A. R. Nott

COLLINS

training manual

Product Support Division | Collins Radio Company, Dallas, Texas

Fundamentals of Single Sideband

Copy no. Extracted from 597 0332 00

1 February 1959

training manual

Fundamentals of Single Sideband

Product Support Division | Collins Radio Company, Dallas, Texas

TABLE OF CONTENTS

TABLE OF CONTENTS (CONT)

TABLE OF CONTENTS (CONT)

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS (CONT)

LIST OF ILLUSTRATIONS (CONT)

LIST OF TABLES

CHAPTER 1

INTRODUCTION TO SINGLE SIDEBAND

1. NEED FOR SINGLE SIDEBAND

The need for single-sideband communication systems has arisen because present day radio communications require faster, more reliable, spectrum conservative systems. The quantity of commercial and military traffic is presently so great in the high-frequency (2 to 30 mc) spectrum that it has become necessary to restrict the use of this spectrum to those services which cannot be accommodated by other means. Land-lines, microwave links, and uhf scatter propagation are employed to relieve the load from the high-frequency spectrum. In many instances, these provide a better and more reliable service.

There are, however, many communication services which need the propagation characteristics obtainable only in the high-frequency range. Among these are ship-to-shore communications, air-to-ground communications, and the many military and naval systems which require independence, mobility, and flexibility. Since high-frequency spectrum space is limited, it is essential that the best possible use be made of the space available. This means that communication systems must use a minimum bandwidth, that the guard bands between channels to allow for frequency drift and poor selectivity be minimized, and that spurious radiation be kept to a very low value to avoid interference between services. In addition to this, a more reliable signal is desirable if not essential. Single-sideband communication systems in their present state of development provide these assets.

2. WHAT SINGLE SIDEBAND MEANS

A single-sideband (SSB) signal is an audio signal converted to a radio frequency, with or without inversion. For instance, an intelligible voice signal contains audio frequencies over the range of 300 to 3000 cycles per second (cps). If this audio signal is converted to a radio frequency by mixing it with a 15 mc r-f frequency, the resultant sum frequencies cover the range of 15,000,300 to 15,003,000 cps. Such a signal is an SSB signal without inversion and is referred to as an upper sideband, because it occupies the spectrum space above the r-f conversion frequency. Note that the 15 mc carrier is not included in the range of the SSB signal. The above example does not indicate the presence of a difference frequency. However, when the voice signal is mixed with the r-f frequency, a difference frequency does develop which covers the range from 14,999,700 to 14,997,000 cps. This signal is also an SSB signal but is an SSB signal with inversion . This SSB signal is referred to as a lower sideband signal because it occupies the spectrum space below the r-f conversion frequency. Figure 1-1 illustrates the position of the SSB signal in the r-f spectrum

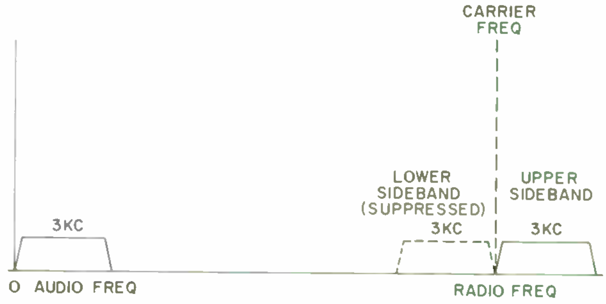


Figure 1-1. Location of SSB Signal in R-F Spectrum

From the above description of the SSB signal, it is apparent that only one sideband signal need be transmitted to convey the intelligence. Since two sideband signals are obtained from the mixing process, it is also necessary to remove one sideband before transmission. To receive the SSB signal, it is necessary to convert the SSB signal back to the original audio signal. This requires identical transmitter and receiver conversion frequencies. In the past, a low-power, pilot carrier was transmitted for automatic frequency control (afc) purposes to provide this end. However, with present day frequency stabilities (1 cps at 10 mc in ground and 10 cps at 10 mc in mobile equipment) the need for afc and pilot carriers is eliminated.

Several methods of sideband communication are in use or under development. The "single-sideband" method as the term is used throughout this book refers to the method which is, perhaps, more accurately termed "single-sideband, suppressed carrier." In this method, only one sideband is transmitted and the carrier is suppressed to the point of non-existence. To demodulate the single-sideband signal requires conversion of the signal with a locally-generated signal close to the proper frequency but with no phase relationship required. In the "single-sideband, pilot carrier" system only one sideband is transmitted, but a low-level carrier of sufficient amplitude for reception is also transmitted. To demodulate this signal, the pilot carrier is separated from the sideband in the receiver, then amplified and used as the conversion frequency to demodulate the sideband signal. In another method, the pilot carrier is used for automatic frequency control of the receiver. In the "double-sideband" (DSB) system, both the upper and lower sidebands of the signal are transmitted with the carrier suppressed to the point of nonexistence. To demodulate the double sideband requires insertion of a locally-generated carrier of both the proper frequency and the proper phase. This system depends upon an automatic frequency and phase control, derived from the double-sideband signal, for control of the locally-generated carrier. In the "single-sideband, controlled carrier" system only one sideband is transmitted, but a carrier which varies inversely with the signal level is also transmitted. This allows an appreciable average carrier level for automatic-frequency-control without reducing the sideband power below the full transmitter rating.

3. HISTORICAL DEVELOPMENT OF SINGLE-SIDEBAND COMMUNICATION SYSTEMS

Although SSB transmission has only received publicity in the last few years, the knowledge of the sideband and the development and use of SSB techniques have progressed over the last 40 years. The acoustical phenomenon of combining two waves to produce sum and difference waves carried over into electric-wave modulation. The presence of the upper and lower sidebands in addition to the carrier frequency were tacitly assumed to exist but were not concretely visualized in the earliest modulated transmissions. Recognition that one sideband contained all the signal elements necessary to reproduce the original signal came in 1915. It was then, that at the Navy Radio Station at Arlington, Va., that an antenna was tuned to pass one sideband well, even though the other was attenuated.

From 1915 until 1923, the physical reality of sidebands was vigorously argued with the opponents contending that sidebands were mathematical fiction. However, the first trans-Atlantic radiotelephone demonstration in 1923 provided a concrete answer. This system employed an SSB signal with a pilot carrier. Single sideband was used in this system because of the limited power capacity of the equipment and the narrow resonance bands of efficient antennas at the low frequency (57 kc) used. By 1927 trans-Atlantic SSB radiotelephony was open for public service.

The first overseas system was followed by short-wave systems, 3 to 30 mc, which transmitted double sideband and carrier because SSB development did not permit practical SSB transmission in this frequency range. However, SSB techniques were employed in various telephony applications and in various multiplexing systems. It has not been until recently that equipment developments have permitted the advantages of SSB communication to be fully exploited. These developments have been in the fields of frequency stability, filter selectivity, and low-distortion linear power amplifiers. These developments have led to military and commercial acceptance of SSB communication systems. There are presently available several radio amateur and commercial SSB radio sets, fixed-station SSB exciters up to 45 kw linear power amplifiers, and airborne transceivers capable of reliable communications with unlimited range. Some of these equipments, especially the military equipments, are provided with automatic frequency selection and automatic tuning to further enhance their value as reliable, easily operated systems.

4. BASIC FUNCTIONAL UNITS OF A SINGLE-SIDEBAND TRANSMITTING SYSTEM

Some of the basic functional units of an SSB system have been previously mentioned. Figure 1-2 shows these units in their functional relationship for an SSB transmitter.

The audio amplifier is of conventional design. Audio filtering is not required because the highly selective filtering which takes place in the SSB generator attenuates the unnecessary frequencies below 300 cps and above 3000 cps. It should be noted that a voice signal is used only as a convenience for explanation. The input signal may be any desired intelligence signal and may cover all or any part of the frequency range between 100 and 6000 cps. The upper limit of the input audio signal is determined by the channel bandwidth and the upper cutoff frequency of the filter in the SSB generator. The lower limit of the input audio signal is determined by the lower cutoff frequency of the filter in the SSB generator.

The SSB generator produces the SSB signal at an i-f frequency. The most familiar way to produce the SSB signal is to generate a double-sideband (DSB) signal and then pass this signal through a highly selective filter to reject one of the sidebands. The SSB signal is generated at a fixed i-f frequency because highly selective circuits are required. The highly selective filter requirements for the filter method of SSB generation are met by either crystal or mechanical filters. Both of these filters have been improved in performance and reduced in size and cost to make their application practical.

The generated SSB signal at a fixed i-f frequency then goes through mixers and amplifiers where it is

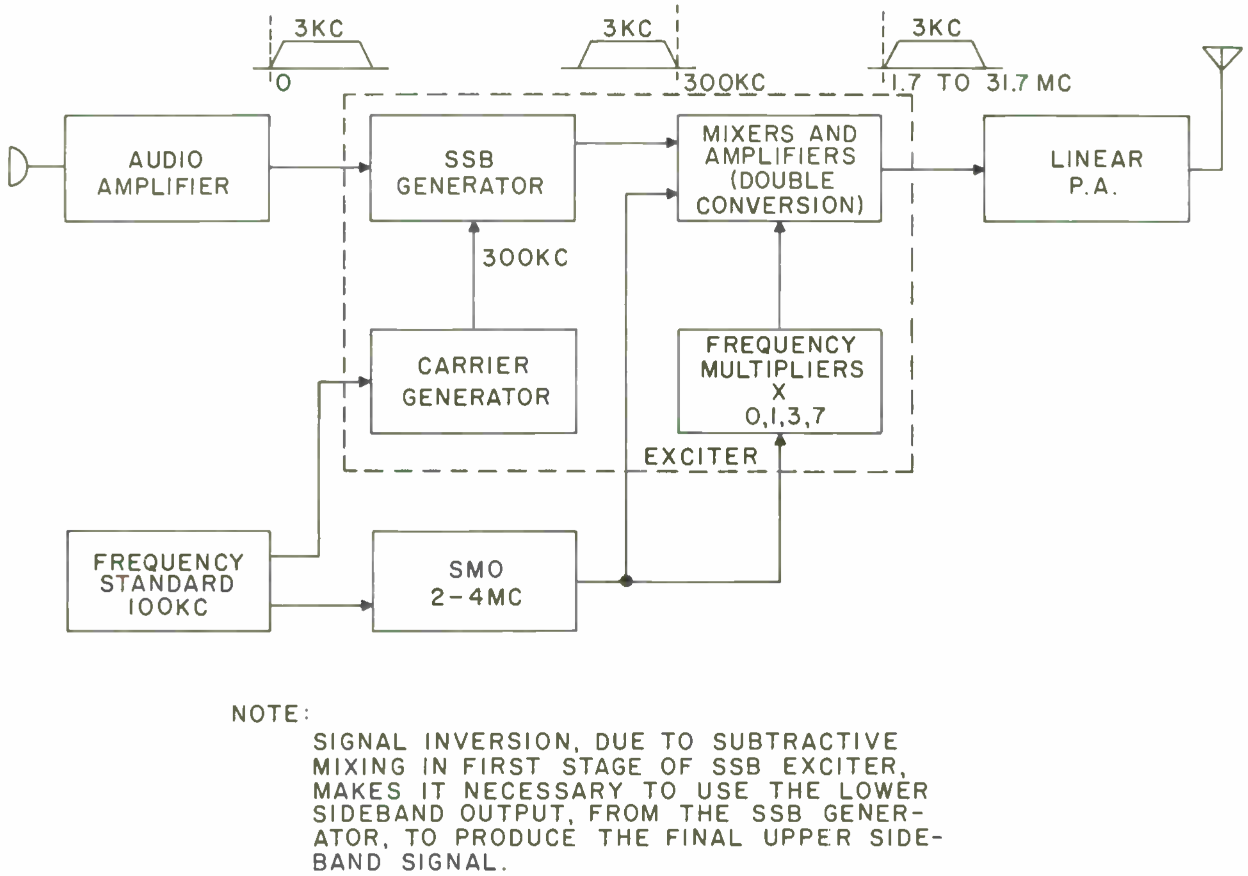


Figure 1-2. Functional Units of an SSB Transmitting System

converted up in frequency to the transmitted r-f frequency. Two stage conversion is shown with the second conversion frequency being a multiple of the first conversion frequency. The frequency conversions required to produce the r-f frequency produce sum and difference frequencies as well as higher order mixing products inherent in mixing circuits. However, the undesired difference frequency or the undesired sum frequency, along with the higher order mixing products, is attenuated by interstage tuned circuits.

The SSB exciter drives a linear power amplifier to produce the high power r-f signal. A linear power amplifier is required for SSB transmission, because it is essential that the plate output r-f signal be a replica of the grid input signal. Any nonlinear operation of the power amplifier will result in intermodulation (mixing) between the frequencies of the input signal. This will produce not only undesirable distortion within the desired channel but will also produce intermodulation outputs in adjacent channels. Distortion in the linear power amplifier is kept low by the design choice of power amplifier tubes, their operating conditions, and use of r-f feedback circuits. The low distortion obtainable in modern linear power amplifiers is not essential to the SSB system nor is it essential for good voice transmission, but it is essential to minimize the guard band between channels and thereby permit full utilization of the spectrum space.

Because an SSB system without a pilot carrier demands an extremely stable frequency system, the frequency standard and stabilized master oscillator (smo) are extremely important. The standard frequency is obtained from a crystal oscillator with the crystal housed in an oven. Since the stability of the crystal frequency depends directly upon the stability of the oven temperature, stable thermal control of the oven is necessary. This thermal control of the oven is obtained by using heat-sensitive semiconductors in a bridge network. Any variation in the oven temperature, then, is indicated and corrected by an unbalance in the control bridge. This system will limit changes in oven temperature to 0.001°C. Such oven stability will provide a standard frequency which will vary no more than 1 cps in 10 mc per day when used in fixed-station equipment and no more than 10 cps in 10 mc per day when used in mobile station equipment.

The carrier generator provides the i-f carrier used to produce the fixed i-f SSB signal, and the smo provides the necessary conversion frequencies to produce the r-f SSB signal. The frequencies developed in these units are derived from or phase locked to the single standard frequency so that the stability of the standard frequency prevails throughout the SSB system. Choice of the fixed i-f frequency and the conversion frequencies to obtain the r-f frequency is an extremely important design consideration. Optimum operating frequencies of the various circuits must be considered as well as the control of undesirable mixing products. The frequency scheme shown is the result of extensive study and experimental verification. It produces minimum spurious output in the high-frequency range (2 to 32 mc). The use of harmonically related conversion frequencies in the mixer permits the frequency range to be covered with a single 2 to 4 mc oscillator, a very practical range for obtaining high oscillator stability. Use of the 300 kc fixed i-f frequency is the optimum operating frequency for the mechanical filter required in the SSB generator.

The foregoing discussion may give the erroneous impression that only single channel communication is possible with an SSB system. Quite the opposite is true. To add additional channels to the SSB system requires only additional circuits in the SSB generator. One method is to use the upper sideband of one signal and the lower sideband of the other signal. Figure 1-3 shows the circuit for producing these two channels and the location of each channel with respect to the carrier frequency. It should be noted that with this method a twin sideband is transmitted, and that the signal in the lower sideband is inverted. Another method of adding channels is shown in figure 1-4. Different fixed i-f frequencies, one raised 4 kc from the original, are injected into separate SSB generators, and the upper sideband is filtered from each output. This produces two channels both using the upper sideband. It should be realized that as additional channels are added to the system, less transmitter power output is available for each channel.

The SSB transmitter is designed for linear operation from the audio input amplifier through the output power amplifier. That is, the transmitter faithfully transmits the original input intelligence with negligible distortion. This distortion-free system is ideally suited for the transmission of multiplex and Kineplex signals, because the original pulses are transmitted without distortion of their wave shape.

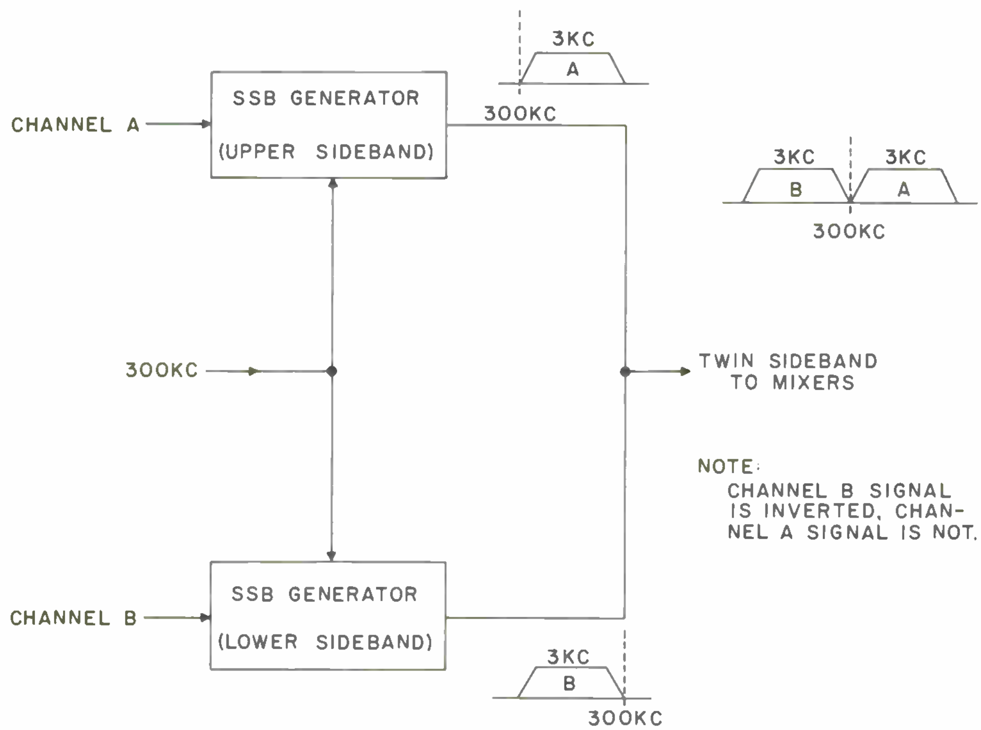


Figure 1-3. Generation of the Twin-Channel Sideband Signal

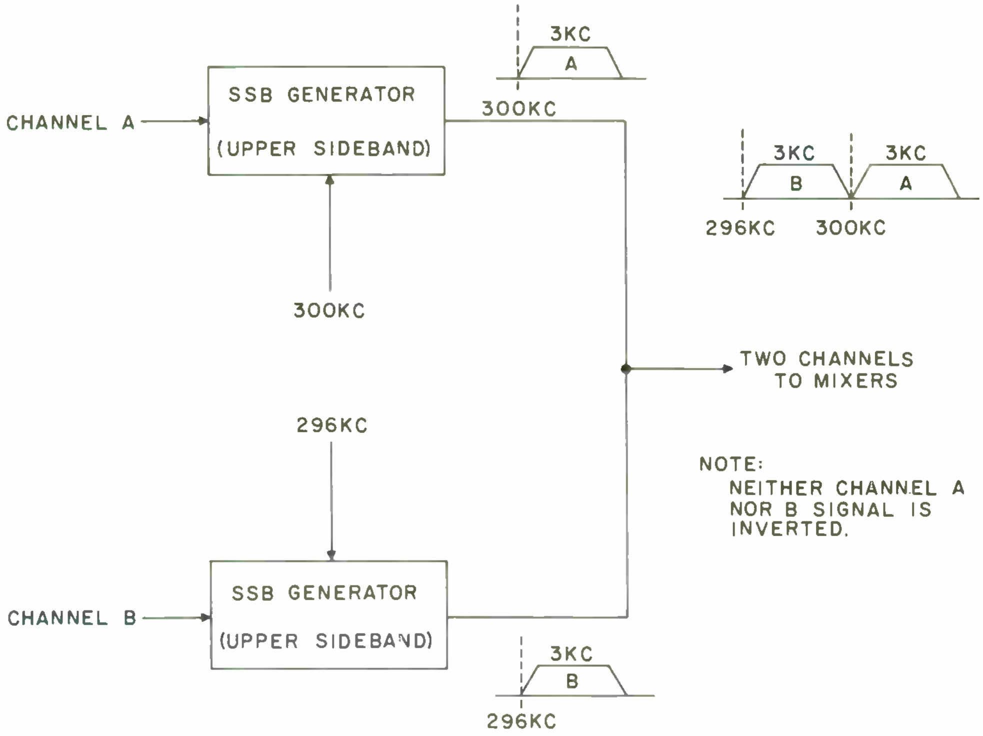


Figure 1-4. Generation of Two Channel SSB Signal

5. BASIC FUNCTIONAL UNITS OF A SINGLE-SIDEBAND RECEIVING SYSTEM

To receive the SSB signal requires a heterodyning system which will convert the radio-frequency signal back down to its original position in the audio spectrum. The basic functional units of such a receiver are shown in figure 1-5. It can be seen that the SSB receiver is almost identical to a conventional heterodyne receiver except for the detection circuit. The r-f signal is amplified and converted down in frequency to a fixed i-f frequency. Then a final fixed i-f injection frequency is required to bring the signal down to its original position in the audio spectrum.

Many of the units of an SSB receiver are identical with units of the SSB transmitter as can be seen by comparing figure 1-2 with figure 1-5. The frequency standard, carrier generator, and smo are identical. The double conversion mixer and amplifier unit of the receiver can be made identical to the double conversion mixer and amplifier unit of the transmitter. This similarity of functions permits the construction of transceivers with much of the circuitry used for both receiving and transmitting by merely adding switching to reverse the direction of signal flow. By using dual purpose units and adding switching to reverse the direction of the signal, equipment size, weight, cost, and power consumption are substantially reduced.

6. COMPARISON OF SSB WITH AM.

a. POWER COMPARISON OF SSB AND AM.

There is no single manner which can be used to evaluate the relative performance of AM. systems and SSB systems. Perhaps the most straightforward manner to make such a comparison is to determine the transmitter power necessary to produce a given signal-to-noise (s/n) ratio at the receiver for the two systems, under ideal propagating conditions. Signal-to-noise ratio is considered a fair comparison, because it is the s/n ratio which determines the intelligibility of the received signal. Figure 1-6 shows such a comparison between an AM. system and an SSB system where 100 percent, single-tone modulation is assumed.

Figure 1-6A shows the power spectrum for an AM. transmitter rated at 1 unit of carrier power. With 100 percent sine-wave modulation, such a transmitter will actually be producing 1.5 units of r-f power. There is .25 unit of power in each of the two sidebands and 1 unit of power in the carrier. This AM. transmitter is compared with an SSB transmitter rated at

.5 unit of peak-envelope-power (PEP). Peak-envelope-power is defined as the rms power developed at the crest of the modulation envelope. The SSB transmitter rated at .5 unit of peak-envelope-power will produce the same s n ratio in the output of the receiver as the AM. transmitter rated at 1 unit of carrier power.

The voltage vectors related to the AM. and SSB power spectrums are shown in figure 1-6B. The AM. voltage vectors show the upper and lower sideband voltages of .5 unit rotating in opposite directions around a carrier voltage of 1 unit. For AM. modulation, the resultant of the two sideband voltage vectors must always be directly in phase or directly out of phase with the carrier so that the resultant directly adds to or subtracts from the carrier. The resultant shown when the upper and lower sideband voltage are instantaneously in phase produces a peak-envelope-voltage (PEV) equal to twice the carrier voltage with 100 percent modulation. The .5 unit of voltage shown in each sideband vector produces the .25 unit of power shown in A, .25 unit of power being proportional to the square of .5 unit of voltage. The SSB voltage vector is a single vector of .7 unit of voltage at the upper sideband frequency. The .7 unit of voltage produces the .5 unit of power shown in A.

The r-f envelopes developed by the voltage vectors are shown in figure 1-6C. The r-f envelope of the AM. signal is shown to have a PEV of 2 units, the sum of the two sideband voltages plus the carrier voltage. This results in a PEP of 4 units of power. The PEV of the SSB signal is .7 unit of voltage with a resultant PEP of .5 unit of power.

When the r-f signal is demodulated in the AM. receiver, as shown in figure 1-6D, an audio voltage develops which is equivalent to the sum of the upper and the lower sideband voltages, in this case 1 unit of voltage. This voltage represents the output from the conventional, diode detector used in AM. receivers. Such detection is called coherent detection because the voltages of the two sidebands are added in the detector. When the r-f signal is demodulated in the SSB receiver, an audio voltage of .7 unit develops which is equivalent to the transmitter upper sideband signal. This signal is demodulated by heterodyning the r-f signal with the proper frequency to move the SSB signal down in the spectrum to its original audio position.

If a broadband noise level is chosen as .1 unit of voltage per 6 kc bandwidth, the AM. bandwidth, the same noise level is equal to .07 unit of voltage per 3 kc bandwidth, the SSB bandwidth. This is shown in figure 1-6E. These values represent the same noise power level per kc of bandwidth: that is, $.1^2/6$ equals $.07^2/3$. With this chosen noise level, the s/n ratio for the AM. system is 20 log s/n in terms of voltage, or 20 db. The s/n ratio for the SSB system is also 20 db, the same as for the AM. system. The 1/2 power unit of rated PEP for the SSB transmitter, therefore, produces the same signal intelligibility as the 1 power unit rated carrier power for the AM.

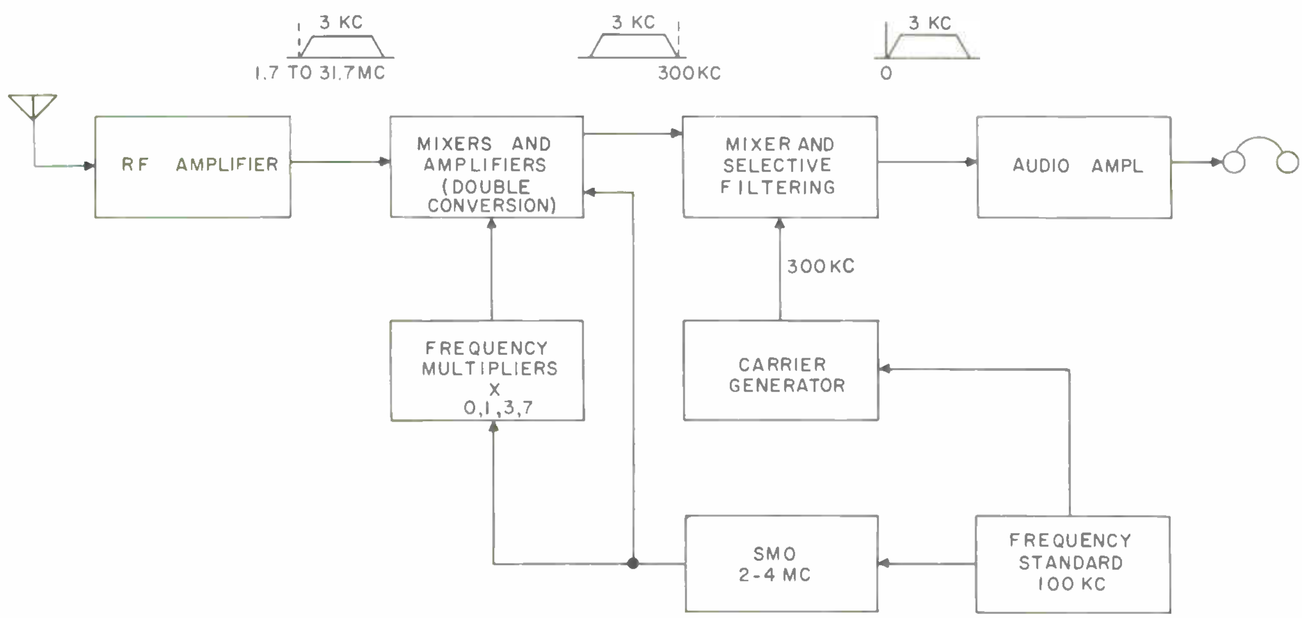


Figure 1-5. Functional Units of an SSB Receiving System

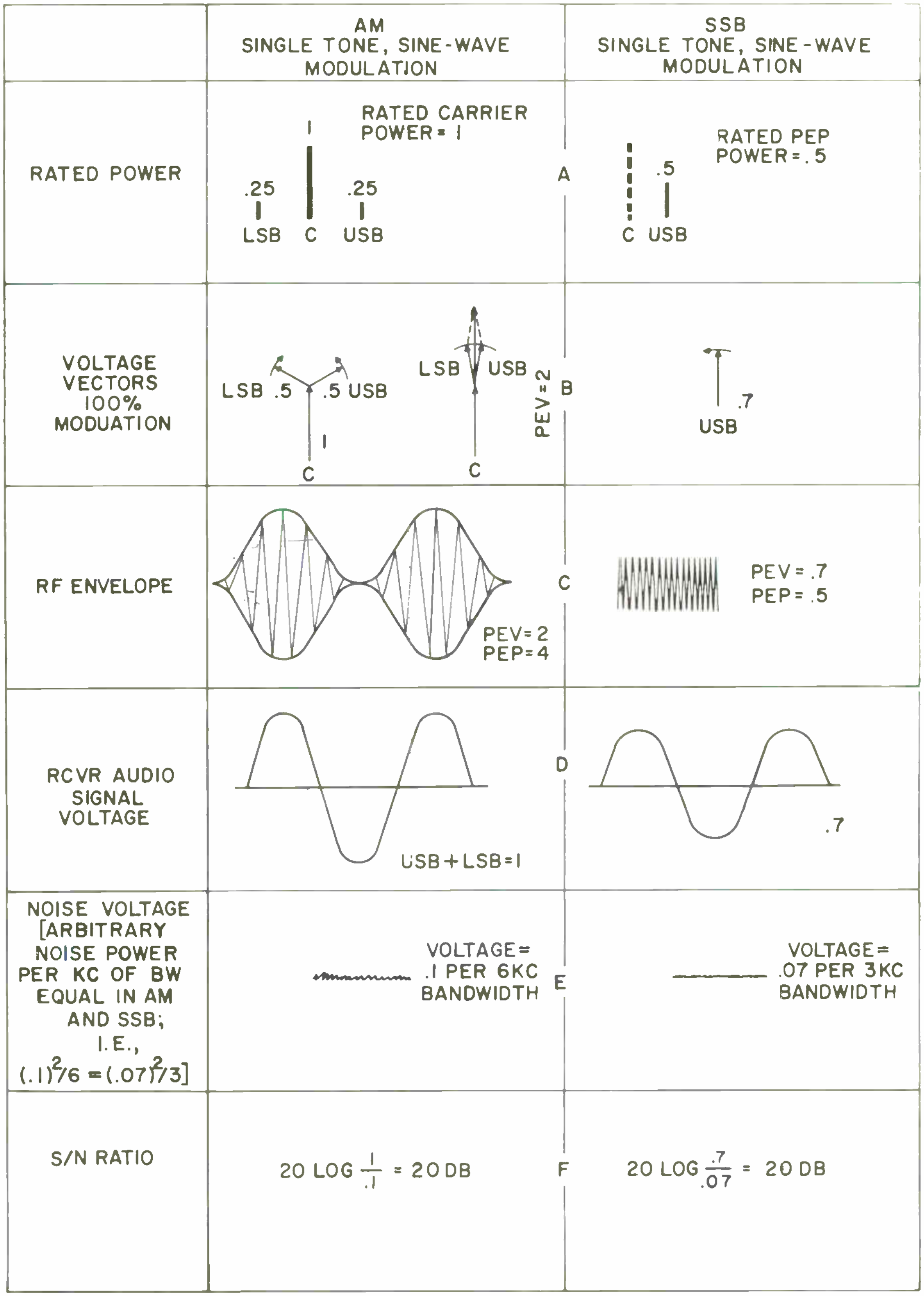


Figure 1-6. SSB and AM. Comparison with Equal Signal-to-Noise Ratio

transmitter. This conclusion can be restated as follows:

Under ideal propagating conditions but in the presence of broadband noise, an SSB and AM. system perform equally (same s/n ratio) if the total sideband power of the two transmitters is equal. This means that an SSB transmitter will perform as well as an AM. transmitter of twice the carrier power rating under ideal propagating conditions.

b. ANTENNA VOLTAGE COMPARISON OF SSB AND AM.

Of special importance in airborne and mobile installations where electrically small antennas are required, is the peak antenna voltage. In these installations, it is often the corona breakdown point of the antenna which is the limiting factor in equipment power. Figure 1-6C shows the r-f envelopes of an SSB transmitter and an AM. transmitter of equal performance under ideal conditions. The peak-envelope-voltage produced by these two transmitters is shown to be in the ratio 2 for the AM. transmitter to .7 for the SSB transmitter. This indicates that for equal performance under ideal conditions, the peak antenna voltage of the SSB system is approximately 1/3 that of the AM. system.

A comparison between the SSB power and the AM. power which can be radiated from an antenna of given dimensions is even more significant. If an antenna is chosen which will radiate 400 watts of peak-envelope-power, the AM. transmitter which may be used with this antenna must be rated at no more than 100 watts. This is true because the PEP of the AM. signal is four times the carrier power. An SSB transmitter rated at 400 watts of PEP, all of which is sideband power, may be used with this same antenna. Compared with the 50 watts of sideband power obtained from the AM. transmitter with a 100-watt carrier rating.

c. ADVANTAGE OF SSB WITH SELECTIVE FADING CONDITIONS

The power comparison between SSB and AM. given in the previous paragraph is based on ideal propagation conditions. However, with long distance transmission, AM. is subject to selective fading which causes severe distortion and a weaker received signal At times this can make the received signal unintelligible. An AM. transmission is subject to deterioration under these poor propagation conditions, because all three components of the transmitted signal, the upper sideband, lower sideband, and carrier must be received exactly as transmitted to realize fidelity and the theoretical power from the signal. Figure 1-7 shows the deterioration of an AM. signal with different types of selective fading.

The loss of one of the two transmitted sidebands results only in a loss of signal voltage from the demodulator. Even though some distortion results, such a loss is not basically detrimental to the signal, because one sideband contains the same intelligence as the other. However, since the AM. receiver operates on the broad bandwidth necessary to receive both sidebands, the noise level remains constant even though only one sideband is received. This is equivalent to a 6 db deterioration in s/n ratio out of the receiver. Although the loss of one of the two sidebands may be an extreme case, a proportional deterioration in s/n ratio results from the reduction in the level of one or both sidebands.

The most serious result of selective fading, and the most common, occurs when the carrier level is attenuated more than the sidebands. When this occurs, the carrier voltage at the receiver is less than the sum of the two sideband voltages. When the carrier is attenuated more than the sidebands, the r-f envelope does not retain its original shape, and distortion is extremely severe upon demodulation. This distortion results upon demodulation because a carrier voltage at least as strong as the sum of the two sideband voltages is required to properly demodulate the signal. The distortion resulting from a weak carrier can be overcome by use of the exalted carrier technique whereby the carrier is amplified separately and then reinserted before demodulation. In using the exalted carrier, the carrier must be reinserted close to the original phase of the AM. carrier.

Selective fading can also result in a shift between the relative phase position of the carrier and the sidebands. An AM. modulation is vectorally represented by two counter-rotating sideband vectors which rotate with respect to the carrier vector. The resultant of the sideband vectors is always directly in phase or directly out of phase with the carrier vector. In an extreme case, the carrier may be shifted 90° from its original position. When this occurs, the resultant of the sideband vectors is ±90° out of phase with the carrier vector. This results in converting the original AM. signal to a phase modulated signal. The envelope of the phase modulated signal bears no resemblance to the original AM. envelope and the conventional AM. detector will not produce an intelligible signal. Any shift in the carrier phase from its original phase relationship with respect to the sidebands will produce some phase modulation with a consequential loss of intelligibility in the audio signal. Such a carrier phase shift may be caused by poor propagating conditions. Such a carrier phase shift will also result from using the exalted carrier technique if the reinserted carrier is not close to its original phase, as previously mentioned.

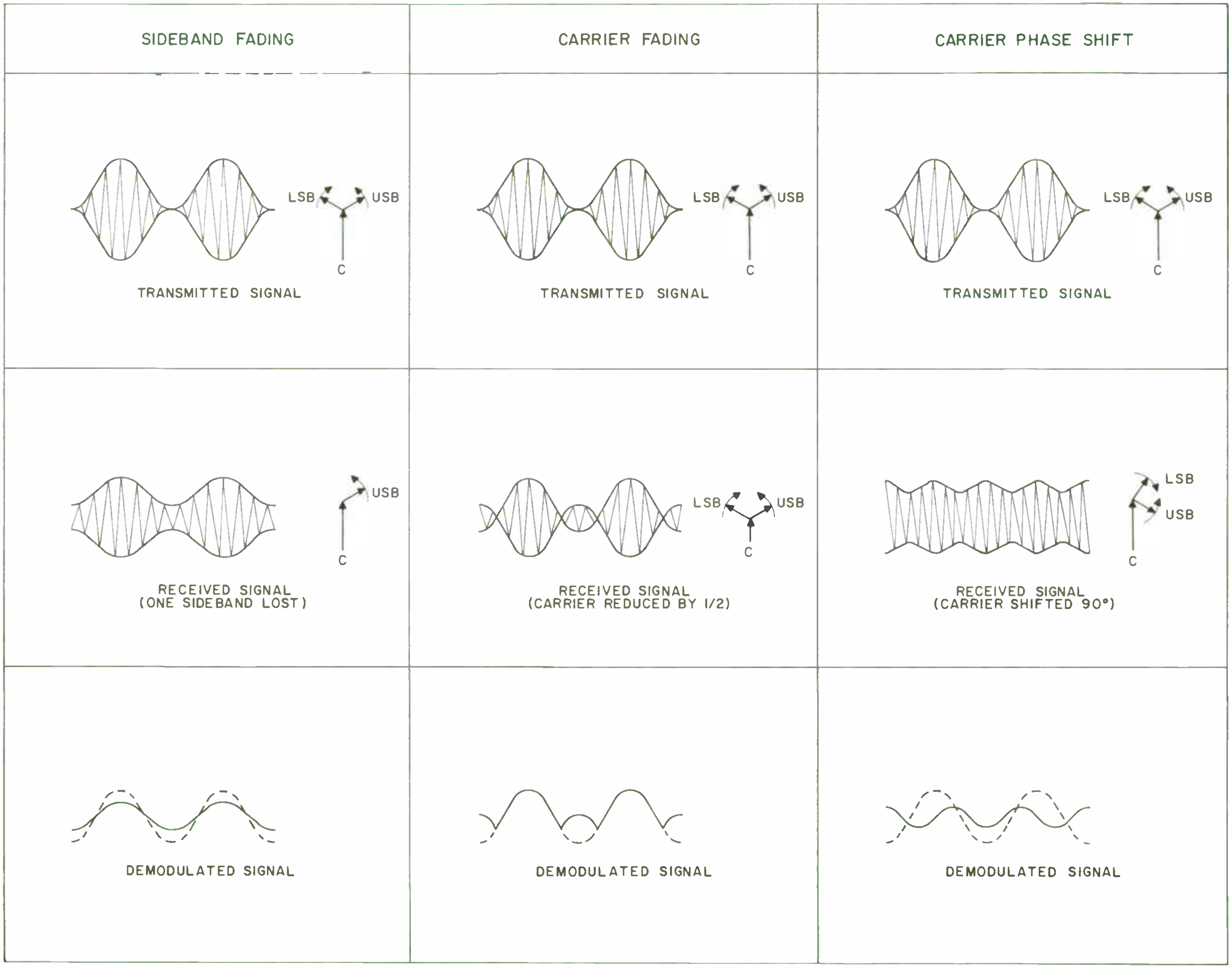


Figure 1-7. Deterioration of an AM. Signal with Selective Fading

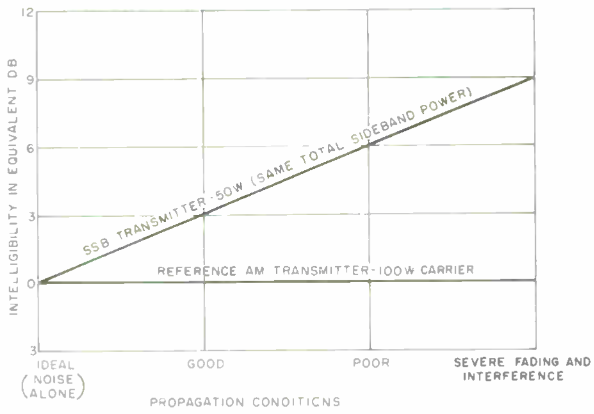


Figure 1-8. Relative Advantage of SSB over AM. with Limiting Propagating Conditions

An SSB signal is not subject to deterioration due to selective fading which varies either the amplitude or the phase relationship between the carrier and the two sidebands in the AM. transmission. Since only one sideband is transmitted in SSB, the received signal level does not depend upon the resultant amplitude of two sideband signals as it does in AM. Since the receiver signal does not depend upon a carrier level in SSB, no distortion can result from loss of carrier power. Since the receiver signal does not depend upon the phase relationship between the sideband signal and the carrier, no distortion can result from phase shift. Selective fading within the one sideband of the SSB system only changes the amplitude and the frequency response of the signal. It very rarely produces enough distortion to cause the received signal or voice to be unintelligible.

d. COMPARISON OF SSB WITH AM. UNDER LIMITING PROPAGATING CONDITIONS

One of the main advantages of SSB transmission over AM. transmission is obtained under limiting propagating conditions over a long-range path where communications are limited by the combination of noise, severe selective fading, and narrow-band interference. Figure 1-8 illustrates the results of an intelligibility study performed by rating the intelligibility of information received when operating the two systems under varying conditions of propagation.[1] The two transmitters compared have the same total sideband power. That is, a 100 watt AM. transmitter puts 1/4 of its rated carrier power in each of two sidebands, while a 50 watt SSB transmitter puts its full rated output in one sideband. This study shows that as propagation conditions worsen. and interference and fading become prevalent, the received SSB signal will provide up to a 9 db advantage over the AM. signal. The result of this study indicates that the SSB system will give from 0 to 9 db improvement under various conditions of propagation when total sideband power in SSB is equal to AM. It has been found that 3 of the possible 9 db advantage will be realized on the average contact. In other words, in normal use, an SSB transmitter rated at 100 watts (PEP) will give equal performance with an AM. transmitter rated at 400 watts carrier power. It should be pointed out that in this comparison the receiver bandwidth is just enough to accept the transmitted intelligence in each case and no speech processing is considered for SSB transmission.

e. COMPARISON OF AIRBORNE HIGH-FREQUENCY SYSTEMS

Figure 1-9 shows a comparison in weight, volume, input power, effective output power, and peak antenna voltage between Radio Set AN/ARC-38 and Radio Set AN/ARC-58. These sets are both airborne transceivers operating in the 2 to 30 mc, high-frequency range. The AM. set, AN/ARC-38, is rated at 100 watts r-f output, and the SSB set, AN/ARC-58, is rated at 1000 watts r-f output.

The effective output power of the SSB transceiver is shown to be 16 db higher than the AM. transceiver. This 16 db is equivalent to a power advantage of 40 to 1, which is an enormous advancement in the communication ability of an airborne system. In addition to the power advantage of the SSB system of significance in airborne equipment is the more efficient use of the antenna with the SSB system.

f. SUMMARY

For long-range communications in the low-, medium-, and high-frequency ranges, SSB is well suited because of its spectrum and power economy and because it is less susceptible to the effects of selective fading and interference than is AM. The principal advantages of SSB result from the elimination of the high-energy AM. carrier and from improved performance under unfavorable propagating conditions. On the average contact, an SSB transmitter will give equal performance to an AM. transmitter of four times the power rating. The advantage of SSB over AM. is most outstanding under unfavorable propagating conditions. For equal performance, the

[1] J. F. Honey, "Performance of AM. and SSB Communications," Tele-Tech, September 1953.

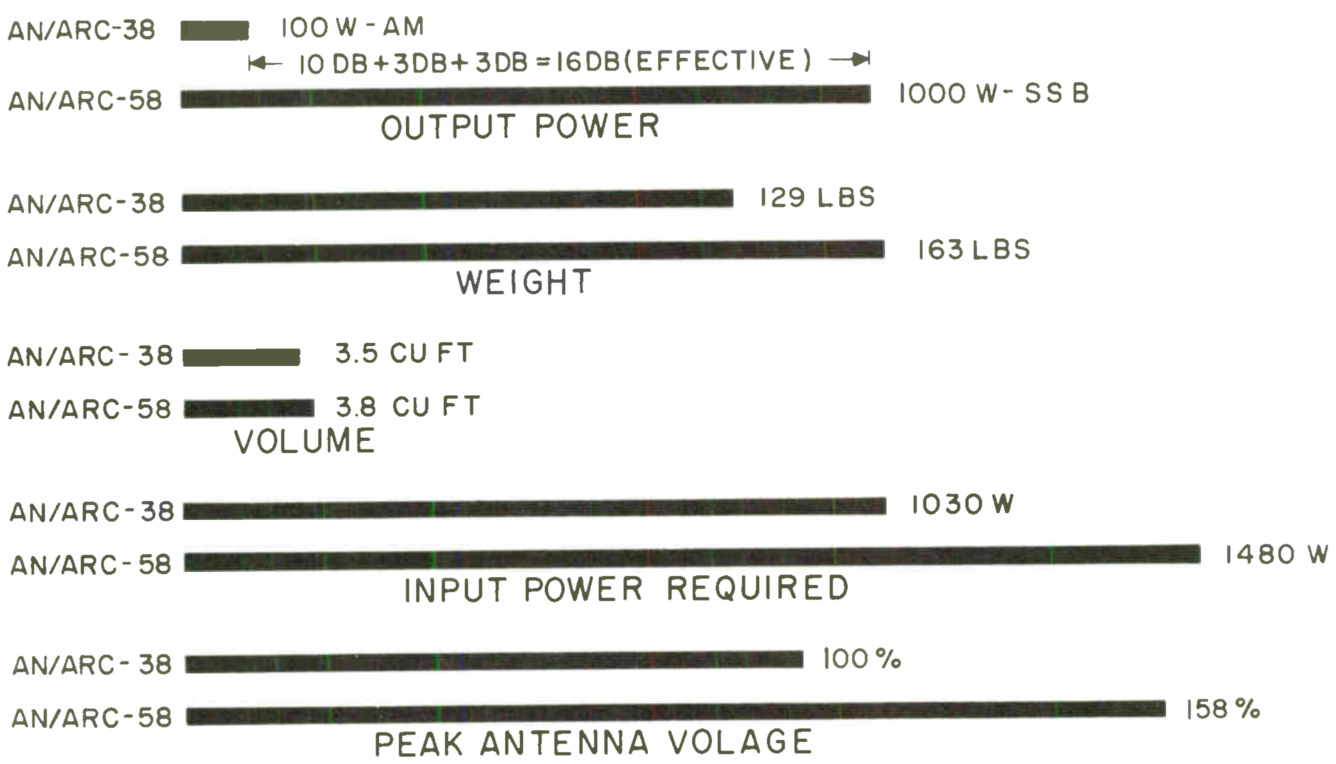


Figure 1-9. AN/ARC-38 and AN/ARC-58 Comparison

size, weight, power input, and peak antenna voltage of the SSB transmitter is significantly less than the AM. transmitter.

7. COMPARISON OF SSB WITH FM

Although much experimental work has been done to evaluate the performance of SSB systems with AM. systems, very little work has been done to evaluate the performance of SSB systems with FM systems. However, figure 1-10 shows the predicted result of one such study based on a mobile FM system as compared to a mobile SSB system of equal physical size.[1] The two systems compared also used the same output tubes to their full capacity so that the final r-f amplifiers dissipated the same power during normal speech loading. The study is complicated by evaluating the effects of speech processing, such as clipping and pre-emphasis, with its resultant distortion. Such speech processing is essential in the FM system but has little benefit in the SSB system.

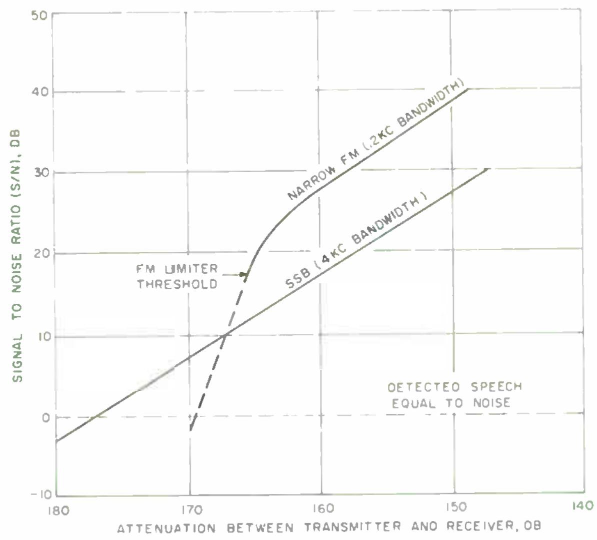


Figure 1-10. SSB Performance Compared with FM

[1] H. Magnuski and W. Firestone, "Comparison of SSB and FM for VHF Mobile Service," Proceedings of the IRE, December 1956.

Figure 1-10 shows the signal-to-noise ratio in decibels on the y-axis and the attenuation between transmitter and receiver in decibels on the x-axis. This graph indicates that with between 150 to 160 db of attenuation between the transmitter and receiver, a strong signal, the narrow-band FM system provides a better s/n ratio than the SSB system. Under weak signal condition, from 168 and higher db of attenuation between transmitter and receiver, the s/n ratio of the FM system falls off rapidly, and the SSB system provides the best s/n ratio. This fall-off in the FM s/n ratio results when the signal level drops below the level required for operation of the limiter in the FM receiver.

The conclusions which can be drawn from figure 1-10 are as follows: (1) For strong signals, the FM system will provide a better s/n ratio than the SSB system. However, this is not an important advantage because when the s/n is high, a still better s/n ratio will not improve intelligibility significantly. (2) For weak signals, the SSB system will provide an intelligible signal where the FM system will not. (3) The SSB system provides three times the savings in spectrum space as the narrow-band FM system.

8. NATURE OF SINGLE-SIDEBAND SIGNALS

a. INTRODUCTION

As defined in paragraph 2, chapter 2, a single-sideband signal is an audio signal converted to a radio frequency, with or without inversion. To facilitate illustrating the manner and the results of this conversion, it is necessary to use pure sine-wave tones, rather than the very complex waveforms of the human voice. For this reason single tones or combinations of two or three tones are generally used in the following discussion.

b. THE SSB GENERATOR

The most familiar SSB generator consists of a balanced modulator followed by an extremely selective mechanical filter as shown in figure 1-11. The balanced modulator produces basically two output frequencies: (1) An upper sideband frequency equal to the injected i-f frequency plus the input audio frequency. (2) A lower sideband frequency equal to the injected i-f frequency minus the input audio frequency. Theoretically, the injected i-f frequency is balanced out in the modulator so that it does not appear in the output.

It should be especially noted that the generation of undesirable products occur in any mixing operation as well as the generation of the desired products. The equipment must be so designed to minimize the generation of undesirable products and to attenuate those undesirable products which are generated. This is accomplished by designing good linear operating characteristics into the equipment to minimize the generation of undesirable frequencies and by choosing injection frequencies which will facilitate suppression of undesirable frequencies.

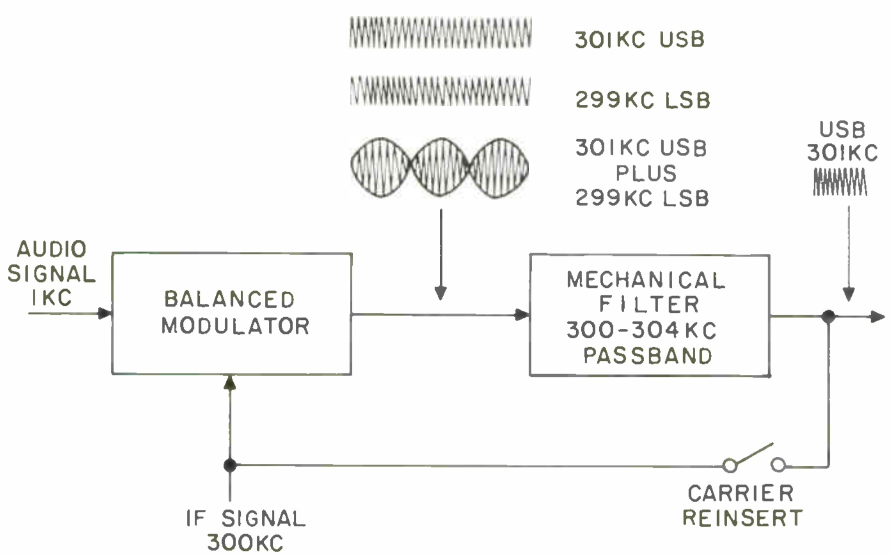


Figure 1-11. Filter-Type SSB Generator

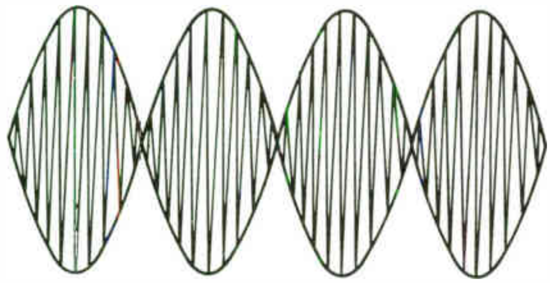


Figure 1-12. Single-Tone, Balanced Modulator Output

It should also be noted that the i-f carrier injected into the balanced modulator is only theoretically canceled from the output. Practical design considerations determine the extent to which the carrier can be balanced out. Present balanced modulators, using controlled carrier leak to balance out uncontrolled carrier leak, result in carrier suppression of from 30 db to 40 db below the PEP of the sidebands. Further suppression of the carrier by the SSB filter results in an additional 20 db of carrier suppression. Total carrier suppression of from 50 db to 60 db can, therefore, reasonably be expected from the transmitter system.

c. GENERATING THE SINGLE-TONE SSB WAVEFORM

The most fundamental SSB waveform is generated from the single audio tone. This tone is processed through the SSB generator to produce a single i-f frequency. As pointed out in paragraph 4, chapter 1, the SSB signal is actually generated at an i-f frequency and is subsequently converted up in frequency to the transmitted r-f frequency. It is the generation of the SSB signal at the i-f frequency with which we are concerned.

Figures 1-12 and 1-13 show the waveforms obtained in a filter-type SSB generator. The audio tone injected into the balanced modulator is 1 kc and the i-f frequency injected is 300 kc. The output from the balanced modulator contains the 299 kc lower sideband and 301 kc upper sideband frequencies. These two sideband frequencies, being of equal amplitude, produce the characteristic half sine-wave envelope shown in figure 1-12. The repetition rate of this envelope with a 1-kc tone is 2 kc, the difference between the two frequencies represented by the envelope. This i-f signal, which contains both the upper sideband and lower sideband signal, is called a double-sideband signal (DSB).

