network B:	2.0151	10.9107	63.7315	383.6157	2117.3661	17593.0876

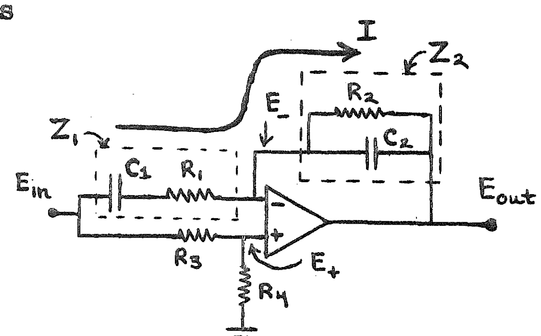
<u>APPENDIX B: The Use of Higher Order Networks:</u>

The term "Higher Order Network" implies. among other things, that the active element (transistor, op-amp, etc.) is going to be made to do several jobs at one time. Originally, the application of higher order networks was useful for reducing the cost of the active elements, and the price that was paid was that the passive components had to be more precise, and performance was more sensitive to tolerance variations and drift of these passive components. With today's drastic reduction in the cost of semiconductors compared to their passive counterparts, we can begin to think the other way. This is particularly true in the case where an individual is not particularly concerned with reducing the parts count in the one unit he is going to build, and is more interested in being able to build it quickly and easily, and adjust it for proper performance without special equipment. For these reasons, we have taken a look at the simpler "brute force" method. However, we would be remiss if we did not at least look at higher order networks, to see why they do not work well in this case for the individual builder, if for no other reason.

It should be pointed out that much of what was done in the main text still applies here - the basic idea is the same and pole positions will still be the same; we are only making amplifiers handle two or more poles each.

The best known of the higher order methods was presented by Albersheim and Shirley[4] and in an abridged form by Shirley[7]. An interesting reduction of the pole position calculations to an algebraic form is an important feature of this work. Four, six, and eight pole networks are discussed. The main drawbacks of these circuits is the extremely low component tolerances required, and the interaction of components, making it difficult to determine what needs tuning and how it might be done. This makes the circuits difficult for the individual user who lacks shelves of precision components and adequate test equipment.

Another interesting second order network is given by Lloyd[8]. This two pole network as shown at the right. Note that this is essentially a differential amplifier, and we can apply the s-plane analysis as follows:



$$E_+ = (R_4/(R_3+R_4)) \cdot E_{in} = K \cdot E_{in} = E_-$$

$$I = [E_{in} - K \cdot E_{in}]/Z_1 = \frac{(E_{in} - K \cdot E_{in})sC_1}{1 + sC_1R_1}$$

$$= \frac{E_{in}\ (1 - K)\ sC_1}{1 + sC_1R_1}$$

$$E_{out} = KE_{in} - I \cdot Z_2 = KE_{in} - \frac{E_{in}(1-K)sC_1}{(1 + sC_1R_1)} \cdot \frac{R_2}{1 + sC_2R_2}$$

$$T(s) = (E_{out}/E_{in}) = \frac{K(1+sC_1R_1)(1+sC_2R_2) - (1-K)sC_1R_2}{(1+sC_1R_1)(1+sC_2R_2)}$$

$$\frac{K(1-sC_1R_1)(1-sC_2R_2) + 2sKC_1R_1 + 2sKC_2R_2 - (1-K)sC_1R_2}{(1+sC_1R_1)(1+sC_2R_2)}$$

$$\frac{K(1-sC_1R_1)(1-sC_2R_2)}{(1+sC_1R_1)(1+sC_2R_2)}$$ Second order all-pass provided that:

$$2sKC_1R_1 + 2sKC_2R_2 - (1-K)sC_1R_2 = 0$$

$$\text{or } K = \frac{1}{2(R_1/R_2) + 2(C_2/C_1) + 1}$$

This is a nice form because the poles are stated in terms of RC values. Two such networks can form the A and B networks of a 90°PDN. With poles selected for a 90°PDN, the K factor will come out the same for both A and B, and hence, the K divider can be the same one. These could, with appropriate pole placement, be cascaded for an eight pole network, 12 pole network, etc. However, the problem of selecting precision components remains here too, and a simple allignment procedure is not obvious. The possibility of selecting correct R and C combinations by the first order method of Fig. 16, and substituting them into a second order circuit might be considered.

This has given a quick look at second order and higher order networks. In this case, where precision results are necessary, we have seen that the higher order networks have a theoretical advantage, but not necessarily a practical one when it actually gets down to building a device. In many cases, we may be using a bulldozer to crack a walnut, but for lack of a hammer ....

REFERENCES:

1. S. Darlington, "Realization of a Constant Phase Difference" <u>Bell Systems Tech. Jour.</u> <u>29</u> pg. 94 (1950)
2. N.O. Johannsson "Wideband Two-Phase Networks" <u>Wireless Engineer</u> Aug-Sept. 1950, pg. 237
3. D.K. Weaver "Design of RC Wide-Band 90-Degree Phase Difference Network" <u>Proc IRE</u> <u>42</u> pg. 671, April. 1954
4. W.J. Albersheim and F.R. Shirley, "Computation Methods for Broad-Band 90° Phase-Difference Networks" <u>IEEE Trans on Ckt. Theory</u> CT-16, pg. 189 (1969)
5. S. Bedrosian, "Normalized Design of 90° Phase-Difference Networks" <u>IRE Trans on Ckt. Theory</u> <u>CT-7</u>, pg. 128, June 1960
6. H. Bode "Solid State Audio Frequency Spectrum Shifter" AES Preprint No. 395 (1965). H. Bode and R.A. Moog "A High-Accuracy Frequency Shifter for Professional Audio Applications" AES Preprint No. 865 (D-6), (1972)
7. F.R. Shirley, "Shift Phase Independent of Frequency" <u>Electronic Design</u> 18 Sept. 1, 1970, pg. 62
8. A.G. Lloyd, "Here's a Better Way to Design a 90° Phase-Difference Network" <u>Electronic Design</u> #15, pg. 78 (1971)

CALCULATION OF THE POLES OF 90° PHASE DIFFERENCE NETWORKS [after Weaver]

1. Select F_ℓ and F_h, the lower and upper ends of the bandwidth.

2. Calculate:

$$k = [1 - (F_\ell/F_h)^2]^{1/2}$$

$$L = (1/2)\frac{1 - \sqrt{k}}{1 + \sqrt{k}}$$

$$Q' = L + 2L^5 + 15L^9 + \cdots$$

$$Q = e^{\pi^2/\log_e(Q')}$$

3. Select a phase error that can be allowed, and consult the graph below to determine the minimum corresponding number of poles "n".

4. Choose two networks, A and B. If n is even, there will be n/2 poles in each network. If the number of poles is odd, put an extra one in network A.

5. Compute the angles for A and the poles p_a:

$$\phi_a(r) = (45°/n)(4r - 3) \qquad \text{for } r = 1, 2, \ldots (n/2) \text{ or } [(n+1)/2]$$

$$\phi'_a(r) = \text{ARCTAN}\left((Q^2-Q^6) \text{ Sin } 4\phi_a(r) \;/\; [1 + (Q^2+Q^6) \text{ Cos } 4\phi_a(r)]\right)$$

$$p_a(r) = [1/\sqrt{(F_\ell/F_h)}] \text{ TAN}[\phi_a(r) - \phi'_a(r)]$$

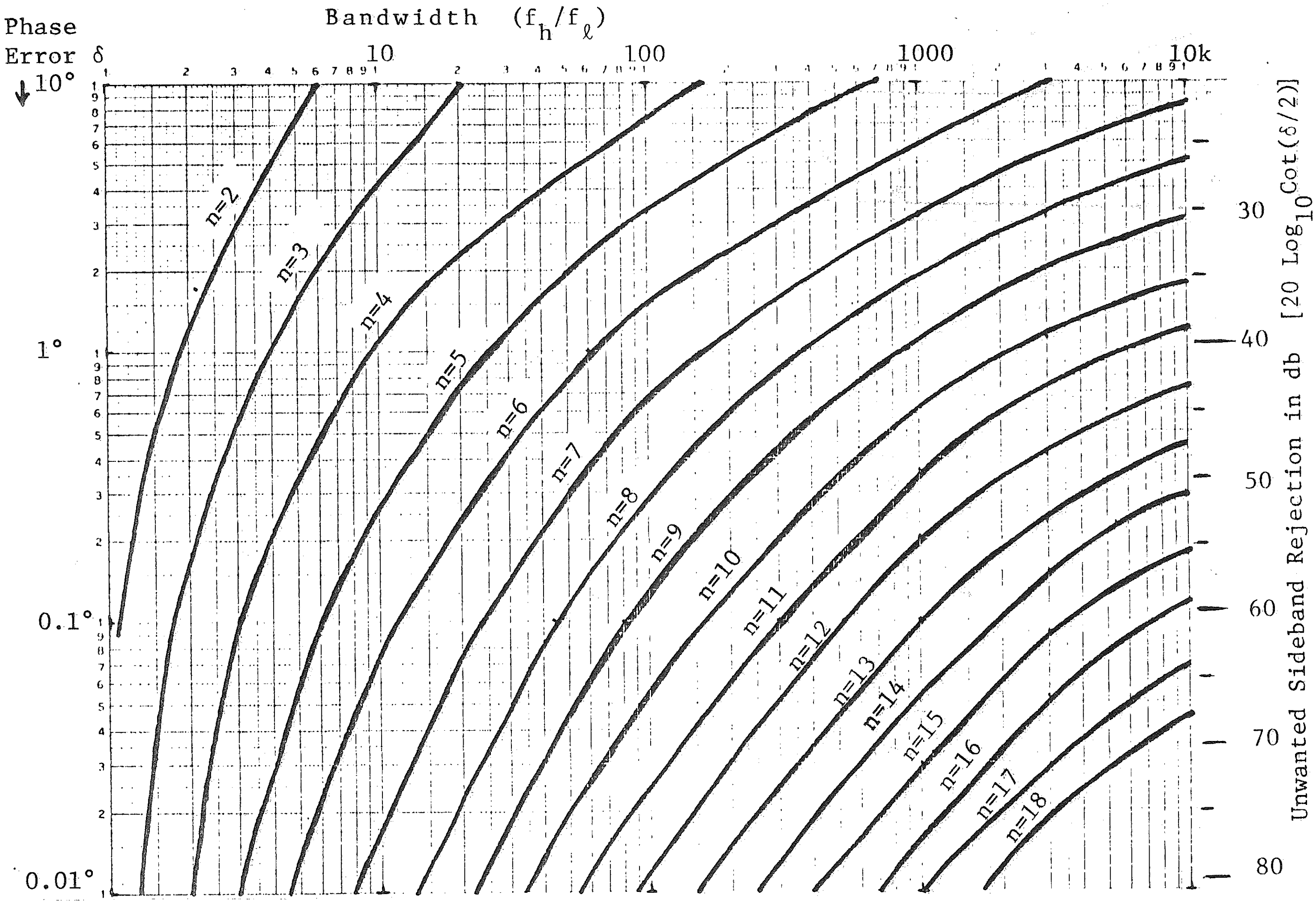
and the angles for B with its poles p_b:

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

$$\phi'_b(r) = \text{ARCTAN}\left((Q^2-Q^6) \text{ Sin } 4\phi_b(r) \;/\; [1 + (Q^2+Q^6) \text{ Cos } 4\phi_b(r)]\right)$$

$$p_b(r) = [1/\sqrt{(F_\ell/F_h)}] \text{ TAN}[\phi_b(r) - \phi'_b(r)]$$

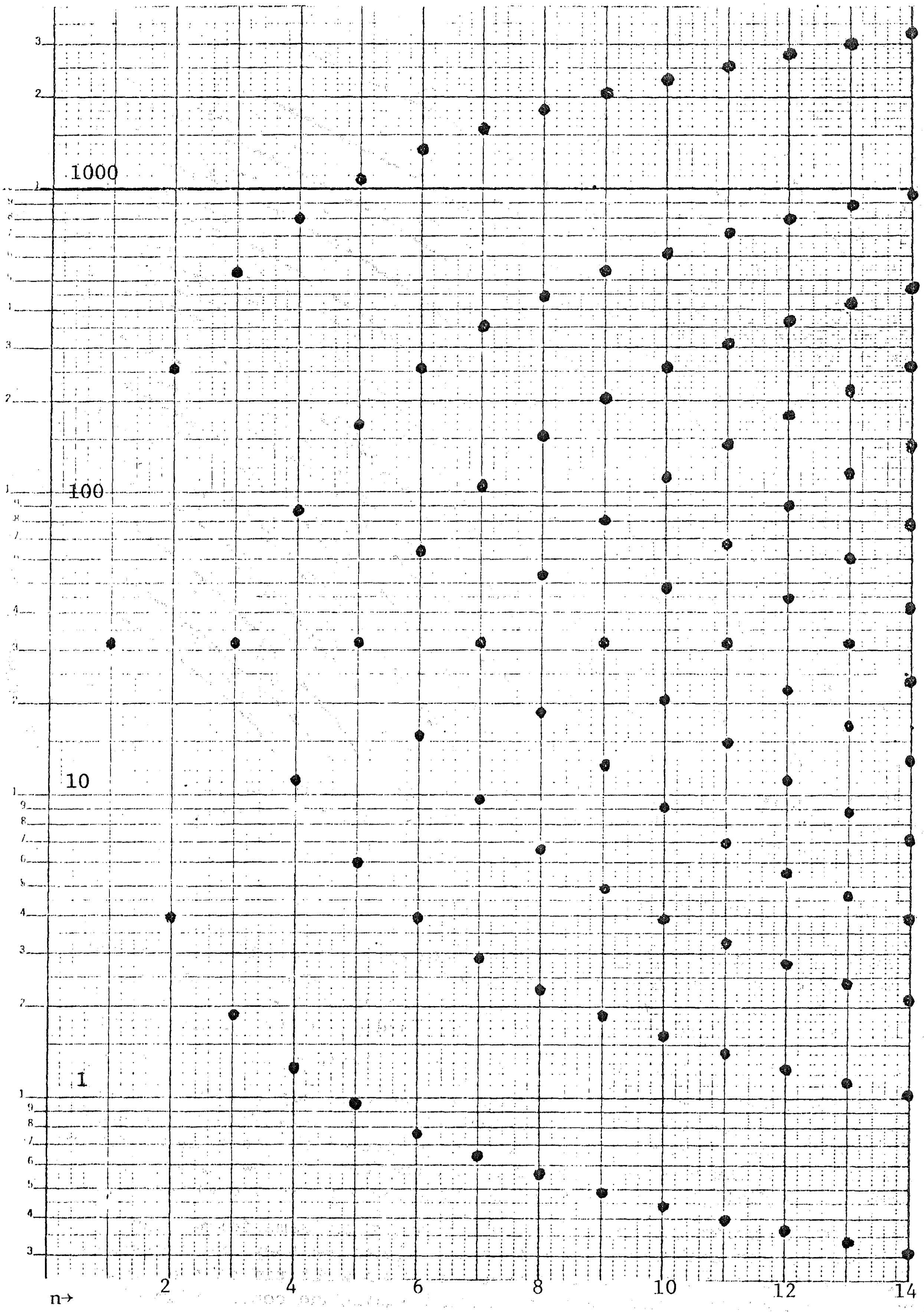
6. The above poles are normalized in terms of F_ℓ. To get the actual poles for the networks, multiply by F_ℓ.



Design Chart for Number of Poles (n) for a Given Normalized Bandwidth and Allowable Phase Error. The Unwanted Sideband Rejection is also Plotted along with the Phase Error. [After Bedrosian]

DETERMINING THE REQUIRED ACCURACY:

When it comes to determining the required accuracy of multipliers and 90°PDN's, the two major considerations are "carrier rejection" and "unwanted sideband rejection." Carrier rejection is largly a matter of the properties of the multipliers. This comes in when you apply two signals f_1 and f_2 to a multiplier and find that in addition to f_1+f_2 and f_1-f_2 you also get a little f_1 and a little f_2. Generally, the higher of the two frequencies will be the one that is noticed, and this is the reason for the term carrier rejection. The point is that if this gets through the multiplier, it will find its way to the output of the frequency shifter. For example, if it is the program signal that is the highest, this will come through the multiplier as a $\mathrm{Sin}(\omega_p t + \phi)$ in one case and a $\mathrm{Cos}(\omega_p t + \phi)$ in the other case where ϕ is the phase shift across the multiplier which is the same for either of the multipliers and for either of the quadrature signals since they are the same frequency. There is nothing that the summers do that will get rid of the program signal once it is through, since they are in quadrature, and no combination of addition or subtraction will cause them to cancel. Thus, carrier rejection is important in multiplier selection for frequency shifters.



Normalized Pole Positions for a bandwidth f_h/f_ℓ = 1000. Note the symmetry that shows up in this plot of Log(Pole positions) vs. Linear n. The poles were calculated according to Weaver's method.

The second factor is the accuracy of the phase difference network, and this determines the unwanted sideband rejection. This can be seen by considering the effect of an error angle δ. If the actual quadrature program signals are:

$$\mathrm{Sin}(\omega_p t) \qquad \text{and} \qquad \mathrm{Cos}(\omega_p t + \delta)$$

then the multiplier equations become:

$$\mathrm{Sin}(\omega_p t)\ \mathrm{Sin}(\omega_s t) = (1/2)[\mathrm{Cos}(\omega_p t-\omega_s t) - \mathrm{Cos}(\omega_p t+\omega_s t)]$$

$$\mathrm{Cos}(\omega_p t + \delta)\ \mathrm{Cos}(\omega_s t) = (1/2)[\mathrm{Cos}(\omega_p t+\omega_s t+\delta) + \mathrm{Cos}(\omega_p t-\omega_s t + \delta)]$$

We saw that for the normal downshift, we added the two outputs. Adding the two above equations and using trig identities for the sum of two cosines (and the corresponding equation for the difference of two cosines), we get the following form for the downshift output:

$$\mathrm{Cos}(\omega_p t - \omega_s t + \delta/2)\mathrm{Cos}(\delta/2) - \mathrm{Sin}(\omega_p t + \omega_s t + \delta/2)\mathrm{Sin}(\delta/2)$$

The phases $\delta/2$ are not important for the time varying signals. Note that the first term is thus the desired downshift, and its amplitude is $\mathrm{Cos}(\delta/2)$. The second term is the unwanted upshift, and its amplitude is $\mathrm{Sin}(\delta/2)$. We see that for small values of δ, the unwanted sideband is small. The unwanted sideband rejection is therefore (in db):

$$20\ \mathrm{Log}_{10}[\mathrm{Cos}(\delta/2)/\mathrm{Sin}(\delta/2)] = 20\ \mathrm{Log}_{10}\mathrm{Cot}(\delta/2)$$

Some typical values are:	20db rejection	$\delta = 11°$
	40db rejection	$\delta = 1°\ 10'$
	60db rejection	$\delta = 0°\ 8'$

The unwanted sideband rejection is plotted along with the phase error on the design chart for selecting the number of poles for the network. For most frequency shifters, 40 db rejection or better should be chosen.

* * * * * * * * * * * * * * * * *

Final Comments: Before ending this discussion, it should be mentioned that a new type of 90° network may prove very useful in the future. This is the Hilbert transform formed from a tapped bucket brigade delay line in a transversal filter configuration. Such devices can be used in place of the standard 90° PDN's. The advantages of the transversal filter is that it should lend itself nicely to integration. This could well reduce a fairly massive curcuit full if capacitors to a single chip.

CHAPTER 6B

PITCH AND ENVELOPE FOLLOWERS

CONTENTS:

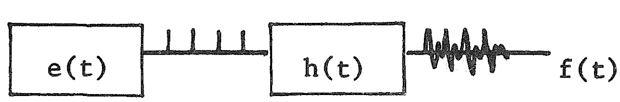
INTRODUCTION

Two common parameters of a musical sound that we would like to extract and use are pitch and amplitude. The devices for doing this are called pitch extractors or pitch-to-voltage converters, and envelope followers respectively. Pitch extraction is an extremely difficult problem, and at present cannot be solved in the most general case. Envelope following is basically a matter of employing some sort of rectifier and then low-pass filtering. However, there is a compromise on the time constant of the low-pass filter which tends to make this simple sort of follower useless in the limits of low frequency and rapidly varying envelope. The pitch extraction problem and the envelope following problems are in some ways inseparable. It is usually necessary to have the amplitude to recover pitch, and if the pitch is known, the time constants of the envelope follower can be optimized.

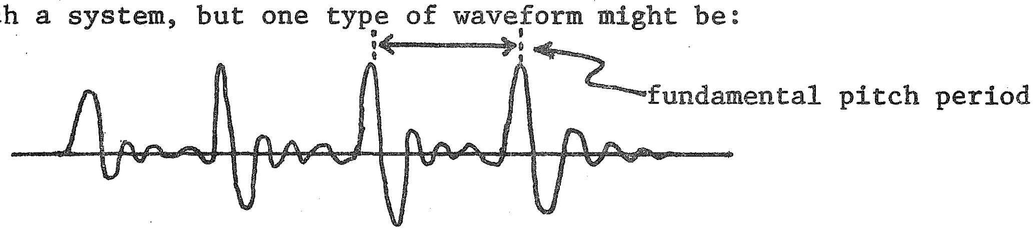
We are interested in recovering the pitch and amplitude of real (live) sounds. These are two items of "information" in the sounds of real instruments or the human voice. Pitch recovery is not the same as frequency recovery. Pitch is more general and deals with waveforms that have a complexity of information that far exceeds the amount of information in a waveform to which a clearly defined frequency can be assigned. A pitch extractor should give the frequency of an electronic signal however.

PITCHED SIGNAL PRODUCTION

We first have to look at a block diagram of a system for producing a musical sound in an acoustical manner (mechanical as opposed to electrical):



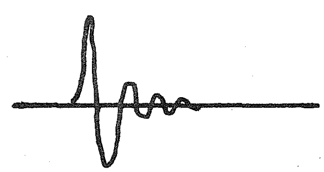
Here, e(t) is a periodic excitation while h(t) is a system function that can be associated with the resonances of the system. In a violin for example, e(t) would be the relatively pure vibration of the strings while h(t) would represent certain resonances (formants) of the violin box. The output f(t) is a complex signal that contains both the pitch information [from e(t)] and the formant information [from h(t)]. It is difficult to suppose that there is any sort of typical waveform that results from such a system, but one type of waveform might be:



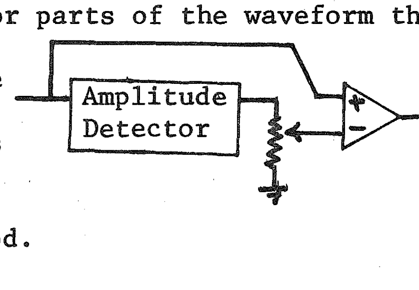
The main peaks may be considered to result from the excitation while the fine structure can be considered the ringing of the formant filters. A typical frequency recovery method is to count zero crossings. The above example shows that a zero crossing count will be much too high and will not recover the fundamental pitch period.

PITCH EXTRACTION BASED ON SPECTRAL INFORMATION OR PATTERN RECOGNITION

Since the fundamental pitch is a spectral component (usually the lowest, except in a few artificial test cases), we can consider trying to separate it out by filtering. The problem is that if the filter is sharp enough to separate the fundamental from the harmonics, it is too sharp to pick it up unless it happens to be very close to begin with. If the filter scans up, there is a good chance the fundamental will be detected, but there may be detectable delays before the fundamental is acquired. Phase-locked-loops will be plagued by harmonic locking, slewing, and loss and recapture problems, and are generally unsuccessful. It is thus more useful to abandon methods that are based on the perception of pitch by spectral information, and consider an alternative theory that is often required to explain pitch perception experiments - the idea that the ear-brain does a pattern analysis on the waveform.

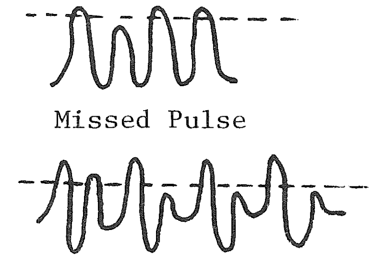


In the example waveform f(t) that was drawn above, we saw that the fundamental pitch occured as a repetition of segments of the general form shown at the right. This sort of thing is an easy thing for the eye to pick out, and is called autocorrelation. The eye picks out the fundamental building block. Even though they are not all exactly the same, they are close enough. They look enough like each other to be recognized. The successive segments are thus correlated with themselves - hence the term autocorrelation for the perception process. The ear may do something very similar on the time waveform f(t). The use of actual autocorrelating devices such as are possible with various types of delay lines is a promising method of pitch extraction by pattern analysis. Here we will use a more restricted method based on the fact that the excitation source produces amplitude prominences in the waveform f(t).



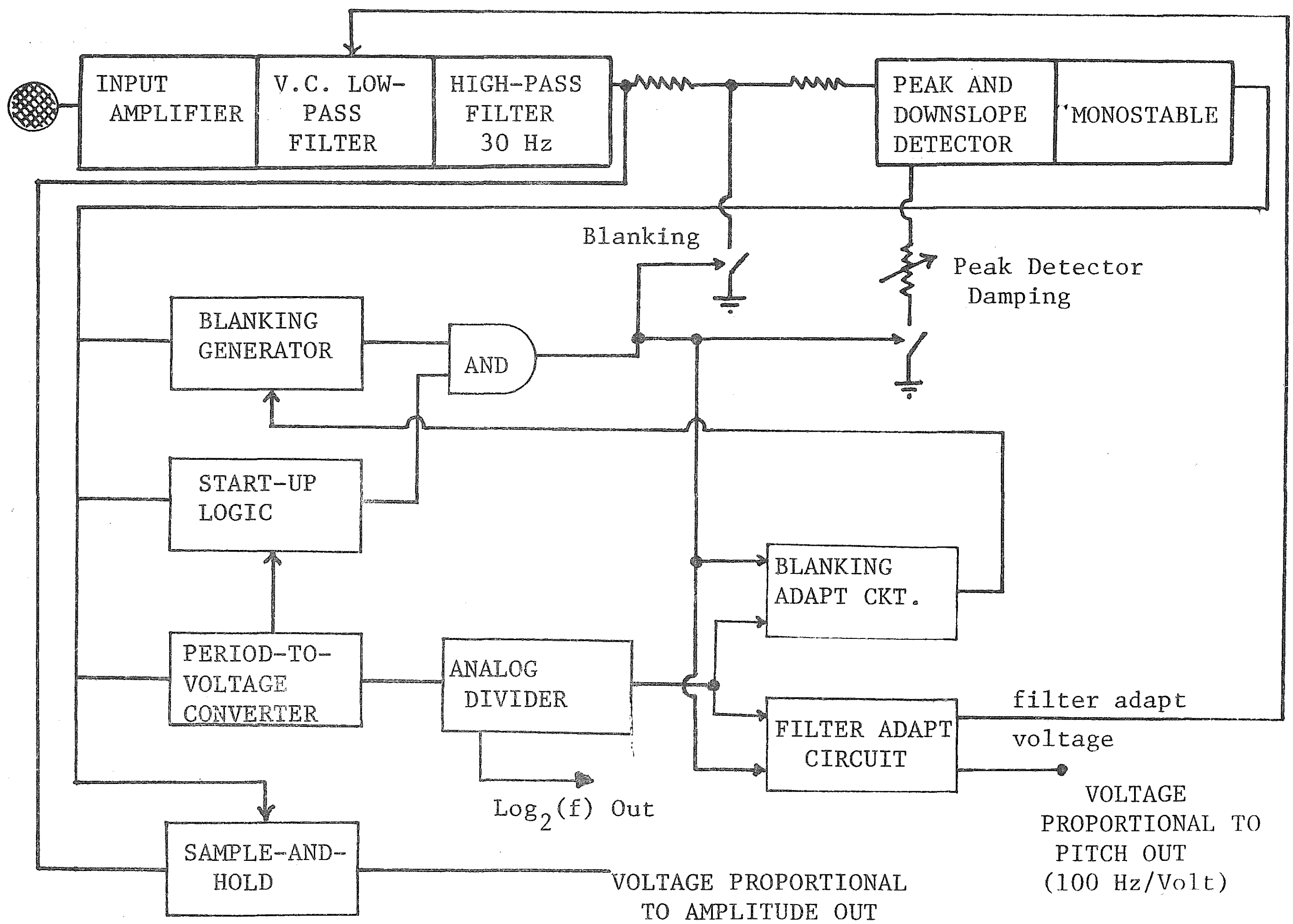
The amplitude peaks will in general determine the overall amplitude. Thus, if we know the amplitude, we can pick off the peaks by looking for parts of the waveform that approach the maximum amplitude. This method is known as center clipping and can be implemented with an amplitude detector, attenuator, and a comparator is indicated at the right. This method is limited in several ways. (1) It is no better than the amplitude detector is. (2) It is impossible to find an optimum clipping level since there are often level changes that occur for only one pitch period.

(3) Transients provide special problems. Setting the clipping level is a matter of the actual attenuation of the detected amplitude, and the damping. If the two are too high, an actual pitch pulse may be missed. If the two are too low, secondary pulses may be picked up. Experiments show that the two cases will overlap, so there is no way to set a clipping level based on amplitude detection and damping. The two faults, missed pulse and extra pulse are illustrated at the right.

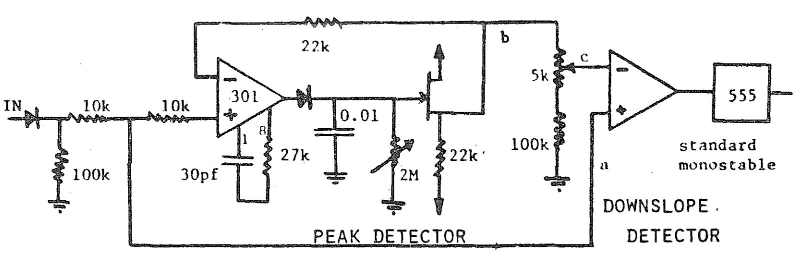


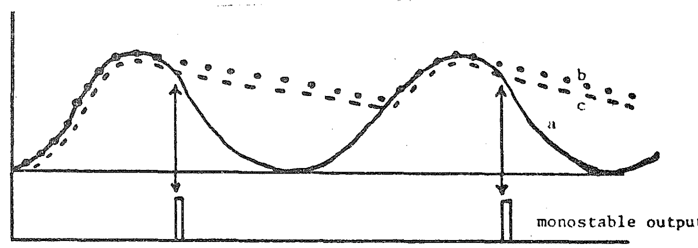
AN EXPERIMENTAL PITCH EXTRACTOR AND ENVELOPE FOLLOWER

Below we will describe an experimental device that is intended to extract the pitch and recover the envelope as well. The device as shown is not a satisfactory electronic music module. It does work quite well, but even an accuracy that is 99% or greater is not enough since the errors may make their presence known as sharp and annoying transitions in the music. There are some who believe that live sounds are too complex for a simple device to follow, and there may be some basis in information theory for believing this to be true. Thus, to pick up live sounds, the musical engineer may choose one of two restricted approaches. First, he can try to optimize the parameters of the detector for the specific device to be monitored and for a specific range. This can be quite successful in many cases. Secondly, he can try to obtain a signal that is closer to the excitation source e(t) where the system function h(t) has not yet had a chance to make the sound more interesting (e.g., string pickups). The device below may be a useful starting point for either approach.



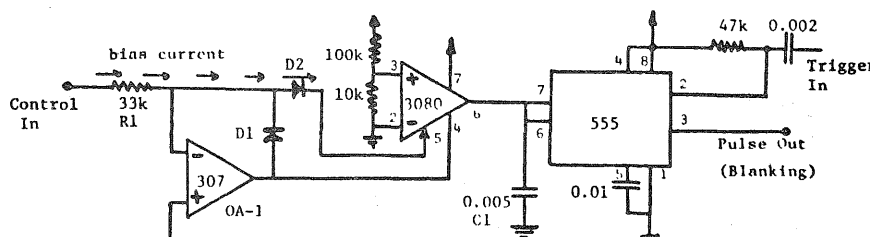
The device shown works on peaks in the waveform f(t). It is similar to center clipping, but has the following additional features: (1) The clipping is replaced by a peak detector and a downslope detector, and thus does not rely on attenuation and damping to detect peaks. (2) The device has a blanking feature that is triggered after a peak is detected which effectively makes the device insensitive to any input features for a period of time proportional to the last pitch period detected. (3) The device has an input filter which is voltage tunable so that it can automatically adapt slightly to the input once the initial pitch readings are found. Thus, the overall effect is much like a flywheel - the system gets going and looks for only peaks that meet the general requirements of the pattern. Of coures, there is a danger in making such a simple machine too smart. It may look so hard for something that it will make its own. To prevent this, the device has a system of start-up logic. None the less, the device can be fooled (i.e., the flywheel can get going at half the correct speed or twice the speed - equivalent to harmonic locking), or momentary dropouts that the ear would ignore or smooth over may leave the system hanging.



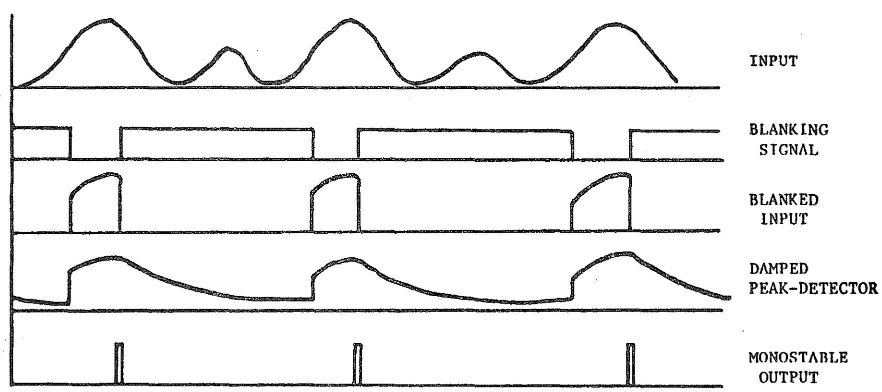


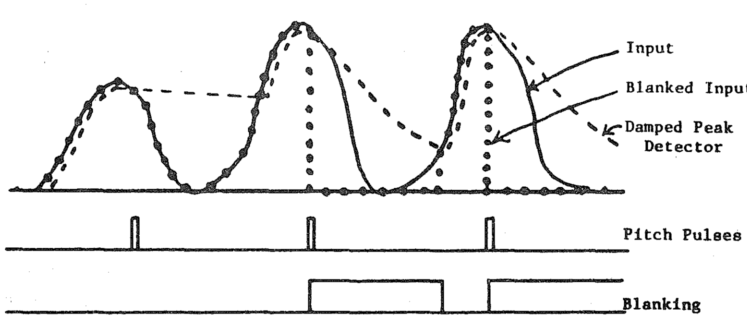
WAVEFORMS OF PEAK DETECTOR AND DOWNSLOPE DETECTOR

It is not possible here to give all the circuit details of the experimental device. We give here a few circuits and typical waveforms. First is shown the peak detector and downslope detector (above) and its waveforms. Note that a pulse is produced just after the input waveform starts down. The voltage-controlled monostable is the heart of the blanking circuit. It produces a pulse width that is inversely proportional to the control voltage. Thus, when controlled by a voltage proportional to frequency (output of the analog divider), it produces a pulse width proportional to the pitch period. This proportion can be varied from about 10% to 100%. Operation of the blanking circuit is indicated by the waveform drawings which show how a secondary peak is effectively removed. A logic circuit controls the start up phase and prevents blanking until an initial reading on the pitch (two pulses) is obtained. This generally works because the first few amplitude peaks rise in succession. For additional circuits, see EN#53-EN#55.



VOLTAGE-CONTROLLED MONOSTABLE





START-UP PHASE OF PITCH EXTRACTOR

CHAPTER 6C

MODULES EMPLOYING ANALOG DELAY LINES

CONTENTS:

INTRODUCTION

There are a large number of devices for electronic music that can be best implemented with analog delay lines. At the time of this writing, analog delay lines are just becoming practical for delays that are useful at audio frequencies (from 1 ms on up).

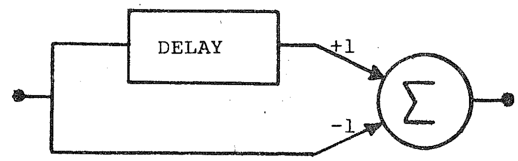
Two such devices are phasing units and artificial reverberation units. In the past, the need for such devices has been met by phasing simulators (phase shifting filters), and by mechanical delays (e.g., coiled springs) for reverberation. Neither of these is completely satisfactory since the most logical implementation is through the use of a time delay, and the engineer has been forced to work with something similar. In addition to the applications above, there are numerous other potential uses.

We will not be discussing here the actual circuits for analog delays. These are available as devices referred to as "bucket brigade delay lines" or "charge coupled devices," but we can also use "digital delay lines" where the audio signals are first digitized, fed down a digital shift register, and reconverted to audio. The analog delay line is conceptually very simple - you put a signal in one end, and it comes out the other end later. A tape delay line is also possible, but the useful devices for the future will be the all electronic devices that are not subject to mechanical wear.

Two circuits for the ITT type TCA350 have been given: J. Hall has shown a circuit in EN#48. F. Hinkle has shown another circuit in Electronics, Aug. 7, 1975. The TCA350 is a 185 stage bucket brigade MOS shift register (analog delay line).

APPLICATIONS USING MIXTURES OF DELAYED AND ORIGINAL SIGNALS

A number of interesting applications arise when the delayed signal is mixed with the original version, or with a version delayed by a different amount of time. A simple and useful example is the "comb filter" formed from a first-order non-recursive (no feedback) network as shown at the right. The summers used in this circuit and in later circuits are just op-amp summers. The coefficients are set by adjusting the summing resistors. A simple method of analysis of the comb filter starts with the trig identity:



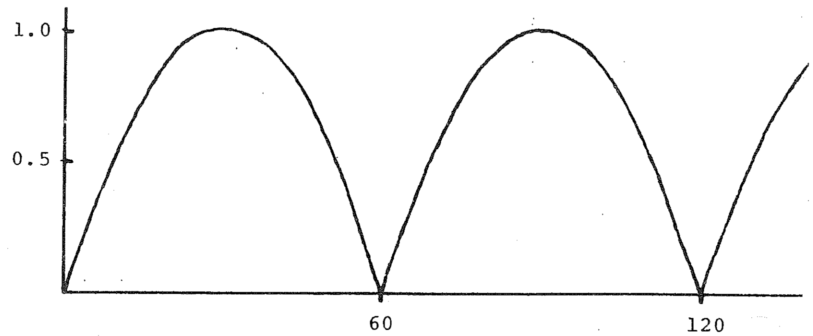
Simple Comb Filter

$$\sin A - \sin B = 2 \sin[(A-B)/2] \cos[(A+B)/2]$$

Denoting the delay time as τ, the output of the network is:

$$\sin(\omega t+\omega\tau) - \sin(\omega t) = 2 \sin(\omega\tau/2) \cos[\omega t + (\omega\tau)/2]$$

The amplitude of the output is determined by $\sin(\omega\tau/2)$. Thus, the comb filter has nulls when $\omega = 2n\pi/\tau$, and the network is useful for removing a fundamental and all overtones from a signal. If a delay time of 1/60 sec. is used for example, nulls will fall at 60 Hz, 120 Hz, 180 Hz, etc as indicated at the right.



If both the inputs to the summer are made positive, a comb filter results that is useful for removing all odd harmonics. This can be demonstrated in a manner analogous to the above calculation, but a second method, impulse analysis, will be used for illustration. The impluse response of a delay line is very simple - an impulse enters at time t and comes out at time $t+\tau$. Thus, for an impulse $\delta(t)$, the response is $\delta(t-\tau)$. For the comb filter with both inputs positive, the total impulse response is $\delta(t) + \delta(t-\tau)$. The transfer function for the network is the Fourier transform of the impulse response (Chapter 1d). The FT of an impulse is just 1, and that of the delayed impulse $\delta(t-\tau)$ is $e^{j\omega\tau}$. [The FT of a delayed signal is just e^{st} times the undelayed FT. For this reason, e^{st} is referred to as the unit delay operator and is often denoted as "z" in the literature on digital signal processing]. The system transfer function is thus:

$$T(j\omega) = 1 + e^{j\omega\tau}$$

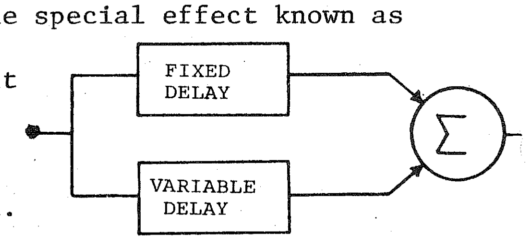
The magnitude of the transfer function is the frequency response:

$$|T(j\omega)| = [T(j\omega)\cdot T(-j\omega)]^{1/2} = [(1+e^{j\omega\tau})\cdot(1+e^{-j\omega\tau})]^{1/2} = (2+2\cos\omega\tau)^{1/2} = 2\cos(\omega\tau/2)$$

This comb filter has nulls when $\omega = (2n-1)\pi/\tau$, and is thus used to remove odd harmonics. The delay in this case would be set to half the value of the delay used to remove all harmonics.

PHASING

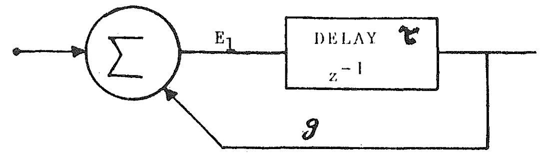
One application of the comb filter technique is to achieve the special effect known as "phasing" or "flanging" in the music trade. For this effect, it is useful to add a second fixed delay as shown at the right so that the variable delay can be swept through zero relative to the fixed delay. Sweeping this delay slowly (on the order of 1/4 Hz) produces moving nulls in the comb filter response that produce the effect. Phasing is discussed in Chapter 2e.



Comb filter structure for "Phasing" effect.

RECURSIVE STRUCTURES

A recursive discrete time filter is one in which output samples are fed back to the input summer. A first order recursive network is shown at the right. This type of filter has an infinitely extended impulse response as a result of the feedback. The impulse response is denoted by h(t).



First-order recursive network.

$$h(t) = \delta(t - \tau) + g\delta(t - 2\tau) + g^2\delta(t - 3\tau) + \dots\dots$$

Taking the FT, the transfer function is:

$$T(j\omega) = e^{-j\omega\tau} + ge^{-2j\omega\tau} + g^2e^{-3j\omega\tau} + \dots\dots$$

$$= \frac{e^{-j\omega\tau}}{1 - ge^{-j\omega\tau}}$$

where the sum of the series was obtained from the formulas for a geometric series. An alternative method of analysis for this network is to use the z-transform notation where $z^{-1} = e^{-s\tau}$.

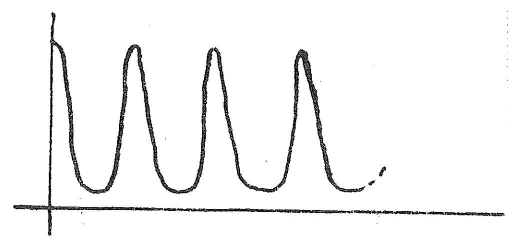
$$E_{out} = E_1\ z^{-1} = E_{in}\ z^{-1} + g\ E_{out}\ z^{-1}$$

$$T(z) = z^{-1}/(1 - gz^{-1})$$

$$T(j\omega) = e^{-j\omega\tau}/(1 - ge^{-j\omega\tau})$$

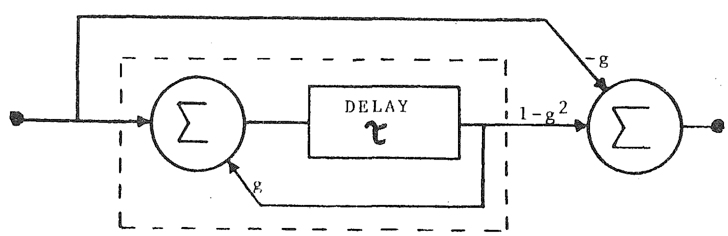
$$|T(j\omega)| = 1/(1 + g^2 - 2g \operatorname{Cos} \omega\tau)^{1/2}$$

This response is a comb filter with sharp resonant peaks as indicated at the right. When g is set to one, the network oscillates. However, the main interest in this network is as part of an overall network for artificial reverberation. The possibility for artificial reverberation is implied by the infinite duration (but decaying) impulse response.