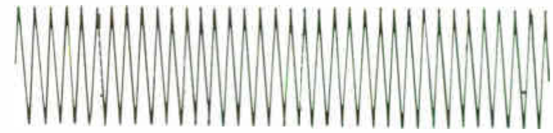


Figure 1-13. Single-Tone Balanced Modulator Output After Filtering Out the LSB

By passing the DSB signal through a highly selective filter with a 300 kc to 303 kc passband, the upper sideband signal is passed while the lower sideband signal is attenuated. The 301 kc signal which remains is the upper sideband signal and appears as shown in figure 1-13. Note that the SSB signal remaining is a pure sine wave when a single-tone audio signal is used for modulation. This SSB signal is displaced up in the spectrum from its original audio frequency by an amount equal to the carrier frequency, in this case 300 kc. This SSB signal can be demodulated at the receiver only by converting it back down in the frequency spectrum. This is done by mixing it with an independent 300 kc i-f signal at the receiver.

d. GENERATING THE SSB WAVEFORM OF A SINGLE TONE WITH CARRIER

From the single-tone SSB signal without carrier, it is a simple step to generate the single-tone SSB signal with carrier. This is done by reinserting the carrier after the filtering operation, as shown in figure 1-11. When the carrier reinserted is of the same amplitude as the SSB signal, the waveform shown in figure 1-14 results. Note that this waveform is similar to the double-sideband signal obtained directly out of the balanced modulator, as shown in

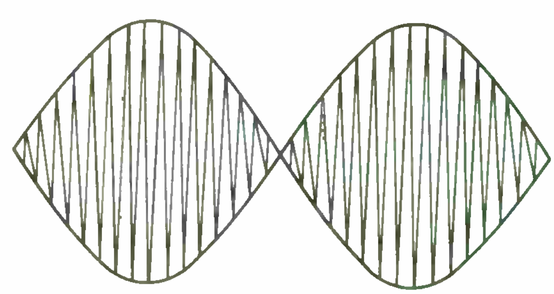


Figure 1-14. Single-Tone SSB Signal with Carrier--Carrier Equal in Amplitude to Tone

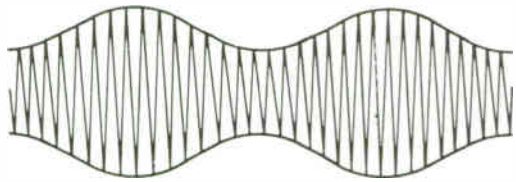


Figure 1-15. Single-Tone SSB Signal with Carrier--Carrier 10 DB Below Tone

figure 1-12. However, the frequency components of the two waveforms are not the same. The frequency components of the SSB signal with carrier are 301 kc and 300 kc when a 1-kc audio signal is used. A voice SSB signal with full carrier can be demodulated with a conventional diode detector used in AM. receivers without serious distortion or loss of intelligibility.

If the reinserted carrier is such that the carrier level is less than the level of the single-tone SSB signal, the waveform shown in figure 1-15 results. To successfully demodulate this signal, the carrier must be separated, amplified, exalted, and reinserted in the receiver, or locally supplied. The separate carrier amplification should be sufficient to raise the reinserted carrier to a level greater than the level of the sideband signal. The waveform shown in figure 1-15 represents the waveform used in the SSB with pilot carrier systems. The exalted carrier technique is used to demodulate such a signal.

e. GENERATING THE TWO-TONE SSB WAVEFORM

The two-tone SSB waveform is generated by combining two audio tones and then injecting this two-tone signal into the balanced modulator. One sideband is then suppressed by the filter, leaving the SSB waveform shown in figure 1-16. This two-tone SSB signal is seen to be similar to the single-tone DSB signal as well as the SSB signal with full carrier. However,

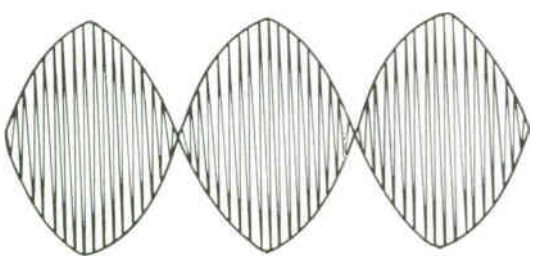


Figure 1-16. Two-Tone SSB Signal-Tones of Equal Amplitude

the two-tone SSB signal contains a different two frequencies than either of the other two. In the two-tone SSB signal shown in figure 1-16, 1 kc and 2 kc audio signals of equal amplitude are injected into the balanced modulator. After filtering, this results in a two-tone SSB signal containing frequencies of 301 kc and 302 kc. If a pilot carrier is reinserted with the two-tone test signal, the pilot carrier will be indicated by the appearance of a sine-wave ripple on the two-tone waveform. This waveform is shown in figure 1-17.

The generation of this two-tone envelope can be shown clearly with vectors representing the two audio frequencies, as shown in figure 1-18. When the two vectors are exactly opposite in phase, the envelope value is zero. When the two vectors are exactly in phase, the envelope value is maximum. This generates the half sine-wave shape of the two-tone SSB envelope which has a repetition rate equal to the difference between the two audio tones.

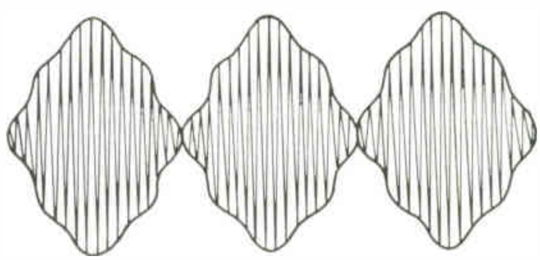
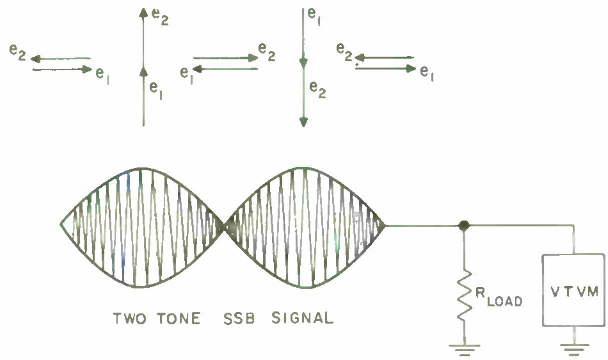


Figure 1-17. Two-Tone SSB Signal with Small Reinserted Pilot Carrier

The two-tone SSB envelope is of special importance because it is from this envelope that power output from an SSB system is usually determined. An SSB transmitter is rated in peak-envelope-power output with the power measured with a two equal-tone test signal. With such a test signal, the actual watts dissipated in the load are one-half the peak-envelope-power. This is shown in figure 1-18. When the half sine-wave signal is fed into a load, a peak-reading, rms-calibrated vtvm across the load indicates the rms value of the peak-envelope-voltage. This voltmeter reading is equal to the in-phase sum of $e_1 + e_2$, where e_1 and e_2 are the rms voltages of the two tones. Since in the two-tone test signal e_1 equals e_2, the PEP equals $(2e_1)^2/R$ or $(2e_2)^2/R$. The average power dissipated in the load must equal the sum of the power represented by each tone, $e_1^2/R + e_2^2/R$, $4e_1^2R$ or $4e_2^2/R$. Therefore, with a two equal-tone SSB test signal, the average power dissipated in the load is equal



$V_{vtvm} = (e_1 + e_2)$,

with e_1 and e_2 in phase and rms values

$PEP = V_{vtvm}^2/R_{load} = 4e_1^2/R$ or $4e_2^2/R$,

where $e_1 = e_2$

$P_{average} = e_1^2/R + e_2^2/R = 2e_1^2/R$ or $2e_2^2/R$

Therefore: (1) $PEP = V_{vtvm}^2/R$

(2) $P_{average} = 1/2\ PEP$

(3) $P_{tone\,1}$ or $P_{tone\,2} = 1/4\ PEP$

Figure 1-18. Power Measurements from Two-Tone SSB Test Signal

to 1/2 of the PEP, and the power in each tone is equal to 1/4 of the PEP. Peak-envelope-power can be determined from the relationship "$PEP = V^2_{vtvm}/R$;" the average power can be determined from the relationship "$P_{average} = 1/2\ V^2_{vtvm}/R$." This is true only where the vtvm used is a peak-reading, rms-calibrated voltmeter. Similar measurements can be made using an a-c ammeter in series with the load instead of the vtvm across the load.

The above analysis can be carried further to show that with a three equal-tone SSB test signal, the power in each tone is 1/9 of the PEP, and the average power dissipated in the load is 1/3 the PEP. These relationships are true only if there is no distortion of the SSB envelope, but since distortion is usually small, its effects are usually neglected.

f. GENERATING THE SQUARE WAVEFORM

Transmitting an audio square wave at a radio frequency imposes severe requirements on any transmitting system. This is true because the square wave is composed of an infinite number of odd-order harmonics of the fundamental frequency of the square wave. Therefore, to transmit such a signal without distortion requires an infinite bandwidth, an infinite spectrum. This, of course, is impossible because tuned circuits will not pass an infinite bandwidth. The idealized SSB square wave, where all frequency components are present, shown in figure 1-19, indicates that the SSB signal requires infinite amplitude as well as infinite bandwidth. This occurs because the harmonically related SSB components will add vectorally to infinity when the modulating signal switches from maximum positive to maximum negative and vice versa. This infinite amplitude is not present in an AM. envelope, because the AM. envelope contains both sidebands with the frequency components in one sideband counter-rotating vectorally from the frequency components in the other sideband. The result is, then, when the resultant amplitude of one sideband is plus infinity; the resultant amplitude of the other sideband is minus infinity, which produces a net amplitude of zero.

Figure 1-20 shows the SSB envelope which results from severely clipping a 300 cps sine wave. The clipping level is such that the modulating signal is essentially a square wave. In generating the SSB envelope from the modulating signal, all harmonics above the ninth are removed by the highly selective SSB filter. This illustration shows the maximum peaking that can occur with a 300 to 3000 cps passband. All harmonics of an input sine wave above 1000 cps will be removed by the SSB filter. The output envelope for input frequencies above 1000 cps is constant in amplitude and 27% higher than the clipping level. Since AM. has this 27% rise also, the maximum theoretical difference is 1.8 between AM. and SSB. Fortunately it is not necessary to transmit a square wave audio signal. A clipped speech wave

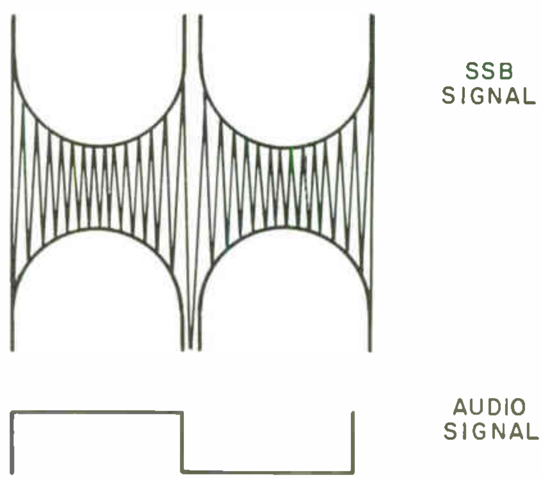


Figure 1-19. Square Wave SSB Signal--All Frequency Components Present

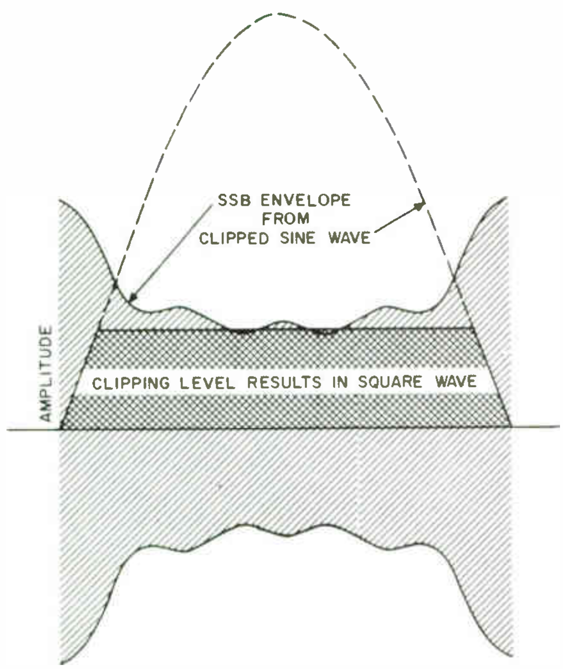


Figure 1-20. SSB Envelope Developed from 300 CPS, Clipped Sine-Wave (Harmonics above 9th Attenuated)

cannot be represented by an audio square wave because of their vastly different relationship of frequency components. Data and teletype signals are not fed to SSB transmitters as square waves either. A tone such as 1500 cps may be keyed on and off or frequency shifted at, say, a 40 cps rate. The keying wave shape at 40 cps may be a square wave but this produces double sidebands ±40 cps, ±120 cps and ±200 cps from the 1500 cps tone. These sidebands are transmitted along with the 1500 cps "carrier" tone and no envelope peaking results.

g. GENERATING THE VOICE WAVEFORM

The human voice produces a complex waveform which can be represented by numerous frequency components of various amplitudes and various instantaneous phase relationships. No human voice is exactly like another voice, but statistical averages concerning the frequencies and amplitudes in the human voice can be determined. The average power level of speech is relatively low when compared to the peak power level. An audio frequency waveform of an ā sound is shown in figure 1-21. This same ā sound, raised in frequency, is shown in figure 1-22 as it appears as an SSB signal. From the "Christmas-tree" shape of these waveforms, it is evident that the peak power, which is related to the peak voltage of a waveform is considerably higher than the average power.

Over-all transmission efficiency depends upon the average power transmitted, while transmitter power is limited to the peak power capability of the transmitter. Therefore, for voice transmission, it behooves the transmitter designer to use speech-processing circuits which will increase the average power in the voice signal without increasing the peak power. This can be done in three different ways: (1) by clipping the power peaks, (2) by emphasizing the low-power, high-frequency components of the speech signal and attenuating the high-power, low-frequency components of the speech signal, and (3) by using automatic-gain-control circuits to keep the signal level near the maximum capability of the transmitter.

Figure 1-23 shows a power vs frequency distribution curve for the average human voice, after filtering below 200 cps and above 3000 cps. This curve shows that the high-power components of speech are concentrated in the low frequencies. Fortunately, it is the low-frequency components of speech which contribute little to intelligibility since these frequencies are concentrated in the vowel sounds. The low frequencies, therefore may be attenuated without undue loss

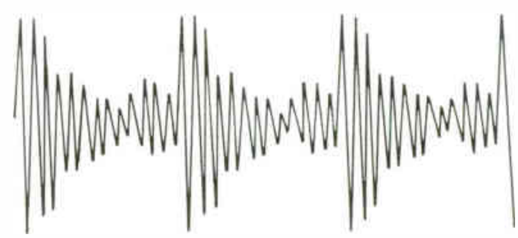


Figure 1-21. Voice Signal at Audio Frequency--ā Sound

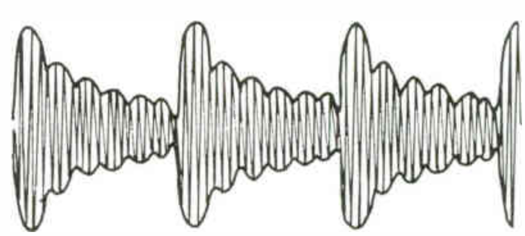


Figure 1-22. SSB Voice Signal--ā Sound

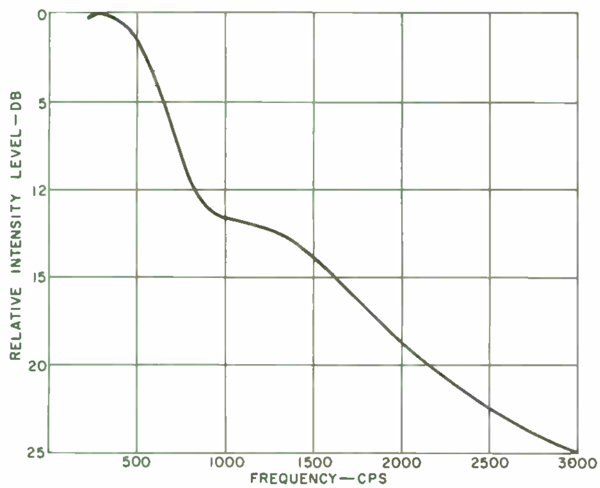


Figure 1-23. Power Distribution in Speech Frequencies--Low and High Frequencies Removed

of intelligibility of the speech. The low-power, high-frequency components present in a voice signal can be pre-emphasized to increase the average power of the signal. Since it is the high-frequency components which are predominate in the consonant sounds, some emphasis of the high frequencies will improve intelligibility. However, to emphasize the high frequencies sufficiently to raise the average power level significantly would require compatible de-emphasis at the receiver to prevent loss of fidelity.

A voice signal consists of many frequency components and the resultant peaking is less. For this reason audio peak clipping can be used with substantial effectiveness. Clipping the SSB signal at i-f is somewhat better because there are fewer clipping products within the passband. As a result the voice signal sounds better and is more intelligible. Additional filtering following the i-f clipper is required to remove out-of-band clipping products, however. Automatic load control is more effective and more simply accomplished than speech clipping.

Speech-processing methods are being reinvestigated in relationship to SSB transmission to determine the most suitable method or combination of methods. Several circuits are presently used in SSB transmitters which effect some speech processing, although the primary purpose of most of these circuits is to process the input signal to prevent overdriving the power amplifier. These circuits include the following: (1) Automatic-load-control to maintain signal peaks at the maximum rating of the power amplifier. (2) Speech compression, along with some clipping, to maintain a constant signal level to the single-sideband generator. (3) Highly-selective filters used in filter-type SSB exciters attenuate some of the high-power, low-frequency components of the voice signal. There are also several speech processing circuits under investigation which, if effective and practical, will be used to improve the efficiency of voice transmission. These circuits include (1) increased audio clipping with additional filtering to remove the harmonics generated, (2) reduction of the power level of frequencies below 1000 cps by shaping the audio amplifier characteristics for low-frequency roll-off, and (3) use of speech clipping at an i-f level where the generated harmonics can be more easily filtered. See paragraph 2-2a for input signal processing circuits used in an SSB exciter.

9. MECHANICAL FILTERS

Both SSB transmitters and SSB receivers require very selective bandpass filters in the region of 100 kc to 500 kc. In receivers, a high order of adjacent channel rejection is required if channels are to be closely spaced to conserve spectrum space. In SSB transmitters, the signal bandwidth must be limited sharply in order to pass the desired sideband and reject the other sideband. The filter used, therefore, must have very steep skirt characteristics and a flat bandpass characteristic. These filter requirements are met by LC filters, crystal filters, and mechanical filters. Until recently, crystal filters used in commercial SSB equipment were in the 100-kc range. These filters have excellent selectivity and stability characteristics, but their large size makes them subject to shock or vibration deterioration and their cost is quite high. Newer crystal filters are being developed which have extended frequency range and are smaller. These newer crystal filters are more acceptable for use in SSB equipment. LC filters have been used at i-f frequencies in the region of 20 kc. However, generation of the SSB signal at this frequency requires an additional mixing stage to obtain a transmitting frequency in the high-frequency range. For this reason, LC filters are not widely used. The recent advancements in the development of the mechanical filter have led to their acceptance in SSB equipment. These filters have excellent rejection characteristics, are extremely rugged, and are small enough to be compatible with miniaturization of equipment. Also to the advantage of the mechanical filter is a Q in the order of 10,000 which is about 100 times the Q obtainable with electrical elements.

Although the commercial use of mechanical filters is relatively new, the basic principles upon which they are based is well established. The mechanical filter is a mechanically resonant device which receives electrical energy, converts it into mechanical

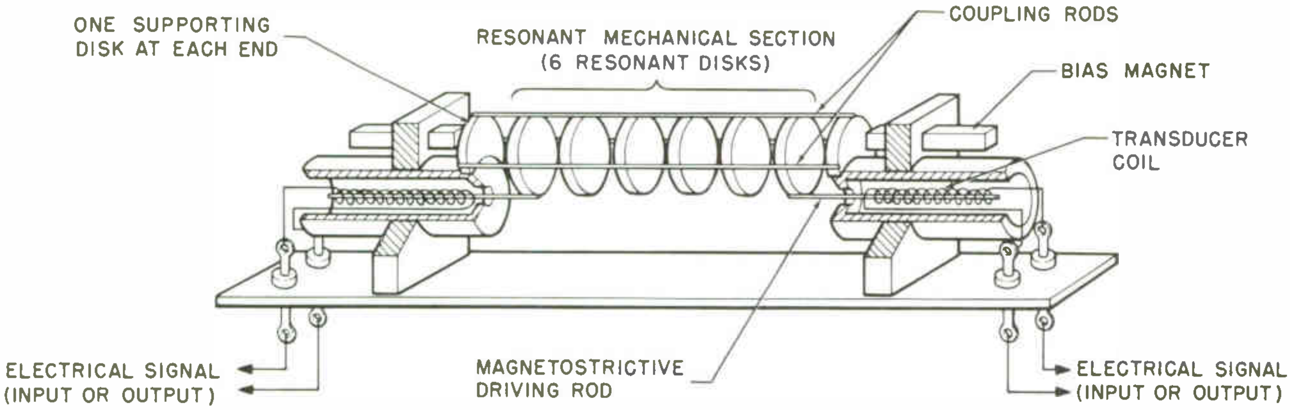


Figure 1-24. Elements of a Mechanical Filter

vibration, then converts the mechanical energy back into electrical energy at the output. The mechanical filter consists of basically four elements: (1) an input transducer which converts the electrical input into mechanical oscillations, (2) metal disks which are mechanically resonant, (3) coupling rods which couple the metal disks, and (4) an output transducer which converts the mechanical oscillations back into electrical oscillations. Figure 1-24 shows the elements of the mechanical filter, and figure 1-25 shows the electrical analogy of the mechanical filter. In the electrical analogy the series resonant circuits L_1C_1 represent the metal disks, the coupling capacitors C_2 represent the coupling rods, and the input and output resistances R represent the matching mechanical loads.

The transducer, which converts electrical energy into mechanical energy and vice versa, may be either a magnetostrictive device or an electrostrictive device. The magnetostrictive transducer is based on the principle that certain materials elongate or shorten when in the presence of a magnetic field. Therefore, if an electrical signal is sent through a coil which contains the magnetostrictive material as the core, the electrical oscillation will be converted into mechanical oscillation. The mechanical oscillation can then be used to drive the mechanical elements of the filter. The electrostrictive transducer is based on the principle that certain materials, such as piezoelectric crystals, will compress when subjected to an electric current. In practice, the magnetostrictive transducer is more commonly used. The transducer not only converts electrical energy into mechanical energy and vice versa; it also provides proper termination for the mechanical network. Both of these functions must be considered in transducer design.

From the electrical equivalent circuit, it is seen that the center frequency of the mechanical filter is

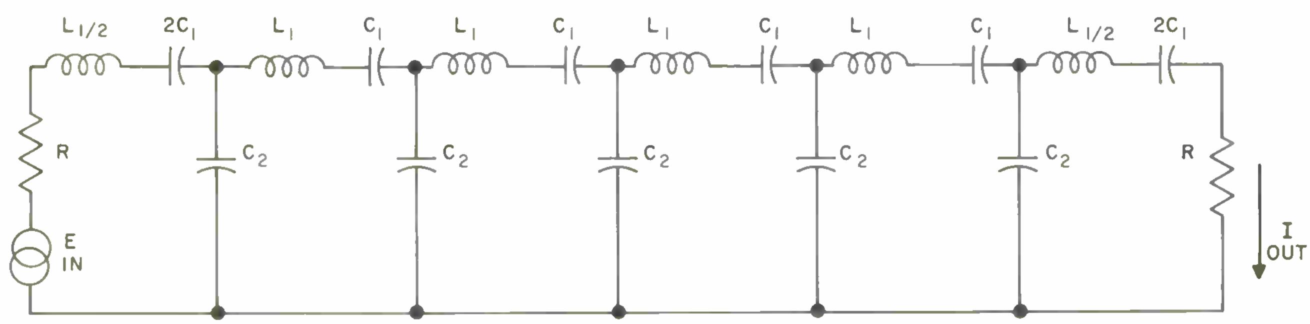


Figure 1-25. Electrical Analogy of a Mechanical Filter

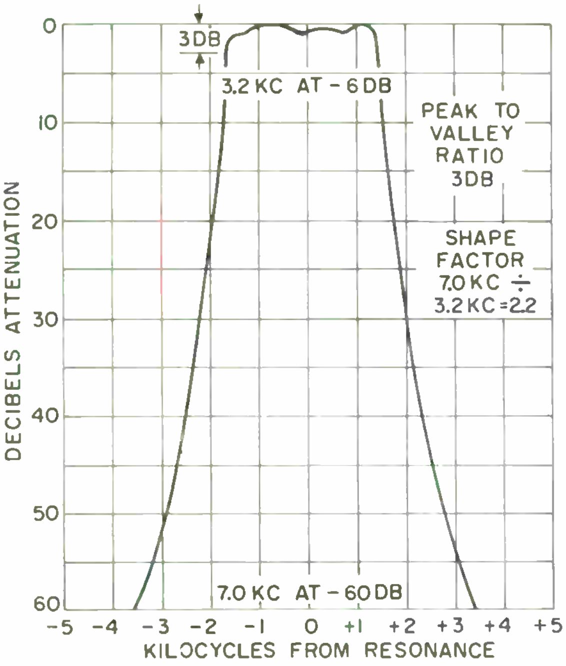


Figure 1-26. Mechanical Filter Characteristic Curve

determined by the metal disks which represent the series resonant circuit L_1C_1. In practice, filters between 50 kc and 600 kc can be manufactured. This by no means indicates mechanical filter limitations, but is merely the area of design concentration in a relatively new field. Since each disk represents a series resonant circuit, it follows that increasing the number of disks will increase skirt selectivity of the filter. Skirt selectivity is specified as shape factor which is the ratio (bandpass 60 db below peak)/(bandpass 6 db below peak). Practical manufacturing presently limits the number of disks to eight or nine in a mechanical filter. A six-disk filter has a shape factor of approximately 2.2, a seven-disk filter a shape factor of approximately 1.85, a nine-disk filter shape factor of approximately 1.5. The future development of mechanical filters promises even a faster rate of cutoff.

In the equivalent circuit, the coupling capacitors C_2 represent the rods which couple the disks. By varying the mechanical coupling between the disks, that is, making the coupling rods larger or smaller, the bandwidth of the filter is varied. Because the bandwidth varies approximately as the total area of the coupling wires, the bandwidth can be increased by either using larger or more coupling rods. Mechanical filters with bandwidths as narrow as 0.5 kc and as wide as 35 kc are practical in the 100 kc to 500 kc range.

Although an ideal filter would have a flat "nose" or passband, practical limitations prevent the ideal from being obtained. The term "ripple amplitude" or "peak-to-valley ratio" is used to specify the nose characteristic of the filter. The peak-to-valley ratio is the ratio of maximum to minimum output level across the useful frequency range of the filter (figure 1-26). A peak-to-valley ratio of 3 db can be obtained on a production basis by automatic control of materials and assembly. Mechanical filters with a peak-to-valley ratio of 1 db can be produced with accurate adjustment of filter elements.

Spurious responses occur in mechanical filters due to mechanical resonances other than the desired resonance. By proper design, spurious resonances can be kept far enough from the passband to permit other tuned circuits in the system to attenuate the spurious responses.

Other mechanical filter characteristics of importance include insertion loss, transmission loss, transfer impedance, input impedance, and output impedance. Since the input and output transducers of the mechanical filter are inductive, parallel external capacitors must be used to resonate the input and output impedances at the filter frequency. With such capacitors added, the input and output impedances are largely resistive and range between 1000 ohms to 50,000 ohms. The insertion loss is measured with both the source and load impedance matched to the input and output impedance of the filter. The value of insertion loss ranges between 2 db and 16 db, depending upon the type of transducer. The transmission loss is an indication of the filter loss with

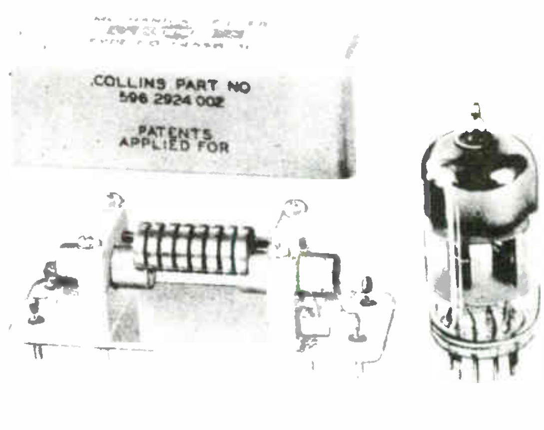


Figure 1-27. Size Comparison Between a Mechanical Filter and a Miniature Tube

source and load impedances mismatched. The transmission loss is of importance when using a mechanical filter in pentode i-f amplifiers where both source and load impedance are much greater than the filter impedances. The transfer impedance is useful to determine the over-all gain of a pentode amplifier stage which utilizes a mechanical filter. The transfer impedance of the filter multiplied by the transconductance of the pentode gives the gain of the amplifier stage.

The physical size of the mechanical filter makes it especially useful for modular and miniaturized construction. Figure 1-27 shows a mechanical filter compared with a miniature tube. The mechanical filter is about 1 inch square by 3 inches long. More recent development has resulted in a smaller tubular filter which is about 1/2 inch in diameter by 1-3/4 inches long.

Mechanical filter types other than the disk type are presently being used. These include the plate type which is a series of flat plates assembled in a ladder arrangement. Another type which has recently been developed is the neck-and-slug type. This filter consists of a long cylinder which is turned down to form the necks which couple the remaining slugs. All mechanical filters are similar in that they employ mechanical resonance. Mechanical filters differ in that they employ various modes of mechanical oscillation to achieve their purpose. They may also use different types of transducers.

Collins 618T Single Sideband Transceiver Provides Communication on All Aeronautical Bands in the 2 to 30 Mc Range with 400 Watts PEP

CHAPTER 2

SINGLE-SIDEBAND EXCITERS

1. SINGLE-SIDEBAND EXCITER CONSIDERATIONS

The single-sideband exciter must translate the incoming audio frequency signal to a band of frequencies in the r-f range. A single-sideband exciter is, in fact, a complete transmitter in itself. It must generate a radio-frequency sideband from an audio input signal, translate this r-f sideband to the final output frequency, and provide sufficient amplification to drive the r-f power amplifier. A functional diagram of a typical single-sideband exciter is shown in figure 2-1.

To generate the r-f sideband of frequencies, the single-sideband exciter uses low-level modulation and obtains the desired output level through the use of linear amplifiers. Low-level modulation is used since the carrier and unwanted sideband must be suppressed. The best suppression is obtained at a fixed low frequency since the problems involved in building a high-level balanced modulator, capable of working over a wide frequency range, appears to be insurmountable.

The most desirable performance characteristics of a single-sideband exciter would be the ability to generate the desired sideband, completely suppress the undesired sideband, and suppress the carrier. Practical design permits suppressing the undesired sideband and carrier frequencies by more than 40 db.

Careful consideration must be given to the amount of frequency spectrum space occupied by the generated signal. The band of side frequencies is normally held to 4 kc in single-sideband exciters for communication purposes.

The two basic systems for generating single-sideband signals are the filter system, shown in figure 2-2A, and the phase shift system, shown in figure 2-2B.

a. FILTER SYSTEM

The filter system uses a band-pass filter having sufficient selectivity to pass one sideband and reject the other. Filters having such characteristics are normally constructed for relatively low frequencies, below 500 kc, but recent developments in crystal filter research has produced workable filters at 5 megacycles. The carrier generator output is combined with the audio output of a speech amplifier in a balanced modulator. The upper and lower sidebands appear in the output, but the carrier is suppressed. One of the sidebands is passed by the filter and the other rejected, so that a single-sideband signal is applied to the mixer. The signal is mixed with the output of a high-frequency r-f oscillator to produce the desired output frequency. The problem of undesired mixer products arising in the frequency conversions of single-sideband signals becomes important. Either balanced modulators or sufficient selectivity must be used to attenuate these frequencies in the output and minimize the possibility of unwanted radiations.

b. PHASE SHIFT SYSTEM

The principle involved in the generation of a single-sideband signal by the phase shift method,

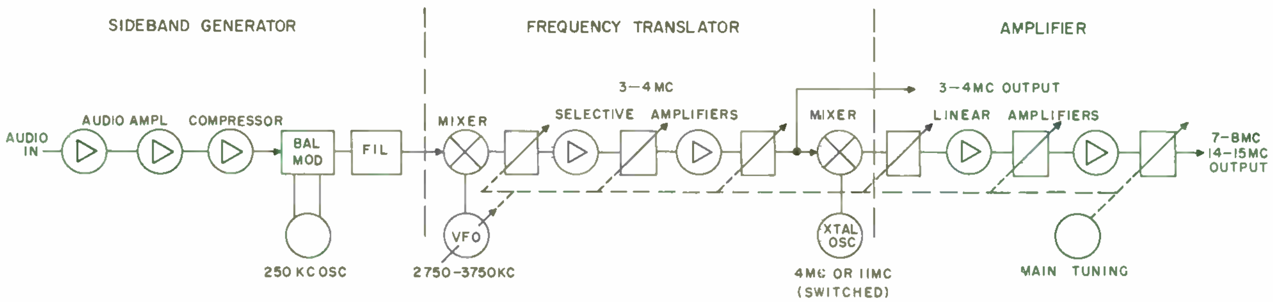


Figure 2-1. Typical Single-Sideband Exciter, Functional Diagram

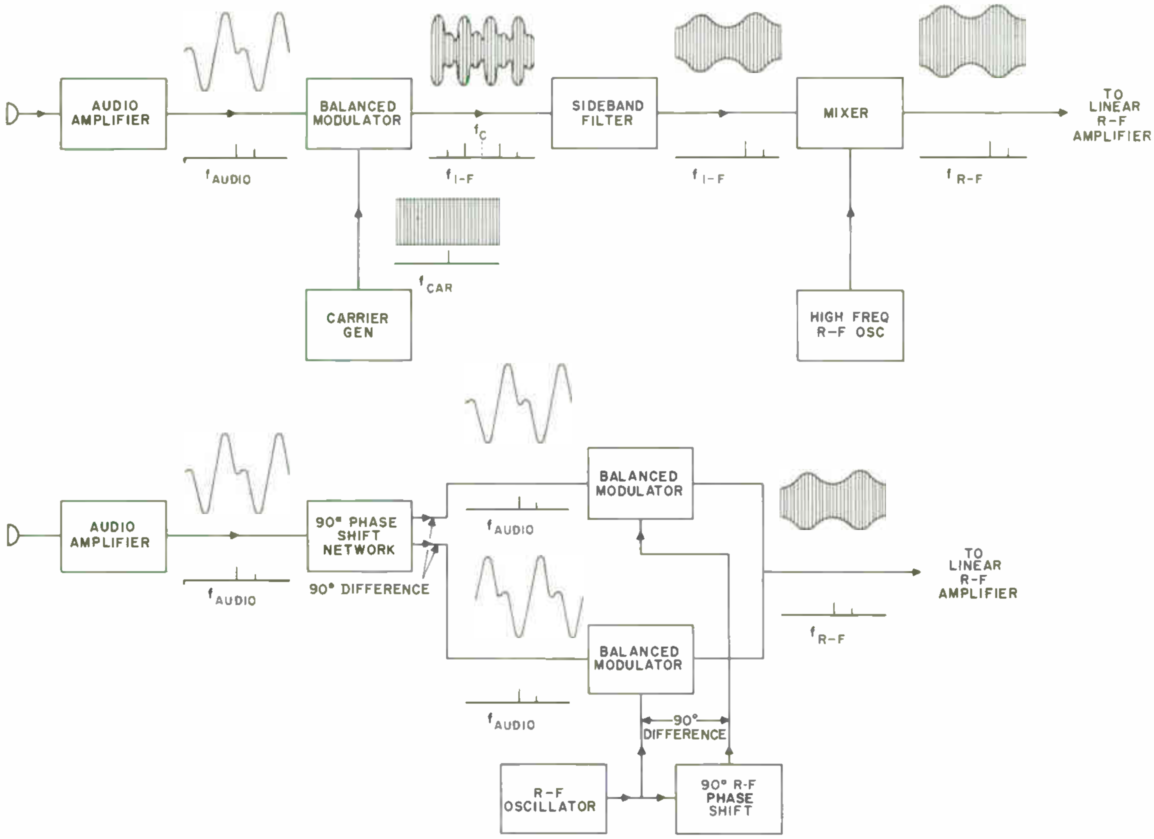


Figure 2-2. Basic Single-Sideband Generator, Block Diagram

shown in figure 2-2B, is centered about two separate simultaneous modulation processes and the combination of the modulation products. The audio signal is split into two components that are identical except for a phase difference of 90 degrees. The output of the r-f oscillator (which may be at the operating frequency if desired) is also split into two separate components having a 90-degree phase difference. One r-f and one audio component are combined in each of two separate balanced modulators. The carrier is suppressed in the modulators, and the relative phases of the sidebands are such that one sideband is balanced out while the other sideband is accentuated in the combined output. If the output from the balanced modulator is of sufficient amplitude, such a single-sideband exciter can work directly into the antenna, or the power level can be increased in a following linear amplifier.

2. THE SINGLE-SIDEBAND GENERATOR

The sideband generator processes the input audio signal, generates the r-f sideband in a modulator, selects the desired sideband while suppressing the unwanted sideband, and suppresses the carrier. The circuits which perform these functions are shown in the single-sideband generator portion of figure 2-1. The audio input wave must be amplified, amplitude limited, and shaped before being applied to the modulator circuits. Sideband generation is accomplished by using this audio input signal to vary the amplitude of a carrier wave in a modulator. The desired sideband is selected from the modulator output using frequency discrimination or phase discrimination. The carrier wave is suppressed by using balanced modulators or rejection filters.

a. INPUT SIGNAL PROCESSING

Processing of the audio input signal is an important part of single-sideband generating. If the input signal is a tone, or group of tones, of constant amplitude, such as the signal from a data gathering device, only a limited degree of processing will be required. However, if the input audio signal is a voice signal, rather elaborate input processing circuits must be designed to obtain optimum results.

The amount of amplification required depends upon the output capability of the source of the audio signal and the input signal requirements of the modulator. Modulators require an audio signal in the range of .1 to 1 volt at impedances of 200 ohms for diode modulators or several hundred-thousand ohms for vacuum tube modulators. The output of a microphone may be from 100 to 1000 times less than the .1 to 1 volt range. Telephone line levels will also be considerably less than the required level. To obtain efficient utilization of the transmitter power amplifier, the applied driving signal should be as close to maximum without exceeding the overload level. To avoid driving the power amplifier into overload, it is necessary to adjust gain to the point where maximum output is obtained with the maximum input signal.

When the input signal is made up of extreme variations, such as a peak level to average level of 4:1, the average tránsmitted power level will only be 1/4 the maximum output the transmitter is capable of furnishing. This analogy is illustrated in figure 2-3. An effort must be made to compress the dynamic range of the human voice to make it more compatible with the electrical characteristics of a communications system. The two methods most commonly used to reduce these amplitude variations are compression and clipping circuits.

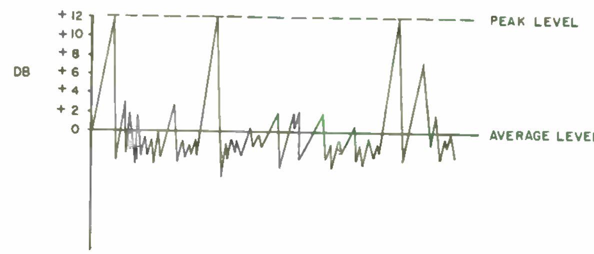


Figure 2-3. Peak-to-Average Level Variations of Speech

(1) COMPRESSOR CIRCUIT

A compressor is an automatic variable gain amplifier whose output bears some consistent relation to its input; for example, a one db rise in output for a two db rise in input. This circuit has very low steady state distortion. Common compressors use some type of feedback loop that samples the output of the amplifier and regulates the gain of the stage The time constants of this type circuit are necessarily slow to prevent oscillation, motorboating, and distortion. The attack time, the time necessary to reach steady state condition after a sudden rise in input level, will be several milliseconds. The release time, the

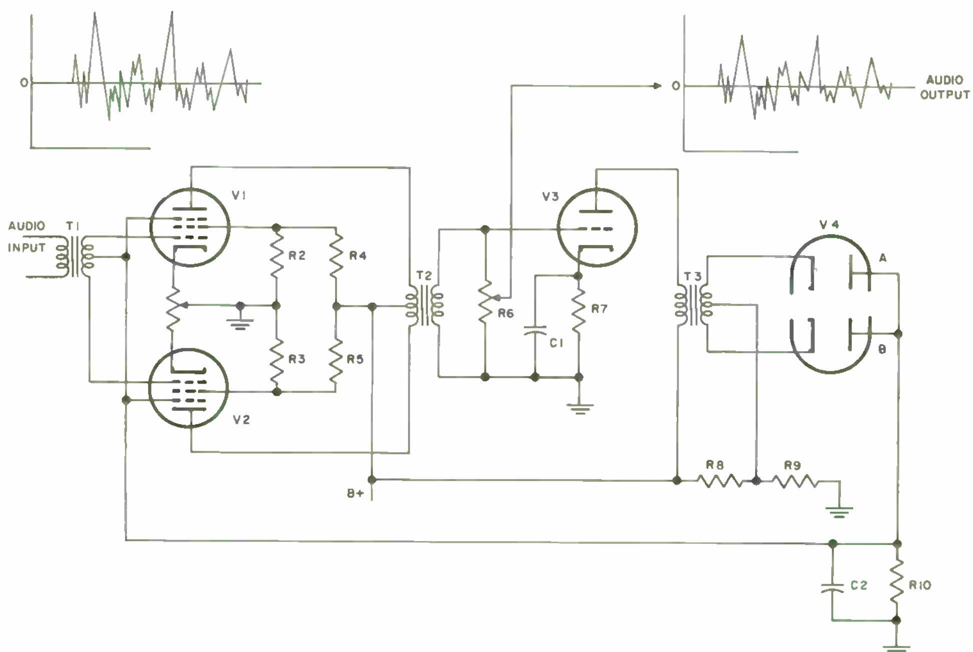


Figure 2-4. Compressor Circuit

time necessary to reach a steady state condition after a sudden drop in input level, will be several seconds. Compression of about 10 db is usually considered as an acceptable maximum value.

Operation of the compressor circuit, illustrated in figure 2-4, is such that the d-c bias voltage applied to the control and suppressor grids of the push-pull stage is in direct proportion to the amplitude of the signal passing through the circuit. If a large amplitude signal is impressed on the control grids of V1 and V2, such as a large amplitude low frequency, the signal is amplified and appears across transformer T3. As the audio signal swings positive at the top of the secondary of the transformer, tube V4B conducts, since the bottom of the secondary is negative with respect to ground, and the resultant current flow creates a bias voltage drop across resistor R10. The negative voltage on the control and suppressor grids of V1 and V2 reduces the gain of the tubes to limit the excursion of the audio signal. Conversely, as the audio signal swings negative at the top of the secondary of transformer T3, tube V4A conducts. Since the plates of the rectifiers are in parallel, the bias voltage is produced on both positive and negative going portions of the audio signal.

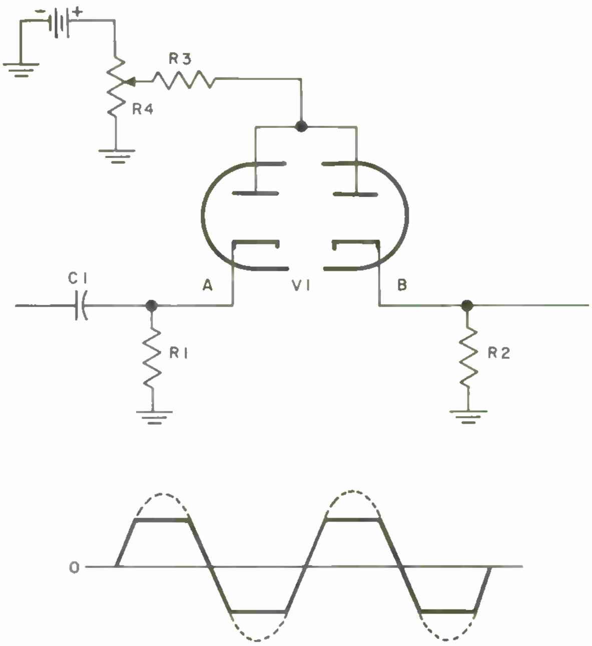


Figure 2-5. Clipper Circuit

(2) CLIPPER CIRCUIT

The clipper circuit, illustrated in figure 2-5, prevents the amplitude of a signal from exceeding a preset level. Its time constants are practically instantaneous, and it functions on each cycle of a wave. Distortion is very high, which results in loss of individuality in speaking and broadening of the spectrum occupied by the speech. Low-pass filters are usually used in conjunction with clippers to limit the spectrum and reduce distortion. The advantages of clipping is simplicity of circuit design and its ability to prevent overmodulation. The ability to prevent overmodulation results from its extremely fast attack on a wave after it exceeds the threshold. A well-designed clipper has no overshoot and an extremely fast release. A weak signal following one cycle after a wave that is heavily clipped will not be limited. This means that a weak consonant that follows a loud vowel in human speech will be given full amplification, although the preceding vowel was severely clipped. This amplifying of weak sounds in relation to soft sounds is referred to as consonant amplification.

The clipper circuit, illustrated in figure 2-5, serves as an instantaneous voltage amplitude limiter at a predetermined point on the positive and negative going portions of the audio signal. As the cathode of V1A swings positive, the tube will conduct until the potential on the cathode reaches the potential of the plate. The current flow through resistor R3 causes a voltage drop across R3 which is alternately reenforcing and bucking the plate voltage in exact response to the applied audio signal. This action causes the current through V1B to vary as the plate voltage varies and the signal in the output, across resistor R2, will be the same as the signal at the input, across resistor R1. When tube V1A is cut off, due to the cathode becoming more positive than the plate, there is no change of the plate voltage applied to V1B, and the current through the tube is held at a constant point. When the signal starts negative, the current variations through V1B will follow the current variations through V1A until the current through V1A becomes sufficiently great to cause the negative voltage drop across resistor R3 to equal the applied d-c plate potential. At this point the plate of V1B is no longer positive with respect to the cathode, and V1B ceases to conduct. The net result of this action is the clipping (or limiting) of the positive and negative peaks of an audio signal at a value predetermined by the setting of potentiometer R4.

(3) FREQUENCY RESPONSE SHAPING

The energy contained in a voice signal is confined principally to frequencies below 1000 cycles per second. Most of this energy is used to produce the vowel sounds which contribute little to intelligibility. The energy used to produce the consonant sounds is largely high frequency in content and is very important in intelligibility. An improvement in intelligibility will result if the frequency response of the audio input signal circuits is modified to amplify the high frequencies more than the low frequencies.

b. MODULATORS FOR SINGLE SIDEBAND

The r-f sideband is obtained by combining the audio signal obtained from the processing circuits and an r-f carrier wave in an amplitude modulator. There are many types of modulators, but they can be grouped into two main functional divisions: (1) those in which the modulation is dependent on the polarity of the modulating signal, and (2) those where the modulation is dependent on the instantaneous waveform of the modulating signal. For practical reasons, it is more convenient to group modulator circuits in the following three categories, based on the circuit components: (1) rectifier modulators (2) multielectrode vacuum tube modulators, and (3) nonlinear reactance modulators. Each group has its advantages and disadvantages, and these control the extent of their use. Because one of the characteristics of a modulator is frequency changing or frequency translating, this type of modulator is used in the frequency changing portion of a single-sideband exciter.

(1) RECTIFIER MODULATORS

Rectifier modulators have several advantages which make them particularly useful for single-sideband generation. Their great advantage is high stability in comparison with vacuum tube modulators. They require no heating elements, and therefore no power is required and no heat has to be dissipated. They can be made quite compact, have long life, and require little maintenance. Rectifier modulators may be one of three general types, ring, series, or shunt. These type names refer to the manner in which the diodes are connected in the circuit. In all circuits, the rectifiers are made to work like switches by using a large r-f switching signal which greatly exceeds the audio signal level. These modulators are almost invariably connected as balanced modulators so that, as nearly as possible, there is no output of the r-f switching voltage in the modulator output terminals

The basic circuits of the ring, shunt, and series modulators are shown in figures 2-6A, 2-7A, and 2-8A. It must be assumed that the rectifiers are capable of switching at zero voltage from an infinite back resistance to a zero forward resistance and back again. The basic signal circuits may then be represented by the equivalents shown in figures 2-6B, 2-7B, and 2-8B. These equivalent signal circuits are shown for any half-cycle of the carrier voltage, with switches shown in place of the rectifiers. Practical rectifiers are not ideal, but will have a finite forward and backward resistance. If it is assumed that the carrier frequency is several times that of the input signal, the resulting output waveforms are as shown in figures 2-6C, 2-7C, and 2-8C.

The output of these modulators consists of a series of pulses whose polarity and repetition frequency are determined by the switching or carrier voltage, and whose amplitude is controlled by the input audio signal. A spectrum analysis of these output signals reveals the presence of an upper sideband and a lower sideband displaced about the switching or carrier frequency. A similar set of sidebands is placed about the second harmonic of the carrier frequency and some other undesired products higher in frequency.

The ring modulator has the highest efficiency, being capable of twice as much output voltage as the shunt or series modulator. Where carrier balance is important, a split ring modulator may be used in which it is possible to balance independently the two sets of diodes. The shunt modulator has the unique ability of being able to handle input and output terminations of the unbalanced, one-side-grounded type.

Rectifier balanced modulators are capable of a high performance; however, if they are to retain this performance for long periods of time, they must be carefully made of good quality, accurately matched components. Initial carrier balance exceeding 40 db may be readily obtained, but it is difficult to retain this degree of carrier suppression if the environmental conditions are severe. The level of third order intermodulation products can be held to 50 db below the desired sideband output signal.

(2) MULTIELECTRODE VACUUM TUBE MODULATORS

Multielectrode vacuum tube modulators are flexible and used in a wide variety of applications in addition to generating sidebands. They are capable of giving conversion gain rather than loss as the case in rectifier modulators. However, they are quite unstable as to gain and impedance which makes them undesirable in balanced modulators. Since they employ vacuum tubes they require power, dissipate heat, and have relatively short life compared with rectifier modulators. Vacuum tube modulators, employing modulating functions dependent on the instantaneous amplitude of the modulating signals, are basically one of two types: a product modulator, or a square law modulator.

In a product modulator the output signal is proportional to the two input signals. In single-sideband application the input signals would be the carrier signal and the modulating signal. An example of such a product modulator is a double grid vacuum tube. The carrier voltage is applied to one grid, and the modulating signal applied to the other. Modulation takes place due to the combined action of the grids on the plate current. It is important to realize that nonlinearity is not necessary, and modulation will take place even if each grid has a linear mutual characteristic.

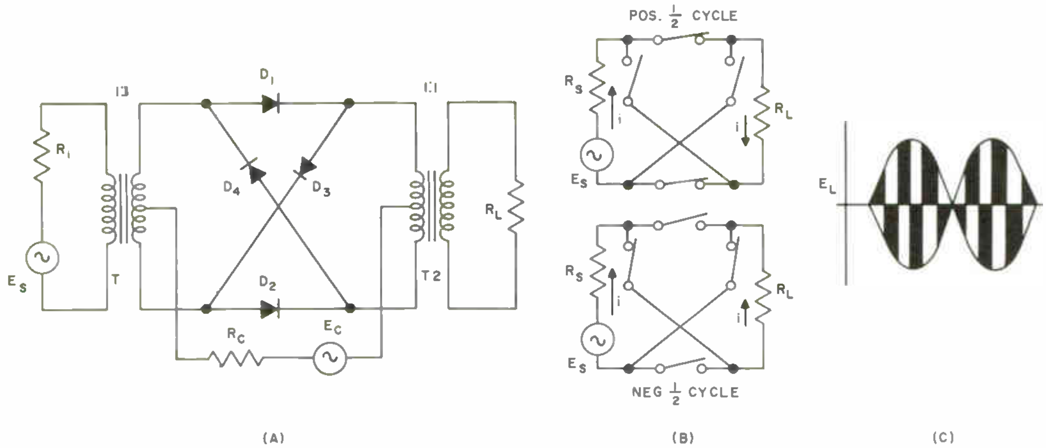


Figure 2-6. Basic Ring Modulator Circuits

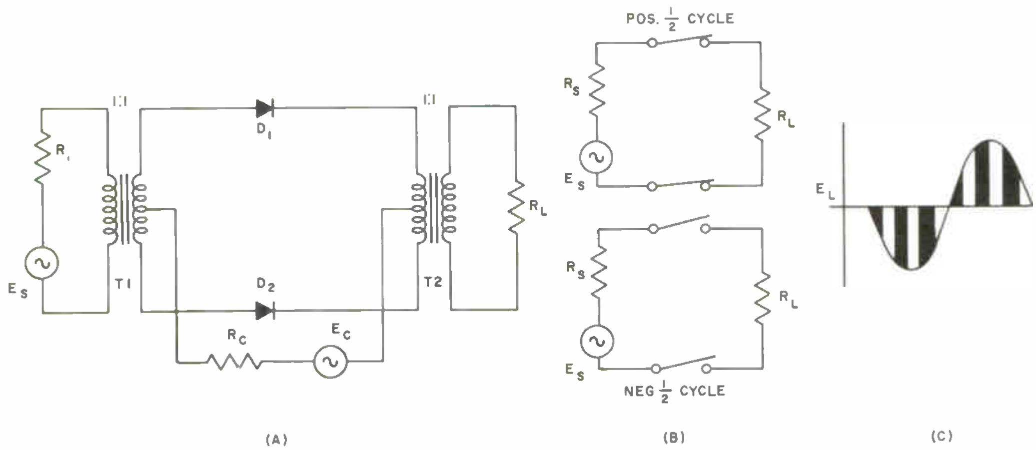


Figure 2-7. Basic Shunt Modulator Circuits

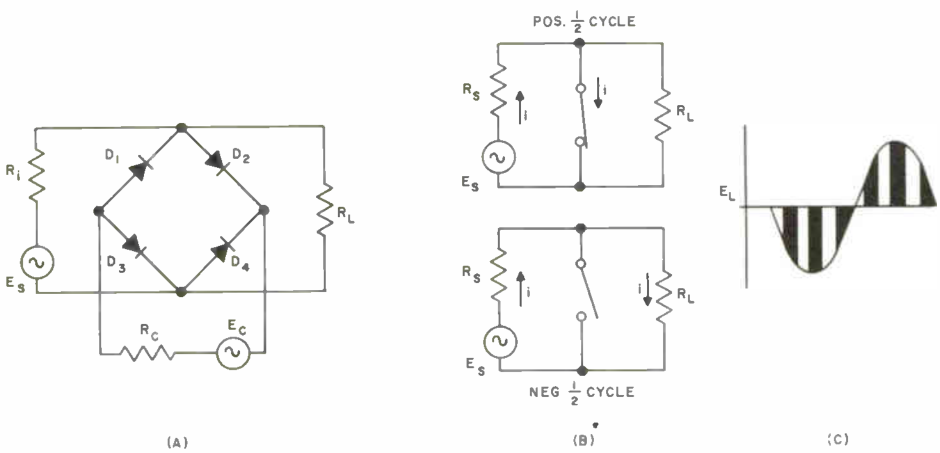


Figure 2-8. Basic Series Modulator Circuits

In contrast to the product modulator is a square law modulator in which modulation takes place directly because of a nonlinearity. An example of a square law modulator is a triode vacuum tube in which the shape of the plate current versus grid voltage curve has at least second order curvature or square law. This characteristic is possessed by all vacuum tubes and is the cause of distortion in amplifiers. If the curvature is purely square law, it can be shown that the output signal will contain only the desired sum or difference frequency and no other products except the second harmonics of the input signals. Product modulators and square law modulators are particularly useful in frequency changers because they generate a minimum of unwanted products.

Vacuum tube modulators in which the modulating function is dependent on the polarity of the modulating signal are large signal devices that have high efficiency but also generate considerable amounts of spurious signals. An example of such a modulator is a plate modulated triode operated class C. The modulating signal is used to vary the plate voltage applied to the class C amplifier. The resulting output is a series of pulses recurring at the carrier frequency rate and with amplitude proportional to the modulating signal. The tuned output circuit is necessary to suppress the harmonics of the signal. The double grid vacuum tube can also be used as a switching type modulator by increasing the amplitude of the signal applied to one of the grids. This signal can be large enough to drive the plate current of the tube to cutoff in one direction and saturation in the other, resulting in an output signal somewhat similar in waveform to that of a rectifier modulator.