REVERBERATION DEVICES

An all-pass network developed by Schroeder (M. R. Schroeder, "Natural Sounding Artificial Reverberation," JAES 10 #3, July 1962) is shown at the right. The impulse response of this network will be denoted h'(t).



All-Pass network useful for artificial reverberation.

$$h'(t) = -g\delta(t) + (1 - g^2)\ h(t)$$

where h(t) is the impulse response of the first order recursive network above which is contained within the dotted lines of the all-pass. This gives a transfer function:

$$T(j\omega) = e^{-j\omega\tau}\ [1 - ge^{j\omega\tau}]/[1 - ge^{-j\omega\tau}]$$

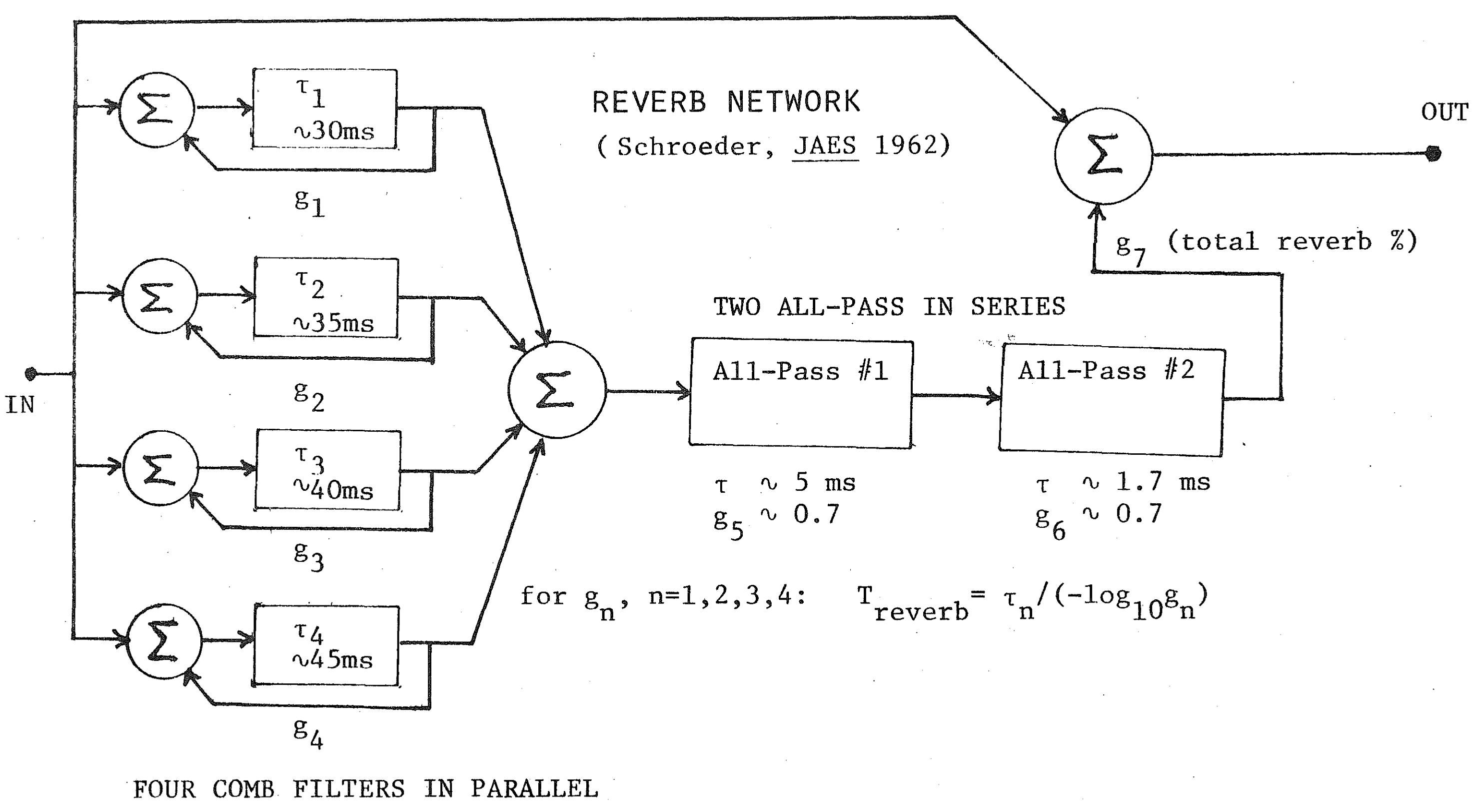
$$|T(j\omega)| = 1$$

Schroeder examined various structures of the all-pass filters to achieve sufficient echo density for artificial reverberation that had a realistic sound. Studies of reverberation indicate that an echo density building up to about 1000/second is required to avoid a feeling of "flutter." Also, a reverberation time on the order of 2 seconds (time to decay by 60 db) is found in natural reverberation. For a loop gain g in the resonator of the all-pass, the response decays by $20 \log_{10} g$ for each trip around the loop (time τ). This means that the reverberation time of the all-pass is:

$$T_{reverb} = \frac{60 \text{ db}}{(20 \log_{10} g) \text{ db}} \tau$$

For sufficient echo density, τ would have to be 1 ms, and for 2 seconds of reverb, g would have to be very close to one. This is not possible in a practical resonator. For an assured stable condition, g should be only about 0.7. Schroeder thus used series all-pass units with delay times of successive units on the order of (but not exactly equal to) 1/3 the delay of the preceding unit. This greatly increases the echo density, and five such series units gives the necessary density and time with practical all-pass units.

Since all real rooms do not have a perfectly flat frequency response, one can consider the use of some "coloration" of the reverberation. There is a connection between reverberation time and the number of natural modes of the room. Schroeder used a number of delay line comb filters in parallel to give the same number of natural peaks, and followed this with series units for echo density. An example overall reverb network is indicated below:



CHAPTER 6D

TRANSFORM DEVICES

CONTENTS:

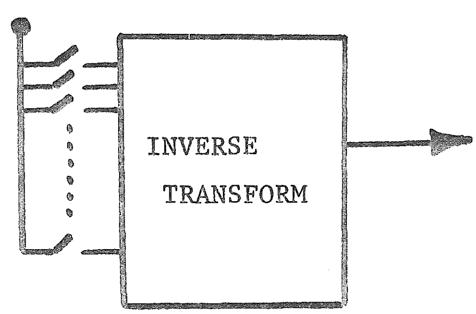
INTRODUCTION

Music moves in two dimensions that are the variables of a Fourier Transform pair: time and frequency. Both time and frequency are also dimensions in which signal processing can occur. It is helpful at times to consider processing done in the different domains, and at times the mathematics of the processing is greatly simplified in one domain - hence the utility of the mathematical Fourier Transform apparatus.

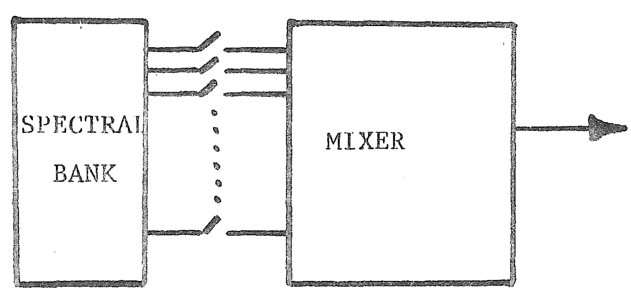
It is useful to consider the use of devices that actually do these transforms. In this way, a number of interesting devices can be considered. We will first consider a few possible devices and their more conventional formulations.

COMPARISON OF TRANSFORM APPROACHES AND THE TRADITIONAL APPROACHES

Additive synthesis can be approached through the use of an inverse transform (frequency domain to time domain). This might for example be an inverse Fourier transform which is fed spectral information and gives out a waveform. This basic idea is shown below along with the more conventional approach which employs a source bank of harmonic components.



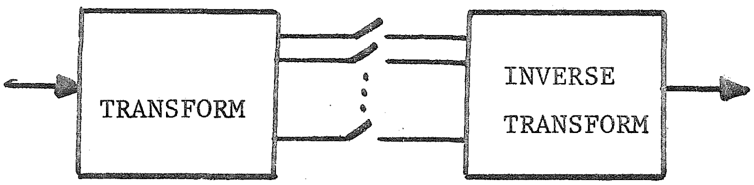
TRANSFORM DEVICE



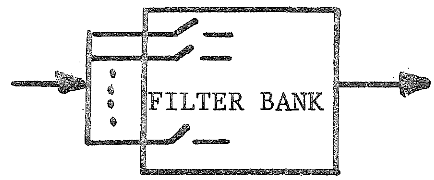
CONVENTIONAL ADDITIVE SYNTHESIS

In the transform approach, inputs would be fed to the inverse transform through some form of control (shown here as switches). These could select spectral points corresponding to the pitches of the scales we want to use, or they might select the overtones of a waveform.

A scheme that does both an analysis (from time to frequency) and a resynthesis (from frequency components to time) is shown below along with its more common counterpart which is formed from a bank of filters or perhaps just one VCF:



TRANSFORM DEVICE



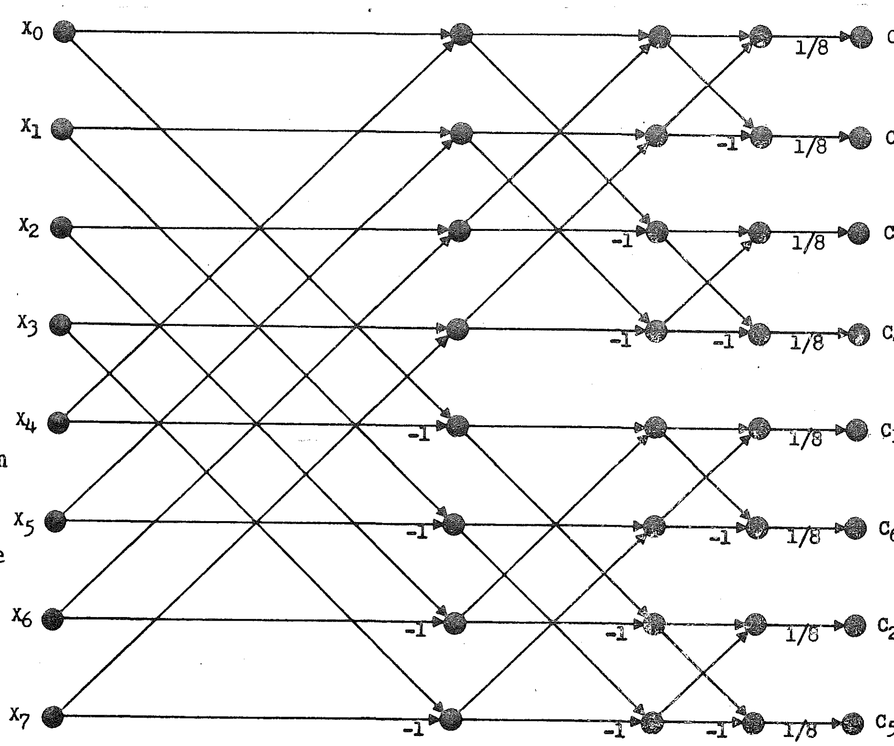
CONVENTIONAL SUBTRACTIVE SYNTHESIS

In the case of the transform, certain of the controls (shown as switches) can be set to control the harmonic contents of the waveform. Thus, this analysis and resynthesis technique is analogous to subtractive synthesis. This may be best seen by considering that if switches are left open, certain spectral components will be blocked, and this corresponds to filtering.

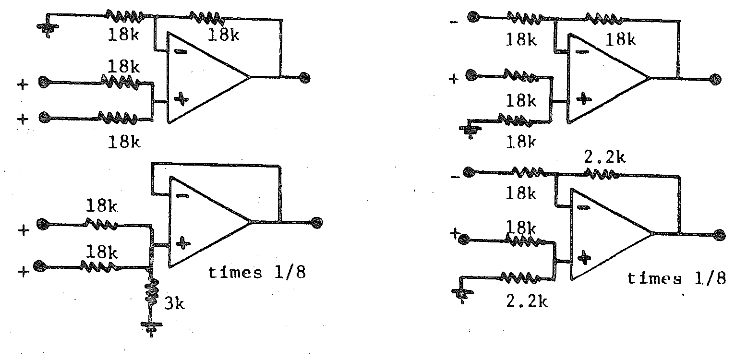
A HADAMARD TRANSFORM DEVICE

In realizing an actual transform, there is a decision to be made as to the type of transform, the number of points (in the discrete transform), and the implementation. Use of the Fourier transform (or the Fast Fourier Transform - FFT) is possible, but this would involve the handling of complex arithmetic. The FFT is easily implemented by computer, but except for a limited number of points, it is too slow to run in real time. A real time network can be formed by doing the Hadamard transform by an analog method. The Hadamard transform (HT) is a device which gives a spectrum in terms of Walsh functions rather than sinusoidals, and is a fast computer method similar to the FFT except it does not use complex arithmetic. In fact, the only arithmetic to be done is a multiplication by +1 or -1, and a summation, and these are easy jobs for op-amps. The HT, like the FFT has a "Flow Graph" called a "butterfly" as shown at the right. Each of the nodes of the network can be realized by using the op-amp differential amplifiers or summers as shown below. The transform network shown involves only 8 points, and does analysis only over a range corresponding to three octaves, but is large enough to examine the method. The inverse transform uses the same basic network.

It is a curious fact that when we choose to analyze a waveform by means of Walsh functions and an analog method that the HT may not be the most efficient. The discrete Walsh transform (DWT) just happens to be more efficient for small networks.

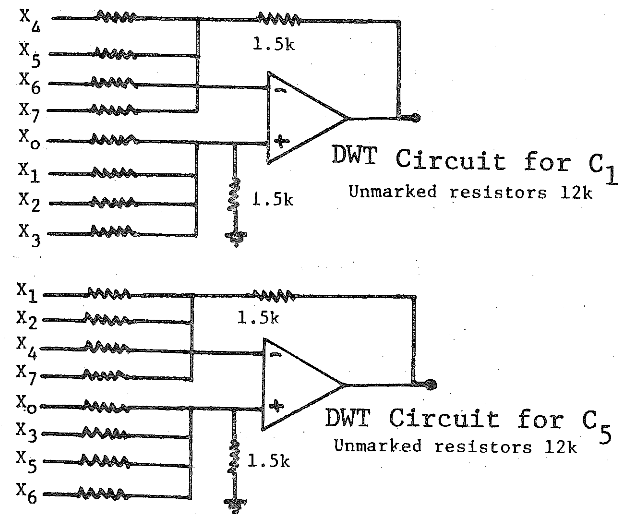


The DWT uses a matrix multiplication method (see reprint paper in chapter 9d). Since all the elements of the matrix are either +1 or -1, the matrix-vector product used to evaluate the coefficients becomes a simple summation and can be performed by op-amp circuits of the type shown at the right. For networks of practical size, this approach saves many op-amps. For networks of many points (perhaps 64 or 128), the original HT method may be a saving as even though more op-amps are required, far fewer resistors will be.

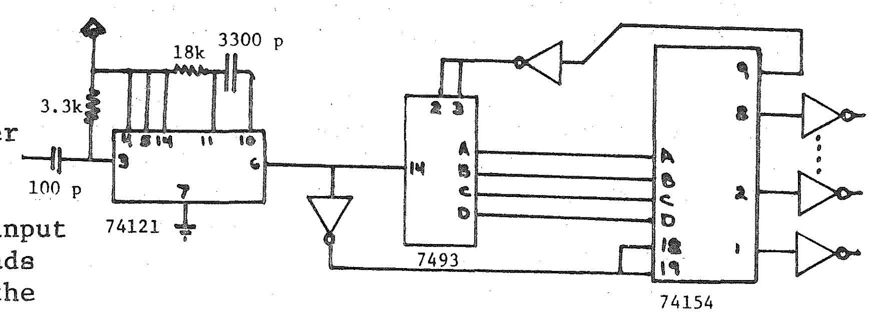


OP AMP CIRCUITS FOR THE HT NODES

The networks implemented as shown do analysis in space - the various coefficients and data points appear at various positions in the network. To be useful for processing audio signals, the samples must be sequenced in time. This involves the use of a sequencing circuit and a sample and hold for each of the voltages to be used. A convenient circuit for the sequencer uses the 74154 data distributer as indicated at the right. This circuit provides a short pulse for the triggers to the sample-and-holds at each of eight outputs. As the eighth output falls, another 74121 circuit fires and provides the main transfer signal for the coefficients.

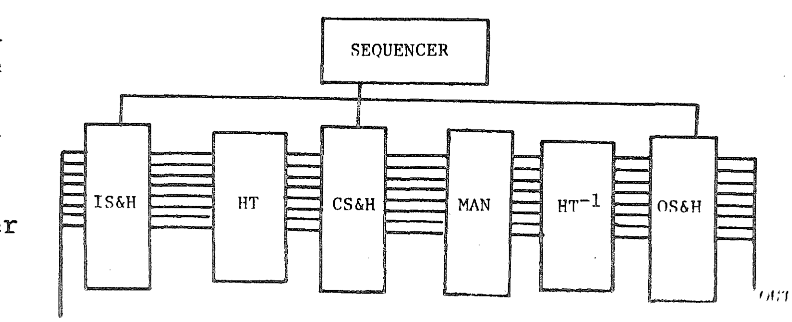


OP AMP CIRCUITS FOR DWT



SEQUENCER FOR HT NETWORK

The overall Hadamard transform scheme is indicated at the right and below. The operation of the sequencer and connection of the sample-and-hold circuits should be noted. The eight input sample-and-holds have a common input and are clocked in sequence. This loads eight input samples. On the fall of the eights enable pulse to the eighth sample-and-hold, a single pulse transfers all the coefficients (of the eight input samples) to storage sample-and-holds known as the coefficient sample-and-holds. Once the coefficients are stored, they are subjected to some form of manipulation and then fed to the HT^{-1} (inverse HT) which resynthesizes the first frame in a manipulated from. When the next sequencer pulse arrives, it reloads the first input sample-and-hold, but this does not change the stored coefficients of the previous frame. At the same time that the first input is reloaded, the first output sample of the resynthesized frame is unloaded. All output sample-and-hold circuits have



IS&H - Input sample-and-holds (8)
CS&H - Coefficient Sample-and-holds (8)
MAN - Coefficient manipulators
OS&H - Output sample-and-holds (8)

OVERALL HT ANALYSIS-SYNTHESIS SCHEME

a common hold capacitor and a common buffer. The output sample points are thus unloaded as new inputs to the HT network are being loaded. The output voltage of the whole scheme is the last HT^{-1} voltage sampled. Observe that if the coefficients were passed directly without manipulation, and circuit errors can be neglected, the output waveform is an exact replica of the input delayed by eight samples.

The device is a "sequency" filter. It removes certain Walsh function components, or alters their amplitude as the coefficients inside the network are manipulated. It is useful to manipulate the coefficients with analog multipliers so that each of the coefficients can be enveloped into the inverse transform.

The relationship between the clocking frequency and the input signal frequency alters the results greatly. It can be seen for example that if the clocking frequency is much greater than the input frequency, each of the input voltages in an 8-sample "frame" will be fairly close. In this case, the main part of the network energy will be concentrated in the C_o or DC coefficient. If the frequency of the input is 1/8 the clocking frequency, the set of coefficients are constant in time. Near integer multiples produce complicated beating effects.

As a practical matter, the network coefficients can be examined to see which ones have significant energy. Then two approaches are possible. The user can either work with the high energy coefficients, or block these and work with the residual activity by adding amplification. In the former case, the results will be somewhat predictable. In the latter case, it is difficult to estimate what the residuals are, and since they are of much lower amplitude, system errors may contribute much to the output as well. In either case, useful sound combinations may result.

Additional outputs are possible. Squaring the coefficients with the multipliers for example gives a Walsh "Power spectrum," and the inverse transform gives a form of autocorrelation. It is also possible to convert the coefficients to a Fourier form.

The experimental device has proven useful for producing unique sounds of various types. For the most part however, an approach that is tried for a certain effect may give something altogether different. Thus, further experimentation on the control of the network is indicated. Furthermore, it may be useful to use a network of larger size (16, 32, or 64 points) to give a greater bandwidth. For more details on the present device, see: B. Hutchins,"Application of a Real-Time Hadamard Transform Network to Sound Synthesis," JAES July/Aug (1975).

CHAPTER 7A

DESIGN OF CONTROLLERS

CONTENTS:

INTRODUCTION

It is through the controller that the electronic music system becomes useful to the musician. The fact is that to date relatively little research into new types of controllers has been done. For the most part, more or less standard musical controllers have been adapted for electronic music. There is surely a lot that can be done in this area to aid the musician.

Much of what is covered here requires additional circuitry of the type discussed in the next chapter to give the proper interfacing with electronic music modules. At times, it is difficult to say what part is the controller and what part is the interface. Thus this chapter and the next are best read as a unit. Sequencers are another type of controller and are discussed in chapter 7c. Parameter recovery devices, which are really a "translational" type of controller (pitch and envelope followers for example) are discussed in chapter 6b.

KEYBOARDS

Many musicians prefer or demand a standard electronic organ type of keyboard. These are the easiest type of controller for a keyboard player to use as they have the standard key spacing and feel. This type of keyboard would be extremely difficult to build. Even for an experienced machinist with a properly equiped workshop, this would probably require more time than it would be worth. In addition, some special spring contact wires (special alloy and gold plating) are required. All and all it comes down to the fact that if you want this type of keyboard, you had better be prepared to pay something like $100 for a 4-5 octave model. The keyboards generally used are made by Pratt Read Corp. of Ivoryton, CT 06442 and meet AGO (Americal Guild of Organist) specifications. The individual buyer will have to obtain one from a dealer most likely.

There are numerous ways of rigging up makeshift keyboards with mechanical contacts. None of these are as reliable as the standard keyboard. If they are made mechanically sound, eventually they will suffer from corrosion of the contacts. One source of gold plated wire that will help things is the clippings from gold plated transistor leads.

It is possible to use the standard keyboard format and use touch controlled keys. The basic design starts with a keyboard design etched on a printed circuit board. This is then attached to electronic circuitry that detects the presence of a finger in contact with the etched "key." Since these circuits are usually capacitive (see following chapter), it is generally possible to paint the keys in the standard black and white manner, and still have the circuitry work properly.

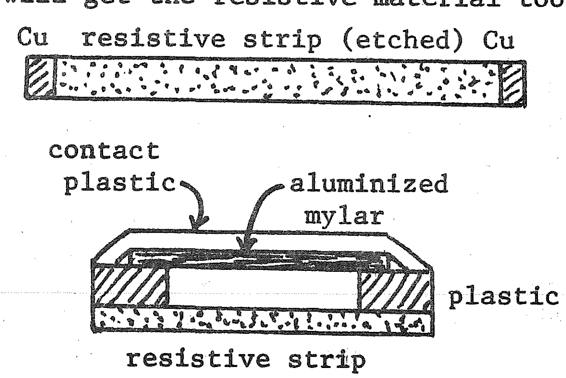
Finally, there is a wide range where experimental keyboards can be used. These may be specially designed arrangements, or keyboards designed for non-musical purposes (such as a calculator keyboard). Keyboards like those for calculators may be very reliable when used in music systems. However, the obvious drawback is that they are not natural for most musicians. However, some composers prefer to work with unfamiliar keyboards as they tend to prevent automatic learned responses from transferring standard structures to their music.

CONTINUOUS CONTROLLERS

When you push a key on a keyboard, you have your note. The keyboard is set for a discrete set of pitches, and while it can be easily tuned up or down, once set there are only as many pitches available as there are keys. A second type of controller works with a continuous range of pitches. The simplest type of controller of this type is the potentiometer. This can be tuned for any of the keyboard pitches or any inbetween. However, for any sort of wide range controller, the pitches cannot be set exactly because the user cannot set the pot knob accurately enough. Thus, as with other continuous instruments such as the violin, exact tuning depends on aural feedback from the instrument to the players ear. However, this sort of controller is also often used for pitch glides or for note "bending" where the exact pitches are not important.

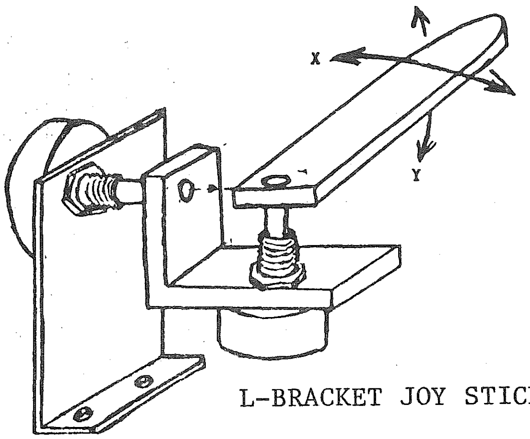
One popular continuous controller is the linear "Ribbon Controller" which can be thought of as a continuous keyboard or a slide pot without a slide. Pressure is applied somewhere along a strip and contact is made between a conductive pickup and a resistive strip carrying a constant current. This transfers out a voltage proportional to the distance of the point of contact. With such a device, vertical action produces constant voltages (selected from a continuous set) while horizontal motion maintaining pressure produces glides.

Rossum has described a simple ribbon controller (EN#31, pg. 9). The controller uses a 1" by 18" strip of "Micaply Ohmega" (Mica Corp, 4031 Elenda St., Culver City, CA 90230). This material is a copper covered resistive material. The copper can be etched off (must use chromic acid for this, ferric chloride will get the resistive material too) leaving short contact strips on the ends. Plastic strips about 1/4" wide are mounted along the sides lengthwise, and a strip of aluminized mylar is placed over the strips and fastened in place with a strip of contact plastic. The general idea is suggested by the drawings at the right. The resistive strip is then powered by applying 5 volts along its length, and an interface circuit of the type described in the next chapter can be used. If desired, the strip can be driven by a constant current source instead of just the 5 volt supply.



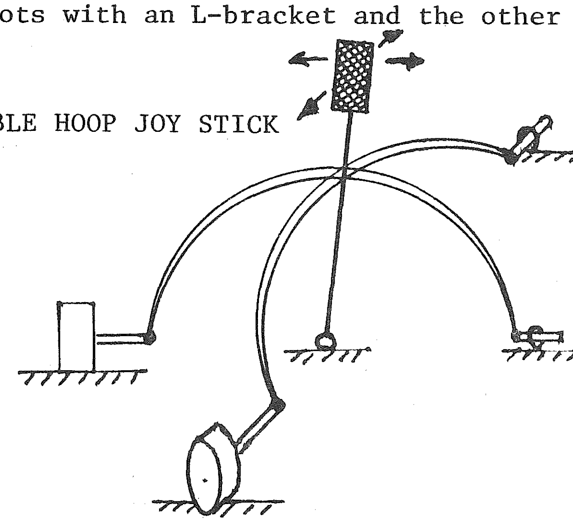
Craig Anderton has described a "Dial Controller" formed by adding a larger type of handle to a pot shaft, and adding a calibration dial (EN#34 (19)). The device is best formed with a log pot since this spreads the notes and makes them easier to play. The best note spread is found to be about two octaves for a single turn.

A "Joy Stick" (or X-Y Controller) is a continuous, two-output controller that can be considered a two-dimensional pot. A mechanical arrangement is used to convert left-right motion of the top of the stick to rotational motion for one pot, and forward-backward motion is converted to rotational motion for a second pot. Two arrangements for doing this are common. One connects the pots with an L-bracket and the other uses two semicircular double hoops as shown below:



L-BRACKET JOY STICK

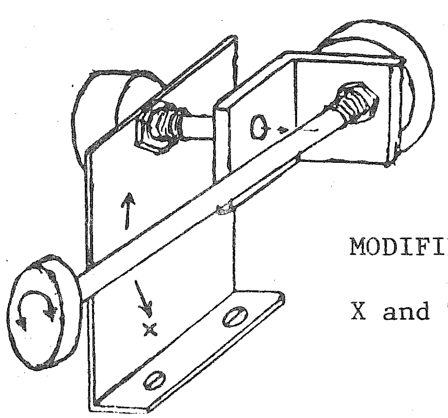
DOUBLE HOOP JOY STICK



The pots are used as voltage dividers by applying a voltage across them and removing the control voltage from the pot wiper. It may be best in this case to use a log pot and use the section of the pots range that has the most resistance per degree of rotation. This is because the joy stick has much less than full range of rotation for the pot. In any case, scaling and ranging units can be added to the output.

The joy stick gives two voltages which are continuous. They can be varied as rapidly as the hand can move. If it is desired to have the device silent between settings, a gate and/or trigger pushbutton can be added to the general setup. For example, this might be mounted on the top of the stick and operated with the thumb.

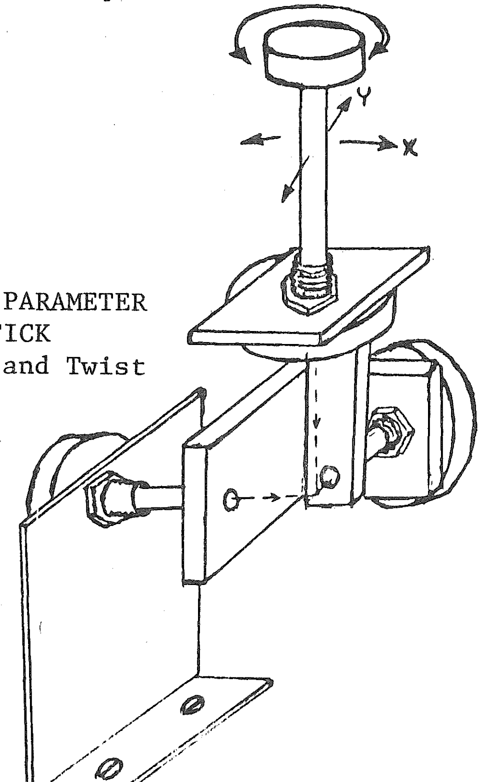
A number of variations on the basic joy stick are possible and are indicated below. The one directly below removes one of the rotational motions and replaces it with a twist, as in the dial controller. The one below and to the right adds a twist motion to form a third parameter control.



MODIFIED JOY STICK

X and Twist Motions

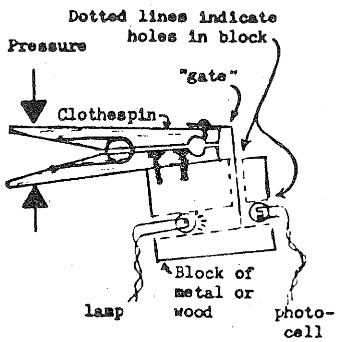
THREE PARAMETER
JOY STICK
X, Y, and Twist

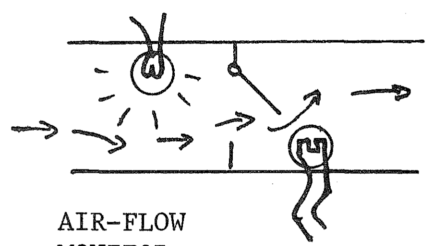


MISCELLANEOUS CONTROLLERS

Other controllers are possible, and these are often intended to monitor some physical parameter. A number of these are discussed in connection with the interface units of the next chapter. These use such things as time delay and pressure to arrive at controllers that will possibly allow the musician more control over expression by more properly monitoring the way the control elements were activated, not just which one and when (as is the case of any switch closing).

One additional physical quantity that is useful is light. Light can activate an electronic system by causing a voltage to be generated in a photocell, or by changing the resistance of a photoresistor. This can be made quantitative by controlling the light; first by controlling the base level (by channeling the light and keeping ambient light out), and secondly by a mechanical device that is controlled by the user. With such a system, a material can be compressed to block greater and greater amounts of light, and this can be a measure of pressure as well as a time record of pressure. Two other devices have been described. The "electric clothespin" was suggested by Bill Hemsath as a device that could be controlled by the pressure of the teeth. The basic idea is indicated by the diagram below. A second device uses air flow to move a membrane out of the way and let more light strike a light sensitive cell in a manner suggested by the drawing on the right below:





Another type of controller uses the mechanical vibrations of a conventional instrument. Typically this might pick up a strings vibrations, and may be fed to a pitch-to-voltage converter first, or the signal might drive processing circuits. A magnetic type of pickup is common for this purpose. The instrument is outfitted with conducting strings and the part of the string that is in a magnetic field generates a voltage proportional to its velocity. This voltage may be integrated to give a voltage proportional to the strings position if desired. Another type of string pickup has been described by C. H. Agren ("Photoelectric Vibration Probe for String Instruments," Electronic Engineering, Dec. 1974, pg. 18).

A number of programmable controllers involve some sort of moving paper tape or chart. These are similar to sequencer devices. One outstanding possibility of such a device is that a composer may be able to write on the paper with a code that will be properly interpreted by sone electronic pickup as the paper moves. This sort of writing is not unlike the process of score writing. An example of this sort of approach has been described by Brent Gabrielsen ("A Programmable Control Device for Analog Synthesizers," AES Preprint # 1000 (S-3), Sept. 1974).

Finally, with the advent of computer control of synthesizers, controllers for computers will become directly available. There are a wide variety of such devices. Of particular interest are "Light Pens" which effectively write on the faces of cathode ray tubes, and certain devices which pick up the patterns written with a regular type of pen. Such devices make certain types of programming very natural and easy to interpret.

CHAPTER 7B

DESIGN OF CONTROLLER INTERFACES

CONTENTS:

INTRODUCTION

A controller interface is a device which changes a useful control action into an appropriate voltage level and/or waveform to control the music synthesis system. In some cases, (e.g., a pot voltage divider) it is difficult to say what is the controller and what is the interface. In general, we shall find the controller to be mechanical parts while the interface is electrical. We shall often be concerned with keyboards for example, and shall consider the mechanical keys and switches to be the controller while anything else that is needed is the interface.

We are most often concerned with three types of controlling signals. (1) Control voltages - those which have a value that controls the parameters of the synthesis system. (2) Gates - timing signals that have a start, a duration, and an end. (3) Triggers - short pulses defining one point in time, often the same as the onset of a gate. However, triggers may occur at other times as well, and might well be considered as "retriggers."

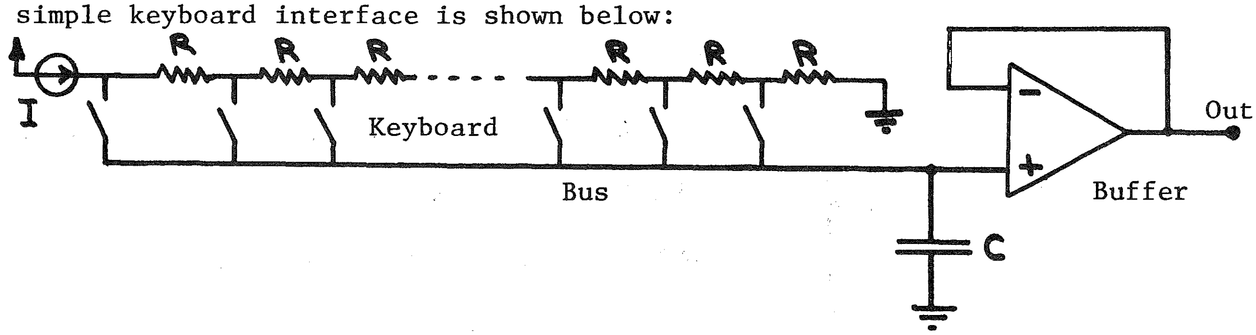
Inputting of other controllers is often possible and can be done as a matter of convenience. Inputting of sequencers may be a special case as some sequencers may provide their own gates and triggers. In others, design can be simplified if a keyboard interface is used.

KEYBOARD INTERFACES

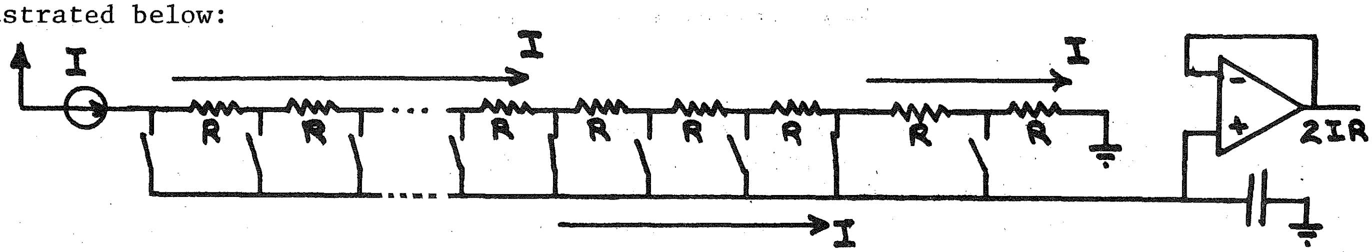
The voltage divider is a convenient way of obtaining a set of discrete voltages. The necessary set of voltages for a keyboard depends on the module which the keyboard is to control. When the driven module has an exponential response, the set of voltages are all equally spaced to provide intervals that have the same ratio. For a 12 tone scale and a 1-volt per octave response, the required voltages are 1/12th of a volt apart. We could thus generate a set of voltages by using a string of resistors of value R, and driving the string with a current (1/12)/R.

For a practical keyboard system, the string will be composed of one resistor for each key, and the resistors will usually be mounted right on the keyboard switches. The divided voltages thus appear on one terminal of the keyboard switches while the other terminals are all connected to a common bus. On the other side (the bus) it is the usual practice to include a sample-and-hold (S&H) circuit. The reason for this is quite simple: On a traditional keyboard instrument, it is generally the case that the tone continues without a change of pitch as the decay phase proceeds following the lifting of a key. If it were not for the S&H on the keyboard bus, the pitch information would be lost - the VCO would go to its uncontrolled state.

A simple keyboard interface is shown below:



The string is driven by a current source as shown. This assures that if more than one key is pushed down, the lowest one will sound. This is because the current in the line will be shorted through the bus if more than one key is pressed. Only the voltage that is dropped across the resistors between the bus and the ground will appear. This is illustrated below:



The S&H is formed by simply attaching a storage capacitor to the keyboard bus and buffering the capacitor voltage. Current flows from the string to charge the capacitor to the divider voltage. The output impedence of the divider string is equal to the total resistance below the tapped point. Thus for a resistor string of 100 ohm resistors on a 49 note keyboard, the maximum output impedence of the keyboard would be 4.9k. The RC time constant for the sampling section should be only a few milliseconds. A value of C = 1 mfd would give RC = 4.9k·1 mfd or about 5 ms as the maximum charging time constant. This is acceptable, and the 1 mfd capacitor makes possible a fairly long storage time with a good buffer. In many cases, it may be possible to use a smaller capacitor however.

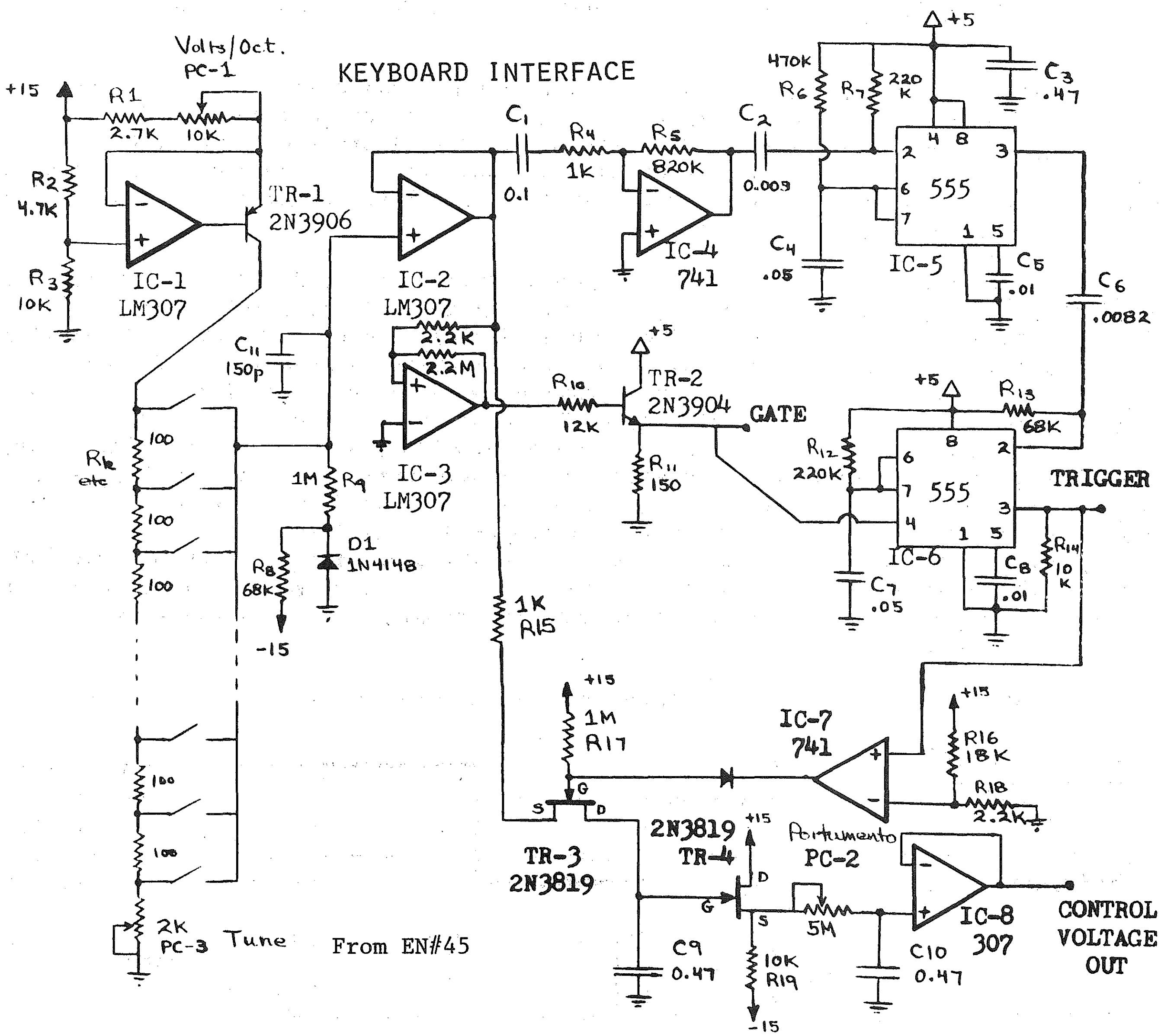
In general for synthesizer use, it is necessary to obtain timing signals from the keyboard in addition to the control voltage. This is sometimes done with a second set of switches on the keys. However, all the necessary timing information is present in the control voltage. The three items of information available in the control voltage are:

1) Which key is down. This is given by the magnitude of the control volatge.

2) At least one key is down - the gate signal. Given whenever the bus voltage exceeds some minimum value.

3) The key that is down is changing - the trigger signal. This is obtained when the bus voltage has a sufficiently rapid rate of change.