Modulators using nonlinear reactances, instead of rectifiers or vacuum tubes, have not seen much use in high-frequency equipments due to the lack of materials usable at high frequencies. With suitable components such as titanate capacitors and ferrite-core inductors now available, it is probable that such modulators will be used more frequently in the future.

c. SUPPRESSING THE CARRIER

The carrier frequency can be suppressed or nearly eliminated by the use of a balanced modulator. The basic principle in any balanced modulator is to introduce the carrier in such a way that current at the carrier frequency in the output circuit cancels out.

(1) VACUUM TUBE BALANCED MODULATORS

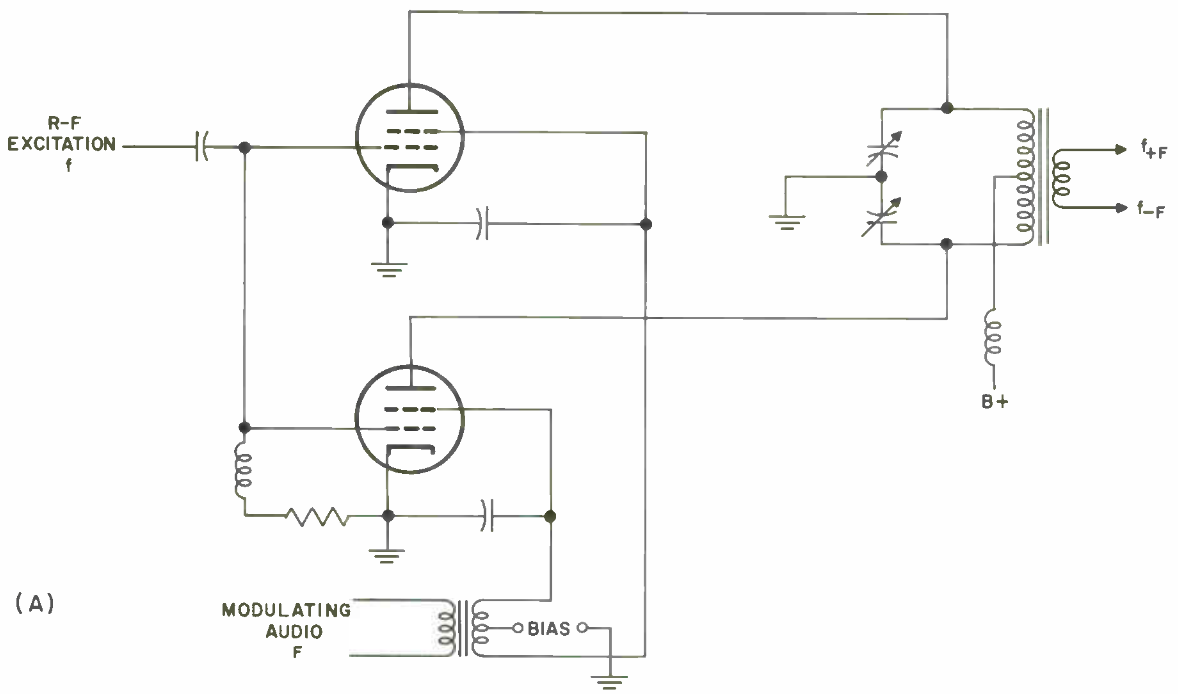
The requirements stated above are satisfied by introducing the audio in push-pull and the r-f drive in parallel, and connecting the output of the tubes in push-pull, as shown in figure 2-9A. Balanced modulators can also be connected with the r-f drive and audio inputs in push-pull and the output in parallel (figure 2-9B) with equal effectiveness. The choice of a balanced modulator circuit is generally determined by constructional considerations and the method of modulation preferred by the builder. Screen-grid modulation is shown in the examples in figure 2-9, but control grid or plate modulation could be used with the same result. In balanced modulator vacuum tube circuits, there will be no output with no audio signal because the circuits are balanced. The signal from one tube is balanced or canceled in the output circuit by the signal from the other tube. The circuits are thus balanced for any value of parallel audio signal. When push-pull audio is applied, the modulating voltages are of opposite polarity, and one tube will conduct more than the other. Since any modulation process is the same as mixing, sum and difference frequencies (sidebands) will be generated. The modulator is not balanced for the sidebands, and they will appear in the output. The amount of carrier suppression obtained is dependent upon the matching of the two tubes and their associated circuits. Normally, two tubes of the same characteristics can be adjusted to give at least 30 db of carrier suppression without further filtering. Balance is difficult to maintain since tube characteristics change with age and supply voltage variations. Since in suppressed carrier single-sideband transmission it is desirable to suppress the carrier at least 40 db, the selective filter following the balanced modulator is used for further carrier suppression.

(2) DIODE BALANCED MODULATORS

The operation of diode balanced modulators was discussed in paragraph 2.b.(1) and figures 2-6, 2-7, and 2-8. The equivalent circuits illustrated in figures 2-6B, 2-7B, and 2-8B do not present the carrier balancing action of the modulators. This action may be analyzed from the basic circuits shown in figures 2-6A, 2-7A, and 2-8A. Using the ring modulator as an example, the carrier currents may be as shown in figures 2-10A and 2-10B. Assume that the carrier generator voltage is such that D1 and D2 conduct. The current flow will be through R_c, T1, D1 and D2, T2, and back to the generator. The current through the two windings of the output transformer T2 are out of phase and will cancel. On the next carrier half-cycle, D3 and D4 conduct, and the phases of all currents are changed by 180 degrees. The output currents are again out of phase. Therefore, no carrier voltage appears across the output load, R_L. Carrier currents may be similarly traced in the shunt and series circuits to show the balancing action of the carrier currents.

d. SIDEBAND SELECTION

It is a property of all modulators that the output consists of a pair of sidebands symmetrically disposed on either side of the carrier frequency. Since the objective is to transmit only a single sideband, a



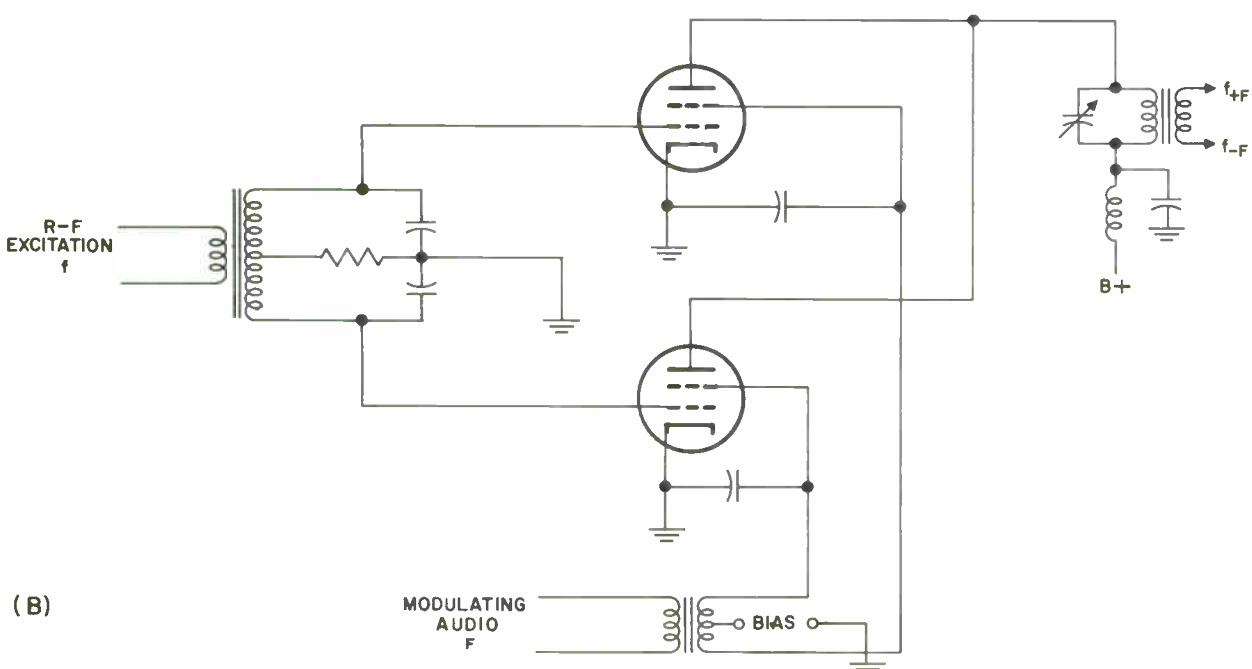
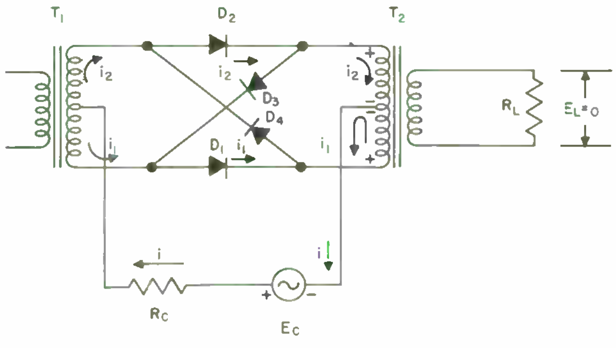


Figure 2-9. Vacuum Tube Balanced Modulators



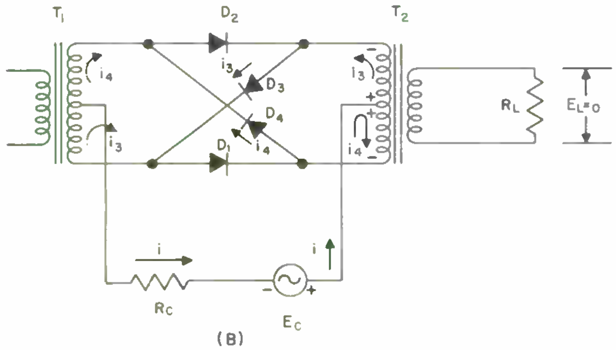


Figure 2-10. Ring Modulator Carrier Current Paths

means must be found to select the desired sideband and suppress the undesired sideband. This may be accomplished through the use of one of two techniques: the technique of frequency discrimination (filtering) or the technique of phase discrimination (phase shift).

(1) FILTER SYSTEM OF SIDEBAND SELECTION

The frequency discrimination method uses a frequency selective filter to select the desired sideband. This is possible because the modulating wave is usually confined to a restricted band of frequencies, and this frequency band is separated from the carrier by an appreciable amount. The rapid increase of attenuation required of a sideband selecting filter in order that it may adequately suppress the unwanted sideband is a decisive factor in filter design. Components having a high rate of change of impedance with frequency, or high Q, must be used. The requirement is a certain amount of attenuation in a given number of cycles. For a given frequency of filter operation and a certain degree of sideband suppression, the quality or Q factor of the components making up the sideband filter is determined. This means that for low-frequency sideband selection a lower Q element may be used, conversely for high-frequency sideband selection high Q elements are required. Inductors and capacitors have low Q factors and can be successfully used for sideband filters only at relatively low frequencies, up to about 50 kilocycles per second. Small metal plates and quartz crystal plates on the other hand have extremely high Q factors and can be used to build sideband filters capable of operating at higher frequencies. Mechanical filters or metal plate filters have been built to operate up to 600 kilocycles, and crystal filters have been made to work at frequencies as high as 5 megacycles.

Removing the unwanted sideband through the use of a selective filter has the advantage of simplicity and good stability. The suppression unwanted sideband is determined by the attenuation of the sideband selecting filter. The stability of sideband suppression is determined by the stability of the elements used in constructing the sideband filter. This stability can be quite high because it is possible to use materials that have very low temperature coefficient of expansion.

(2) PHASE SHIFT METHOD OF SIDEBAND SELECTION

In the phase shift method, two balanced modulators are used, and the exciting signals to these modulators are arranged to have phase relationships such that when the outputs of these two modulators are combined, the desired sideband components are reinforced and the unwanted components are canceled out. Into modulator number 1 the modulating signal and the carrier signal are fed directly. Into modulator number 2 these signals are fed after first being passed through networks which shift the relative phase of these signals 90°. In other words, the modulating signal fed into modulator number 2 is shifted 90° with respect to the phase of the audio signal fed into modulator number 1. Similarly the phase of the carrier voltage fed into modulator number 2 is shifted 90° relative to the phase of the carrier voltage fed to modulator number 1. If these phase relationships are maintained over the desired modulating signal frequency range, the action is to suppress completely one set of sidebands and to reinforce the opposite set in the output of the combining circuit. If balanced modulators are used, the carrier signal will not appear in the output. Practically speaking, it is very difficult to design a phase shifting network that will perform according to the above restriction for the modulating signal phase shift network. However, if a separate phase shifting network is inserted in each modulating signal input circuit, it is possible to maintain a phase difference of 90° between the two network outputs over the required signal frequency range. The action of this circuit with these networks is identical with that in which a 90° phase shift is used in one branch alone.

The phase shift method of single-sideband generation does not require a rapid change of discrimination in a narrow frequency interval; therefore, it can be used to generate a single-sideband signal which can have extremely low-frequency components.

Since no selective filter is required, it is possible to generate the single sideband at the operating frequency with no frequency conversion being required. The degree of suppression of the undesired sideband is dependent on the accuracy with which the undesired sideband components are canceled. To obtain complete cancellation, it is necessary to maintain accurately the phase shifts and amplitudes of the signals applied to the combining network. These requirements place a very stiff specification on the phase shift and amplitude control properties of the circuit. The circuit is also somewhat more complex since two modulators are required.

3. TRANSLATION TO THE OPERATING FREQUENCY

The single-sideband signal is translated to the operating frequency by the use of one or more frequency changers. These frequency changers perform their function through the modulation process which is identical with that used to generate the sideband signal. The sideband signal is used to modulate a high-frequency carrier whose frequency is such that the upper or lower sideband is on the desired operating frequency. As a result of this modulation process, the sideband signal will be shifted to a new frequency that is either the sum of the carrier and sideband frequencies or the difference between the carrier and sideband frequencies. It is important to realize that if the lower sideband of the translation modulation process is selected, an inversion of the sideband signal occurs. That is to say, an upper sideband signal will be converted into a lower sideband signal. Another important consideration is the frequency accuracy and stability of the carrier used in the modulation process since any error in the carrier frequency is passed on to sideband signal in exact proportion. The translation system consists of two major components: the modulator (commonly called a mixer) and the carrier (commonly called oscillator signal).

All modulators previously described can be used for frequency changing. Vacuum tube modulators are used almost exclusively in this service, because these circuits have considerable gain and generate a minimum of spurious products. It is these spurious products which exert the most influence on the design of the frequency translation system.

a. SPURIOUS MIXER PRODUCTS

In order to show how undesired frequencies may be generated in a mixer stage, consider the case where signal and oscillator voltages are applied to the same grid of a mixer tube. In order that the desired sum or difference frequency be generated, it is necessary that the plate current versus grid voltage characteristic have some nonlinearity or curvature. The components of the plate current will be the d-c, signal, oscillator, signal second harmonic, oscillator second harmonic, and the sum and differences of the signal and oscillator. It is necessary to eliminate all the products except the desired sum or difference product by filtering. To obtain the desired sum or difference product, we would desire a tube in which the characteristic curve had only second order curvature. Unfortunately, all practical tubes have characteristic curves having higher order curvature. This higher order curvature contributes additional unwanted frequency components into the output current. Sometimes the frequency of these unwanted components is far removed from the desired output frequency, and they are easily filtered out, but often these frequencies are very nearly equal in frequency to the desired signal frequency, and they will fall within the passband of the filter used in the mixer output circuit. The amplitude or strength of these undesired mixer products varies from tube type to tube type and tube to tube and with a given tube will change when the operating point is varied. Consequently, it is not surprising that tube designers have not been particularly successful in designing tubes having the desired second order curvature to the exclusion of any higher order curvature. The circuit designer, therefore, must select his mixer tubes by means of a series of experiments in which the amplitudes of these undesired mixer products are measured. The result of such an experimental determination of mixer product amplitudes is shown in table 2-1.

It can be seen that there are several undesired products that are greater in amplitude than the desired signal and a considerable number that are weaker than the desired signal. Furthermore, it can be seen that as the order of the mixer product involved increases, its amplitude decreases. The presence of undesired mixer products in the output of the frequency translation system may be minimized through intelligent selection of the signal and oscillator frequencies. The problem of frequency selection is relatively simple where the operating frequencies are fixed. Where the operating frequency must be varied, the problem becomes more complex; and if continuous operation over wide frequency ranges is required, the problem is exceedingly complicated. In an attempt to simplify the problem, circuit designers have resorted to charts in which the frequency of the spurious mixer products is plotted with respect to the signal and oscillator frequencies. Such a chart is shown in table 2-2.

The following example illustrates the spurious product problem. In this example, it is desired to produce a single-sideband transmitter capable of operating on the amateur 20, 40, and 80 meter bands. These bands are 3.5 to 4.0 megacycles, 7.0 to 7.3 megacycles, 14.0 to 14.35 megacycles. Assume that the single-sideband signal has been generated at 250 kilocycles carrier frequency. The lowest frequency band 3.5 to 4.0 megacycles can be covered by mixing

TABLE 2-1. CALCULATED FREQUENCY PRODUCTS CONTAINED IN THE PLATE CURRENT OF A 12AU7 TRIODE MIXER

CALCULATED FREQUENCY PRODUCTS CONTAINED IN THE PLATE CURRENT OF A 12AU7 TRIODE MIXER

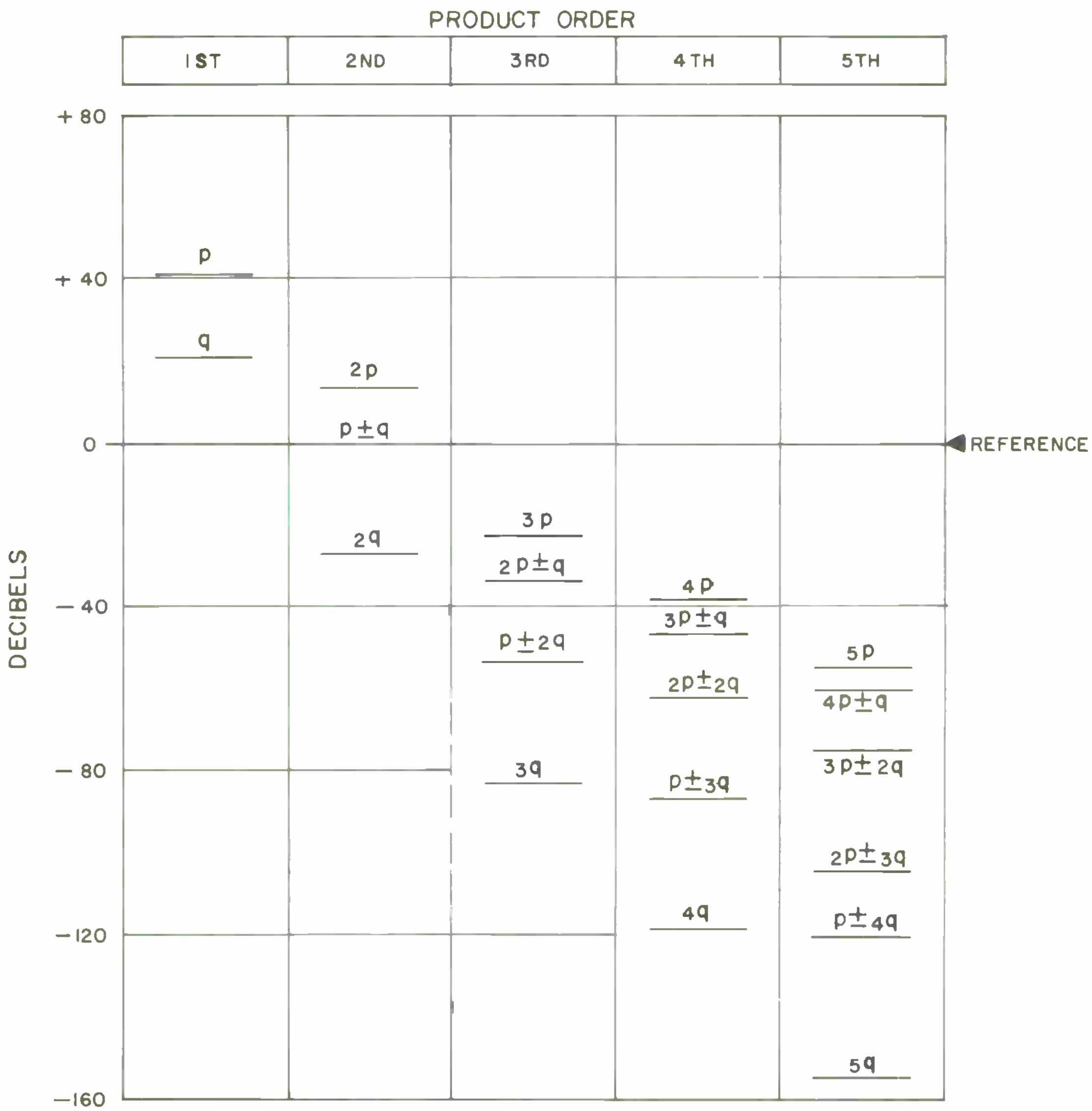
$e_{osc} = P\cos pt = 2V\,rms$ $e_{sig} = Q\cos qt = .2V\,rms$

$E_b = 250V$ $E_k = 10V$ $E_{bb} = 415V$ $R_L = 10K$

TABLE DERIVED FROM POWER SERIES EXPANSION

WHERE $e_{in} = P\cos pt + Q\cos qt$

ZERO DB REFERENCE IS MAGNITUDE OF $(p \pm q)$

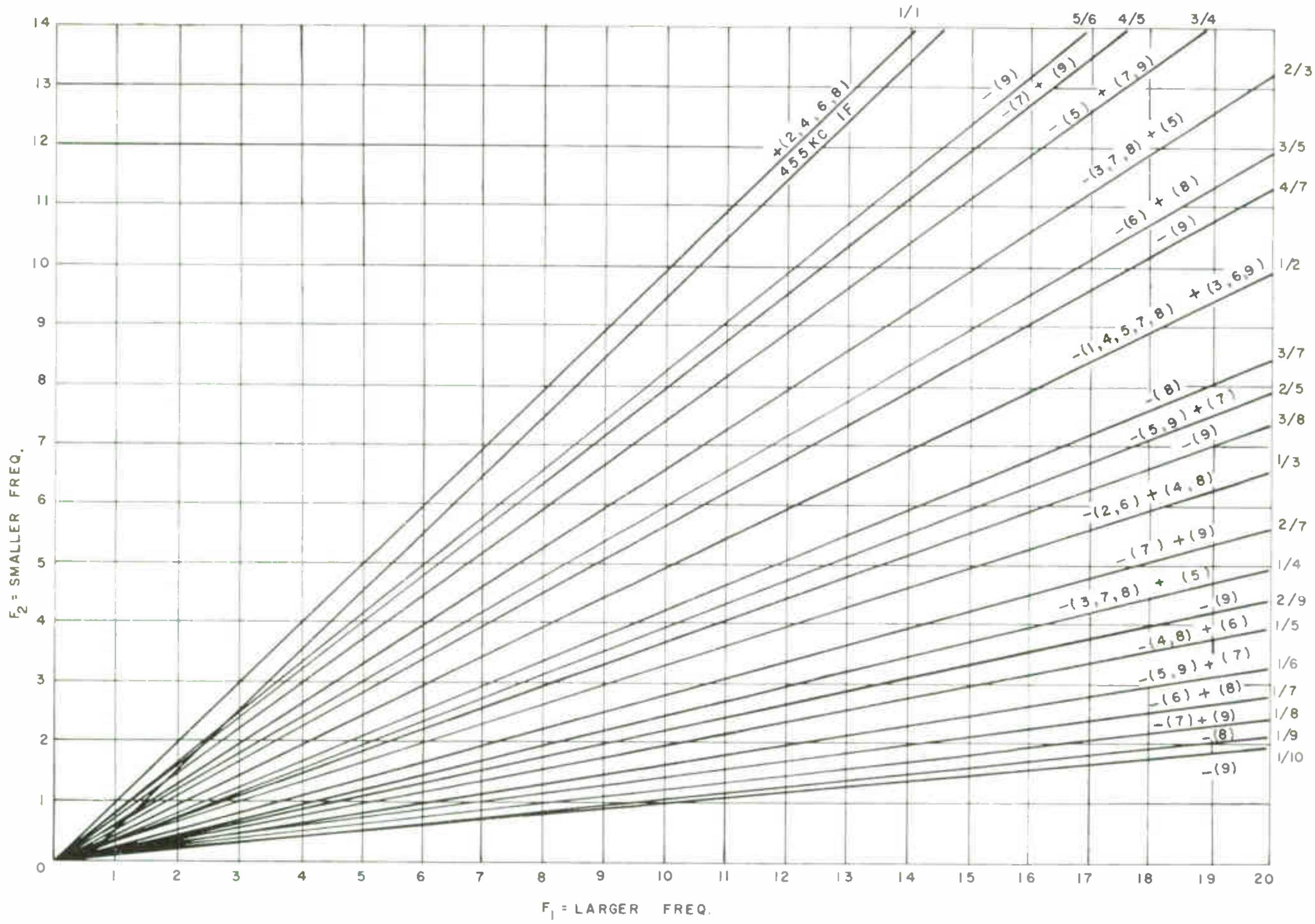


COEFFICIENTS OF POWER SERIES FROM REPORT BY

DR. V. W. BOLIE DUAL TRIODE MIXERS 7/23/53

$a_1 = 3.47 \times 10^{-4}$, $a_2 = 1.47 \times 10^{-5}$, $a_3 = 2.2 \times 10^{-7}$, $a_4 = 3.7 \times 10^{-8}$, $a_5 = 5.7 \times 10^{-9}$

TABLE 2-2. SPURIOUS RESPONSE CHART



$F_2 \sim F_1$

ORDER	1	2	3	4	5	6	7	8	9
1/1		•20 •02		• 3 •51		•24 •42		•35 •53	
1/2	10		•12 •30	31	32	•33 •51	52	53	•54 •72
1/3		20		•22 •40		42 51		•53 •71	
1/4			30		•32 •50		52	71	
1/5				40		•42 •60		62	
1/6					50		•52 •70		72
1/7						60		•62 •80	
1/8							70		•72 •90
1/9								80	
1/10									90
2/3			21		•23 •41		43	53	

$F_2 \sim F_1$

ORDER	1	2	3	4	5	6	7	8	9
2/5				41			•43 •61		63
2/7							61		•63 •81
2/9									81
3/4					32		•34 •52		54
3/5						42		•44 •62	
3/7								62	
3/8									72
4/5							43		•45 •63
4/7									63
5/6									54

• INDICATES SUM MIXING
OTHER — DIFF MIXING

this single-sideband signal with the output of a variable frequency oscillator tunable from 3.25 megacycles to 3.75 megacycles. Due to the large difference between the signal and oscillator frequencies, there are no difficulties with undesired mixer products. However, the oscillator signal is only 250 kilocycles removed from the output frequency and must be filtered by means of a band-pass filter or else balanced out through the use of a balanced modulator. As it is quite difficult to build a modulator which can retain balance over a wide frequency range, it is necessary to resort to a combination of both methods to obtain suppression of this spurious frequency of at least 60 db. As the operating frequency increases, it becomes difficult to suppress this product, and some other method must be found. This is because the selectivity required in the tuned circuits is so high as to be impracticable. A solution to the problem is to use a second conversion following the first. In this mixer, the 3.5 to 4.0 megacycle output of the first conversion is mixed with the output of a crystal oscillator at 3.3 megacycles. This crystal frequency is chosen rather than 3.5 megacycles because with a 7.0 megacycle output frequency, there is a crossover of the second harmonic of the 3.5 megacycle signal with the desired output. A closer look at the frequencies involved, however, reveal that even with this crystal frequency, the second harmonic of the crystal at 6.6 megacycle is only 400 kilocycles removed from the low-frequency desired output, and the 7.4 megacycle second harmonic of the input signal is only 400 kilocycles on the other side of the desired output frequency. Selectivity of a very high order would be required to reduce these spurious signals satisfactorily. Some relief can be obtained by extending the range of the variable frequency oscillator used in the first mixer so that the output frequency from the first conversion runs from 3.0 to 4.0 megacycles. Now, a crystal oscillator

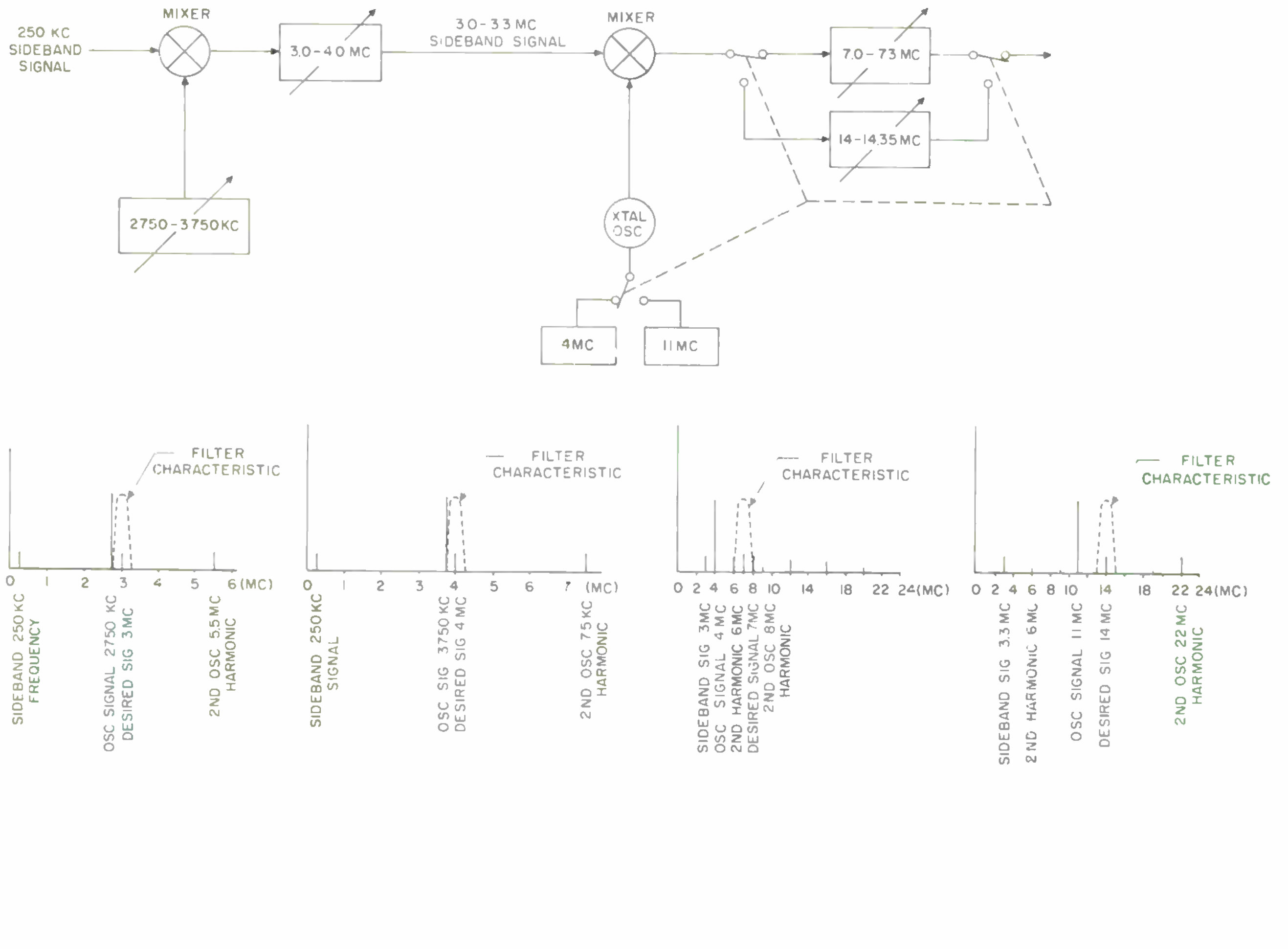


Figure 2-11. Selectivity Considerations in Frequency Translators

frequency at the second mixer of 4 megacycles can be used. With this frequency the range of the first converter system of 3.0 to 3.3 megacycles can be used to cover the 7.0 to 7.3 megacycle band. With this frequency scheme, the second harmonic of the crystal oscillator is at least 700 kilocycles removed from the desired operating range, and the second harmonic of the signal frequency ranges from 100 to 700 kilocycles below the desired output frequency range. These can be filtered adequately with relatively simple filters. According to the spurious frequency chart, a seventh order crossover occurs at the low end of the band when the third harmonic of 3 megacycles mixes with the fourth harmonic of 4 megacycles to yield a frequency equal to the output frequency. However, the level of a seventh order spurious signal is sufficiently low that it may be neglected, providing sufficient attention is paid to the selection of the mixer tube and its operating point. In considering frequencies to be used to cover the 14 to 14.35 megacycle band, one notes that if the difference product is selected, a clear region exists in the spurious chart between the 1/5 line and the 1/6 line. However, if such a scheme were adopted, the dial scale for this band would be reversed with respect to the two lower frequency bands, a distinctly unattractive feature. Fortunately, it is possible to use the sum product if a crystal frequency of 11 megacycles is used. The ninth order crossover occurring near the low end of the range is of no consequence since it is of negligible amplitude.

b. OSCILLATOR REQUIREMENTS

It must be realized that the frequency stability of the output signal is dependent on the frequency stability of the carrier frequency and the oscillator outputs used in the frequency changers. The total frequency error is the sum of the error in all three of these oscillators. This oscillator frequency error has two aspects. First, there is the accuracy of the calibration of the frequency involved. The second aspect of the frequency error is that of stability or a drift. If the equipment is to be operated and continuously monitored by skilled operators, it is possible to get by with rather large errors in both calibration and drift. In come cases, it is possible to use equipment in which the calibration error is relatively large, but the frequency drift is quite small. In this case, it may be possible to carry out effective communication with such an equipment, providing an operator is available to make the initial tuning adjustment. In some cases where it is desired to operate the equipment on many channels by remote control and with relatively unskilled operators, it is necessary to provide equipment with a high degree of performance both with respect to calibration and drift. Authorities tend to disagree as to the allowable frequency error which can be tolerated in a single-sideband communication system used for voice communication. As the error increases, the naturalness of the reproduced speech suffers first. If the error is such as to place the reinserted carrier on the high side of the original carrier frequency, the voice becomes lower pitched. If the error is such as to place the reinserted carrier on the low frequency side, the voice becomes higher pitched. As the error is increased, a point is reached where intelligibility is degraded. The frequency error at which this occurs is approximately 100 cycles per second. When the single-sideband transmitter is used to transmit narrow-band telegraph or teletype signals, it is sometimes necessary to maintain accuracy considerably higher than that required for voice transmission.

The stability that can be obtained from oscillators is an important factor in the design of a frequency translation system. Typical oscillator frequency errors are shown in tabular form in tables 2-3 and 2-4. In table 2-3, the frequency error is the long-term frequency error. Calibration, drift, and aging all contribute to the long-term frequency error. The errors listed in table 2-3 are typical for a term of several months. The short-term frequency errors are shown in table 2-4. The short-term frequency

TABLE 2-3. TYPICAL OSCILLATOR LONG-TERM FREQUENCY ERROR

OSCILLATOR TYPE	ERROR %	ERROR CPS 3 mc	10 mc	30 mc
Variable Frequency Oscillator	.05	1500	5000	15,000
Crystal Oscillator	.005	150	500	1500
Temperature Controlled Crystal Oscillator	.001	30	100	300
Precision Standard Oscillator	.0001	3	10	30

TABLE 2-4. TYPICAL OSCILLATOR SHORT-TERM FREQUENCY ERROR

OSCILLATOR TYPE	ERROR PPM	ERROR CPS		
		3 mc	10 mc	30 mc
Variable Frequency Oscillator	20	60	200	600
Crystal Oscillator and Temperature Controlled Crystal Oscillator	1	3	10	30
Precision Standard Oscillator	.01	.03	.1	.3

error is principally that of frequency drift, although in some cases aging is rapid enough to have some effect. From the data shown in these tables, it can be seen that single-sideband equipments using variable frequency oscillators would require manual operation and frequent attention. For an equipment to meet the stability and accuracy requirements for quick frequency selection by remote control by unskilled operators, it is necessary to use the form of oscillator known as a stabilized master oscillator, in which a variable frequency oscillator is stabilized by comparing its frequency with that of a frequency derived from a reference standard oscillator. Oscillators of this type are described in a later chapter.

4. AMPLIFICATION

In order that the single-sideband exciter have useful output, it is necessary to provide amplification of the sideband signal. Since the output power will be used to drive the power amplifiers of the system, the power output of the exciter is determined by system considerations. The driving power required is usually quite small since it is customary to use high gain tetrode tubes in most linear amplifiers. As a result, the power output of the single-sideband exciter may be limited to a fraction of a watt. This power level can be readily obtained through the use of receiving type tubes.

Pentode receiving tubes designed for use as r-f or i-f amplifiers in receivers may be used for low-level voltage amplifier stages. Low grid-to-plate capacitance is a necessary requirement for linear r-f amplifiers, since positive feedback through this path increases distortion. High mutual conductance is a useful characteristic because the required gain is then obtained with a minimum of stages. Receiving-type power pentodes may be used to obtain moderate power output, although most types suffer from having relatively large grid-to-plate capacitance. Tubes designed for video power amplifier use are best suited for use in linear r-f power amplifiers.

a. TUNED CIRCUITS

Tuned circuits used in a single-sideband linear amplifier perform a dual function. A tuned circuit provides a suitable load impedance for the amplifier stage, so that the amplifier may provide sufficient voltage amplification. Secondly, this tuned circuit acts as part of a selective filter which acts to suppress unwanted mixer products generated in the frequency translation system. To obtain sufficient selectivity, it is quite often necessary to use double-tuned and even triple-tuned circuits in order that the required selectivity be obtained.

b. LINEAR AMPLIFICATION

It is necessary to use linear amplifiers in a single-sideband transmitter in which low level modulation is used. The single-sideband system is an amplitude modulated system and once the modulation is performed, the amplitude relationships of the sideband components must be faithfully maintained. The principal distortion component encountered in tuned linear amplifiers is the third order intermodulation product. This product is so called because its generation depends on the third order curvature of the input-output amplifier characteristic. Unlike audio linear amplifiers which must handle a wide frequency range, tuned radio-frequency linear amplifiers seldom have difficulty with products generated due to second order curvature, such as sums and difference frequencies and harmonics. These frequencies usually fall far outside the tuned passband and are suppressed accordingly. On the other hand, the third order intermodulation products are always close to the desired frequency band, and many of the products actually fall within the desired passband. These intermodulation products are generated whenever there are two tones, or frequencies, within the amplifier passband whose frequencies are sufficiently close together that the second harmonic of one will mix with the other to yield a third frequency within the tuned amplifier passband. The amplitude of these spurious products can be controlled by limiting the input signal amplitude so that operation of the tube is always over a linear part of its input-output characteristic. It is readily possible to obtain sizable voltage and power amplification using receiving type tubes and still limit the amplitude of the third order intermodulation product to a level more than 50 db below the desired.

An important consideration in transmitters using linear power amplifiers is that the amplifier be driven with sufficient signal and yet not be overdriven to cause excessive intermodulation distortion. There are many factors which tend to cause the output of a single-sideband exciter to vary. The gain of the amplifier stages changes from one frequency channel to another due to the impedance of the tuned circuits, used as the load impedances in these stages, varying with frequency. Tube gain characteristics change from tube to tube and from time to time as the tubes age. Changes in temperature and other environmental factors can also cause changes in amplifier gain. An effective way to cope with these variations is to sample the driving voltage of the power output amplifier with a rectifier, and to use the resulting d-c to control the gain of one or more amplifier stages in the exciter. This control voltage may be used to control the gain of amplifier stages using remote cutoff characteristic tubes similar to those used in receiver r-f amplifiers on which automatic gain control is used.

5. SUMMARY

The single-sideband exciter consists of three major sections: a single-sideband generator, a frequency translator, and a voltage power amplifier. In the sideband generator the audio input signal is processed by the use of amplification, amplitude limiting, and frequency energy distribution. The processed signal is then converted into an r-f sideband in a modulator. The desired signal or sideband is selected and the unwanted sideband suppressed, using the technique of frequency discrimination or phase discrimination. The desired sideband is then translated to the desired range of operating frequency by means of a frequency translation system. The desired output level is obtained through the use of linear amplifiers.

CHAPTER 3
SINGLE-SIDEBAND RECEIVERS

1. SINGLE-SIDEBAND RECEIVER CONSIDERATIONS

The operation of a single-sideband receiver is, in a limited sense, the reverse of the process carried out in a single-sideband exciter. The received single-sideband radio-frequency signal is amplified, translated to a low i-f frequency, and converted into a useful audio-frequency signal. The reception of a high-frequency single-sideband signal is essentially the same as receiving a high-frequency AM. signal. Receivers are invariably of the superheterodyne type to provide high sensitivity and selectivity. The absence of a carrier in the received SSB signal accounts for the principal difference between single-sideband and AM. receivers. In order to recover the intelligence from the single-sideband signal, it is necessary first to restore the carrier. This local carrier must have the same relationship with the side-band components as the initial carrier used in the exciter modulator. To achieve this, it is a stringent requirement of single-sideband receivers that the oscillator which produces the reinserted carrier have extremely good frequency accuracy and stability. The total frequency error of the system must be less than 100 cycles per second, or the intelligibility of the received signal will be degraded.

The single-sideband receiver must be able to select a desired signal from among the many signals which populate the high-frequency band. Good selectivity becomes an essential requirement when signals of considerable variance in amplitude are spaced close together in the frequency spectrum. The sensitivity of the receiver must be sufficient to recover signals which are of very low amplitude, almost lost in the noise which is constantly present in the antenna. These requirements determine the design of the front end, or r-f section, of the receiver.

Double conversion superheterodyne circuits are used in present day receivers. The principal advantages of such circuits are the extra image rejection obtained and the decided improvement in frequency stability. This can be achieved by using a crystal oscillator in the high-frequency conversion and injecting the tunable oscillator at a lower frequency conversion where its error has less effect.

In a conventional receiver, the audio intelligence is recovered from the radio-frequency signal by means of an envelope detector. This detector may be a simple diode rectifier. This same diode detector, provided with a local carrier, can also be used to recover the audio signal from a single-sideband suppressed carrier signal; however, the amplitude of the local carrier must be quite high in order that the intermodulation distortion be kept low. Better performance, particularly with respect to distortion, may be obtained by using product demodulators to recover the audio signal.

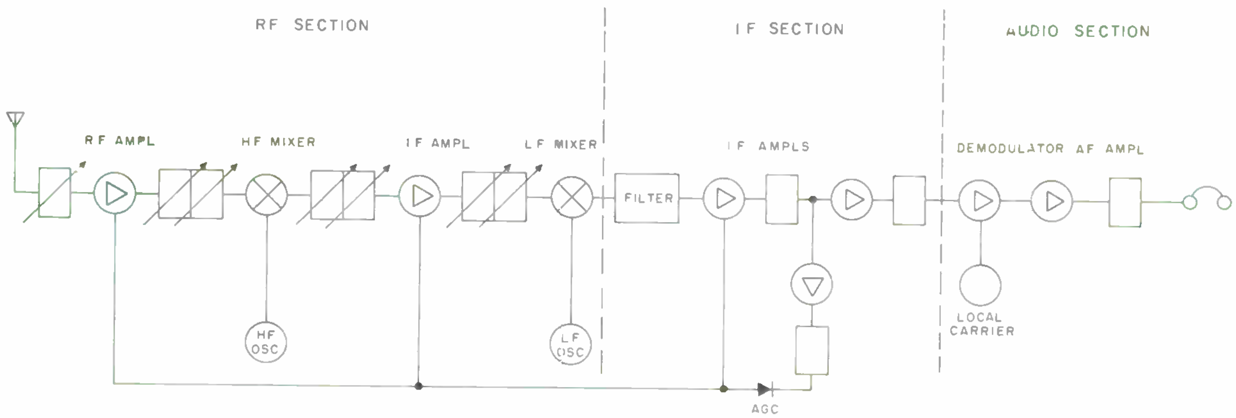


Figure 3-1. Typical Single-Sideband Receiver, Functional Diagram

The characteristics of the automatic gain control system of a single-sideband receiver must be somewhat different from those of a receiver designed for conventional AM. signals. The conventional avc system provides the automatic gain control by rectifying the carrier signal, since this carrier is relatively constant and does not vary in amplitude rapidly. This avc system can have a relatively long time constant. In a receiver for single-sideband suppressed carrier signals, the agc rectifier must be of a quick acting type because the signal amplitude is changing very rapidly and frequently disappears altogether.

Single-sideband receivers have three main sections: a radio-frequency section, an intermediate-frequency section, and an audio-frequency section. These sections of a typical single-sideband receiver are shown in figure 3-1. The principal requirement of the r-f section is to select the desired signal in the antenna and translate this signal to a lower r-f frequency (intermediate frequency) with a minimum of distortion and generation of spurious signals. The intermediate-frequency section provides selectivity and amplification. The audio section recovers the audio-frequency intelligence and provides the necessary audio-frequency amplification.

2. R-F SECTION

The r-f section consists of an r-f amplifier and one or more mixers which translate the signal to the intermediate frequency. The use of an r-f amplifier as the first stage of a receiver provides increased sensitivity and reduction in spurious responses. Increased sensitivity results from the lower inherent noise of amplifiers compared with mixers. Spurious signals are reduced because increased r-f filtering can be used without degrading the signal-to-noise ratio. The amplification provided by the r-f amplifier offsets the losses inherent in the passive filter circuits.

The sensitivity of a receiver is usually defined as the minimum signal with which a 10 db signal-plus-noise-to-noise ratio may be obtained. This definition has a practical basis because it recognizes the fact that ultimately the noise level is the limiting factor in readability, and that a signal 10 db stronger than the noise level is acceptacle for voice communication. Maximum receiver sensitivity is not determined by the gain of the receiver but by the magnitude of the receiver noise. The three sources of noise which contribute to the noise level of a receiver are the antenna to which the receiver is connected, the input resistance of the receiver, and the grid circuit of the first tube used in the receiver. If the gain of the first amplifier is low, it is possible that the noise of the second tube in the receiver can also have some effect on the receiver over-all noise level.

a. NOISE SOURCES

A noise voltage will be present across the terminals of any conductor due to the random motion of electrons. This random electron motion is known as thermal agitation noise and is proportional to resistance of the conductor and its absolute temperature. All noise currents and voltages are random fluctuations and occupy an infinite frequency band. The actual magnitude of noise voltage which affects a device is proportional to the bandwidth of the device. The noise voltage can be calculated with the following equation.

$$E_n = \sqrt{4\ KTBR}$$

where E_n = rms noise voltage
K = Boltzmann's Constant, $1.38\ (10)^{-23}$
T = absolute temperature
B = bandwidth in cps
R = resistance in ohms

If the antenna to which the receiver is connected could be placed in a large shielded enclosure, there would exist at the terminals of this antenna a noise voltage equal to the thermal agitation noise of a resistor which is equal to the radiation resistance of the antenna. Even if a receiver could be built with no internal sources of noise, noise would still be introduced into the receiver from the antenna, and weak signals would have to compete with this noise.

Additional noise signals originating in atmospheric disturbances, the sun and other stellar sources, and in electrical machinery increase the noise threshold below which even a perfect receiver could not detect signals. In the h-f band from 2-30 mc, this threshold is usually much higher than that set by receiver internal noise sources. However, the external noise threshold is subject to many variations, and it is possible that under certain favorable combinations of conditions, the receiver noise could be a factor at frequencies in the upper half of the band. For this reason, the noise generated in the receiver circuits is an important consideration.

The internal noise generated in a receiver is conveniently described by a number called the noise figure. The noise figure is expressed as the ratio in decibels between the noise level of the receiver to the noise level of a so-called perfect receiver, in which all the noise is assumed to be generated in the antenna by thermal agitation. A perfect receiver in which the input circuit is designed to match the antenna resistance has a noise figure of 3 db.

The sources of noise in a receiver are the input circuit resistance, the first tube grid circuit, and the second tube grid circuit if the first tube gain is low.

These noise sources are shown in schematic form in figure 3-2. For convenience, the tube noise is usually expressed as being equivalent to the noise generated in a resistance of the proper value, referred to as the equivalent noise resistance of the tube. A tube having a low value of equivalent noise resistance is a low noise tube. Equivalent noise resistances of a number of tubes are listed in table 3-1. From the figures shown, it can be seen that triodes have lower noise than pentodes, and amplifiers have lower noise than mixers. Tube noise, although important, is not a decisive factor in tube selection. Pentode tubes offer advantages over triodes as amplifiers, since they have very low grid-to-plate capacitance and give large gain without neutralization. Pentagrid mixers require less oscillator power, have excellent isolation between signal and oscillator circuits, and give high conversion gain.

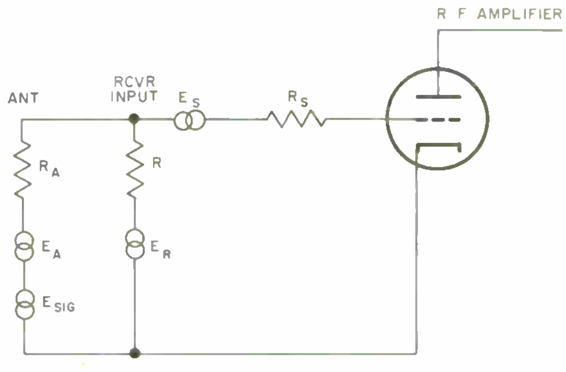


Figure 3-2. Noise Sources in R-F Section of Single-Sideband Receiver

TABLE 3-1. EQUIVALENT TUBE NOISE RESISTANCE

Type	Application	gm or gc	Calculated Req
2C51	Triode Amplifier	5500	455
6AC7		11200	220
6AH6		11000	230
6AN4		10000	250
6BK75		6100	410
6BQ7A		6400	390
6BZ7		6800	370
6J4		11000	230
6J6		5300	470
6T4		7000	360
6U8		8500	295
12AT7		6600	380
12AU7		2200	1140
12AX7		1600	1560
12BH7		3100	810
5687		10000	250
5842		24000	105
6386		4000	625
6AG5	Pentode Amplifier	5000	1650
6AH6		9000	720
6AK5		5100	1880
6AK6		2300	8800
6AU6		5200	2660
6BA6		4400	3520
6BC5		5700	1350
6BD6		2350	13800
6BH6		4600	2360
6BJ6		3800	3860
6BZ6		6100	1460
6CB6		6200	1440
6U8		5200	2280

TABLE 3-1. EQUIVALENT TUBE NOISE RESISTANCE (Cont)

Type	Application	gm or gc	Calculated Req
2C51	Triode Mixer	1375	2900
6AN4		2500	1600
6J4		2750	1450
6J6		1575	2540
6T4		1750	2290
12AT7		1650	2430
12AU7		550	7280
12BH7		775	5170
5687		2500	1600
6386		1000	4000
6AG5	Pentode Mixer	1250	6600
6AK5		1280	7520
6BA6		1100	14080
6BC5		1425	5400
6BZ6		1525	5840
6U8		1300	9120
6X8		2100	7780
6BA7	Pentagrid Converter	950	61700
6BE6		475	174000
6SA7		450	240000
6SB7Y		950	61700

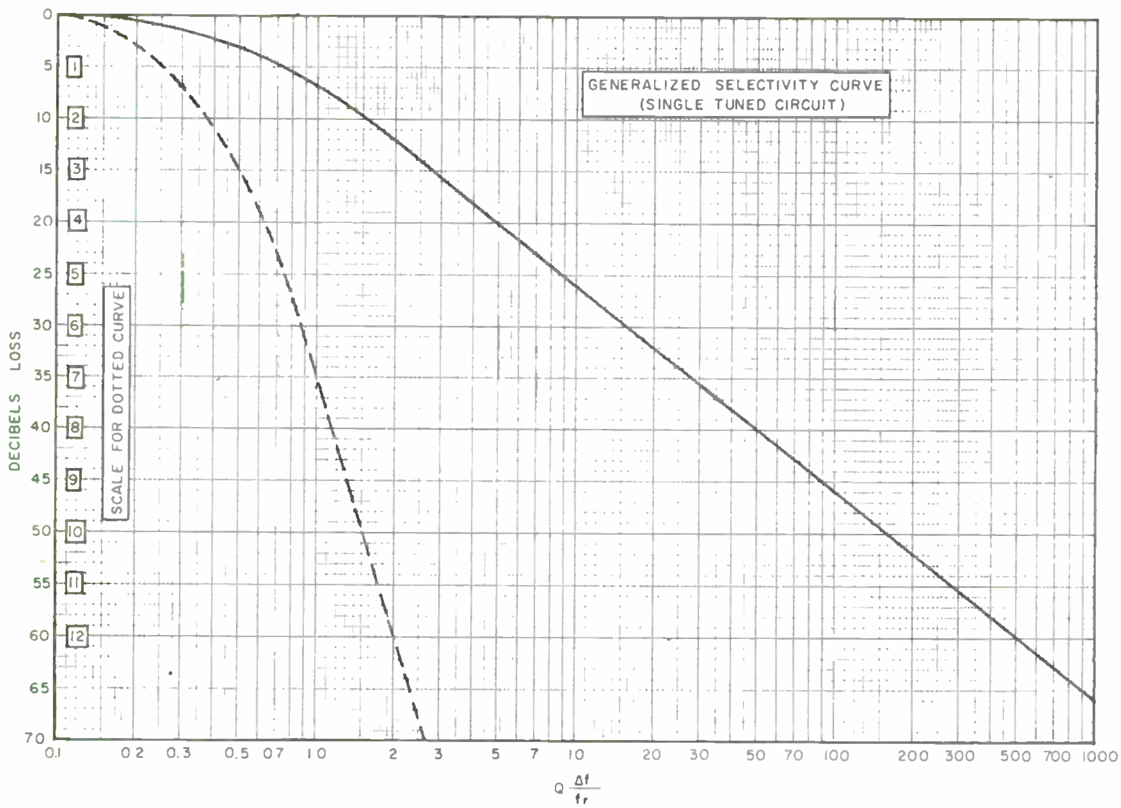


Figure 3-3. Generalized Selectivity Curve

b. R-F SELECTIVITY

For minimum spurious responses, it would be best to provide all the selectivity ahead of the amplifiers in the receiver. This is impractical for several reasons. First, the h-f band spans a range of frequencies in which filters having the required selectivity would be large and difficult to tune. Furthermore, they would have such high insertion loss that the noise figure would be seriously degraded. In some applications where the noise figure can be sacrificed and preselection is a necessity, r-f filters are used. Usually a single-tuned circuit is used between the antenna and the r-f amplifier grid. The selectivity required to suppress adequately the various spurious signals is provided by a tuned filter between the r-f amplifier and the mixer. The tuned filter may consist of several parallel-tuned LC circuits interconnected by mutual inductance or capacitance. The number of tuned elements required depends on the Q factor, frequency, and attenuation required. A universal selectivity curve relating these factors is a convenient tool and is shown in figure 3-3.

c. MIXERS

The r-f signal is translated in frequency from the operating frequency to the intermediate frequency by means of modulation in circuits commonly called mixers. This process has been previously described in detail in chapter 2. The problems encountered in using mixers in receivers is slightly different from those encountered in exciters. Referring to figure 3-4, it can be seen that to translate a desired signal of 1500 kc to an intermediate frequency of 500 kc, a local oscillator having a frequency of 2000 kc can be used. Going further into the example, it can be seen that there are several other signals that can enter the i-f amplifier through the mixer. Some of these signals are listed in figure 3-4. The response at 2500 kc is called the image response and is usually the most troublesome. The higher order responses are attenuated in the mixer tube. Careful selection of a tube and the operating point is necessary to obtain the maximum possible suppression of these responses.

Careful selection of frequencies used in the i-f amplifier is necessary to avoid spurious responses. These responses occur whenever the spurious response frequency coincides with the desired frequency. This type of response is referred to as a crossover, tweet, or birdie, and is illustrated in figure 3-5. A signal of 1001 kc when mixed with an oscillator signal of 1500 kc yields a desired signal of 499 kc. Due to the nonlinearity of the mixer, another product is generated which has frequency equal to the difference between the second harmonic of the signal and the oscillator frequencies. Both these signals are passed by the i-f filter because they are only 3 kc apart. These signals will be demoduled by the audio section to yield an audio output (or tweet) of 3 kc in addition to the usual desired output. Spurious responses are minimized if the intermediate frequency is kept as low as possible consistent with good image rejection.

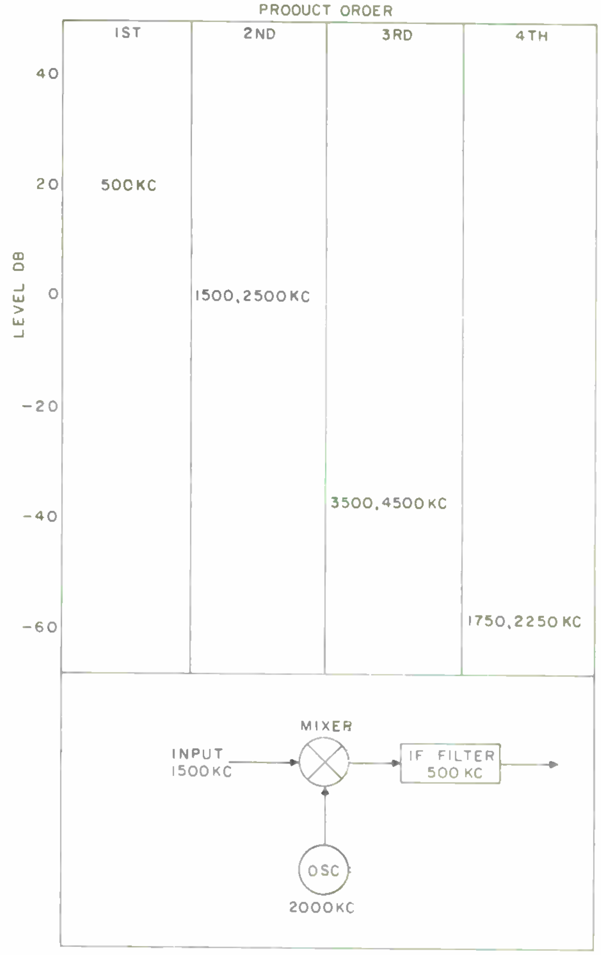


Figure 3-4. Typical Receiver Mixer Spurious Response

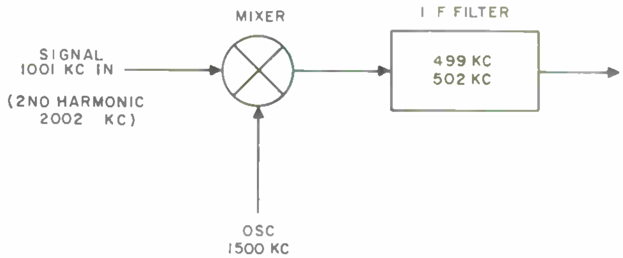


Figure 3-5. Receiver Mixer Crossover Response

As the range over which the receiver must operate is increased, it becomes increasingly difficult to find frequency schemes which are reasonably free of spurious responses. In order to keep these responses attenuated when covering the h-f band (2-30 mc), it is necessary to resort to double conversion or the use of two intermediate-frequency sections. Single conversion is then used on the low frequencies, and the second conversion is brought into use at high frequencies.

The use of double conversion makes possible an improvement of frequency stability through the use of a crystal-controlled high-frequency oscillator. Tuning is accomplished by providing a variable first intermediate frequency ganged to the tunable low-frequency oscillator. As shown in table 2-3, the frequency stability of crystal oscillators are many times better than that obtainable from tunable oscillators. Furthermore, the tuning rate remains the same on the high-frequency bands as it is on the low-frequency bands. On the low-frequency band, the r-f amplifier feeds directly into the tunable i-f circuit, retaining the favorable ratio of signal to i-f frequencies.

For the best sensitivity, it is desirable to have as much gain as possible ahead of the mixers. This would insure that the signal level would be strong enough to override completely the noise from the mixer. From the standpoint of strong signals, it is desirable to have low amplification until the selectivity of the receiver is effective. This would keep the level of strong adjacent channel signals from becoming high enough to overload the initial stages of the receiver. These requirements for no amplification ahead of the selective filter for strong signal reception, and high gain in the r-f amplifier for weak signal reception, conflict, and a compromise is necessary.

When a receiver is tuned to a weak signal and a strong signal is present outside the passband of the i-f selective filter, a type of interference known as cross modulation can exist. The selectivity of the r-f section circuits is not so good as the i-f selective filter, and there is a region near the operating frequency in which strong signals are accepted by the r-f section. Due to the sharp selectivity of the i-f circuits, these signals are not passed by the i-f amplifier and, therefore, do not produce automatic gain control voltage. As a result, these large interfering adjacent channel signals are amplified along with the weak desired signal by the r-f amplifier. When these interfering signal voltages are large enough to drive the amplifier and mixer tubes into nonlinear operation, they cause modulation of the desired signal. To minimize the generation of cross-modulation interference, it is necessary to very carefully select the tubes used in the r-f section. The application of automatic gain control bias is helpful since as the desired signal level increases, the gain of the r-f amplifier can be decreased, reducing the amplification of the interfering signals as well as the desired signal. It is necessary that the tube used in the r-f amplifier retain its linearity with the application of variable bias. It is interesting to note that cross modulation is not as troublesome in single-sideband reception. As an example, if both the undesired adjacent channel signal and the desired signal are conventional AM. signals with full carrier, the modulation of the undesired signal is readily transferred to the desired signal through the process of cross modulation. Effectively the modulation on the undesired signal is modulated onto the carrier of the desired signal. This undesired modulation is passed through the receiver as readily as the desired modulation, and considerable interference results. In the case of single-sideband suppressed carrier reception, there is no carrier present to be modulated, and therefore, the modulation is applied to each of the sideband signal components. As the single-sideband signal consists of a number of relatively weak components, this undesired modulation is spread. Furthermore, when the single-sideband signal is demodulated, the interfering signal is merely recovered as noise and is not as troublesome.

3. I-F SECTION

The intermediate-frequency section contains the frequency selective filter elements and the principal amplifier stages.

a. SELECTIVITY

Consideration must be given to the bandwidth of the receiver as well as the transmitter if the advantages offered by single-sideband communications are to be realized. Optimum receiver selectivity occurs when the noise bandwidth (6 db point) is wide enough to pass the required intelligence, and the skirt bandwidth (60 db point) is narrow enough to reject an unwanted signal in the adjacent communication channel. Extremely steep skirts on the selectivity curves are required to obtain this optimum passband. Ideally, the ratio of the 60 db to 6 db bandwidths should be 1. See curve 3 in figure 3-6. This figure shows the selectivity obtainable from a Collins Mechanical Filter and also from three pairs of double-tune, slightly overcoupled i-f transformers (coil Q's of 150). These curves are superimposed for comparison and show how nearly the mechanical filter selectivity curve approaches the ideal selectivity curve.

Selectivity performance has generally been made by comparing the shape factor, which is the ratio of the 60 db to the 6 db bandwidths. This basis of evaluation has developed from the problem of avoiding adjacent channel interference. While it is customary to define receiver performance in terms of shape factor, it is not always adequate. It can be shown that better shape factors are easier to obtain in wide-band systems than in narrow-band systems. The shape

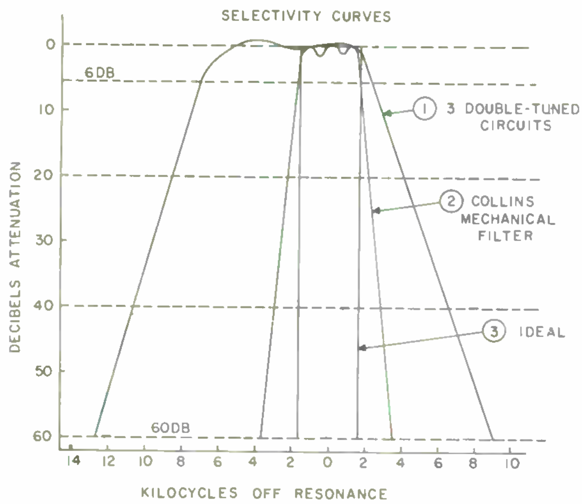


Figure 3-6. Selectivity Comparison

factor is a good comparison if the selectivity curves being compared have the same nose bandwidth. A better method of specifying the performance of a selective system is to define the selectivity in terms of the nose bandwidth and the decibel attenuation per kilocycle on the slopes of the selective curve.

A receiver having an i-f selectivity as in curves 2 or 3 will have a 3 db advantage over a receiver having a selectivity curve as in 1 when receiving an SSB signal whose bandwidth is 3 kc. This is due to both the receiver bandwidth and the input noise power being cut in half. In addition, interference is reduced because the receiver passband is narrow, thus permitting a large percentage of clear signals.

It is desirable to place the selective filter in the circuit ahead of the amplifier stages so that strong adjacent channel signals are attenuated before they can drive amplifier tubes into the overload region. These filters are very similar to the filters used in sideband generators for selecting the desired sideband while rejecting the undesired sideband. Electro-mechanical elements, piezo elements, and inductance capacitance elements can be used in these filters. In one respect, the requirements for these filters are different from those of sideband selecting filters used in the exciter. In order for the receiver to have good rejection to strong adjacent channel signals, it is necessary for the filter used in the receiver to have the ability to reject signals outside the passband to a much higher degree. Attenuation of 60 db or more is necessary for this purpose, and greater rejection is required under some conditions when receiving extremely weak signals. Since single-sideband transmission occupies one-half the bandwidth of a conventional AM. signal, the i-f filter need be only one-half the bandwidth.

b. AMPLIFICATION

The amplifier portion of the intermediate-frequency section consists of the necessary amplifiers to build up the signal to a level suitable for the demodulator. This amplifier consists of cascaded class A linear amplifier stages using remote cutoff pentode tubes. Tuned circuits may be used to provide the load resistance for these stages. The selectivity of these tuned circuits is helpful in improving the over-all receiver selectivity, especially at frequencies which are down on the skirt of the selectivity curve. Some types of filters have spurious responses outside the passband which can be suppressed in this manner.

c. AUTOMATIC GAIN CONTROL

A factor to be carefully considered in single-sideband receiver design is the use of automatic gain control. The basic function of the automatic gain control is to keep the signal output of the amplifier constant and thus hold constant audio output for changing signal levels. This automatic gain control is also applied to amplifiers in the r-f section. However, it is important to delay the application of avc voltage to the r-f amplifier until a suitable signal-to-noise ratio is reached. Conventional AM. systems are generally not usable since they operate on the level of the carrier. This carrier is suppressed in single sideband. Automatic gain control systems must be used which obtain their information directly from the modulation envelope. Refer to figure 3-7. This can be done with conventional diode rectifiers and additional amplification. This may be a d-c amplifier or an a-c amplifier using the i-f frequency. Special care must be taken to isolate the agc system from the reinserted carrier since it is a large signal of the same frequency as the i-f signals. This problem can be avoided by developing the agc voltage from the audio signal. In either case, the time constant of the system is very important. The control must be rapid enough to prevent strong signals from coming through too loud at first and yet be slow enough not to follow the syllabic variation of normal speech. One solution to this time constant problem is to use a fast change, slow discharge type of circuit. Circuits having a charge time of 50 milliseconds and discharge time of 5 seconds have proven successful. Consideration should also be given to dual time constant circuits having a ratio of 100 to 1. Such a circuit allows the rapid signal changes to develop a control voltage across one RC network and the slow signal variation to develop a control voltage across another RC circuit. These two voltages can then be applied in series or to different stages to give the desired control characteristics. Such a dual time constant circuit is similar to a rapid agc system used in conjunction with a manual gain control.

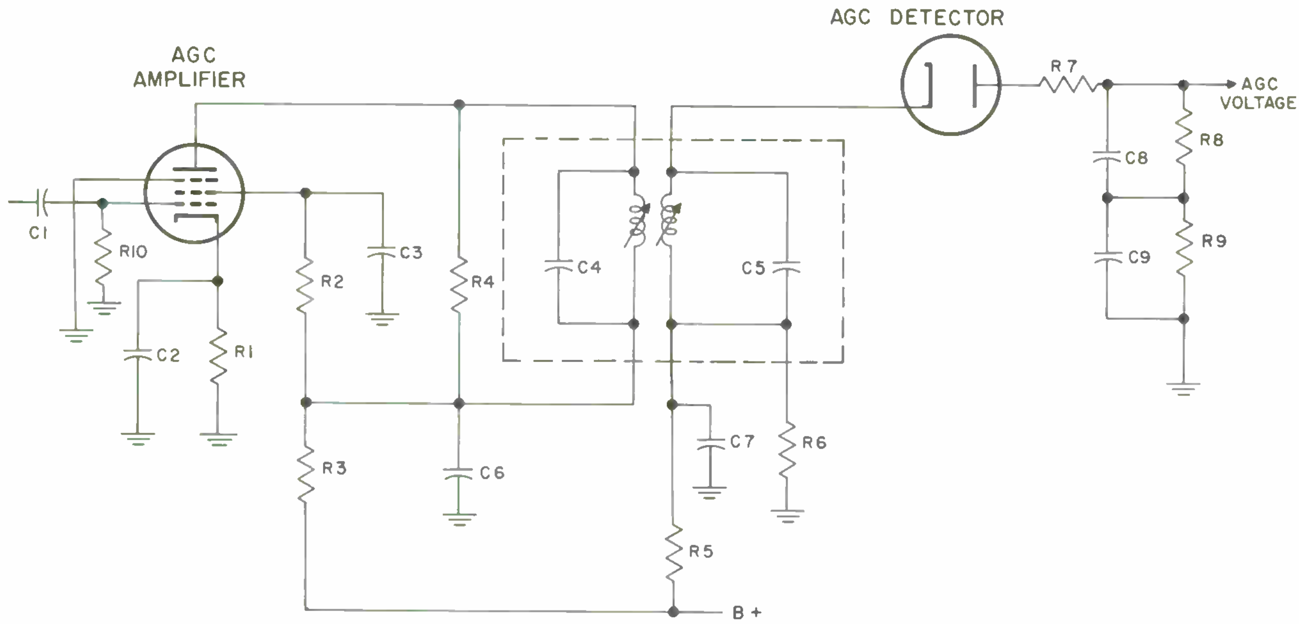


Figure 3-7. AGC Circuit

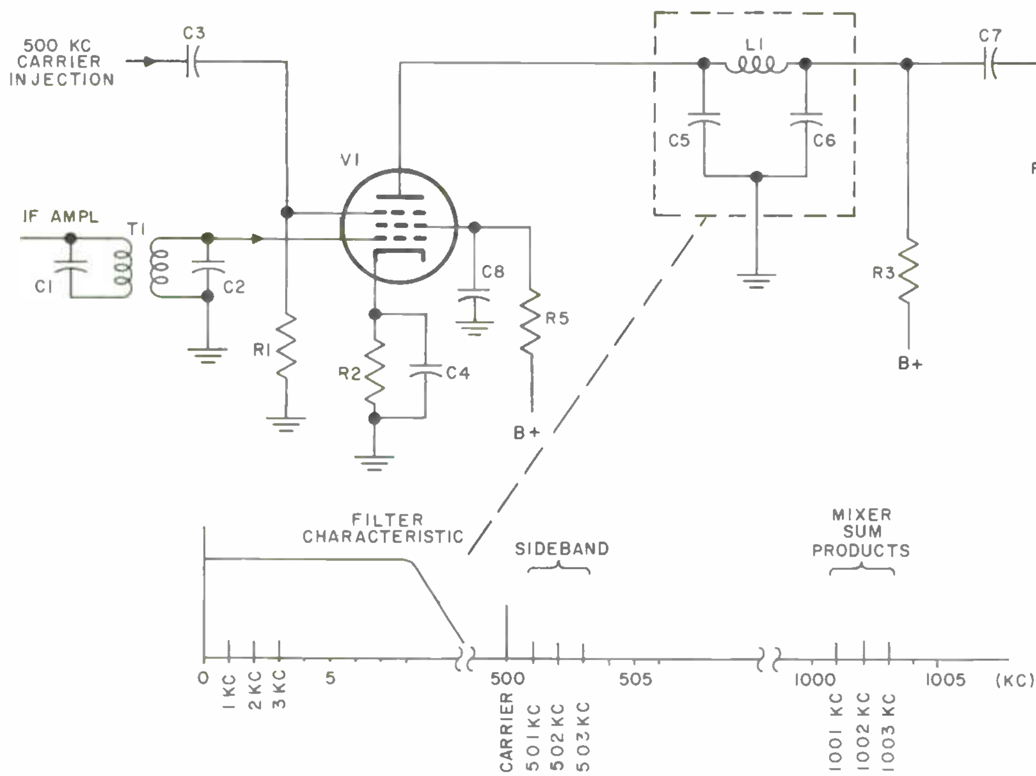


Figure 3-8. Product Demodulator

4. AUDIO SECTION

The information carried by the single-sideband signal is recovered and amplified to a level suitable for the audio output circuits. The circuits used to recover this audio intelligence perform the same function as the modulator in the exciter, and therefore, the same circuits can be used. The single-sideband is first combined with a local carrier. The local carrier must have a proper frequency relationship with the sideband components for faithful reproduction of the original audio signal. In the demodulator, the single-sideband signal is used to modulate the local carrier. The demodulator output consists of an audio signal and several r-f outputs. These signals are easily filtered by passing the output of the modulator through an audio low-pass filter. It is necessary to maintain the proper frequency relationship between the sideband signal and the local carrier. If the received signal is an upper sideband, the carrier frequency is below this sideband; and if the received signal is a lower sideband, the carrier frequency is above the sideband signal. If a receiver must be used to receive either upper or lower sidebands, it is necessary to provide a means of changing the relative position of the carrier with respect to the sideband. One way of accomplishing this is to use two sideband filters in the i-f section and a single carrier at the demodulator. The desired sideband is then selected by switching in the proper filter. A single filter can be used for dual sideband reception by providing a means of shifting the local carrier from one side of the i-f filter passband to the other and then retuning the oscillators in the r-f section.

a. THE PRODUCT DEMODULATOR

Product demodulator circuits are preferred in single-sideband reception, because they minimize intermodulation distortion products present in the audio output signal and do not require large local carrier voltages. Figure 3-8 shows a typical product demodulator. The sideband signal from the i-f amplifier is applied to the control grid of the dual control pentode tube, V1, through transformer T1. The carrier is applied to the other control grid. The desired audio output signal is recovered across resistance R4 in the demodulator plate circuit. Since the plate current of the demodulator is controlled by both grids acting simultaneously, the plate current will contain frequencies equal to the sum and difference between the sideband and carrier frequency. There will also be components of plate current having a frequency equal to the carrier frequency and the sideband frequency. These components are suppressed by means of a low-pass filter (L1, C5, and C6), and the desired audio signal is passed to the audio amplifier. The frequency spectrum presentation in figure 3-8 shows the principal components which will be present in the demodulator plate current. In this example, it is assumed that the sideband signal consists of three components having frequencies of 501, 502, and 503 kc. The carrier frequency is 500 kc. The plate current components consist of three audio-frequency components of 1, 2, and 3 kc and three r-f components of 1001, 1002, and 1003 kc as well as the carrier and original sideband frequencies. By constructing a low-pass filter in the plate circuit, consisting of L1, C5, and C6, it is possible to filter out all frequencies except the difference frequencies. By this method the audio frequency has been recovered from the i-f sideband signal.

CHAPTER 4
STABILIZED MASTER OSCILLATORS

1. TECHNICAL REQUIREMENTS

The frequency accuracy requirements for single-sideband communications are very precise when compared with most other communications systems. A frequency error in carrier reinsertion of 20 cps or less will give good voice reproduction. Errors of only 50 cps result in noticeable distortion, and intelligibility is impaired when the frequency error is 150 cps or greater.

There are significant frequency errors introduced by the propagation medium and by Doppler shifts due to relative motion between transmitter and receiver in aircraft communications. In h-f skywave transmission, the Doppler shifts caused by the motion of the ionosphere introduce frequency shifts of several cycles per second. Doppler shift due to relative motion amounts to one part in 10^6 for every 670 miles per hour difference in velocity between the transmitting and receiving station. At a carrier frequency of 20 megacycles, communicating from a jet aircraft to ground, the frequency shift will be approximately 20 cps. Inasmuch as this represents approximately half of the desired maximum frequency error, the errors introduced by the transmitting and receiving equipment must be comparatively small. This dictates a design goal in the vicinity of $\pm 1/2$ part in 10^6 in both ground and aircraft installations.

Present day trends demand that communications be established on prearranged frequencies without searching a portion of the spectrum in order to obtain netting, and therefore, the figure of ± 0.1 part in 10^6 presents the required absolute accuracy rather than short term stability. Most military and some commercial applications demand that operation be obtained on any one of the seven thousand SSB voice channels in the h-f band. A channel frequency generator having an absolute accuracy of ± 0.1 part in 10^6 ($\pm 0.00001\%$) and providing either continuous coverage or channelized coverage in steps no greater than 4 kc is required in many SSB systems.

2. AFC VS ABSOLUTE FREQUENCY CONTROL

To meet the stringent frequency control requirements, early h-f single-sideband systems utilized various methods of automatic control of the reinserted carrier at the receiver. Either a pilot tone or carrier was transmitted along with the sideband components, and the receiver frequency was synchronized with the transmitter frequency. No stabilization of the transmitter frequency was used other than that obtained by using crystal-controlled oscillators.

The first single-sideband radiotelephony system did not use automatic frequency control and was able to accomplish its purpose because the operating frequency of about 60 kc was low enough that oscillators were then available with sufficient stability. Although oscillators have long been available with sufficient frequency stability and accuracy for use in high-frequency single-sideband equipment, these oscillators were bulky, fragile, and limited in frequency channels. They were used principally as laboratory frequency standards. Improvements in the crystal art, development of circuit technique, and new components have made available the means to obtain h-f receivers and transmitters capable of multichannel operation with sufficient frequency accuracy and stability for independent operation of the receiver.

The advantages obtained through the use of independent absolute frequency control are considerable. The bandwidth required for a communication channel is minimized since there need be no allotment for the synchronizing signal and the frequency tolerance. The relationship between transmitter and receiver carriers is absolute and indestructible and is immune to any type or degree of interference, resulting in maximum fidelity of the received signal. Even in the extreme cases where Doppler effects introduce sufficient frequency shift to upset the system making some form of automatic frequency correction necessary, the use of absolute frequency control assures that the bandwidth and, therefore, the interference susceptibility of the afc circuit will be minimized.

3. DEVELOPMENT OF FREQUENCY CONTROL

It is of some interest to trace the development of frequency control circuits and the technical and economic forces that caused their evolution. In the early days of radio the tunable LC oscillator provided a simple and serviceable answer to the problem of generating channel frequencies. The lower frequency end of the spectrum and amplitude modulation were in use and the spectrum was not unduly crowded.

Later crowding of the spectrum was alleviated by closer channel spacing and expansion into the higher frequency regions. The increased frequency accuracy required was provided by crystal oscillators, and a multiplicity of channels was provided by a like number of crystals. In World War II the logistics of delivering the right crystal to the right place at the right time became untenable.

It became apparent to those involved in multichannel equipment that the simple MOPA circuit would no longer provide desired flexibility. A choice of one of hundreds of channels was required at the flick of a switch, guard bands were narrowed, vhf bands were pressed into more extensive service, and under these forces, the multiple crystal synthesizer soon evolved. The principle was simple: the output frequencies of several crystal oscillators were mixed together to produce the desired output frequencies. Each oscillator was provided with a means of selecting one of ten or more cyrstals so that a large number of channel frequencies may be synthesized. This principle is illustrated in figure 4-1.

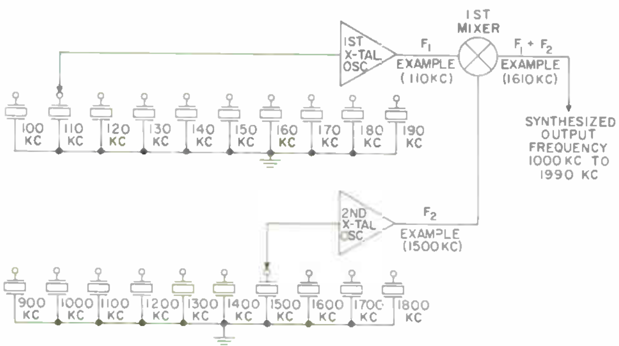


Figure 4-1. Multiple Crystal Frequency Synthesizer

It would be technically and economically unfeasible to maintain all the crystals in a multiple crystal synthesizer to the required accuracy. It would be more practical to place all the stability requirements in one or, at the most, several highly stable oscillators. From this challenge has emerged several operationally satisfactory types of single crystal synthesizers.

4. FREQUENCY SYNTHESIZERS

The frequency synthesizer is basically a circuit in which harmonics and subharmonics of a single standard oscillator are combined to provide a multiplicity of output signals which are all harmonically related to a subharmonic of the standard oscillator. A simple block diagram of such a synthesizer is shown in figure 4-2. A great advantage of this circuit is that the accuracy and stability of the output signal is essentially equal to that of the standard oscillator. The problems involved in building a single frequency oscillator of extreme precision are much simpler than those

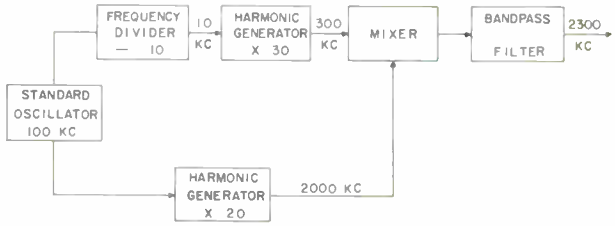


Figure 4-2. Single Crystal Frequency Synthesizer

associated with multifrequency oscillators. Furthermore, as techniques improve, the stability of the synthesizer is readily improved because it is necessary only to replace the standard oscillator to obtain improved precision. The primary difficulty encountered in the design of the frequency synthesizer is the presence of spurious signals generated in the combining mixers. Extensive filtering and extremely careful selection of operating frequencies are required for even the simplest circuits. Spurious frequency problems increase rapidly as the output frequency range increases and the channel spacing decreases.

a. HARMONIC GENERATORS

The generation of higher harmonics of signals from low-frequency sources is a rather difficult problem when carried to higher order harmonics. To obtain stable signals which are exact multiples of a low frequency, several schemes can be used. An ordinary class C amplifier can be used for harmonics up to the ninth. Diode clippers yielding square or rectangular waveforms provide much higher harmonics, but have limited amplitude capability. A blocking oscillator synchronized to the reference frequency generates short, sharp pulses which contain considerable harmonic energy. A particularly effective harmonic generator can be devised using a keyed oscillator (see figure 4-3). In this circuit, the low-frequency reference signal is shaped by a clipper to provide an off-on keying signal which is used to turn on and off a free-running oscillator tuned to the approximate frequency of the desired harmonic of the keying signal repetition frequency. The resulting oscillator output is a train of r-f pulses. If the keying signal is sharply defined and the oscillator starts oscillation uniformly, each pulse will begin on the same r-f phase. The output waveform will then be as shown in figure 4-3. The spectrum of this wave consists of a number of components having various amplitudes grouped around the oscillator free-running frequency. The frequency of each component is an exact integral multiple of the keying signal repetition frequency.

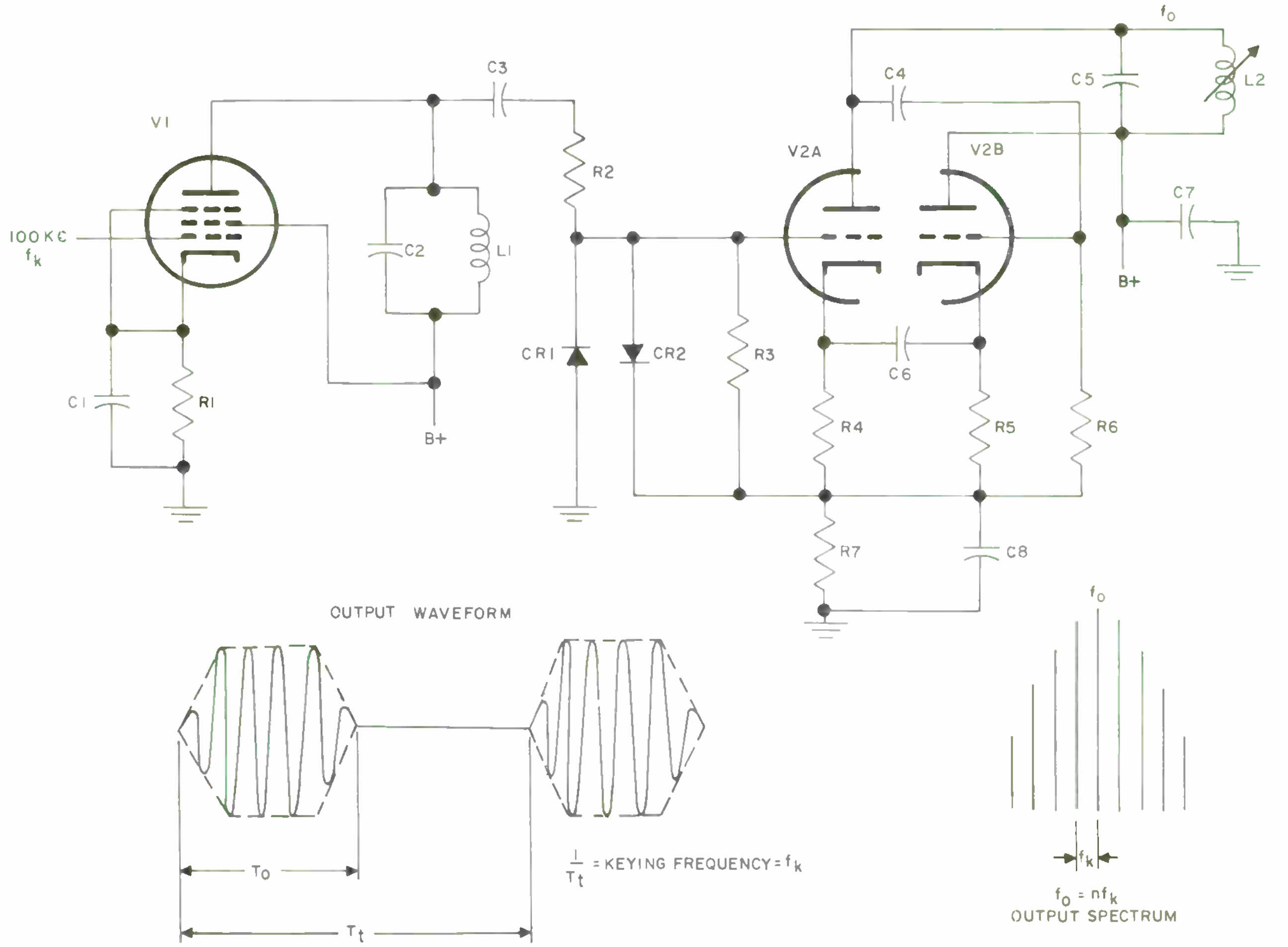


Figure 4-3. Harmonic Generator

b. HARMONIC FREQUENCY SELECTORS

The problem now is to select the desired harmonic while rejecting the adjacent undesired harmonic. Such a selection requires very sharp filters, as the frequency range increases and the spacing between harmonics decreases. By means of an additional mixer it is possible to relieve this situation. Such an arrangement is shown in figure 4-4. In this case it is desired to select higher order harmonics from a one kilocycle source. The one kilocycle reference signal is applied to a harmonic generator, the output of which is tuned to approximately 2.4 megacycles. A considerable number of one kilocycle harmonics will be contained within the harmonic generator output. This harmonic generator output is fed to mixer number one along with a local oscillator of 1945 kilocycles. The desired output of 455 kilocycles is selected by means of a mechanical filter having a bandwidth of less than one kilocycle. This signal is fed to mixer number two along with the output of the same oscillator used to drive mixer number one, and the desired product of 2.400 megacycles is selected in the output filter. To select the adjacent one kilocycle harmonic of the reference signal, the local oscillator is moved to a frequency one kilocycle higher. The desired output is then 2.401 megacycles. The stability of the output signal is dependent entirely upon the stability of the one kilocycle reference signal. The local oscillator frequency accuracy need only be such as to keep the desired 455 kilocycle signal within the passband of the mechanical filter.

5. THE STABILIZED MASTER OSCILLATOR

It is possible to retain the advantage of the frequency synthesizer and avoid many of the spurious frequency

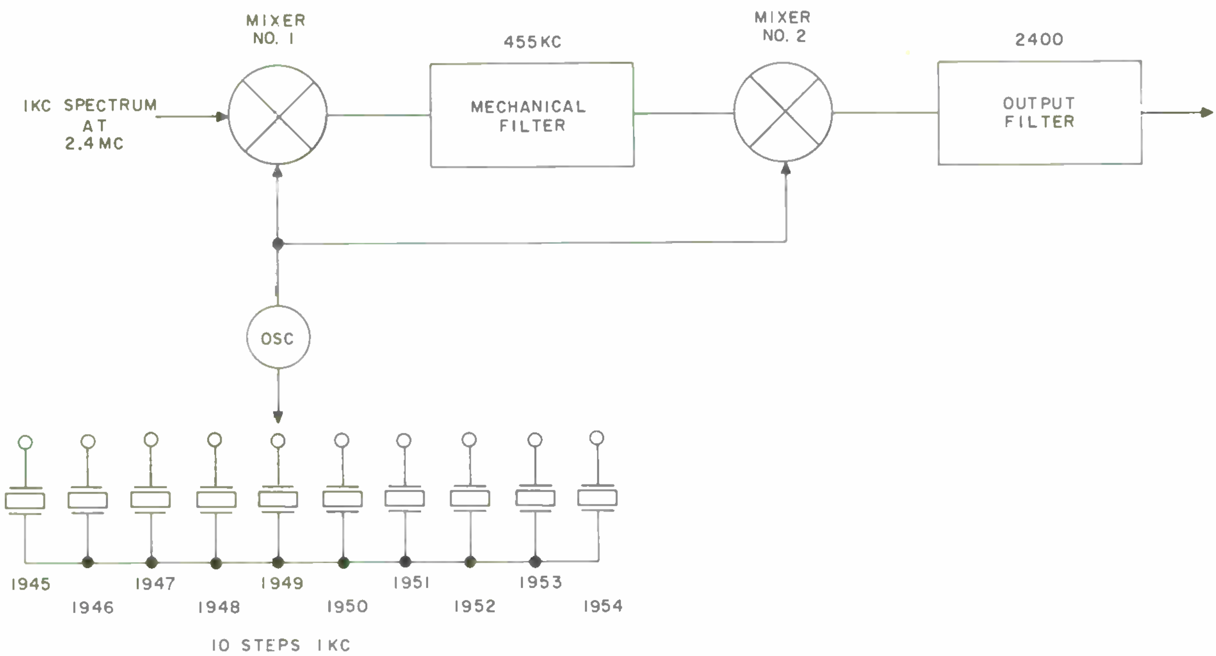


Figure 4-4. Harmonic Frequency Selector

problems by using the synthesizer to provide a reference signal to control the frequency of a variable frequency master oscillator. Such a circuit has come to be known as a stabilized master oscillator (frequently referred to by its initials smo). The basic elements of a stabilized master oscillator circuit are the master oscillator, reactance control, and discriminator (see figure 4-5). The frequency of the master oscillator is determined by the inductance and capacitance of elements L1 and C1. The frequency of oscillation may be manually changed by varying the capacitance of C1, or electronically changed by varying the permeability of the core on which L1 is wound.

a. FREQUENCY DISCRIMINATOR

The operation of the frequency discriminator is such to provide a d-c output signal whose amplitude and polarity is determined by the relationship between the input signal frequency and the frequency to which the discriminator is tuned. The frequency discriminator consists of a double-tuned transformer and two diode rectifiers (see figure 4-6). The transformer is used to supply signals to the two rectifiers, the outputs of which are series connected. The coupling capacitor C1 places the centertap of the secondary at the same r-f potential as the plate end of the primary winding. As a result of this connection the voltage applied to each diode is the sum of one-half the secondary voltages plus the voltage appearing across the primary. The voltage applied to each diode is shown in the vector diagrams below the circuit diagram. The action of the discriminator depends on the fact that the phase of the voltage developed across the secondary of the discriminator transformer will vary as the frequency of the applied signal is varied above and below the transformer resonant frequency. Referring to the vector diagrams (figure 4-6) it can be seen that if the applied signal frequency is equal to the discriminator frequency, the equal voltages are applied to each discriminator diode and the d-c output of the discriminator is zero (figure 4-6B). If the applied frequency is higher than the discriminator frequency, the voltage applied to diode one exceeds that applied to diode two, and the resulting d-c output is positive (figure 4-6A). If the applied signal is lower than the discriminator frequency, the voltage applied to diode one exceeds that applied to diode two, and the resulting d-c output is negative (figure 4-6C). This direct current output is applied to a reactance control device.

b. REACTANCE CONTROL

The reactance control provides the means by which the direct current output of the discriminator is made to alter the inductance or capacitance of the

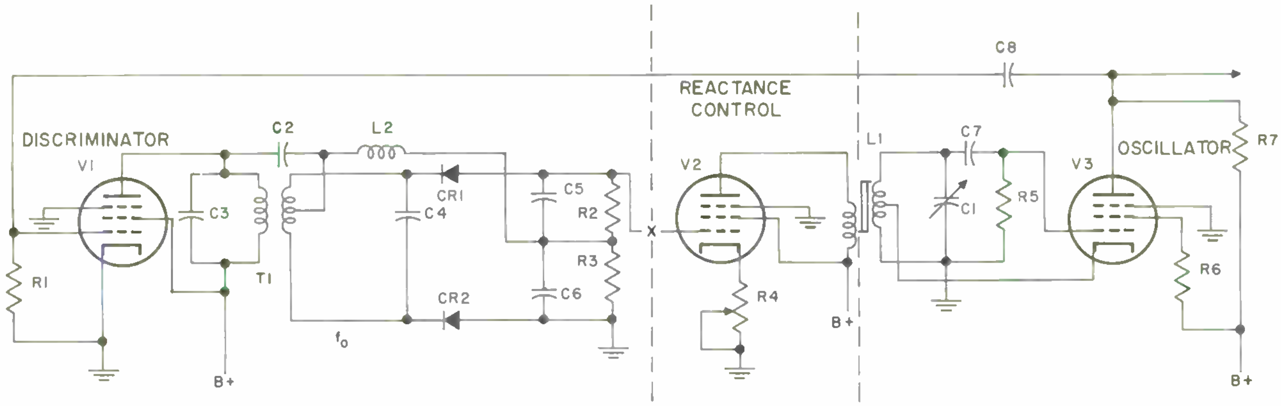


Figure 4-5. Simple Stabilized Master Oscillator with Frequency Discriminator

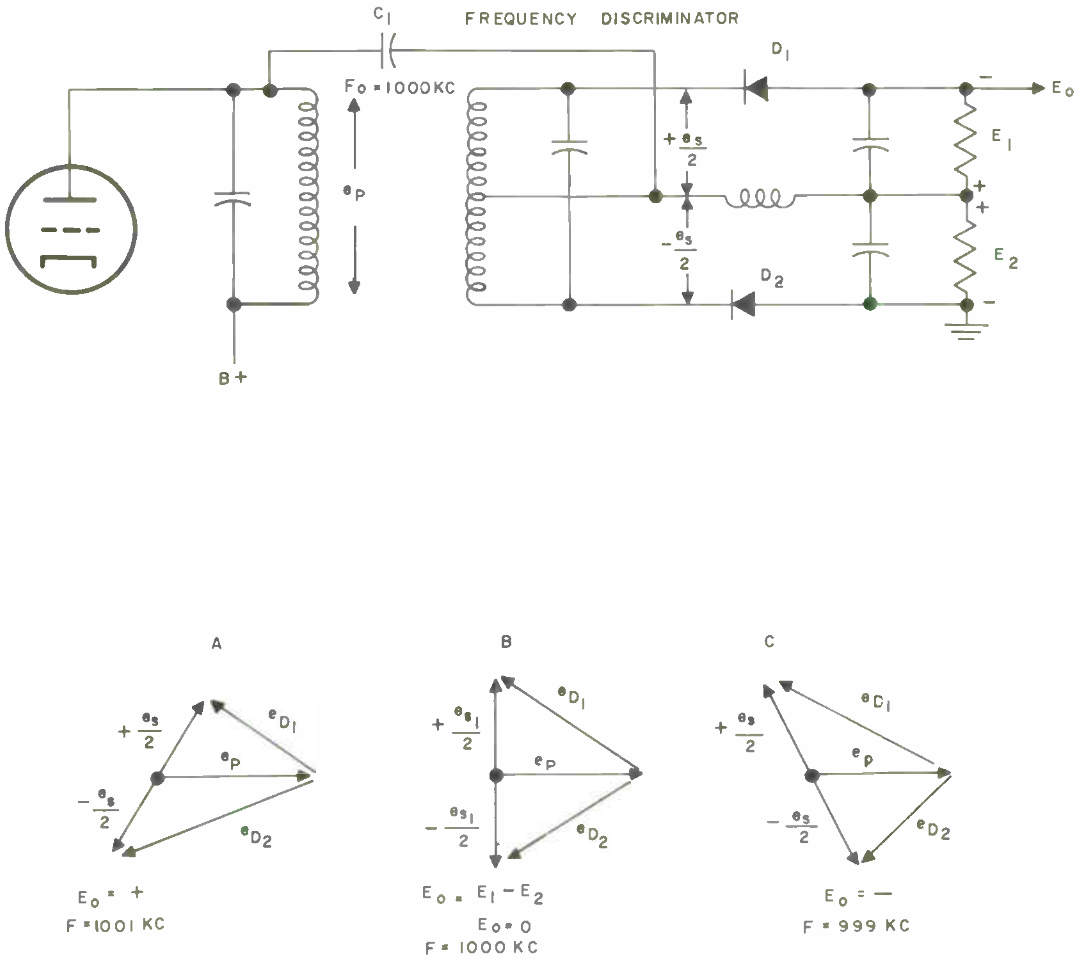


Figure 4-6. Frequency Discriminator

tuning elements of the master oscillator. Devices that have been used for this are reaotance tube circuits, saturable reactors (i.e. current-sensitive inductors), voltage-sensitive capacitors, and motor-driven variable capacitors. The saturable reactor is used in the example given as its operation is easily understood. The saturable reactor consists of an inductor wound on a core material having magnetic permeability which is a nonlinear function of the magnetizing force. Such a reactor will have inductance which can be changed by varying the current in its winding or through an auxiliary control winding (see figure 4-7). The change in permeability will be the same for either polarity of magnetizing force. For this reason it is necessary to resort to fixed magnetic bias to obtain inductance change that will reverse polarity when the external magnetizing force polarity reverses. The magnetic bias may be obtained from a permanent magnet or from a bias current in the control winding as is the case in the example shown.

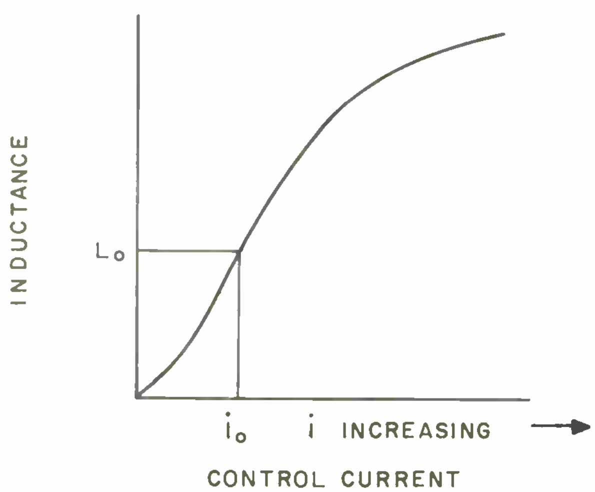


Figure 4-7. Variable Inductor Response Curve

c. BASIC SMO OPERATION

The manner in which the stabilized master oscillator circuit operates may be described in two conditions, open-loop and closed-loop. If the control is opened at the grid of the reactance control tube and the tuning of the oscillator varied with the discriminator tuning fixed at f_0, the output voltage of the discriminator will follow the open-loop curve shown in figure 4-8. In this curve the discriminator voltage is plotted on the vertical scale versus oscillator frequency on the horizontal scale. If the master oscillator frequency differs from the discriminator frequency when the loop is closed, the master oscillator frequency will be pulled toward the discriminator frequency provided the proper polarity of discriminator and control device has been observed. It is important to realize that perfect correction cannot be achieved unless there is an infinite amount of amplification of the error signal from the discriminator. This can be seen by examining the discriminator output when the loop is closed. If perfect correction had somehow been achieved, the discriminator output would be zero. Obviously such cannot be the case as there must be some signal applied to the reactance control to correct the oscillator frequency error. The closed-loop frequency error will depend on the master oscillator error and on the gain of the control loop. The performance of the closed-loop is shown by the dashed curve in figure 4-8.

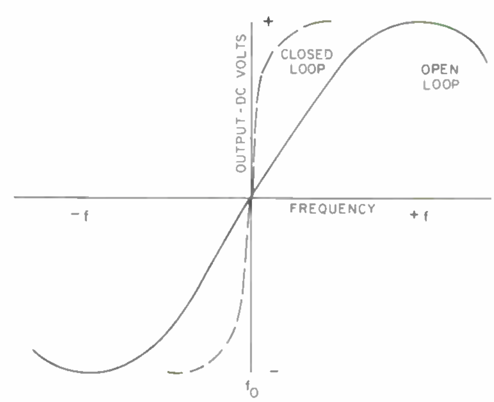


Figure 4-8. Discriminator Frequency Response Curve

In the example shown, the oscillator frequency is effectively compared with the discriminator frequency. There are two fundamental defects in this stabilizing system: (1) there is a residual frequency error and, (2) the stability obtainable from the discriminator is limited. The over-all accuracy of a system using this principle in the h-f band is approximately 50 PPM, insufficient for SSB service.

Both of these shortcomings can be eliminated by utilizing phase deviation rather than frequency deviation error signals in the control-loop. To do this, the frequency discriminator is replaced by a phase discriminator with a reference signal derived from a standard oscillator (see figure 4-9). The stabilized master oscillator is then locked in frequency synchronization with the reference oscillator, and the error signal is the phase angle between the two oscillator voltages. As the frequency of the controlled oscillator drifts away from the reference frequency,

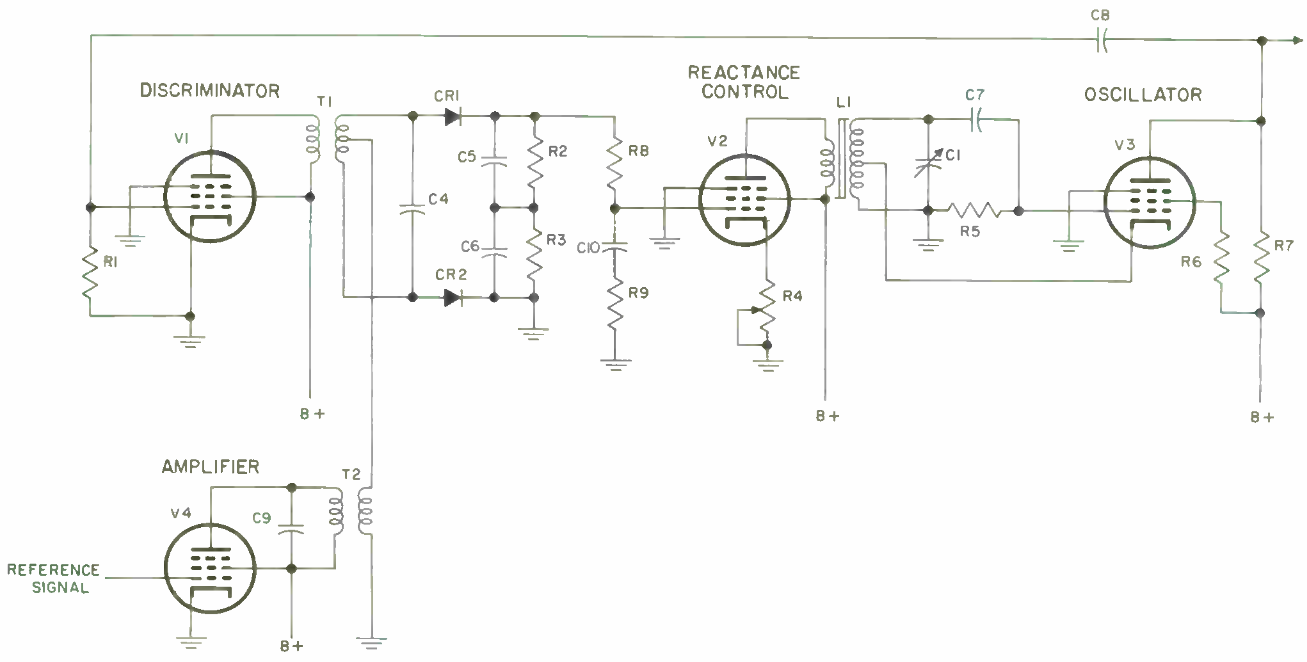


Figure 4-9. Simple Stabilized Master Oscillator with Phase Discriminator

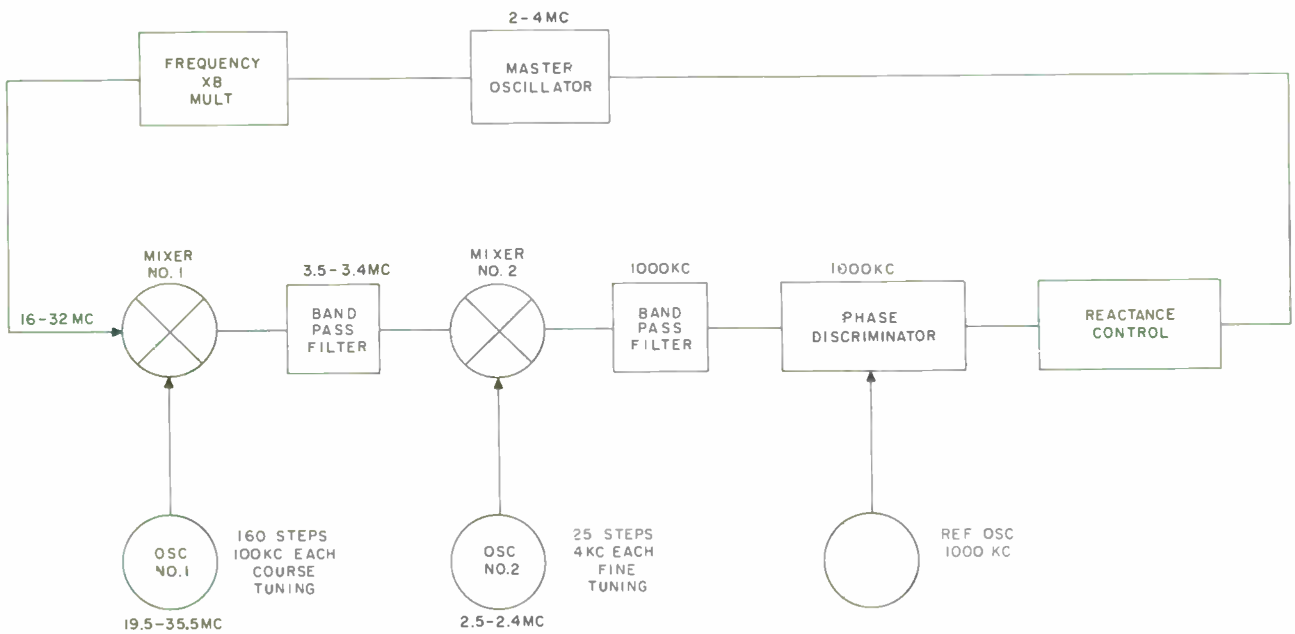


Figure 4-10. Block Diagram of Basic Stabilized Master Oscillator

the phase angle between the two voltages increases to provide the correcting voltage necessary to operate the reactance control. The stability of the system now is completely dependent on the stability of the reference oscillator. The stabilization loop is a feedback system and, as a result, careful attention must be paid to gain and phase shift if stable operation is to be obtained. If the gain at the frequency at which phase shift around the loop is 180° is unity, oscillation will result. The low-pass filter network in the grid of the reactance control tube provides the necessary control of gain to avoid oscillation by reducing the gain at the critical frequency.

d. MULTIPLE FREQUENCY SMO OPERATION

Although the stabilized master oscillator described is capable of operating on one frequency only, the circuit can be extended to operate on additional frequencies. To accomplish this, a double-mixer frequency translation system is designed to feed the discriminator (see figure 4-10). By this means, the reference frequency is translated upward in frequency to the range 16 to 32 megacycles. As long as oscillator two is fixed in frequency, the reference signal can be translated to frequencies separated 100 kilocycles between 16.0 and 32.0 megacycles, 16.0, 16.1, 16.2, etc. If the frequency of oscillator one is fixed and oscillator two is varied, the reference signal will be translated in steps of four kilocycles each over an interval of 96 kilocycles, (16,000, 16004, 16008, etc.). In this way the reference frequency can be translated to any one of 4000 frequencies between 16 and 32 megacycles, spaced four kilocycles apart.

The master oscillator operates from two to four megacycles, this range being best suited to covering the h-f band using both fundamental and harmonics. For synchronization with the reference, the master oscillator output frequency is multiplied by eight so that the master oscillator signal frequency range corresponds to the reference frequency range. Under these conditions the master oscillator fundamental output frequency will be stabilized on 1/2 kilocycle intervals over the range two to four megacycles. More channels can be synthesized by adding another mixer stage or by increasing the number of steps used at each mixer.

The accuracy of the stabilization obtained by the system described above depends on the accuracy of the frequencies used at the translating mixers. To obtain the greatest accuracy, all of these frequencies are derived from a single source; a standard reference oscillator of extremely high accuracy and having great stability.

The use of a phase error signal in the control loop insures that the residual error of the stabilized oscillator will be measured in terms of degrees of phase angle between controlled and reference oscillators rather than cycles of frequency difference if only a frequency discriminator were used.

CHAPTER 5
FREQUENCY STANDARDS

1. INTRODUCTION

Because frequency is defined in terms of cycles per second or events per unit time, frequency control and timekeeping are inseparable. Any measurement of frequency can be only as accurate as the time unit used. Thus in order to determine the accuracy of a frequency standard, its period of oscillation must be compared to a time standard of known accuracy. The best secondary time standard consists of a frequency standard which drives a cycle counter or clock. The accuracy of this time standard can be then determined by comparing it with the primary time standard which is the mean solar day, that is, the time required for the earth to complete one revolution about its axis.