Below is shown an interface circuit that requires only one keyboard bus and yet supplies the control voltage, the gate, and the trigger:



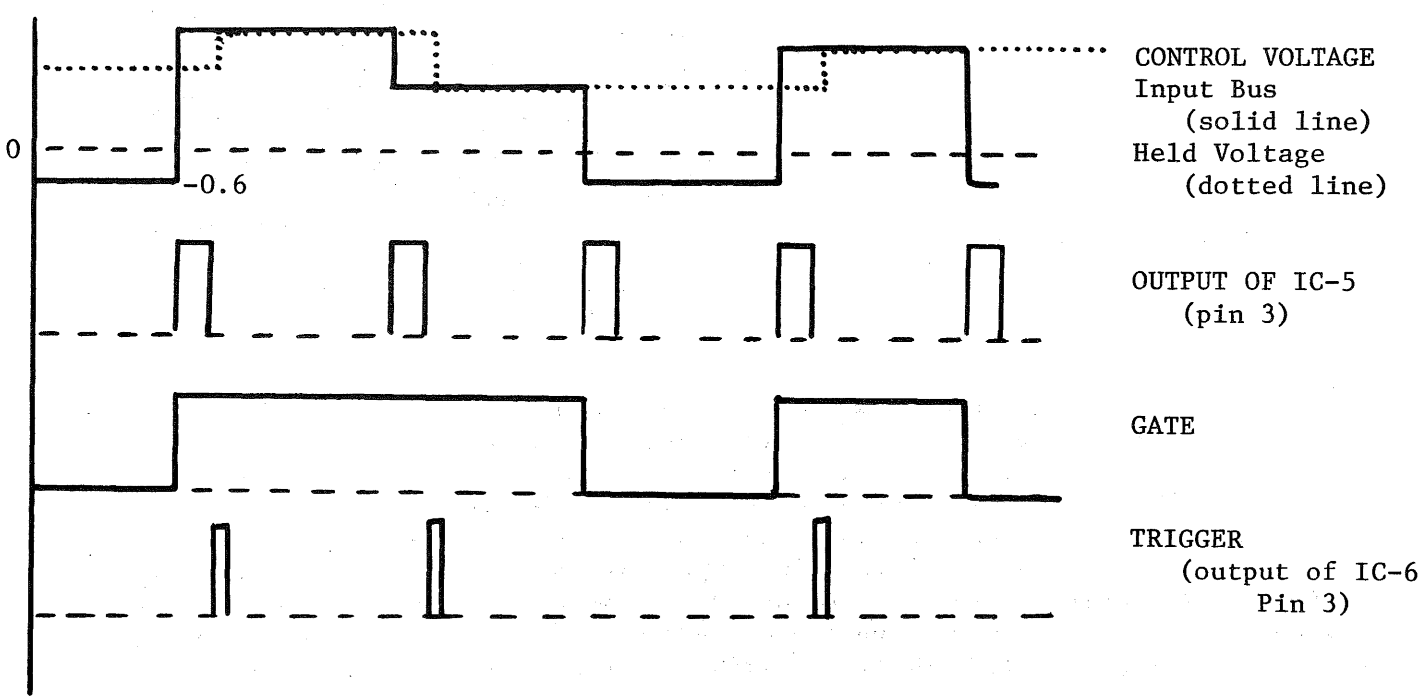
IC-1 serves as a current source for the resistor string. The current source will work with any number of keys for a total keyboard voltage up to 7 or 8 volts. Thus it will work with 3, 4, 5, and 6 octave standard keyboards. All the resistors between the keys should be the same value on the order of 100 ohms (20 - 200 ohms preferred) and should have a tolerance of 1%. The current source trimmer PC-1 is adjusted to supply a current so that 1/12 volt is dropped across each resistor. It may be necessary to change the value of R1 and PC-1 somewhat for some values of keyboard resistors. The voltage between +15 and the emitter of TR-1 is about 5 volts, giving a current through R1 and PC-1 of 5/(R1+PC 1). This is the emitter current since no current flows into the - input of IC-1. For a high-beta transistor, the collector current is the same as the emitter current. Thus, the same current is supplied to the resistor string.

IC-2 serves as a buffer for the keyboard bus voltage. Capacitor C11 is not a storage capacitor - it just keeps noise off the bus line - it is too small to hold much of a charge. R8 and D1 hold the buffer at about -0.6 volts when no key is down. When a key goes down, the output impedence of the keyboard string is much less than R9 = 1 meg so the keyboard rules. When the voltage on the bus line exceeds ground level (going from -0.6 to some positive value) IC-3 goes high, and TR-2 provides a gate level of +5 volts.

IC-4 and IC-5 form a differentiator and one shot combination. IC-4 triggers IC-5 whenever a voltage changes by 1/12th of a volt or greater and changes rapidly (as fast as IC-2 slews). While IC-5 only triggers when a negative transition occurs, this is certain to occur since contact bounce of the mechanical contacts of the keyboard typically make and break something like 100 times before settling. Once triggered, pin 3 of IC-5 remains high for about 25 ms.

When pin 3 of IC-5 falls, a 15 ms trigger occurs at pin 3 of IC-6, provided that the gate is high (pin 4 of IC-6 is enabled). This trigger is available to control an envelope generator. It is also coupled through IC-7 to the gate of TR-3. This causes the S&H to sample. The keyboard bus voltage is thus stored as capacitor C9 is charged from the output of IC-2. C9 should be a polyethylene or polystyrene dielectric type for best results. The voltage on C9 is buffered by TR-4 (source follower). There is a voltage drop on the order of 1/2 volt across the source follower, but this is seldom a problem since it is a constant offset. The source voltage is the correct held voltage for all practical purposes. The final circuitry PC-2 and IC-8 is a portamento circuit. The voltage on the source of TR-4 charges C10 through PC-2 as an exponential ramp. The final voltage output appears at the output of IC-8. The bias current through IC-8 flows through PC-2, so there may be some pitch shift as PC-2 is adjusted. If this is a problem, IC-8 can be replaced with a type 308 or a type 536 op-amp. TR-4 can be replaced with a 536 as a follower if desired.

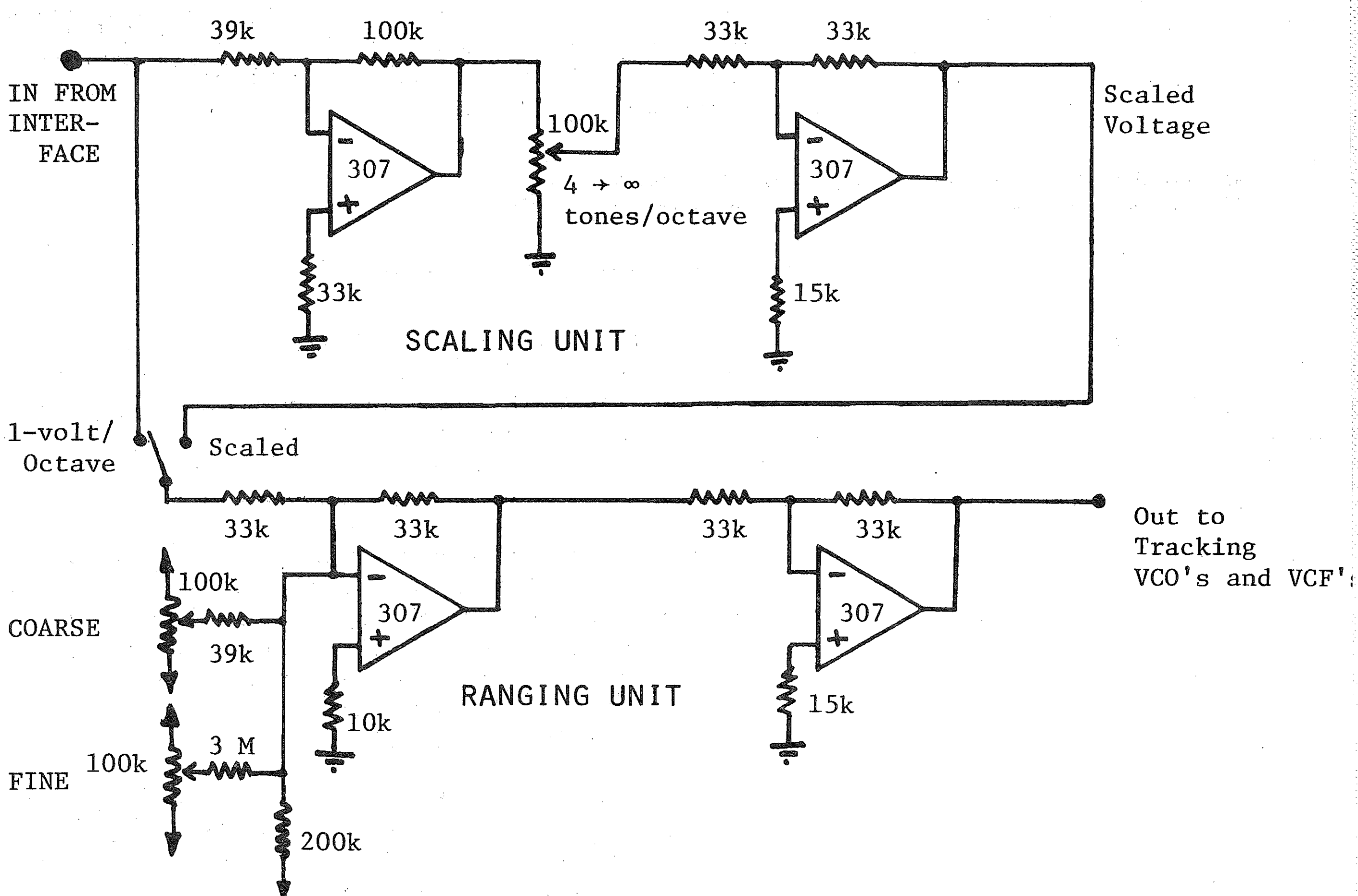
The timing diagram for the interface unit is shown below:



The timing diagram illustrates the need for the delay provided by IC-5. When the bus voltage changes from some upper keyboard value to its standby value of -0.6, this corresponds to the lifting of the last key in a sequence for example. This transition triggers IC-5, as does any change exceeding 1/12 volt. Yet this does not indicate a new voltage to be sampled - we want to hold on to the last actual keyboard voltage, not the -0.6 volt standby. Thus the delay provided by IC-5 lets the gate go low before IC-6 (the trigger) is triggered. Since the gate is low, the trigger is inhibited. The trigger is therefore the AND function of the gate and the falling edge of the output of IC-5. Note that when the input voltage changes and the gate does not go low (the player slides from key to key, not lifting them completely between notes) triggers will appear to retrigger the attached envelope generators. This provides some envelope indication that the tone has changed, not just a pitch change. The time scale on the timing diagram has been greatly exaggerated to show the relationship of the timing signals.

Note that there are only three wires that connect the keyboard to the interface, the current line, the bus line, and ground. This means that the keyboard may be connected with a three terminal connector such as a stereo phone plug. In fact, if this is done, it is possible to use the interface unit for general input signals as we shall see below. The tune control PC-3 can be mounted on the keyboard and this allows the performer to turn the pitch up or down to tune to other instruments. It is probably best to make PC-1 a trim control rather than a panel control. The interface can be tuned up for a 12-tone scale and held there. It is important to be able to change the voltage scaling to use different musical scales, but this can be done with additional scaling controls, and the 12-tone scale can be regained easily without need for careful resetting of PC-1.

A control voltage scaling and ranging unit is shown below:

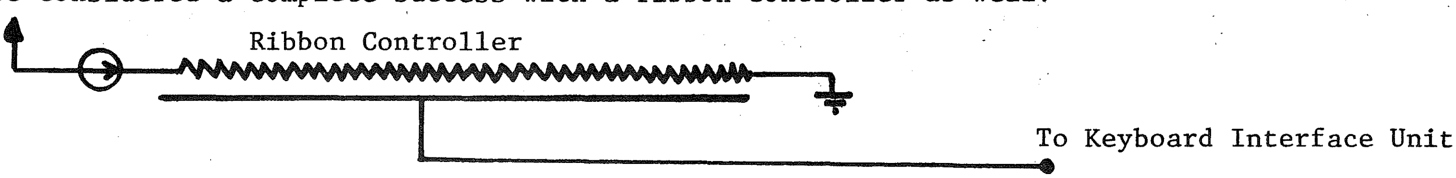


The purpose of the scaling unit is to produce responses other than 1 volt/octave. There is often an interest in producing scales of more than 12 tones per octave, with 19 and 31 tones per octave being popular. This is basically an attenuation process. At times a sharper response may be desired, and that is the reason for the initial amplification before the attenuating pot. The ranging unit is used when it is desired to change the pitch level of the overall patch with a single knob. In some patches there may be three or more VCO's and a VCF or two that all track the keyboard voltage. If the user wants to change the pitch level of the whole system, it would be necessary to adjust many different knobs and carefully adjust the exact tracking. With the ranging unit, it should be possible to achieve tracking once and then run the pitch level up and down with the ranging unit controls. Note that if the ranging unit is used, it would be easiest to adjust the volts/octave control on the interface to 1-volt/octave while monitoring the voltage at the output of the ranging unit. In this way, it will not be necessary to critically select the resistors in the ranging unit.

USE OF KEYBOARD INTERFACE WITH OTHER INPUTS

The keyboard interface above was designed to operate in what would be considered the expected manner with a standard keyboard. It can also be used to handle a number of other controllers, sequencers, or general voltage inputs.

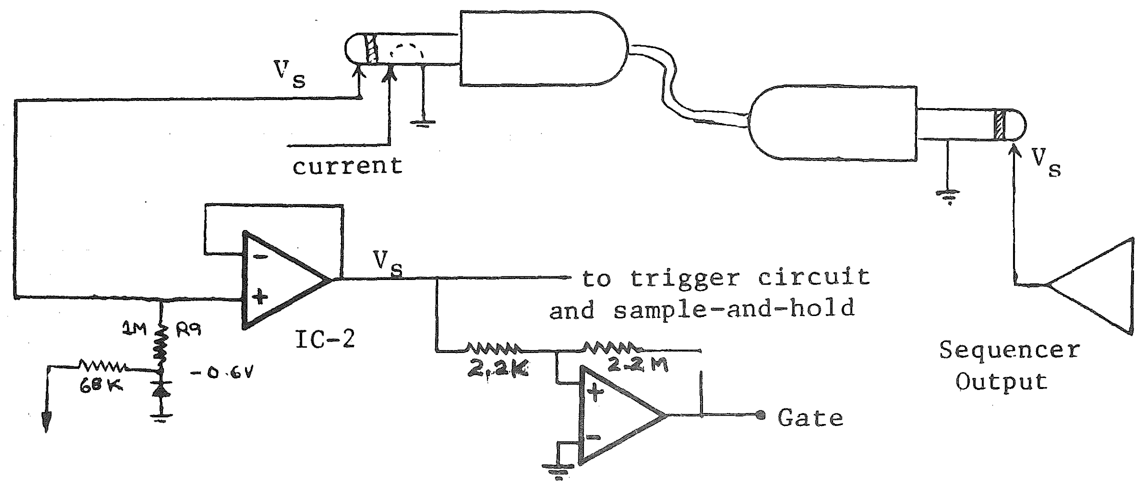
The keyboard interface will work with ribbon controllers in a fairly natural manner. When the ribbon is pressed to the resistive strip, the interface works just as it would if a switch contact were closed - it produces a gate and a trigger. When this contact is lifted, the gate disappears, as in the keyboard case. The difference comes in when we consider a glide from one voltage to another. In general, this will not have a fast enough change of voltage to trigger the differentiator. Therefore, no trigger will appear. This is probably the best thing however, as it would be difficult to define a time when the trigger should occur during the glide. In the case of the keyboard, this is not a problem as the voltages are discrete. It seems that the keyboard interface may be considered a complete success with a ribbon controller as well.



A joy stick is a similar problem as far as glides are concerned. If we input the voltage from the joy stick, gates will appear when the voltage exceeds zero, and in general no triggers will appear. The joy stick voltage will be crossing zero instead of jumping rapidly across as is the case with the kyeboard and the ribbon controller. This is the reason that IC-3 is configured as a weak Schmitt trigger rather than a standard comparator.

Sequencers may be input through the keyboard interface as well. It responds to a sequence of voltages just as it does the the keyboard bus voltage. An output will result when the sequencer voltage exceeds ground level. Triggers will appear in the interface if the sequencer voltage changes discretely by amounts on the order of 1/12 volt or larger. In the same way, any voltage can be input to the interface. Slowly varying voltages will produce only gates while those with discrete steps can be expected to produce triggers as well. In some cases where triggers are not produced but are needed, the rising edge of the gate can be used to supply triggers to an envelope generator.

One convenient way of setting up the interface so that it can be used with a general type of input is shown at the top of the next page. A standard stereo phone jack is used for the keyboard. When a different type of voltage is to be input, it is a simple matter of of using a standard phone plug. The current in this case is shorted to ground but since it is supplied by a current source, this is no problem.



DIGITAL INTERFACES

Digital methods are particularly useful where a discrete set of voltages are used. For example, when a standard 1 volt/octave system is used with a 12-tone equally tempered scale, all the voltages are 1/12 volt apart. For example, a standard 61 note keyboard (5 octaves plus a note) can be represented by 6 bits of digital information (2^6=64). Such a system starts with a scanning device that rapidly checks for a key that is down. When it finds a down key, the corresponding digital word is latched and fed to a D/A converter. Additional logic circuits can be used to give gate and trigger signals, and also select only the lowest key if more than one is pressed.

The digital interface of this type has several outstanding advantages and possibilities. (1) No sample-and-hold is required. The digital latching holds the output voltage without drift. (2) Bounce suppression of the mechanical contacts of the keyboard is easily built into the latching mechanism. (3) The scanning and digital logic system is the start of a polyphonic keyboard system. The major drawbacks are (1) the added expense of the system, (2) only discrete voltages can be handled, and (3) the device is hard wired to the keyboard and is not available for other controllers.

DIGITAL COUNTING INTERFACES

Numerous methods have been proposed and used to generate the actual tones for a musical device by a digital counting method. One method suggested by D. Gossel ["Generation of Musical Intervals by a Digital Method," Phillips Tech. Rev 26 170 (1965)] starts with a high frequency oscillator and divides the oscillator by a series of numbers (in the range of 100-200) selected so that frequencies at the output of the dividers approximates the musical scale tones with reasonable accuracy. The digital counter is a divide-by-n counter where n changes with the scale tone. This is accomplished by using a programmable divide-by-n. These typically come in 4 bit units that divide by numbers from n=2 to n=16 according to how four programming pins are programmed. Programming is accomplished typically by grounding the pins as needed. A diode matrix can be formed so that when a key on a keyboard is depressed, the correct pins are grounded for that tone. Typically, two of these divider chips (DM8520, 74193, etc.) are cascaded to give 8 bits so that n can be selected from numbers from 2 to 256 by programming 8 pins. The most accuracy is usually obtained by working toward the high end, i.e., in the upper octave from n = 126 to n = 256. The actual numbers used can be obtained from the data of R. Stapelfeldt [JASA 46 pg. 478, 1969] who has used a computer program to optimize the numbers for minimum error. Note that for practical circuits, the upper frequency must be in the MHz range. It is only necessary to generate the top octave if more than one octave are used. Lower octaves are obtained by using flip-flops. The choices of divisors for minimum error are shown below for an 8 bit a 10 bit, and a 12 bit counter (results from Staplefeldt):

OPTIMUM DIVISORS FOR EQUAL-TEMPERED SCALES

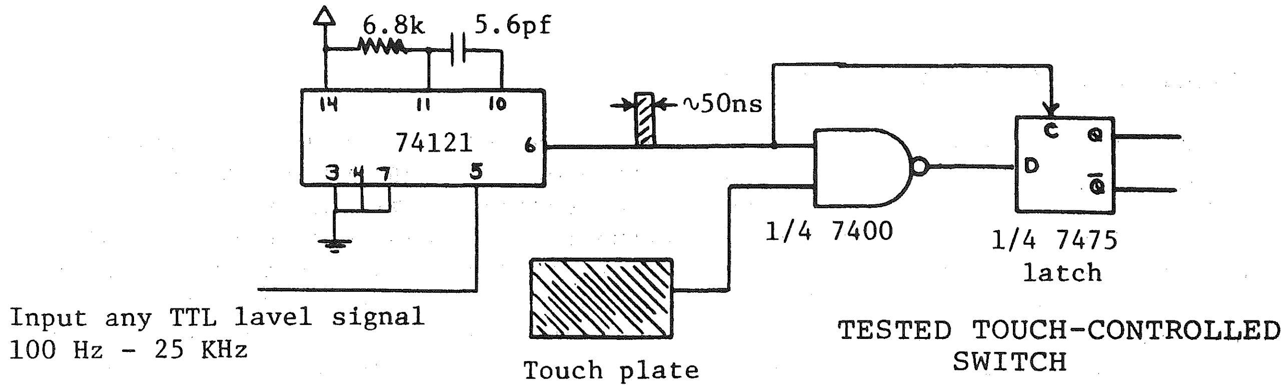
TONE (C scale)	8 Bits	10 Bits	12 Bits
C	116	508	1951
B	123	538	2067
A#	130	570	2190
A	138	604	2320
G#	146	640	2458
G	155	678	2604
F#	164	718	2759
F	174	761	2923
E	184	806	3097
D#	195	854	3281
D	207	905	3476
C#	219	959	3683
C	232	1016	3902
Max % Error	0.15	0.03	0.007

The divide-by n method and two other methods using digital circuits are discussed by R. Cotton ["Tempered Scale Generation from A Single Frequency Source," JAES 20 #5, pg. 376 (1972)]. Two additional methods of digital scale generation have been examined by G. Small ["Synthesis of Musical Scales Using Noninteger Frequency Division," JAES 21 #4 pg. 261 (May 1973) and "Rate-Feedback Binary Counter in Muisca Scale Generation," JAES 21 #9 pg. 702 (Nov. 1973)].

TOUCH CONTROL

Touch control or touch activation should be distinguished from the term "touch sensitive." A touch controlled device is a zero travel switch. It is used where a mechanical switch might be used, yet activation is through the touch of a finger, not by mechanical motion and contact of metallic parts. In this sense, the actual controller is just a "touch pad," and the only electrical design is for the interface. The term "touch sensitive" is used where some physical parameter of the contact of a mechanical device is recovered and used as a control. The recovered parameter could be velocity of a key, force on a key, final pressure on a key, or some combination. This requires some device that does travel, not a touch control pad. Thus, the two terms are quite different.

Touch controlled devices using the type 555 timer can be found in chapter 3e and in EN#31, pg. 5 (circuit of Craig Anderton). Another inexpensive touch control circuit was described in EN#52, pg. 10 based on a principle described by D. Cockerell in Electronics, Feb 20, 1975, pg. 108. The basic idea is to provide a pulse of approximately 50 ns to one input of a TTL NAND gate while the other effectively floats connected to a touch plate. When the pulse arrives at the first input, it immediately goes high. The other input starts to rise also as the stray capacitance is charged by leakage current. The output goes low as soon as both inputs are high. When a human finger is placed on the touch pad, the additional capacitance slows the charge rate of the second input. The falling edge of the pulse triggers the latch. Thus by properly setting the pulse width, it is possible to catch the NAND gate output either high or low depending on whether the pad is touched or not. A working circuit is shown on the next page:



TESTED TOUCH-CONTROLLED SWITCH

Note that the clocking frequency is really just an updating rate. Therefore, this frequency is not critical. Note also that the 74121 can drive any number of NAND gates (using buffers as needed), so that part of the circuit need not be repeated for each key. Thus a key requires just 1/4 of a 7400 and 1/4 of a 7475. It may be necessary to adjust the monostable time slightly (small variations in either the 6.8k resistor or the 5.6 pf capacitor) to get reliable operation on all keys. It would be best to use 7400 gates from the same lot to get them matched closely, although this was not necessary in several experimental circuits. For a start, the finger should touch the metal directly. The device is capacitively operated however, and it is possible to lacquer or paint the keys and still have it work. A little more pressure or increased contact area between finger and the plate may be needed in this case.

TOUCH SENSITIVITY TRANSLATORS

The basic method of velocity sensing involves two sets of contacts per key and the two must make or break in sequence. A logic arrangement determines when only one contact is made, and an output goes high. When the second contact is made, the output of the logic goes low. The width of the output pulse is inversely proportional to the velocity. The pulse width is converted to a voltage using an integrator and sample-and-hold if necessary. A voltage proportional to the velocity can be obtained by using an analog divider. In some arrangements, the voltage proportional to the pulse width is subtracted from some reference. While not strictly proportional to the velocity in this case, the effect is in the right direction. A problem that sometimes comes up with this sort of device is that the mechanical spacing of the contacts differs for different keys and the same velocity will produce different control voltages for different keys.

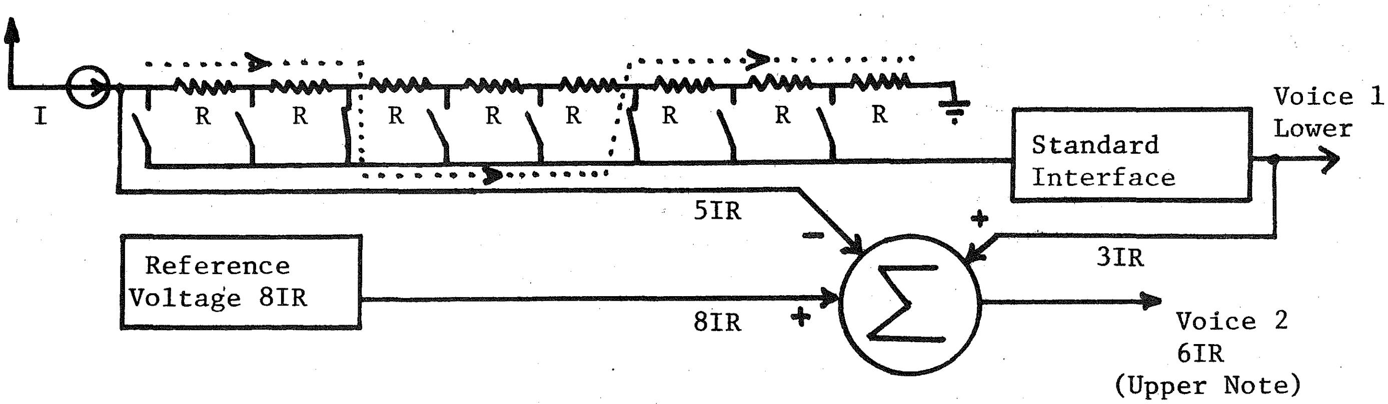
Pressure sensing is a matter of adding some material to the keyboard so that a different voltage is output for different key pressure. Possible arrangements use conductive foam of the type used to ship CMOS IC's. Increased pressure on this foam gives a lower resistance. Some devices employ light that falls on a photocell. This light may be blocked partially by some resilient material (e.g., similar to weather stripping) that blocks more light as it is compressed. Some experimental devices use a plastic tube under the row of keys. Typically this might contain some fluid that would be compressed as a key is lowered and pressed against it. The pressure of the liquid would be monitored at one end by a pressure to voltage transducer.

In other instruments that do not use a keyboard, similar translators may be used. Wind instruments for example may use some means of measuring air pressure or air flow. Air flow can be measured by having a flexible membrane move to let air pass and at the same time let more light fall on a photocell. Many devices which translate physical quantities into control voltages are under investigation to find those which work properly without imposing stylistic restrictions on the performer.

POLYPHONIC CONTROLS AND SYSTEMS

The standard synthesizer as a solo voice just as is a trumpet for example. Thus if we want to make a polyphonic instrument, the most straightforward procedure is to put a synthesizer on each key of the keyboard (each string of the guitar, etc.). To do this, at present, this means that the actual synthesizer for each key must be simple and inexpensive. An area of compromise is to have a limited set of synthesizer voices (complete synthesizer systems including VCO, VCF, and VCA with associated envelope generators) since, for example, the keyboard player seldom uses more than 6 to 8 voices and 10 would seem to be an upper limit unless the player is in the habit of pressing several keys with one finger or hitting tone clusters with the hand or forearm. If a limited number of voices are used for a larger number of keys, some method of assigning voices to keys as required is necessary. This means that the keyboard must be scanned for down keys and logic circuits circuits must assign down keys to unused synthesizer voices, and release voices when the key to which they are assigned is lifted. This is an inherently digital system. For most practical systems, the available voices must be identical (all trumpets or all tubas, not a mixture). If different voices are used, some sort of assignment must be used so that the desired voice comes up at the desired time. This might be something like a position rule (upper voice and lower voice) so that when there are two voices, the upper one is a flute and the lower is a trombone. The problem with such a system is that some decision must be made when there is only one voice (even during instantaneous releases of keys in one voice) so that uncontrolled responses do not occur and/or the player's style is inhibited.

We describe below a two voice system that is often used with analog systems. The system, while clever, is subject to the limitions of the position rule. The engineer should be familiar with the method because it is commonly used, not because we are suggesting its use. The principle is illustrated by the example below:



The example shows an 8 note system. We assume that two keys are down. Since a current source is used, the standard interface gives the expected voltage of the lower note (3IR). The 6th key is also down, and we would want a voltage of 6IR from this. Since the current that would flow through three of the 8 resistors is shorted through the keyboard bus, the voltage at the top of the string is only 5IR, not 8IR as it is when only one key is down. We can however have a voltage reference 8IR available. With this arrangement, and the summation network as shown, the output is in fact the voltage 6IR. This does work and can be made quite accurate. However, it is the dynamic timing conditions, not he basic principles that make this difficult to use.

CHAPTER 7c

DESIGN OF SEQUENCERS

CONTENTS:

INTRODUCTION

There are a wide variety of devices that are called sequencers. In fact, the variety is so great that two persons may be talking about two completely different things and think each other crazy when they discuss how they use them. Basically the sequencer is a device for producing a series of events, where the term events refers to something that occurs in time with certain musical properties. We will find it necessary here to form a rather formal operating definition of sequencers.

Since there is such a wide variety of sequencers, and because people have a wide variety of applications for sequencers, we will just give some basic ideas on the design of sequencers. None of the circuits given are to be considered complete - there is just to much that can be done on an individual basis. The designer must use his imagination and consider the needs of the users.

THE CHARACTERIZATION OF SEQUENCERS

The essential features of a sequencer are:

1) The production of an ordered series (sequence) of "events" in time. An event in turn has the following properties:
   a) A start that is indicated by the appearance of some one of the properties of the event.
   b) Properties during the event (such as voltages that may be changing). These properties may have translations in musical terms - e.g., one of the voltages may control pitch.
   c) An end point, followed by a (possibly zero) waiting time to the next event.

An example of an event would be the appearance of a control voltage of 4 volts at time equal to 10 seconds, and the disappearance of the voltage at time equal to 11 seconds. The event need not have a duration that can be humanly detected - it may be only a short trigger pulse that is detected by an electronic circuit and used to initiate some related event (such as a sequencer used to produce triggers for an envelope generator).

2) The capability of programming the output events so that they appear automatically after an initial activation.

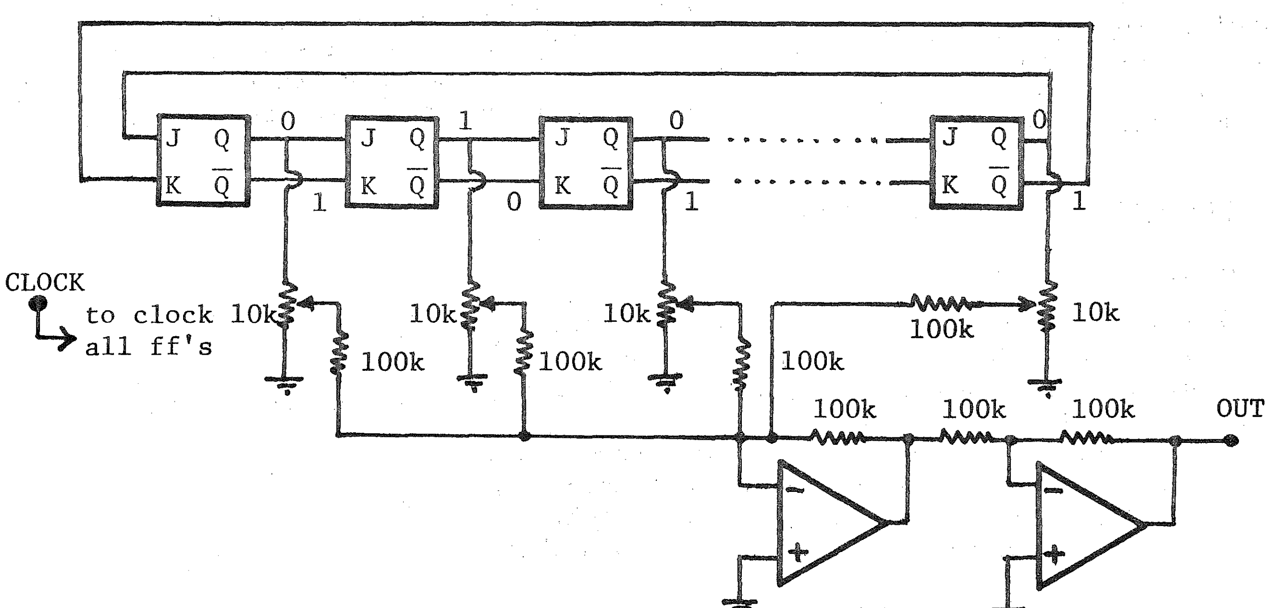
3) Provisions for using the events to drive additional equipment for a variety of musical purposes.

4) A known degree of control over the programming of the sequence. That is, we want to know if the output is completely defined, or random; if the events are all of equal duration or variable; if the outputs when converted to pitch will be restricted to equal tempered tuning or not, etc.

The above definition may seem unnecessarily formal, but it does provide a basis for logical sequencer design and use. For example, a keyboard meets the first condition, but not the second, as a human hand is required for each step, and for each repetition of the sequence. If the action of the hands is recorded (as in a "player piano") then the result is a sequencer. A recorded piano piece would meet conditions 1 and 2, but not 3 as the device would be used directly (i.e., you would listen to the music). If on the other hand, the recorded piano piece were fed to a pitch-to-voltage converter and the voltage were used to drive a VCO, the overall scheme could be considered a sequencer. Condition 4 is necessary as a practical matter, as we would like to know the types of sequences we can expect so that it can be applied to the production of music. Since there are so many types of sequencers, the term "sequencer" is not enough to tell if it will be useful for the type of music the composer has in mind.

BUCKET BRIGADE SEQUENCERS

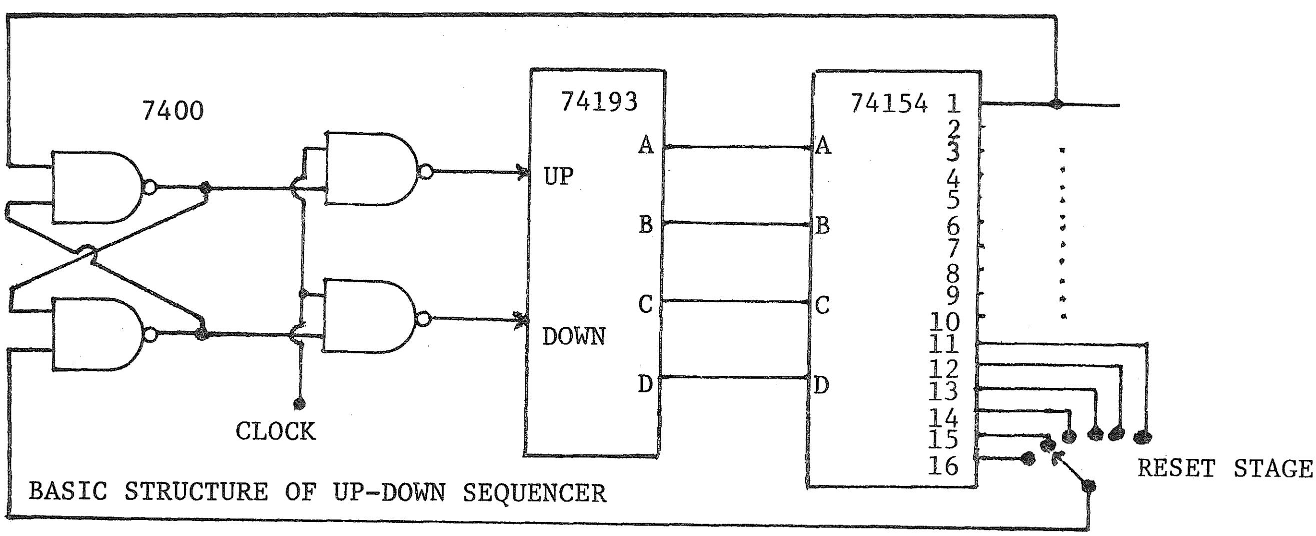
The first type of sequencers developed used digital logic counters in a "bucket brigade" manner. This is simply a counter circuit consisting of n stages with only one high stage at any one time. As pulses arrive to clock the counter, the high state moves down the line one step at a time. The clocking pulses may be regular or irregular. A position along the line has associated with it a potentiometer to select a portion of the voltage of the high state to be output. Such a circuit can be constructed from flip-flops as shown below:



This is a simple arrangement, but is not especially accurate since it sums the TTL voltages which are imprecise It may be satisfactory if the voltages are summed by ear, or diodes can be placed in the lines between the Q outputs and the pots. Another approach is to use CMOS for the counter.

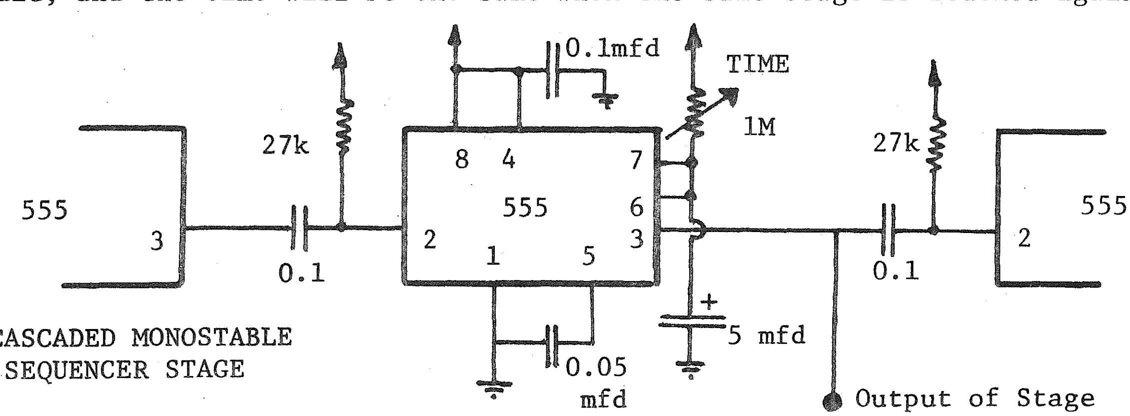
A saving in chip count can be achieved if packaged shift registers like the 74164 is used. Note that the sequencer above would be reset by using the asynchronous controls on the flip-flops. Typically, the first flip-flop would be set high while all the others are set low. This would be achieved with a pushbutton for example. If the 74164 were used, this provides 8 flip-flops, but there is only access to the clear of these. Thus, it would be necessary to start the register with something like a 7476 which gives two flip-flops with preset and clear.

Another popular approach employs a 4-bit binary counter and the type 74154 4-line to one line decoder. The basic idea is shown below. The counter could be a simpler 7494, but the 74193 allows counting in either direction (up from 1, 2, ...16, 1, 2, ... or down 16, 15, ... 1, 16, 15.....). With the control arrangement shown, the counter runs up to a value set by the reset stage switch, then goes back down to 1 and starts up again and so on. It is easy to add switches to give up only, down only, or up-down. The outputs of the 74154 are all high except for the one that is addressed. Thus to use this as a bucket brigade, inverters should be added to the outputs. This is also a good idea because it eliminates loading effects of the control circuitry from reaching the voltage output. Open collector inverters can be used to advantage.



BASIC STRUCTURE OF UP-DOWN SEQUENCER

Another type of bucket brigade sequencer is one formed by cascading monostable circuits so that one goes high, and when it falls it triggers the next one high, and so on. This sequencer has two unusual properties. First, the times can be set independently and on a continuous scale, and the time will be the same when the same stage is reached again.



CASCADED MONOSTABLE SEQUENCER STAGE

To do this with a digital counter would require synchronization of the counter and the clock, which is possible, but difficult. The second property is that two sequences can exist on the line chasing each other but in general, never meeting. This does not mean that the device operates as two sequencers, but that a higher level of complexity can be achieved with more correlation than would occur by just adding the outputs of two sequencers It is also possible to add a flip-flop to the line so that the two chasing sequences will be treated in a different manner. This will likely cause them to collide and one will be annihilated. The basic sequencer stage used for the cascaded monostable sequencer is shown at the bottom of the previous page (see also EN#35).

THE SAMPLE-AND-HOLD AS A SEQUENCER

A standard sample-and-hold module can be used to produce sequences. The S&H is used to sample some waveform in response to a clocking command. It thus produces a sequence of voltages that depend on the input waveform and the time at which the sampling commands arrive. Thus, a moderate variety of sequences can be obtained. Popular choices for the waveforms used as the input and trigger are shown below:

INPUT	SAMPLE COMMAND	TYPE OF SEQUENCE
Periodic	Periodic(integer multiple of input)	Periodic
Periodic	Periodic	Patterned (long term periodic)
White Noise	Periodic	Randon voltages, regular time
Periodic	White noise (with higher threshold)	Random time, but voltage is predictable once time is determined
White noise	White Noise (with higher threshold)	Random voltage, Random Time

The reason for demanding a higher threshold when white noise is used as a sampling command is that this means that the sampling rate will be lower. If the white noise were used directly, the sampling rate would be too fast (see 2d (6)).

DIGITAL SEQUENCERS

An "all digital" sequencer is basically a digital memory which can be programmed in various ways and recalled later. The memory generally stores two parameters, pitch and time. Pitch can be stored by just 6 or 7 bits if only an equally-tempered scale is demanded (as with devices programmed by reading a standard keyboard). More bits are needed (perhaps 12) if pitches are to be close enough that the ear hears them as continuous. Time is likewise quantized, hopefully to the point where the quantization is not detectable. Two approaches are possible for the storage of time. Each of the time intervals can be filled with a pitch value, or the total length of time that a pitch exists can be recorded as a digital word giving the number of basic intervals consumed. Fullmer (EN#54) indicates that the latter approach is much more efficient in the general case.

Such a sequencer can be programmed in various ways. Having the device simply read the keyboard as it is played is a useful technique and one that is easy to learn. In addition, it is useful if the sequence can be later edited to make minor changes if this is necessary.

It is possible that in the near future digital sequencers will be available as functions that can be obtained by properly programming a small computer. This is a useful approach as computer can be used for a variety of other purposes as well.

CHAPTER 8A

OVERALL DESIGN CONSIDERATIONS

CONTENTS:

INTRODUCTION

In chapters 5a through 7c, we have been concerned with the design of the smaller units of larger systems. For the most part, these have been electrical circuits that were organized into functional modules. At some point, all these have to be integrated into an overall larger system - even if this is only a matter of choosing how many of each particular module that are to be built and installed.

In fact, the overall design of the system may be the most difficult, and the one requiring the most input from musicians. Perfectly designed modules that are not used to make music are of no use.

There are three levels of overall design that the musical engineer is involved with. The first is the integration of electrical components to perform certain functions. This is the general process of modular design, particularly for the more complex modules. The second level is the selection of component modules to form a viable music system, not just a sound synthesis system. The final level is the setup of the system to perform a certain musical function. Typically this might occur when the engineer is called back to help the musician set up a patch on the system. All of these may have a very similar procedural base. The final degree of complexity is comparable in all three cases. The difference of level occurs because once the lower design is accomplished, it can be ignored except for the final result. The design of a VCO may have been trying, but once it is done, the engineer then has to take the waveforms for granted and work them into a musical system.

DESIGN STEPS

STEP 1: The first step in design, as with most things, is to define the problem. For musical engineering, this involves two things: the musical requirements, and an evaluation of these requirements in terms of what is known about musical engineering. For example, a musician wants an instrument to produce a just tuned scale. Producing the instrument is the problem, but the engineering evaluation may show that different approaches are possible in terms of the accuracy with which the ear can detect pitch and the way the instrument is to be used.

STEP 2: The second step is a literature search (or the equivalent gathering of information). Has the problem been solved? What is known about the problem? Can the problem be solved at all? In the search, has anything been learned that causes us to go back to step 1 for a reaccessment?

STEP 3: A trial design is the next step. This can be on paper, computer simulation, actual circuit, or whatever is appropriate. Assuming this is successful, we are led to the next step.

STEP 4: Step 4 is construction and testing for basic function (e.g., the VCO is built and it does oscillate). This is the same as any field of engineering.

STEP 5: Step 5 is a musical evaluation, a fairly difficult step. (e.g., the VCO does or does not track accurately enough for musical use.) Also, a device that meets the requirements of one musician may not meet those of another.