Time measurement has always been based on astronomical phenomena. Days and years are determined by the relative motion of the earth with respect to the sun. However, to co-ordinate events and to make precise measurements of physical phenomena, a device that can divide the day into accurate, shorter intervals is required. The search for accurate timing devices started in prehistoric time. It followed two separate lines: first, those devices which derive time directly from astronomical observations; and second, independent mechanisms and devices for measuring time intervals. The first type started with the casual observation of the position of the sun, progressed to the sundial, and culminated in the modern Zenith tube. The second type started with devices based on restricted flow. The first of these were the noncycling types, such as the sand clocks, water clocks, time candles, and time lamps. Typically, sand clocks or hourglasses had inaccuracies of 4,000 seconds per day. These were followed by automatic recycling types using escapement mechanisms controlled by friction and inertia, such as the Verge and Foliot balance. Clocks of this type varied 1,000 seconds per day. With the discovery of resonance phenomena, that is, an oscillating system in which energy is alternately stored in the form of kinetic and potential energy, much more accurate time measurements were made possible. The first device using resonance phenomena to measure time was the pendulum clock which ultimately attained an accuracy of .002 second per day. The pendulum was followed by the hairspring and balance which attained an accuracy of .2 second per day, and the electrically activated tuning fork which attained an accuracy of .008 second per day. The quartz crystal, which followed the tuning fork as the resonator in time and frequency standards, has an accuracy of 1 part in 10^9 per day or .0001 second per day and is the most widely used control element in modern time and frequency standards. Due to variations in the rotation of the earth, the short-term accuracy of the quartz crystal is better than that of the mean solar day. Seasonal variations of several milliseconds and yearly variations as great as 1.6 seconds in the mean solar day have been observed. On a long-term basis, however, the length of the mean solar day is increasing at the rate of only .00164 second per century. In order to achieve long-term accuracy, the standard must remain in constant operation; and since mechanical devices such as the quartz crystal clock will run for only a few years, they will probably not replace astronomical phenomena as a primary time standard. The most recent development is the use of devices based on atomic or molecular resonance. These have attained short-term accuracy equal to that of the quartz crystal, but their long-term accuracy is expected to be considerably better.

Technical and economic forces have led to the development of more and more accurate frequency control circuits. In the early days of radio, the tunable LC oscillator provided a simple and serviceable answer to the problem of generating channel frequencies. The lower frequency end of the spectrum and amplitude modulation were used, and the spectrum was not unduly crowded. Later crowding of the spectrum led to closer channel spacing and expansion into the higher frequency regions. This in turn required more accurate frequency control. Crystal oscillators provided the required accuracy, but many crystals were required to provide the required number of channels. During World War II, it became almost impossible to deliver the right crystal to the right place at the right time. After the war, users of communication equipment demanded a choice of hundreds of channels at the flick of a switch. In order to meet the demand for spectrum space, guard bands were narrowed, and the vhf bands were put to more extensive use. All of these forces led to the development of the multiple-crystal frequency synthesizer (figure 5-1) in which the output frequencies of several crystal oscillators were mixed together to produce the desired output frequencies, providing many more channels than the number of crystals used. Present crowding of the spectrum and increasing demand for communication channels now indicate that some method of further decreasing the spectrum space required for each channel must be found. Single sideband is a solution to this problem. The use of single

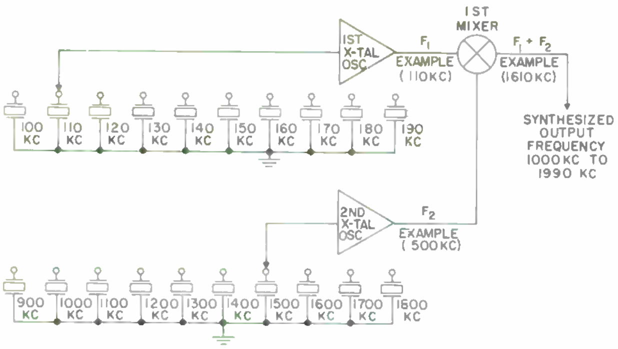


Figure 5-1. Multiple-Crystal Frequency Synthesizer

sideband, however, requires that channel frequencies be maintained within ±1/2 part per million. Maintaining all of the crystals in a multiple-crystal synthesizer to the required accuracy is impractical; therefore, all the stability requirements must be concentrated in one or, at the most, several highly stable oscillators. The solution to this problem is the single-crystal frequency synthesizer. Figure 5-2 is a block

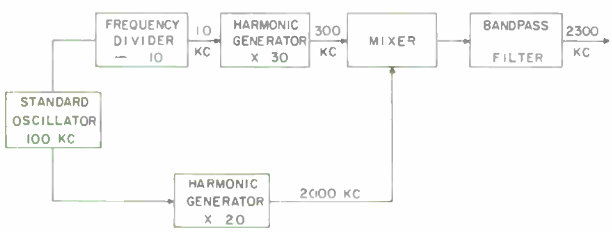


Figure 5-2. Single-Crystal Frequency Synthesizer

diagram of a typical single-crystal frequency synthesizer. Basically it is a circuit in which harmonics and subharmonics of a single standard oscillator are combined to provide a number of output signals which are all harmonically related to a subharmonic of the standard oscillator. In this system the accuracy and stability of the output signals are equal to that of the standard oscillator and as techniques improve, the stability of the synthesizer can be improved by replacing only the standard oscillator.

The best laboratory standards now available have aging rates of approximately 1 part in 10^9 per month and short-term variations of several parts in 10^{11}. Operational standards have several orders of magnitude greater instability. Typical examples are shown in the curves of figures 5-3 and 5-4. In figure 5-3, the dots represent errors derived from direct time comparison with WWV, and the crosses represent errors derived from time comparison with WWV after correction according to WWV's time correction bulletin.

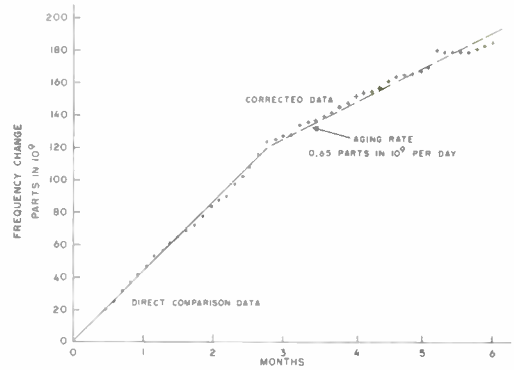


Figure 5-3. Typical Long-Term Stability

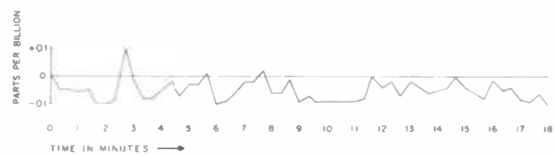


Figure 5-4. Typical Short-Term Stability

2. DETERIORATION OF GENERATED FREQUENCY

Although frequency standards in use today have accuracies of 1 part in 10^8 or better, serious errors can be introduced in the transmission and reception of the signal. These errors are caused by Doppler shift, shifts due to propagation characteristics, and shifts due to equipment circuitry.

a. EFFECT OF DOPPLER SHIFT

Relative motion between receiving and transmitting stations causes premature or delayed reception of individual cycles of the transmitted signal. Since the speed of propagation of radio signals is equal to the speed of light or 186,000 miles per second, cycles of the transmitted signal will be received 1 millisecond earlier or later for every 186 miles of change in transmission path length. A change in transmission path length at a rate of 670 miles per hour results in a frequency shift due to Doppler effect of 1 part in 10^6. Figure 5-5 shows an aircraft approaching a radio transmitter. In the formula shown in the figure, v = velocity of the aircraft and C = speed of light. If the aircraft is approaching at a velocity of 670 miles per hour or 0.186 miles per second, then the ratio v/C = 0.186/186,000 or 1/1,000,000. Thus the ratio of frequency change to transmitted frequency ($\Delta f/f$) is $1/10^6$. If the transmitter is operating on a frequency of 10 mc, then the frequency as received at the aircraft will be 10 mc plus 10 cps. If the transmitter were also in an aircraft flying toward the first aircraft at a velocity of 670 miles per hour, the frequency error would be doubled because the relative velocity would be the sum of the velocities of the two aircraft or 1,340 miles per hour. In ship to ship communication or in communication between ground vehicles, Doppler shifts of 1 part in 10^7 or greater are possible. Doppler shifts due to antenna sway caused by the pitch and roll of a ship are of the order of ±3 parts in 10^9. Extreme examples of Doppler shift are the case of back-pack radios in which the Doppler shift while the operator is walking is 5 parts in 10^9, and the case of the IGY satellite, where signals transmitted by the satellite will suffer frequency shifts of up to 30 parts per million. Signal transit time in the case of a jet aircraft traveling 670 miles per hour and communicating with a fixed station changes at the rate of 3.5 milliseconds per hour, and in the case of battleships communicating with each other the signal transit time can change 0.2 milliseconds per hour.

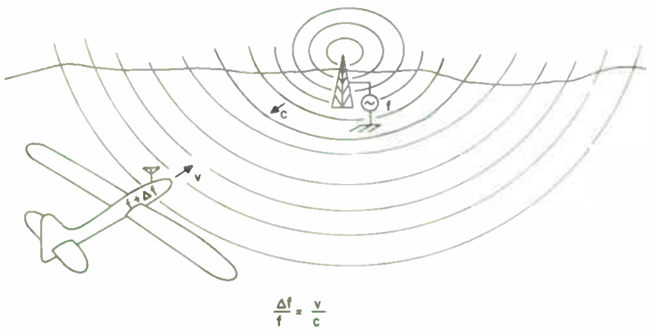


Figure 5-5. Doppler Frequency Shift in Aircraft

b. EFFECT OF PROPAGATION CHARACTERISTICS

Low-frequency waves tend to follow the curvature of the earth, and the length of the transmission path is not seriously affected by atmospheric or ground conditions. Errors introduced by the propagation medium at low frequencies are only about ±3 parts in 10^9 in frequency and ±40 microseconds in transit time. In the high-frequency bands, however, reflections from the ionosphere are used for long-range communications. Frequency variations of ±2 parts in 10^7 and transit time variations of ±1 or 2 milliseconds can be introduced by changes in path length due to movement of the reflection point in the ionized layer and variations of the skip distance. Errors introduced in vhf and uhf scatter propagation are not well known, but available data indicate that they may be several parts in 10^8 in frequency and several hundred microseconds in transit time.

c. EFFECTS OF EQUIPMENT CIRCUITRY

Transmitter or receiver circuit elements when subjected to mechanical vibration or temperature changes can cause temporary frequency shifts by temporarily shifting the phase of the signal. Phase advancement of 360 degrees in one second adds 1 cycle per second to the frequency of a signal. Thus, a phase shift change of 1 degree per second imposed on a 100 kilocycle signal would cause a temporary frequency shift of 3 parts in 10^8. Mechanical vibration of tuning elements causes phase shifts which even under laboratory conditions may cause frequency shifts as great as 1 part in 10^8. Under operating conditions severe mechanical vibration and temperature changes may be encountered which, if not compensated for, would cause excessive frequency errors. Therefore, in precision work, mechanically rigid components must be used in all tuned circuits.

3. MEASUREMENT TECHNIQUES

a. TIME COMPARISON

Accurate comparisons of time and frequency using radio communication are difficult because of variations in propagating mediums. Present methods are based on time measurements taken over a long period so that these variations average out. By taking time measurements from WWV over a period of 20 days, accuracies of 1 part in 10^9 can be attained in the 2 mc to 30 mc bands. At 16 kc, the same accuracy can be attained in approximately one day because the variations in the propagating medium have less effect on the low-frequency signal. Figures 5-6 and 5-7 are block diagrams of time comparison systems suitable for fixed station use. On shipboard errors introduced by changes in signal transit time due to relative motion between stations must be taken into account to achieve the same accuracies as are attained in fixed station use.

Figure 5-6 shows a system using an aural indication of synchronization of a clock, controlled by a local oscillator, with the time signals transmitted by WWV. The oscillator operates at 100 kc; this frequency is divided by 100, and the resulting 1000 cps signal operates the synchronous clock. The clock operates a switch which closes once each second. A receiver tuned to WWV's signal is used to detect the

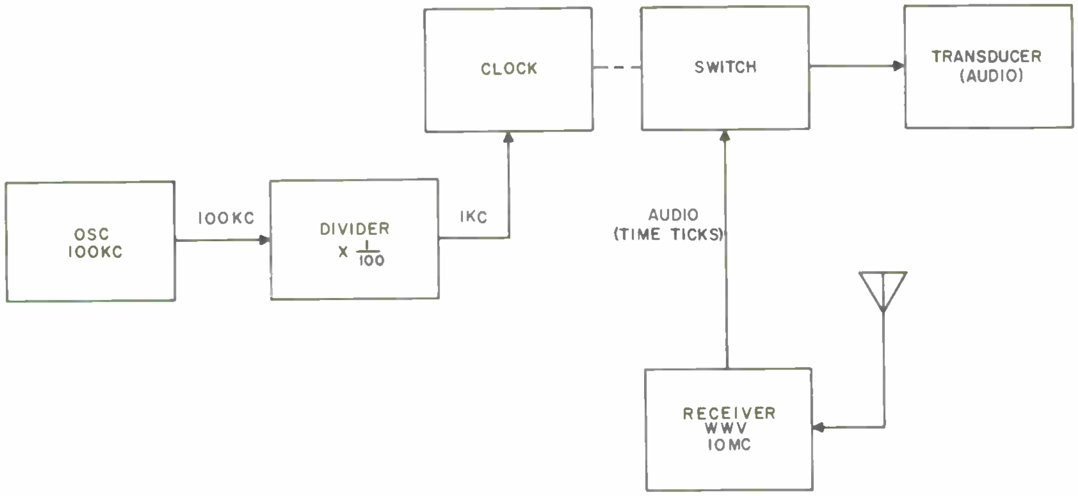


Figure 5-6. Time Comparison System, Aural Indication

time signals which are in the form of clock ticks. These clock ticks consist of 5 cycles of a 1000 cps tone, transmitted at the rate of one tick per second. The ticks are coupled to a loud-speaker through the clock operated switch which can be adjusted to close each time a tick is received. Once adjusted, the switch will continue to close in synchronism with the reception of the clock ticks as long as the frequency of the oscillator remains exactly 100 kc. If the oscillator frequency changes, the speed of the clock will also change, and the switch closures will slowly drift out of synchronization. A calibrated dial is used to adjust the synchronization of the switch daily to permit only the last cycle of the clock tick to pass. The stability of the oscillator can be determined to an accuracy of 12 parts in 10^8 by calculations based on the amount of adjustment required in one day. The accuracy of measurement can be increased to 1.2 parts in 10^9 by basing the calculations on the amount of adjustment required in a period of 100 days.

Figure 5-7 is a block diagram of a chronoscope. The 100 kc signal from the oscillator is divided to 10 cps and applied to the vertical and horizontal plates of the cathode-ray tube through phase shifting networks to produce a circular trace on the scope screen. A 1

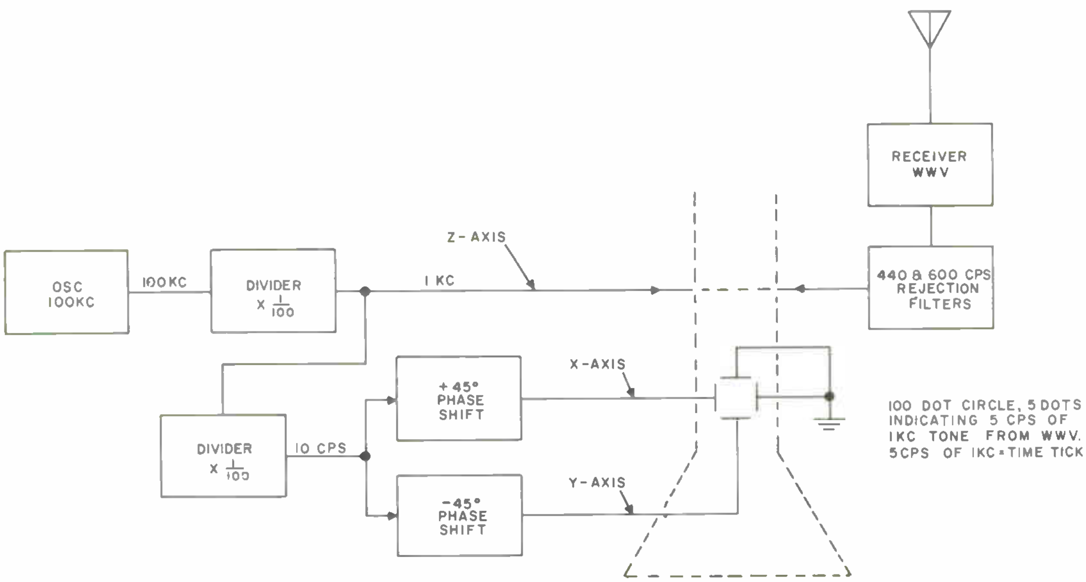


Figure 5-7. Time Comparison System, Visual Indication, Chronoscope

kc signal derived from the same oscillator is applied to the intensity control grid to break this solid circle into 100 dots. A receiver tuned to WWV supplies the clock ticks to the same control grid producing five additional dots somewhere on the 100 dot circle depending upon the relative phase of the 1 kc signal from the oscillator and the 5 cycles of 1 kc which comprise the clock tick. If the phase relationship remains constant, the 5 dot pattern on the screen will remain fixed; but if the phase changes, the pattern will move around the 100 dot circle at a rate determined by the rate of phase change. The rate of movement in turn indicates the magnitude of frequency error. With this system the frequency of the oscillator can be determined to an accuracy of 1.2 parts in 10^8 in one day or 1.2 parts in 10^9 in 10 days.

b. FREQUENCY INTERCOMPARISON

Short-term stabilities of oscillators can be determined by intercomparison of the frequencies of two or more oscillators. When only two oscillators are compared, only the relative stabilities of the oscillators with respect to each other can be determined. Statistical data which will indicate the short-term stability of an individual oscillator can be obtained by intercomparing the frequencies of three or more oscillators two at a time.

Figure 5-8 illustrates a system using two oscillators operating at frequencies differing by, nominally, 1 cps. One oscillator operates at 1 mc, and the other operates at 1 mc plus 1 cps.

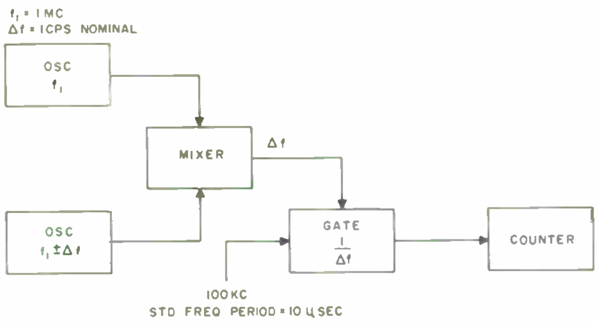


Figure 5-8. Frequency Intercomparison System Using Frequency Counter

Their outputs are mixed and the difference frequency, 1 cps, is used to control a gate circuit. A 100 kc standard frequency is applied to the gate and the number of cycles of this standard frequency which are counted at the output indicates the length of time the gate is open. This in turn is the period of the difference frequency controlling the gate. Thus if the difference frequency is exactly 1 cycle per second, the gate will be open exactly 1 second, and the counter will count exactly 100,000 cycles. If the difference frequency is more than 1 cycle per second, the counter will count less than 10^5 cycles, and if the difference frequency is less than 1 cycle per second, the counter will count more than 10^5 cycles. This system will indicate the relative stability of one oscillator with respect to the other to 1 part in 10^{11}.

Figure 5-9 illustrates a system using two oscillators adjusted to operate at frequencies differing by 0.6 cps. The frequency of each oscillator is multiplied by 100 before mixing so that the resultant beat note is 60 cps. This beat note is recorded on a power line frequency recorder to give a continuous indication of relative stability between the two oscillators with an accuracy of 5 parts in 10^{10}.

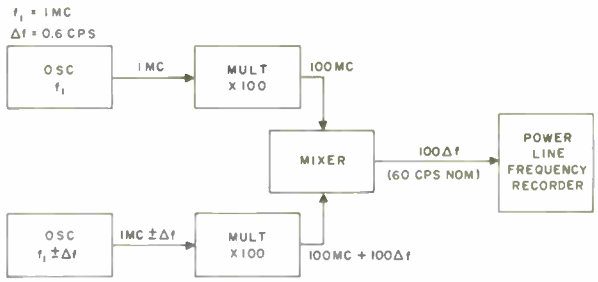


Figure 5-9. Frequency Intercomparison System Using Power Line Frequency Recorder

c. PHASE INTERCOMPARISON

Figure 5-10 illustrates a system wherein the relative phase of two oscillators operating at the same frequency is measured. If the relative phase as indicated by the phase comparison meter changes 360 degrees in one second, then the difference in frequency of one oscillator with respect to the other is 1 cps;

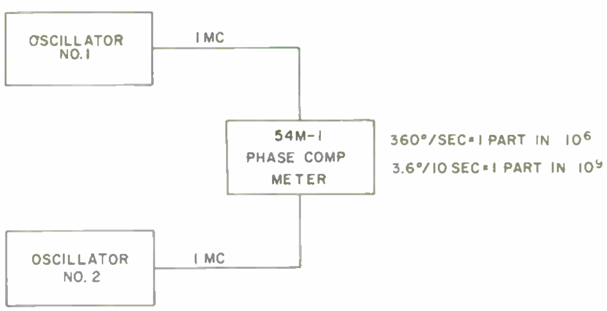


Figure 5-10. Phase Intercomparison System

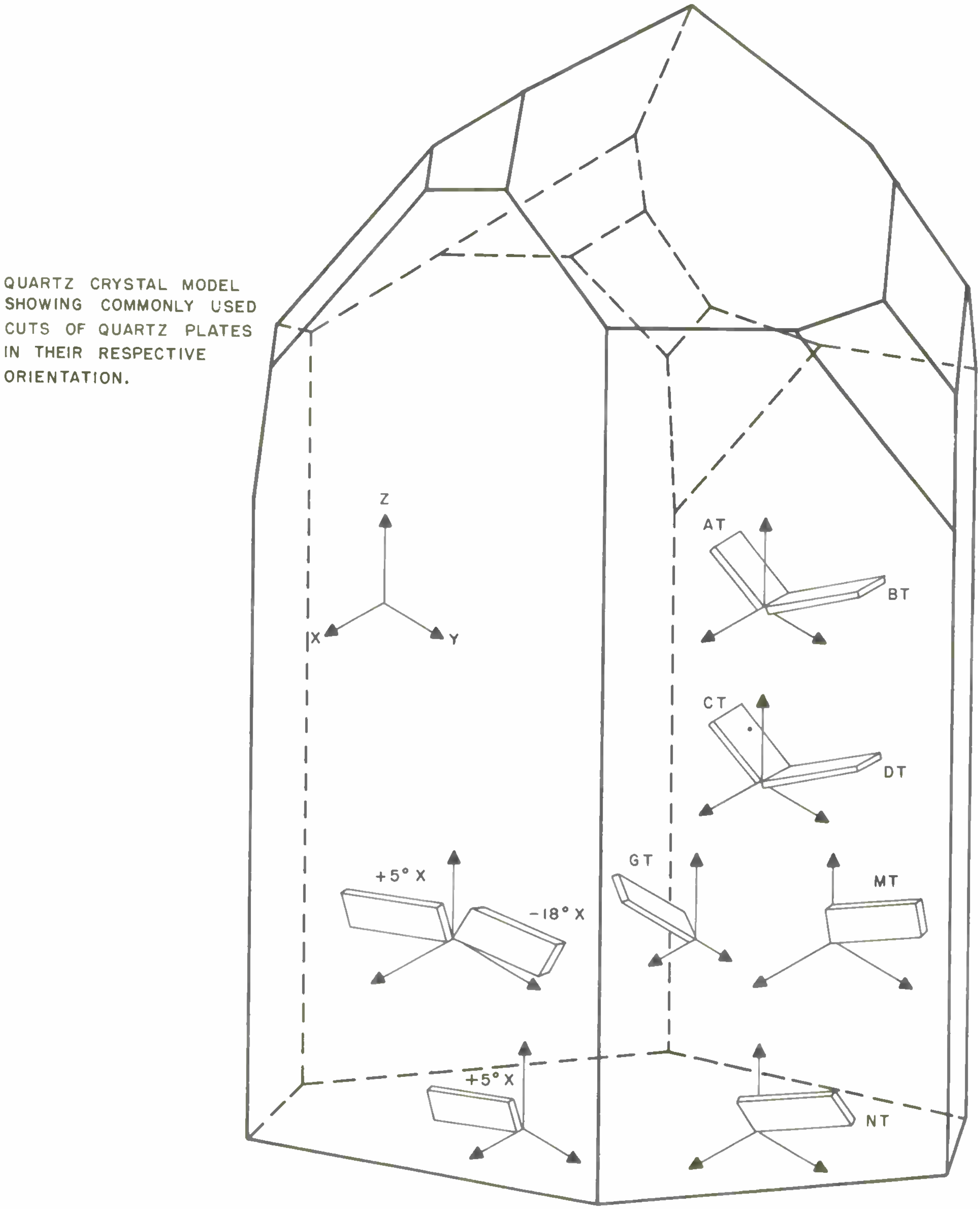


Figure 5-11. Quartz Crystal Showing Types of Cuts

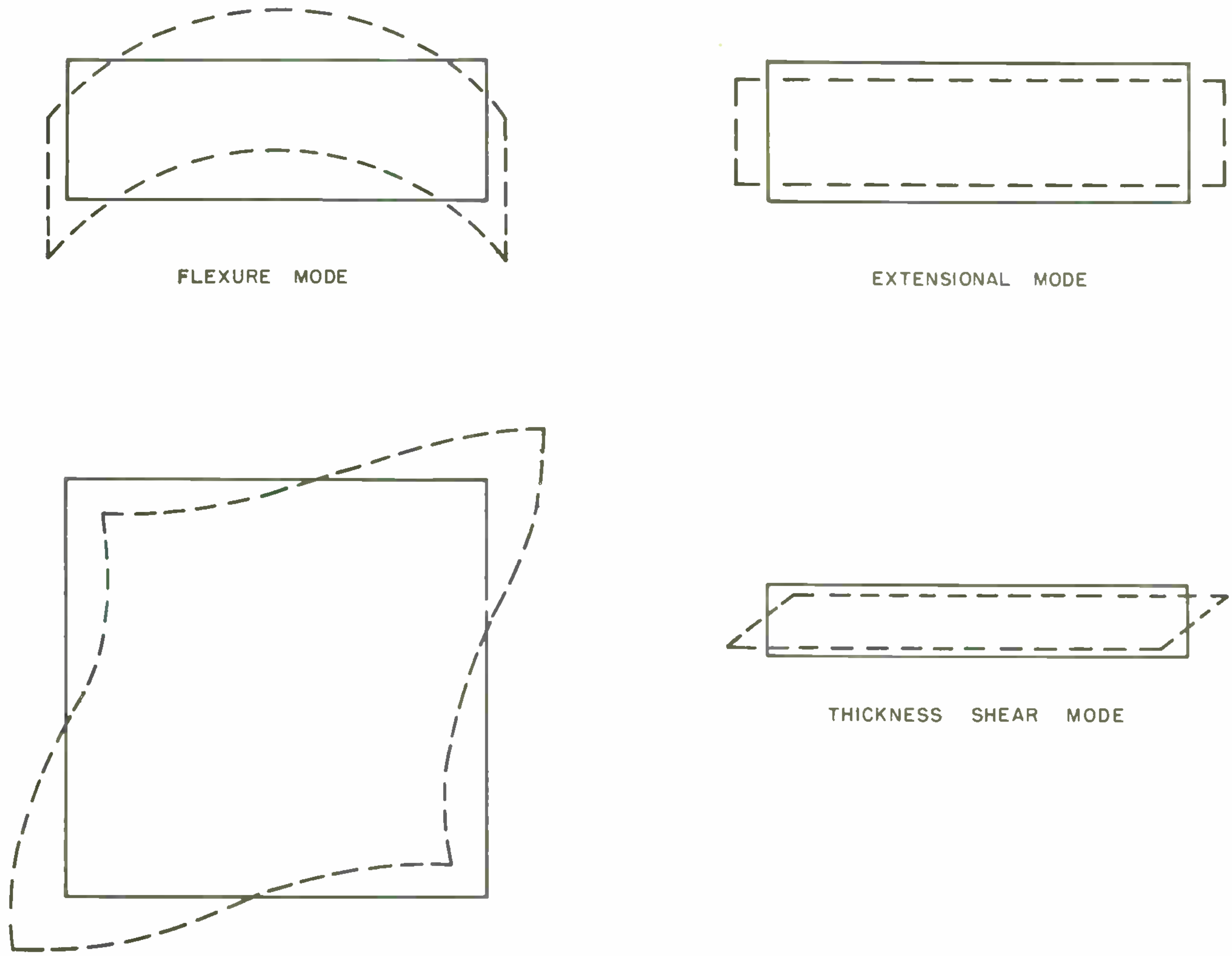


Figure 5-12. Modes of Vibration

or if the oscillators are operating on a nominal frequency of 1 mc, the difference is 1 part in 10^6. If the phase changes only 3.6 degrees in 10 seconds, the difference is only 1 part in 10^9. Thus, small differences in frequency can be measured and recorded. The resulting record will be an indication of the relative short-term stabilities of the oscillators.

4. QUARTZ RESONATOR THEORY

a. CONSTRUCTION AND OPERATION

Quartz resonators are electromechanical devices having extremely high Q's and stable resonant frequencies and are used as resonant circuits in electronic oscillators and filters. Quartz is a piezo electric material, that is, mechanical deformation of the quartz causes an electric charge to appear on certain faces, and conversely, application of voltage across the quartz causes a mechanical deformation. The quartz resonator unit generally consists of a crystalline quartz bar or plate provided with electrodes and suitably mounted in a sealed holder. The mounting structure supports the bar or plate at nodal points in its vibrational pattern so that damping of the mechanical vibrations with resultant degradation of Q is minimized. The bar or plate is cut from the mother crystal at a carefully controlled angle with respect to the crystallographic axes and finished to close dimensional tolerances. The quartz bar or plate has a mechanical resonant frequency determined by its dimensions. This resonant frequency changes with temperature, but by properly orienting the angle at which the blank is cut from the mother crystal, this temperature coefficient can be minimized. Commonly used orientations have been given designations, such as AT, CT, and DT cuts. Figure 5-11 illustrates the relationship between these cuts and the crystallographic axes. Proper orientation of electrodes on the quartz plate provides electric coupling to its mechanical resonance. The electrodes usually consist of a metal plating which is deposited directly on the surface of the quartz plate. Connections from these electrodes to external circuits are usually made through the mounting structure. After the blank has been cut from the mother crystal, it is reduced in thickness by successive stages of lapping until it is within etching range of the specified frequency. After thorough cleaning, the blank is etched to final frequency for pressure-mounting or to the preplating frequency if the electrodes are to be metal plated and the unit wire mounted. After the etching process, the blank is again thoroughly cleaned. After cleaning, the blank is base plated, recleaned, and wire mounted in a clean, moisture free, hermetically sealed holder designed to support the crystal unit against the effects of vibration. After mounting, the metal plated quartz plate is adjusted to the precise final frequency by additional plating. Typical metals used for plating are gold, silver, aluminum, and nickel. Silver is the metal most often used. Quartz crystal units vibrate in different modes depending upon the principal resonant frequency, the three most common modes being flexure, extensional, and shear. In high-frequency precision type units, the shear mode is used. Figure 5-12 illustrates the various modes of vibration.

b. CHARACTERISTICS

Two terminal plated quartz resonators may be represented by the electrical equivalent circuit shown in figure 5-13. The series arm consisting of R_1, L_1, and C_1 represents the motional impedance of the

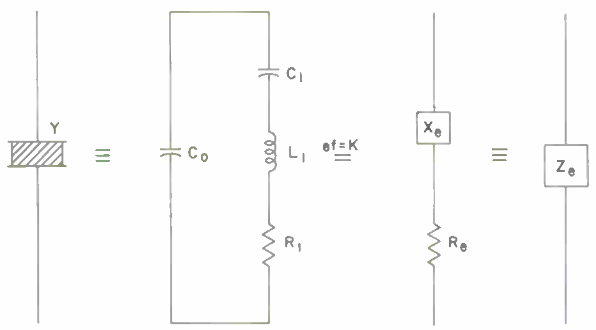


Figure 5-13. Quartz Resonator Equivalent Circuit

quartz plate while C_o represents the electrode capacitance, C_2, plus the holder capacitance, C_h. At a single frequency this can be simplified to an effective reactance, X_e, in series with an effective resistance, R_e. These impedances are a function of frequency as shown in the impedance versus frequency curve illustrated in four views in figure 5-14. The frequency f_s

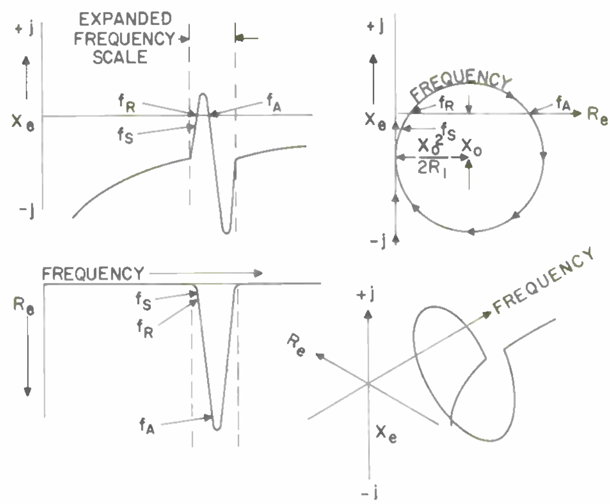


Figure 5-14. Impedance Versus Frequency Curve

is the resonant frequency of the series arm and f_R is the resonant frequency of the quartz resonator unit. The antiresonant frequency of the resonator, f_A, is only a fraction of one per cent higher than f_R. At frequencies removed from f_A by about one per cent, the resonator appears to be a capacitor having a value C_o. Resonators have a number of responses of lesser degree which are usually called unwanted responses. However, certain responses that are approximately harmonically related to the main response are called overtones and are used to control the frequency of vhf oscillators. The equivalent circuit values of a resonator can be controlled to about ±10% except for the series arm resistance, R_1. The resonant frequency can be controlled to close tolerances by close dimension control in the construction of the quartz plate. The resonator performance in a particular application can be calculated if the values of the equivalent circuit are given. If a capacitance C_x is added in series with the resonator and this combination operated at its series resonant frequency f_x, the following formulae hold.

$$X_e = \frac{1}{2\pi f_x C_x}$$

$$R_e = \frac{X_o}{2R_1} - \sqrt{\frac{X_o^4}{4R_1^2} - (X_o + X_x)^2} \quad \text{if } X_o^2 \gg R_1^2$$

$$R_e \approx \left(\frac{C_o + C_x}{C_x}\right)^2 R_1$$

$$R_1 \approx \left(\frac{C_x}{C_o + C_x}\right)^2 R_e$$

$$f_x \approx f_s \left[1 + \frac{C_1}{2(C_o + C_x)}\right]$$

$$\frac{df_x}{f_x} \approx - \frac{C_1}{2(C_o + C_x)^2}\, dC_x$$

The frequency range over which a quartz resonator operates best is determined by the type of cut. Each type of cut has its own optimum frequency range as determined by the physical dimensions of the resonator plate. The following table lists the different cuts and their normal frequency range.

TABLE 5-1

FREQUENCY RANGE OF QUARTZ RESONATORS

Cut	Normal Frequency Range
Fundamental AT	500 kc to 20 mc
3rd Overtone AT	10 mc to 60 mc
5th Overtone AT	30 mc to 80 mc
7th Overtone AT	60 mc to 120 mc
CT	300 kc to 800 kc
DT	200 kc to 500 kc
NT	16 kc to 100 kc
+5°X	90 kc to 300 kc
Bounded +5°X	1.2 kc to 10 kc

Temperature characteristics of quartz resonators are determined mainly by the orientation of the cut with respect to the crystallographic axes. The frequency versus temperature characteristics are shown in figure 5-15. The peaks of the parabolic shaped curves can be moved so as to appear at any desired temperature by changing the orientation of the cut slightly, and the S-shaped curve of the AT cut resonator can be tipped up or down by the same technique. It is seldom possible to adjust a quartz resonator to an exact resonant frequency at a specified temperature. Normal finishing tolerance for commercial units is about ±20 parts in 10^6. However, in precision resonators, finishing tolerances as low as 1 part in 10^6 have been achieved.

The resonant frequency of a resonator and the resistance of the series arm are, to some extent, a function of the amplitude of vibration or the power dissipated in the resonator. Below a current of about 100 microamperes, the frequency and resistance are essentially constant. As the current exceeds this critical value, the series arm resistance, R_1, increases; and the resonator frequency changes as the square of the current. In AT cut elements, the frequency increases about 0.1 part in 10^6 per milliwatt per mc. At still higher values of current, the frequency drifts considerably because of self-heating and

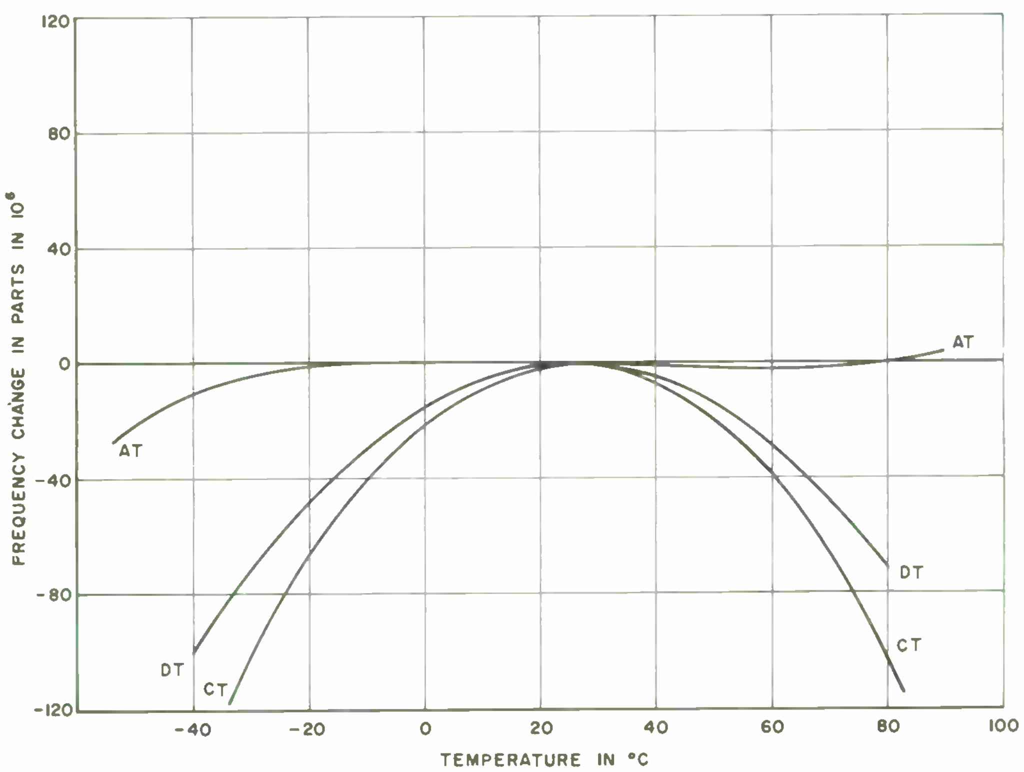


Figure 5-15. Frequency Versus Temperature Characteristics

finally the resonator fractures because of the large amplitude of vibration. Also, coupling of harmonically related modes of vibration can occur because the vibrations are not linear at the higher amplitudes. This coupling degrades the Q of the wanted response, and since these other modes usually have poor temperature coefficients, the Q depends upon both ambient temperature and resonator current. This Q degradation is known as an activity dip. In AT cuts, the unwanted responses within several per cent of the desired frequency are usually higher in frequency than the desired response. These can become prominent enough to control the frequency in oscillator applications. The resonator frequency also changes some with time due to surface contamination of the quartz and to sublimation of the plated electrodes. The actual amount of change depends on the cut, design, cleanness, and construction of the resonator unit. The rate of aging generally increases rapidly with temperature and is sometimes 100 times greater at 40°C than at 0°C. Therefore, aging is more rapid in oven controlled units. At present, the aging in commercial high-frequency AT cut crystal units is about 40 parts in 10^6 per year at 85°C. However, in precision, oven controlled resonators aging rates as low as 1 part in 10^9 per month have been achieved. Normal aging rates for precision units are 1 part in 10^8 per day. Recent studies on the aging rate of quartz resonators indicate that their stability is improved by very low temperature operation. Figure 5-15.1 shows that stability on the order of 1 part in 10^{10} per day can be achieved by operating commercial grade crystals at 4°K.

c. CONSTRUCTION OF PRECISION RESONATORS

Figure 5-16 shows the construction and mode of vibration of a precision crystal resonator, 5th overtone AT cut. The blank is made circular with one spherical surface and one flat surface. In a crystal of this shape, all of the mechanical vibration takes place near the center of the plate and the edges remain dormant. Thus, supports can be attached to the edges of the plate without degrading Q through damping of the vibrations. The quartz plate is usually given a high polish

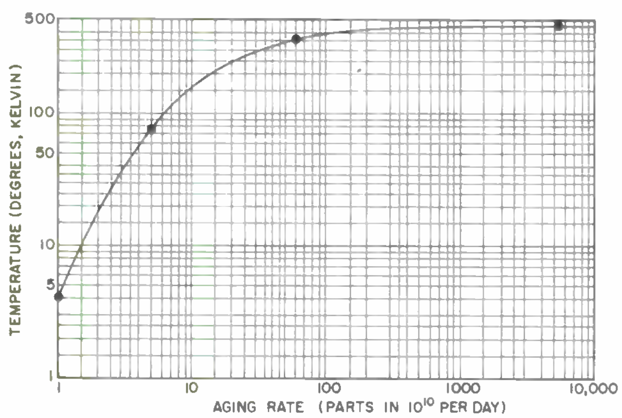


Figure 5-15.1. Quartz Resonators, Aging Rate versus Temperature

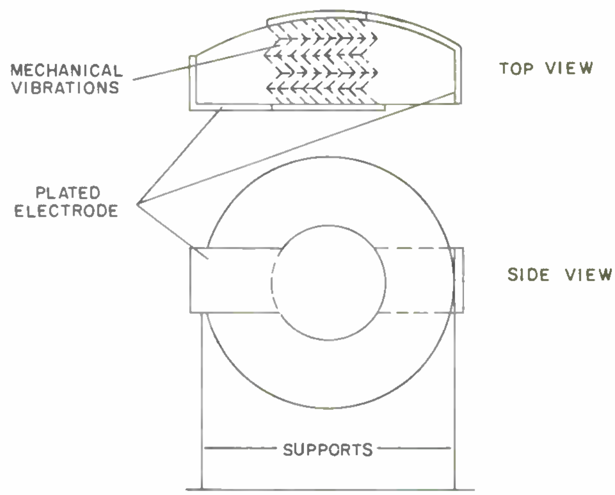


Figure 5-16. Precision Quartz Resonator, Construction and Mode of Vibration of 5th Overtone AT Cut

which may be followed by a brief etching operation before the electrodes are plated on. The plating operation is performed in a vacuum in order to minimize contamination and after plating, the unit is sealed in an evacuated glass or metal envelope. The mode of vibration used is the 5th overtone in thickness shear. Use of this mode of vibration greatly decreases the volume to effective surface ratio and at the same time reduces the effective surface area exposed to contamination since ten effective surfaces, consisting of the five interfaces resulting from 5th mode operation, are inside the crystal. Typical applications for these units are in 2.5 mc and 5 mc frequency standards. The resonant frequency of 5th overtone AT cut crystals is not affected by shock and vibration below the level that permanently damages the mounting structure.

5. OSCILLATOR THEORY

a. GENERAL THEORY

Oscillator operation can be analyzed on a feedback basis wherein the oscillator consists of an amplifier with a frequency selective device which couples energy at the desired frequency from the output back to the input. When the circuit is adjusted so that the amplifier supplies energy at the desired frequency sufficient to overcome the losses in the feedback path, the circuit oscillates and generates a signal at a frequency controlled by the resonant frequency of the feedback path. If energy is to be coupled out of the oscillator and used to drive other devices, the amplifier must supply this energy in addition to that required to overcome the losses in the feedback path.

In crystal-controlled oscillators a quartz resonator network provides the coupling from the output of the amplifier to its input. Because of its high Q, the resonator operates as a highly selective feedback network with extremely high attenuation of frequencies on either side of its resonant frequency. Thus the frequency of oscillation cannot deviate appreciably from the resonator frequency. Since the output of the feedback network is the input of the amplifier, the total phase shift around the loop must be zero. For this reason the resonator must compensate for phase shifts in the rest of the oscillator circuit and these phase shifts will affect the frequency stability of the circuit.

Another method of analysis is that based on the negative resistance theory. Figure 5-17a is an

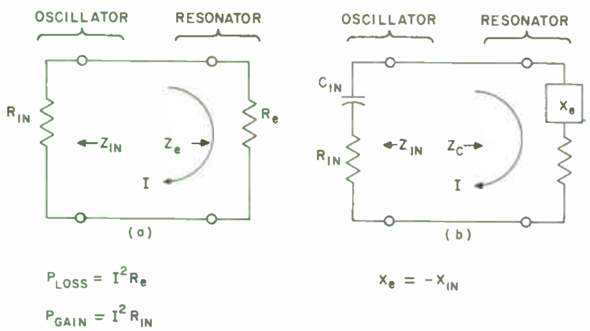


Figure 5-17. Equivalent Circuit of a Crystal-Controlled Oscillator

equivalent circuit of an oscillator operating at series resonance. The input impedance of the oscillator is a negative resistance, R_{in}, and the resonator has an effective resistance, R_e. The power loss in the resonator is I^2R_e, and the power supplied by the oscillator is I^2R_{in}. If the power gain is greater than the power loss, oscillations will build up; and if the power gain is less than the power loss, oscillations will die out. The negative oscillator input resistance is a function of the current I so that R_{in} will decrease as oscillations build up, until R_{in} equals R_e and a stable amplitude is reached. A more general case is illustrated in figure 5-17b in which the oscillator input impedance has a capacitive component. It is standard practice to make this input capacitance, C_{in}, 32 uuf at frequencies above 500 kc and 20 uuf at frequencies below 500 kc. The entire network oscillates at a frequency such that $X_e = -X_{in}$, and the equations for power loss and gain given for figure 5-17a still apply. All types of oscillators can be analyzed in this manner except that in a few special cases the reactive component of oscillator input impedance may be inductive.

Mathematical analysis consists of replacing the resonator with an imaginary test voltage generator and solving for $Z_{in} = \frac{E_{in}}{I_{in}} = R_{in} + jX_{in}$. Figure 5-18 illustrates a method of calculating the power dissipation in

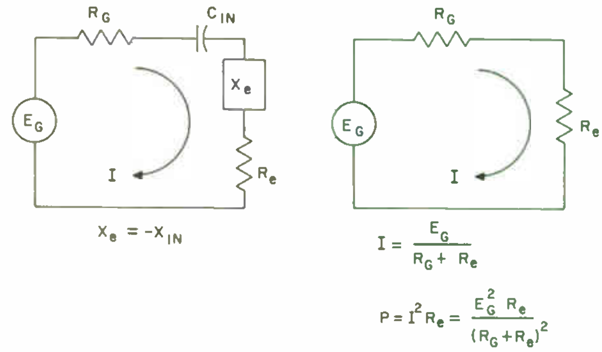


Figure 5-18. Calculation of Resonator Power Dissipation

the resonator for vacuum-tube saturation limiting. In the first equivalent circuit the oscillator is represented by a generator, E_G, with a series generator resistance, R_G. Since the resonator reactance, X_e, equals the negative reactance, $-X_{in}$, of the input capacitance, C_{in}, these two reactances cancel, and the equivalent circuit is reduced to the second circuit shown in figure 5-18. After the grid voltage on the oscillator tube reaches the value that saturates the tube, the generator voltage E_G remains relatively constant and independent of grid drive. Then resonator current is given by $I = \frac{E_G}{R_G + R_e}$, and the power dissipated in the resonator is given by $P = I^2R_e = \frac{E_G{}^2R_e}{(R_G + R_e)^2}$.

b. TYPICAL OSCILLATOR CIRCUITS

Quartz crystal resonators have two resonant frequencies. At one frequency they exhibit antiresonant characteristics, and at a slightly lower frequency they exhibit series resonant characteristics. At frequencies between these two the resonator reactance is inductive, and at frequencies outside this range the reactance is capacitive. The design of the oscillator circuit determines in which part of the reactance characteristic it will be used. In the oscillator represented by figure 5-19, the series resonant response is used. The amplifier is designed so that the total phase shift from amplifier input to output is zero. Since the total phase shift around the complete loop, including the feedback network, must be zero, the feedback network must also have zero phase shift. If the feedback network is to have zero phase shift, it

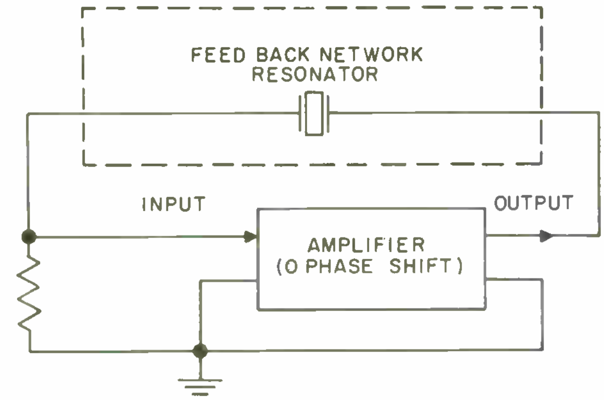


Figure 5-19. Basic Crystal Oscillator Operating the Resonator at Series Resonance

must be resistive at the frequency of oscillation. The quartz resonator which forms the feedback network is resistive at two frequencies, its antiresonant frequency and its series resonant frequency. Since the resonator is in series with the feedback path, the frequency at which it offers the least resistance to the signal is its series resonant frequency, and this will be the frequency of oscillation.

Figure 5-20 shows the basic Pierce oscillator circuit, an equivalent circuit, and a vector diagram showing the phase relationships, neglecting circuit losses. In this oscillator the feedback network operates at antiresonance, but the resonator operates at a point between its series resonant frequency and its antiresonant frequency where it is sufficiently inductive to resonate with C_p and C_g in series. The generator voltage $-uE_g$ is the grid voltage multiplied by the gain of the tube. Since the circuit representing

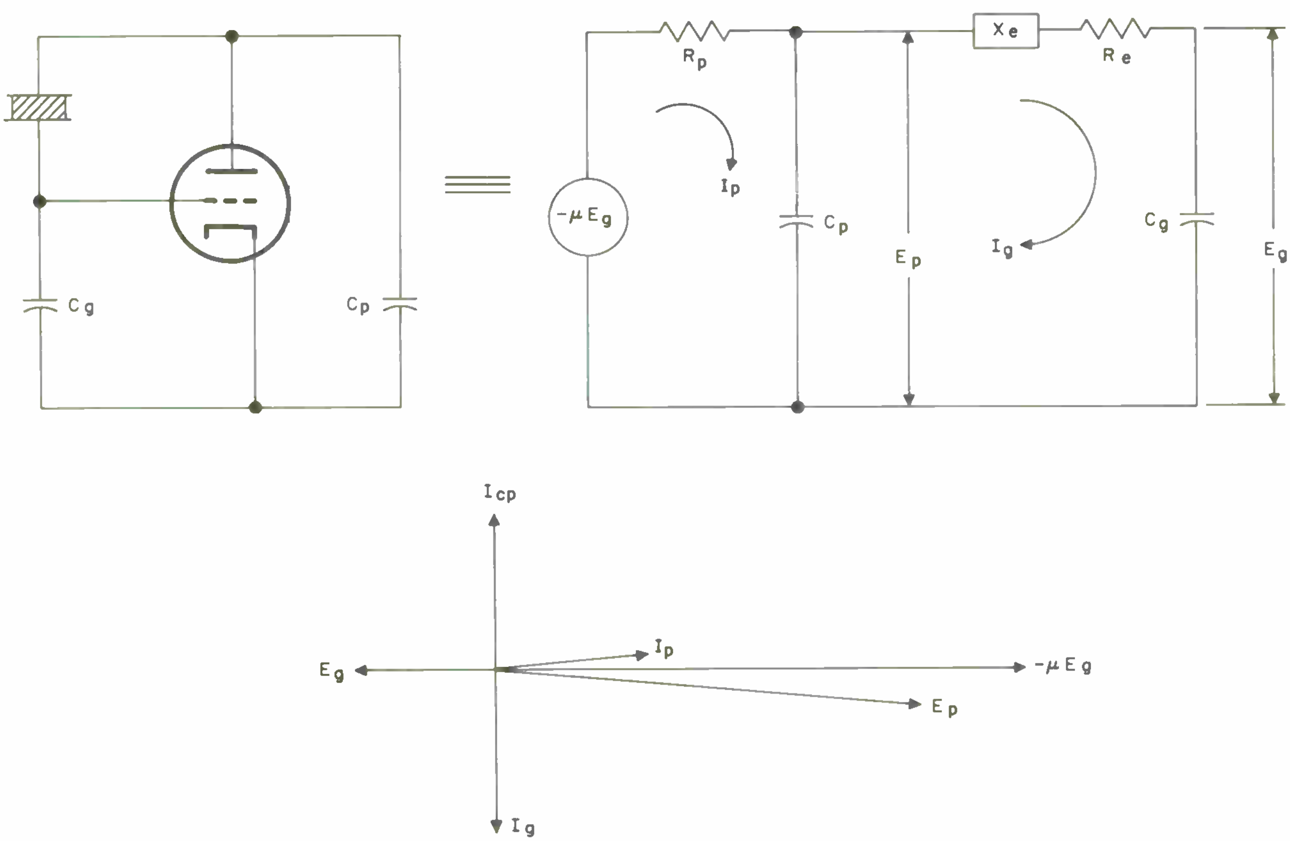


Figure 5-20. Basic Pierce Oscillator

the generator load is resonant, I_p will be in phase with $-uE_g$; and since the branch consisting of X_e, R_e, and C_g is inductive, I_g will lag $-uE_g$ by 90°. The voltage, E_g, developed across C_g will lag I_g by 90°. Therefore, E_g will lag $-uE_g$ by 180°, and since the tube introduces another 180° phase shift, the condition that there be zero phase shift around the loop is satisfied.

c. PRECISION CRYSTAL-CONTROLLED OSCILLATORS

In the design of precision oscillators, several precautions must be taken to minimize instabilities. The construction of the quartz resonator itself was described in paragraph 4 of this chapter. Additional precautions to be observed in the use of quartz resonators in precision oscillators are listed below.

(1) The components that make up the resonant circuit must be placed in a controlled environment.

(2) The amplitude of oscillation must be controlled to avoid instabilities caused by nonlinearity in the vibration pattern of the resonator at high amplitudes.

(3) Phase instabilities in the active amplifying portion of the oscillator must be held to a minimum.

(4) External circuitry must be isolated from the oscillator so that reactive components are not reflected back into the resonant circuit to cause instability.

(5) The Q of the resonator should be high so that loop phase shifts can be compensated for with minimum change in resonator operating frequency. In addition, high Q makes possible low coupling between the resonator and the active amplifying portion of the oscillator, thus minimizing the effect of the active network on the resonator.

(6) Nonlinearities in the active amplifying portion of the oscillator cause harmonic distortion. Adjacent harmonics are mixed together in the same or other nonlinear portion of the circuit after having passed around the feedback network. The fundamental frequency component thus produced is usually not phase stable and causes phase instability in the oscillator. Therefore, the amplifier must be operated on the linear portion of its characteristic.