STEP 6: The final step, assuming all others are successful, is to add refinements. This is frequently necessary with musical engineering, mainly because it is most difficult to access the device before it is built. For this reason, the musical engineer should show an excess of foresight. In practical terms, this may be a matter of leaving an extra op-amp on the circuit board, or drilling an extra hole in the panel in the event that something extra will be required.

SELECTING THE NUMBER OF MODULES

The selection of the number of modules and exact types can be very difficult. Some users know exactly what they need - other have no idea. Probably the most important thing is to try a few systems and see what seems useful, and then be prepared to expand an individual system as is necessary rather than getting it exactly right the first time.

As a general rule, you never have too many VCO's or VCA's, and usually not too many envelope generators. VCF's are more specifically oriented toward subtractive synthesis. Other devices depend on the user. It is probably well to have one of each of the common special devices (balanced modulators, sample-and-hold, noise sources) to see how they work.

As a rough guide, we show at the top of the next page some typical requirements. The basic system would be for the builder just starting out - what he really needs to get a feel for what is going on. The medium user system is for the serious experimenter and the individual composer. The large studio system is for the academic studio or very serious (and financially successful!) composer. These figures could well be altered by the use of computers in the system. Very soon, computers may be practical at the medium user level.

Musical engineers persuing special lines of research will have special needs and may have to overbuild one part of the system. This may make it less suitable for the typical musician.

SUGGESTED NUMBER OF MODULES

Module	Basic System	Medium User	Large Studio
VCO	2	4 - 6	7 - 10
VCA	1	2 - 3	4 - 5
VCF	1	2 - 3	4 - 6
Envelope Generators	1 - 2	3 - 6	6 - 10
Multipliers	1	2	4
Sample and Hold	1	2	4
Noise Source	1	2	4
Controllers	1	2	3 - 6
Frequency Shifters	0	0 - 1	1
Pitch Follower	0	0 - 1	1
Envelope Follower	0	0 - 1	1
Analog Delay Lines	0	1	1 - 3
Filter Banks	0	1	2 - 5
Sequencers	0 - 1	1 - 2	2 - 5
Computers	0	0 - 1	1

* * * * * * * * * * * * * * * * * *

BREAKING DESIGN DEADLOCKS

The designer who has never found himself up against a brick wall as fars as making progress with his design is indeed lucky. This will typically occur at the lowest level (modular design) and the highest (setup for specific musical purposes). At the middle level, he can always use a table such as the one above, or at least a decision can be made and work can go on. The real problem is when there seems to be no place to turn. The following suggestions may help:

1) Apply the principles of troubleshooting, starting with the (false?) premise that the thing should work, and that there is only some incidental fault. This should help pinpoint the problem.

2) Do an additional literature search - even from a seemingly unrelated field or from a different point of view.

3) Get another view from a person who is capable of solving problems on the same general level that you are, and who is in a position to appreciate your problem. They may pick up something you missed. Also, such a person may make what seems to be a silly suggestion (and it may be in fact silly) but in the process of telling him why it won't work, the original designer may gain insight into the problem.

4) Write the project up. Write down what works, what does not work, and why certain approaches are not possible. Write this as though you were defending your design. Write it in detail. This is helpful for making sure you understand everything you have done so far. Often I have been writing up projects for EN and find myself headed back to the test bence kicking myself for not doing something the easy way.

DOCUMENTATION

Documentation can be as important to a design as it is to a finished project. The individual designer should not be fooled into thinking he knows enough about his design that he will never forget it.

Once a basic design is worked out, a complete sketch should be made. If this is not done, certain parts of the circuit will probably be left out and forgotten until it gets to the testing stage. These diagrams need not be particularly neat, but they should be complete. Instead of redrawing circuits from published projects, a reference or a photocopy will serve. If the designer is not familiar with the published circuit however, redrawing the circuit may be very useful.

Another question that comes up is in regard to the format for a set of working plans. Such things as circuit board sketches and wiring lists may be useful if the builder is someone unfamiliar with the design. However, the original designer is most likely familiar enough with his design that he can work from the schematic. Some builders working from schematics like to go over the schematic with a colored marker as they put in connections. Working diagrams to be used by the system designer in his construction should use the terms that are most familiar to him in terms of the electrical functions of components. Thus, it is more useful to denote something as the "Q" output of a flip-flop rather than pin 15 of IC-5 for example. Working diagrams should give both where this is necessary. On the other hand, many builders know the base diagram of a typical op-amp inside out and the pin numbers just confuse the diagram.

Assuming the design goes well and everything works, the next step in documentation depends on what is to become of the actual device. If the designer is going to use it himself, it may be enough for him to just take a few minutes to go over his working diagrams and make a few notes. If it is to be turned over to another user, it is not absolutely necessary that the diagrams be any neater, but any individual conventions used in the documentation should be spelled out. Projects for publication are a special problem. The diagrams will probably have to be made neater, and some work will have to be done on description. Specific problems that will come up will probably be in the notation of parts. There will likely be several parts in the design that are non-standard. They may have come off a surplus board or be a precision part that just happened to be used even though it was not necessary. If the part is listed as it really is, some builders may waste months trying to find it. Here the designer will have to use his own judgement as to whether other parts will obviously work, or if they must be tried.

Probably the most important thing about documenting a design is to realize that accuracy and completeness are more important than neat drawings and typewritten descriptions. A component notation scribbled in and a note scratched in even poor handwriting is infinitely more useful than one that never got in because the designer did not have time to redraw the circuit or have the page retyped. Too many essential changes are never noted because the designer intended to do it in a "professional" manner later.

CHAPTER 8B

CONSTRUCTION

CONTENTS:

INTRODUCTION

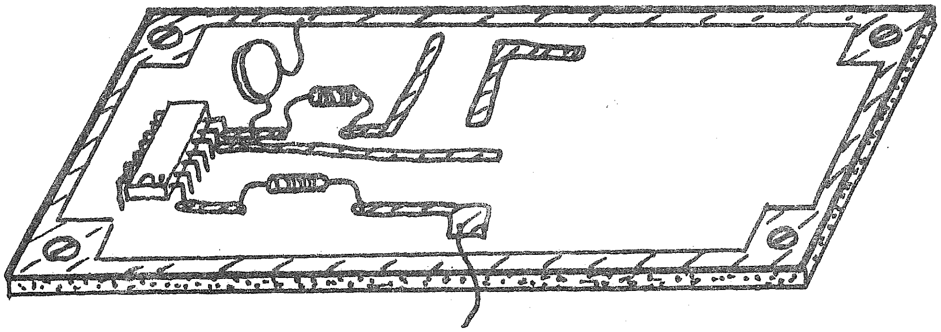
There are surely as many individual construction procedures as there are individual builders. The technique described below is a "topside" technique, all components are mounted on the top (foil) side of the board. No drilling is used. The outstanding advantage of this is that everything is in plain sight. The major disadvantage is that things are not as compact as they could be with construction on both sides. It will not be suitable for all requirements, but for the individual one-time construction of single modules, it is very useful. It is cheap. It is fast. It is mechanically sound and allows easy alteration.

PREPARATION OF CIRCUIT BOARDS

The first step is to select the proper sized piece of blank circuit board. It is useful to roughly lay out the IC's and any large components on the board to get some idea as to whether or not the board is the proper size. Next, the board should be cleaned of tarnish. Steel wool and scouring powder will do this. Complete the cleaning with circular motion to make sure that all the scratches are not lined up.

Next, consult the schematic diagram and any sketches you have made and lay out the IC's on the board. Once the position is correct, use a pencil to mark the positions of the IC leads. Then you are ready to begin painting the board.

The paint to be used should be some sort of fast drying lacquer and should be thinned with lacquer thinner so that the cover is very light. The brush used should be a very fine artist type, and it should not be clean. Old paint should be allowed to dry in it to stiffen it up a little. This will make painting with



GENERAL IDEA OF THE TOPSIDE CONSTRUCTION METHOD

it more like writing with one of those felt marker pens. You then start to paint the board and do your design as you go. Do this with some care as if you accidently make a connection that is not correct, it is difficult to wipe the paint clean enough so that it will not interfere with the clean etching out of the accidental connection. If you do make such a mistake, it is best to let it dry, and scratch the paint off with a sharp object. It is quite likely that you will at some point "paint yourself into a corner" and find that some necessary connection can't be made with the paint. This is the point where you must accept the fact that a wire jumper or two will be necessary. Once the board wiring pattern is painted, you will want to enlarge some of the ground areas so that you will have some place to drill the board for mounting.

When the paint is dry, the board is ready for etching. A useful technique is to float the board on top of the Ferric Chloride etchant. This type of etchant is available from electronic supply outlets or hobby type electronic stores. It can be poared into one of those polyethylene food storage boxes and covered when not in use. Floating the board is not too difficult if it is set down carefully. However, the board should not be set down perfectly flat, as this will almost certainly cause bubbles to be trapped underneath, and you will have to etch twice. Instead, set the board down at a very slight angle so that the etchant flows along the lower surface smoothly. Then level the board off and release it, and it should float. Leave it there for about 20 minutes, then remove it and rinse it off. If you find any areas of incomplete etching, put it back again. When the etching is complete, rince the board thoroughly.

To remove the paint, use the lacquer thinner. Drop on a few drops. If this seems to dissolve the paint completely, hold the board over your paint can and let drops of the thinner run off the board into the paint can. This serves three purposes: (1) it cleans the board, (2) it recycles your paint, and (3) the thinner is returned to the paint and keeps it easy to use the next time.

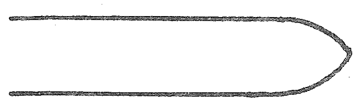
When most of the paint is off the board, return to the steel wool and scouring powder and clean the board again. Rince the board off thoroughly and then dry it without delay.

The board is now ready to use. Several preliminary steps may be considered first however. First, any holes should be drilled. Secondly, you may want to coat the copper traces with solder. This is a simple manner of just melting a thin cover of solder on the board. It makes it easier to solder components to, prevents tarnish, and will often repair any hair line cracks in the copper traces. If you get too much solder on, turn the board over and remove it by allowing the solder to run down the soldering iron by gravity. You may want to leave a heavier coating of solder on any thin grounding traces to lower the resistance.

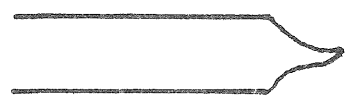
SOLDERING AND SOLDERING IRONS

This is a good point to talk about soldering and soldering irons, and this really applies to any construction technique. Buy a good grade of rosin core solder - do not skimp on what you pay for solder, as if you do you will pay for it many times over in frustration and lost time. I recommend "ERSIN Multicore" 60% tin, 40% lead in 1/16" diameter or less. Kester's "Resin 5" in the same small diameter is a good second choice.

For printed circuit work or other construction techniques, you need a small low wattage "pencil" type soldering iron. The inexpensive "Ungar" irons are a good choice. These irons consist of a handle, a heating element, and a tip which screws into the end. Select a heating element in the range of 27 watts for IC work. Select a copper tip with a conical end, not a wedge end. Higher wattage elements are not necessary and will likely just burn up the tips faster, and burn off the rosin flux in the solder before it has a chance to do its work. IC's can generally take a lot of heat, so it is not necessary to use a low wattage element for that reason, but the point is that if you can't easily work with a 27 watt element, you are probably using too much solder. The soldering iron tips do burn away, and that's the reason they do screw out and are replaceable. When they start to burn away, the do not do so uniformly, but form pits, rough spots and even hooks on the end. It is important to realize that the best sort of tip is one that is rounded as shown below:



Properly Filed Tip



Badly Burned Concave - Needs Filing

With this type of tip, solder tends to run off the tip and onto the connection. As the tip burns away, it usually becomes concave, and solder tends to stick to the iron. It is thus important to frequently (at least every hour) remove the tip, file it round again, and retin it. It is not necessary to cool the iron to do this. The tip can be removed with a pair of pliers and it will cool rapidly. Also, it is a good idea to loosen the tip and retighten it from time to time between reshaping. This assures that it will not set up and become difficult to remove, and also breaks up any oxide layer that developes that makes heat transfer to the tip incomplete.

WIRING OF CIRCUIT BOARDS

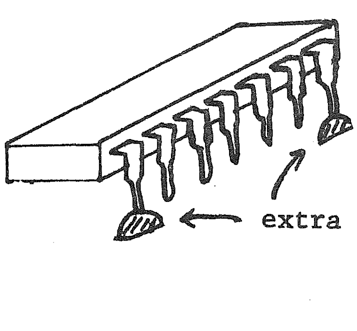
Now, returning to the board you have prepared, the next step is to mount all the IC's except for any CMOS which should be mounted after everything else. To mount an 8-pin "mini-DIP" IC, you add a little extra solder to the board at the places where the IC leads will come in contact with the board. Then hold the IC in position and draw the soldering iron tip along the leads at the point where they contact the board. This should solder all the leads on one side at one time. Removing the IC later should it become necessary is just about as simple. The story is a little different with 14-pin of 16-pin IC's as there are more leads on one side and it is hard to heat them all rapidly enough so that the solder is in a liquid state on all of them at one time. Repeating the procedure outlined for the 8-pin IC will work for mounting, but removal is difficult. Thus for 14 or 16 pin IC's, I recommend mounting them a little above the board. To do this, apply small mounds of solder to the board under the corner pins. The IC is then placed above these mounds and the corners are soldered without forcing the leads down through the solder to contact the board. The other leads can be soldered by allowing solder to flow between the soldering iron and the IC pin and onto the circuit board trace. The IC is thus mounted a little above the board. Check this carefully for good connections. The advantage of this comes when it is necessary to remove the IC (and sometime, you will have to remove such an IC). The board is held so that solder will run away from the lead and onto the iron. Done carefully, you can lift the IC off. The clearance of the IC above the board should be only about 1/32 of an inch.

Next, the PC board wiring can be completed. The two most important tools here are clips: alligator, and nail. The ordinary nail clipper is an excellent wire cutter for small components. The alligator clip serves to hold small components while soldering. (Hold a 1/4 watt resistor with 1/4 inch leads in your fingers while soldering and you will see it gets too hot to hold.) The alligator clip can be used to hold the component body, or its leads, and is also very useful for removing components. Choose an alligator clip that has jaws that don't quite match properly, and this may also be an excellent wire stripper as well.

If you have already tinned the entire board, components can be mounted by bending and cutting the leads to the proper length, tinning the ends of the leads, and then simply touching the leads to the board and heating briefly with the soldering iron. Since you only have two hands, and you have three things to hold (component, solder, and soldering iron), you have to develop a system. Fortunately, all electrical components have at least two leads. This means that you can just "tack-solder" one of the leads to hold the component in place. This is the purpose of tinning both the board and the leads of the component, as there will be enough solder on the two to hold. Once tacked in place, you can release the component and pick up the solder. Reheat the connection and add a little solder to complete the connection on one side. When this hardens, reheat and add solder to the other lead.

In many cases, it is necessary to run wires from points on the circuit board to the panel controls. In such cases, tinning becomes very important as you will not be able to rely on the other end of the wire to hold the first end in place. It is therefore a good idea to leave wider traces or provide soldering "pads" at the points where wires must be connected.

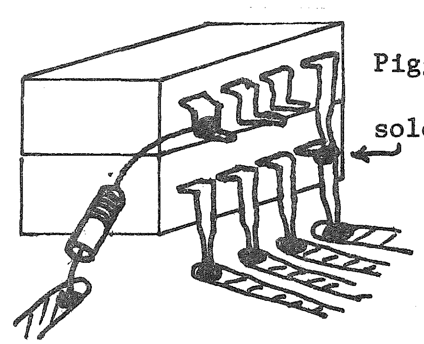
A very useful extra "component" is the wire jumper which you use to take care of the connections that you couldn't make with circuit board traces. If these jumpers are long, use insulated wire. For short ones, you have all the short jumpers you need all over your bench - the component lead trimmings. These trimmings can also be used for other purposes as well. Connect these at one end at various test points and the wire sticking up is a useful test terminal that can be removed later if desired. It is also an excellent idea to put in some intentional breaks in some of the traces and bridge these with wire jumpers when you mount the other components. The purpose of this is to open up the line later if this is necessary for troubleshooting. They thus form breakable test points.



Elevated Mount for

14 pin or 16 pin DIP

extra solder elevates corners



Piggy-Back Mount

solder common pin

Sooner or later, you may have to make use of the "piggy-back IC". If you forget to leave room for an op-amp somewhere, or later need to add one for some other purpose, you may be able to mount it above an existing IC. The basic idea is to bend up all leads of the IC that have different connections from the lower IC. There will usually be at least two leads in common (the power supply connections). These are soldered to the lower IC, and components are soldered between board traces and the IC leads. For some reason, this often works out very well if the existing IC position is carefully chosen.

PANEL PREPARATION AND WIRING

The choice of panel type, material, and the panel layout is a matter of personal choice and design. Unlike other electronic circuits, electronic music modules will in general have a very large number of controls and will require a good deal of panel space relative to the number of components on the circuit board. Correspondingly, there will be a good deal of panel wiring to do. In fact, it may take as long to wire the panel as it did to do the actual circuit board.

The recommended panel material is aluminum, 1/8 inch thick or a little thicker. It is suggested that enough planning be done so that the entire panel can be drilled at one time. This is much easier, and avoids the danger that later drill turnings may fall into completed circuitry and cause shorts. Once all the holes are drilled, all loose drill turnings should be broken free. It is not necessary to remove hole rims resulting from the drilling or to countersink the holes. In fact, a little burr may improve the electrical contact between the panel and the mounting hardware. Bear in mind that most of the necessary grounds are available as contacts to the panel.

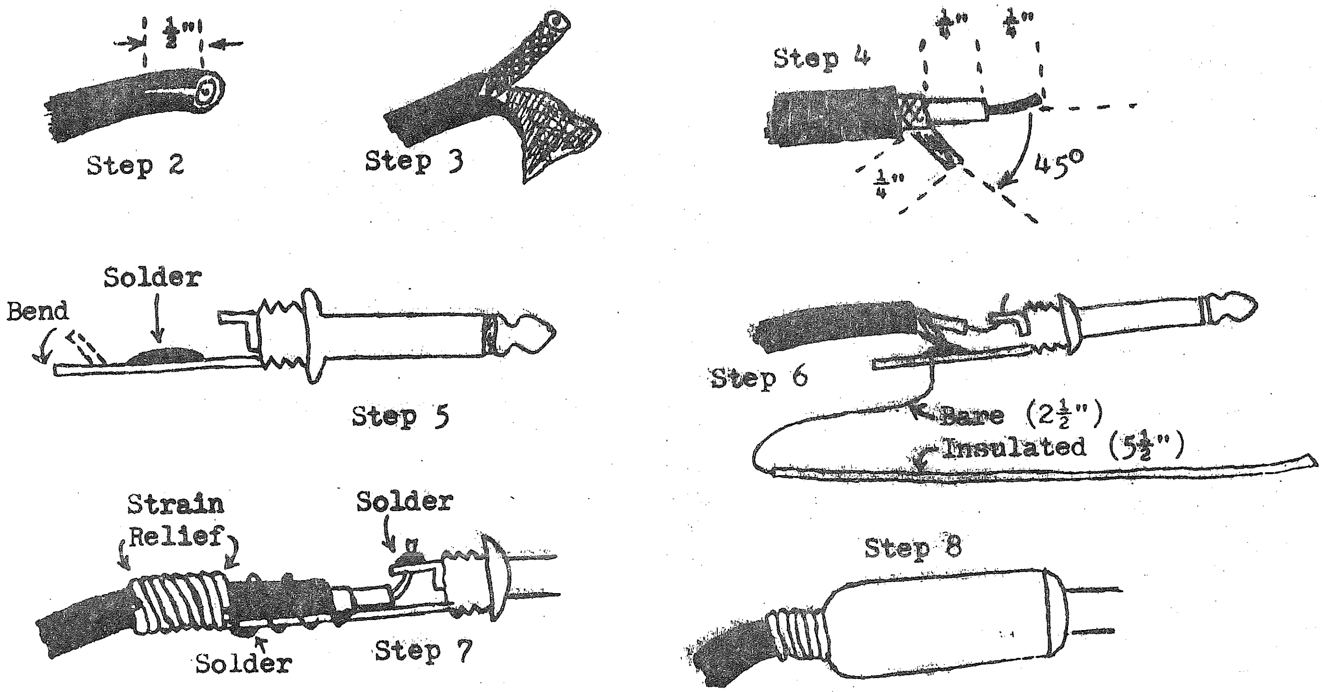
The next step is to do all the wiring on the panel. This should be done neatly and as simply as possible, and with solid rather than stranded hookup wire. This means that all on-panel wires can be kept out of the way when it comes to wiring between the panel and the circuit board. Finally, the wiring between the panel and the circuit board can be done. This should be done with stranded wire unless the panel and the circuit board are and will always be mechanically fixed relative to each other. Keep these wires to a minimum. It is often the case that a single point on the circuit board runs to several panel points. There should be only one wire from the board to the panel in this case, and the other connections are on-panel connections. When there are two wires which may be interchanged, run them as a twisted pair. This will be the case for example when wires run to a pot as a variable resistor. [The easy way to make twisted pairs is to cut two long lengths of wire and string them across the room, attaching the far end securely. Stretch them parallel and place the other ends in an electric drill. Press the trigger and you get neat and easy twisted pairs.]

LIVING WITH PATCH CORDS

Probably the most reliable patching system is the one employing standard 1/4" phone plugs and phone jacks. The best way to obtain reliable patch cords is to buy the cables with the plugs molded on. However, for a number of reasons (expense, need for a custom item, etc.) you may want to make up your own. Also, most studios have cables around that are in need of repair. The following is a technique for attaching plugs to cables that works very well. (see diagrams on next page).

1) First of all, take the phone plug apart and slide the plastic handle over the cable so that you don't forget it later.

2) Next, use a sharp knife to put a lengthwise slit in the outer insulation for the last 1/2 inch of cable. It is hard to avoid cutting the shield wires while doing this, so cut them too if necessary, it won't matter later.



3) Pull back the outer insulation and keep on pulling, <u>tearing</u> the outer insulation another 3/8 inch beyond the end of the slit. Cut off the pulled back insulation, being careful not to cut the shield wires.

4) Unravel the outer shield wires. Many of the small wires you cut while making the initial slit will fall away. Twist up the remaining shield, tin with solder, and trim to 1/4 inch. Carefully expose 1/4 inch of the center conductor, leaving 1/4 inch of the inside insulation. Twist and tin the inside conductor too.

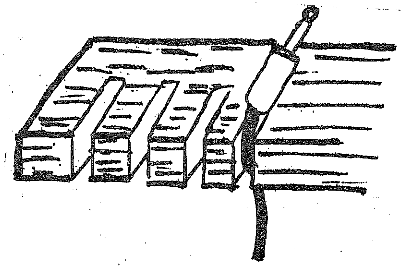
5) Remove the screws (if any) from the phone plug tabs. Bend down the larger tab as shown. Put a blob of solder along the length of this tab as shown.

6) Prepare a 8 inch length of #22 solid hookup wire as shown, removing about 2-1/2 inches of insulation. Push the center conductor of the cable up through the hole in the smaller tab (but do not solder yet). Now solder the shield to the solder blob, and also solder the bare end of the #22 wire slightly to the rear of this point as shown.

7) Wrap the bare wire as tightly as you can around the cable and the larger tab. Plan this so that the first loop is clear of the center conductor, and so that the last bare loop just passes over the empty screw hole of the larger tab. Continue winding the insulated part of the wire down the cable to form a sort of "strain relief." Now, the center conductor has been pulled down through the hole in the smaller tab, and can be soldered as no more tension will be applied to it. Also, solder the bare wire at the empty screw position just enough to hold it. Do not solder this further up as the heat may melt the center insulator and cause a short.

8) Bring up the plastic handle, and slide it over the strain relief. If it is not fairly tight, back it off and wrap the strain relief with a turn or two of plastic electricians tape, and repeat if necessary for a tight fit.

A couple of other hints: 1) When making up cables, a good way to hold the plug is to insert it into a jack. You can use any handy jack, or mount one on your bench for this purpose. 2) Storing patch cords is always a problem. One way which a lot of people use is shown at the right. You simply take a length of 1×4 or 1×6 lumber and make saw slits in it. Then knock out the wood inbetween with a screwdriver blade. The lumber is then mounted so that the cables can hang down.



CHAPTER 8c

TROUBLESHOOTING

CONTENTS:

INTRODUCTION

When is troubleshooting necessary? Quite simply, whenever something is not working properly. This may be when a new module design fails to work as predicted, when one that was working quits, when a patch out of a book does not work when implemented, or when a new idea fails to give the expected results.

Troubleshooting is the process of finding out what is wrong - not the process of correcting the trouble. In many cases, correcting the problem once it is found is quite trivial. In other cases, there is no way of fixing it - the idea just does not work out.

The process of troubleshooting is basically the same in all cases, and the lessons learned troubleshooting a car for example may be useful for electronics. Lessons learned troubleshooting a patch may help in troubleshooting a module. Efficient troubleshooting is essential for efficient operations. In some cases (e.g., when putting on a demonstration), efficient troubleshooting is essential to keeping one's sanity!

CHECKING THE OBVIOUS

The obvious things to check start of course with the proverbial "is it plugged in?" and go on through such things as wiring errors, missing parts, poor soldering, solder bridges (shorts), hot IC's, and other basics. A good check list might be:

- /__/ Are any IC's hot (really hot, not just warm)? If so, turn the power off and skip the power checks below.
- /__/ Are the power supply voltages correct (or is the supply shorted)?
- /__/ Do all the IC's have the correct power supply voltages at their supply pins (checked above the solder)?
- /__/ Check all the ground points for a good ground. A good way to do this is to touch the scope probe with your finger causing 60 Hz hum to be displayed on the screen. Touching this to a ground should make the hum disappear.
- /__/ Check for wiring errors. Do this thoroughly and methodically. If possible, have someone else check your wiring, or check your own as though it were someone else's.
- /__/ Check for a bad IC if this is fairly easy. However, relatively few problems will be caused by bad IC's unless you got as "unbelievable" surplus "bargain."
- /__/ Check any suspicious solder connections.
- /__/ Do you really understand how the device is supposed to be used. Does it need some input or control signal that you are not giving it? Is some control set wrong?
- /__/ Were corrections to the circuit published later?

LEARNING WHAT SHOULD BE HAPPENING

The most important thing about troubleshooting is to know how the circuit is supposed to work. Few of us can resist the temptation of just building the circuit and trying it. Once in a while, it works. Often it does not, and after you check the obvious with no improvement, you have to face the fact that it is now necessary to know how the thing is supposed to work.

This means reading circuit descriptions of just figuring it out, but it has to be done. In a larger sense, learning the electronics of the circuit won't do you any harm.

ISOLATING THE TROUBLE

This is a vary natural approach that you use everyday, but it can usually be done more systematically, and there are ways you can be fooled.

You probably tested the device by checking for an output, and this was in some way unsatisfactory. But if for example, it was the sine wave output that you checked, and there was no signal at all, you won't start troubleshooting the sine wave shaper since you know this needs a triangle wave to drive it. You check for the triangle at the input of the sine shaper - because it probably isn't there.

Eventually, you trace the circuit back, either step by step, or by checking a few likely and important points. Hopefully, you will find a place where the circuit is working properly, and this should pretty well isolate the problem. If no working area is found, it is probably the case that something is locked up, and all voltages remain constant.

Such a lock-up is typically caused by an op-amp output that is at + or - the supply voltage while it should be a signal balanced around ground (due to negative feedback). If you locate this op-amp. it will probably be the case that the differential input voltage is not zero (because it is a good op-amp and its output is pinned). Then the problem is finding out why the differential input is not zero, why the negative feedback loop has failed.

There are ways in which the isolation method can fail or fool you. A very common case is where there is an overall feedback loop and everything around the loop is malfunctioning (naturally). If possible, you have to break the loop and operate the circuit or parts of it open loop until the real fault is found. In any case, don't be fooled into thinking that because the first stage in the loop is not working properly that this is the offending stage. There is no real starting point in such a loop. If you can't break the loop, you have to troubleshoot "on the run."

SOME TROUBLESHOOTING RULES

Don't look for complicated explanations - just simple ones. Bad parts are rare unless you have done something wrong in the wiring that will destroy a good part. Changing parts can take a lot of time. Construction errors are much more probable.

Don't hesitate to make the effort necessary for a crucial test. If you know of a sure test, even if it will take quite a bit of effort and time, do it.

By all means take frequent breaks away from the actual troubleshooting. Sit down with a pencil and paper and try to figure out what is going on and a few new tests to make. Take the time to think the thing out rather than just probe around at random.

If you have a fault in one part of a complex system, and it occurs just after you made a change in another part of the system, even if it seems impossible, there is a 99% chance that the fault is due in some way to the change you made. This is a hard lesson to learn, except by encountering it yourself frequently.

Count the IC's in the circuit. If there are 5 or fewer, the circuit should work the first time or should be easy to fix. If there are more than 10, there will likely be one or more malfunctions. A good bet might be one fault for each 10 IC's if you made the whole circuit yourself from scratch.

CHECKING CERTAIN MODULES

VCO's: VCO's generally have no loops except in the main oscillator. If there is no output at any of the waveshapers, the main oscillator is probably not working. Once this is isolated, it may be best to isolate the current source to the oscillator from the main oscillator. This may be a simple matter of using a substitute current source (such as a single resistor to one of the power supplies). This should isolate the trouble in either the current source, the main oscillator, or in one of the waveshapers.

VCA's: VCA's are generally quite simple and it is unlikely that anything except an obvious fault will occur. If it does, it may be the lock-up of some op-amp output (due to lack of feedback, typically a transistor in the feedback loop will be malfunctioning). Another possible problem is an overdriven input.

VCF's: A VCF like the state variable is a complete loop. If there is a break in the loop, the whole loop will likely lock up with various points pinned at one supply or the other. Voltage-controlled Q stages may malfunction and lock up the loop. Other types of VCF's are usually not as subject to loop problems (because they do not rely on the loop for DC stability). Such filters may have faults in one stage that propogate to the output, but this is usually easy to check, particularly if the first stage is working and can serve as an example. The current source may be isolated as in the VCO.

Envelope Generators: The "AR" part of an envelope generator is likely an open loop and is easy to troubleshoot. The "AD" part is a closed loop, and the analysis is complicated because there are generally comparators in the loop that work at power supply voltages anyway. One useful trick is to trigger the stage from an oscillator so that both hands can be used for troubleshooting.

TROUBLESHOOTING DESIGNS AND PATCHES

Most new designs start with many familiar building blocks that are more or less standard circuits and relatively fool-proof. The first step in troubleshooting a new design is to be sure these are working properly. In troubleshooting a patch, the user must be certain all the modules (and the patch cords) are working. It would be a shame to have a good design discarded because an inverting amplifier was not working, or a good new patch ruled out because a patch cord was open.

Beyond this, the designer can troubleshoot various combinations of building blocks to make sure that they work together. If the entire system still fails he has to work with various contrived test conditions and/or a theoretical analysis to see if there is a correctable problem of if the idea must really be discarded.

SPECIAL TECHNIQUES AND PROBLEMS

One advantage of modular systems is that you often have more than one of the same circuit. If one of them works while the other does not, you have in the working model the perfect tool to troubleshoot the non-working one. This is extremely useful for isolating sections that are properly working. On the other hand, if a non-working section is found, this does not in itself indicate that the fault lies there. There might be a fault in some other stage that is causing the trouble. The entire thing must be checked.

One special and all too familiar problem is the "intermittent." This is the sort of thing that works some of the time only. If it can be made to work or not work according to some action that the user can control (for example, if the circuit board is pressed in one point it works), then it should be quite easy to fix as the working and non-working conditions can be examined. If it can not be controlled, the troubleshooter may be in for a real test of his skill and a supreme test of his patience. Very likely, the unseen factor is heat. Heat can often be applied with a soldering iron, or with a heat gun or hair dryer, and this may make the fault appear. In any case, this sort of problem is one which has to be corrected as it is the most annoying during use.

CHAPTER 9A

MATHEMATICAL TABLES

CONTENTS:

INTRODUCTION

It is the purpose here to give useful math data that would be fairly difficult to find elsewhere. Available space limits the amount of data that can be included here, so many common and useful tables are not included because they are readily available in standard references, or are standard functions of pocket calculators.

Those needing additional data may consult the standards such as:

CRC Math Tables [Chemical Rubber Company, Yearly Editions]

M. Speigel, Mathematical Handbook of Formulas and Tables, Schaum's Outline Series, McGraw-Hill (1968)

M. Abramowitz & I. Stegun, Handbook of Mathematical Functions, Dover (1965)

CONSTANTS

π = 3.14159 26535 89793 23846 .....

e = 2.71828 18284 59045 23536 ..... (base of natural log)

$\log_e 2$ = 0.69314 71805 59945 30941 .....

$2^{1/12}$ = 1.05946 309.. ....

k_B = 8.6167×10^{-5} electron-volts/°K (Boltzmann's Constant) [at room temperature, approximately 300°K; k_BT is approximately 26 millivolts]

q = 1.6019×10^{-19} coulombs (Charge of the electron) [Note that q = 1 when k_B is given in eV/°K as it is above]

CONVERSION FACTORS

1 radian = 57.29577 95131 ..... degrees

1 foot = 0.3048 meters

1 ounce = 0.02834952 kilograms

°Fahrenheit = 32 + (9/5)°Celsius [°Celsius = °Centegrade = °Kelvin - 273]

FRACTIONAL ROOTS OF TWO ($2^{1/n}$ for n = 1 to 60)

n	$2^{1/n}$	n	$2^{1/n}$	n	$2^{1/n}$
1	2.00000000	21	1.03355778	41	1.01704974
2	1.41421356	22	1.03200827	42	1.01664043
3	1.25992104	23	1.03059554	43	1.01625032
4	1.18920711	24	1.02930223	44	1.01587808
5	1.14869835	25	1.02811382	45	1.01552251
6	1.12246204	26	1.02701805	46	1.01518251
7	1.10408951	27	1.02600448	47	1.01485709
8	1.09050773	28	1.02506421	48	1.01454533
9	1.08005973	29	1.02418956	49	1.01424638
10	1.07177346	30	1.02337389	50	1.01395947
11	1.06504108	31	1.02261143	51	1.01368390
12	1.05946309	32	1.02189714	52	1.01341899
13	1.05476607	33	1.02122660	53	1.01316414
14	1.05075663	34	1.02059590	54	1.01291879
15	1.04729412	35	1.02000160	55	1.01268242
16	1.04427378	36	1.01944064	56	1.01245454
17	1.04161601	37	1.01891028	57	1.01223471
18	1.03925922	38	1.01840809	58	1.01202250
19	1.03715504	39	1.01793188	59	1.01181753
20	1.03526492	40	1.01747969	60	1.01161944

RULES FOR LOGS AND EXPONENTIALS

$x^a \cdot x^b = x^{a+b}$ $\qquad x^a/x^b = x^{a-b}$ $\qquad (x^a)^b = x^{ab}$

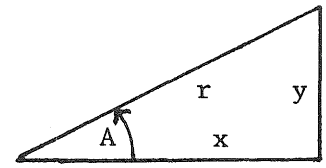
$\mathrm{Log}_a(xy) = \mathrm{Log}_a(x) + \mathrm{Log}_a(y)$ $\qquad \mathrm{Log}_a(x/y) = \mathrm{Log}_a(x) - \mathrm{Log}_a(y)$

$\mathrm{Log}_a x^p = p\,\mathrm{Log}_a x$ $\qquad \mathrm{Log}_a x = \mathrm{Log}_b x/\mathrm{Log}_b a$

$\mathrm{Log}_e x = 2.3026\,\mathrm{Log}_{10} x$ $\qquad \mathrm{Log}_{10} x = 0.43429\,\mathrm{Log}_e x$

$\mathrm{Log}_2 x = 1.442695\,\mathrm{Log}_e x$

TRIG IDENTITIES



Sin A = y/r Cos A = x/r Tan A = y/x

Csc A = r/y Sec A = r/x Cot A = x/y

by the Pythagorean theorem: $\sin^2 A + \cos^2 A = 1$

$\sin(A \pm B) = \sin A \cos B \pm \cos A \sin B$ $\cos(A \pm B) = \cos A \cos B \mp \sin A \sin B$

$\sin A + \sin B = 2 \sin\frac{A+B}{2}\cos\frac{A-B}{2}$ $\sin A - \sin B = 2 \cos\frac{A+B}{2}\cos\frac{A-B}{2}$

$\cos A + \cos B = 2 \cos\frac{A+B}{2}\cos\frac{A-B}{2}$ $\cos A - \cos B = 2 \sin\frac{A+B}{2}\sin\frac{B-A}{2}$

$\sin A \cdot \sin B = (1/2)[\cos(A-B) - \cos(A+B)]$

$\cos A \cdot \cos B = (1/2)[\cos(A-B) + \cos(A+B)]$

$\sin A \cdot \cos B = (1/2)[\sin(A-B) + \sin(A+B)]$

See also: Chapter 2b, pg. 4

RELATION BETWEEN EXPONENTIALS AND TRIG FUNCTIONS

Euler's Identity: $e^{j\theta} = \cos\theta + j\sin\theta$ where $j = \sqrt{-1}$

Thus: $\sin\theta = \dfrac{e^{j\theta} - e^{-j\theta}}{2j}$ $\cos\theta = \dfrac{e^{j\theta} + e^{-j\theta}}{2}$

RELATIONSHIPS WITH BESSEL FUNCTIONS

$\sin x = 2[J_1(x) - J_3(x) + J_5(x) - \cdots]$ $\cos x = J_o(x) - 2J_2(x) + 2J_4(x) - \cdots$

$1 = J_o(x) + 2J_2(x) + 2J_4(x) + \cdots$ $x = 2[J_1(x) + 3J_3(x) + 5J_5(x) + \cdots]$

$1 = J_o^2(x) + 2J_1^2(x) + 2J_2^2(x) + \cdots$ (conservation of energy, linear FM)

$\cos(x \sin y) = J_o(x) + 2[J_2(x)\cos(2y) + J_4(x)\cos(4y) + \cdots\cdots]$

$\sin(x \sin y) = 2[J_1(x)\sin(y) + J_3(x)\sin(3y) + J_5(x)\sin(5y) + \cdots\cdots]$

$\cos(x \cos y) = J_o(x) - 2[J_2(x)\cos(2y) - J_4(x)\cos(4y) + J_6(x)\mathrm{Cox}(6y) - \cdots\cdots]$

$\sin(x \cos y) = 2[J_1(x)\cos(y) - J_3(x)\cos(3y) + J_5(x)\cos(5y) - J_7(x)\cos(7y) + \cdots]$

$I_n(x) = j^{-n}J_n(jx)$ where j = square root of -1

TRIG FUNCTIONS (Angles in Degrees)

ANGLE	SINE	COSINE	TANGENT	ANGLE	SINE	COSINE	TANGENT
0	0.000	1.000	0.000				
1	0.018	1.000	0.018	46	0.719	0.695	1.036
2	0.035	0.999	0.035	47	0.731	0.682	1.072
3	0.052	0.999	0.052	48	0.743	0.669	1.111
4	0.070	0.998	0.070	49	0.755	0.656	1.150
5	0.087	0.996	0.088	50	0.766	0.643	1.192
6	0.105	0.995	0.105	51	0.777	0.629	1.235
7	0.122	0.993	0.123	52	0.788	0.616	1.280
8	0.139	0.990	0.141	53	0.799	0.602	1.327
9	0.156	0.988	0.158	54	0.809	0.588	1.376
10	0.174	0.985	0.176	55	0.819	0.574	1.428
11	0.191	0.982	0.194	56	0.829	0.559	1.483
12	0.208	0.978	0.213	57	0.839	0.545	1.540
13	0.225	0.974	0.231	58	0.848	0.530	1.600
14	0.242	0.970	0.249	59	0.857	0.515	1.664
15	0.259	0.966	0.268	60	0.866	0.500	1.732
16	0.276	0.961	0.287	61	0.875	0.485	1.804
17	0.292	0.956	0.306	62	0.883	0.470	1.881
18	0.309	0.951	0.325	63	0.891	0.454	1.963
19	0.326	0.946	0.344	64	0.899	0.438	2.050
20	0.342	0.940	0.364	65	0.906	0.423	2.145
21	0.358	0.934	0.384	66	0.914	0.407	2.246
22	0.375	0.927	0.404	67	0.921	0.391	2.356
23	0.391	0.921	0.425	68	0.927	0.375	2.475
24	0.407	0.914	0.445	69	0.934	0.358	2.605
25	0.423	0.906	0.466	70	0.940	0.342	2.747
26	0.438	0.899	0.488	71	0.946	0.326	2.904
27	0.454	0.891	0.510	72	0.951	0.309	3.078
28	0.470	0.883	0.532	73	0.956	0.292	3.271
29	0.485	0.875	0.554	74	0.961	0.276	3.487
30	0.500	0.866	0.577	75	0.966	0.259	3.732
31	0.515	0.857	0.601	76	0.970	0.242	4.011
32	0.530	0.848	0.625	77	0.974	0.225	4.331
33	0.545	0.839	0.649	78	0.978	0.208	4.705
34	0.559	0.829	0.675	79	0.982	0.191	5.145
35	0.574	0.819	0.700	80	0.985	0.174	5.671
36	0.588	0.809	0.727	81	0.988	0.156	6.314
37	0.602	0.799	0.754	82	0.990	0.139	7.115
38	0.616	0.788	0.781	83	0.993	0.122	8.144
39	0.629	0.777	0.810	84	0.995	0.105	9.514
40	0.643	0.766	0.839	85	0.996	0.087	11.430
41	0.656	0.755	0.869	86	0.998	0.070	14.301
42	0.669	0.743	0.900	87	0.999	0.052	19.081
43	0.682	0.731	0.933	88	0.999	0.035	28.636
44	0.695	0.719	0.966	89	1.000	0.018	57.291
45	0.707	0.707	1.000	90	1.000	0.000	∞

TABLES OF BESSEL FUNCTIONS

Tables of Bessel functionf and modified Bessel functions follow. The table of J_n ranges from n = 0 to 12 for argument 0 to 12. The table of modified bessel finctions is needed for FM calculations for exponential VCO's (chapter 2c). These are given in terms of $\text{Log}_2 x$ as the argument, so the tables are also useful for $\text{Log}_2 x$.

BESSEL FUNCTIONS OF THE FIRST KIND $J_n(x)$

The initial decimal points are not shown in the tables below. All numbers have magnitude lest than or = 1. For example, the fourth number under J_1 shown as 1483 is actually 0.1483

x	J_0	J_1	J_2	J_3	J_4	J_5	J_6	J_7	J_8	J_9	J_{10}	J_{11}	J_{12}
0.0	1.0	0	0	0	0	0	0	0	0	0	0	0	0
0.1	9975	0499	0012	0	0	0	0	0	0	0	0	0	0
0.2	9900	0995	0050	0002	0	0	0	0	0	0	0	0	0
0.3	9776	1483	0112	0006	0	0	0	0	0	0	0	0	0
0.4	9604	1960	0197	0013	0001	0	0	0	0	0	0	0	0
0.5	9385	2423	0306	0026	0002	0	0	0	0	0	0	0	0
0.6	9120	2867	0437	0044	0003	0	0	0	0	0	0	0	0
0.7	8812	3290	0588	0069	0006	0	0	0	0	0	0	0	0
0.8	8463	3688	0758	0102	0010	0001	0	0	0	0	0	0	0
0.9	8075	4059	0946	0144	0016	0001	0	0	0	0	0	0	0
1.0	7652	4401	1149	0196	0025	0002	0	0	0	0	0	0	0
1.1	7196	4709	1366	0257	0036	0004	0	0	0	0	0	0	0
1.2	6711	4983	1593	0329	0050	0006	0001	0	0	0	0	0	0
1.3	6201	5220	1830	0411	0068	0009	0001	0	0	0	0	0	0
1.4	5669	5419	2074	0505	0091	0013	0002	0	0	0	0	0	0
1.5	5118	5579	2321	0610	0118	0018	0002	0	0	0	0	0	0
1.6	4554	5699	2570	0725	0150	0025	0003	0	0	0	0	0	0
1.7	3980	5778	2817	0851	0188	0033	0005	0001	0	0	0	0	0
1.8	3400	5815	3061	0988	0232	0043	0007	0001	0	0	0	0	0
1.9	2818	5812	3299	1134	0283	0055	0009	0001	0	0	0	0	0
2.0	2239	5767	3528	1289	0340	0070	0012	0002	0	0	0	0	0
2.1	1666	5683	3746	1453	0405	0088	0016	0002	0	0	0	0	0
2.2	1104	5560	3951	1623	0476	0109	0021	0003	0	0	0	0	0
2.3	0555	5399	4139	1800	0556	0134	0027	0004	0001	0	0	0	0
2.4	0025	5202	4310	1981	0643	0162	0034	0006	0001	0	0	0	0
2.5	-0484	4971	4461	2166	0738	0195	0042	0008	0001	0	0	0	0
2.6	-0968	4708	4590	2353	0840	0232	0052	0010	0002	0	0	0	0
2.7	-1424	4416	4696	2540	0950	0274	0065	0013	0002	0	0	0	0
2.8	-1850	4097	4777	2727	1067	0321	0079	0016	0003	0	0	0	0
2.9	-2243	3754	4832	2911	1190	0373	0095	0020	0004	0001	0	0	0
3.0	-2601	3391	4861	3091	1320	0430	0114	0025	0005	0001	0	0	0
3.1	-2927	3009	4862	3264	1456	0493	0136	0031	0006	0001	0	0	0
3.2	-3202	2613	4835	3431	1597	0562	0160	0038	0008	0001	0	0	0
3.3	-3443	2207	4780	3588	1743	0637	0188	0047	0010	0002	0	0	0
3.4	-3643	1792	4697	3734	1892	0718	0219	0056	0012	0002	0	0	0
3.5	-3801	1374	4586	3868	2044	0804	0254	0067	0015	0003	0001	0	0
3.6	-3918	0955	4448	3988	2198	0897	0293	0080	0019	0004	0001	0	0
3.7	-3992	0538	4283	4092	2353	0995	0336	0095	0023	0005	0001	0	0
3.8	-4026	0128	4093	4180	2507	1098	0383	0112	0028	0006	0001	0	0
3.9	-4018	-0272	3879	4250	2661	1207	0435	0130	0034	0008	0002	0	0
4.0	-3971	-0660	3641	4302	2811	1321	0491	0152	0040	0009	0002	0	0
4.1	-3887	-1033	3383	4333	2958	1439	0552	0176	0048	0011	0002	0	0
4.2	-3766	-1386	3105	4344	3100	1561	0617	0202	0057	0014	0003	0001	0
4.3	-3610	-1719	2811	4333	3236	1687	0688	0232	0067	0017	0004	0001	0
4.4	-3423	-2028	2561	4301	3365	1816	0763	0264	0078	0020	0005	0001	0
4.5	-3205	-2311	2178	4247	3484	1947	0843	0300	0091	0024	0006	0001	0
4.6	-2961	-2566	1896	4171	3594	2080	0927	0340	0106	0029	0007	0002	0
4.7	-2693	-2791	1556	4072	3693	2214	1017	0382	0122	0034	0008	0002	0
4.8	-2404	-2985	1211	3952	3780	2347	1111	0429	0141	0040	0010	0002	0
4.9	-2097	-3147	0853	3811	3853	2480	1209	0479	0161	0047	0012	0003	0001
5.0	-1776	-3276	0466	3648	3912	2611	1310	0534	0184	0055	0015	0004	0001
5.1	-1443	-3371	0121	3466	3956	2740	1416	0592	0209	0064	0017	0004	0001
5.2	-1103	-3432	-0217	3265	3985	2865	1525	0654	0237	0074	0021	0005	0001
5.3	-0758	-3460	-0547	3046	3996	2986	1637	0721	0267	0086	0024	0006	0001
5.4	-0412	-3453	-0867	2811	3991	3101	1751	0791	0300	0099	0029	0007	0002
5.5	-0068	-3414	-1173	2561	3967	3209	1868	0866	0337	0113	0034	0009	0002
5.6	0270	-3343	-1464	2298	3926	3310	1986	0945	0376	0129	0039	0011	0003
5.7	0599	-3241	-1737	2023	3866	3403	2104	1027	0418	0147	0045	0013	0003
5.8	0917	-3110	-1990	1738	3788	3486	2223	1113	0464	0166	0053	0015	0004
5.9	1220	-2951	-2221	1446	3691	3559	2341	1203	0513	0188	0061	0017	0005
6.0	1506	-2767	-2429	1148	3576	3621	2458	1296	0565	0212	0070	0020	0005
6.1	1773	-2559	-2612	0846	3444	3671	2574	1392	0621	0238	0080	0024	0006
6.2	2017	-2329	-2769	0543	3294	3708	2686	1491	0681	0266	0091	0028	0008
6.3	2238	-2081	-2899	0240	3128	3731	2795	1592	0744	0297	0104	0032	0009
6.4	2433	-1816	-3001	-0059	2945	3740	2900	1696	0810	0330	0118	0037	0011
6.5	2601	-1538	-3074	-0353	2748	3736	2999	1801	0880	0366	0133	0043	0013
6.6	2740	-1250	-3119	-0641	2537	3716	3093	1908	0954	0405	0150	0049	0015
6.7	2851	-0953	-3135	-0918	2313	3680	3180	2015	1031	0446	0168	0057	0017
6.8	2931	-0652	-3123	-1185	2077	3629	3259	2122	1111	0491	0189	0065	0020
6.9	2981	-0349	-3082	-1438	1832	3561	3330	2230	1194	0539	0211	0073	0023
7.0	3001	-0047	-3014	-1676	1578	3479	3392	2336	1280	0589	0235	0083	0027
7.1	2991	0252	-2920	-1896	1317	3380	3444	2441	1368	0643	0262	0094	0031
7.2	2951	0543	-2800	-2099	1051	3266	3486	2543	1459	0700	0290	0106	0035
7.3	2882	0826	-2656	-2281	0781	3137	3516	2643	1553	0760	0321	0120	0040
7.4	2786	1096	-2490	-2442	0510	2993	3535	2739	1648	0823	0354	0135	0046
7.5	2663	1352	-2303	-2581	0238	2835	3541	2831	1744	0889	0390	0151	0052
7.6	2516	1592	-2097	-2696	-0031	2663	3535	2919	1842	0958	0425	0168	0059
7.7	2346	1813	-1875	-2787	-0297	2478	3516	3001	1940	1031	0469	0188	0067
7.8	2154	2014	-1638	-2853	-0557	2282	3483	3076	2039	1105	0512	0209	0076
7.9	1944	2192	-1389	-2895	-0810	2075	3436	3145	2137	1183	0559	0231	0086
8.0	1717	2346	-1130	-2911	-1054	1858	3376	3206	2235	1263	0608	0256	0096
8.1	1475	2476	-0864	-2903	-1286	1632	3301	3259	2331	1346	0659	0283	0108
8.2	1222	2580	-0593	-2869	-1507	1399	3213	3303	2426	1430	0714	0311	0121
8.4	0960	2657	-0320	-2811	-1713	1161	3111	3337	2518	1517	0771	0342	0135
8.4	0692	2708	-0047	-2730	-1903	0918	2996	3362	2608	1605	0831	0375	0150
8.5	0419	2731	0223	-2626	-2077	0671	2867	3376	2694	1694	0894	0410	0167
8.6	0146	2728	0488	-2500	-2233	0424	2725	3379	2775	1785	0960	0448	0185
8.7	-0125	2697	0745	-2355	-2369	0176	2571	3371	2853	1876	1028	0488	0205
8.8	-0392	2641	0993	-2190	-2485	-0070	2406	3351	2925	1967	1099	0530	0226
8.9	-0653	2559	1228	-2007	-2581	-0313	2230	3319	2991	2058	1172	0575	0249
9.0	-0903	2453	1448	-1809	-2655	-0550	2043	3275	3051	2149	1247	0622	0274
9.1	-1142	2324	1653	-1598	-2707	-0782	1847	3218	3103	2238	1324	0672	0301
9.2	-1367	2174	1840	-1374	-2736	-1005	1644	3149	3148	2327	1404	0725	0329
9.3	-1577	2004	2008	-1141	-2743	-1219	1432	3068	3185	2413	1484	0780	0360
9.4	-1768	1816	2154	-0900	-2728	-1422	1215	2974	3214	2496	1567	0837	0392
9.5	-1939	1613	2279	-0653	-2691	-1613	0993	2868	3233	2577	1650	0897	0427
9.6	-2090	1395	2380	-0403	-2633	-1790	0768	2750	3243	2655	1735	0959	0464
9.7	-2218	1166	2458	-0153	-2553	-1953	0540	2620	3242	2728	1820	1024	0503
9.8	-2323	0928	2512	0097	-2453	-2099	0311	2480	3231	2797	1905	1091	0544
9.9	-2403	0684	2542	0343	-2333	-2229	0082	2328	3110	2860	1990	1160	0588
10.0	-2459	0435	2546	0584	-2196	-2341	-0145	2167	3179	2919	2075	1231	0634
10.1	-2490	0184	2527	0817	-2042	-2434	-0368	1996	3135	2971	2159	1304	0682
10.2	-2496	-0066	2483	1040	-1871	-2508	-0587	1817	3081	3016	2241	1379	0732
10.3	-2477	-0313	2416	1252	-1687	-2562	-0800	1630	3015	3054	2322	1455	0785
10.4	-2434	-0555	2327	1450	-1491	-2596	-1006	1436	2939	3085	2401	1532	0840
10.5	-2366	-0789	2216	1632	-1283	-2611	-1203	1236	2851	3108	2477	1611	0898
10.6	-2276	-1012	2085	1799	-1067	-2604	-1390	1031	2751	3122	2551	1690	0957
10.7	-2164	-1224	1936	1948	-0844	-2578	-1566	0822	2642	3128	2621	1770	1019
10.8	-2032	-1422	1769	2077	-0615	-2532	-1730	0610	2521	3124	2686	1850	1083
10.9	-1881	-1603	1586	2186	-0383	-2467	-1880	0397	2390	3111	2748	1930	1149
11.0	-1712	-1768	1390	2273	-0150	-2383	-2016	0184	2250	3088	2804	2010	1216
11.1	-1528	-1913	1183	2340	0082	-2281	-2136	-0029	2100	3056	2855	2089	1285
11.2	-1330	-2039	0966	2383	0311	-2161	-2241	-0239	1941	3013	2901	2167	1356
11.3	-1121	-2143	0741	2405	0536	-2026	-2328	-0447	1775	2960	2940	2243	1428
11.4	-0902	-2225	0512	2404	0753	-1875	-2399	-0649	1601	2896	2972	2318	1501
11.5	-0677	-2284	0279	2381	0963	-1711	-2451	-0846	1421	2823	2998	2390	1575
11.6	-0446	-2320	0046	2336	1162	-1534	-2485	-1036	1234	2739	3015	2460	1650
11.7	-0213	-2333	-0185	2270	1349	-1347	-2501	-1218	1043	2645	3025	2527	1726
11.8	0020	-2323	-0413	2183	1523	-1150	-2498	-1390	0849	2541	3027	2590	1802
11.9	0250	-2290	-0635	2076	1682	-0945	-2477	-1552	0651	2427	3020	2649	1877
12.0	0477	-2234	-0849	1951	1825	-0735	-2437	-1703	0451	2304	3005	2704	1953

MODIFIED BESSEL FUNCTIONS $I_n(x)$

Table of x, $Log_2(x)$, $I_o(Log_2x)$, $I_1(Log_2x)$

Last page contains $I_2(x)$ and $I_3(x)$

x	Log_2x	$I_o(Log_2x)$	$I_1(Log_2x)$
0,00	0,00000	1,00000	0,00000
0,01	0,00693	1,00001	0,00346
0,02	0,01386	1,00004	0,00693
0,03	0,02080	1,00010	0,01040
0,04	0,02773	1,00019	0,01386
0,05	0,03466	1,00030	0,01733
0,06	0,04159	1,00043	0,02080
0,07	0,04852	1,00058	0,02427
0,08	0,05545	1,00076	0,02774
0,09	0,06239	1,00097	0,03121
0,10	0,06932	1,00120	0,03468
0,11	0,07625	1,00145	0,03815
0,12	0,08318	1,00173	0,04162
0,13	0,09011	1,00203	0,04510
0,14	0,09704	1,00235	0,04858
0,15	0,10397	1,00270	0,05205
0,16	0,11091	1,00307	0,05554
0,17	0,11784	1,00347	0,05902
0,18	0,12477	1,00389	0,06250
0,19	0,13170	1,00434	0,06599
0,20	0,13863	1,00481	0,06948
0,21	0,14556	1,00530	0,07297
0,22	0,15249	1,00582	0,07647
0,23	0,15943	1,00636	0,07996
0,24	0,16636	1,00693	0,08346
0,25	0,17329	1,00752	0,08697
0,26	0,18022	1,00813	0,09047
0,27	0,18715	1,00877	0,09398
0,28	0,19408	1,00943	0,09750
0,29	0,20101	1,01012	0,10101
0,30	0,20795	1,01084	0,10453
0,31	0,21488	1,01157	0,10806
0,32	0,22181	1,01233	0,11159
0,33	0,22874	1,01312	0,11512
0,34	0,23567	1,01393	0,11865
0,35	0,24260	1,01476	0,12219
0,36	0,24953	1,01562	0,12574
0,37	0,25647	1,01651	0,12929
0,38	0,26340	1,01742	0,13284
0,39	0,27033	1,01835	0,13640
0,40	0,27726	1,01931	0,13996
0,41	0,28419	1,02029	0,14353
0,42	0,29112	1,02130	0,14711
0,43	0,29806	1,02233	0,15069
0,44	0,30499	1,02339	0,15427
0,45	0,31192	1,02447	0,15786
0,46	0,31885	1,02557	0,16146
0,47	0,32578	1,02671	0,16506
0,48	0,33271	1,02786	0,16867
0,49	0,33964	1,02904	0,17228
0,50	0,34658	1,03025	0,17590
0,51	0,35351	1,03148	0,17953
0,52	0,36044	1,03274	0,18316
0,53	0,36737	1,03402	0,18680
0,54	0,37430	1,03533	0,19045
0,55	0,38123	1,03666	0,19410
0,56	0,38816	1,03802	0,19776
0,57	0,39510	1,03940	0,20143
0,58	0,40203	1,04081	0,20510
0,59	0,40896	1,04225	0,20878
0,60	0,41589	1,04371	0,21247
0,61	0,42282	1,04519	0,21617
0,62	0,42975	1,04670	0,21987
0,63	0,43668	1,04824	0,22359
0,64	0,44362	1,04980	0,22731
0,65	0,45055	1,05139	0,23104
0,66	0,45748	1,05301	0,23477
0,67	0,46441	1,05465	0,23852
0,68	0,47134	1,05631	0,24227
0,69	0,47827	1,05801	0,24604
0,70	0,48520	1,05972	0,24981
0,71	0,49214	1,06147	0,25359
0,72	0,49907	1,06324	0,25738
0,73	0,50600	1,06504	0,26118
0,74	0,51293	1,06686	0,26499
0,75	0,51986	1,06871	0,26881

x	Log_2x	$I_o(Log_2x)$	$I_1(Log_2x)$
0,76	0,52679	1,07059	0,27264
0,77	0,53373	1,07249	0,27648
0,78	0,54066	1,07442	0,28032
0,79	0,54759	1,07638	0,28418
0,80	0,55452	1,07836	0,28805
0,81	0,56145	1,08037	0,29193
0,82	0,56838	1,08241	0,29582
0,83	0,57531	1,08447	0,29972
0,84	0,58225	1,08656	0,30363
0,85	0,58918	1,08868	0,30756
0,86	0,59611	1,09083	0,31149
0,87	0,60304	1,09300	0,31543
0,88	0,60997	1,09520	0,31939
0,89	0,61690	1,09743	0,32336
0,90	0,62383	1,09968	0,32734
0,91	0,63077	1,10196	0,33133
0,92	0,63770	1,10427	0,33533
0,93	0,64463	1,10661	0,33935
0,94	0,65156	1,10898	0,34337
0,95	0,65849	1,11137	0,34741
0,96	0,66542	1,11380	0,35147
0,97	0,67235	1,11625	0,35553
0,98	0,67929	1,11872	0,35961
0,99	0,68622	1,12123	0,36370
1,00	0,69315	1,12377	0,36781
1,01	0,70008	1,12633	0,37193
1,02	0,70701	1,12892	0,37606
1,03	0,71394	1,13154	0,38020
1,04	0,72087	1,13419	0,38436
1,05	0,72781	1,13687	0,38853
1,06	0,73474	1,13958	0,39272
1,07	0,74167	1,14232	0,39692
1,08	0,74860	1,14508	0,40114
1,09	0,75553	1,14788	0,40537
1,10	0,76246	1,15070	0,40961
1,11	0,76940	1,15356	0,41387
1,12	0,77633	1,15644	0,41815
1,13	0,78326	1,15935	0,42244
1,14	0,79019	1,16230	0,42674
1,15	0,79712	1,16527	0,43106
1,16	0,80405	1,16827	0,43540
1,17	0,81098	1,17130	0,43975
1,18	0,81792	1,17437	0,44412
1,19	0,82485	1,17746	0,44851
1,20	0,83178	1,18059	0,45291
1,21	0,83871	1,18374	0,45732
1,22	0,84564	1,18693	0,46176
1,23	0,85257	1,19014	0,46621
1,24	0,85950	1,19339	0,47068
1,25	0,86644	1,19667	0,47516
1,26	0,87337	1,19998	0,47966
1,27	0,88030	1,20332	0,48418
1,28	0,88723	1,20669	0,48872
1,29	0,89416	1,21009	0,49327
1,30	0,90109	1,21353	0,49785
1,31	0,90802	1,21699	0,50244
1,32	0,91496	1,22049	0,50705
1,33	0,92189	1,22402	0,51168
1,34	0,92882	1,22759	0,51632
1,35	0,93575	1,23118	0,52099
1,36	0,94268	1,23481	0,52567
1,37	0,94961	1,23847	0,53037
1,38	0,95655	1,24216	0,53510
1,39	0,96348	1,24589	0,53984
1,40	0,97041	1,24964	0,54460
1,41	0,97734	1,25344	0,54938
1,42	0,98427	1,25726	0,55419
1,43	0,99120	1,26112	0,55901
1,44	0,99813	1,26501	0,56385
1,45	1,00507	1,26894	0,56871
1,46	1,01200	1,27289	0,57360
1,47	1,01893	1,27689	0,57850
1,48	1,02586	1,28091	0,58343
1,49	1,03279	1,28498	0,58838
1,50	1,03972	1,28907	0,59334
1,51	1,04665	1,29320	0,59833
1,52	1,05359	1,29737	0,60335
1,53	1,06052	1,30157	0,60838
1,54	1,06745	1,30580	0,61344
1,55	1,07438	1,31007	0,61852
1,56	1,08131	1,31437	0,62362
1,57	1,08824	1,31871	0,62874
1,58	1,09517	1,32309	0,63389
1,59	1,10211	1,32750	0,63906
1,60	1,10904	1,33195	0,64426
1,61	1,11597	1,33643	0,64947
1,62	1,12290	1,34095	0,65471
1,63	1,12983	1,34551	0,65998
1,64	1,13676	1,35010	0,66527
1,65	1,14369	1,35473	0,67058

x	Log_2x	$I_o(Log_2x)$	$I_1(log_2x)$
1.66	1.15063	1.35940	0.67592
1.67	1.15756	1.36410	0.68128
1.68	1.16449	1.36884	0.68667
1.69	1.17142	1.37362	0.69209
1.70	1.17835	1.37844	0.69753
1.71	1.18528	1.38329	0.70299
1.72	1.19222	1.38818	0.70848
1.73	1.19915	1.39311	0.71400
1.74	1.20608	1.39808	0.71954
1.75	1.21301	1.40309	0.72511
1.76	1.21994	1.40814	0.73070
1.77	1.22687	1.41322	0.73632
1.78	1.23380	1.41834	0.74197
1.79	1.24074	1.42351	0.74765
1.80	1.24767	1.42871	0.75335
1.81	1.25460	1.43395	0.75909
1.82	1.26153	1.43923	0.76485
1.83	1.26846	1.44455	0.77063
1.84	1.27539	1.44991	0.77645
1.85	1.28232	1.45532	0.78229
1.86	1.28926	1.46076	0.78817
1.87	1.29619	1.46624	0.79407
1.88	1.30312	1.47177	0.80000
1.89	1.31005	1.47733	0.80596
1.90	1.31698	1.48294	0.81195
1.91	1.32391	1.48859	0.81797
1.92	1.33084	1.49428	0.82402
1.93	1.33778	1.50001	0.83010
1.94	1.34471	1.50579	0.83622
1.95	1.35164	1.51161	0.84236
1.96	1.35857	1.51747	0.84853
1.97	1.36550	1.52337	0.85474
1.98	1.37243	1.52932	0.86097
1.99	1.37936	1.53530	0.86724
2.00	1.38630	1.54134	0.87354
2.01	1.39323	1.54741	0.87987
2.02	1.40016	1.55354	0.88624
2.03	1.40709	1.55970	0.89264
2.04	1.41402	1.56591	0.89907
2.05	1.42095	1.57216	0.90553
2.06	1.42789	1.57846	0.91203
2.07	1.43482	1.58481	0.91856
2.08	1.44175	1.59120	0.92512
2.09	1.44868	1.59763	0.93172
2.10	1.45561	1.60411	0.93835
2.11	1.46254	1.61064	0.94502
2.12	1.46947	1.61722	0.95172
2.13	1.47641	1.62384	0.95846
2.14	1.48334	1.63050	0.96524
2.15	1.49027	1.63722	0.97205
2.16	1.49720	1.64398	0.97889
2.17	1.50413	1.65079	0.98577
2.18	1.51106	1.65764	0.99269
2.19	1.51799	1.66455	0.99965
2.20	1.52493	1.67150	1.00664
2.21	1.53186	1.67850	1.01367
2.22	1.53879	1.68555	1.02073
2.23	1.54572	1.69265	1.02784
2.24	1.55265	1.69980	1.03498
2.25	1.55958	1.70700	1.04216
2.26	1.56651	1.71425	1.04938
2.27	1.57345	1.72155	1.05664
2.28	1.58038	1.72890	1.06394
2.29	1.58731	1.73630	1.07127
2.30	1.59424	1.74375	1.07865
2.31	1.60117	1.75125	1.08607
2.32	1.60810	1.75881	1.09353
2.33	1.61503	1.76641	1.10102
2.34	1.62197	1.77407	1.10856
2.35	1.62890	1.78178	1.11614
2.36	1.63583	1.78954	1.12376
2.37	1.64276	1.79736	1.13145
2.38	1.64969	1.80523	1.13913
2.39	1.65662	1.81315	1.14688
2.40	1.66356	1.82113	1.15467
2.41	1.67049	1.82916	1.16251
2.42	1.67742	1.83724	1.17038
2.43	1.68435	1.84538	1.17830
2.44	1.69128	1.85358	1.18627
2.45	1.69821	1.86183	1.19427
2.46	1.70514	1.87014	1.20233
2.47	1.71208	1.87850	1.21043
2.48	1.71901	1.88692	1.21857
2.49	1.72594	1.89539	1.22676
2.50	1.73287	1.90392	1.23499
2.51	1.73980	1.91251	1.24327
2.52	1.74673	1.92116	1.25160
2.53	1.75366	1.92986	1.25997
2.54	1.76060	1.93862	1.26839
2.55	1.76753	1.94745	1.27686

x	Log_2x	$I_o(Log_2x)$	$I_1(Log_2x)$
2.56	1.77446	1.95633	1.28537
2.57	1.78139	1.96527	1.29394
2.58	1.78832	1.97426	1.30255
2.59	1.79525	1.98332	1.31121
2.60	1.80218	1.99244	1.31992
2.61	1.80912	2.00162	1.32868
2.62	1.81605	2.01086	1.33749
2.63	1.82298	2.02016	1.34634
2.64	1.82991	2.02953	1.35525
2.65	1.83684	2.03895	1.36421
2.66	1.84377	2.04844	1.37322
2.67	1.85070	2.05799	1.38229
2.68	1.85764	2.06760	1.39140
2.69	1.86457	2.07728	1.40056
2.70	1.87150	2.08702	1.40978
2.71	1.87843	2.09682	1.41905
2.72	1.88536	2.10669	1.42838
2.73	1.89229	2.11662	1.43776
2.74	1.89923	2.12662	1.44719
2.75	1.90616	2.13668	1.45667
2.76	1.91309	2.14681	1.46622
2.77	1.92002	2.15701	1.47581
2.78	1.92695	2.16727	1.48546
2.79	1.93388	2.17760	1.49517
2.80	1.94081	2.18800	1.50493
2.81	1.94775	2.19847	1.51475
2.82	1.95468	2.20900	1.52463
2.83	1.96161	2.21960	1.53456
2.84	1.96854	2.23027	1.54455
2.85	1.97547	2.24102	1.55460
2.86	1.98240	2.25183	1.56471
2.87	1.98933	2.26271	1.57488
2.88	1.99627	2.27366	1.58511
2.89	2.00320	2.28468	1.59539
2.90	2.01013	2.29578	1.60574
2.91	2.01706	2.30694	1.61614
2.92	2.02390	2.31818	1.62661
2.93	2.03092	2.32949	1.63714
2.94	2.03785	2.34088	1.64773
2.95	2.04479	2.35233	1.65838
2.96	2.05172	2.36387	1.66910
2.97	2.05865	2.37547	1.67987
2.98	2.06558	2.38715	1.69072
2.99	2.07251	2.39891	1.70162
3.00	2.07944	2.41074	1.71259
3.01	2.08637	2.42265	1.72362
3.02	2.09331	2.43464	1.73472
3.03	2.10024	2.44670	1.74589
3.04	2.10717	2.45884	1.75712
3.05	2.11410	2.47106	1.76841
3.06	2.12103	2.48336	1.77978
3.07	2.12796	2.49573	1.79121
3.08	2.13490	2.50819	1.80271
3.09	2.14183	2.52073	1.81427
3.10	2.14876	2.53334	1.82591
3.11	2.15569	2.54604	1.83761
3.12	2.16262	2.55882	1.84939
3.13	2.16955	2.57168	1.86123
3.14	2.17648	2.58462	1.87315
3.15	2.18342	2.59764	1.88513
3.16	2.19035	2.61075	1.89719
3.17	2.19728	2.62394	1.90932
3.18	2.20421	2.63722	1.92152
3.19	2.21114	2.65058	1.93379
3.20	2.21807	2.66403	1.94614
3.21	2.22500	2.67756	1.95856
3.22	2.23194	2.69118	1.97105
3.23	2.23887	2.70489	1.98363
3.24	2.24580	2.71868	1.99627
3.25	2.25273	2.73256	2.00899
3.26	2.25966	2.74653	2.02179
3.27	2.26659	2.76059	2.03466
3.28	2.27352	2.77474	2.04761
3.29	2.28046	2.78898	2.06064
3.30	2.28739	2.80330	2.07375
3.31	2.29432	2.81772	2.08694
3.32	2.30125	2.83224	2.10020
3.33	2.30818	2.84684	2.11355
3.34	2.31511	2.86154	2.12698
3.35	2.32204	2.87633	2.14048
3.36	2.32898	2.89121	2.15407
3.37	2.33591	2.90619	2.16774
3.38	2.34284	2.92126	2.18149
3.39	2.34977	2.93643	2.19533
3.40	2.35670	2.95170	2.20925
3.41	2.36363	2.96706	2.22326
3.42	2.37057	2.98252	2.23734
3.43	2.37750	2.99807	2.25152
3.44	2.38443	3.01373	2.26578
3.45	2.39136	3.02948	2.28012

x	$Log_2 x$	$I_o(Log_2 x)$	$I_1(Log_2 x)$
3,46	2,39829	3,04534	2,29456
3,47	2,40522	3,06129	2,30908
3,48	2,41215	3,07735	2,32369
3,49	2,41909	3,09351	2,33839
3,50	2,42602	3,10977	2,35317
3,51	2,43295	3,12613	2,36805
3,52	2,43988	3,14260	2,38302
3,53	2,44681	3,15917	2,39808
3,54	2,45374	3,17584	2,41323
3,55	2,46067	3,19262	2,42847
3,56	2,46761	3,20951	2,44380
3,57	2,47454	3,22650	2,45923
3,58	2,48147	3,24360	2,47476
3,59	2,48840	3,26081	2,49037
3,60	2,49533	3,27812	2,50609
3,61	2,50226	3,29555	2,52190
3,62	2,50919	3,31308	2,53780
3,63	2,51613	3,33073	2,55380
3,64	2,52306	3,34849	2,56990
3,65	2,52999	3,36636	2,58610
3,66	2,53692	3,38434	2,60240
3,67	2,54385	3,40243	2,61880
3,68	2,55078	3,42064	2,63530
3,69	2,55772	3,43897	2,65190
3,70	2,56465	3,45741	2,66860
3,71	2,57158	3,47596	2,68540
3,72	2,57851	3,49463	2,70231
3,73	2,58544	3,51342	2,71932
3,74	2,59237	3,53233	2,73644
3,75	2,59930	3,55136	2,75366
3,76	2,60624	3,57051	2,77098
3,77	2,61317	3,58977	2,78842
3,78	2,62010	3,60916	2,80596
3,79	2,62703	3,62867	2,82360
3,80	2,63396	3,64831	2,84136
3,81	2,64089	3,66806	2,85922
3,82	2,64782	3,68794	2,87720
3,83	2,65476	3,70795	2,89529
3,84	2,66169	3,72808	2,91349
3,85	2,66862	3,74834	2,93180
3,86	2,67555	3,76873	2,95022
3,87	2,68248	3,78924	2,96876
3,88	2,68941	3,80988	2,98741
3,89	2,69634	3,83065	3,00617
3,90	2,70328	3,85156	3,02505
3,91	2,71021	3,87259	3,04405
3,92	2,71714	3,89376	3,06317
3,93	2,72407	3,91506	3,08240
3,94	2,73100	3,93649	3,10176
3,95	2,73793	3,95805	3,12123
3,96	2,74486	3,97976	3,14083
3,97	2,75180	4,00160	3,16054
3,98	2,75873	4,02357	3,18038
3,99	2,76566	4,04569	3,20034
4,00	2,77259	4,06794	3,22042
4,01	2,77952	4,09033	3,24063
4,02	2,78645	4,11286	3,26096
4,03	2,79339	4,13554	3,28142
4,04	2,80032	4,15835	3,30201
4,05	2,80725	4,18131	3,32272
4,06	2,81418	4,20442	3,34357
4,07	2,82111	4,22767	3,36454
4,08	2,82804	4,25106	3,38564
4,09	2,83497	4,27460	3,40687
4,10	2,84191	4,29829	3,42824
4,11	2,84884	4,32213	3,44974
4,12	2,85577	4,34611	3,47137
4,13	2,86270	4,37025	3,49314
4,14	2,86963	4,39454	3,51504
4,15	2,87656	4,41898	3,53708
4,16	2,88349	4,44357	3,55925
4,17	2,89043	4,46832	3,58157
4,18	2,89736	4,49323	3,60402
4,19	2,90429	4,51828	3,62661
4,20	2,91122	4,54350	3,64935
4,21	2,91815	4,56888	3,67222
4,22	2,92508	4,59441	3,69524
4,23	2,93201	4,62010	3,71840
4,24	2,93895	4,64596	3,74171
4,25	2,94588	4,67197	3,76516
4,26	2,95281	4,69815	3,78876
4,27	2,95974	4,72450	3,81250
4,28	2,96667	4,75101	3,83640
4,29	2,97360	4,77768	3,86044
4,30	2,98053	4,80452	3,88463
4,31	2,98747	4,83154	3,90898
4,32	2,99440	4,85872	3,93347
4,33	3,00133	4,88607	3,95812
4,34	3,00826	4,91359	3,98293
4,35	3,01519	4,94128	4,00788

x	$Log_2 x$	$I_o(Log_2 x)$	$I_1(Log_2 x)$
4,36	3,02212	4,96915	4,03300
4,37	3,02906	4,99719	4,05827
4,38	3,03599	5,02541	4,08370
4,39	3,04292	5,05380	4,10929
4,40	3,04985	5,08238	4,13504
4,41	3,05678	5,11113	4,16095
4,42	3,06371	5,14006	4,18703
4,43	3,07064	5,16917	4,21326
4,44	3,07758	5,19847	4,23967
4,45	3,08451	5,22795	4,26623
4,46	3,09144	5,25761	4,29297
4,47	3,09837	5,28746	4,31987
4,48	3,10530	5,31750	4,34694
4,49	3,11223	5,34772	4,37418
4,50	3,11916	5,37814	4,40159
4,51	3,12610	5,40874	4,42917
4,52	3,13303	5,43954	4,45693
4,53	3,13996	5,47053	4,48486
4,54	3,14689	5,50171	4,51297
4,55	3,15382	5,53309	4,54125
4,56	3,16075	5,56467	4,56971
4,57	3,16768	5,59644	4,59835
4,58	3,17462	5,62842	4,62717
4,59	3,18155	5,66059	4,65617
4,60	3,18848	5,69296	4,68536
4,61	3,19541	5,72554	4,71472
4,62	3,20234	5,75832	4,74428
4,63	3,20927	5,79131	4,77401
4,64	3,21620	5,82451	4,80394
4,65	3,22314	5,85791	4,83405
4,66	3,23007	5,89152	4,86436
4,67	3,23700	5,92534	4,89485
4,68	3,24393	5,95938	4,92554
4,69	3,25086	5,99363	4,95642
4,70	3,25779	6,02809	4,98749
4,71	3,26473	6,06277	5,01876
4,72	3,27166	6,09767	5,05023
4,73	3,27859	6,13278	5,08189
4,74	3,28552	6,16812	5,11376
4,75	3,29245	6,20367	5,14582
4,76	3,29938	6,23945	5,17809
4,77	3,30631	6,27546	5,21057
4,78	3,31325	6,31169	5,24324
4,79	3,32018	6,34814	5,27613
4,80	3,32711	6,38483	5,30922
4,81	3,33404	6,42175	5,34252
4,82	3,34097	6,45889	5,37603
4,83	3,34790	6,49627	5,40975
4,84	3,35483	6,53389	5,44369
4,85	3,36177	6,57174	5,47783
4,86	3,36870	6,60983	5,51220
4,87	3,37563	6,64816	5,54678
4,88	3,38256	6,68672	5,58158
4,89	3,38949	6,72553	5,61661
4,90	3,39642	6,76459	5,65185
4,91	3,40335	6,80389	5,68731
4,92	3,41029	6,84343	5,72301
4,93	3,41722	6,88322	5,75892
4,94	3,42415	6,92327	5,79506
4,95	3,43108	6,96356	5,83144
4,96	3,43801	7,00411	5,86804
4,97	3,44494	7,04491	5,90487
4,98	3,45187	7,08597	5,94194
4,99	3,45881	7,12728	5,97924
5,00	3,46574	7,16886	6,01678

x	$I_2(x)$	$I_3(x)$
0.2	0.0050	0.000167
0.4	0.0203	0.00135
0.6	0.0464	0.00459
0.8	0.0843	0.0111
1.0	0.1358	0.0222
1.2	0.2026	0.0393
1.4	0.2876	0.0644
1.6	0.3940	0.0998
1.8	0.5260	0.148
2.0	0.6890	0.213
2.2	0.8891	0.297
2.4	1.1111	0.407
2.6	1.4338	0.552
2.8	1.7994	0.728
3.0	2.2452	0.956
3.4	3.4495	
3.8	5.2323	
4.2	7.8683	
4.6	11.761	
5.0	17.505	

CHAPTER 9B

ENGINEERING REFERENCES

CONTENTS:

RESISTOR COLOR CODE

COLOR	NUMERAL	NUMBER OF ZEROS (Third Band)
Black	0	0 (for 10 ohms to 99 ohms)
Brown	1	1 (for 100 ohms to 999 ohms)
Red	2	2 (for 1k to 9.9k)
Orange	3	3 (for 10k to 99k)
Yellow	4	4 (for 100k to 999k)
Green	5	5 (for 1 Meg to 9.9 Megs)
Blue	6	6 (for 10 Meg to 22 Meg)
Violet	7	7 (not generally used)
Grey	8	8 (not generally used)
White	9	9 (not generally used)

Silver for fourth band indicates 10% tolerance
Gold for fourth band indicates 5% tolerance
Any additional bands are reliability code

BUTTERWORTH POLYNOMIALS (FACTORED TO FIRST AND SECOND ORDER)

FILTER ORDER:	POLYNOMIAL:
1	$(s+1)$
2	$(s^2 + 1.414s + 1)$
3	$(s+1)(s^2 + s + 1)$
4	$(s^2 + 0.765s + 1)(s^2 + 1.848s + 1)$
5	$(s+1)(s^2 + 0.618s + 1)(s^2 + 1.618s + 1)$
6	$(s^2 + 0.518s + 1)(s^2 + 1.414s + 1)(s^2 + 1.932s + 1)$
7	$(s+1)(s^2 + 0.445s + 1)(s^2 + 1.247s + 1)(s^2 + 1.802s +1)$
8	$(s^2 + 0.390s + 1)(s^2 + 1.111s + 1)(s^2 + 1.663s + 1)(s^2 + 1.962s + 1)$

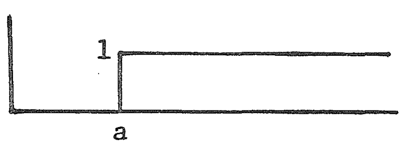
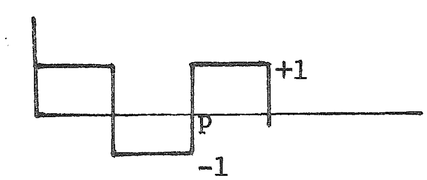
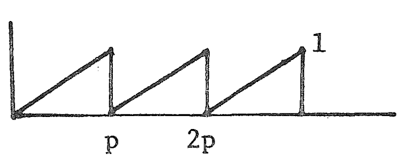
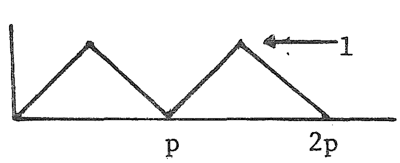
The table below gives the ratios of 5% resistor values. It is useful for choosing a pair of resistors to achieve a needed gain for an op-amp configuration for example. Two things should be kept in mind however. First, note that the resistors have a tolerance of 5% (or 10% of these are used for the values without the *) and therefore, the ratio error is 10%. Thus, in many cases it does not make a lot of difference which of several close values is used. Secondly, note that since the selection of values is approximately exponential, the ratios tend to cluster around certain values (diagonally right and down).

	10	11*	12	13*	15	16*	18	20*	22	24*	27	30*	33	36*	39	43*	47	51*	56	62*	68	75*	82	91*
10	1	.909	.833	.769	.667	.625	.556	.500	.455	.417	.370	.333	.303	.278	.256	.233	.213	.196	.179	.161	.147	.133	.122	.110
11*	1.10	1	.917	.846	.733	.688	.611	.550	.500	.458	.407	.367	.333	.306	.282	.256	.234	.216	.196	.177	.162	.147	.134	.121
12	1.20	1.09	1	.923	.800	.750	.667	.600	.545	.500	.444	.400	.364	.333	.308	.279	.255	.235	.214	.194	.176	.160	.146	.132
13*	1.30	1.18	1.08	1	.867	.813	.722	.650	.591	.542	.481	.433	.394	.361	.333	.302	.277	.255	.232	.210	.191	.173	.159	.143
15	1.50	1.36	1.25	1.15	1	.938	.833	.750	.682	.625	.556	.500	.455	.417	.385	.349	.319	.294	.268	.241	.221	.200	.183	.165
16*	1.60	1.45	1.33	1.23	1.07	1	.889	.800	.727	.667	.593	.533	.528	.444	.410	.372	.340	.313	.286	.258	.235	.213	.195	.175
18	1.80	1.64	1.50	1.38	1.20	1.13	1	.900	.818	.750	.667	.600	.545	.500	.462	.419	.383	.353	.321	.290	.265	.240	.220	.198
20*	2.00	1.82	1.67	1.54	1.33	1.25	1.11	1	.909	.833	.741	.667	.606	.556	.512	.465	.426	.392	.357	.323	.294	.267	.244	.220
22	2.20	2.00	1.83	1.69	1.47	1.38	1.22	1.10	1	.917	.815	.733	.667	.611	.564	.512	.468	.431	.393	.355	.324	.293	.268	.242
24*	2.40	2.18	2.00	1.85	1.60	1.50	1.33	1.20	1.09	1	.889	.800	.727	.667	.615	.558	.511	.471	.429	.387	.353	.320	.293	.264
27	2.70	2.45	2.25	2.08	1.80	1.69	1.50	1.35	1.23	1.13	1	.900	.818	.750	.692	.628	.574	.529	.482	.435	.397	.360	.329	.297
30*	3.00	2.73	2.50	2.31	2.00	1.88	1.67	1.50	1.36	1.25	1.11	1	.909	.833	.769	.698	.638	.588	.536	.484	.441	.400	.366	.330
33	3.30	3.00	2.75	2.54	2.20	2.06	1.83	1.65	1.50	1.38	1.22	1.10	1	.917	.846	.767	.702	.647	.589	.532	.485	.440	.402	.363
36*	3.60	3.27	3.00	2.77	2.40	2.25	2.00	1.80	1.64	1.50	1.33	1.20	1.09	1	.923	.837	.766	.706	.643	.581	.529	.480	.439	.396
39	3.90	3.55	3.25	3.00	2.60	2.44	2.17	1.95	1.77	1.63	1.44	1.30	1.18	1.08	1	.907	.830	.765	.696	.629	.574	.520	.475	.429
43*	4.30	3.91	3.58	3.31	2.87	2.69	2.39	2.15	1.95	1.79	1.59	1.43	1.30	1.19	1.10	1	.915	.843	.768	.694	.632	.573	.524	.473
47	4.70	4.27	3.92	3.62	3.13	2.94	2.61	2.35	2.14	1.96	1.74	1.57	1.42	1.31	1.21	1.09	1	.922	.839	.758	.691	.627	.573	.516
51*	5.10	4.64	4.25	3.92	3.40	3.19	2.83	2.55	2.32	2.13	1.88	1.70	1.55	1.42	1.31	1.19	1.09	1	.911	.823	.750	.680	.623	.560
56	5.60	5.09	4.67	4.31	3.73	3.50	3.11	2.80	2.55	2.33	2.07	1.87	1.70	1.56	1.44	1.30	1.19	1.10	1	.903	.824	.747	.683	.615
62*	6.20	5.64	5.16	4.77	4.13	3.88	3.44	3.10	2.82	2.58	2.30	2.07	1.88	1.72	1.59	1.44	1.32	1.22	1.11	1	.912	.827	.756	.681
68	6.80	6.18	5.67	5.23	4.53	4.25	3.78	3.40	3.09	2.83	2.52	2.27	2.06	1.89	1.74	1.58	1.45	1.33	1.21	1.10	1	.907	.829	.747
75*	7.50	6.82	6.25	5.77	5.00	4.69	4.17	3.75	3.41	3.13	2.78	2.50	2.27	2.08	1.92	1.74	1.60	1.47	1.34	1.21	1.10	1	.915	.824
82	8.20	7.45	6.83	6.31	5.47	5.13	4.56	4.10	3.73	3.42	3.04	2.73	2.48	2.28	2.10	1.91	1.74	1.61	1.46	1.32	1.21	1.09	1	.901
91*	9.10	8.27	7.58	7.00	6.07	5.69	5.06	4.55	4.14	3.79	3.37	3.03	2.76	2.53	2.33	2.12	1.94	1.78	1.63	1.47	1.34	1.21	1.23	1

TABLE OF 5% RESISTOR RATIOS

* indicates 5% value only, rest in 5% or 10% tolerances

LAPLACE TRANSFORM PAIRS $[F(s) = \int_0^\infty f(t)\, e^{-st} dt]$

f(t)	F(s)
1	$1/s$
t	$1/s^2$
e^{at}	$1/(s-a)$
$\mathrm{Cos}(at)$	$s/(s^2+a^2)$
$\mathrm{Sin}(at)/a$	$1/(s^2+a^2)$
$e^{bt}\mathrm{Cos}(at)$	$(s-b)/[(s-b)^2+a^2]$
$t\ \mathrm{Sin}(at)/2a$	$s/(s^2+a^2)^2$
$t\ \mathrm{Cos}(at)$	$(s^2-a^2)/(s^2+a^2)^2$
$\mathrm{Sin}(at)/t$	$\mathrm{Tan}^{-1}(a/s)$
$\delta(t-a)$	e^{-as} (=1 for a=0)
te^{at}	$1/(s-a)^2$
	$\frac{1}{s}e^{-as}$
	$\frac{1}{s}$ Tanh(ps/4)
	$\frac{1}{ps^2} - \frac{e^{-ps}}{s(1-e^{-ps})}$
	$\frac{2}{ps^2}$ Tanh(ps/4)

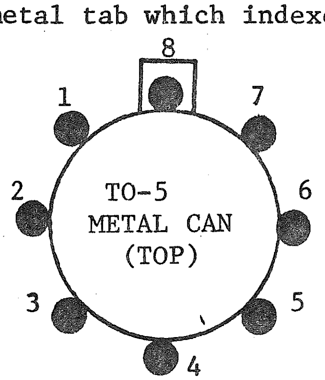
CHAPTER 9C

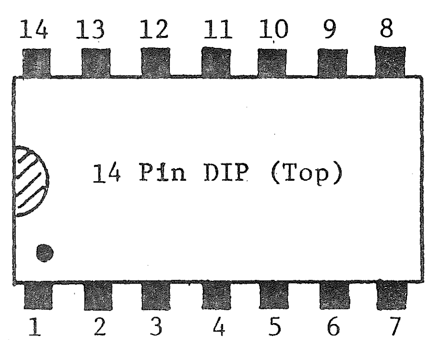
ELECTRONIC COMPONENT DATA

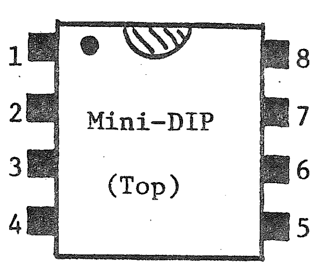
CONTENTS:

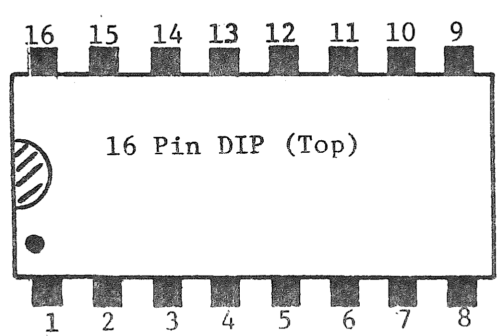
INTRODUCTION

This chapter will be concerned with the base diagrams of various IC and other semiconductor components. The reader should become familiar with the various common packages for IC's. We will be concerned with the "Round Metal Can" or TO-5 package, the 8-Pin Dual Inline (DIP) or Mini-DIP, the 14-Pin DIP, and the 16-Pin DIP. We will show views from the <u>top</u>. Note that the DIP packages feature a semicircular indentation in one end, and a dot to mark pin 1, while the TO-5 can has a metal tab which indexes the highest numbered pin.