Figure 5-21 illustrates the principle of operation of the Meacham oscillator. The resonator, Y1, operates at its series resonant frequency and thus offers a low resistance and zero phase shift to the frequency of oscillation. The opposite leg of the bridge circuit is an incandescent lamp which when cold also has a low resistance. Resistors R1 and R2 have about the same resistance as the effective resistance of the resonator at its series resonant frequency. This condition exists when oscillations start. The coupling between the amplifier and the feedback loop is relatively tight, and there is a large amount of positive feedback. As oscillations build up, the

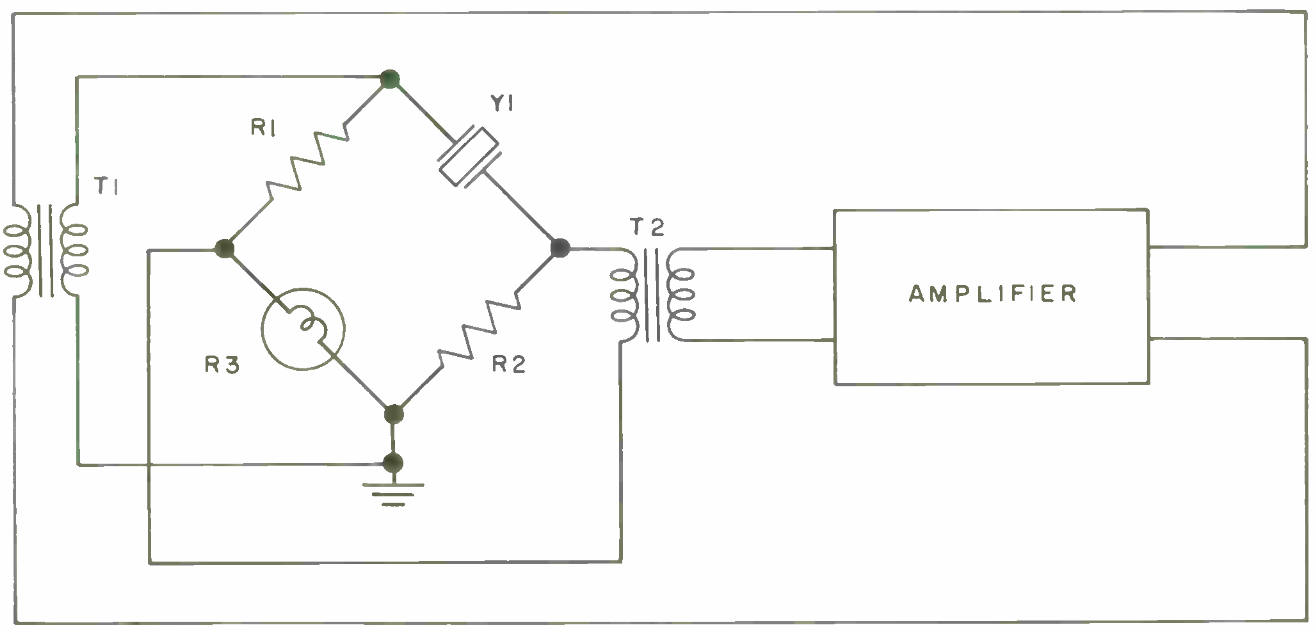


Figure 5-21. Meacham Oscillator

lamp, R3, is heated by the r-f current, and its resistance increases until it is almost equal to that of R1. As the lamp resistance increases, the bridge approaches balance; the positive feedback is reduced until the bridge is almost in balance, and the residual positive feedback is just sufficient to sustain oscillation. This oscillation is usually used only at frequencies below 1 mc because of the difficulties of obtaining transformers that do not cause phase instabilities at the higher frequencies.

Figure 5-22 is a schematic diagram of a typical Pierce oscillator used in high precision frequency standards. The 1 mc quartz resonator, V1, is a fundamental AT cut crystal, sealed in an evacuated glass envelope. Its temperature coefficient is only several parts in 10^7 per degree centigrade. The components that make up the resonant circuit, Y1, R1, C1, C2, and C3 are housed in an oven in which the temperature is held constant to better than .01°C. Capacitor C1 and resistor R1 hold the d-c voltage impressed on the resonator, Y1, to a minimum. The resonator has a minimum Q of 1 million, thus capacitors C2 and C3 can be made large to bypass effectively the plate and grid of the tube to ground and reduce the coupling between the resonator and the active portion of the circuit to a low value. In addition, all frequency controlling components are isolated from other circuit elements by shielding. Capacitor C4 is a precision variable capacitor which provides a small range of adjustment of the resonant frequency of the circuit. The total range of adjustment is about 4 cps at the nominal operating frequency of 1 mc. The output of the oscillator is coupled to an untuned buffer stage which isolates the oscillator from succeeding stages. Two stages of amplification follow the buffer stage and provide additional isolation. The amplitude of oscillation is controlled by negative voltage developed in the grid circuit of the last amplifier stage. Cathode bias on this tube delays development of negative voltage until the signal applied to the grid reaches a predetermined level. When this level is reached, the resultant negative voltage couples to the grid of the oscillator tube through an RC filter increasing grid bias, and thus reducing the tube gain, and limiting the amplitude of oscillation to a low level. This automatic amplitude control system holds the operating power level in the quartz resonator to less than .1 microwatt. This type of oscillator has attained short-term stability of better than 1 part in 10^{10} and long-term stability of better than 1 part in 10^9 per day. Oscillators are now being designed to use the 5th overtone AT cut crystal. These are expected to have even better stabilities than oscillators using the fundamental AT cut crystal.

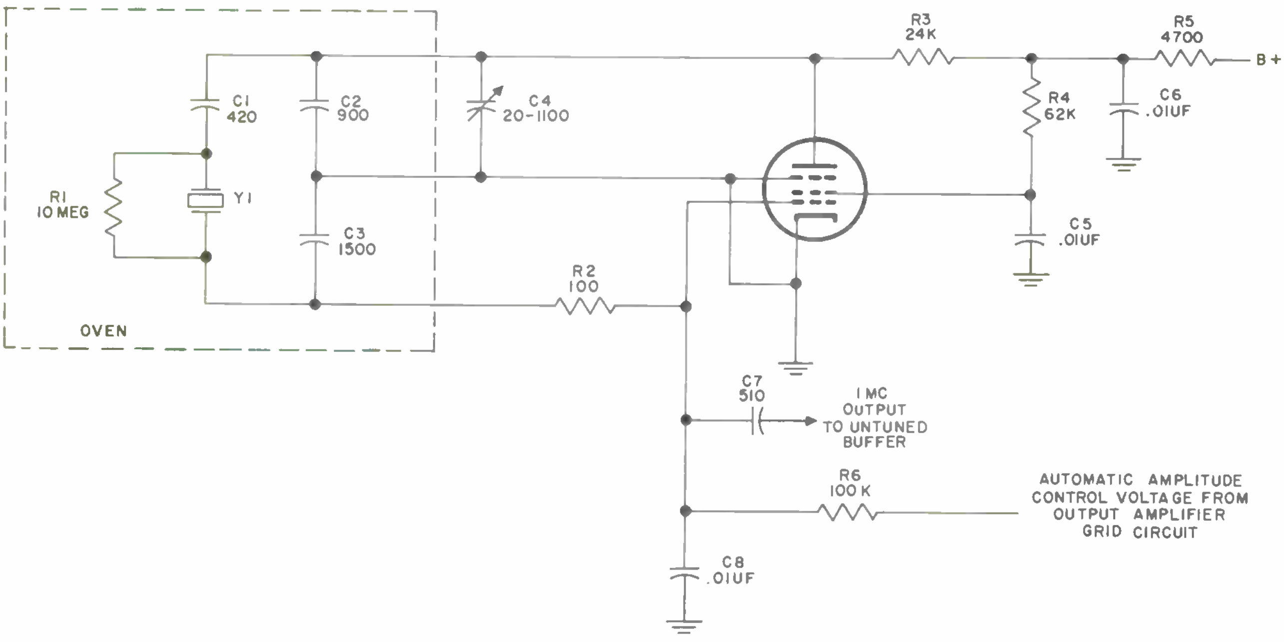


Figure 5-22. Precision Pierce Oscillator

6. OVEN THEORY

a. GENERAL THEORY

Since all quartz resonators have some variation of frequency with temperature, the resonator must be kept at a constant temperature in order to achieve maximum stability. In most frequency standards, this is accomplished by placing the resonator in an oven and then maintaining the oven temperature at a level somewhat higher than the ambient temperature surrounding it. The six items listed below make up a typical oven.

(1) The resonator or device to be temperature controlled

(2) The oven heater

(3) A device for controlling the power delivered to the heater

(4) A temperature sensing element

(5) A heat sink (ambient temperature around oven)

(6) Thermal insulation or thermal resistance

The operation of the oven can be compared to the operation of an electrical bridge circuit as illustrated in figure 5-23. The arms of the bridge R1, R2, R3,

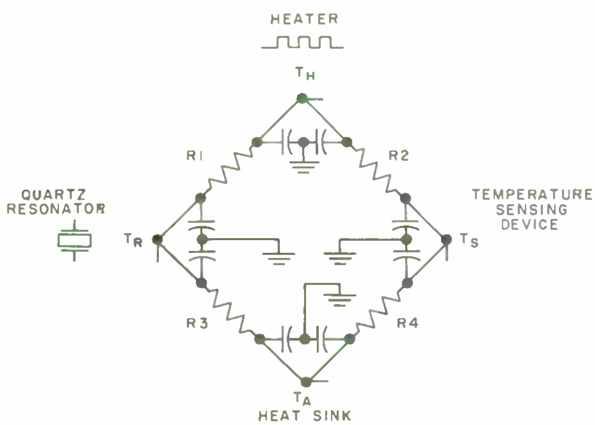


Figure 5-23. Oven Operation Equivalent Electrical Circuit

and R4 represent thermal resistance, that is, resistance to heat flow. The temperatures T_H, T_R, T_A, and T_S are analogous to electrical potentials at the points indicated. T_H is the heater temperature; T_R is the resonator temperature; T_A is the temperature surrounding the oven, and T_S is the temperature of the sensing element. The heat storage or thermal capacities of the materials in the heat flow path are analogous to electrical capacitance. The temperature of the heater T_H, is regulated by the sensing element through a servo system so that the temperature T_S remains constant. To maintain the temperature T_R of the resonator constant regardless of variations in T_A, T_R must equal T_S, that is, the bridge must be balanced or R1/R3 must equal R2/R4. Conditions for balance are less critical if R1 and R2 are made very small as compared to R3 and R4, since then T_H, T_S, and T_R will be more nearly equal. The time lag between T_H and T_S, caused by the time constant of the thermal resistance R2 and its associated thermal capacities, causes the servo system to hunt and this in turn causes the temperature T_S to cycle. Reducing the time constant of R2 and its associated capacities to a very low value eliminates this cause of hunting. However, another cause of hunting is the operating differential of thermostats. When these are used as temperature sensing devices, the time constant of R1 and its associated capacities must be made long in order to filter out variations in T_R caused by hunting in the servo system. If proportional control is used, the time lag due to operating differential in the sensing element is eliminated, and the time constant of R2 and its associated capacities can be made very low to eliminate hunting. If the servo system is free of hunting, then the time constant of R1 and its associated capacities can be made low, and if at the same time the time constants in the R3 leg and the R4 leg are made long, T_S and T_R will be on an isothermal line with T_H. Ovens using this system can maintain temperature within .01°C.

b. TYPICAL PRECISION OVEN

Figure 5-24 illustrates the construction of a typical precision oven. In this oven, the resonator, the heater, and the temperature sensing element are all in an isothermal space. The resonator in its sealed envelope is housed in an aluminum cylinder upon which the heater is wound. Because of the high heat conductivity of aluminum and because the resonator is almost completely surrounded by aluminum, the temperature of the resonator is nearly identical to that of the aluminum enclosure. The heater is wound on this enclosure and tightly coupled to it thermally. Thus the resistance R1 in figure 5-23 and the thermal time constant between T_H and T_R are nearly zero. The heater is constructed so that it is also the temperature sensing element, making R2 in figure 5-23 and the time constant between T_H and T_S nearly zero. The resistances R3 and R4 are made very large by housing this assembly in a vacuum bottle. The vacuum bottle is enclosed in a second aluminum cylinder which makes T_A uniform on all sides of the oven.

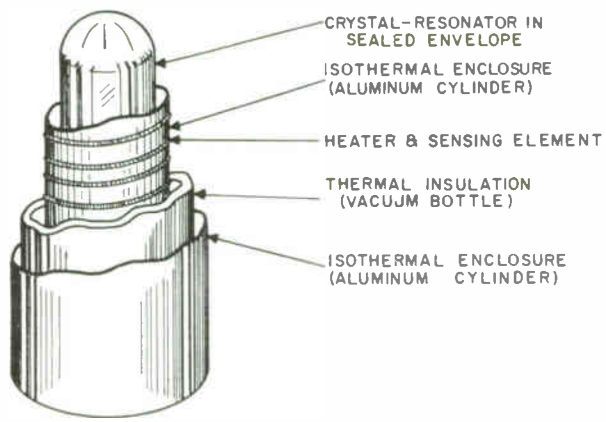


Figure 5-24. Typical Oven Construction

Since the heater, the resonator, and the temperature sensing element have been placed in an isothermal space well isolated from the ambient temperature, the only remaining requirement is to maintain the heater at a constant temperature. The circuit of figure 5-25 satisfies this requirement. The bridge circuit, HR601, performs two functions. It is the heating element for the oven and the control element for the oven oscillator. Two arms of the bridge are made of nickel wire, and the other two arms are made of Low Ohm wire. The arms are of selected lengths so that their resistances at the desired oven temperature are almost equal. When the oven temperature is low, the nickel wire has less resistance than the Low Ohm wire, and terminals 5 and 7 of the secondary winding of T601 see less resistance to ground than do terminals 4 and 6. At the same time, terminals 4 and 6 of T601 see less resistance to the feedback path than do terminals 5 and 7. As a result, the alternating current flowing through the bridge is applied as positive feedback to the first amplifier stage. Under these conditions, the circuit oscillates at an amplitude determined by the amount of bridge unbalance which in turn is controlled by the temperature Thus, proportional control is provided, and the thermal lag inherent in thermostatic devices is eliminated. The power supplied to HR601 by the oscillator heats the oven; and as the temperature approaches the desired level, the bridge approaches balance reducing the amount of feedback until at the desired temperature, it is just sufficient to sustain oscillation. When the oven reaches this steady state condition, the oven control oscillator supplies just enough power to the heater to replace the heat lost to the surrounding medium and maintains the oven temperature constant within .01°C. If for any reason the temperature of the oven rises above the desired level, the bridge becomes unbalanced in the opposite direction and resultant negative feedback prevents oscillation.

7. FREQUENCY DIVIDER THEORY

In order to obtain maximum stabilities, standard signals must be generated at higher frequencies than the lowest frequency required for use in the equipment. In order to obtain the lower frequencies, frequency dividers must be used, and if the divided frequency is to have the same stability as the original, these dividers must be under control of the standard. Figure 5-26 is a block diagram of a typical divider. The equipment contains two regenerative dividers which divide their input frequencies by 10. With a 1 mc input, this circuit provides outputs at 1 mc, 100 kc, and 10 kc. The principles of operation of the two dividers are identical except for the frequencies involved; therefore, only the 1 mc to 100 kc divider will be discussed. When the 1 mc signal supplied to

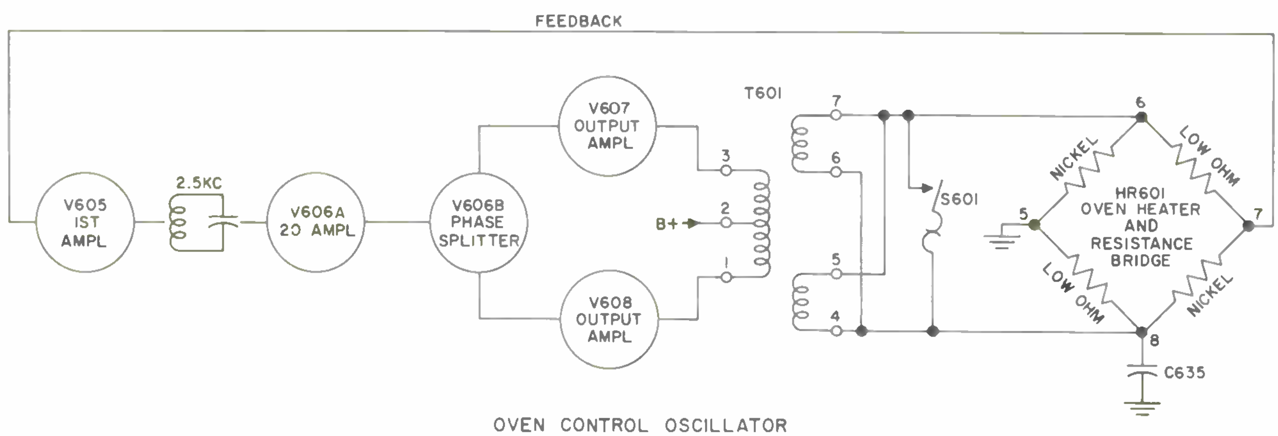


Figure 5-25. Oven Control Oscillator

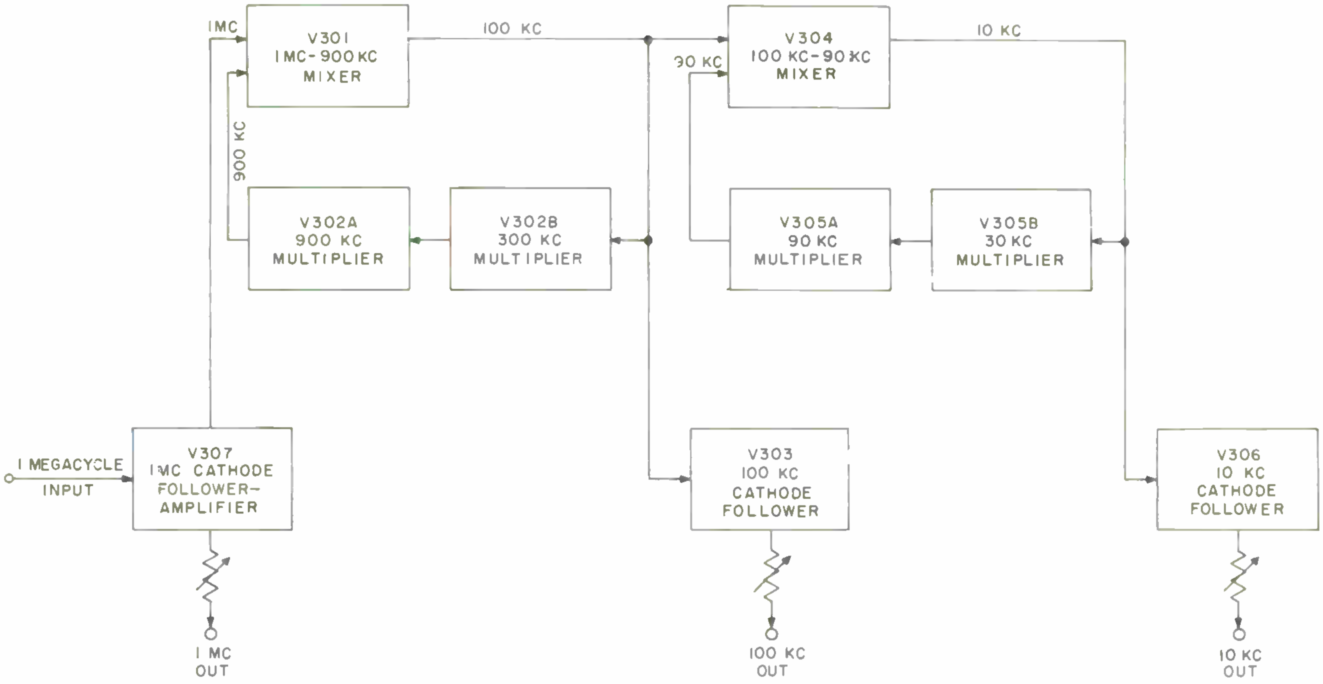


Figure 5-26. 8U-1 Frequency Divider, Block Diagram

the injection grid of mixer V301 is large enough to make the circuit sufficiently regenerative, noise energy at 900 kc appearing at the signal grid of V301 mixes with the 1 mc signal to produce sufficient 100 kc signal to drive multiplier V302B. This circuit multiplies the 100 kc signal by 3 producing a 300 kc signal which drives a second multiplier V302A. The 300 kc signal is again multiplied by 3 to produce a 900 kc signal for mixing in V301. The 100 kc signal thus produced is under complete control of the 1 mc signal. If the 1 mc injection falls below the threshold level, the loop gain of the circuit falls below the level required to maintain operation, and no output is available. The second divider circuit operates from the 100 kc signal in the same way to produce a 10 kc signal. The cathode followers used to couple the signals to external circuits provide isolation.

8. SYSTEM CONSIDERATIONS

In single-sideband communications, the total frequency shift in the system, both transmitting and receiving, should not exceed 50 cps. At 20 mc this requires a system stability of 2.5 parts in 10^6, including errors introduced by the propagating medium, Doppler shifts, and errors in terminal equipment. To assure total system stability of 2.5 parts in 10^6 over a period of months without readjustment, stabilities of 1 part in 10^8 per day are required in the frequency standards. Frequency, time, and phase stability requirements in other systems, such as Kineplex*, are even more severe. In the Kineplex system, 22 millisecond pulses are transmitted. Each pulse is a reference for the succeeding pulse. Short-term phase stability for this system must be within a few degrees over a 44 millisecond period. Frequency accuracy must be ±.5 cps to prevent deterioration of the signal. Errors of ±1 cps cause noticeable distortion and ±3 cps is the practical limit of permissible frequency error. Time accuracy within ±1 millisecond would provide a signal with no noticeable deterioration, but ±5 milliseconds is the practical limit of time error. Thus, in these systems, total system frequency stabilities of 6 parts in 10^7 and total system time stabilities of 1 part in 10^8 are required. In mobile systems, Doppler shifts, and time variations due to changing transmission path lengths must be compensated for either by automatic correction circuits or by manual readjustment of local equipment.

* Registered in U. S. Patent Office

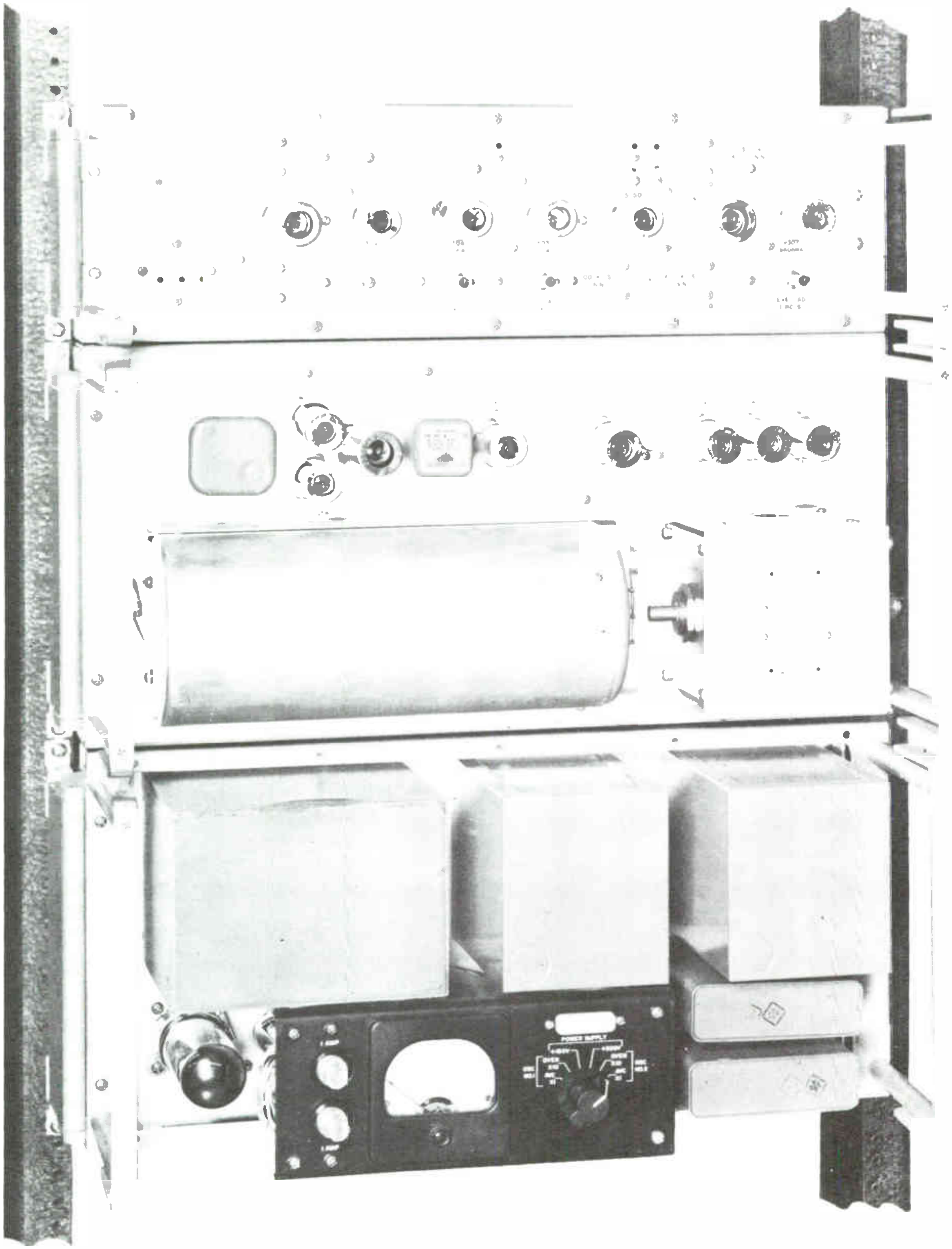


Figure 5-27. Typical Secondary Frequency Standard, Covers Removed

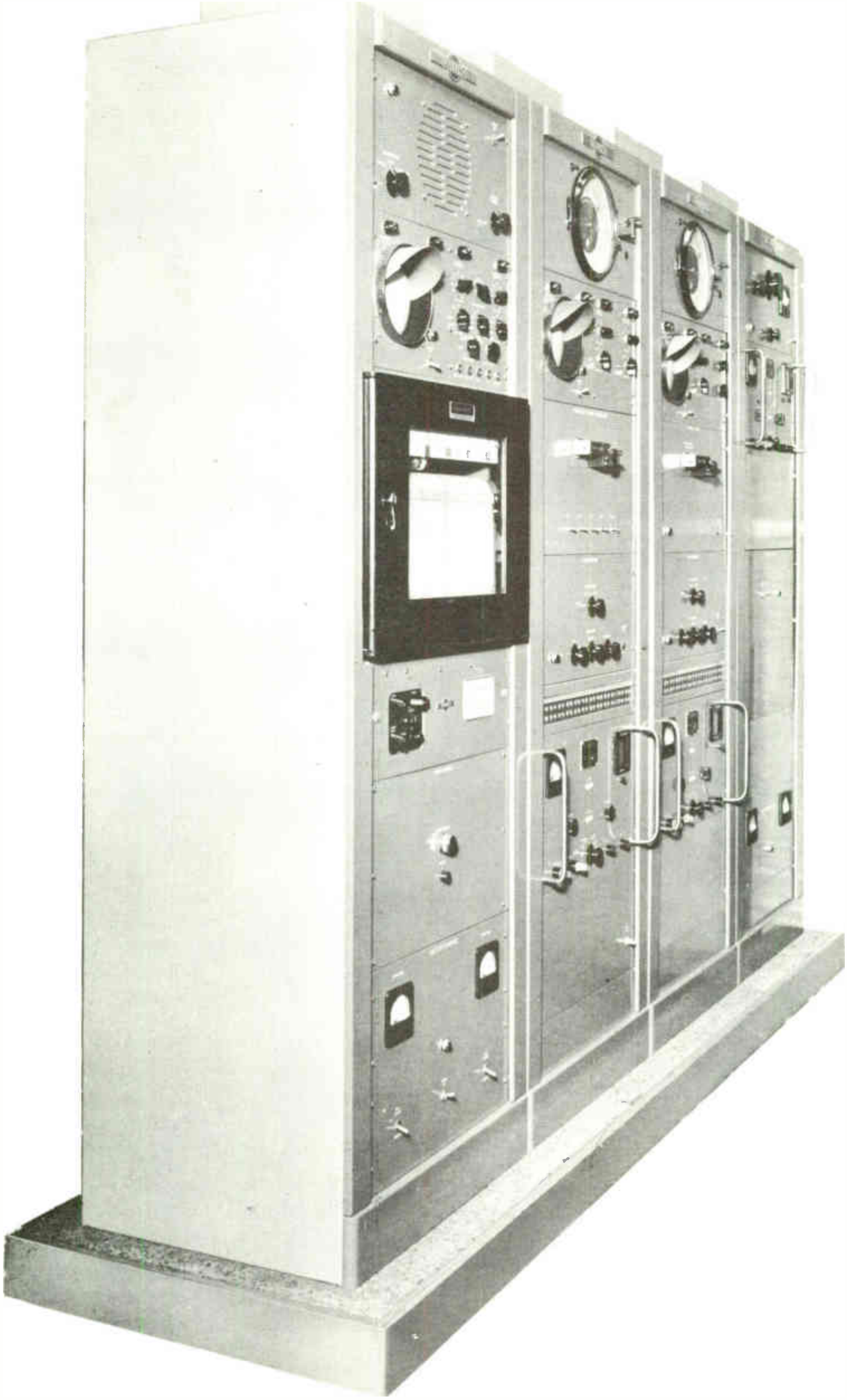


Figure 5-28. Collins Radio Company Primary Frequency Standard

CHAPTER 6
PRINCIPLES OF SERVOMECHANISMS

1. DEFINITIONS

A servomechanism is most commonly defined as a feedback control system of which at least one element is mechanical in nature. Voltage regulators for power supplies, automatic volume control and automatic frequency control circuits used in radio equipment and thermostats used to regulate temperature in home heating equipment and in various electrical appliances are examples of feedback control systems. Power steering, gun turret positioning devices, and airplane autopilots are examples of servomechanisms.

In all feedback control systems, the quantity to be controlled is measured in some manner. This measured value is compared to a desired, or reference, value to form an error signal, and the controlling action is governed by some function of the error signal. Feedback control devices may contain electrical, mechanical, pneumatic, hydraulic, and other types of elements. Frequently, a human operator is included in the feedback loop.

2. A TYPICAL POSITION SERVO SYSTEM

Figure 6-1 illustrates a typical position follow-up type of servomechanism. This type of device can be used for repeating the position of a shaft at a remote point. For example, in an airborne radio transmitter, it could make the shaft of a precision oscillator follow a dial in the pilot's control box. Follow-up servos are also used to repeat the shaft position of a delicate instrument at a shaft where a large amount of torque is needed. In this case, the purpose of the servo is to provide torque amplification.

In figure 6-1 the reference input R is a shaft position. A voltage proportional to the shaft position is obtained from a linear potentiometer connected across a battery. This voltage is mixed with a voltage proportional to the controlled variable to form the error signal E which is then amplified and applied to the winding of a motor. The motor shaft is coupled through a gear train to the load, which in most cases is a friction device, although it sometimes contains a

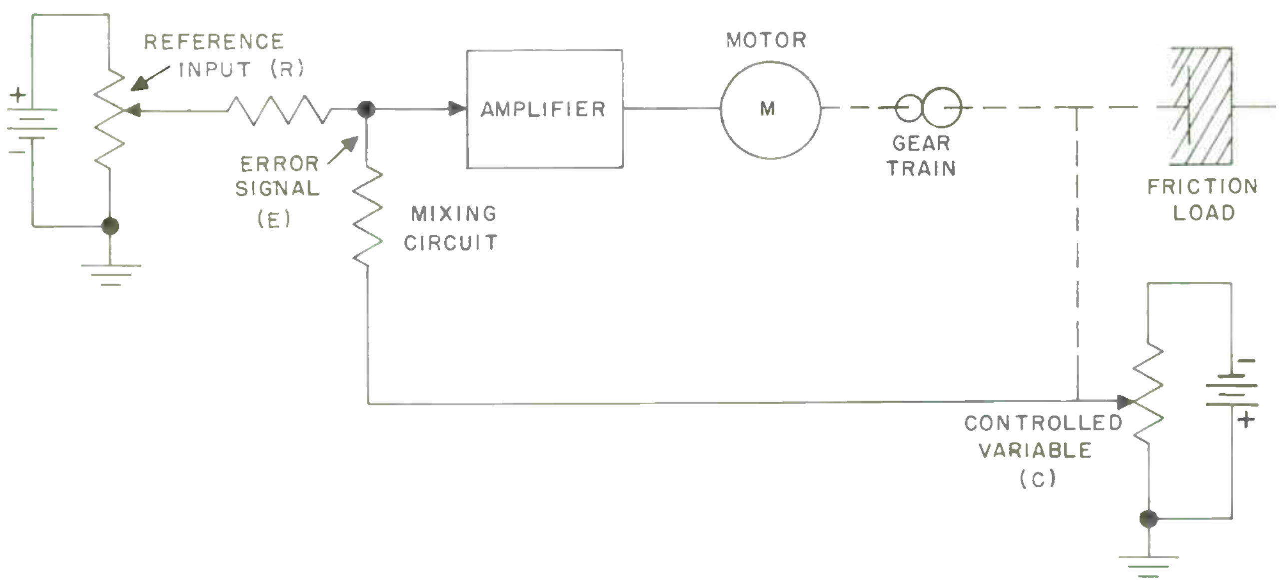


Figure 6-1. Position Follow-up Servomechanism

significant amount of inertia. Coupled to the load shaft is another potentiometer which produces a voltage C, proportional to the position of the output shaft. The mixing circuit subtracts the controlled voltage C from the reference voltage R to obtain error signal E. The amplifier drives the motor in such a way that if E is positive the motor turns in one direction, and if E is negative the motor reverses. When E is zero, the motor stops. Below the saturation level of the motor and amplifier, the motor speed is proportional to the error signal. A device of this kind is called a proportional controller.

Suppose that when power is applied to the circuit, the reference input R is greater than the controlled variable C so that E is positive. If the motor is connected so that a positive E causes the motor to turn in such a direction as to increase C, then as the motor turns, E will decrease, and the motor will slow down until R and C become equal. Then E is zero, and the motor stops. If the battery voltages feeding the reference input and the control variable potentiometers are equal, and the electrical angles of the two potentiometers are equal, this null condition will occur only when the load shaft is at the same angular position as the reference input shaft. If some external force, such as vibration, displaces the output shaft so as to increase C, E will become negative, and the motor will apply torque to the load in a direction to decrease C. This torque is proportional to the displacement, and the result is similar to the effect of a spring. In a feedback device of the type described above, the motor will run as long as there is an error signal sufficient to overcome the load friction. Consequently, the amount of residual error in this type of servo depends only on the amount of friction in the load, and not upon the value of c.

In contrast to this type of behavior is a class of feedback control devices typified by the voltage regulator circuit of figure 6-2. In this case, the controller is a tube whose plate to cathode resistance varies in proportion to its grid voltage. The feedback signal C, obtained from a voltage divider across the regulated output, is compared in a mixer circuit with a reference voltage obtained from a voltage reference tube. The resulting error signal E is amplified and applied to the grid of the controller. If the supply voltage suddenly drops, C will be reduced and E increased. The resulting change in the grid voltage of the controller will reduce its effective resistance and raise the output voltage. However, the output voltage cannot become quite as high as it was before the supply voltage dropped, for if it did, E would return to its original value, and the resistance of the controller would become the same as it was before the decrease in the supply voltage. This condition could be corrected by replacing the controller with a motor driven rheostat and connecting the motor to the amplifier. In this case, as soon as the supply voltage dropped, the resulting error voltage would be amplified and applied to the motor causing it to drive the rheostat to reduce the resistance and increase the output voltage. The motor would continue to run until the error signal went to zero, at which time the output voltage must be up to its original value.

The accuracy with which a positioning servo can repeat the shaft position R is in most cases limited by the amount of torque required to move the friction load. Figure 6-3 shows the torque versus error signal characteristic of a typical servo with the motor stalled. The figure shows that a certain error voltage must exist to produce enough torque to move the load shaft against the starting friction. The battery feeding the controlled variable potentiometer determines how many volts of error signal E will be produced per degree displacement of the controlled variable shaft. Increasing this battery voltage will increase the stiffness of the system at the load shaft.

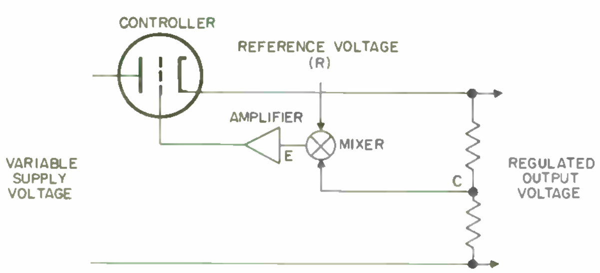


Figure 6-2. Voltage Regulator

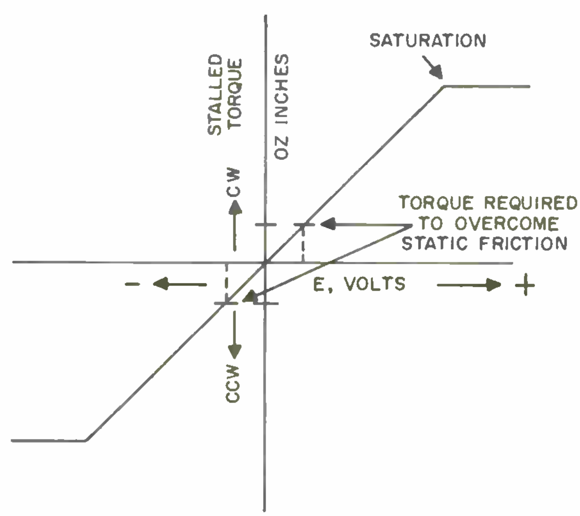


Figure 6-3. Idealized Torque versus Voltage Curve

Stiffness is defined as the reaction torque at the load shaft divided by displacement of the shaft. The greater the stiffness of the system, the smaller will be the residual error.

If the controlled variable shaft is required to follow the reference shaft when it is moving at a constant speed, the motor must turn the output shaft at the same speed as the reference input shaft. Since a certain amount of power from the motor is required to overcome the running friction of the load, there must be a fixed difference between R and C sufficient to produce an error signal capable of driving the motor at the required speed. This difference between R and C indicates that the controlled variable shaft lags behind the reference input shaft by a certain number of degrees, although both are traveling at the same speed. This type of error, called a dynamic tracking error, has a magnitude at any speed of reference input determined by the velocity constant of the servo. To obtain the velocity constant, the shaft of the controlled variable potentiometer is uncoupled from the gear train and displaced one degree from the null position, producing an error signal. The speed of the output shaft is measured, and the ratio of this speed in degrees per second to the error in degrees required to produce it is the velocity constant of the servo system. In figure 6-1, increasing either the gain of the amplifier or the voltage of the battery across the controlled variable potentiometer will increase the velocity constant as well as the stiffness of the servo, so that increasing the over-all gain of the system will decrease both the static friction error and the dynamic tracking error of the system.

3. SERVO STABILITY REQUIREMENTS

A servo system such as that of figure 6-1 may be represented by the block diagram shown in figure 6-4.

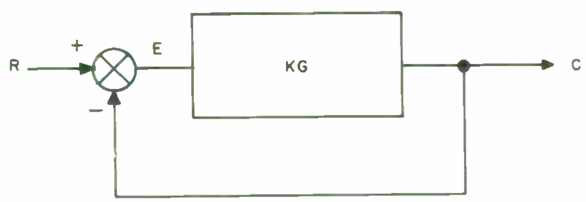


Figure 6-4. Servo Loop Block Diagram

The box labeled KG represents the amplifier, the motor, and the gear train. K is the gain constant of the system, in this case the velocity constant of the entire loop. It includes the amplifier gain, the motor velocity constant, and the gain of the reference input and controlled variable potentiometers. The constant K is independent of the frequency of the applied signals, but G is an expression that describes the frequency response or time response of the amplifier and motor to error signals. From figure 6-4,

$$E = R - C, \text{ and} \tag{1}$$

$$C = KGE \tag{2}$$

Solving (2) for E:

$$E = \frac{C}{KG} \tag{3}$$

Substituting (3) into (1):

$$\frac{C}{KG} = R - C \tag{4}$$

$$C = RKG - CKG \tag{5}$$

$$C(1 + KG) = RKG \tag{6}$$

Hence, the effective gain of the closed loop is:

$$\frac{C}{R} = \frac{KG}{1 + KG} \tag{7}$$

Since the amplifier and motor must be built with physically realizable components, the function G represents a certain finite bandwidth. The inertia of the motor will tend to slow down its response to high-frequency error signals, and since a roll-off in frequency response is accompanied by a phase shift, there will be some frequency at which the controlled variable C will lag the error signal E by 180 degrees. At this frequency the quantity KG becomes negative, and if KG approaches -1, the denominator of equation 7 approaches 0 so that $\frac{C}{R}$ approaches infinity. Physically, this can be interpreted to mean that an output C is obtained with no input R. This is the condition under which the loop will oscillate and is known as the Nyquist stability criterion. Because of the finite bandwidth of G, this condition for stability imposes a limitation on the value of K that may be used.

Now that the condition required for stability has been developed mathematically, the system of figure 6-1 may be examined to see what happens when a servomechanism is unstable. The higher the velocity constant, which includes amplifier gain and the voltage of the battery driving the controlled variable potentiometer, the greater will be the motor speed at any given value of error signal. Because the motor, gear train, and load possess inertia, the system in motion has kinetic energy equal to $J\omega^2$. In order to stop the motor, this energy must be dissipated. Because an inertia cannot be made to move instantaneously, the response of motor speed to a step of error signal

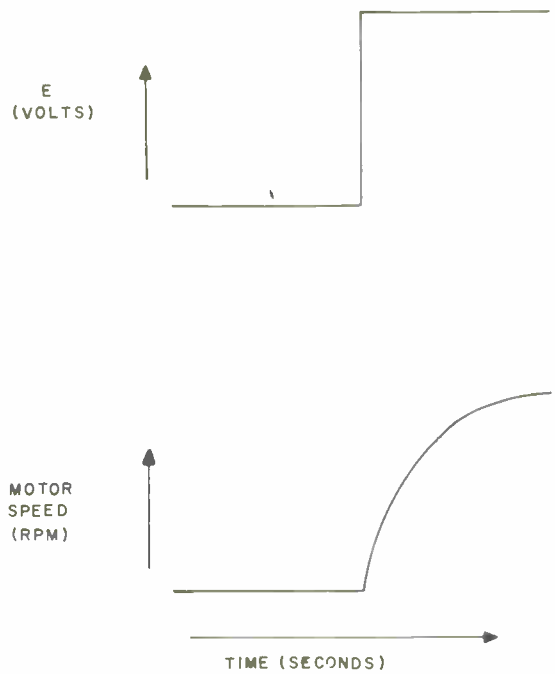


Figure 6-5. Response of Motor and Amplifier to Step Input

voltage into the amplifier is as given in figure 6-5. At the beginning of the step, a step of torque is applied to the rotor, producing acceleration. However, as the motor builds up speed the friction in the load, motor bearings, and gear train dissipates an increasing amount of energy. Eventually, the motor reaches a speed at which the amount of power supplied to the motor winding by the amplifier equals the total amount of power dissipated in the friction load, the motor and gear train bearing friction, and in the copper loss in the motor winding. At this point, the motor speed remains constant. If the system gain is low, the stiffness at the load will be quite small. As the output shaft approaches the null position, the motor torque drops off rapidly enough that the friction can dissipate all the kinetic energy in the motor. Consequently, the response of the closed loop system to a step input will be as shown in figure 6-6b. This condition of the servo loop is referred to as overdamped. If the gain is increased, increasing the stiffness, an oscillatory condition is reached in which the friction load cannot dissipate the energy stored in the inertia by the time the error signal gets to zero. In this case, the motor will overshoot the null position, and the position feedback signal will produce a reverse torque, causing the motor to overshoot in the opposite direction. This oscillation will continue with less energy imparted to the system on each oscillation, until a point is reached where the total system energy at the null is zero. This system is underdamped and has the response to a step of R as shown in figure 6-6c. If the system gain is made sufficiently large, the stiffness is so great that the amplifier is able to add more kinetic energy to the motor in each successive cycle of oscillation than the friction load can dissipate. This condition results in divergent oscillations, as shown in figure 6-6d.

4. STABILIZING METHODS

In many cases a servo may be satisfactorily damped by merely adding some sort of velocity-proportional friction device to absorb the energy stored in the inertia. Automobile shock absorbers are an example of this type of damper, and small instrument servomotors are frequently equipped with a drag cup fastened to the motor shaft and turning in a fixed magnetic field. Such dampers produce a torque proportional to velocity. Friction dampers reduce the velocity constant of a system because the motor requires more voltage to run at a given speed. However, the damping effect allows the gain to be made up in the amplifier, so that for a given velocity constant a greater stiffness may be realized, and this reduces the static error in the system.

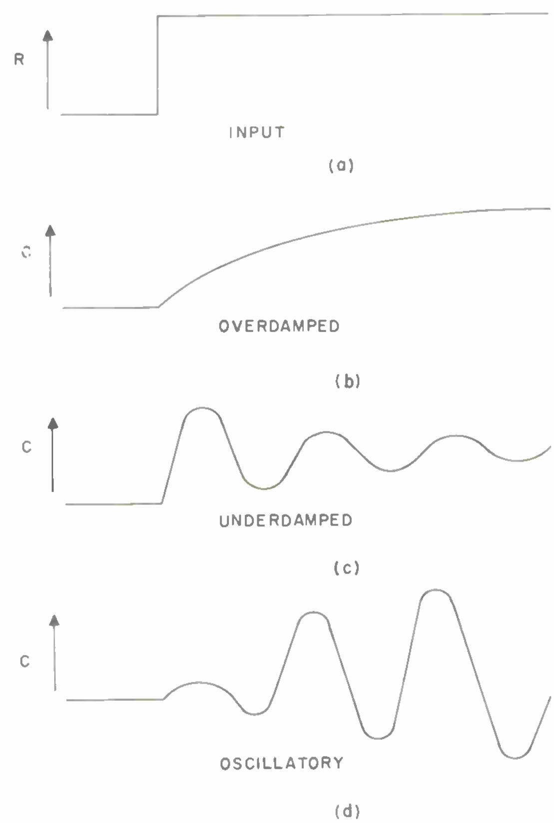


Figure 6-6. Response of Follow-Up Servo to Step Input

Another method of damping a servo is to use rate feedback, as shown in figure 6-7b. A generator is coupled directly to the motor shaft, and the output of the generator is a voltage directly proportional to the motor speed. If this voltage is fed back inversely to the amplifier, it results in a torque proportional to speed, the same as would be obtained with a friction or drag cup damper. However, since the subtracting is done at a low signal level, the amplifier is not required to supply any more power than when it is driving a motor with no load, and the motor does not have to be so large. Also, there is no requirement for dissipating the motor's energy when it is running. Figure 6-7a shows the response of motor speed to a step of error voltage, similar to figure 6-5. Let us assume for illustration that in the steady state condition .1 volt of E is required to make the motor turn at 1000 rpm. Very little torque is available to accelerate the motor initially because of the small error signal, and hence the rise time is rather large. In figure 6-7b, the generator fastened to the motor shaft puts out 1 volt per 1000 rpm. When excited with a step function E of 1.1 volts, the error signal ϵ will at first be 1.1 volts, since the motor starts at zero speed and the generator output is initially zero. As the motor picks up speed, the generator voltage which is an exponential will subtract from the 1.1 volts initial value, and in the steady state condition, when the motor reaches 1000 rpm, ϵ will be 1.1 volts minus the 1 volt generator output, which leaves .1 volt. The voltage going into the amplifier for the steady state condition is the same as in figure 6-7a, but the 1.1 volt spike present in the amplifier input in figure 6-7b produces eleven times more acceleration torque at the motor, reducing the rise time. Because the amplifier gain K is the same in both cases and because the rate generator output is 0 when the motor is stalled, the stalled torque for a given error signal is the same in both cases. Because of the much larger value of E required to get 1000 rpm with a rate generator, the velocity constant for this case will be 1/11 that of the motor alone. If in figure 6-7b, the amplifier gain K were multiplied by 11, the two systems would then have the same velocity constant but the system with the rate generator would have 11 times the stiffness, and the positioning accuracy would be 11 times as great.

Another method of stabilizing a servo and allowing an increase in its stiffness is to use a lead network. Figure 6-8 shows a lead network and its transient response to a step input. If 1 volt is suddenly applied as E_{in}, the entire 1 volt will appear as E_{out} because the voltage across C cannot be changed instantaneously. As the capacitor charges up, E_{out} will drop exponentially and approach the value it would have if C were not present, which is $E \frac{R_2}{R_1 + R_2}$. This transient response is similar to that of ϵ in figure 6-7b. Figure 6-9 shows how a lead network is used to accomplish a result similar to that obtained with a rate generator.

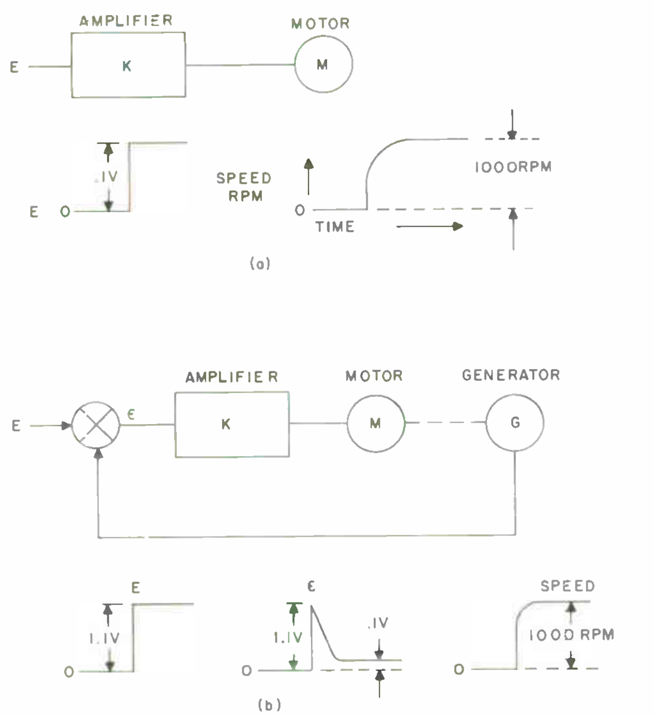


Figure 6-7. Effect of Rate Generator on Response Time

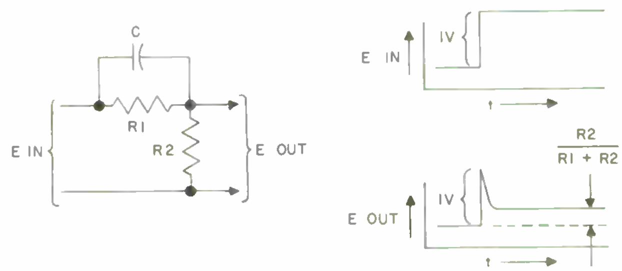


Figure 6-8. Lead Network

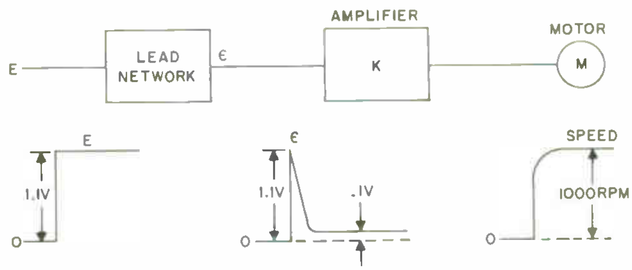


Figure 6-9. Effect of Lead Network on Response Time

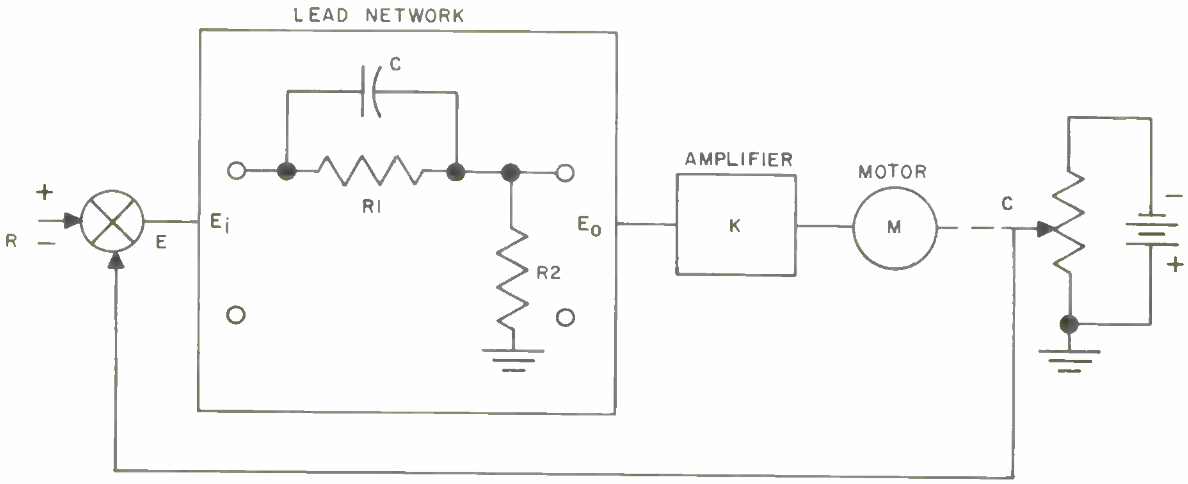


Figure 6-10. Position Follow-up Servo with Lead Network

If $\frac{R_2}{R_1 + R_2}$ is made equal to 1/11, a voltage of 1.1 volt applied at E will produce the wave form shown at ϵ and this will produce a rapid acceleration of the motor to a speed of 1000 rpm. The capacitor in the lead network must be chosen to produce the same time constant in the signal at ϵ as that of figure 6-7b.

The transfer function of the lead network of figure 6-8 may be derived by considering it a voltage divider with the parallel combination of R_1 and C in the top leg and R_2 in the bottom leg. Thus the output impedance is

$$Z_o = R_2$$

The parallel impedance of R_1C is

$$\frac{R_1 \frac{1}{pC}}{R_1 + \frac{1}{pC}} = \frac{R_1}{1+pR_1C}$$

and the total series impedance of the divider is

$$Z_t = R_2 + \frac{R_1}{1+pR_1C} \quad \text{Hence}$$

$$\frac{E_o}{E_i} = \frac{Z_o}{Z_t} = \frac{R_2}{R_2 + \frac{R_1}{1+pR_1C}} = \frac{R_2(1+pR_1C)}{R_1+R_2+pR_1R_2C}$$

$$= \frac{R_2}{R_1+R_2} \; \frac{1+pR_1C}{1+p\frac{R_1R_2}{R_1+R_2}C}$$

This expression may be rewritten as follows:

$$\frac{E_o}{E_i} = \frac{R_2}{R_1+R_2} + \frac{R_2}{R_1+R_2} \; \frac{\left(R_1 - \frac{R_1R_2}{R_1+R_2}\right)Cp}{1+\frac{R_1R_2}{R_1+R_2}Cp}$$

$$\frac{E_o}{E_i} = \frac{R_2}{R_1+R_2} + \frac{R_1^2 R_2 C}{\left(R_1+R_2\right)^2} \; \frac{p}{1+\frac{R_1R_2}{R_1+R_2}p}$$

$$\frac{E_o}{E_i} = K_1 + K_2 \frac{Tp}{1+Tp}$$

therefore $\quad E_o = K_1E_i + K_2 \dfrac{Tp}{1+Tp} E_i$

Thus, the lead network behaves like a straight feed with a gain of K_1 plus a high-pass filter with a gain of K_2. The network differentiates low frequency signals. When connected in a closed position loop, as shown in figure 6-10, a lead network provides the sum of a position signal and a differentiated position signal, so that the rate of change of ϵ is used, producing an effect similar to that of a rate generator except that R is differentiated as well as C.