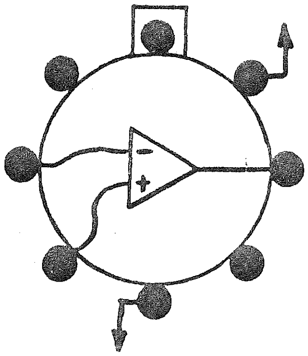


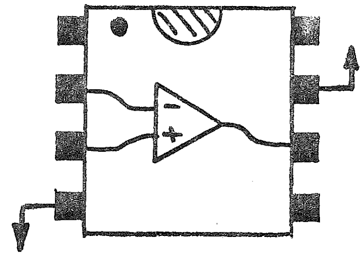




LINEAR IC'S

The two pin configurations for the IC op-amp below are very common and probably should be memorized. They use either the TO-5 or the Mini-DIP and the pins have the same numbers in either case. The common pins are: Pin 2 (- in), Pin 3 (+ in), Pin 4 (- supply), Pin 6 (output) and Pin 7 (+ supply).





The remaining pins (1, 5, and 8) are either unused, or are used for frequency compensation and/or offset null. Below is a list of common op-amps that use the above pin configuration which lists the uses of pins 1, 5, and 8.

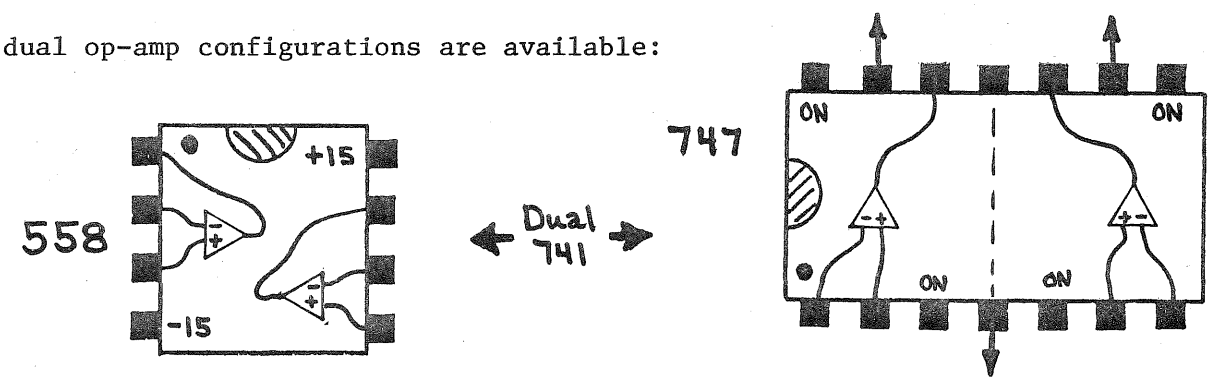
OP-AMP TYPE	Pin 1	Pin 5	Pin 8
101, 201, 301	FC ON	ON	FC
107, 207, 307	NC	NC	NC
108, 208, 308	FC	NC	FC
531	ON	ON	FC
536	ON	ON	NC
556	ON	ON	NC
709	FC	FC	FC
740	ON	ON	NC
741	ON	ON	NC
748	FC ON	ON	FC

FC = Frequency Compensation

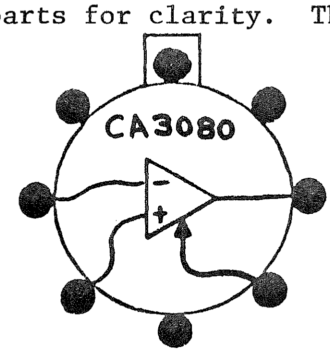
ON = Offset Null

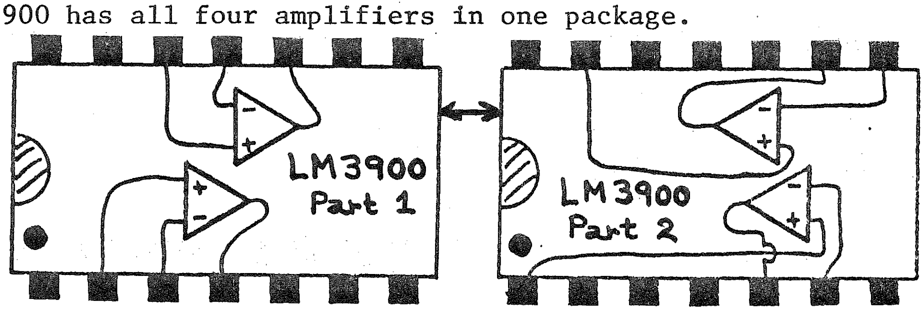
NC = No Connection

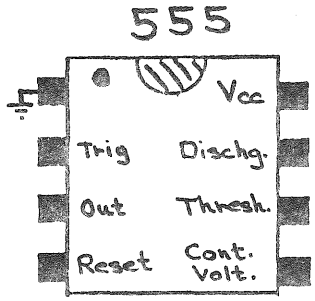
A number of dual op-amp configurations are available:



Below are the pin configurations of the CA3080 OTA and the LM3900 CDA. The 3080 is the standard op-amp except for the control pin on pin 5. Note that the LM3900 is given in two parts for clarity. The LM3900 has all four amplifiers in one package.

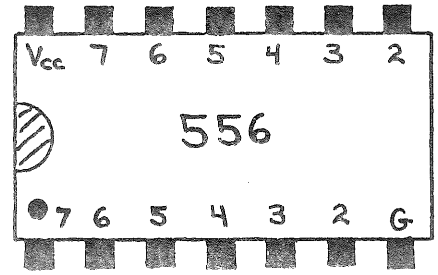


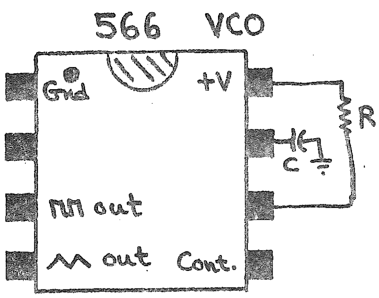


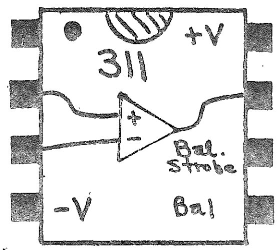


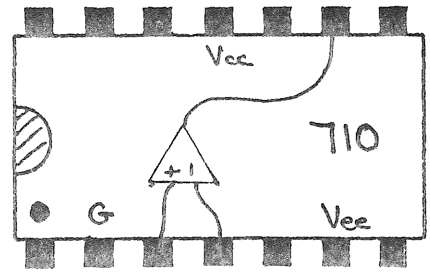
TIMERS

The pin arrangement for the 555 is shown at the left. The 556 is a dual timer. Its pins are listed in terms of the function of the 555, as this is the way they are most easily related.

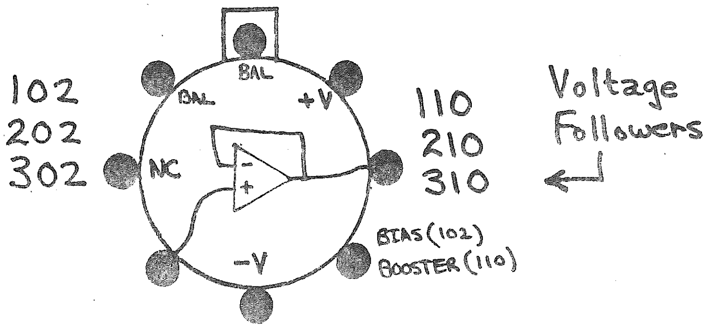


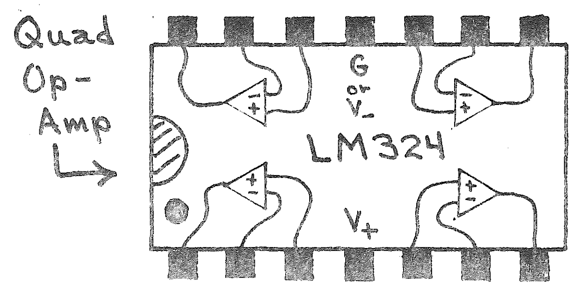






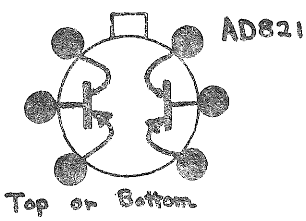
Comparators

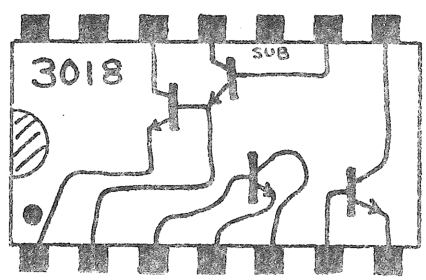


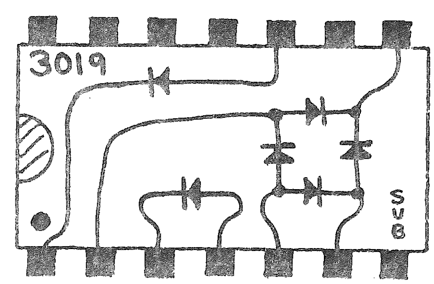


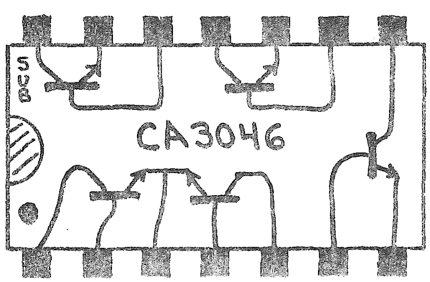
ARRAYS

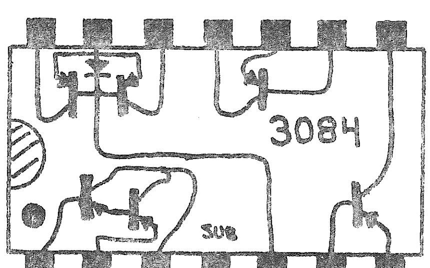
The AD821 is a PNP matched pair. The CA3018 and CA3046 (Also as CA3045 and the CA3086) are NPN arrays. The CA3019 is a diode array (Chapter 5c) while the CA3096 is a mixed PNP NPN array. The CA3084 is useful as a multiple current mirror (Ch. 5d).

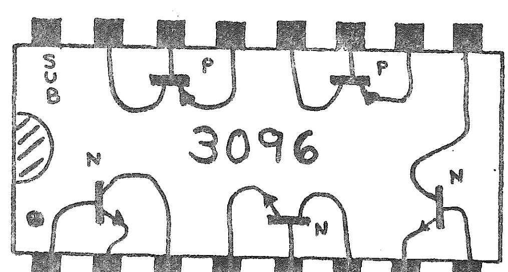












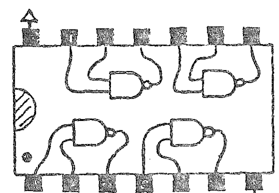
DIGITAL IC'S

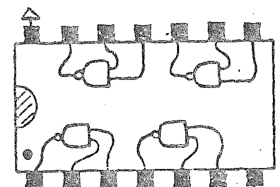
The listing below covers mainly the 7400 series and the 4000 series CMOS circuits. Most of the base diagrams of the simpler circuits are given, and a reference is given for base diagrams not given. Some digital IC's are so complex that a base diagram alone is of little help anyway, so these base diagrams are not given. A rather complete list of type numbers, descriptions, and base references is given.

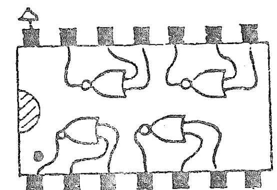
The 7400 TTL also includes in some cases a listing of output type as TPO (totem pole - normal TTL), OCO (open collector), or TSO (tri-state output). The references listed for base diagrams are: SIG = Signetics Digital Linear, MOS, Data Book; TI-1 = Texas Instruments The TTL Data Book, and TI-2 = Supplement to the TTL Data Book; NAT = National's CMOS Integrated Circuits (red cover) and RCA is a reference to any of several RCA listings of 4000 series CMOS. A full description of 74C00 CMOS is found in the National book.

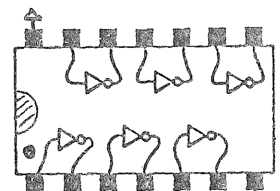
In many cases, the base diagram alone will not tell the designer all he needs to know, except for the simpler gates. Thus the user should have available some of the above references, and use this listing as an overview or to locate pin numbers when the application is clear.

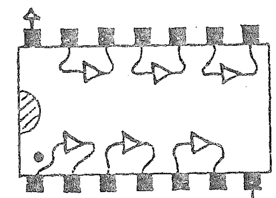
7400	Quad 2-Input NAND Gate	- - -	TPO	This Book
7401	Quad 2-Input NAND Gate	- - -	OCO	This Book
7402	Quad 2-Input NOR Gate	- - -	TPO	This Book
7403	Quad 2 Input NAND Gate	- - -	OCO	As 7400, This Book
7404	Hex Inverter	- - - -	TPO	This Book
7405	Hex Inverter	- - - -	OCO	As 7404, This Book
7406	Hex Inverter	- - - -	OCO 30V	As 7404, This Book
7407	Hex Buffer	- - - -	OCO 30V	This Book
7408	Quad 2-Input AND Gate	- - -	TPO	This Book
7409	Quad 2-Input AND Gate	- - -	OCO	As 7408, This Book
7410	Triple 3-Input NAND	- - -	TPO	This Book
7411	Triple 3-Input AND	- - -	TPO	This Book
7412	Triple 3-Input NAND	- - -	OCO	As 7410, This Book
7413	Dual 4-Input NAND Schmitt Trigger	- -	TPO	As 7420, This Book
7414	Hex Schmitt Trigger Inverter	- -	TPO	As 7404, This Book
7415	Triple 3-Input AND	- - -	OCO	As 7411, This Book
7416	Hex Inverter	- - - -	OCO 15V	As 7404, This Book
7417	Hex Buffer	- - - -	OCO 15V	As 7407, This Book
7420	Dual 4-Input NAND Gate	- - -	TPO	This Book
7421	Dual 4-Input AND Gate	- - -	TPO	This Book
7422	Dual 4-Input NAND Gate	- - -	OCO	As 7420, This Book
7423	Dual Expandable 4-Input NOR with Strobe	-	TPO	TI-1, pg. 67
7425	Dual 4-Input NOR with Strobe	- -	TPO	TI-1, pg. 67
7426	Quad 2-Input NAND	- - -	OCO 15V	As 7400, This Book
7427	Triple 3-Input NOR	- - -	TPO	This Book
7428	Quad 2-Input NOR (Buffer, 3 times drive of 7402)		TPO	As 7402, This Book
7430	8-Input NAND Gate	- - -	TPO	This Book
7432	Quad 2-Input OR Gate		TPO	This Book
7433	Quad 2-Input NOR Gate (Buffer)		OCO	As 7402, This Book
7437	Quad 2-Input NAND (Buffer, 3 times drive of 7400)		TPO	As 7400, This Book
7438	Quad 2-Input NAND (Buffer)		OCO	As 7400, This Book
7439	Quad 2-Input NAND (Buffer)		OCO	As 7401, This Book
7440	Dual 4-Input NAND (Buffer)		TPO	As 7420, This Book
7441	BCD-to-Decimal Decoder/Driver			This Book
7442	BCD-to-Decimal Decoder Output low for address high)			This Book
7443	Excess 3-to-Decimal Decoder			SIG 2-46
7444	Excess 3-Gray-to-Decimal Decoder			SIG 2-48
7445	BCD-to-Decimal Decoder/Driver		OCO	As 7442, This Book
7446	BCD-to-Seven Segment Decoder Driver		OCO 30V	This Book
7447	BCD-to-Seven Segment Decoder Driver		OCO 15V	As 7446, This Book

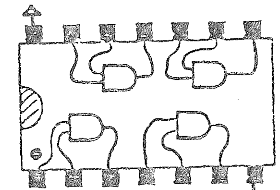

7400

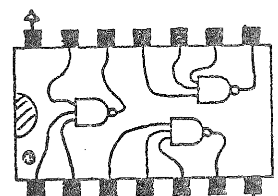

7401

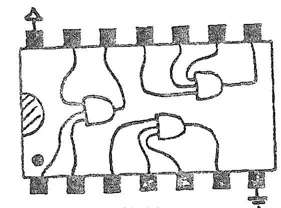

7402

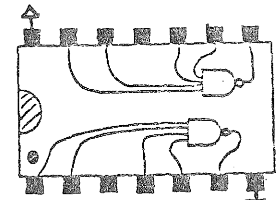

7404

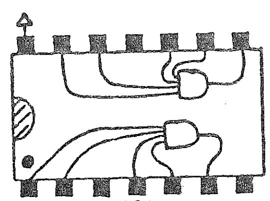

7407

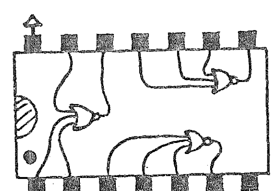

7408

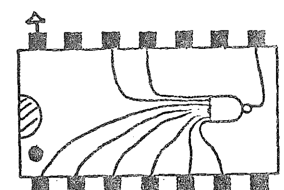

7410

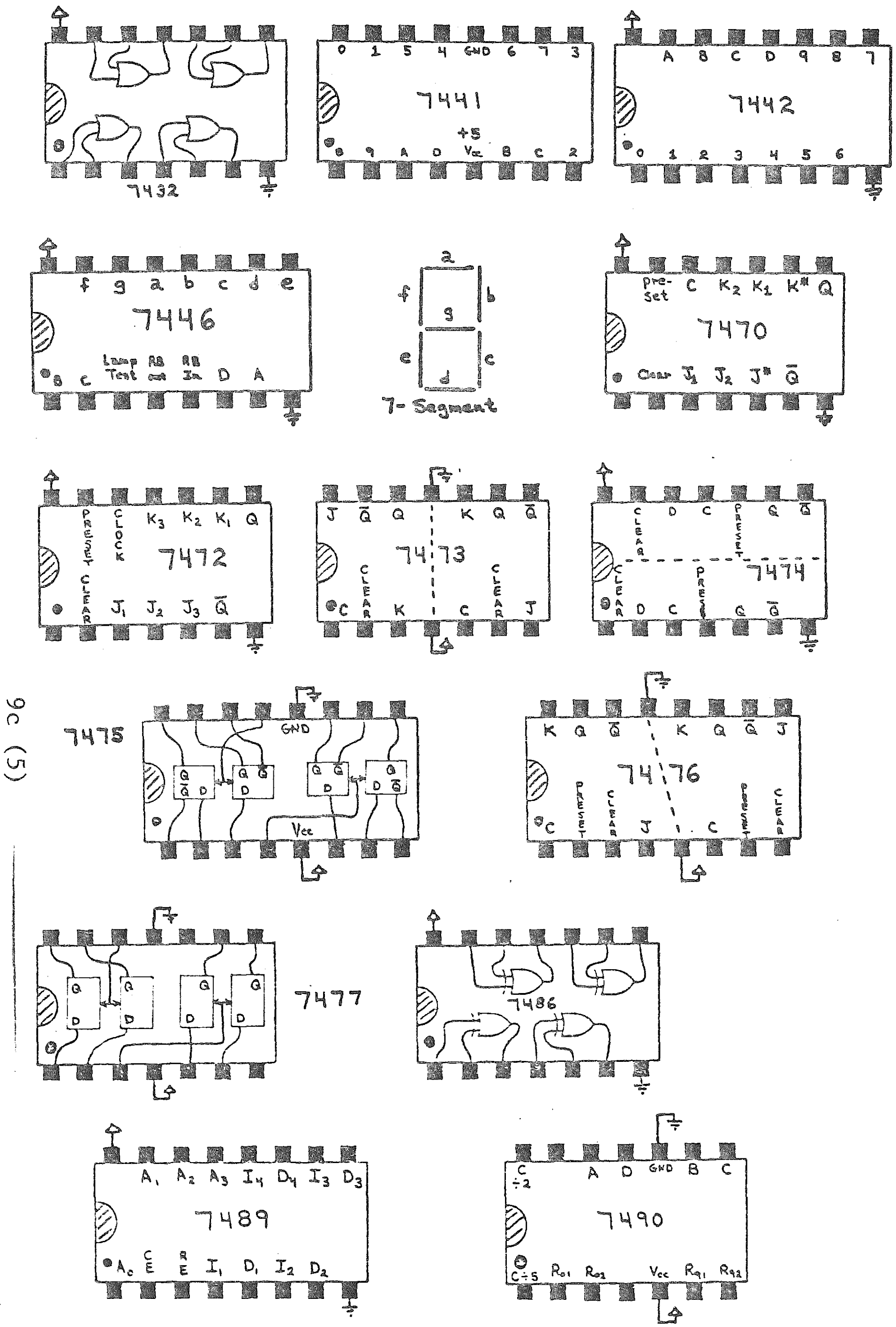

7411


7420

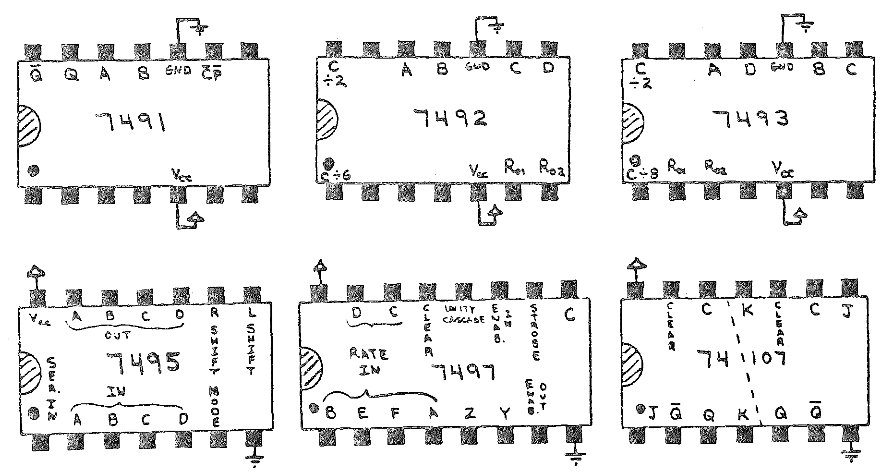

7421


7427

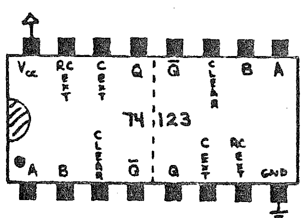

7430

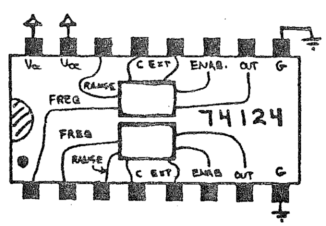


7448	BCD-to-Seven Segment Decoder Driver (LED Drive)		As 7446, This Book
7449	BCD-to-Seven Segment Decoder Driver		TI-2, pg. S-96
7450	Expandable Dual 2-Wide 2-Input AND-OR-Invert Gate	TPO	SIG. Pg. 2-60
7451	Expandable Dual 2-Wide 2-Input AND-OR-Invert Gate		SIG. Pg. 2-60
7452	Expandable 4-Wide AND-OR-Gate		TI-1, Pg. 71
7453	4-Wide 2-Input AND-OR-Invert Gate		SIG. Pg. 2-62
7454	4-Wide 2-Input AND-OR-Invert Gate		SIG. Pg. 2-62
7455	2-Wide 4-Input AND-OR-Invert Gate		TI-1, Pg. 73
7460	Dual 4-Input Expander		SIG. Pg. 2-64
7461	Triple 3-Input Expander		TI-1, Pg. 73
7462	4-Wide AND-OR Expander		TI-1, Pg. 74
7464	4-2-3-2-Input AND-OR-Invert Gates	TPO	TI-1, Pg. 74
7465	4-2-3-2-Input AND-OR-Invert Gates	OCO	TI-1, Pg. 74
7470	AND Gated JK Flip-Flop		This Book
7471	AND-OR Gated J-K MS Flip-Flop with Preset		TI-1, Pg. 75
7472	AND-Gated J-K MS Flip-Flop with Preset and Clear	TPO	This Book
7473	Dual J-K MS Flip-Flop with Clear	TPO	This Book
7474	Dual D-Type Edge Triggered Flip-Flop	TPO	This Book
7475	Quad Bistable Latch		This Book
7476	Dual J-K MS Flip-Flop with Preset and Clear	TPO	This Book
7477	Quad Bistable Latch		This Book
7478	Dual J-K Flip-Flop with Preset and Clear(same CLK)		TI-1, Pg. 77
7480	Gated Full Adder		SIG. Pg. 2-83
7481	16 Bit Active Element Memory		TI-1, Pg. 190
7482	2-Bit Binary Full Adder		TI-1, Pg. 195
7483	4-Bit Binart Full Adder		SIG. Pg. 2-85
7484	16-Bit Active Element Memory		TI-1, Pg. 190
7485	4-Bit Magnitude Comparators		SIG. Pg. 2-87
7486	Quad 2-Input Exclusive-OR Gate		This Book
7487	4-Bit True/Complement Zero/One Element		TI-1, Pg. 214
7488	256-Bit Read Only Memory		SIG. Pg. 2-92
7489	256-Bit Read/Write Memory		This Book
7490	Decade Counter		This Book
7491	8-Bit Shift Register		This Book
7492	Divide-by-12 Counter		This Book
7493	Divide-by-16, 4-Bit Binary Counter		This Book
7494	4-Bit Shift Register (Parallel In, Serial Out)		SIG. Pg. 2-104
7495	4-Bit Right-Shift/Left Shift Register		This Book
7496	5-Bit Shift Register		SIG. Pg. 2-109
7497	Synchronous 6-Bit Binary Rate Multiplier		This Book
7498	4-Bit Data Selectors/Storage Register		TI-1, Pg. 253
7499	4-Bit Right-Shift/Left-Shift Register		TI-1, Pg. 257
74100	4-Bit Bistable Latch		SIG. Pg. 2-111

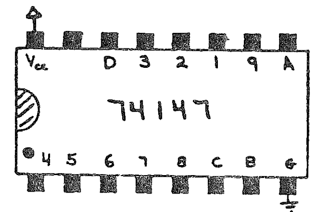


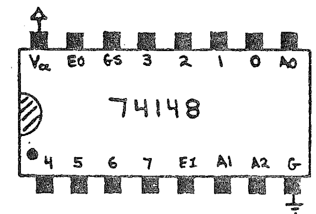
74101	AND-OR-Gated J-K Negative Edge Triggered Flip-Flop		TI-1, Pg. 78
74102	AND Gated J-K Negative Edge Triggered Flip-Flop		TI-1, Pg. 78
74103	Dual J-K Negative Edge Triggered Flip-Flop		TI-1, Pg. 78
74106	Dual J-K Negative Edge Triggered Flip-Flop		TI-1, Pg. 79
74107	Dual J-K MS Flip Flop (7473 with different base)		This Book
74108	Dual J-K Negative Edge Triggered Flip-Flop		TI-1, Pg. 79
74109	Dual J-K Positive Edge Triggered Flip-Flop		TI-1, Pg. 80
74110	AND Gated J-K MS Flip-Flop with Data Lockout		TI-1, Pg. 80
74111	Dual J-K MS Flip-Flop with Data Lockout		TI-1, Pg. 80
74112	Dual J-K Negative Edge Triggered Flip-Flop		TI-1, Pg. 81
74113	Dual J-K Negative Edge Triggered Flip-Flop		TI-1, Pg. 81
74114	Dual J-K Negative Edge Triggered Flip-Flop		TI-1, Pg. 81
74116	Dual 4-Bit Latch with Clear		TI-1, Pg. 261
74120	Dual Pulse Synchronizer/Driver		TI-1, Pg. 265
74121	Monostable Multivibrator		This Book, Pg. 3e (2)
74122	Retriggerable Monostable Multivibrator with Clear		This Book, Pg. 3e (2)
74123	Dual Retriggerable Monostable Multivibrator		This Book
74124	Dual Voltage Controlled Clock		This Book, TI-2, Pg. S-62
74125	Quad Bus Buffer Gates	TSO	SIG. Pg. 2-125
74126	Quad Bus Buffer Gates (Inverted Control)	TSO	SIG. Pg. 2-125
74128	50 Ohm Line Driver		TI-1, Pg. 83
74132	Quad Schmitt Trigger NAND		As 7400, This Book
74133	13-In NAND Gate		TI-1, Pg. 84
74134	12-In NAND Gate	TSO	TI-1, Pg. 84
74135	Quad EX-OR/NOR Gate		TI-1, Pg 269
74136	Quad 2-In EX-OR	OCO	As 7486, This Book
74138	3-To-8 Line Decoder		TI-1, Pg. 274
74139	Dual 2-to-4 Line Decoder		TI-1, Pg. 274
74140	Dual 4-In NAND 50 Ohm Driver		As 7420, This Book
74141	BCD to Decimal Decoder/Driver		As 7441, This Book
74142	BCD Counter/4-Bit Latch/BCD Decoder/Driver		TI-1, Pg. 281
74143	4-Bit Counter/Latch/7-Segment LED Driver		TI-1, Pg. 283
74144	4-Bit Counter/Latch/7-Segment Lamp Driver	OCO	TI-1, Pg. 283
74145	BCD-to-Decimal Decoder Driver (80 Ma)		As 7442, This Book
74147	10 Line to 4 Line Priority Encoder		This Book
74148	8 Line to 3 Line Priority Encoder		This Book
74150	16 Line to 1 Line Data Selectro/Multiplexer		This Book
74151	8 Line to 1 Line Data Selector/Multiplexer		This Book
74152	8 Line to 1 Line Data Selector/Multiplexer		SIG. Pg. 2-139
74153	Dual 4 Line to 1 Line Data Selector/Multiplexer		SIG. Pg. 2-142
74154	4 Line to 16 Line Decoder/Demultiplexer	TPO	This Book
74155	Dual 2 Line to 4 Line Decoder/Demultiplexer	TPO	SIG. Pg. 2-146
74156	Dual 2 Line to 4 Line Decoder/Demultiplexer	OCO	SIG. Pg. 2-146
74157	Quad 2 Input Data Selector/Multiplexer		SIG. Pg. 2-150
74158	Quad 2 Input Data Selector/Multiplexer (Inverted)		SIG. Pg. 2-150
74159	4-Line to 16 Line Decoder/Demultiplexer	OCO	TI-1, Pg. 323
74160	Synchronous 4-Bit Counter		SIG. Pg. 2-152
74161	Synchronous 4-Bit Counter		SIG. Pg. 2-152
74162	Synchronous 4-Bit Counter		SIG. Pg. 2-152
74163	Synchronous 4-Bit Counter		SIG. Pg. 2-152
74164	8-Bit Parallel Out Serial Shift Register		This Book
74165	Parallel Load 8-Bit Shift Register		SIG. Pg. 2-161

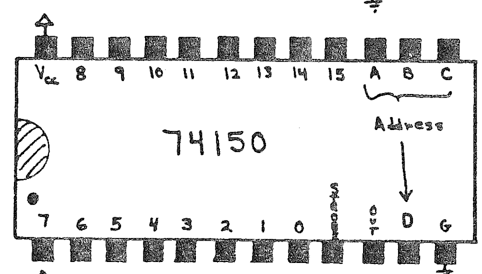


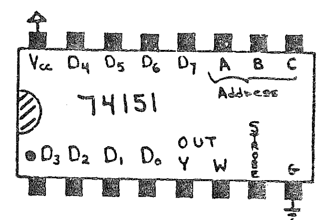


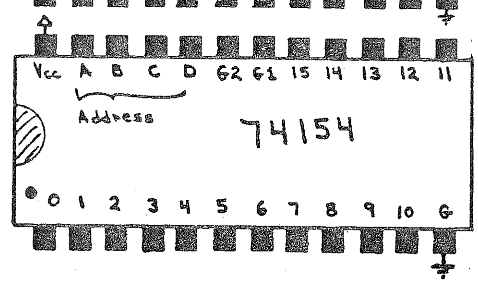
74166	8-Bit Shift Register		SIG. Pg. 2-163
74167	Synchronous Decade Rate Multiplier		This Book
74168	Synchronous 4-Bit Up/Down Counter		TI-2, Pg. S-193
74169	Synchronous 4-Bit Up/Down Counter		TI-2, Pg. S-193
74170	4 by 4 Register File		SIG. Pg. 2-166
74172	16-Bit Multiple Port Register		SIG. Pg. 2-169
74173	4 Bit D Type Register	TSO	TI-1, Pg. 361
74174	Hex D-Flip-Flop with Clear		This Book
74175	Quad D-Type Edge Triggered Flip-Flop		This Book
74176	35 MHz Presettable Decade Counter/Latch		TI-1, Pg. 369
74177	35 MHz Presettable Binary Counter/Latch		TI-1, Pg. 369
74178	4-Bit Parallel Access Shift Register		TI-1, Pg. 375
74179	4-Bit Parallel Access Shift Register (Inverted out)		TI-1, Pg. 375
74180	8 Bit Odd/Even Paraty Generator/Checker		SIG. 2-178
74181	High Speed Arithmetic Unit		SIG. Pg. 2-180
74182	Look Ahead Carry Generator		SIG. Pg. 2-184
74183	Dual Carry-Save Full Adder		TI-1, Pg. 396
74184	BCD-to-Binary Converter		TI-1, Pg. 398
74185	Binary-to-BCD Converter		TI-1, Pg. 398
74186	512 Bit Programmable ROM		TI-1, Pg. 405
74187	1024 Bit ROM (256 by 4)		TI-1, Pg. 410
74188	256 Bit Programmable ROM		TI-1, Pg. 414
74189	64 Bit Random Access Memory	TSO	TI-2, Pg. S-211
74190	Synchronous Up/Down Counter, BCD		This Book
74191	Synchronous Up/Down Counter, 4-Bit Binary		This Book
74192	Synchronous Up/Down Counter (BCD)		This Book
74193	Synchronous Up/Down Counter, 4-Bit		This Book

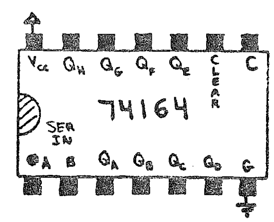


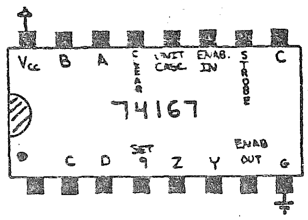


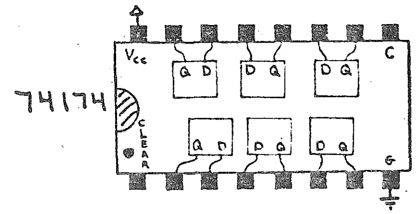


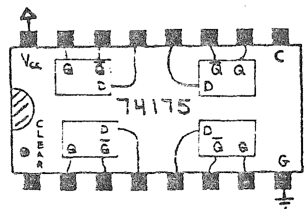


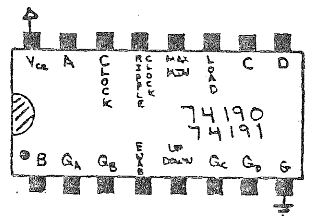


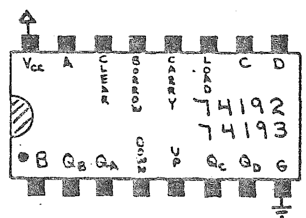


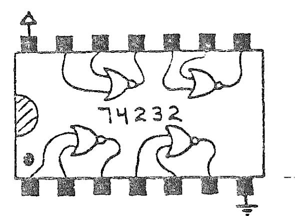












9C (7)

74194	4-Bit Bidirectional Universal Shift Register		SIG. Pg. 2-199
74195	4-Bit Parallel Access Shift Register		SIG. Pg. 2-201
74196	30 MHz Presettable Decade Counter/Latch		TI-1, Pg. 451
74197	30 MHz Presettable Binary Counter Latch		TI-1, Pg. 451
74198	8 Bit Shift Register		SIG. Pg. 2-203
74199	8 Bit Shift Register		SIG. Pg. 2-205
74200	256 Bit Read/Write Memory	TSO	TI-1, Pg. 463
74201	256 Bit Random Access Memory	TSO	TI-2, Pg. S-231
74206	256 Bit Read/Write Memory	OCO	TI-1, Pg. 470
74221	Dual Monostable Multivibrator with Schmitt Trigger		TI-2, Pg. S-69
74232	Quad NOR Schmitt Trigger		This Book
74246	BCD-to-7 Segment Decoder/Driver	OCO	TI-2, Pg. S-233
74247	BCD-to-7 Segment Decoder/Driver	OSO	TI-2, Pg. S-233
47248	BCD-to-7 Segment Decoder/Driver		TI-2, Pg. S-233
74249	BCD-to-7 Segment Decoder/Driver	OSO	TI-2, Pg. S-233
74251	Data Selector/Multiplexer	TSO	TI-1, Pg. 473
74253	Dual 4-Line to 1-Line Data Selector	TSO	TI-1, Pg. 480
74257	Quad 2-Line to 1-Line Data Selector/Multiplexer	TSO	TI-1, Pg. 483
74258	Quad 2-Line to 1-Line Data Selector/Multiplexer	TSO	TI-1, Pg. 483
74260	Dual 5-Input NOR Gates		TI-1, Pg. 84
74261	2-Bit by 4-Bit Parallel Binary Multiplier		TI-2, Pg. S-248
74265	Quad Complimentary Output Elements		TI-2, Pg. S-77
74266	Quad 2-Input EX-NOR Gates	OCO	TI-1, Pg. 486
74270	2048-Bit Read-Only Memory	OCO	TI-2, Pg. S-254
74271	2048-Bit Read-Only Memory	OCO	TI-2, Pg. S-254
74273	Octal D-Type Flip-Flop with Clear		TI-2, Pg. S-260
74274	4-Bit By 4-Bit Binary Multiplier	TSO	TI-2, Pg. S-260
74275	7-Bit Slice Wallace Trees	TSO	TI-2, Pg. S-260
74278	4-Bit Cascadable Priority Encoder		TI-1, Pg. 488
74279	Quad S-R Latches		SIG. Pg. 2-209
74280	9-Bit Odd/Even Parity Generator/Checker		TI-1, Pg. 488
74281	4-Bit Parallel Binary Accumulator		TI-2, Pg. S-271
74283	4-Bit Binary Full Adder with Fast Carry		TI-1, Pg. 494
74285	4-Bit by 4-Bit Parallel Binary Multiplier		TI-1, Pg. 496
74287	1024 Bit Programmable ROM	TSO	TI-2, Pg. S-279
74289	64-Bit RAM	OCO	TI-2, Pg. S-283
74290	Decade Counter		TI-1, Pg. 499
74293	Binary Counter		TI-1, Pg. 499
74295	4-Bit Right-Shift Left-Shift Register	TSO	TI-1, Pg. 502
74298	Quad 2 Input Multiplex with Storage		TI-1, Pg. 505
74299	8-Bit Universal Shift/Storage Register		TI-2, Pg. S-301
74351	Dual Data Selector/Multiplexer	TSO	TI-2, Pg. S-305
74365	Hex Bus Driver	TSO	TI-2, Pg. S-85
74366	Hex Bus Driver	TSO	TI-2, Pg. S-85
74367	Hex Bus Driver	TSO	TI-2, Pg. S-85
74368	Hex Bus Driver	TSO	TI-2, Pg. S-85
74370	2048 Bit ROM	TSO	TI-2, Pg. S-309
74371	2048 Bit ROM	TSO	TI-2, Pg. S-309
74381	Arithmetic Logic Unit/Function Generator		TI-2, Pg. S-313
74386	Quad 2-Input EX-OR Gate		TI-2, Pg. S-315
74387	1024 Bit PROM		TI-2, Pg. S-317
74393	Dual 4-Bit Binary Counter		TI-2, Pg. S-321
74395	4-Bit Cascadable Shift Register	TSO	TI-2, Pg. S-325
74490	Dual 4-Bit Decade Counter		TI-2, Pg. S-328
74670	4 by 4 Register File	TSO	TI-2, Pg. S-332

74C00 CMOS

The 74C00 CMOS series are functional equivalents of the 7400 TTL series. Available 74C00 CMOS IC's are:

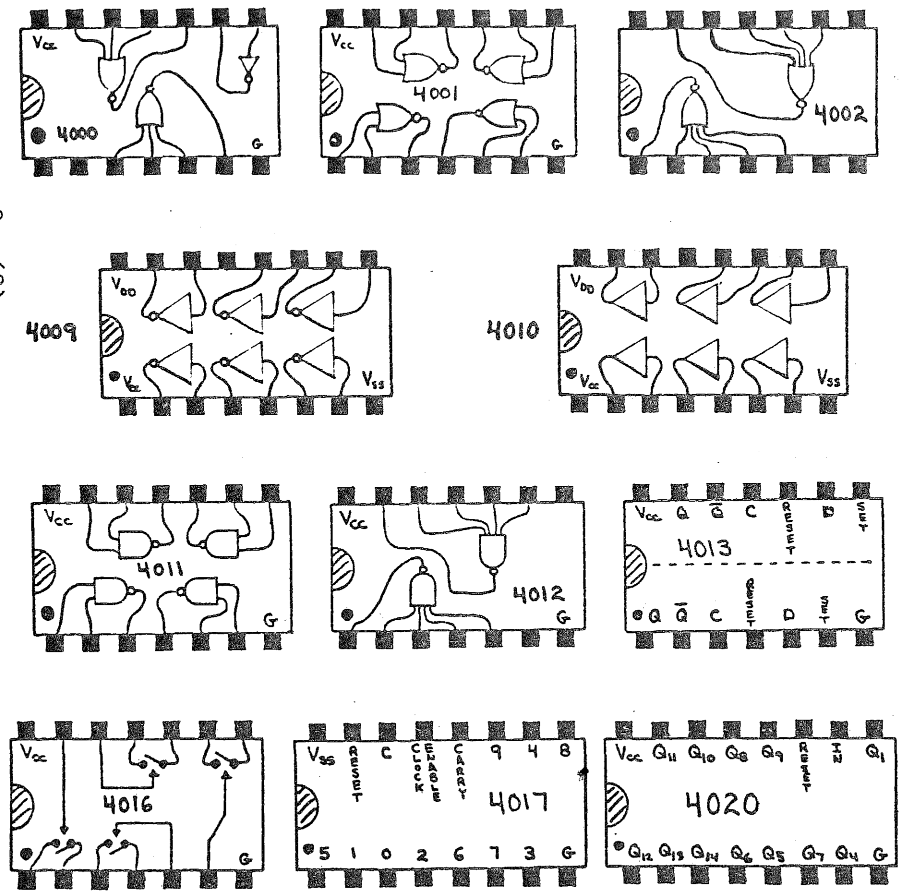
74C00	74C48	74C95	74C165
74C02	74C73	74C107	74C173
74C04	74C74	74C151	74C174
74C08	74C76	74C154	74C175
74C10	74C83	74C157	74C192
74C14	74C85	74C160	74C193
74C20	74C86	74C161	74C195
74C30	74C89	74C162	74C200
74C32	74C90	74C163	74C221
74C42	74C93	74C164	

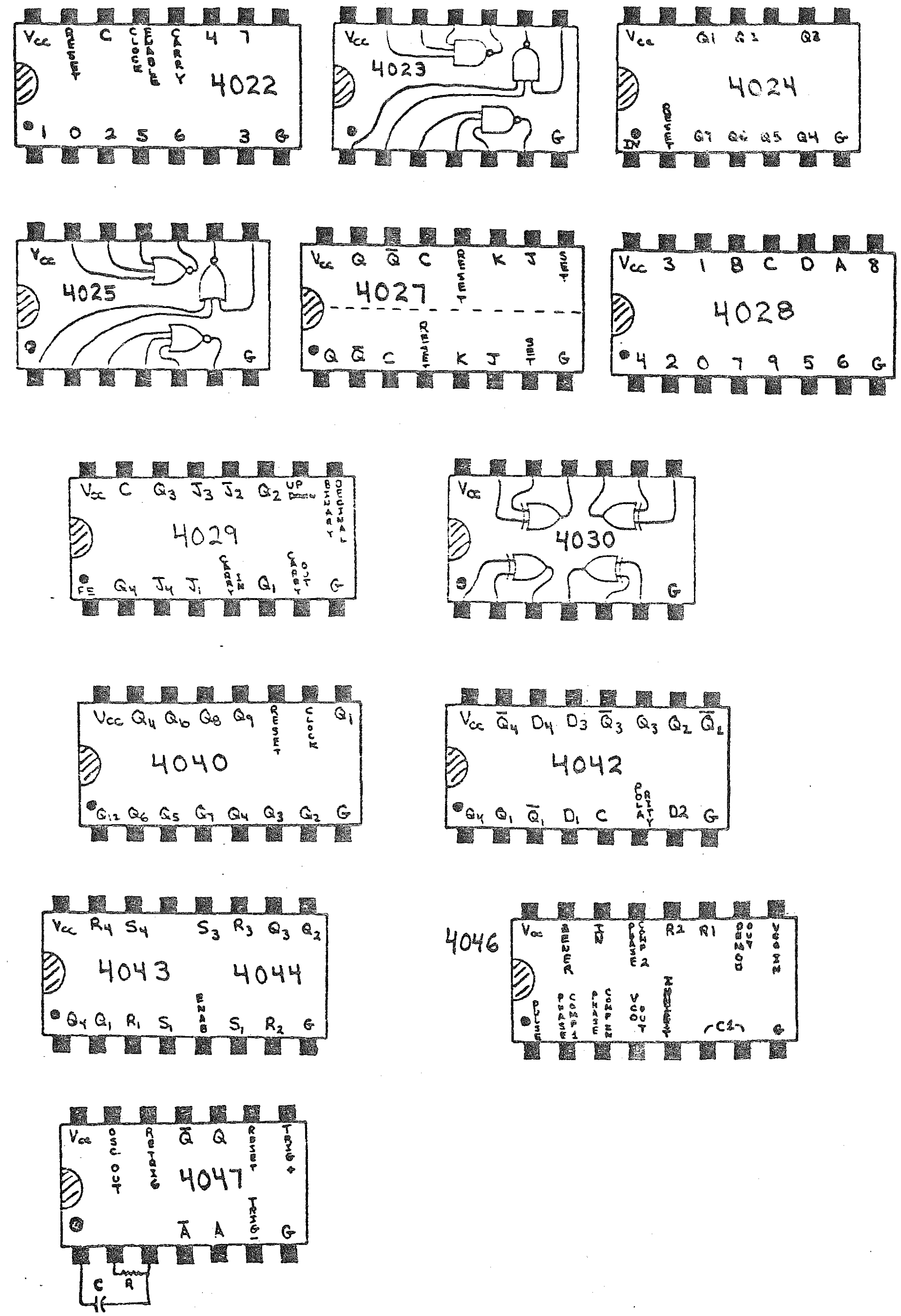
In addition, there is a 74C900 series of which the 74C901 (Hex inverting TTL Buffer) and the 74C902 (Hex Non-Inverting TTL Buffer) are of particular interest.