5. COMPONENTS AND CIRCUITS

In the preceding discussion of the position follow-up servomechanism of figure 6-1, it was assumed that the reference input and controlled variable voltages, the amplifier, and the motor were all direct-current components. In practice, 400 cycle or 60 cycle carrier systems are more commonly used for small, low power servo systems in radio communication equipment. The use of an a-c carrier system simplifies the design of the amplifier, and since an amplifier bandwidth of 40 or 50 cycles usually suffices, the d-c operating point of the individual stages is of no consequence. Because of the low-frequency requirement, the junction transistor is well suited to servo work. Where 20 watts or more of amplifier output is required, a magnetic amplifier or saturable reactor driven by a transistor preamplifier may be used. Some servo amplifiers employ high performance magnetic amplifiers for all stages.

The most commonly used type of servomotor is the two phase induction motor, and frequently a two phase induction generator is built into the same case and coupled to the motor shaft. The schematic diagram for a typical motor generator appears in figure 6-11. The reference phase of the motor must be driven with a voltage 90° out of phase with the voltage on the control phase in order to obtain torque. If the control phase voltage leads the reference phase voltage, the motor will turn in one direction, and if the control voltage lags the reference voltage, the motor will turn in the opposite direction. Thus if the error signal source is a 400 cps signal, a phase reversal of the error signal produces a reversal of motor rotation, as required.

In some cases where the servo amplifier output is either in phase or 180 degrees out of phase with the line voltage depending on the sense, the required quadrature relationship between control and reference winding is obtained by means of a phase shift capacitor C which produces a quadrature voltage on the reference winding. Sometimes however, it is more convenient to produce the required 90° phase shift inside the servo amplifier, in which case the reference winding is connected directly to the 400 cps line. When its excitation winding is driven from the 400 cps line, the rate generator produces across its output terminals a voltage proportional to the speed of rotation and either in phase or out of phase with the excitation voltage, depending upon the direction of rotation.

If the purpose of the servomechanism is merely to repeat the position of a shaft or to produce an output shaft position proportional to a voltage used as a reference input, a-c line voltages may be used across reference and controlled variable potentiometers in place of the d-c voltages used in the illustration of figure 6-1. In this case, as the controlled variable voltage increases and becomes larger than the reference input, the phase of the error signal voltage reverses and reverses the direction of rotation of the two-phase servomotor. In this way a carrier servo system may be constructed in which all variables are represented by 400 cps a-c voltages.

In some transmitter tuning servos and antenna matching networks, it is desired to have the servomotor and gear train position a mechanical tuning element such as a variable capacitor to a position such that the phase shift imposed upon an r-f signal by the tuned circuit is zero. An r-f phase discriminator circuit of the type used for detection of FM signals may be used to obtain a d-c voltage proportional to the magnitude of the phase shift through the r-f circuit, and of polarity determined by whether the output leads or lags the input. If this d-c error information is to be fed into a carrier type servomechanism, it may be converted to a-c by means of an electromechanical chopper connected as in figure 6-12. The chopper consists of a vibrating reed and a pair of contacts which form a single-pole double-throw switch. The reed is excited from a magnetic coil and is driven from the 400 cps line. In most cases the action of the reed contact is not in phase with the excitation voltage fed to the coil, so that a phase shift network must be used on the coil

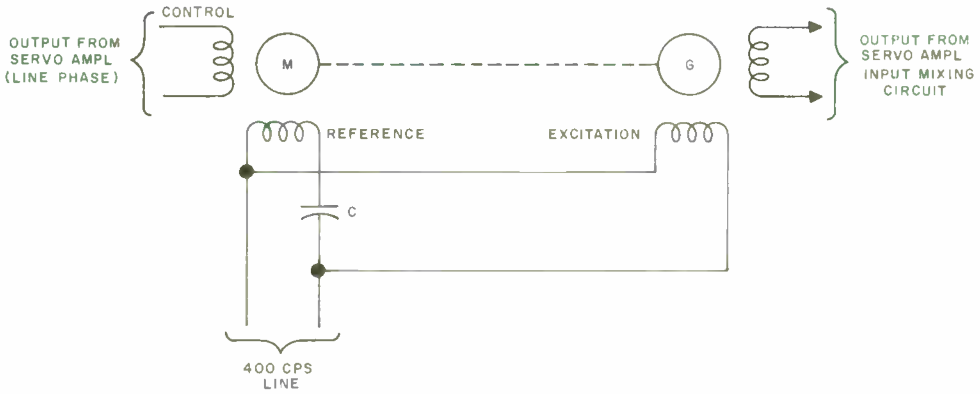


Figure 6-11. Motor-Generator Schematic Diagram

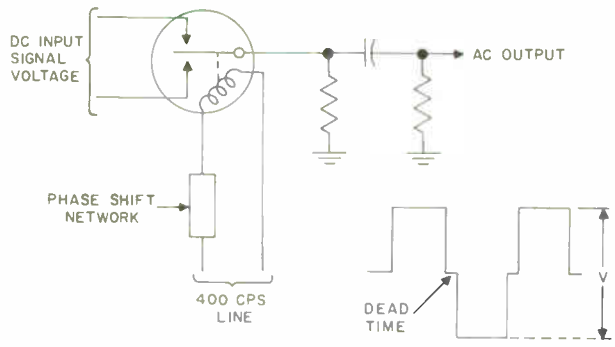


Figure 6-12. Electromechanical Chopper Circuit

to make the reed contact action either in phase with or in quadrature with the line, as required. In the chopper output wave form shown in figure 6-12, the peak to peak voltage of the square wave is the amplitude of the d-c voltage connected across the contacts.

Another type of position transmitter frequently encountered is the synchro. Figure 6-13 illustrates a typical synchro error circuit. The synchro transmitter may be thought of as being a transformer with a single primary, the rotor winding, and three secondaries, the three stator windings. The stator windings are placed with their axes 120° apart. The rotor winding induces an a-c voltage in each of the stator windings proportional to the cosine of the angle between the rotor winding axis and the respective stator winding axes. The stator windings of the control transformer are connected directly across the stator windings of the transmitter. Therefore, the same voltages will exist across each of the control transformer stator windings as are induced in the corresponding stator windings of the transmitter, and the field pattern set up inside the core of the control transformer will be a replica of the field pattern in the transmitter. If the control transformer rotor winding axis is lined up with this field, the error signal voltage developed across the rotor terminals will be maximum. As the control transformer rotor is turned, the error signal voltage will decrease and become zero when the rotor axis makes an angle of 90° with the field pattern set up by the stator windings. Further rotation of the rotor will produce an increasing error voltage of reversed phase. Thus, if the rotor of the transmitter is actuated by the reference input shaft of a position follow-up system (figure 6-1) and the control transformer rotor is coupled to the controlled variable shaft, the voltage developed across the rotor of the control transformer may be used as an error signal E to be fed into the servo amplifier, and the rotor of the control transformer will follow the transmitter rotor.

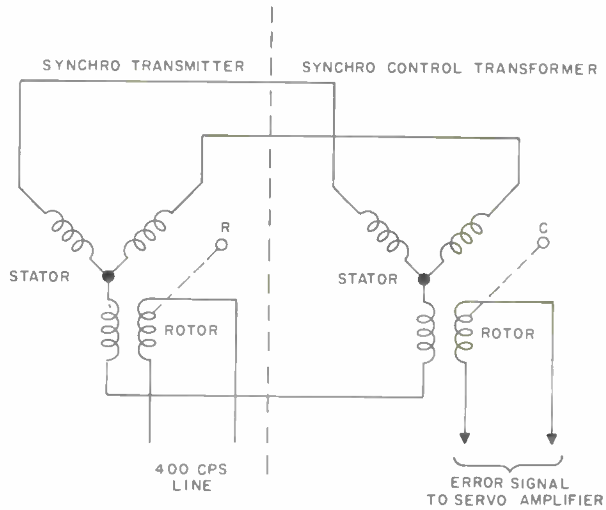
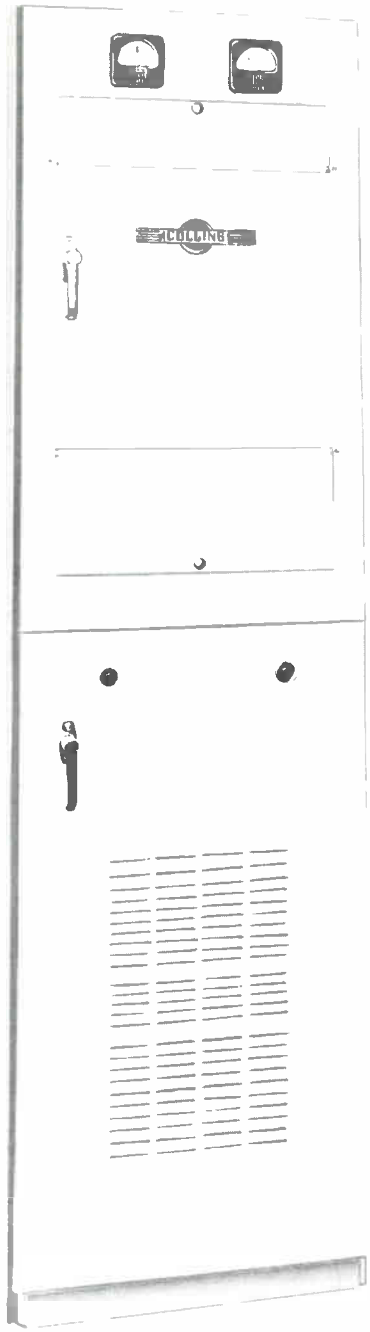


Figure 6-13. Synchro Error Circuit



Collins 204F Power Amplifier, a Three Stage Linear Amplifier with an Output of 2.5 Kw PEP

CHAPTER 7

R-F LINEAR POWER AMPLIFIERS

1. INTRODUCTION

The r-f power amplifier of the SSB transmitter receives a low power level, radio-frequency SSB signal from the exciter. The function of the power amplifier is to raise the power level of the input signal without changing the signal. That is, the envelope of the output signal must be a replica of the envelope of the input signal. A power amplifier which will perform this function is, by definition, a linear power amplifier.

2. POWER AMPLIFIER CLASSIFICATION

Radio-frequency amplifiers are classified A, B, and C according to the angle of plate current flow; that is, the number of degrees of plate current flow during a 360° r-f cycle. Class A amplifiers have a continuous plate current flow and operate over a small portion of the plate current range of the tube, as shown in figure 7-1. This class amplifier is used for amplification of small signals for low distortion. Its efficiency in converting d-c plate power input into r-f power output is quite low, usually less than 35 per cent, but this is seldom of major importance where small signals are amplified.

Class B amplifiers have their grids biased to near plate current cutoff so that plate current flows for approximately 180° of the r-f cycle, as shown in figure 7-2. Amplifiers operated with appreciably more than 180° of plate current flow but less than 360° are called class AB amplifiers. Both class AB and class B operation is used in the high-power stages of r-f linear amplifiers to achieve higher efficiency and maximum output power with low distortion. Plate efficiency depends upon the tube used and the operating conditions selected, with efficiencies in the range of 50 to 70 per cent obtainable. The distinction between class B and class AB is somewhat arbitrary since both operate over more than 180° but less than 360°. However, the class AB amplifier draws appreciably more static plate current than the class B amplifier, which draws only a small static plate current.

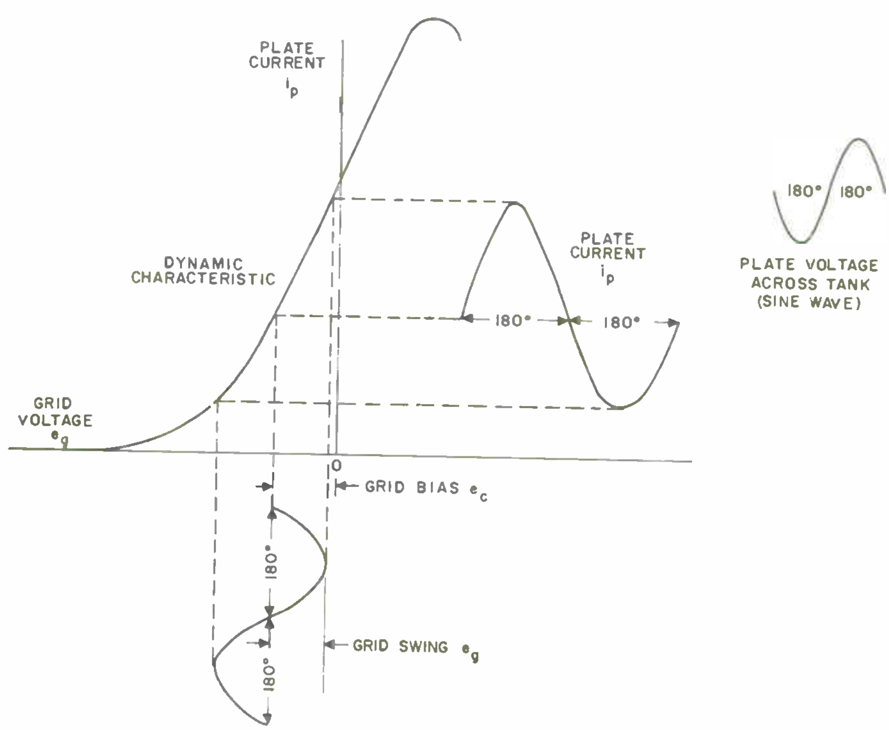


Figure 7-1. Class A Tube Operation

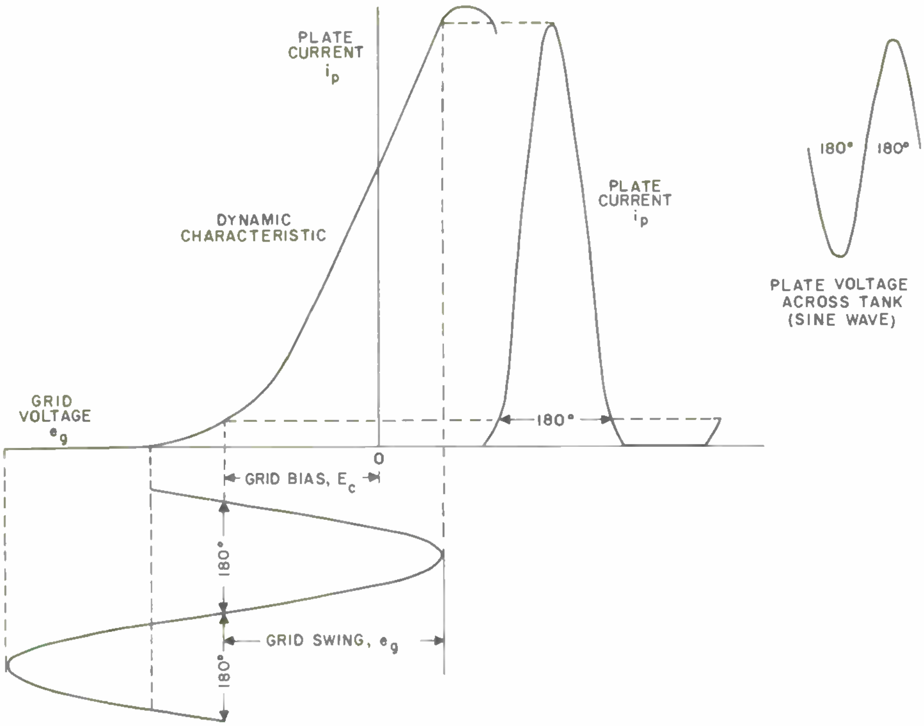


Figure 7-2. Class B Tube Operation

The class C amplifier, as shown in figure 7-3, is biased well beyond cutoff so that plate current flows less than 180° of the r-f cycle. The principal advantage of the class C amplifier is high plate efficiency, from 65 to 85 per cent, but class C amplifiers are not suited for SSB use because they are not linear amplifiers and will not respond to low-level input signals.

A subscript number is commonly added to the amplifier class designator to indicate whether or not the tube is operated in the positive grid region over part of the cycle. For example, class AB_1 indicates that the grid never goes positive so that no grid current is drawn. Class AB_2 indicates that the grid does go positive so that grid current is drawn. Because class A amplifiers are nearly always operated without grid current, and because class C amplifiers are nearly always operated with grid current, subscript designators are omitted unless they are operated to the contrary of the usual practice.

3. R-F POWER AMPLIFIER TUBES

Conventional grid-controlled power amplifier tubes are classified according to the number of elements they have. Until fairly recently, the triode which has a control grid in addition to the cathode and anode was the only transmitting-type tube available in the medium and high power sizes. (The cathode is sometimes called the filament because the cathode is usually directly heated in high power tubes, and the anode is often called the plate.) Tetrodes, which have a screen grid between the control grid and plate, have recently become available. The screen grid provides an accelerating potential to the electron stream and also provides an electrostatic shield between the anode and the control grid. The two grids of most transmitting-type tetrodes provide a beaming action to the electron stream which improves the tube characteristics. This beaming action reduces the d-c screen current and increases the control of the control grid. Power pentodes up to 1 kw have an additional grid, a suppressor grid, located between the screen and the anode. In some beam power tubes this element may consist of beam forming plates which, in general, give an improved plate characteristic when the plate voltage swings below or in the region of the d-c screen voltage.

Triode power amplifier tubes have the advantage of simplicity, low cost, and availability in all sizes. In general, they require a large amount of driving power. Also, since their grid is exposed directly to the plate, there is considerable capacitive coupling from the plate to the grid within the tube. This plate-to-grid capacitance must be accurately neutralized in r-f linear power amplifiers. The amplification factor of triode tubes ranges from 4 to 5 for low mu tubes, to twenty for medium mu tubes to fifty for high

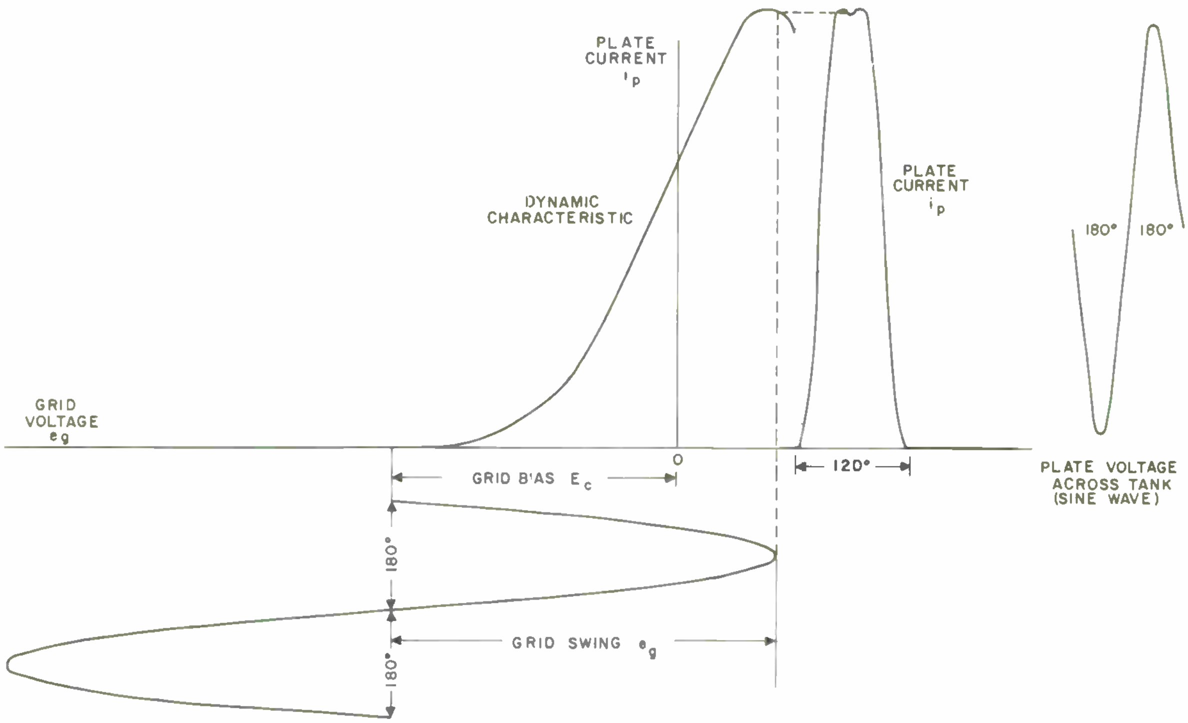


Figure 7-3. Class C Tube Operation

mu tubes. Generally only the low and medium mu triodes can be used for linear power amplifier circuits. Therefore, a large grid swing is required to obtain the power amplification available from the tube.

In tetrode power amplifier tubes, the screen grid acts as an electrostatic shield between the plate and control grid which reduces the plate-to-grid capacitance. This reduces the neutralization required to as little as one-hundredth of that required for a triode. However, since the gain of the tetrode tube is so much higher than that of the triode, neutralization of the small residual plate-to-grid capacitance of the tetrode is still required for the best, high-gain linear performance. Because of the high gain of the tetrode, the tube requires relatively low drive to obtain high power output. This advantage allows fewer stages to be used to obtain a given power output.

Pentode construction is used in most small receiving size power tubes and in some cases in power tubes up to 1 kw output. In small tubes, pentodes provide good performance with low plate voltage, and in larger tubes, pentodes give improved efficiency because the r-f plate swing can be increased some. The pentode has disadvantages in that it is more complex than the tetrode, is more expensive, and requires extra circuitry for the suppressor grid. These disadvantages have limited the development of the pentode power tubes. At the present time, the pentode has little advantage over the well designed tetrode.

For r-f linear amplifier operation, the following features are desirable in the power amplifier tubes:

(1) High gain

(2) Low plate-to-grid capacitance

(3) Good efficiency

(4) Linear characteristics which are maintained without degradation at all frequencies in the desired operating range

The needs for power amplifier tubes in the vhf and uhf ranges have spurred development of tubes suitable for operation at those frequencies. This has resulted in tubes with better performance in the h-f (3 to 30 mc) range. A typical comparison can be made between the type 813 tube and the type 4X250B tube which are in the same power class. The small compact design of the 4X250B tube results in short lead lengths, better screening, closer element spacing and much higher performance which can be maintained easily over the h-f range. The ceramic construction, rather than glass, of an increasing number of new tubes promises to result in a more rugged and longer lifed tube. Ceramic sealed tubes which are now available include the RCA-6118 which is smaller than the 4X150A, the Eimac 4CX300A which has characteristics similar to the 4X250B, an all ceramic version of the

4X250B, the Eimac 4CX5000A which is capable of 10 kw of r-f output, and an RCA super-power, shielded-grid tube that will deliver 500 kw of r-f output. Tube manufacturers have additional types of power amplifier tubes under development which promise better performing tubes for the near future.

The Collins Radio Company has chosen to use high gain tubes of those types considered to be the best compromise of desired characteristics. At low signal levels, such as exist in exciters, conventional receiver-type r-f amplifier tubes are used. For delivering .1 watt output from exciters, the type 6CL6, which is a miniature 9-pin tube, is generally used. The 6CL6 is also frequently used to excite type 4X250B power amplifier tubes. The 4X250B tube is used in small, compact equipment for power levels of 1 kw by paralleling three, and for power levels of 500 watts by paralleling two. The type 4CX5000A is used for power levels of from 5 kw to 10 kw. This tube is used to obtain power levels up to 45 kw by paralleling four of them.

4. BASIC LINEAR POWER AMPLIFIER CIRCUITS

a. GENERAL

For linear operation, r-f power amplifiers may be operated class A or class AB. The amplifiers used are quite conventional, being either grid driven or cathode driven (grounded grid) type amplifiers. However, the design considerations are extremely stringent to produce maximum linearity for a given tube in a given circuit. The tube operating point must be discreetly chosen and precisely maintained, neutralization must be as effective as possible, r-f feedback circuits are often used, and input and output impedances must be held as constant as possible. Generally class A pentode power amplifiers are employed in low-level power stages to preserve linearity in these stages while producing enough power to drive the higher level stages. Class AB_1 or AB_2, triode or tetrode power amplifiers are employed in the high-level power stages to obtain the desired power output.

b. GRID DRIVEN TRIODE POWER AMPLIFIER

Figure 7-4 is a simplified schematic of a typical grid driven triode power amplifier. This amplifier, operating class AB_1, produces up to 2.5 kw using the type 3X3000A-1 triode. The triode tube, having a large plate-to-grid interelectrode capacitance, always requires neutralization to prevent oscillation when used in the grid-driven circuit. The only types of triodes capable of class AB_1 operation are the low amplification factor types, such as the 3X3000A-1. Due to the low amplification factor, very high r-f grid excitation voltage is required, on the order of 1000 volts for the 3X3000A-1. A similar tube suitable for class AB_2 operation is the 3X2500A-3 which has an

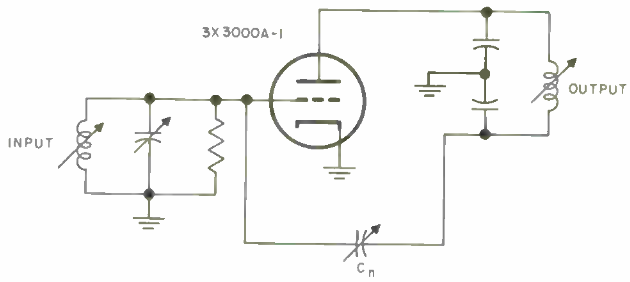


Figure 7-4. Grid Driven, Plate Neutralized Triode Power Amplifier

amplification factor of 20. This medium-mu triode requires less grid swing, but it requires grid driving power for class AB_2 operation. Neutralization, of course, is still required.

A swamping resistor is used in the grid circuit to maintain a constant input impedance to the stage and for stability. When the stage is operated class AB_2, the grid current represents a varying load to the driving source. By adding the swamping resistor, the grid current drawn represents only a small portion of the total grid load so that the driver load impedance is relatively constant. The swamping resistor does increase the required driving power. The swamping resistor also improves stability by affording a low impedance to ground for regenerative feedback through the plate-to-grid capacitance.

c. CATHODE DRIVEN TRIODE POWER AMPLIFIER

Figure 7-5 is a simplified schematic of a typical cathode driven (grounded grid) triode power amplifier. This amplifier, operating class AB_2 produces 4 to 5 kw using the type 3X2500A triode. In the cathode driven amplifier, the control grid is at r-f ground and the signal is fed to the cathode. The main advantage

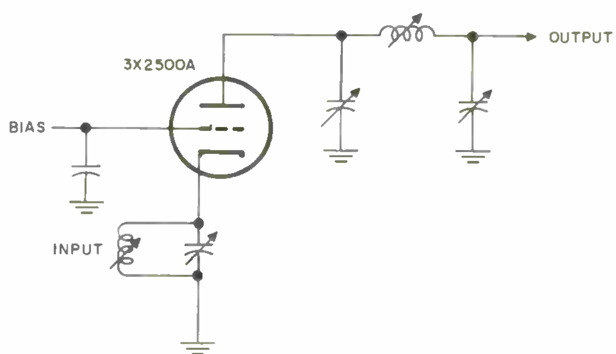


Figure 7-5. Cathode Driven Triode Power Amplifier

of operating the triode in this manner is that the control grid becomes an effective screen between the plate and the cathode making neutralization seldom necessary. The small values of plate-to-cathode capacity have very little effect on the input signal because the input circuit impedance is usually quite low. Since neutralization is not required, triodes with an amplification factor of 20, such as the 3X2500A, can be used. Another advantage of the cathode driven power amplifier is that the feedthrough power is an effective load across the input circuit, making swamping resistors unnecessary. The main disadvantages of this circuit are that a large driving power is required and that power gains of from six to ten are all that can be realized. Most of the power required for driving, however, feeds through the stage and appears in the plate circuit so that it is not lost. The cathode driven circuit is a convenient circuit to use when high power has already been developed and needs another step up.

d. GRID DRIVEN TETRODE POWER AMPLIFIER

Figure 7-6 is a simplified schematic of a grid driven tetrode power amplifier. This amplifier, operating class AB_1 produces 250 watts per tube using the type 4X250B tetrode. In general, the same design considerations exist for tetrode amplifiers as for triode amplifiers. That is, grid circuit swamping is required to hold the input impedance constant if the tetrode is driven into the grid current region, and neutralization is generally required if the tube is to operate over the entire high-frequency range. However, since the plate-to-grid capacitance is small in the tetrode, neutralization is much simpler. The tetrode amplifier, being a high gain tube, requires relatively little driving power and a relatively small grid swing for operation. This permits the paralleling of tubes with a common input network and a common output network which reduces the number of stages and simplifies tuning. In the tetrode power amplifier, the screen voltage has a very pronounced effect on the dynamic characteristic of the tube. By lowering the screen voltage, the static current required for optimum linearity is lowered. This permits greater plate r-f voltage swing which improves efficiency. The use of lower screen voltage has the adverse effect of increasing the grid drive for class AB_2 operation and lowering the power output for class AB_1 operation. The tetrode tube can be used in the cathode driven circuit and can be so used without neutralization in the high-frequency range.

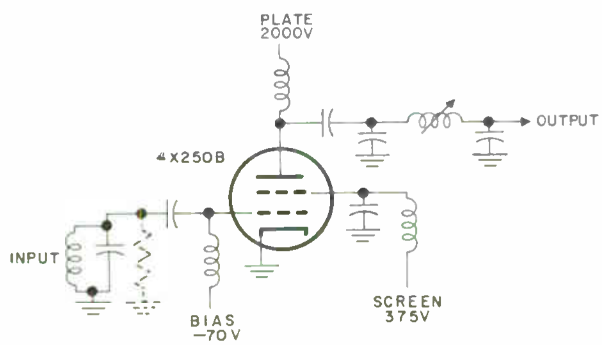


Figure 7-6. Grid Driven Tetrode Power Amplifier

5. POWER AMPLIFIER OUTPUT NETWORKS

a. TANK CIRCUIT CONSIDERATIONS

The plate tank circuit of an r-f power amplifier must perform four basic functions:

(1) It must maintain a sine wave r-f voltage on the plate of the tube.

(2) It must provide a low impedance path from plate to cathode for harmonic components of the plate current pulses.

(3) It must provide part or all of the necessary attenuation of harmonics and other spurious frequencies.

(4) It must provide part or all of the impedance matching from the tube plate to the antenna.

In addition, for many uses the output circuit should be single ended so that it will feed into a 52 ohm coaxial transmission line. A 52 ohm coaxial transmission line is desirable because it prevents stray r-f radiation near the transmitter; it is convenient for coaxial r-f switching; it is a convenient impedance for additional r-f filtering, and because it is ideal for directional wattmeter installation. For simplicity of operation, the output circuit should require a minimum of tuning controls. A direct-coupled network, such as the Pi-L network, is the most suitable network to meet these requirements.

The Q of the plate circuit, of which the tank is a part, must be sufficient to keep the r-f plate voltage close to a sine wave shape. This is often referred to as the "flywheel effect." If the plate circuit Q is insufficient, the r-f waveform may be distorted which will result in low plate efficiency. This loss of efficiency is seldom noticed unless the plate circuit Q is less than 5. A plate circuit Q of at least 10 is known to be sufficient for linear operation and is a recommended minimum.

A power amplifier operating either class AB, B, or C delivers power to the tank circuit by plate current pulses. The harmonic content of these pulses is determined primarily by the angle of plate current flow, the harmonics being greater with a smaller angle of plate current flow. In a linear power amplifier, the second harmonic component can be as great as 6 db below the fundamental at full peak envelope power. The higher order harmonic components drop off rapidly but their magnitude varies greatly, depending upon the pulse shape. These harmonics must be attenuated in the output network so that they are 50 db, 80 db, or even further, below the fundamental component. The Pi-L network will attenuate the second harmonic to about 50 db below the fundamental, which is from 10 db to 15 db more attenuation than can be obtained from the simple Pi network. Where more attenuation is required, external filters of either the low-pass or band rejection type are added. Increasing plate circuit Q increases harmonic attenuation, but since doubling the Q results in only about 6 db more second harmonic attenuation, Q's above 20 are seldom used below 30 mc.

The Pi-L output network is ideally suited to matching a tube load to a 52-ohm coaxial transmission line. Loads with a standing wave ratio as high as 4 to 1 can be matched easily. This can be done with any value of tube load impedance, whereas the simple Pi network has difficulty matching to low load impedance when the tube plate load resistance is high.

The Pi-L network has only four variable elements, and they can be ganged to have only a tuning control and a loading control, as shown in figure 7-7. Since in the Pi-L network, C_2 and L_2 affect loading in the same direction, the extra capacity and inductance range of the elements required to extend the loading range of the circuit is relatively small. For example, the loading control varies about ±25 per cent to match a 52 ohm load with a 4:1 swr. The tuning control varies about ±10 per cent.

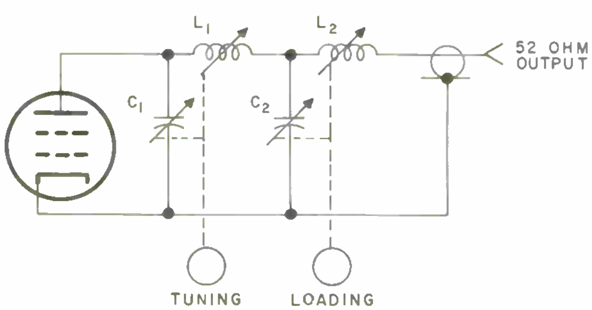


Figure 7-7. Tuning Controls for Pi-L Output Network

b. CIRCUIT LOSSES

Nearly all of the tank circuit loss occurs in the coils. These losses are closely related to the ratio of plate circuit Q to coil Q, but other design considerations enter in. These circuit losses are shown in figure 7-8 for a Pi-L network, which has lower losses than other networks for 50 db of second harmonic attenuation. Resistances r_1 and r_2 represent the equivalent series resistance of the coils determined from coil Q and reactance. Resistance r_q is the equivalent load resistance in series with L_1 and is determined from the relationship

$$r_q = \frac{R_L}{Q^2 + 1} = \frac{R_L}{(R_L/X_C)^2 + 1}$$

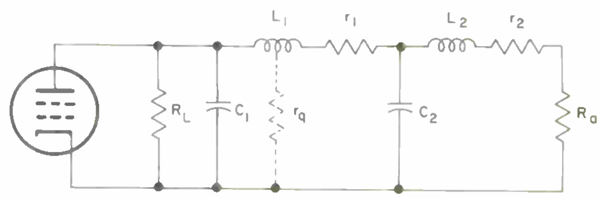


Figure 7-8. Circuit Losses in Pi-L Network

Resistance R_a is the series resistive component of the load. The Pi-L network loss is given by the equation:

$$\text{Per cent loss} = \left(\frac{r_1}{r_q + r_1} + \frac{r_2}{R_a + r_q}\right) 100$$

c. TANK COIL AND CAPACITOR REQUIREMENTS

The frequency range and method of tuning are major factors in determining tank circuit components. Continuously variable coils and capacitors which will cover the entire frequency range without any band switching are the most desirable. However, this is not practical in Autotune transmitters because of the limited torque available to drive the tuning elements and the often short repositioning time specified. With these limitations, bandswitching is almost essential. Where instantaneous frequency change is specified, it is common to switch from one pretuned r-f unit to another and manually tuned circuits are suitable for this purpose. Servo control of the tuning elements permits incorporation of various automatic tuning or prepositioning circuits and is well suited for driving continuously variable elements. A practical way to design a transmitter is to use continuously variable elements that can be operated either manually or by an accessory servo system.

The use of continuously variable elements has the following advantages:

(1) The circuit Q can be kept more uniform across the frequency range.

(2) The circuit losses can be kept to a minimum.

(3) The range of variable coils and capacitors can be less.

(4) A maximum amount of harmonic attenuation is more easily maintained across the frequency range.

Variable vacuum capacitors are widely used in transmitters with power levels of 1 kw and higher. Their added expense is often justified by the added capacity range, small size, and low series inductance, especially where voltages above 2500 volts are employed. Variable tank coils are usually constructed with a rotary coil and either a sliding or rolling contact that traverses the length of the coil as it is rotated. The unused turns are shorted out to keep high voltages from developing in them. The series self-resonant frequency of the shorted-out section must not be near the operating frequency or high circulating currents will develop and cause, appreciable power dissipation.

6. NEUTRALIZATION

a. EFFECTS OF PLATE-TO-GRID CAPACITANCE

The purpose of neutralization is to balance out the effect of plate-to-grid capacitive coupling in a tuned r-f amplifier.

In a conventional tuned r-f amplifier using a tetrode tube, the effective input capacity of the tube is given by the following equation:

$$\text{Input capacitance} = C_{in} + C_{gp}\,(1 + A\cos\Theta)$$

where C_{in} is tube input capacitance

C_{gp} is plate-to-grid capacitance

A is voltage amplification from grid to plate

Θ is phase angle of plate load.

In an unneutralized, 4-1000A tetrode amplifier with a gain of 33, the input capacity of the tube with the plate circuit in resonance is increased 8.1 uuf due to the unneutralized plate-to-grid capacity. This small increase in capacitance is not particularly important in amplifiers where the gain remains constant, but if the gain does vary, serious detuning and r-f phase shift can result. The gain of a tetrode or pentode r-f amplifier operating below plate saturation does vary with loading so that if it drives a following stage into grid current, the loading increases and the gain falls off.

The input resistance of the grid is also affected by the plate-to-grid capacitance. The input resistance is given by the following equation:

$$\text{Input resistance} = \frac{1}{2\pi f\, C_{gp}\,(A\sin\Theta)}$$

This input resistance is in parallel with the grid current loading, grid tank circuit losses, and driving source impedance. When the plate circuit is tuned to the inductive side of resonance, energy is transferred from the plate to the grid circuit through the plate-to-grid capacitance (positive feedback). This introduces negative resistance in the grid circuit. When this shunt negative resistance across the grid circuit is lower than the equivalent positive resistance of the grid loading, circuit losses, and driving source impedance, the amplifier will oscillate. As the plate circuit is tuned to the capacitive side of resonance, the input resistance becomes positive and power is transferred from the grid-to-the-plate circuit. This is why the grid current in an unneutralized tetrode r-f amplifier varies from a low value to a high value as the tank circuit is varied from below to above resonance. If the amplifier is overneutralized, the effect reverses. This effect can be observed in a pentode or tetrode amplifier operating class A or AB_1 by placing an r-f voltmeter across the grid circuit and tuning the plate circuit through resonance.

b. NEUTRALIZING CIRCUITS

Most of the neutralizing circuits developed for use with triodes may be used equally successfully with tetrodes. However, those circuits which require balanced tank circuits for neutralizing purposes only, are undesirable because the trend in r-f power amplifier design is toward single-ended stages.

A conventional grid neutralized amplifier is shown in figure 7-9. Capacitor C_3 balances the grid-to-filament capacity to keep the grid circuit in balance. When $C_1 = C_2$ and $C_n = C_{gp}$, it is readily seen that a signal introduced into the grid circuit will not appear across the plate circuit because the coupling through C_n is equal and opposite to the coupling through C_{gp}.

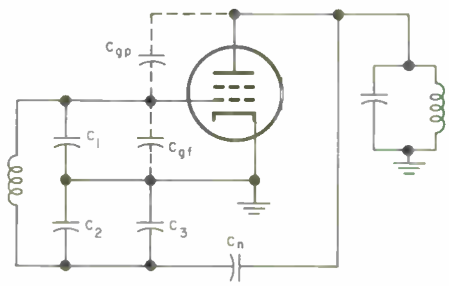


Figure 7-9. Conventional Grid-Neutralized Amplifier

The relationship for no coupling from the grid circuit to the plate circuit is given by the relationship

$$\frac{C_1}{C_2} = \frac{C_{gp}}{C_n}$$

This indicates that the grid tank circuit need not be balanced to ground. If C_2 is made larger, then C_n must be made correspondingly larger. In a tetrode amplifier, C_{gp} is very small (approximately .1 uuf) so that practical values, 5 uuf, can be used for C_n when C_2 is very much larger than C_1.

By placing most of the grid tuning capacitance across the grid tank coil, using the bypass capacitor C from the bottom end of the grid tank circuit to ground for C_2, and using the grid-to-filament capacity for C_1, the modified grid neutralized circuit shown in figure 7-10 results. The relationship for neutralization of this circuit is given by the relationship

$$\frac{C_n}{C} = \frac{C_{gp}}{C_{gf}}$$

This relationship assumes perfect screen and filament bypassing and negligible effect from stray inductance and capacity. This modified grid neutralizing circuit is very effective for neutralizing tetrode power amplifiers and is accomplished with single-ended tuning elements.

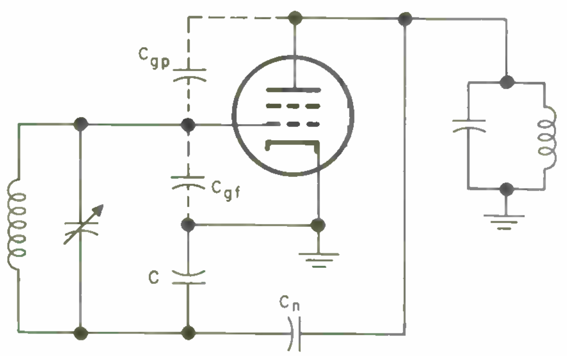


Figure 7-10. Modified Grid-Neutralized Amplifier

c. TESTING FOR PROPER NEUTRALIZATION

When a power amplifier stage is properly neutralized, the power output peaks at the same time the plate current dips. An indication of this simultaneous peak and dip is often the most convenient way of testing for proper neutralization. To perform such a test, the d-c cathode current, or plate current, of the neutralized stage is used to obtain an indication of plate current dip. The power output from the same stage or the grid drive to any succeeding stage is used to obtain an indication of power output. A power amplifier is usually checked for proper neutralization near the high-frequency end of its range where neutralization is more critical.

When the drive to a neutralized stage is so low that a plate current dip is not present, the best way to test for proper neutralization is by injecting a test signal into one circuit and checking for coupling of the signal into another circuit. In the modified grid neutralized circuit shown in figure 7-10, proper neutralization balances out coupling between the input tank circuit and the output tank circuit, but it does not remove all coupling between the plate circuit and the grid-to-cathode circuit. Therefore, a test signal injected into the plate circuit will result in grid-to-cathode signal even with proper neutralization. However, a test signal injected into the plate circuit will not result in a signal in the grid coil with proper neutralization. The presence of a signal in the grid coil can be detected by using an inductive coupling loop. This circuit can also be neutralized by inductively coupling an input signal into the input circuit and adjusting the neutralizing capacitor for minimum signal on the plate circuit.

7. R-F FEEDBACK CIRCUITS

a. INTRODUCTION

An r-f feedback is a very effective means of reducing distortion in a linear power amplifier. Twelve decibels of r-f feedback produces nearly twelve decibels of distortion reduction, and this distortion reduction is realized at all signal levels. However, voltage gain per stage is reduced by the amount of feedback employed, so that with 12 db of feedback the gain is reduced to one-quarter.

b. FEEDBACK AROUND ONE STAGE

Figure 7-11 shows a negative feedback circuit around a one-stage r-f amplifier. The voltage

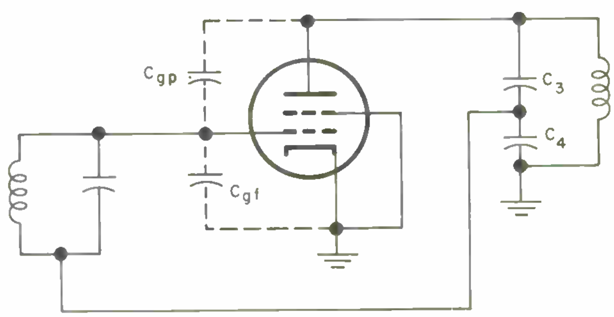


Figure 7-11. One-Stage Feedback with Neutralization

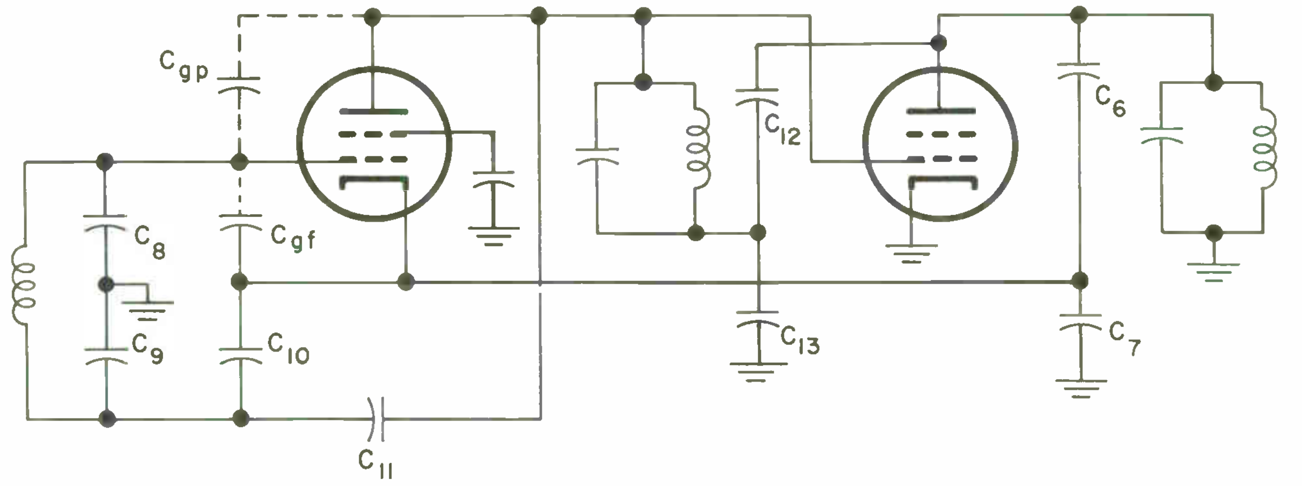


Figure 7-12. Two-Stage Feedback with Neutralization

developed across C_4 is introduced in series with the voltage developed across the grid tank circuit and is in phase opposition to it. The feedback obtainable with this circuit can be varied between zero and 100 per cent by properly choosing the values of C_3 and C_4. It is necessary to neutralize this feedback amplifier, the neutralization requirements being

$$\frac{C_{gp}}{C_{gf}} = \frac{C_3}{C_4}$$

To satisfy the neutralization requirement, it is usually necessary to add capacity from the plate to the grid.

Using this circuit presents a problem in coupling into the grid circuit. Inductive coupling is ideal, but the extra tank circuit complicates the tuning of the power amplifier if several cascaded amplifiers are used with feedback around each. The grid can be capacity coupled to a driver with a high source impedance, such as a tetrode or pentode. However, if this is done, feedback can not be used in the driver because it would cause the source impedance to be low.

c. FEEDBACK AROUND TWO STAGES

Feedback around two r-f stages has the advantage that more of the tube gain can be realized while nearly as much distortion reduction can be obtained. For instance, 12 db feedback around two stages provides about the same distortion reduction as 12 db around each of two stages separately. Figure 7-12 shows a negative feedback circuit around a two stage amplifier with each stage neutralized. The small feedback voltage required is obtained from the voltage divider C_6 and C_7. This feedback voltage is applied to the cathode of the first stage. The feedback divider can be left fixed for a wide frequency range since C_6 is only a few micromicrofarads. For example, if the combined tube gain is 160 and 12 db of feedback is desired, the ratio of C_7 to C_6 may be 400 uuf to 2.5 uuf. Either inductive input coupling or direct capacitive coupling may be used with this circuit, and any form of output coupling can be used.

It is necessary to neutralize the cathode-to-grid capacity of the first tube in the two stage feedback circuit to prevent undesirable feedback coupling to the input grid circuit. The relationship for the circuit which accomplishes this cathode-to-grid neutralization is

$$\frac{C_8}{C_9} = \frac{C_{gf}}{C_{10}}$$

To reduce the voltage across the input tank coil and minimize the power dissipated by the coil, the input circuit can be unbalanced by making C_9 up to five times C_8, as long as C_{10} is increased accordingly. The cathode-to-grid capacity of the first tube can be neutralized by injecting a test signal into the cathode of the tube. The neutralizing bridge is then adjusted for minimum signal as indicated by a detector which is inductively coupled into the input coil.

Except for tubes with very small plate-to-grid capacity, it is necessary to neutralize C_{gp} in both tubes. This neutralization for the second tube is realized by choosing C_{12} and C_{13} so that the ratio C_{12}/C_{13} equals the ratio C_{gp}/C_{gf} in the second tube.

If neutralization of C_{gp} is necessary for the first tube, it is obtained by satisfying the relationship

$$\frac{C_{gp}}{C_{11}} = \frac{C_{gf}}{C_{10}} = \frac{C_8}{C_9}$$

The screen and suppressor of the first stage should be grounded to keep the tank output capacity directly across the interstage circuit. This avoids common coupling between the feedback on the cathode and the interstage circuit.

In a two stage feedback amplifier, the voltage fed back to the cathode of the first stage must be in phase with the grid input signal, measured from grid to ground. If the feedback voltage is not in phase with the grid input signal, the resultant grid-to-cathode voltage increases as shown in figure 7-13. When the output circuit is properly tuned, the resulting grid-to-cathode voltage on the first tube is minimum which will make the voltage across the interstage tank circuit minimum also.

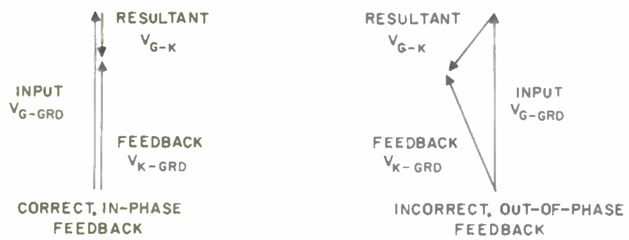


Figure 7-13. Vector Relationship of Voltages for Two-Stage Feedback

8. AUTOMATIC LOAD CONTROL

Automatic load control is a means of keeping the signal level adjusted so that the power amplifier works near its maximum power capability without being overdriven on signal peaks. In AM. systems, it is common to use speech compressors and speech clipping to perform this function. However, in an SSB system these methods are not equally useful because the peaks of the SSB signal do not necessarily correspond with the peaks of the audio signal. Therefore, the most effective means of control is obtained by a circuit which receives its input from the envelope peaks in the power amplifier and uses its output to control the gain of the exciting signal. Such a circuit is an automatic load control (alc) circuit.

Figure 7-14 is a simplified schematic of an alc circuit. This circuit uses two variable gain stages of remote cutoff tubes, such as a 6BA6, operating very similarly to the i-f stages of a receiver with automatic volume control. The grid bias voltage of the variable gain amplifiers is obtained from the alc rectifier connected to the power amplifier plate circuit. The capacity voltage divider steps down the r-f voltage from the power amplifier plate to about 50 volts for the rectifier. A large delay bias is used on the rectifier so that no reduction of gain takes place until the signal level is nearly up to full power capability of the power amplifier. The output of the alc rectifier passes through RC networks to obtain the

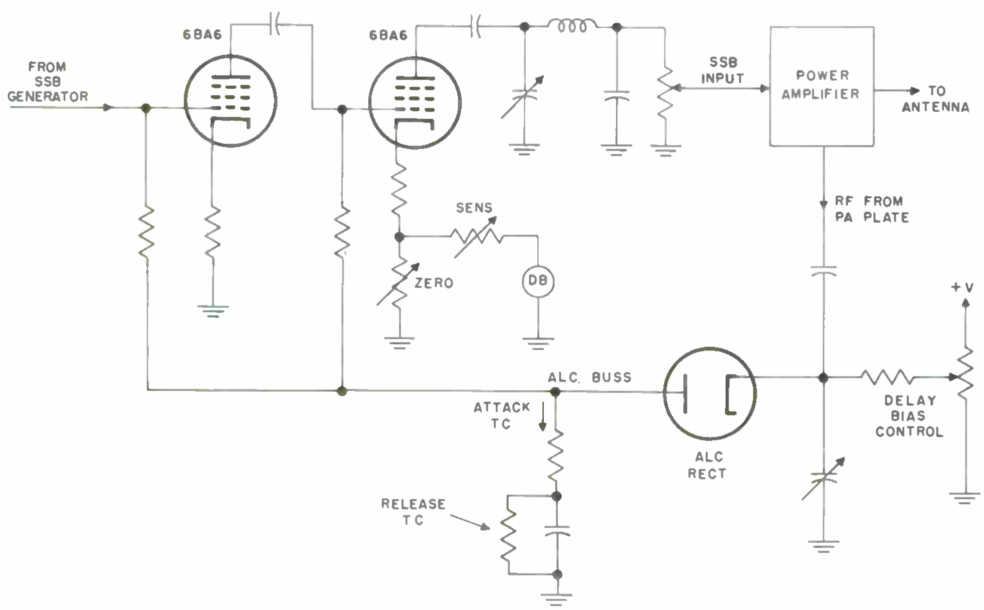


Figure 7-14. Automatic Load Control Circuit

desired attack and release times. Usually a fast attack time, about two milliseconds, is used for voice signals so that the gain is reduced rapidly to remove the overload from the power amplifier. After a signal peak passes, a release time of about one-tenth second returns the gain to normal. A meter calibrated in decibels of compression is used to adjust the gain for the desired amount of load control.

In single channel speech transmission, the alc circuit performs the function of a speech compressor. To do this a range of 12 db is usually provided with control maintained on input peaks as high as 20 db above the threshold of compression. Since the signal level should be fairly constant through the preceding SSB generator, it is unlikely that more than a 12 db range of the alc would be useful. If the signal level varies more than 12 db for the SSB generator, a speech compressor in the input audio amplifier is usually used to limit the range of the signal fed into the SSB generator.

Figure 7-15 shows the effectiveness of the alc circuit in limiting the output signal to the capabilities of the linear power amplifier. An adjustment of the delay bias will put the threshold of compression at the desired level.

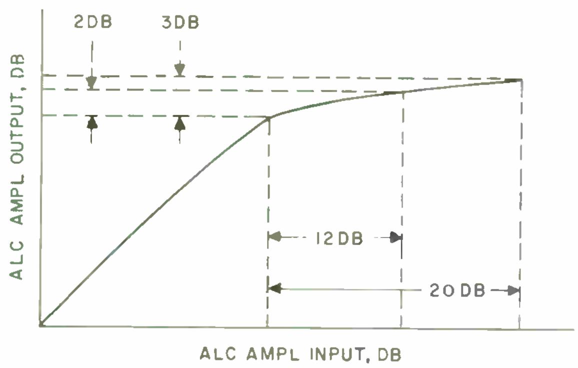


Figure 7-15. Automatic Load Control Performance Curve

9. LINEAR POWER AMPLIFIER TUNING

a. INTRODUCTION

When a power amplifier is operated class C, a pronounced plate current dip and grid current peak are fairly accurate indications of proper tuning. In a linear power amplifier, the use of these indications are limited. For instance, in a class A amplifier there will be no plate current dip; therefore, the class A amplifier output circuit must be tuned for an indication of maximum input to the next stage. In class AB amplifiers, the plate current dip is not always readily detected. This does not mean that conventional tuning procedures will not properly tune a linear amplifier, but tuning a linear amplifier with conventional procedures is much more exacting. One procedure commonly used is to increase the drive to a stage in order to obtain a good plate current dip indication.