4000 Series CMOS

4000	Dual 3-Input NOR Gate plus Inverter	This Book
4001	Quad 2-Input NOR Gate	This Book
4002	Dual 4-Input NOR Gate	This Book
4006	18 Stage Static Shift Register	NAT Pg. 126
4007	Dual Complementary Pait Plus Inverter	NAT Pg. 129
4009	Hex Buffer (Inverting)	This Book
4010	Hex Buffer (Non-Inverting)	This Book
4011	Quad 2-Input NAND Gate	This Book
4012	Dual 4-Input NAND Gate	This Book
4013	Dual D Flip-Flop	This Book
4014	8 Stage Static Shift Register	NAT Pg. 139
4015	Dual 4-Bit Static Register	NAT Pg. 142
4016	Quad Bilateral Switch (Analog Switch)	This Book
4017	Divide-by-10 Counter/Divider	This Book
4018	Presettable Divide-by-N Counter (N= 2 to 10)	RCA
4019	Quad AND/OR Select Gate	RCA, NAT Pg. 154
4020	14-Stage Ripple-Carry Binary Counter/Divider	This Book
4021	8-Stage Static Shift Register	NAT Pg. 159
4022	Divide-by-8 Counter/Divider/Decoder	This Book
4023	Triple 3-Input NAND Gate	This Book
4024	7-Stage Ripple-Carry Counter/Divider	This Book
4025	Triple 3-Input NOR Gate	This Book
4027	Dual J-K MS Flip-Flop	This Book
4028	BCD-to-Decimal Decoder	This Book
4029	Presettable Binary/Decade Up/Down Counter	This Book
4030	Quad EXclusive-OR Gate	This Book

4031	64 Stage Shift Register	RCA
4033	Decade Counter/Divider with 7-Segment Display Outputs	RCA
4034	8-Stage Bidirectional Shift Register	RCA
4035	4-Bit Parallel In/Parallel Out Shift Register	NAT Pg. 186
4036	4 Word by 8-Bit Memory	RCA
4037	Triple AND-OR Bi-Phase Pairs	RCA
4039	4 Word by 8-Bit Memory	RCA
4040	12 Stage Ripple-Carry Binary Counter/Divider	This Book
4041	Qual True/Complement Buffer	RCA
4042	Quad Clocked D Latch	This Book
4043	Quad NOR R/S Latch	This Book
4044	Quad NAND R/S Latch	This Book
4045	21-Stage Clock Timer	RCA
4046	Phase-Locked Loop	This Book
4047	Monostable/Astabel Multivibrator	This Book
4048	Multifunction Expandable 8-Input Gate	RCA
4049	Hex Buffer (Inverting)	As 4009, pin 16 nc
4050	Hex Buffer (Non-Inverting)	As 4010, pin 16 nc
4051	8-Channel Multiplexer	RCA
4052	Differential 4-Channel Multiplexer	RCA
4053	Triple 2-Channel Multiplexer	RCA
4054	4-Line Liquid Crystal Driver	RCA
4055	Single Digit Liquid Crystal Driver	RCA
4056	Single Digit Liquid Crystal Driver	RCA
4061	256-Word by 1 Bit Static RAM	RCA
4066	Quad Bilateral Switch (on resistance 80 ohms)	NAT Pg. 199, As 4016





CHAPTER 9D

REPRINT PAPERS

CONTENTS:

INTRODUCTION

This section is intended mainly so that the individual can insert reprint papers of his own choosing. Two papers are being furnished. The first describes an additive synthesis system and is a reference for chapter 2b. The second is concerned with the generation of sine waves by Walsh functions, but has some calculations in it that should be useful for anyone working with systems that generate sine waves by a digital method. The paper was available at one time as a lending paper L-066.

As a guide for what might be inserted here (there is only a limited amount of room), the reader can consult the additional references listed in the next chapter, but the main guide should be the reader's own needs.

Experimental Electronic Music Devices Employing Walsh Functions

BERNARD A. HUTCHINS, JR.

Ithaca, N. Y. 14850

The complete orthonormal set of Walsh functions is used to generate periodic waveforms and envelope shapes for an additive synthesis electronic music device. The Walsh functions, easily produced by digital circuitry, can be used to generate banks of harmonic and nonharmonic waveforms. A second Walsh function generator forms the basis of a digital envelope controller which can produce a wide variety of simultaneous envelope shapes.

INTRODUCTION: Although investigated by J. L. Walsh in 1923 [1], the set of functions which now bear his name have not found wide application until recent years [2], [3], [4]. Walsh waveforms are rectangular, taking on only the values ± 1 over a basic interval, after which the sequence may be repeated to form a set of periodic functions. Walsh functions form a complete orthonormal set and, therefore, can be employed in waveform synthesis schemes analogous to the Fourier synthesis methods which employ sines and cosines. The ± 1 levels are easily converted to the zero and one levels of digital logic, and numerous schemes for the generation of Walsh functions by digital means have been suggested [5], [6].

PROPERTIES AND GENERATION METHODS

Walsh functions indexed from zero to $2^m - 1$ are defined on a basic interval, such as zero to one, which is subdivided into 2^m equal segments, where m is an integer. The functions indexed by $2^t - 1$, where t is an integer less than or equal to m, are known as Rademacher functions, and are actually a set of square waves in octaves, starting at $+1$, and repeating a total of 2^{t-1} times in the basic interval. The remaining Walsh functions can be generated from the recursion relation:

$$Wal(h) \cdot Wal(k) = Wal(h \oplus k)$$

where the notation Wal(j) denotes the Walsh function of index j, the symbol ($\oplus$) represents modulo-2 addition ($0 \oplus 0 = 0$, $0 \oplus 1 = 1$, $1 \oplus 0 = 1$, and $1 \oplus 1 = 0$), and h and k are represented by their binary equivalents. After converting the Walsh functions to the zero and one logic levels, the indicated multiplication ($\cdot$) in the recursion relation reduces to modulo-2 addition [5] (the logical "EXCLUSIVE-OR" function), suggesting a hardware generation scheme as indicated in Fig. 1 for m = 3. $Wal(0)$ is a constant function. Extension of the generation scheme for larger m is straightforward.

COMPUTER GENERATION OF HIGHER ORDERS

Generation of Walsh functions of higher index (also referred to as higher "sequency" as defined below) is facilitated by a computer program employing the algorithm:

Step 1: Generation of the square waves in positions $2^t - 1$:

Repeat for $k = 1, 2, 3, \ldots\, m$
  Define $L = 2^{m-k}$
    Repeat for $r = 1, 2, 3, \ldots 2^m$
      $p = r/L - 1/2^m$
      If (Highest integer in p)/2 = (An integer)
          Then $W(2^k - 1, r) = 1$
          Else $W(2^k - 1, r) = 0$
    End
End

Step 2: Recursion Relation Implementation

Repeat for $k = 1, 2, 3, \ldots m - 1$
  Define $S = 2^{k+1} - 2^k - 1$
    Repeat for $q = 1, 2, 3, \ldots S$
      Repeat for $r = 1, 2, 3, \ldots 2^m$
        $W(2^{k+1} - q - 1, r) = 1$ if $W(2^{k+1} - 1, r) \neq W(q, r)$
        $= 0$ otherwise
      End
    End
End

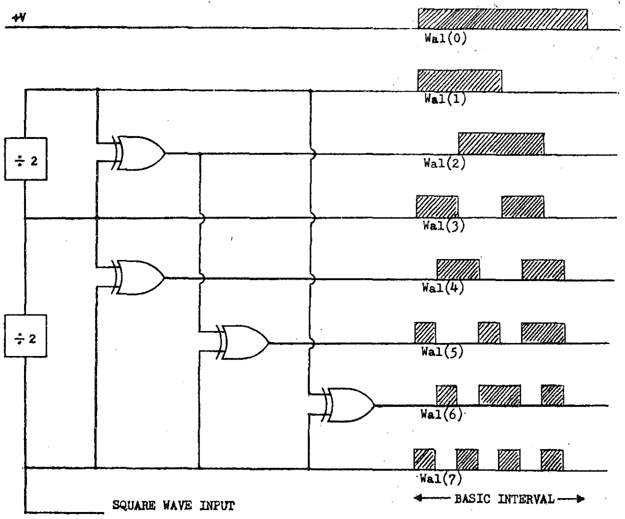


Fig. 1. Digital hardware generation of first eight Walsh functions.

A 2^m by 2^m matrix notation, where the rows of the matrix are Walsh functions in sequency order, is useful. A computer generated Walsh function matrix (W) for $m = 5$ is shown in Fig. 2.

SAL, CAL, AND SEQUENCY

Examination of the matrix in Fig. 2 indicates that successive odd and even indexed Walsh functions are shifted versions of the same sequence, and the notation $Sal(i) = Wal(2i-1)$ and $Cal(i) = Wal(2i)$ is often used in analogy with the sine and cosine notation. The concept of frequency yields to the terminology "sequency," defined as one-half the average number of zero crossings per second (zps) [7]. The ordering of the Walsh functions in sequency order can be regarded in terms of a time-normalized sequency, where the Walsh functions are defined on an open interval so that one does not count the zero crossings at either end of the interval. Corresponding Sal and Cal functions have the same zps, and when regarded as periodic functions, they can be thought of as differing only by a time delay.

WALSH-FOURIER SERIES AND DISCRETE WALSH TRANSFORM

By analogy with the standard Fourier series employing sines and cosines, a corresponding Walsh-Fourier series can be defined [8]. Using a running variable x, the series is:

$$F(x) = \sum_{n=0}^{\infty} C_n \, Wal(n,x)$$

where

$$C_n = \int_0^1 F(x) \, Wal(n,x) \, dx.$$

Since $Wal(n,x)$ takes on only the values ± 1, it not only breaks up the integral into several subintervals of integration, but also effectively moves outside the integral sign. For a Fourier series, it would be necessary to integrate a sine or a cosine times $F(x)$ to obtain the coefficients, but for the Walsh-Fourier series, it is only necessary to be able to integrate $F(x)$. While a smooth curve will never be represented completely by a finite series of Walsh functions, a small amount of low-pass filtering is usually sufficient to remove the sharp corners of the composite waveform. Moreover, many rectangular functions such as sequences and periodically sampled analog signals can be represented exactly by a finite series of appropriately timed Walsh functions. Walsh-Fourier coefficients of such discrete sequences are obtainable as a matrix product $C = WX$, where C is a 2^m dimensioned row vector of Walsh coefficients, W is the 2^m by 2^m Walsh function matrix, as in Fig. 2 for $m = 5$, and X is a 2^m dimensioned column vector of discrete samples. This is referred to as a Discrete Walsh Transform (DWT).

FAST WALSH TRANSFORM

More economical use of computer time can be made by employing the Fast Walsh Transform (FWT) technique, derived from the Hadamard Transform and the Fast Fourier Transform (FFT) techniques [9]. The FWT follows the basic Cooley-Tukey algorithm for the FFT [10], but avoids operations with complex numbers. The FWT transforms N time-sampled data points into N discrete spectrum points. In the case where the N samples constitute a periodic waveform, or are the best approximation to a periodic waveform one could expect from N samples, the spectral points are the Walsh-Fourier coefficients that would be obtained from the DWT. The reduction in the number of computer operations is from approximately N^2 for the DWT to approximately $N \log_2 N$ for the FWT, and the matrix of Walsh functions need not be generated at all. A flow graph for the Fast Hadamard Transform (FHT) as it fits into the overall analysis and synthesis process is shown in Fig. 3 for the case of $m = 3$. The FHT yields the Walsh coefficients C_n, but in a scrambled order. The FWT is obtained by simply rearranging the coefficients in sequency order, the scrambled order being generated by the scheme shown in Table 1.
The scrambled order gives the sequencies of the rows of the Hadamard matrix. Both the flow graph and the ordering schemes are easily extended to larger order.

The FFT and FWT methods suggest possible realiza-

Wal(0)	11111111111111111111111111111111
Wal(1)	11111111111111110000000000000000
Wal(2)	00000000111111111111111100000000
Wal(3)	11111111000000001111111100000000
Wal(4)	00001111111100000000111111110000
Wal(5)	11110000000011110000111111110000
Wal(6)	00001111000011111111000011110000
Wal(7)	11110000111100001111000011110000
Wal(8)	00111100001111000011110000111100
Wal(9)	11000011110000110011110000111100
Wal(10)	00111100110000111100001100111100
Wal(11)	11000011001111001100001100111100
Wal(12)	00110011110011000011001111001100
Wal(13)	11001100001100110011001111001100
Wal(14)	00110011001100111100110011001100
Wal(15)	11001100110011001100110011001100
Wal(16)	01100110011001100110011001100110
Wal(17)	10011001100110010110011001100110
Wal(18)	01100110100110011001100101100110
Wal(19)	10011001011001101001100101100110
Wal(20)	01101001100101100110100110010110
Wal(21)	10010110011010010110100110010110
Wal(22)	01101001011010011001011010010110
Wal(23)	10010110100101101001011010010110
Wal(24)	01011010010110100101101001011010
Wal(25)	10100101101001010101101001011010
Wal(26)	01011010101001011010010101011010
Wal(27)	10100101010110101010010101011010
Wal(28)	01010101101010100101010110101010
Wal(29)	10101010010101010101010110101010
Wal(30)	01010101010101011010101010101010
Wal(31)	10101010101010101010101010101010

Fig. 2. First 32 Walsh functions generated by computer.

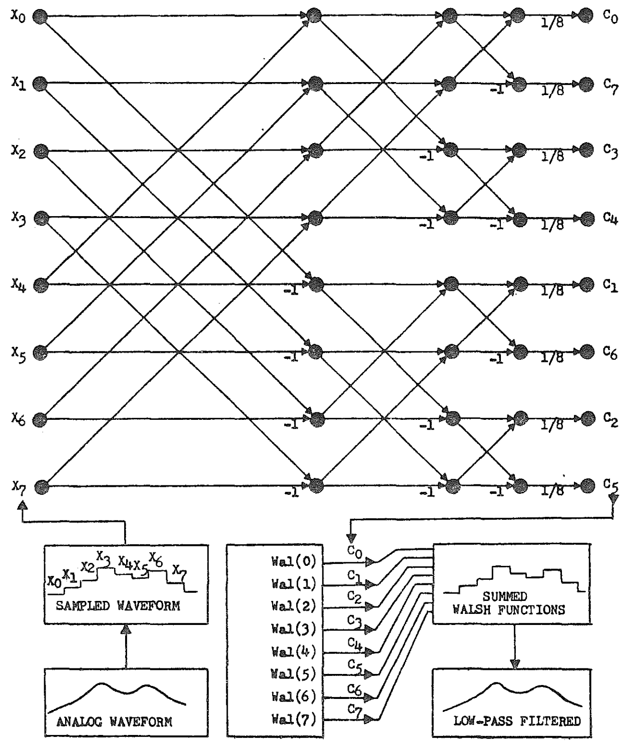


Fig. 3. Overall analysis-synthesis process showing role of Hadamard transform.

tions of polyphonic instruments, filters, and timbre controllers by feeding or controlling spectral information to the inverse transforms by means of some device such as a keyboard. The outputs of the device could be time-ordered for an audible output, and the input spectral points, roughly representing pitch information, could be enveloped "in" at the inputs.

APPLICATIONS

Periodic Waveform Generation

Using the 32 by 32 matrix of Walsh functions, as defined by Fig. 2, Walsh coefficients can be generated by DWT or FWT computer methods. Coefficients for some common waveforms are shown in Fig. 4a. Simple operational amplifier summation techniques are used to sum the appropriate Walsh functions in proportion to their coefficients. Although sequency is not correlated with subjective musical pitch, the composite waveforms have a pitch determined by the basic interval of the Walsh functions, unless the composite waveform is specifically made to repeat more than once in the basic interval. Oscilloscope traces of the Walsh generated sine, sawtooth, and triangle approximations along with the trace of the waveforms after passing through a low-pass filter are shown in Fig. 4b, Fig. 4c, and Fig. 4d respectively. In addition to common waveforms, a wide variety of complex periodic waveforms can be easily obtained. These, while interesting on an oscilloscope face, are no more interesting to the ear than the sawtooth; for example, all periodic waveforms are approximately equally boring to the ear. Low-pass filtering by fixed filters can be used to smooth the synthesized waveforms over moderate ranges of frequency. Tunable or voltage-controlled filters (VCF's) can also be employed.

WALSH HARMONIC BANK

Although the periodic waveforms are not usable directly for music synthesis (unless used with voltage-controlled amplifiers (VCA's), VCF's, etc., in a typical subtractive synthesis system), the Walsh functions can be used as a source of separated (albeit Walsh) harmonics for additive synthesis. The question of the audibility of the relative phase of the Walsh harmonic components of a composite waveform arises here; and, in general, one must allow for a different timbre depending on whether a Sal or corresponding Cal of the same sequency is employed. This is connected with the problem of monaural phase [11], and while there are cases where the phase difference is not important to tone color, it is relatively easy to devise cases where there is a great difference that can be heard when played into an open room as well as over headphones. In cases where it is possible to use less than the full set of 32 Walsh functions, the number of EXCLUSIVE-OR gates needed for the generator can be reduced (from 26 to 18) if only the Sal functions are required, and further reduced (from 18 to 11) if only one function for each sequency (zps) is to be used.

GENERATION OF ENVELOPES

Walsh functions can be used to generate envelope shapes as well as fully periodic waveforms. In this case, a suitable periodic waveform is generated over the basic interval, and can be further altered by a predetermined delay point. An envelope control circuit advances a Walsh function generator from the first segment to a predetermined stopping point, defining the attack and the sustain respectively. A restart of the generator and advance to the last segment defines the decay. Examples of two such envelopes for the $m = 5$ set of Walsh functions (32 segments) are shown in Fig. 5 for a delay at segment 16.

THE BLOCK PULSES $P(16)$ and $P(32)$

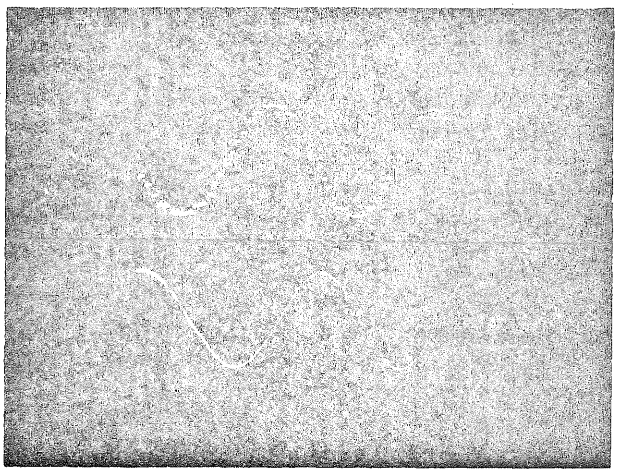
In the generation of envelopes, the segment at 16

Table 1

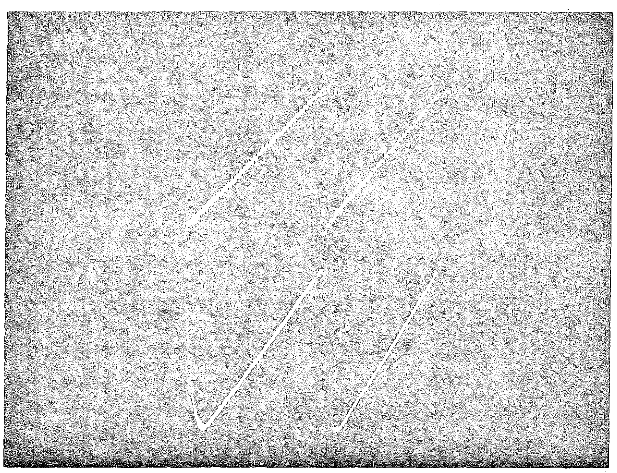
Normal Order			remaining numbers			Scrambled
0	Top is: 0	Y	1	Top is: 1	Z	Y: 0
1	Bottom is: 7		2	Bottom is: 6		7
2	Center: 3		5	Center: 2		3
3	4		6	5		4
4						Z: 1
5						6
6						2
7						5

SINEWAVE		SAWTOOTH		TRIANGLE	
C_1	0.637	C_1	-1.000	C_1	0.500
C_5	-0.264	C_3	-0.500	C_5	-0.250
C_9	-0.053	C_7	-0.250	C_{13}	-0.125
C_{13}	-0.127	C_{15}	-0.125	C_{29}	-0.063
C_{17}	-0.013	C_{31}	-0.063		
C_{21}	0.005	NON-ZERO WALSH COEFFICIENTS IN RANGE 0 - 31			
C_{25}	-0.026				
C_{29}	-0.063				

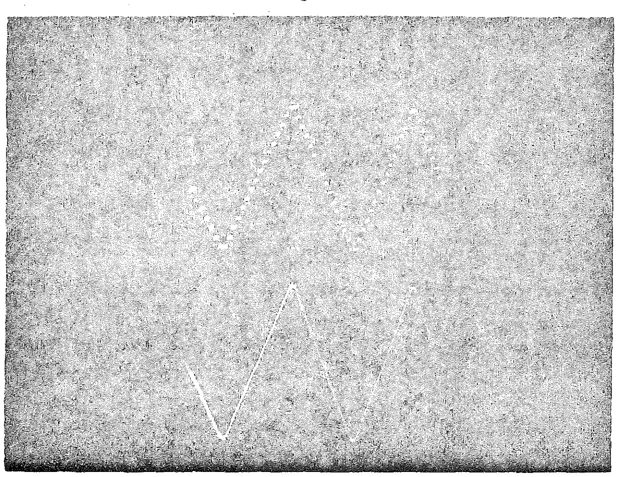
a



b



c



d

Fig. 4**a**. Walsh-Fourier coefficients of common waveforms. **b**. Walsh generated sinewave (top) and with low-pass filtering (bottom). **c**. Walsh generated sawtooth (top) and with low-pass filtering (bottom). **d**. Walsh generated triangle (top) and with low-pass filtering (bottom).

plays an important role. For example, if an upramp reaches its peak at 15 and falls to zero at 16, it is an attack only envelope. However, such a ramp requires the entire available set of 32 Walsh functions and hence, formidable summing problems, and would therefore be inferior to a point-by-point generation method. The ramp peaking at 16 on the other hand requires only nine of the first 32 Walsh functions, but is an attack and sustain envelope. By separating out a block pulse at segment 16, denoted $P(16)$, and using this as a blanking pulse, the attack and sustain envelope may be converted to attack only. Similarly, the easily generated downramp from segment 17 to segment 32 (decay only) can be made sustain and decay by the addition of $P(16)$. $P(16)$ is also needed as the signal-to-end attack, while the block pulse at segment 32, denoted $P(32)$, is the signal-to-end decay and go to a complete rest condition. Furthermore, $P(16)$ represents the entire sustain time, and is useful as a gate for additional effects on the signal during sustain, which otherwise would be a simple periodic waveform unacceptable for long sustain time. While $P(16)$ and $P(32)$ could be obtained as a Walsh-Fourier series, this would require summation of all of the first 32 Walsh functions. Fortunately, they are easily obtained from the Walsh functions using logic gates as indicated in Fig. 7.

OVERALL EXPERIMENTAL SYSTEM

The overall experimental system is shown in Fig. 6.

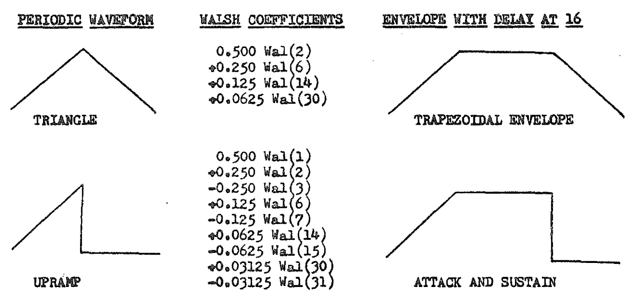


Fig. 5. Generation of envelopes from periodic waveforms by a delay at segment 16.

A voltage-controlled oscillator (VCO) is run four octaves above its normal range. Envelope control signals are obtained in conjunction with a sample-and-hold circuit, which is needed to store the frequency information during decay, i.e., after a key is lifted and the control voltage would normally disappear. Selected Walsh function harmonics are obtained from the Walsh harmonic bank, which is being driven by the VCO, and these are patched into a bank of VCA's. The VCA's are controlled by various envelopes, and the final set of amplitude shaped harmonics is mixed. The output of the mixer is then subjected to low-pass or other desired filtering by means of the VCF which can track the VCO by means of the same control voltage. The system provides time dependent harmonic changes similar to those obtained by VCF's using subtractive synthesis.

THE ENVELOPE CONTROLLER

The envelope controller is shown in Fig. 7. The attack portion of the envelope is triggered by the coinci-

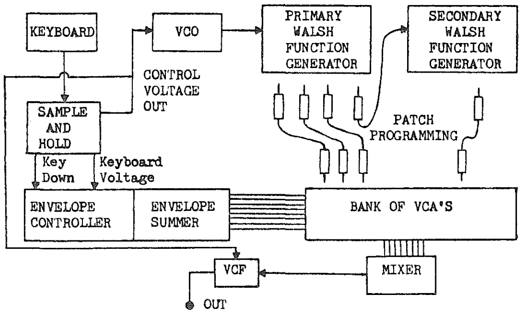


Fig. 6. Overall experimental system.

dence of a key down and a discrete change of control voltage. Thus, an attack envelope is initiated whenever the VCO changes frequency discretely, and hence a new envelope is initiated whenever a different key on the keyboard is depressed regardless of whether or not the first key is lifted completely first. Initiation of attack consists of forcing the generator from the 32 state to the first state. Upon completion of the attack, the generator is stopped by $P(16)$, and remains there until final removal of the key-down signal which forces the generator into state 17, where it is advanced by a separate decay clock to state 32. It is also fairly easy to alter the controller logic so that other modes of envelope timing can be obtained.

THE ENVELOPE SUMMER

The envelope summer is shown in Fig. 8. By adding and subtracting summed coefficients and complete envelopes, seven basic envelopes can be obtained from the envelopes of Fig. 5 plus $P(16)$. The envelopes may be low-passed to remove the sharp steps. Proper use of these envelopes gives some control over both envelope and timbre at the keyboard without touching separate controls. For example, using the trapezoidal envelope, a long attack and sustain can be obtained by always having at least one key pressed down; sharp tapping of the keys gives rapid transition to the decay state for a piano-like decay. By putting one set of Walsh harmonics under attack and sustain envelopes, and a second set under decay envelopes, the same two playing techniques will result in different voices as well as different envelopes.

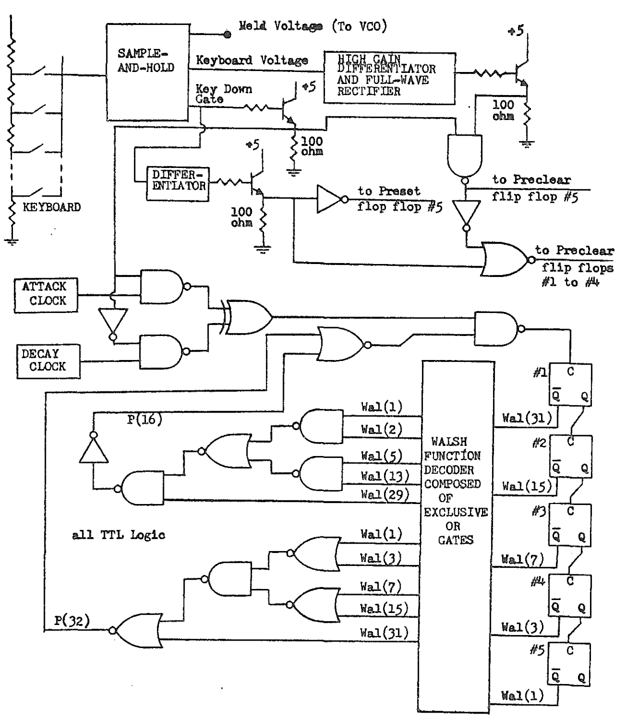


Fig. 7. Envelope controller circuit showing sources of controlling signals and method of obtaining $P(16)$ and $P(32)$ from the Walsh functions. The Walsh functions needed for the envelope summer are removed from the decoder.

NONHARMONIC TONES

The Walsh harmonic bank as discussed above produces only harmonic (in the Fourier sense) overtones, due to the fact that all the Walsh functions of the bank have the same basic interval. The experimental system has available several additional provisions for controlling

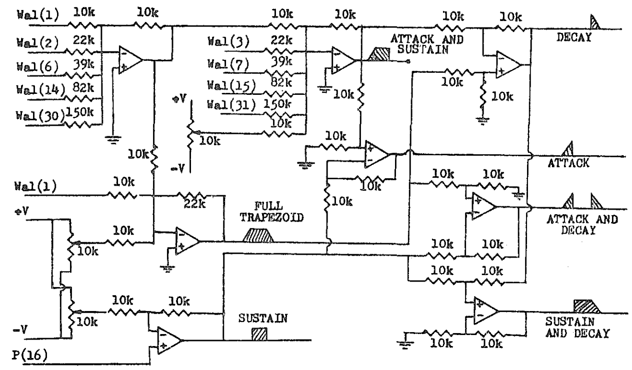


Fig. 8. Practical envelope summer. Seven basic envelopes are available. More complex envelopes require more of the Walsh functions.

timbre by introducing nonharmonic tones. Modulation effects (AM, FM and balanced) can be employed in the conventional manner to produce sidebands, thus altering timbre. Also, a second Walsh harmonic bank can be driven through a symmetric divide-by-n frequency divider (where n is not a power of two), or by a separate but tracking VCO. In the former case, the divide-by-n may be driven by any of the square-wave Walsh functions from the primary bank, and the resulting nonharmonics analyzed in terms of the frequency ratios of the square waves involved. Alternatively, the secondary bank can be driven from any of the more irregular Walsh functions. In this case, the outputs of the secondary bank will not be Walsh functions but may still be useful musically. This latter case is best analyzed in terms of a sequency triggering rate rather than through consideration of frequency ratios.

SEQUENCY TRIGGERING RATE

Suppose for example that $Wal(31)$ is actually a square wave of frequency 1600 Hz (sequency 1600 zps). $Wal(30)$ will then have a sequency of 1500 zps. If these two Walsh functions trigger identical Walsh function generators as indicated in Fig. 9, and these generators trigger on voltage transitions in one direction only, then the average triggering rate is just the sequency. The $m = 5$ Walsh generator involves divide-by-16 circuitry, the divide-by-16 occurring at the $Wal(1)$ output. The sequency of the $Wal(1)$ output of the first generator is therefore 100 zps, while the sequency of the $Wal(1)$

output of the secondary generator is about 93.8 zps. The divide-by-16 has in the mean time reduced the irregularity occuring in the *Wal*(30) waveform to a point where it appears as an error of one part in 16 in every sixteenth half-cycle, which may not be audible [12]. This is nearly a symmetric square wave of frequency 93.8 Hz, so the interval between the *Wal*(1) outputs of the two generators is nearly a just semitone. Analysis of the higher sequency outputs of the secondary generator becomes increasingly difficult as irregularities are not reduced by high integer division. It can be observed that single unit changes in the sequency triggering rate produce large changes in tone color.

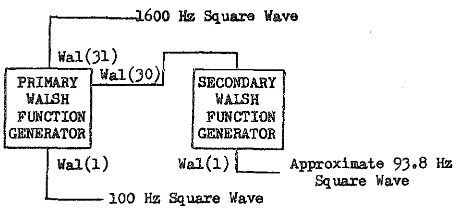


Fig. 9. Sequency triggering rate example showing the approximation of a just semitone.

The sequency drive method can be used, followed by high-integer division circuits to give relatively pure square waves [13] with frequencies proportional to the product of the driving sequency and the output sequency. Various scales and intervals can be investigated by this means. However, the primary interest in the sequency triggering rate is for the nonharmonic effects achieved.

SUMMARY

The use of Walsh functions permits an inexpensive realization of an additive synthesis system through digital-envelope control and digital harmonics. A wide variety of fixed envelopes can be obtained with minimal summation circuitry. Generation of both harmonic and nonharmonic tones using Walsh function generators serves to provide a bank of available waveforms for the additive synthesis process. Careful selection and setting of amplitudes results in sounds with a relatively strong sense of pitch, but the interval of overall periodicity may be greatly extended, and this, along with the time dependent harmonic content of the transients is more demanding on the listener's ear, and hence more demanding of the listener's attention.

ACKNOWLEDGMENT

The author is indebted to Dr. C. Frederick of the Center for Radiophysics and Space Research, Cornell University, for suggesting the use of Walsh functions for music synthesis. The encouragement of Professor W. Ku and his students at the School of Electrical Engineering, Cornell University, is appreciated.

REFERENCES

[1] J. L. Walsh, "A Closed Set of Normal Orthogonal Functions," *American Journal of Mathematics*, vol. 45, p. 5 (1925).

[2] *Applications of Walsh Functions*, Proceedings of 1970 Walsh Function Symposium, AD-707-431, $3.00 from National Technical Information Service, Operations Division, Springfield, Va. 22151.

[3] *Applications of Walsh Functions*, Proceedings of 1971 Walsh Function Symposium, AD-727-000, $3.00 from National Technical Information Service, Operations Division, Springfield, Va. 22151. Also published as a special issue, *IEEE Trans. Electromagn. Compat.*, vol. EMC-13, pp. 75-218 (Aug. 1971).

[4] *Applications of Walsh Functions*, Proceedings of 1972 Walsh Function Symposium, AD-744-650, $6.00 from National Technical Information Service, Operations Division, Springfield, Va. 22151.

[5] H. F. Harmuth, *Transmission of Information by Orthogonal Functions*, 2nd Ed., pp. 90-91 (Springer-Verlag, DDR 1972).

[6] R. Kitai and K. Siemens, "A Hazard-Free Walsh Function Generator," *IEEE Trans. on Instrumentation and Measurement*, vol. IM-21, no. 1, p. 80 (Feb. 1972).

[7] H. F. Harmuth, "A Generalized Concept of Frequency and Some Applications," *IEEE Trans. on Information Theory*, vol. IT-14, no. 3, p. 375 (May 1968).

[8] K. Siemens and R. Kitai, "Digital Walsh-Fourier Analysis of Periodic Waveforms," *IEEE Trans. on Instrumentation and Measurement*, vol. IM-18, no. 4, p. 316 (Dec. 1969).

[9] N. Ahmed and K. R. Rao, "Walsh Functions and Hadamard Transform," Proc. of 1972 Walsh Function Symposium, p. 8.

[10] E. O. Brigham and R. E. Morrow, "The Fast Fourier Transform," *IEEE Spectrum*, vol. 4, pp. 63-70 (Dec. 1967).

[11] V. Lozhkin, "Monaural Phase Effects," *Soviet Physics Acoustics*, vol. 17, no. 1 (July-Sept. 1971). (Review article in English translation).

[12] B. L. Cardozo and R. J. Ritsma, "On the Perception of Imperfect Periodicity," *IEEE Trans. on Audio and Electroacoustics*, vol. AU-16, no. 2, p. 159 (June 1968).

[13] R. B. Cotton, Jr., "Tempered Scale Generation from a Single Frequency Source," *J. Audio Eng. Soc.*, vol. 20, no. 5, p. 377 (June 1972).

THE AUTHOR

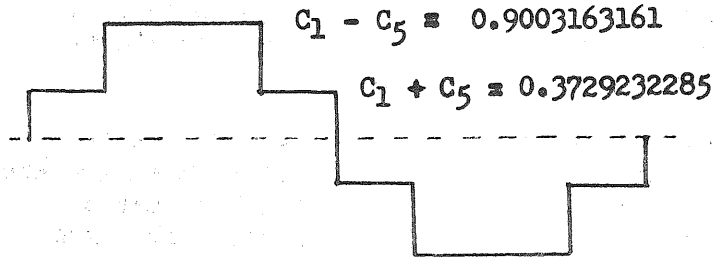
Bernard A. Hutchins, Jr. was born in Rochester, New York in 1945. He received a B.S. degree in Engineering Physics from Cornell University in 1967. After three years of military service, Mr. Hutchins returned to Cornell where he is now pursuing graduate studies in the Department of Applied and Engineering Physics. His main interest in electronic music stems from independent research and activities with *Electronotes Newsletter*. The newsletter is described in "The Forum", *J. Audio Eng. Soc.*, vol. 20, no. 2, p. 129.

SOME NOTES ON THE GENERATION OF SINE WAVES BY WALSH FUNCTIONS

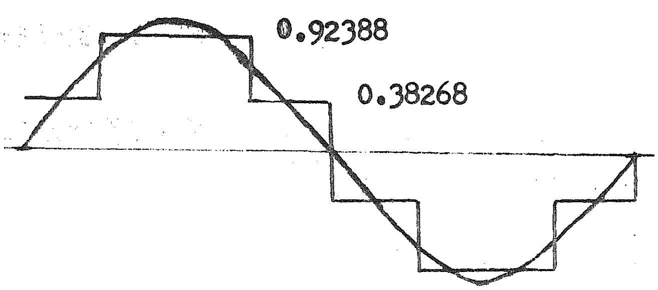
-by B.A. Hutchins June 1973

INTRODUCTION: This paper considers among other things, the degree of approximation obtained by using a truncated Walsh-Fourier series to represent a sine wave, and some of the effects of the initial phase of the sine waves.

1. By evaluating the Walsh-Fourier series by the standard method, keeping only the first two non-zero terms, one gets $C_1 = .63662$ and $C_5 = -.26370$. Plotting up these two terms results in a sine wave approximated by eight segments:



2. As a second method for generating an eight segment sine wave approximation, we can just sample a sine wave at the center of each interval, and use this value. That is, we use the values of the sine function at $22\frac{1}{2}$ degrees and $67\frac{1}{2}$ degrees.



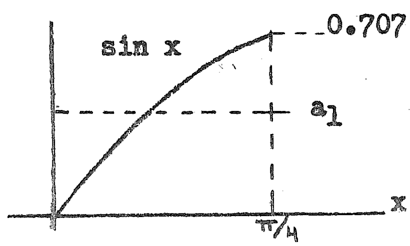
The Walsh-Fourier series of this sampled sine is determined from:

$$C_1' - C_5' = 0.92388$$
$$C_1' + C_5' = 0.38268$$

so: $C_1' = 0.65328$

and: $C_5' = -0.27060$

3. By finding the least-square error in the first segment, it is possible to find the constant level a_1 such that this error is the smallest.



Error = $\sin x - a_1$

Square Error is $(\sin x - a_1)^2$

Representing the total error in the first segment by E:

$$E = \int_0^{\pi/4} (\sin^2 x - 2a_1 \sin x + a_1^2)\, dx$$

giving: $E = 0.143 - 0.586\, a_1 + 0.786\, a_1^2$

taking $dE/da_1 = 0$ gives: $a_1 = 0.3729232285$

Likewise, the level in the second segment for the least-square error is:

$$a_2 = 0.9003163161$$

These are the same as the Walsh-Fourier series values.

4. It is clear that the Walsh-Fourier series and the least-square methods give the same result for the sine wave. It is less clear that the sampling method is also the same. In fact, the levels for the sampled method are proportional to the other levels, but larger by a factor of 1.02617215. Thus, the Walsh series methods and the least square methods will place the levels in the same relative positions as the sampling method. The sine function and the constant levels therefore intersect in the exact middle of the interval. Before considering this further, consider the harmonic content of the eight-segment sine wave approximation.

5. The Fourier harmonic content of Walsh functions can be obtained form the same data used to determine the Walsh-Fourier coefficients of sine wave harmonics. We are interested in Wal(1) and Wal(5). Wal(1) is easy, since it is just a square wave. The Wal(1) waveform therefore contains only odd harmonics, and the intensities go as 1/n, where n is the order of the Fourier harmonic. The series are:

$$\mathrm{Wal}(1) = \frac{2}{\pi}\,\mathrm{Sin}\,x + \frac{2}{3\pi}\,\mathrm{Sin}\,3x + \frac{2}{5\pi}\,\mathrm{Sin}\,5x + \frac{2}{7\pi}\,\mathrm{Sin}\,7x + \ldots\ldots$$

$$\mathrm{Wal}(5) = \frac{-2}{\pi}(\sqrt{2}-1)\,\mathrm{Sin}\,x + \frac{2}{3\pi}(\sqrt{2}+1)\,\mathrm{Sin}\,3x + \frac{2}{5\pi}(\sqrt{2}+1)\,\mathrm{Sin}\,5x$$

$$+\ \frac{-2}{7\pi}(\sqrt{2}-1)\,\mathrm{Sin}\,7x + \frac{-2}{9\pi}(\sqrt{2}-1)\,\mathrm{Sin}\,9x + \frac{2}{11\pi}(\sqrt{2}+1)\,\mathrm{Sin}\,11x\ \ldots\ldots$$

The terms are:

Harmonic	Wal(1)	Wal(5)
1	0.6366197723	-0.2636965438
3	0.2122065907	0.5123120295
5	0.1273239544	0.3073872177
7	0.09094568176	-0.03767093482
9	0.07073553026	-0.02929961597
11	0.05787452476	0.1397214625
13	0.04897075172	0.1182258529
15	0.04244131815	-0.01757976958
17	0.03744822190	-0.01551156140
19	0.03350630380	0.09040800520
21	0.03031522725	0.07318743278
23	0.02767912053	-0.01146506712
25	0.02546479089	-0.01054786175
27	0.02357851008	0.05692355883
29	0.02195240594	0.05299779615
31	0.02053612168	-0.008506340122

6. It is now easy to compute the Fourier harmonic content of the eight segment sine wave. We have simply to take $C_1 \cdot Wal(1) + C_5 \cdot Wal(5)$, where $Wal(1)$ and $Wal(5)$ are represented in terms of their Fourier series. The series is:

$$\frac{8}{\pi^2}(2-\sqrt{2})\,\text{Sin } x + \frac{8}{7\pi^2}(2-\sqrt{2})\,\text{Sin } 7x + \frac{8}{9\pi^2}(2-\sqrt{2})\,\text{Sin } 9x$$

$$+ \frac{8}{15\pi^2}(2-\sqrt{2})\,\text{Sin } 15x + \frac{8}{17\pi^2}(2-\sqrt{2})\,\text{Sin } 17x$$

$$+ \frac{8}{23\pi^2}(2-\sqrt{2})\,\text{Sin } 23x + \frac{8}{25\pi^2}(2-\sqrt{2})\,\text{Sin } 25x + \ldots .$$

The coefficients are:

Harmonic	Fourier Coefficient	% Of First Harmonic
1	0.4748206017	100
7	0.06783151453	14.28571428
9	0.05275784463	11.11111111
15	0.03165470678	6.666666666
17	0.02793062363	5.882352941
23	0.02064437398	4.347826086
25	0.01899282406	4.000000000
31	0.01531679360	3.225806451
33	0.01438850308	3.030303030
39	0.01217488722	2.564102564
41	0.01158099028	2.439024390
47	0.01010256599	2.127659574
49	0 009690216362	2.040816326
55	0.008633101849	1.818181818
57	0.008330185995	1.754385964
63	0.007536834948	1.587301587
65	0.007304932334	1.538461538

7. The harmonic distortion is calculated in terms of the power, hence as the square of the amplitudes above.

$$D_h = \left(1/7^2 + 1/9^2 + 1/15^2 + 1/17^2 + 1/23^2 + 1/25^2 + \ldots \right)^{\frac{1}{2}} / 1^2$$

giving a little less than 23%

8. It can be seen that the three methods in 1, 2, and 3 above give the same result for the best approximation to a sine wave with eight segments. Although not indicated above, the requirement of equal areas above and below the level a_1 give the same result. It is perhaps most surprising that the correct levels fall in the center of the segments, as this is obviously not true for a general waveform.

9. The problem considered above is a special case (eight segments) of the more general problem of the digital generation of sine waves. The output of such digital waveform generators are the same as those that could be obtained with a sample-and-hold and an analog sine wave. Thus, the results of sampling theory apply to the output of the digital generator. The above system would have a sampling equivalent where the sampling frequency is eight times the signal frequency. The spectrum of such a sampled waveform is just the fundamental, the sampling frequence plus or minus the signal frequency, the sampling frequency plus or minus twice the signal frequency, etc. This is consistant with the above results. The overall problem is treated by Y. Neuvo in "The Application of Digital Techniques to a VOR Signal Generator" in <u>IEEE Trans on Aerospace and Electronic Systems</u>, Vol AES-9, No.1, Jan. 1973. His results show that if ω_s is the sampling frequency, and ω_o is the signal frequency, the amplitude of the fundamental is given by:

$$|Y(j\omega_o)| = \frac{\sin \pi\omega_o/\omega_s}{\pi\omega_o/\omega_s}$$

and the amplitude of the higher harmonics are given by:

$$|Y[j(k\,\omega_s \pm \omega_o)]| = |Y(j\omega_o)|\frac{1}{k\,\omega_s/\omega_o \pm 1} \qquad k= 1,2,3,\ \ldots$$

The harmonic distortion can be determined as in 7 above, or by the approximate formula:

$$D_{hd} = \frac{\pi}{\sqrt{3}\,\omega_s/\omega_o} \qquad \text{giving 22.7\% for eight segments}$$

The above figure for distortion applies to infinite precision for the setting of the amplitude levels. When the levels are quantized, additional distortion appears. For N discrete levels, the distortion is:

$$D = \frac{1}{\sqrt{6}\quad N}$$

Additional references are: Ojala et al, "Digital Sinewave Reconstruction for Generating Analog Waveforms" <u>Electronic Engineering</u>, March, 1973, and A. Davies, "Digital Generation of Low Frequency Sinewaves" <u>IEEE Trans. on Instru. and Measur.</u> Vol. IM-18, pp. 97-105, June, 1969.

10. Modulation of the generated sine wave is usually done by an external method. F.M. would be gererated by varying the driving frequency to the Walsh generator. A.M. would best be done on the output waveform from the generator. Voltage-Controlled Oscillators and Voltage-Controlled Amplifiers would handle these functions. A voltage controlled Phase Modualtion system is developed below.

The sine wave was generated from a Walsh-Fourier series starting with Wal(1) and Wal(5), the coefficients being C_1 and C_5. Evidently the series beginning with $-C_1$ and $-C_5$ gives -sin. A corresponding series using Wal(2) and Wal(6) can be used to generate cosines, that is, sine waves out of phase with the first sine wave. In particular, $C_2 = -0.637$ and $C_6 = -0.264$ gives the cosine, and C_2 and C_6 give the -cos. The sinusoidals of arbitrary phase can be generated as a combination of the Wal(1), Wal(5), Wal(9).... series (the odd ordered Sal's) and the Wal(2), Wal(6),Wal(10) .... series (the odd ordered Cal's). This is similar to the process of generating an arbitrary periodic function of arbitrary phase using a combination of sines and cosines. If the delay of a sinusoidal is defined by a phase angle θ, then the coefficients are given in terms of the coefficients of the normal sine,$C_n(0)$, by the following relations:

$C_1(\theta) = C_1(0) \text{ Cos } \theta$ $\qquad$ $C_2(\theta) = C_1(0) \text{ Sin } \theta$

$C_5(\theta) = C_5(0) \text{ Cos } \theta$ $\qquad$ $C_6(\theta) =-C_5(0) \text{ Sin } \theta$

$C_9(\theta) = C_9(0) \text{ Cos } \theta$ $\qquad$ $C_{10}(\theta) = C_9(0) \text{ Sin } \theta$

$C_{13}(\theta) = C_{13}(0) \text{ Cos } \theta$ $\qquad$ $C_{14}(\theta) = -C_{13}(0) \text{ Sin } \theta$

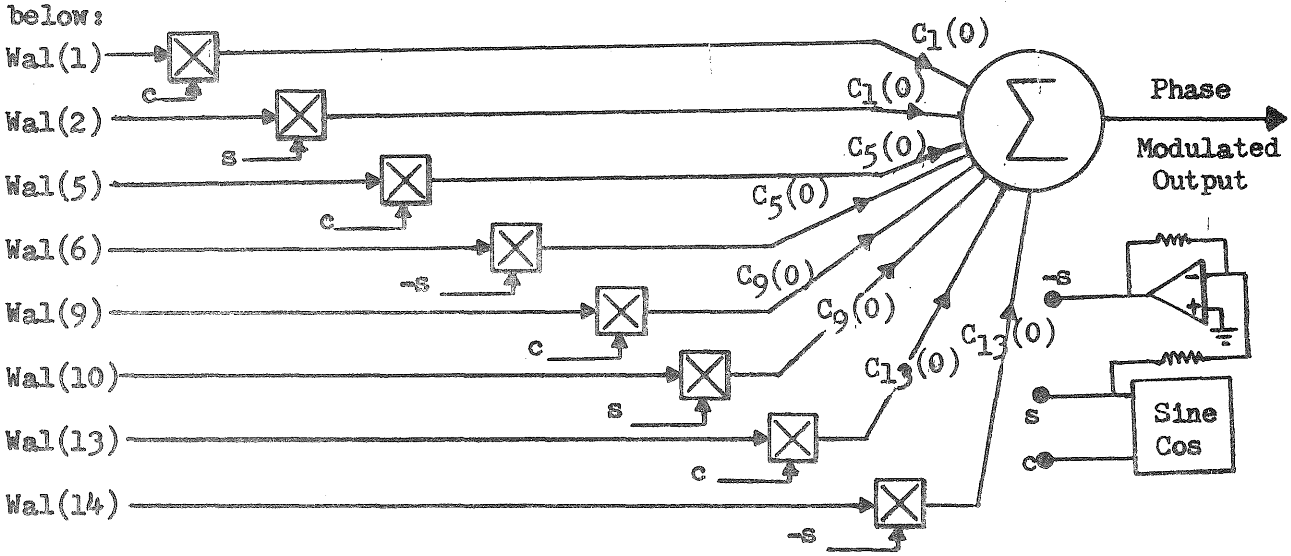
$C_{17}(\theta) = C_{17}(0) \text{ Cos } \theta$ $\qquad$ $C_{18}(\theta) = C_{17}(0) \text{ Sin } \theta$

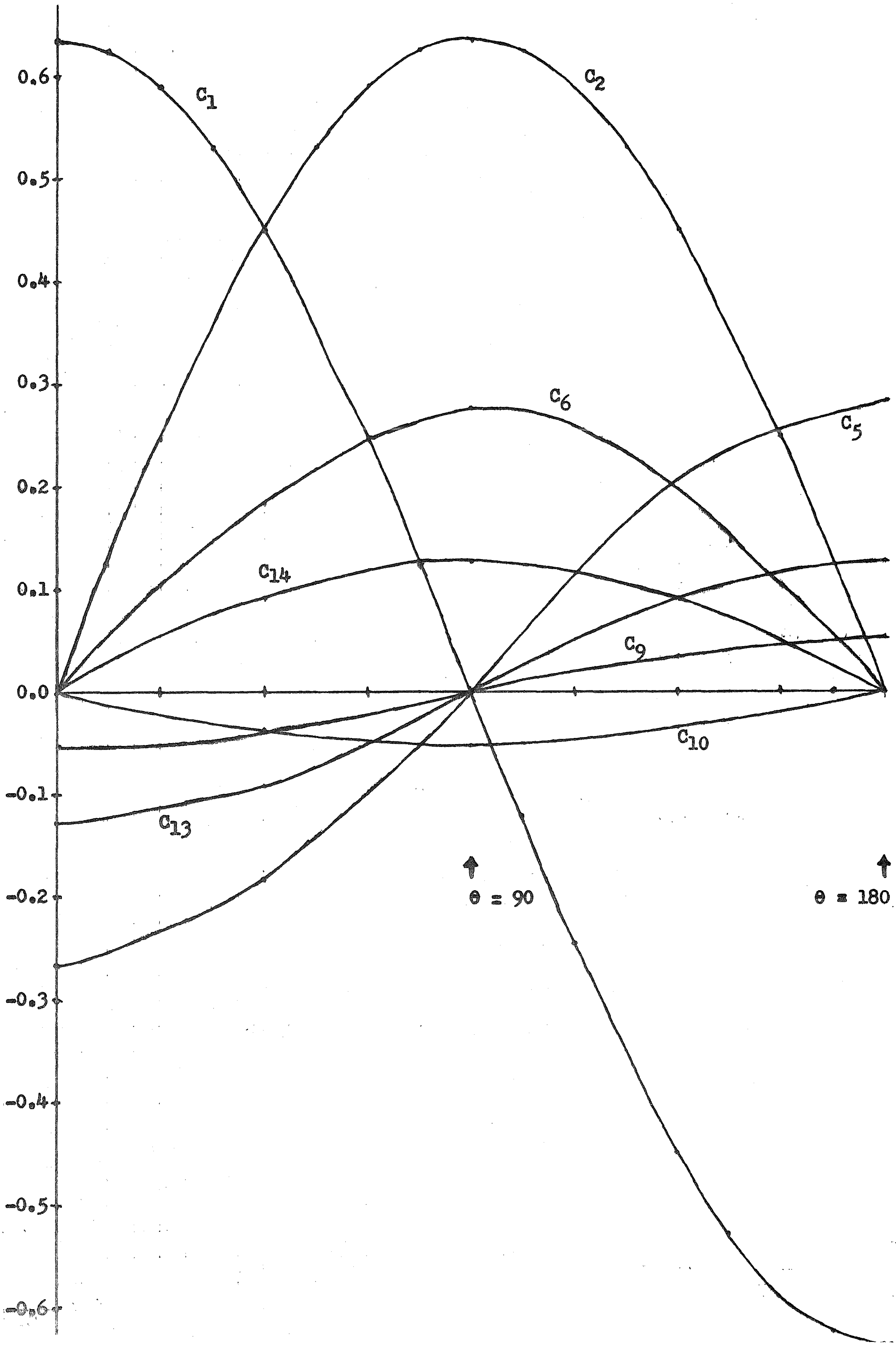
$C_{21}(\theta) = C_{21}(0) \text{ Cos } \theta$ $\qquad$ $C_{22}(\theta) =-C_{21}(0) \text{ Sin } \theta$

$C_{25}(\theta) = C_{25}(0) \text{ Cos } \theta$ $\qquad$ $C_{26}(\theta) = C_{25}(0) \text{ Sin } \theta$

$C_{29}(\theta) = C_{29}(0) \text{ Cos } \theta$ $\qquad$ $C_{30}(\theta) = -C_{29}(0) \text{ Sin } \theta$

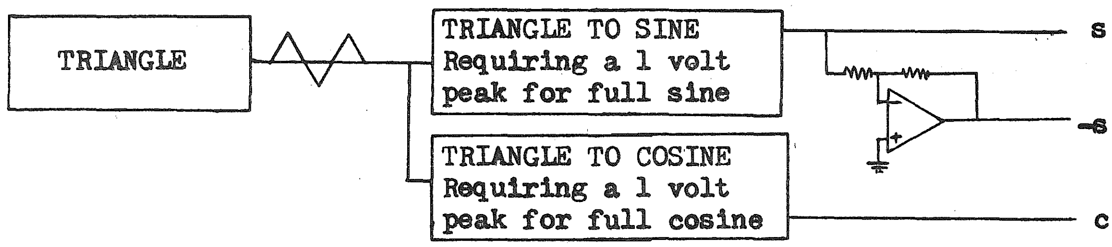
The first eight if these terms are shown plotted on the next page. For the moment, the variation of these coefficients with the Sin of Cos of θ can be regarded as an empirical observation. The application of these results to the design of multiphase oscillators is fairly obvious. Suppose also that a quadrature sinusoidal oscillator is designed by the above or other methods, and this is employed to vary the Walsh coefficients of a second generator. The result would be a form of phase modulation. A possible setup is indicated below:





0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
c_1
c_2
c_6
c_5
c_{14}
c_9
c_{10}
c_{13}
θ = 90
θ = 180

It is interesting to examine the phase modulation system above as it might be driven by the following modulator, composed mainly of triangle to sine converters:

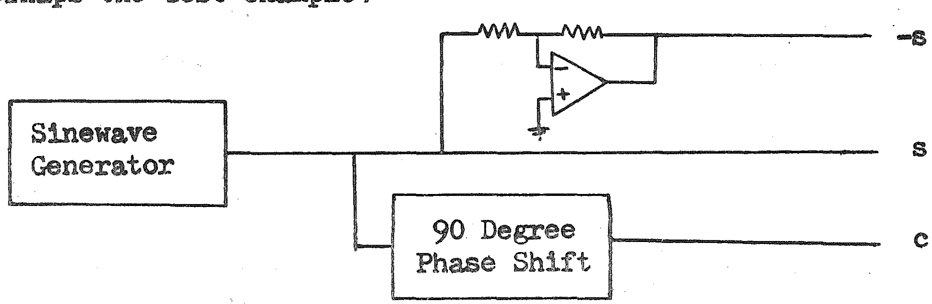


To phase shift the output by an angle θ, it is necessary to provide the functions $\sin\theta$, $-\sin\theta$, and $\cos\theta$. When θ is a linear function of time $\theta = \omega' t$, we need $\sin\omega' t$, $-\sin\omega' t$, and $\cos\omega' t$. The phase angle of the output sinusoidal will therefore be $\phi(t) = \omega' t$.

Suppose first that the triangle generator is providing a signal with a peak less than 1 volt. The sine cinverter will put out a sine from 0 to some max angle θ , where $\theta < 360^\circ$. It will then reverse, go to $\sin(-\theta)$, return to 0, and repeat. The phase of the output sinusoidal will have the same phase θ . In the general case, where the input to the triangle to sine converters may be any function, the output phase will follow the input voltage.

The above system provides a voltage controlled phase (linear with the voltage) and the essential feature is that the trig functions do reverse direction and return. An example of this process is shown in Fig. A, for a signal phase modulated to a maximum of 45° at a frequency $\frac{1}{4}$ the frequency of Wal(1). The output waveform is irregular, and levels are not constant withi the segments since the trig functions are constantly changing. The phase shift is evident however, even in the case where only four Walsh functions have been employed. Driving the triangle to sine converters to 1 volt will result in a 360° phase shift, and larger shifts would be possible if the triangle to cosine converters could keep up. However, while there is only one voltage of the output sine corresponding to a given voltage of the triangle, there are two voltages for the cosine, and this would require additional circuitry, perhaps even to drive the cosine converter beyond 90°.

Continual generating systems for sines and cosines are of course the rule rather than the exception, and a sinewave generator with a 90° phase shift network is perhaps the best example:



Here, the phase shift $\phi(t)$ is continually changing, $\phi(t) = \omega' t$, a linear function of time. The delay in phase is therefore continually changing, and the result is a constant shift to a lower frequency. Such a response is shown in Fig. B. The frequency of Wal(1) is shifted down by the frequency of the modulating signal. The system is perhaps better described as a frequency shifter rather than a phase modulator.

FIGURE "A"

VOLTAGE CONTROLLED PHASE SHIFT
PHASE MODULATION SYSTEM

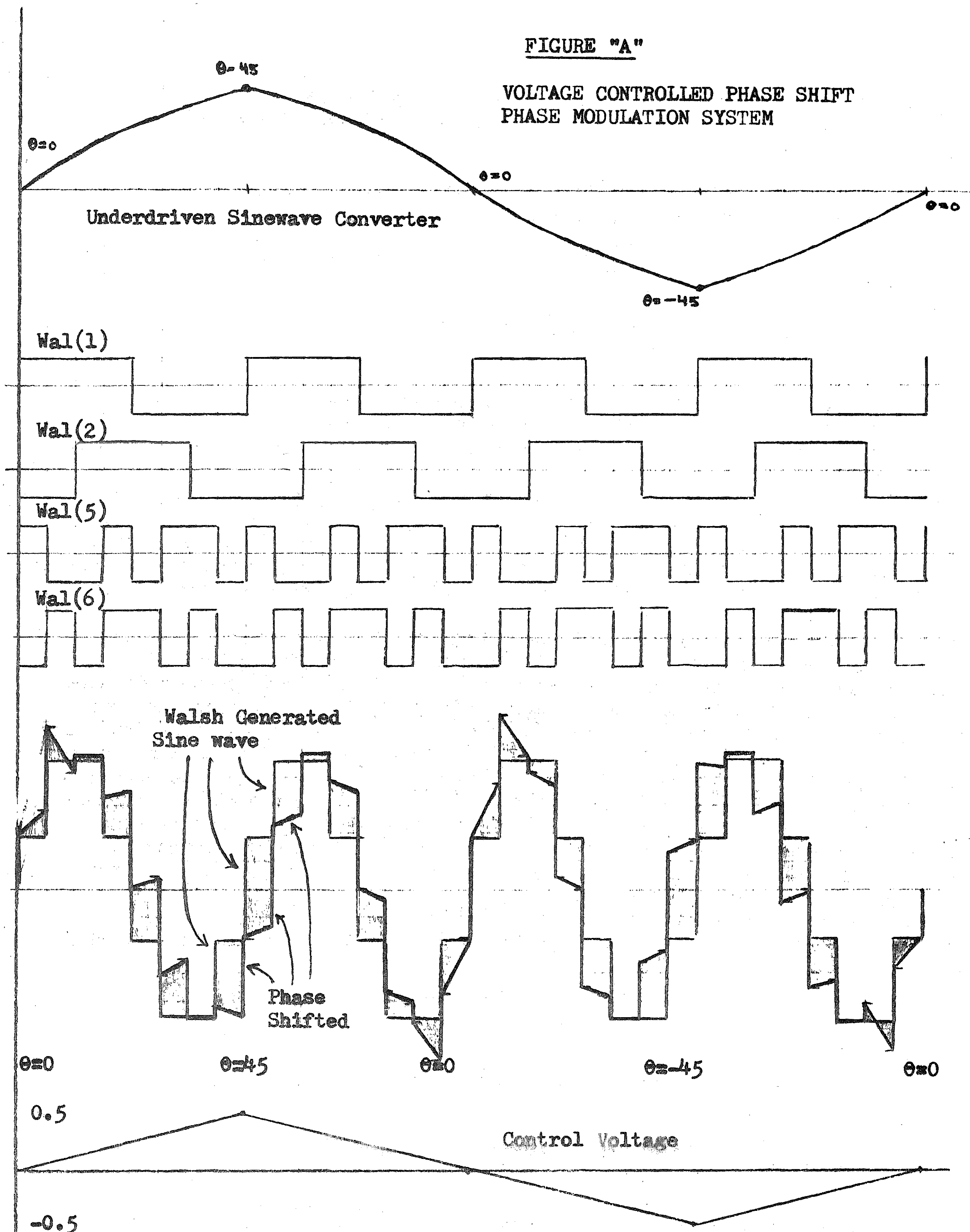
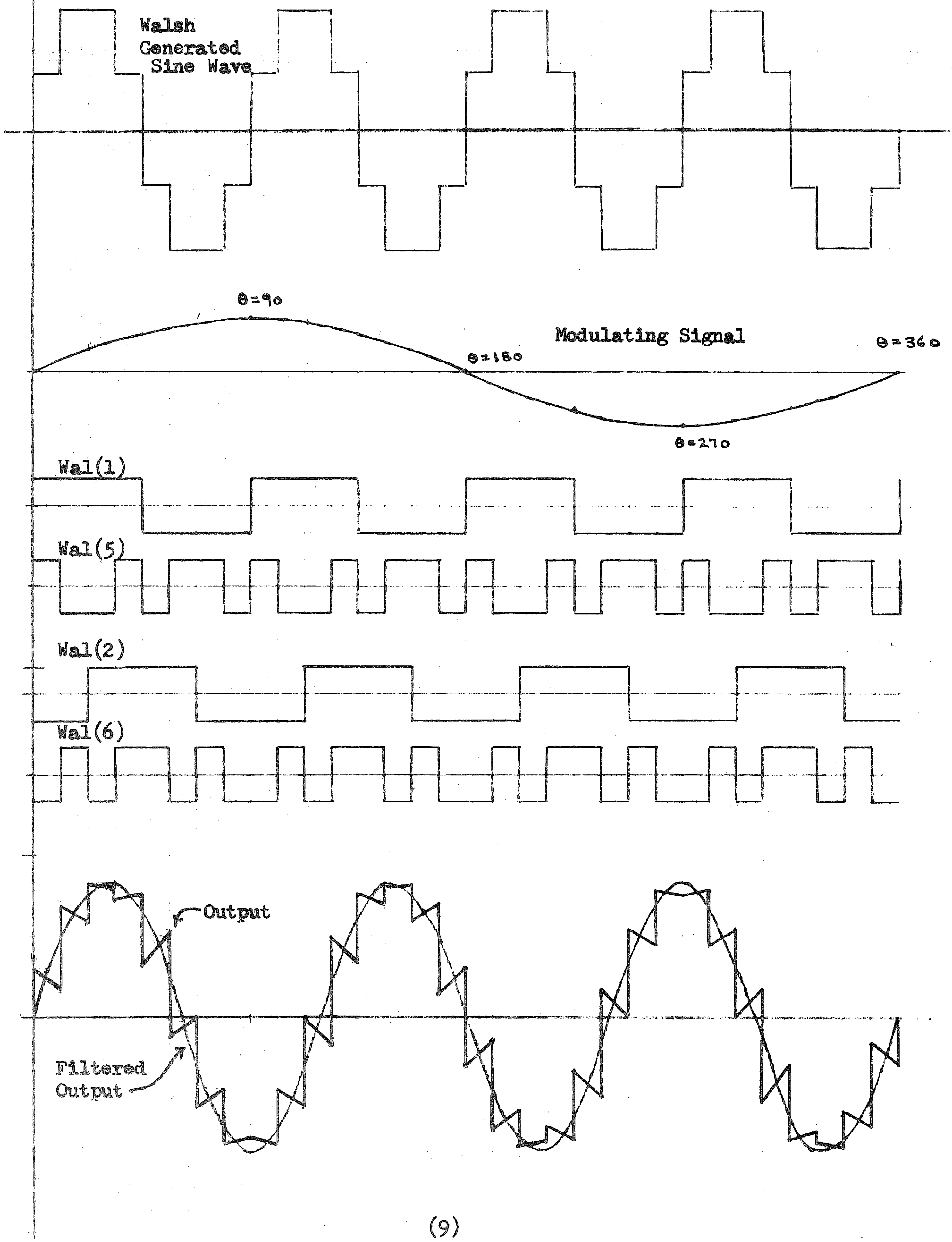


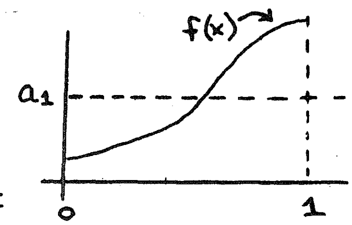
FIGURE "B"
Phase Modulator as a Frequency Shifter



APPENDIX A: The equivalence of the Least Squares, Equal Areas, and Walsh Methods:

LEAST SQUARE ERROR METHOD: The error is $f(x) - a_1$, and the square error is:

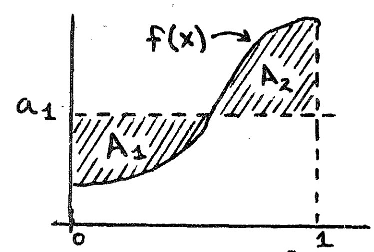
$$E = \int_0^1 (f^2(x) - 2a_1 f(x) + a_1^2)\, dx$$



$$\frac{dE}{da_1} = -2 \int_0^1 f(x)\, dx + 2a_1$$

setting $dE/da_1 = 0$ gives: $$a_1 = \int_0^1 f(x)\, dx$$

EQUAL AREAS ABOVE AND BELOW TO DETERMINE THE LEVEL a_1



$$A_1 = \int_0^{x_1} (a_1 - f(x))\, dx$$

$$A_2 = \int_{x_1}^1 (f(x) - a_1)\, dx$$

setting $A_1 = A_2$ $$\int_0^{x_1} f(x)\, dx + \int_{x_1}^1 f(x)\, dx = a_1 \int_0^{x_1} dx + a_1 \int_{x_1}^1 dx$$

gives: $$a_1 = \int_0^1 f(x)\, dx$$

WALSH FOURIER SERIES METHOD:

Consider the first segment, the level in the first segment is:

$$a_1 = c_0 - c_1 + c_2 - c_3 + \dots\dots\dots c_{2n}$$

Where the c's are the Walsh Fourier coefficients, and the + and - alteration of sign can be seen to result from the sign of the Walsh function in the first segment. An example matrix for n = 3 is indicated below at the right.

$$a_1 = c_0 + c_1 - c_2 \dots$$

$$= \int_0^1 f(x)\, \mathrm{Wal}(0,x)\, dx + \int_0^1 f(x)\, \mathrm{Wal}(1,x)\, dx$$

$$- \int_0^1 f(x)\, \mathrm{Wal}(2,x)\, dx + \dots\dots$$

$$= \int_0^1 f(x)(\mathrm{Wal}(0,x) + \mathrm{Wal}(1,x) - \mathrm{Wal}(2,x) + \mathrm{Wal}(3,x) - \dots\dots)\, dx$$

(from here, use n=3 as an example)

1	1	1	1	1	1	1	1
1	1	1	1	-1	-1	-1	-1
-1	-1	1	1	1	1	-1	-1
1	1	-1	-1	1	1	-1	-1
-1	1	1	-1	-1	1	1	-1
1	-1	-1	1	-1	1	1	-1
-1	1	-1	1	1	-1	1	-1
1	-1	1	-1	1	-1	1	-1

$$a_1 = \int_0^{1/8} f(x)(1 + 1 + 1 + 1 + 1 + 1 + 1 + 1\,)\,dx$$

$$+ \int_{1/8}^{1/4} f(x)(1 + 1 + 1 + 1 - 1 - 1 - 1 - 1)\,dx$$

$$+ \;\ldots\ldots\ldots\ldots \int_{7/8}^{1} f(x)(1 - 1 + 1 - 1 + 1 - 1 + 1 - 1)\,dx$$

The eight ones within each integral are actually the inner products of the first column (a column vector) and the column corresponding to the segment represented by the integral. Except for the first segment, the one we are considering, the pluses and minuses cancel.

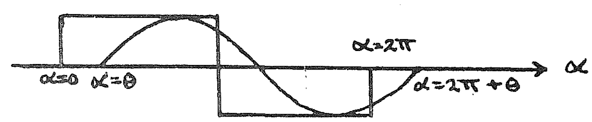
Therefore: $\quad a_1 = 8 \int_0^{1/8} f(x)\,dx$

Except for normalization, this is the same as the results in the first two cases.

APPENDIX B: The Phase Angle Dependence of the Walsh-Fourier Coefficients:

The derivation of the formulas on page 5 is straightforward.

$$C_1(\theta) = \int_\theta^{\theta+2\pi} \sin(\alpha - \theta)\,\mathrm{Wal}(1,\alpha)\,d\alpha$$



$$= \int_\theta^{\theta+2\pi} (\sin\alpha\cos\theta \;-\; \cos\alpha\sin\theta)\,\mathrm{Wal}(1,\alpha)\,d\alpha$$

$$= \cos\theta \int_\theta^{\theta+2\pi} \sin\alpha\,\mathrm{Wal}(1,\alpha)\,d\alpha \;-\; \sin\theta \int_\theta^{\theta+2\pi} \cos\alpha\,\mathrm{Wal}(1,\alpha)\,d\alpha$$

limits on integrals can now be changed to any interval of length 2π since both functions under the integral sign run parallel in their independent variable.

$$= \cos\theta \int_0^{2\pi} \sin\alpha\,\mathrm{Wal}(1,\alpha)\,d\alpha \;-\; \sin\theta \int_0^{2\pi} \cos\alpha\,\mathrm{Wal}(1,\alpha)\,d\alpha$$

The second integral is the product of an odd and an even function and therefore the integral is zero.

$$C_1(\theta) = \cos\theta\; C_1(0)$$

For $C_2(\theta)$, replace $\mathrm{Wal}(1,\alpha)$ above by $\mathrm{Wal}(2,\alpha)$, the first integral then becomes zero, and the integral of $\cos\alpha\,\mathrm{Wal}(2,\alpha)$ is opposite in sign to the integral of $\sin\alpha\,\mathrm{Wal}(1,\alpha)$ so $C_2(\theta) = \sin\theta\; C_1(0)$. Likewise, $C_5 = \cos\theta\; C_5(0)$, and $C_6(\theta) = -\sin\theta\; C_5(0)$, the integrals of $\cos\alpha\,\mathrm{Wal}(6,\alpha)$ being of the same sign as the integral of $\sin\alpha\,\mathrm{Wal}(5,\alpha)$, i.e. "in phase" unlike the case of $C_2(\theta)$. The other coefficients follow in the same way.

CHAPTER 9E

BIBLIOGRAPHY FOR HANDBOOK TOPICS

CONTENTS:

INTRODUCTION

This bibliography is intended to supplement the material and references listed in the text. It lists references from ELECTRONOTES and elsewhere that the reader should be aware of. For the most part, material covered in this book or referenced in the text is not listed again here. For a much larger bibliography, see the "Sourcelist of Electronic Music and Musical Engineering" published in EN#49 - EN#53, which will also soon be available as a reprint.

REFERENCES

Chapter 1a Electronic Music Systems and Their Characterization

T. J. Zingheim, "Introduction to Computer Music Technology," EN#42

M. V. Mathews et al,"Computers and Future Music," Science 183 #4122 Jan 25/1975

Chapter 1b Waveforms, Envelopes, Modules, and Control

B. Hutchins, "Glossary of Electronic Music Terms," EN#47, 48, 50

B. Hutchins, "Starting to Use A Voltage-Controlled EM Synthesizer," S-010

Chapter 1c Basic Mathematics of Musical Engineering

B. Hutchins, "A Little Integral Calculus," EN#16

B. Hutchins, "On Changing Linear Waveforms and Envelopes to Exponentials," EN#26

Chapter 1d Integral Methods of Electrical Engineering

B. Hutchins, --Transform methods-- EN#45, 46

E. Metzger, "Analysis of Musical Sounds by Fourier Transform Methods," JASA 42 #4 (1967) pg. 896

H. P. Hsu, Fourier Analysis, Simon and Schuster Tech. Outlines (1970)

Chapter 1e Topics from Musical Acoustics

B. Hutchins, --various topics-- EN#8, 9, 19

H. Chamberlin, "The Role of Phase in Waveshape and Timbre," EN#25

S. von Hoerner, "Universal Music?" EN#35
J. Backus, The Acoustical Foundations of Music, W. W. Norton (1969)
H. F. Olson, Music, Physics, and Engineering, Dover (1967)
J. G. Roederer, Introduction to the Physics and Psychophysics of Music, Springer-Verlag (1973)

Chapter 1f Basic Principles of Musical Engineering
Various Authors --on philosophy of design and electronic music--EN#6, 9, 36, 40-42
B. Hutchins, "Teaching Electronics for Electronic Music," EN#19
B. Hutchins, "Composer Survey," EN#38
B. Hutchins, "An Interview with Robert Moog," EN#45, 47, 50

Chapter 2a Subtractive Synthesis
B. Hutchins,"One Note in the Life of a Synthesizer," EN#23

Chapter 2b Additive Synthesis
B. Hutchins et al, "Digital Harmonics," EN#12, 13
J. Ball, "The Digital Divider in Music Synthesis," EN#34
B. Hutchins, et al,--Applications of Walsh Functions-- EN#16,S-008, EN#17, 19, 20,23, 24
H. Chamberlin, "Fourier Series Waveform Generator," EN#39, (and AES preprint #1023)

Chapter 2c Generalized Modulations
B. Hutchins, "Sampling a Sawtooth," EN#31
B. Hutchins, A. Gnazzo, --location modulation-- EN#33, 46
B. Hutchins, --formant modulation-- EN#40
H. S. Black, Modulation Theory D. Van Nostrand (1953)

Chapter 2d Control of Musical Structure
B. Hutchins, D. Rossum --patching systems-- EN#24, 52

Chapter 2e Miscellaneous Electronic Music Techniques
C. Anderton, "Classical Phasing," EN#13
B. Hutchins, --techniques with a frequency shifter-- EN#44
M. Goldstein, --video synthesis-- EN#45
W. M. Hartman "Distorters," EN#51
T. Wells & E. Vogel, The Technique of Electronic Music, Univ. Stores Inc., PO Box 7756, Austin, TX 78712

Chapter 3a Basic Applications of: Operational Amplifiers
B. Hutchins, --on substitutions-- EN#35
W. Jung, The IC Op-Amp Cookbook, Sams (1974)

Chapter 3b Basic Applications of: Operational Transconductance Amplifiers
H. A. Witlinger, --RCA 3080 applications notes-- ICAN-6668

Chapter 3c Basic Applications of: Current Differencing Amplifiers
T. Fredriksen et al,--National applications notes on LM3800-- AN-72

Chapter 3d Basic Applications of: Analog Multipliers
--Motorola applications notes on the MC1595--

Chapter 3e Basic Applications of: IC Timers
--Signetics applications of SE/NE555--

Chapter 3f Basic Applications of: Digital Integrated Circuits
D. Lancaster, The TTL Cookbook, Sams, 1974
H. Chamberlin, "Digital Point-by-Point Waveform Generator," EN#34
D. Rossum --using TTL-- EN#35
R. Burhans, "More Applications of Rate Multipliers," EN#51

Chapter 3g Basic Applications of: CMOS Integrated Circuits
--National applications notes: AN-77, AN-88, AN-90, AN-118, MB-18--

Chapter 3h Miscellaneous Musical IC's
--Intersil applications notes on the 8030--

C. Anderton,--use of the 566-- EN#8
B. Hutchins, --phase locked loops-- EN#14

Chapter 4a Circuits Using IC Amplifiers
B. Hutchins, --frequency multipliers-- EN#6, 25
R. Burhans, --level indicator-- EN#44
B. Hutchins, --low freq. oscillator-- EN#48

Chapter 4b Analysis and Design of Active Filters
B. Hutchins, --filters-- EN#18
D. Lancaster, The Active Filter Cookbook, Sams (1975)

Chapter 4c Circuits Using Discrete Semiconductors
B. Hutchins --transistor substitutions--.
A. Veronis, Transistor Thoery for Technicians and Engineers, Tab Book #717 (1975)
--National Semiconductor applications AN-32 (FET applications)--

Chapter 5a Modular Design, General Considerations, Inputs and Outputs
ENS-73 System EN#28, 29, 30, 31
ENS-74 System EN#45, 46
P. Titchener, T. Mikulic --slew limiters-- EN#42

Chapter 5b Voltage-Controlled Oscillator Design
D. Rossum --two circuits-- EN#13, 16
B. Hutchins, T. Mikulic, C. Rogers, M. Suchoff --various comments on exponential converters-- EN#33, 37, 45, 49
D. Rossum --VCO syncing-- EN#51

Chapter 5c Voltage-Controlled Amplifier Design
B. Hutchins, R. Burhans, --simple designs-- EN#41, 44

Chapter 5d Voltage-Controlled Filter Design
D. Rossum, --general comments-- EN#36

Chapter 5e Envelope Generator Design
R. Burhans --4 Bit Envelope Generator-- EN#22
D. Rossum, "On Transient Generators" EN#22
R. Burhans, R. Heslip --simple designs-- EN#44, 45
B. Hutchins, --ADSR and delay unit-- EN#50, 51

Chapter 5f Balanced Modulator Design
T. Oberheim, "A Ring Modulator Device for the Performing Musician," AES Preprint #708 (G=4) (1970)

Chapter 5g Sample-and-Hold Design
D. Rossum, B. Hutchins --designs-- EN#14, 38

Chapter 5h Noise and Random Source Design
D. Rossum, --noise source for ENS-73-- EN#30

Chapter 5i Mixer and Multiple Design
T. Mikulic, --design-- EN#34

Chapter 5j Power Supply Design
D. Rossum --IC regulators-- EN#19
C. Anderton --use of the LM309-- EN#40

Chapter 6a Frequency Shifter Design

Chapter 6b Pitch and Envelope Followers
B. Hutchins, --problems and methods-- EN#43, 53, 54, 55

Chapter 6c Modules Employing Analog Delay Lines
J. Hall, --circuit using TCA350-- EN#48

Chapter 6d Transform Devices

Chapter 7a Design of Controllers

D. Rossum, --keyboards-- EN#21
R. Rae, "The Ideal Keyboard" EN#45, 52
B. Hutchins, D. Rossum, G. Leslie --experimental ribbon controllers--EN#31, 51
B. Hutchins, --digital keyboard ideas-- EN#52
B. Burhans, --digital keyboard-- EN#54
B. Mustacich, --guitar controller-- EN#52

Chapter 7b Design of Controller Interfaces

J. Marshall --TTL Digisyntone-- EN#13
B. Hutchins --on two note keyboards-- EN#31
D. Rossum, --polyphonic keyboards-- EN#32

Chapter 7c Design of Sequencers

B. Hutchins --sequencer using RTL-- EN#1 - EN#3
S. Learner --design-- EN#17 follow up by A. Watson EN#51
B. Hutchins,"Notes on Sequencers" EN#31, 32
B. Hutchins, --design ideas-- EN#35
T. Mikulic, --design-- EN#51

Chapter 8a Overall Design Considerations

B. Hutchins --mechanical standards-- EN#7
G. Sogge --systems theory-- EN#12
D. Rossum --design considerations-- EN#13, 24
H. Chamberlin "Comparison of Methods for Producing Arbitrary Waveshapes and Envelopes EN#23

Chapter 8b Construction

B. Hutchins --substitution of resistors and capacitors-- EN#37
D. Mitchell --making circuit boards-- EN#34

Chapter 8c Troubleshooting

B. Hutchins, --notes on troubleshooting the ENS-73-- S-013
B. Hutchins, --frequency counter-- EN#40

INDEX

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

V

W