In low Q tank circuits, the point of plate current dip is not a true indication of exact resonance because the plate current dip occurs at maximum impedance rather than when the tank circuit is pure resistive. This is especially true for Pi networks and Pi-L networks. For instance, in a network with a Q of ten, the phase angle at maximum impedance is about 17° from unity. Tuning this far from resonance in a linear amplifier with r-f feedback can be much more serious than in a class C amplifier because the phase angle of the feedback voltage is critical.

b. PHASE COMPARISON TUNING

Use of a phase comparator circuit to compare the phase of the input signal to the phase of the output signal affords the most sensitive means of tuning a linear power amplifier stage. This circuit employs a phase discriminator, such as shown in figure 7-16, for phase comparison. A balanced, push-pull voltage is obtained through a 90° phase-shifting network to provide the voltage $E_a + E_b$. In the figure shown, $E_a + E_b$ is in phase with the current in the inductive branch of the grid tank circuit. Since the current in the inductive branch is 90° out of phase with the voltage across the tank circuit, the induced voltage $E_a + E_b$ is also 90°

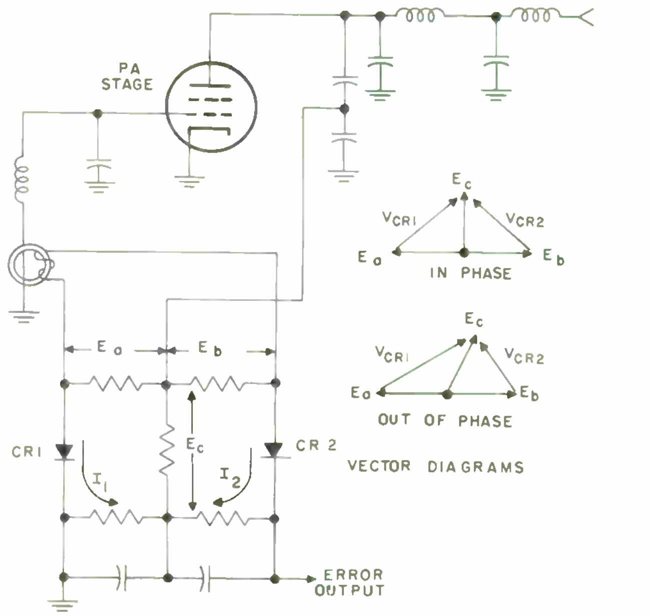


Figure 7-16. Phase Discriminator for PA Tuning

out of phase with the voltage across the tank circuit. From the output of the stage, E_c is obtained. When E_c is exactly 90° out of phase with E_a and E_b, the voltages across the two crystals, CR1 and CR2, are equal in magnitude. Then, the d-c currents in the diode loads are equal and flowing in opposite directions which produces zero output. When E_c is not exactly 90° out of phase with E_a and E_b, the voltages across the two crystals are unequal in magnitude. This will cause the d-c currents in the diode loads to be unequal which will produce an output. The error signal derived from this circuit can be used to operate a zero-center meter for manually tuning the output circuit. When tuned for zero meter indication, the output voltage is exactly 180° out of phase with the input voltage, the condition for true resonance.

The phase discriminator can also be used to obtain an error signal for servo tuning the stage. However, for servo tuning, coarse positioning information is necessary because the phase discriminator responds to harmonic tuning points and because there is insufficient output from the phase discriminator over much of the frequency range. This coarse positioning information can be provided with a coarse follow-up potentiometer which receives information from the exciter frequency control circuits. Such a system requires that the master potentiometer track the tuning curves of the amplifier tank circuits and that sequencing controls be used to initiate and halt coarse positioning at the proper times. Pretuning information can also be derived from the exciter r-f output signal by using a coarse discriminator circuit, such as is shown in figure 7-17. This circuit is a series RC network fed with r-f voltage from the exciter. A servo system then drives the capacitor in the RC bridge to produce zero error signal at the same time it positions a master potentiometer. A second, tuning servo then drives a follow-up potentiometer which is wound to cause the tuning servo to track the tuning curve of the amplifier tank circuit. To automatically tune the amplifier, the error signals from the phase discriminator and the course discriminator can be combined to operate a single servo. The servo system will then operate over the whole frequency range and have a precise zero error signal position, as shown in figure 7-18.

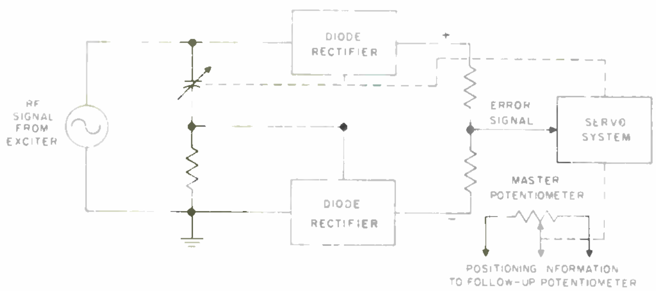


Figure 7-17. Coarse Discriminator for PA Tuning

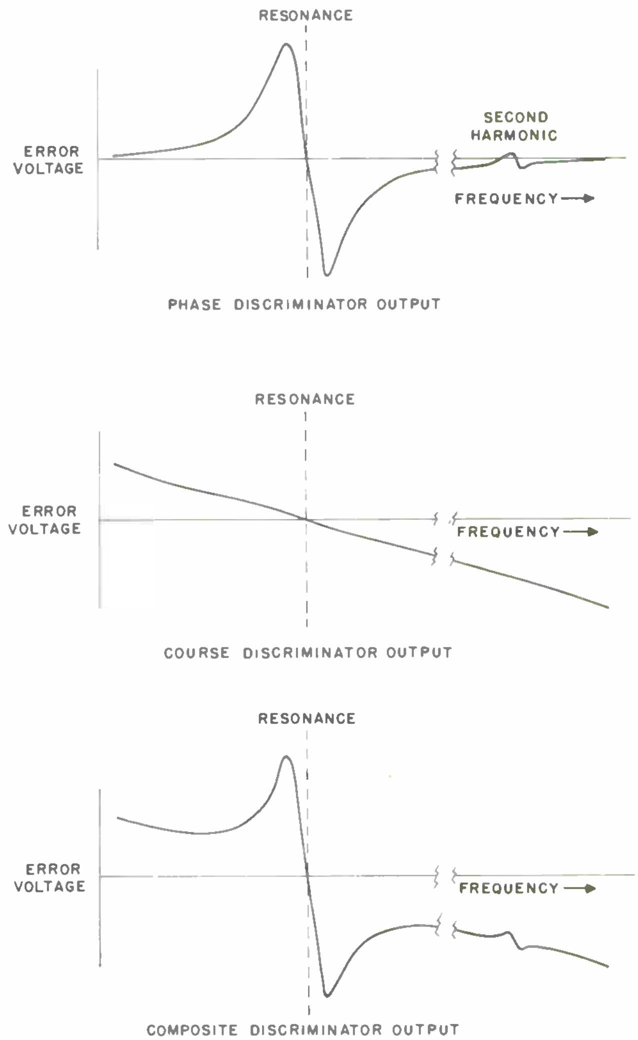


Figure 7-18. Discriminator Output Curves for PA Tuning

c. LOADING COMPARATOR CIRCUIT

Since the voltage gain of a tube is dependent upon the load resistance, a loading comparator circuit, as shown in figure 7-19, can be used to determine proper loading. The loading comparator is designed so that a predetermined ratio between positively rectified grid voltage and negatively rectified plate voltage produces zero error signal output. The power amplifier is then manually or automatically loaded until the error signal output goes to zero. The clamping diode is required so that the circuit will maintain control under light load when the amplifier is driven into plate saturation. In plate saturated operation, the rectified grid voltage will continue to rise with reduced loading while the rectified plate voltage remains relatively constant. This will cause the circuit to lose its sense of direction and result in reducing the load even further. To maintain the sense of direction under this condition, the clamping diode prevents the rectified grid voltage from exceeding a voltage which is proportional to plate current. Therefore, in plate saturated operation, which is similar to class C operation, loading is determined by the ratio of plate current to r-f plate voltage. Proper compromise of the magnitude of the plate, grid, and clamping signal voltages results in a loading comparator that produces proper loading information regardless of the operating conditions, provided the plate circuit is held at resonance.

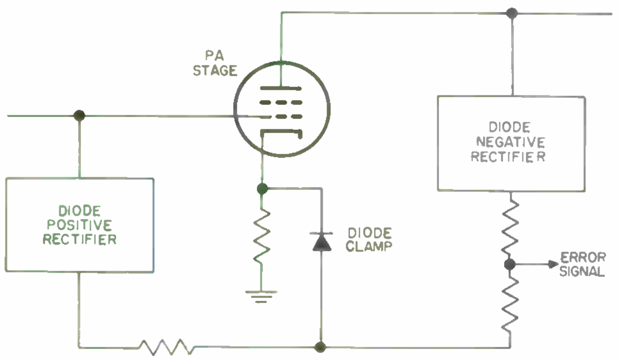


Figure 7-19. Loading Comparator for PA Loading

d. ANTENNA TUNING AND LOADING

The output network of a variable frequency transmitter must be capable of tuning and loading into a transmission line which presents different impedances at different frequencies. This requires output networks which will match a wide range of load impedances with the power amplifier output. In fixed-station equipment, the power amplifier usually works into a transmission line and antenna designed so that the load impedance presented to the amplifier varies over only a limited range. In this case the output network is designed to match the load impedance directly. In mobile and airborne equipment, the power amplifier usually works into a coaxial transmission line terminated with a wide variety of antennas that present unwieldy terminating impedances. In this case an antenna coupler is used which can be located in one of two positions: (1) It can be located near or in the transmitter to provide proper coupling between the transmitter output network and a transmission line which is terminated with a mismatched antenna; (2) It can be located near the antenna to terminate the transmission line properly and provide coupling for maximum power transfer to the antenna. The first method is commonly used in mobile transmitters, and the second method is used in airborne transmitters.

Two power amplifier control functions are required to match properly the load impedances presented to the power amplifier with the power amplifier network. One is a phasing control, or tuning control, which will balance out undesirable reactance and make the load resistive or as nearly resistive as is possible. The other is a load control which will provide the proper terminating impedance. Figure 7-20 shows

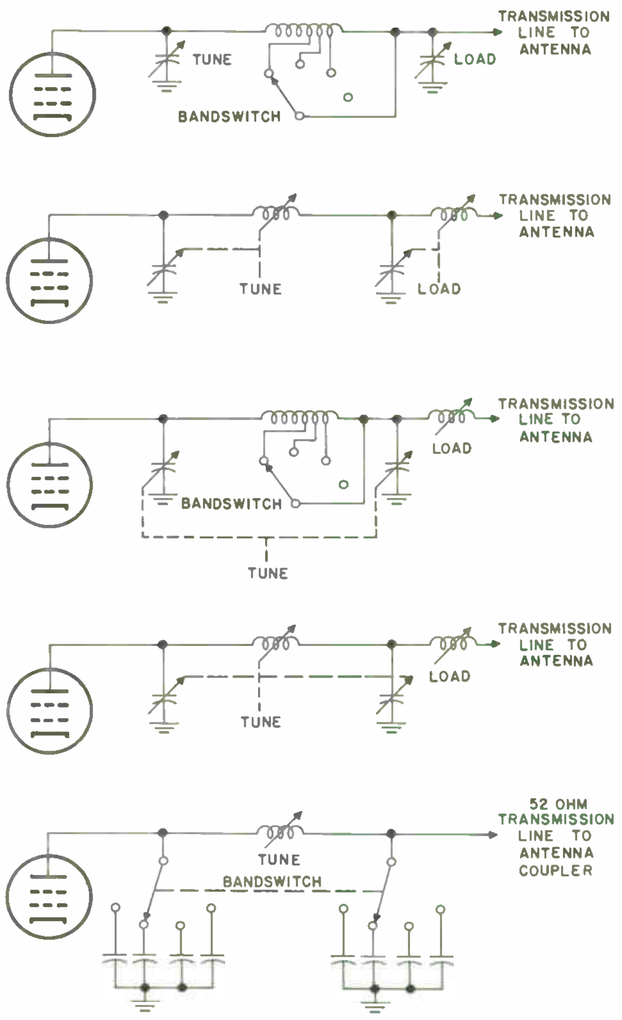


Figure 7-20. Output Network Tuning and Loading Controls

several ways that the output network components can be ganged to provide tuning and loading with two controls. The tuning control is adjusted to produce a plate current dip, which indicates maximum impedance. For more precise tuning and automatic tuning, the phase discriminator circuit is used. The loading control is adjusted to produce a pre-established value of grid voltage and plate current or, in some cases, a pre-established value of screen current and plate current. For more precise loading and automatic loading, the loading comparator circuit is used. The loading and tuning circuits must be so designed that the controls will not lose sense of direction under any circumstances. This is absolutely essential for automatic loading and tuning and is highly desirable for manual loading and tuning.

10. POWER SUPPLIES FOR POWER AMPLIFIERS

Fixed transmitters up to 1 kw usually use a single-phase a-c power source, and larger fixed transmitters usually use a three-phase a-c power source. Mobile equipment may operate from a 6-volt to 28-volt d-c power source using dynamotors or vibrator power supplies to obtain the required high voltages. Airborne equipment usually uses the 400-cycle a-c power source of the aircraft.

In addition to supplying the required d-c voltage and output current, the power supply must have adequate d-c regulation, good dynamic regulation, and low ripple or noise output. Most high-voltage power amplifiers have a varying load characteristic so that good d-c regulation is essential. To reduce ripple and noise, high-voltage filters are used between the rectifier circuit and the power supply load. The filter chokes place a high impedance between the rectifier and the load, making large capacitors necessary in the output side of the filter. These output capacitors supply the rapid variations in load current which are impeded by the filter choke. This is particularly necessary in high-voltage power supplies for linear power amplifier stages.

Vacuum rectifiers can be used for small, low-voltage power supplies which have relatively constant load. Gas-type rectifiers are required where better regulation is necessary. The mercury-vapor rectifier is the most common gas-type rectifier used because it has long life when properly operated. Operating a mercury-vapor rectifier above or below its rated temperature, changes the vapor pressure in the tube and reduces its peak-inverse-voltage capability, making the rectifier more susceptible to arc-back. Equipment which is subject to wide ambient-temperature variations, such as military equipment, uses inert gas rectifiers such as the 3B28 and 4B32. These tubes can be operated in ambient temperatures from -75°C to +90°C, which is frequently a necessary feature. The tube life of the inert gas rectifier, however, is only about one-third the tube life of an equivalent mercury-vapor rectifier. Metallic rectifiers, such as selenium and copper oxide, are frequently used in power supplies delivering less than 100 volts for relay operation, etc.

Rectifier tube life is increased by operating the filaments 90° out of phase with the plate voltage. This minimizes the difference in voltage from each end of the filament to the plate and allows a more uniform emission over the entire filament. A 60° phase difference between the filament and the plate voltage is often used when it is more easily obtained because almost the full advantage of quadrature operation is realized. Tube ratings of some of the larger rectifier tubes are increased for quadrature operation.

Transient voltages and currents which far exceed the steady-state values occur in power supplies when the supply is energized. If these transient peaks exceed the peak-inverse-voltage rating of the tube, an arc-back may result. For this reason, rectifier tubes are often operated so that the normal peak-inverse-voltage does not exceed one-half of the rated peak-inverse-voltage. If this is not possible, a step-start circuit is used which starts the transformer with resistors in series with the primary. After a short time delay, these resistors are shorted out. Some high-voltage rectifiers are started with a resistor in series with the filter capacitor, with the resistor being shorted out after a short time delay. This prevents a transient due to the charging current required to bring the voltage up on the filter capacitor. The added resistance in the circuit prevents excessive current in the rectifier.

11. CONTROL CIRCUITS

Power amplifier control circuits must perform three functions; (1) they must supply circuit control, (2) they must provide equipment protection, and (3) they must provide personnel protection. In small transmitters, the control circuits may consist of nothing more than an on-off switch to supply heater power and a push-to-talk button to apply plate voltage and put the transmitter on the air. In larger equipment, push buttons are usually used to initiate a certain sequence of relay operations which complete a function in the proper manner. Many transmitters, particularly those suitable for remote control, are capable of complete energization from a single push-button control.

The filament on-off switch, or push button, initiates a sequence of functions that applies power to the filaments, starts the cooling system, and energizes time delay circuits that make the power amplifier ready for the application of plate power. When operated to the off position, the power amplifier is shut down.

Filaments of high-power amplifier tubes are energized separately, and, in the case of mercury-vapor tubes, a time delay allows warmup time. The blower is started at the beginning of the starting sequence because the life and reliability of many components is greatly dependent upon operating temperature control. Air interlocks prevent the application of power to high-power tubes before cooling air is present and a blower-off delay maintains cooling air after shutdown. In various power amplifier stages, it is essential that bias voltage be applied before plate or screen voltage is applied. This requires sequencing the application of the bias voltage and the plate voltage as well as interlocks between the two so that the loss of bias voltage will result in removing the plate voltage.

Power amplifier control circuits are sequenced and interlocked so that everything else must be on and functioning before the high-voltage plate transformer is energized. Certain power tetrodes require that screen voltage be applied simultaneously with plate voltage to prevent excessive screen dissipation. To prevent high current and high-voltage transients, plate voltage is often applied through step-start circuits which place resistors for a short time in the power supply circuit.

Medium-power and high-power tubes are nearly always protected from excessive plate current by overload relays. These relays remove the high-voltage primary power if the plate current exceeds a preset value. Many overloads that occur during normal operation will clear themselves when the high voltage is removed. For this reason, large power amplifiers are usually provided with an overload recycle circuit. This circuit brings the power amplifier back on after an overload. If the overload reoccurs, the power amplifier will again shut down. The number of recycles before shutdown can generally be preset with a recycle counting switch.

12. TUBE OPERATING CONDITIONS FOR R-F LINEAR POWER AMPLIFIERS

a. GENERAL

SSB amplifiers provide linear amplification and operating conditions similar to those of audio amplifiers. There is one fundamental difference, however, between audio and r-f linear amplifiers. This is that the input and output voltages of a tuned r-f amplifier are always sine waves because the tuned circuits, if they have adequate Q, make them so. Therefore, the distortion in an r-f amplifier results in distortion of the SSB modulation envelope and not in the shape of the r-f sine wave. This can be restated that distortion in an r-f linear amplifier causes a change in gain of the amplifier when the signal level is varied. The greatest difference between an audio amplifier and an r-f linear amplifier is in the grid driving power requirements when driving into grid current. In the audio amplifier, the driver must supply all of the instantaneous power required by the grid at the peak of the grid swing. To deliver this peak power, the audio driver must be capable of delivering average sine wave power equal to one-half of the peak power. In an r-f linear amplifier, the tank circuit averages the power of the r-f cycle due to its "flywheel" effect so that the driver need only be capable of delivering the actual average power required, and not the peak. With these reservations in mind, examination of the audio or modulator data of a tube will give a good idea of its r-f linear power amplifier operating conditions.

b. CLASS A R-F LINEAR AMPLIFIERS

In low-level amplifiers, where the output signal voltage is less than 10 volts, small receiving type tubes, such as the 6AU6, are very satisfactory for class A service. For voltage levels above 10 volts, the 4X150A is the best choice for class A operation because it has short leads, low plate-to-grid capacitance, and high transconductance. Class A amplifier tubes should be operated in as linear a portion of the plate characteristic curves as is practical. This can be done by inspecting the plate characteristic curves of the tube. Usually the static plate current which results in near maximum plate dissipation is the best. The maximum output voltage should be kept to about one-tenth the d-c plate voltage or less to obtain signal-to-distortion ratios of 50 db or better. The d-c plate voltage regulation for class A operation is seldom of importance and cathode bias and screen dropping resistors are commonly used. Even with tubes such as the 6AU6 and the 4X150A which have short leads and low grid-to-plate capacitance, it is desirable to load the input and output circuits to 5000 ohms when operating up to 30 mc.

c. CLASS AB R-F LINEAR AMPLIFIERS

In the power range from 2 watts to 500 watts, class AB_1 is normally used. This class of operation is very desirable because distortion due to grid current loading is avoided and because high power gain can be obtained. At present, tubes are not available which will give low distortion with good plate efficiency operating class AB_1 at power levels above 500 watts. Therefore, for higher power levels class AB_2 operation is used.

For class AB operating conditions with a given screen voltage and given plate load, there is one value of static plate current which will give minimum distortion. The optimum value of static plate current for minimum distortion is determined by the sharpness of cutoff of the plate current characteristic. Grid bias is then set to produce the optimum static plate current.

This optimum point is determined from the load line on a set of constant current plate current curves. Values obtained from this curve are then plotted to obtain the plate current vs grid voltage curve shown in figure 7-21. This curve is the dynamic characteristic of the tube. By projecting the most linear portion of the curve to intersect with the zero plate current line, the grid bias is determined. This point of intersection is often referred to as the projected cutoff. The static plate current which will flow with this grid bias is the proper static plate current for minimum distortion. This procedure is used in audio amplifier design and is nearly correct for r-f linear amplifier design. Perhaps a more accurate procedure for determining the proper bias for r-f amplifiers is to choose the point Q so that the slope of the curve at Q is one-half the slope of the major linear portion. This will allow the amplifier to operate class A with small signals and deliver power over both halves of the cycle. With a large signal, the tube delivers power over essentially one-half the cycle. Then the change in plate current relative to plate voltage swing over half the cycle will be half as much for small signals as it is for large signals and linear operation is obtained at all signal levels.

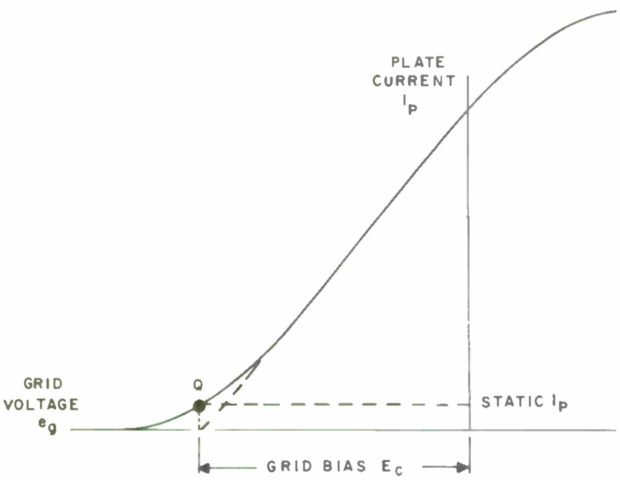


Figure 7-21. Optimum Static Plate Current for Linear Operation

The screen voltage of a tetrode tube has a very pronounced effect on the optimum static plate current, because the plate current of a tube varies approximately as the three-halves power of the screen voltage. For example, raising the screen voltage from 300 to 500 volts will double the plate current. The shape of the dynamic characteristic will stay nearly the same, however, so that the optimum static plate current for minimum distortion is also doubled. A practical limit is reached because high static plate current causes excessive static plate dissipation.

In practice, it is found that the static plate current determined by the above method is so high that plate dissipation is near or beyond the maximum rating of the tube when using desired d-c plate voltage. For example, one of the better medium power triodes for linear amplifier service, the 3X2500A3, requires approximately .5 ampere of plate current for minimum distortion. Using a desirable plate voltage of 5000 volts, static plate dissipation is 2500 watts, which is the maximum rated plate dissipation for the tube. For this reason, it is often necessary to operate the tube below the optimum static plate current, which can be done without causing appreciable distortion. In tetrodes, the optimum static plate current is a function of screen voltage, and the high screen voltages required for class AB_1 operation usually require an excessive amount of plate current for minimum distortion. A choice must then be made between operating the tube at lower than optimum static plate current or using a lower screen voltage and driving the tube into the grid current region, a second principal cause of distortion.

d. ESTIMATING TUBE OPERATING CONDITIONS

The operating conditions of a tube operating class AB in an r-f linear power amplifier can be estimated from the load line on a set of constant plate current curves for the tube, as shown in figure 7-22.

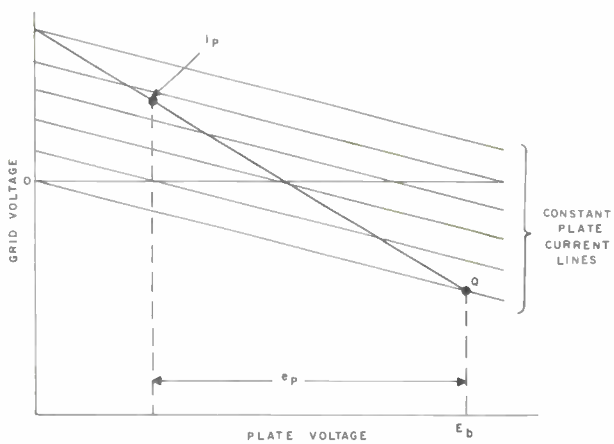


Figure 7-22. Graphical Determination of Tube Operating Conditions

From the end point of the load line, the instantaneous peak plate current, i_p, and the peak plate voltage swing, e_p, can be established. From these two values, the principal plate characteristics can be estimated by using the following relationships for a single-frequency test signal:

d-c plate current, $I_B = \frac{i_p}{\pi}$

plate input watts, $P_{in} = \frac{i_p E_B}{\pi}$

average output watts and PEP, $P_o = \frac{i_p e_p}{4}$

plate efficiency, $Eff = \frac{\pi e_p}{4E_B}$

For a standard two-frequency test signal the relationships are:

d-c plate current, $I_B = \frac{2ip}{\pi^2}$

plate input watts, $P_{in} = \frac{2i_p E_B}{\pi^2}$

average output watts, $P_o = \frac{i_p e_p}{8}$

PEP watts, $P_o = \frac{i_p e_p}{4}$

plate efficiency, $Eff = \left(\frac{\pi}{4}\right)^2 \frac{e_p}{E_B}$

An actual tube with moderate static plate dissipation will have operating characteristics similar to those shown in figure 7-23 for the single-tone and two-tone signals. Plate dissipation and efficiency at maximum signal level are affected little by even rather high values of static plate dissipation. In practice, the peak plate swing is limited to something less than the d-c plate voltage in order to avoid excessive grid drive, excessive screen current, or operation in the nonlinear plate current region. Most tubes operate with an efficiency in the region of 55 to 70 per cent at peak signal level.

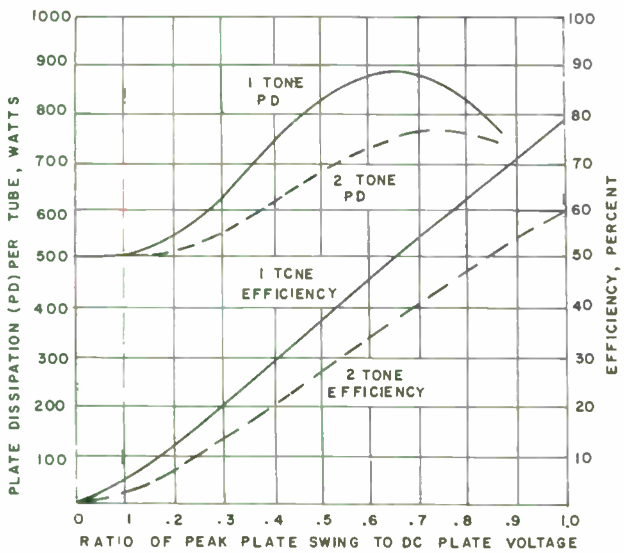


Figure 7-23. Efficiency and Plate Dissipation for Class AB Operation

13. DISTORTION

a. CAUSES OF R-F LINEAR POWER AMPLIFIER DISTORTION

The principal causes of distortion are nonlinearities of the amplifier tube plate current characteristic and grid current loading. In order to confine distortion generation to the last stage or two in a linear power amplifier, all previous stages are operated class A.

The generation of distortion products due to the nonlinear characteristics of the amplifier tube can be derived from the transfer characteristic of the tube, also called the dynamic characteristic, as shown in figure 7-24. The shape of this curve and the choice

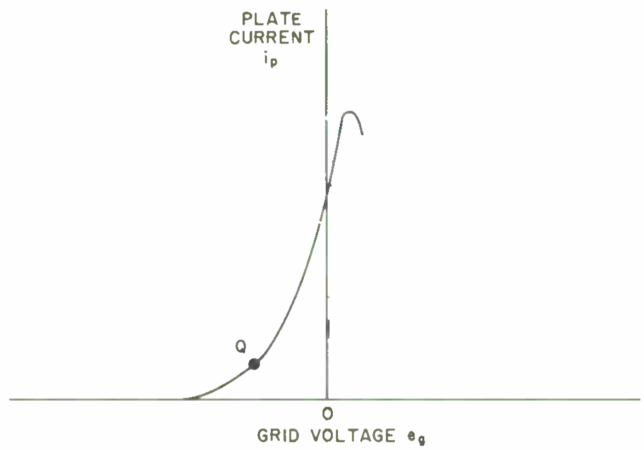


Figure 7-24. Typical PA Tube Transfer Characteristic Curve

of the zero signal operating point, Q, determine the distortion which will be produced by the tube. A power series expressing this curve, written around the zero signal operating point, contains the coefficients of each order of curvature, as shown in the following expression:

$$i_p = k_O + k_1 e_g + k_2 e_g^2 + k_3 e_g^3 + k_4 e_g^4 + k_5 e_g^5 + \ldots k_n e_g^n$$

In this expression, i_p represents instantaneous plate current, k_1, k_2, etc, the coefficients of their respective terms, and e_g the input grid voltage signal. The values for the coefficients are different for every power series written around different zero signal operating points. By making the input signal, e_g, consist of two equal amplitude frequencies with a small frequency separation, the distortion products of concern in linear amplifiers can be obtained. Figure 7-25 shows the spectrum distribution of the stronger plate current components. It is seen that tuned circuits can filter out all products except those which are near the fundamental input frequencies. This removes all of the even-order intermodulation products and the harmonic products. The odd-order intermodulation products fall close to the original frequencies and

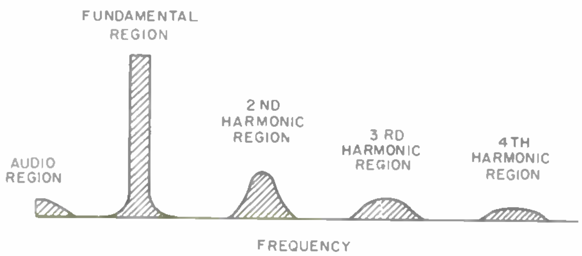


Figure 7-25. Spectrum Distribution of Products Generated in PA Stage

cannot be removed by selective circuits. Figure 7-26 shows these odd-order products which fall within the passband of selective circuits. The inside pair of intermodulation distortion products are third-order, the next fifth-order, seventh-order, etc. The first and most important means of reducing distortion, then, is to choose a tube with a good plate characteristic and choose the operating condition for low odd-order curvature (see paragraph 12c, chapter 7).

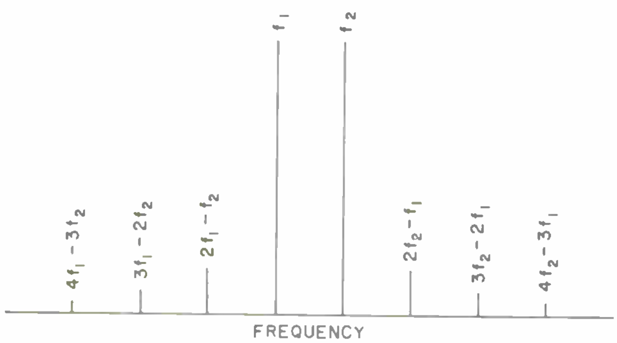


Figure 7-26. Odd-Order Intermodulation Products Causing SSB Distortion

The nonlinearity caused by grid current loading is a function of the regulation of the grid driving source. In general, this regulation with varying load is poor in linear amplifiers. It is common practice to use swamping resistors in parallel with a varying grid load so that the resistance absorbs about ten times the power that the grid of the tube requires. This provides a low, constant driving source impedance and improves linearity at the expense of increased driving power.

The instantaneous plate current of all tubes drops off and causes distortion when the instantaneous plate voltage is low. The main reason for this drop is that current taken by the grid and screen is robbed from the plate. In all but a few transmitting tetrodes, the plate can swing well below the screen voltage before plate saturation occurs. However, when the plate swings into this region, the instantaneous plate current drops considerably. If distortion requirements are not too high, the increased plate efficiency realized by using large plate swings can be realized. However, to minimize distortion, the allowable plate swing may have to be reduced.

b. DISTORTION REDUCTION

There is a need for reduced levels of intermodulation distortion from r-f linear power amplifiers used in SSB systems. This need exists not because the distortion noticeably reduces the intelligibility of the SSB signal, but because distortion products outside of the channel width necessary for transmission of intelligence interferes with adjacent channel transmission. The distortion of some of the early SSB power amplifiers was so great that voice channels were placed a full channel width apart to avoid adjacent channel interference. Recent power amplifier developments permit adjacent channel operation, using power amplifiers with signal-to-distortion ratios of from 35 db to 40 db. However, power amplifiers with signal-to-distortion ratios of from 45 db to 50 db would further increase the utility of single sideband.

There are two basic means of reducing distortion to levels better than is obtainable from available tubes. These are r-f feedback and envelope distortion canceling [1] An r-f feedback is very effective and quite easy to obtain (see paragraph 7, chapter 7). Ten decibels of r-f feedback will produce nearly 10 db of distortion reduction which is realized at all signal levels. Envelope distortion canceling has an inherent weakness because it depends upon envelope detection for its feedback signal. This means that distortion canceling must be instantaneous to be perfect. Since some delay is inherent in the envelope detector and feedback loop, the effectiveness of this circuit depends upon how short the time delay can be made. Development work indicates that a combination of r-f feedback

[1] W. B. Bruene, "Distortion Reduction Means for Single-Sideband Transmitters," IRE Proceedings, Dec. 1956

and envelope distortion canceling will provide more distortion canceling than either method separately. Using 10 db of r-f feedback around all three stages of a 20-kw PEP power amplifier, and a signal synthesized from the input envelope to grid modulate the first stage, a better than 50 db signal-to-distortion ratio has been obtained for all distortion products at any signal level up to the 20-kw PEP.

c. LINEARITY TRACER

The linearity tracer consists of two SSB envelope detectors, the outputs of which connect to the horizontal and vertical inputs of an oscilloscope, as shown in figure 7-27. A two-tone test signal is normally used to supply an SSB modulation envelope, but

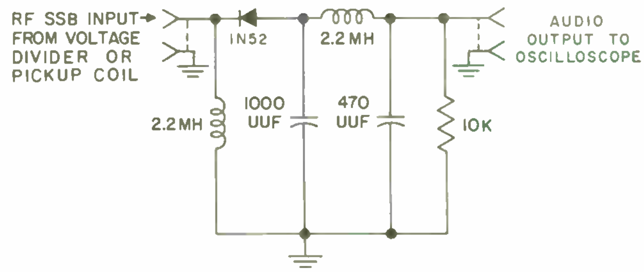


Figure 7-28. Envelope Detector, Schematic Diagram

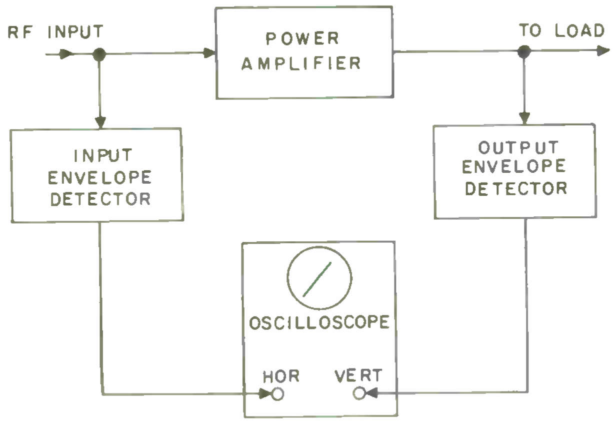


Figure 7-27. Linearity Tracer, Block Diagram

any modulating signal that provides from zero to full amplitude can be used. Even speech modulation gives a satisfactory trace. This instrument is particularly useful for monitoring the signal level and clearly shows when the amplifier is overloaded. It can also serve as a voltage indicator which can be useful for making tuning adjustments. The linearity trace will be a straight line regardless of the envelope shape if the amplifier has no distortion. Overloading, inadequate static plate current, and poor grid circuit regulation are easily detected with the linearity tracer. The instrument can be connected around any number of power amplifier stages, or it can be connected from the output of the SSB generator to the power amplifier output to indicate the over-all distortion of the entire r-f circuit.

A circuit diagram of an envelope detector is shown in figure 7-28. Any type of germanium diode may be used for the detector, but the diodes in each of the two required envelope detectors must be fairly well matched. Using matched diodes cancels the effect

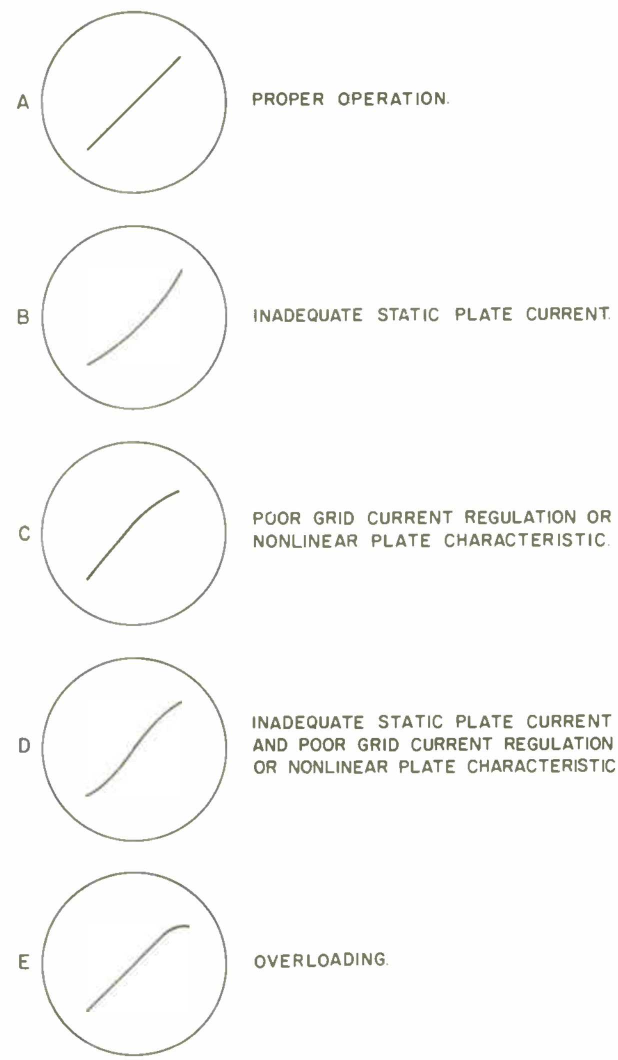


Figure 7-29. Typical Linearity Traces

on the oscilloscope of their nonlinearity at low signal levels. A diode load of from 5K to 10K minimizes the effect of diode differences. Operation of both detectors at approximately the same signal level is important so that diode differences will cancel more exactly. It is desirable to operate the envelope detectors with a minimum of 1 volt input to further minimize diode differences. It is a convenience to build the detector in a small shielded enclosure with coaxial input and output connectors. Voltage dividers can be similarly constructed so that it is easy to patch in the desired voltage stepdown from the voltage sources. A pickup coil on the end of a short coaxial cable can be used instead of voltage dividers to obtain the r-f input signal.

The frequency response and phase-shift characteristics of the oscilloscope vertical and horizontal amplifiers should be the same and should be flat to at least twenty times the frequency difference of the two test tones. Excellent high-frequency characteristics are necessary because the rectified SSB envelope contains harmonics to the limit of the envelope detector's ability to detect them. Inadequate frequency response of the vertical amplifier may cause a little foot to appear at the lower end of the trace. If the foot is small, it may be safely neglected. Another effect which may be encountered is a double trace, but this can usually be corrected with an RC network between one detector and the oscilloscope.

The best way to test the linearity tracer is to connect the inputs of the envelope detectors in parallel. A perfectly straight diagonal trace on the oscilloscope will result if everything is working properly. One of the detectors is then connected to the other source through a voltage divider which will not result in appreciable change in the setting of the oscilloscope gain controls.

Figure 7-29 shows some typical linearity traces which might be observed in linear power amplifier operation. Figure 7-29a indicates proper linear operation. Inadequate static plate current in class A amplifiers, class AB amplifiers, or mixers will result in the trace shown in figure 7-29b. This condition can be remedied by reducing the grid bias, raising the screen voltage, or lowering the signal level through mixers and class A amplifiers. The trace shown in figure 7-29c is caused by poor grid circuit regulation when grid current is drawn, or by nonlinear plate characteristics of the tube at large plate swings. This can be remedied by using more grid swamping or lowering the grid drive. The trace shown in figure 7-29d is a combination of the traces shown in b and c. The trace shown in figure 7-29e is caused by overloading the amplifier. It can be remedied by lowering the signal level.

CHAPTER 8
TEST EQUIPMENT AND TECHNIQUES

1. GENERAL

When a transmitter is operated, three characteristics of the output signal are of prime importance. These are:

(1) Carrier Frequency

(2) Signal level

(3) Undesired power output

The carrier frequency is that frequency which designates the position in the spectrum occupied by the band of frequencies required to transmit the intelligence. Desired signal level is the r-f power confined to those discreet frequencies within this band that are required to transmit the intelligence. Undesired power is the r-f power at frequencies both within this band and outside it, that are not necessary to the transmission of the intelligence and, therefore, interfere with the transmitted signal as well as with other communication channels.

2. FREQUENCY MEASUREMENT

Two commonly used methods of frequency measurement are:

(1) Frequency counter and converter

(2) Receiver and frequency standard

A typical frequency counter for measurements up to 100 megacycles is the Hewlett-Packard 524B with the Hewlett-Packard 525A converter installed. Higher frequency measurements can be made with other converters which can be installed in place of the 525A. The frequency resolution of this instrument is variable by a switch on the front panel to as low as .1 cps, but in normal use, a resolution of ±1 cps is sufficient. The period of the count is the reciprocal of the resolution. For example, for a resolution of 0.1 cps the count period is 10 seconds and for a resolution of 1 cps the count period is 1 second. The readout accuracy of the counter is ±1 in the last digit so that with a resolution of 0.1 cps the accuracy of the reading displayed is ±0.1 cps and with a resolution of 1 cps the accuracy of the reading is ±1 cps. If the frequency to be counted is 10 mc and the resolution is 0.1 cps, then the readout accuracy is plus or minus 1 part in 10^8. The over-all accuracy of the measurement depends upon the accuracy of the time base standard. For example, if the oscillator producing the time base has an accuracy of 1 part in 10^8 then with a readout accuracy of 1 part in 10^8, the over-all accuracy of measurement is 2 parts in 10^8. The frequency counter is designed to count the frequency of a single sine wave of an amplitude of about 1 volt. It will not give a true indication of frequency in the presence of a complex wave since the digitalizing of the data obtained from the incoming signal is accomplished by counting the number of times the instantaneous voltage crosses a given absolute value. Short term errors existing during the period of the count are integrated by the counter, and the average error is included in the frequency display at the end of each count. The internal frequency standard which provides the time base for the count in the HP525A has a stability of approximately 1 part in 10^6 per day, but provision is made for connection of an external standard if higher accuracy is desired.

A second method of frequency measurement using a receiver and associated frequency standard is less accurate than the frequency counter method. If a receiver such as the Collins 51J is used, the internal frequency standard can be calibrated to a standard of known accuracy, and by using the bfo, the kilocycle dial can be calibrated at adjacent, integral 100 kc points on each side of the frequency at which the measurement is to be made. The accuracy of the measurement made with this receiver will be in the order of ±250 cps.

A highly accurate adjustment between two frequency standards can be made by mixing products of the two standards into a receiver and adjusting one standard to agree with the other by observing the beat frequency on the receiver S meter. Best results are obtained when the effect of the two frequencies are equal in amplitude at the detector in the receiver. Where such comparison is made between two frequency standards, the 100 kc output of one standard can be amplified to about 100 volts and applied to the calibrator crystal socket in the receiver to obtain strong 100 kc points throughout the range of the receiver. A transmitter whose frequency standard is to be trimmed is tuned to an integral 100 kc point; the receiver is tuned to that frequency, and the beat between the two observed on the receiver S meter. The level of input to the receiver from the transmitter is adjusted for maximum swing on the S meter and then the transmitter standard is adjusted for zero beat. When using frequency standards with stabilities of 1 part in 10^8 per day or better, beats having periods of approximately 20 seconds can be obtained when comparing signals at 30 mc.

Since the frequency counter will not give an accurate measurement of frequency in the presence of a complex wave, and since measurements of frequency by means of a receiver are confused in the presence of additional signals, it is necessary to disconnect modulation from the transmitter and make all frequency measurements on the reinserted carrier, or on a single tone of known frequency and stability equal to that of the equipment under test.

3. SIGNAL LEVEL

Most measurements of power output presume that the level of undesired power output is small compared to the accuracy required of the measurement of desired power output. Normally, the rms sum of undesired power output will be in the region of 35 to 40 db below desired power output, or about 1% of desired power. If a suitable resistive load is available to terminate the r-f output circuit, a vacuum-tube voltmeter such as the Hewlett-Packard 410B will give a reading of voltage across the load which is reasonably accurate for power computation. This meter is a negative peak reading meter calibrated in terms of rms. It has a very high impedance a-c probe which will not change the reactive characteristics of a 50 ohm line, provided that the meter probe is connected across the line with a minimum of additional shunt capacity or series inductance. When using this method for measuring power, the accuracy of voltage measurement is important since the effect of any error is squared when computing $P = \frac{E^2}{R}$. The most accurate measurement available with the Hewlett-Packard 410B is made with a single tone, although experience has shown that measurement of two equal tones on the r-f output line will give a voltage indication for computing peak envelope power to an accuracy varying between 5 and 10%, the higher accuracy being obtained on the higher ranges of the meter. Table 8-1 and figure 8-1 show comparative measurements made with a vu meter which reads slightly above the average value of the applied signal, a Ballantine 310A which reads the average value, a Ballantine 320 which reads true rms, a Hewlett-Packard 410B which reads negative peaks but is calibrated in rms, and an oscilloscope which was used to obtain the peak to peak voltage of the signal. Measurements were made on a signal containing from 1 to 16 equal audio tones. In table 8-1 the output level of the amplifier was reset for each reading so that the maximum reading on the vu meter was -3.0 vu. Since beats between tones began to effect the meter readings after more than 2 tones

TABLE 8-1. COMPARATIVE METER READINGS 1 TO 16 EQUAL TONES AMPLIFIER OUTPUT HELD CONSTANT ON VU METER

Meter	VU		310A		320	410B			Scope
Reads	Average +		Average		True RMS	Neg Peak			P-P
Calibrated	VU		dbv		dbv	Volts rms			Volts
No. of Tones	Min	Max	Min	Max		Min	Avg	Max	Max
1		-3.0		-1.5	-1.5			0.85	2.6
2		-3.0		-1.7	-0.7			0.23	3.8
3		-3.0		-1.7	-0.8	1.3	1.4	1.5	4.5
4	-4.0	-3.0	-3.0	-1.7	-1.3	1.0	1.4	1.6	5.0
5	-4.0	-3.0	-2.7	-1.8	-1.2	1.1	1.4	1.8	5.5
6	-4.0	-3.0	-3.1	-1.7	-1.0	1.0	1.4	1.9	5.3
7	-4.0	-3.0	-3.0	-1.6	-0.95	1.1	1.5	2.1	6.0
8	-4.0	-3.0	-2.7	-1.8	-1.0	1.1	1.5	2.2	6.2
9	-4.0	-3.0	-3.0	-1.7	-1.0	1.1	1.5	2.3	7.0
10	-3.7	-3.0	-3.0	-1.8	-1.0	1.1	1.6	2.3	7.0
11	-3.7	-3.0	-3.0	-1.5	-1.0	1.05	1.5	2.4	7.0
12	-4.0	-3.0	-2.7	-1.8	-1.0	1.1	1.6	2.3	7.0
13	-3.9	-3.0	-2.7	-1.7	-1.0	1.1	1.6	2.2	7.0
14	-4.0	-3.0	-2.8	-1.8	-1.0	1.1	1.6	2.5	7.2
15	-4.0	-3.0	-3.0	-1.7	-1.0	1.1	1.6	2.5	7.2
16	-4.0	-3.0	-2.8	-1.7	-1.0	1.1	1.6	2.4	7.2

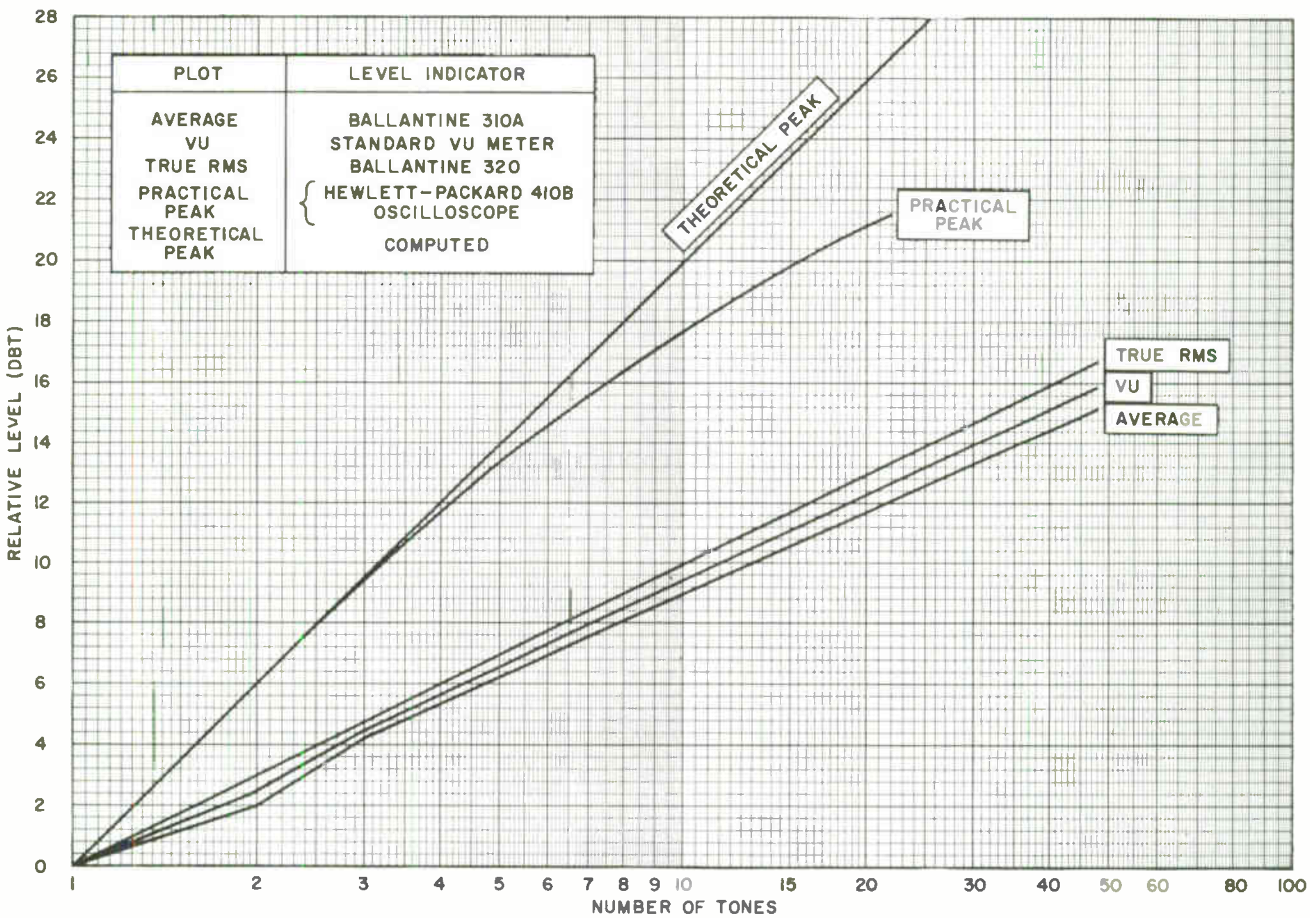


Figure 8-1. DB Level Versus Number of Equal Amplitude Tones for Various Types of Level Indicators

were combined, the minimum and maximum readings were taken each time. Where average readings are given, they are the value indicated by the meter for a major portion of the period during which each reading was taken. Similar readings were taken for figure 8-1. A signal composed of 16 tones of equal amplitude was adjusted to indicate full scale maximum on the vu meter. Then the tones were dropped one at a time and the maximum readings were taken without further level adjustment. All readings were converted to DBT (decibels with 0 db equal to the indication for a single tone on each type of indicator).

The theoretical peak was computed on the basis of a 6 db increase each time the number of tones is doubled and is indicative of theoretical peak envelope power. The true rms indication increases 3 db each time the number of tones is doubled and is indicative of true "heat" power. Statistically, the possibility of the practical peak indication approaching theoretical peak during a given time interval may be computed. The Hewlett-Packard 410B and the oscilloscope agree within a few tenths of a decibel when read for maximum peak indication over a one minute interval, yielding the practical peak curve of figure 8-1 which is indicative of practical peak envelope power. Note the double inflection in the vu and average curves at the two-tone and three-tone points, the divergence of the vu and average curves from true rms above three tones, and the fact that the vu indication remains almost exactly midway between average and rms throughout the graph.

The Ballantine 310A is an average reading vacuum-tube voltmeter calibrated in terms of rms but does not have as high an input impedance as the Hewlett-Packard 410B probe and has a cutoff frequency of about 2 mc. However, since its sensitivity extends to less than 100 microvolts, it is useful for measuring low signal levels at i-f frequencies. Dummy loads such as those manufactured by the Bird Electronic Corporation exhibit resistance tolerance of the order of ±0.5 ohm and present very little reactance to the signal source. Most of these loads are good for measurements into the hundreds of megacycles. If a resistive dummy load is not available, the measurement of voltage or current will not give an accurate indication of power. In these circumstances the calorimetric method of measuring power will give a more accurate indication. The flowmeter and thermometers may be calibrated separately to high accuracies and errors in reading of the flowmeter or thermometers do not appear in a squared term in the calculation of power as do the errors in voltage or current readings. The accuracy of the temperature readings, however, is affected by the temperature at which the calorimeter is operated with respect to the ambient temperature.

Another device for measuring power, which does not require a special load and which can be used while feeding an antenna, is a directional wattmeter. Figure 10-29 is a chart showing the relationship of the standing-wave ratio to forward and reflected power. The wattmeter will measure average power integrated as affected by the time constant of the metering network, and if both forward and reflected power are measured instantaneously, the chart will hold for multiple tones. An envelope detecting voltmeter will read 45% of peak voltage when a two-tone r-f signal is applied. This presumes that the voltmeter reads the average of the r-f and the rms of the envelope. Thus 0.707 of 0.637 is 0.45 or 45% of peak envelope voltage. A modification of directional wattmeters to approach a measurement of peak envelope power is useful for speech and other complex single-sideband signals, since these quantities are of greater importance in single-sideband equipment than average power. With extra capacity to lengthen the time constant of the voltmeter circuit, it is possible to make the meter read approximately 0.8 of the peak envelope power so that it will follow speech peaks more closely.

4. UNDESIRED POWER OUTPUT

Two types of undesired power may be present in the output of a transmitter. These are:

(1) Spurious responses outside the passband

(2) Intermodulation and incidental amplitude and angle modulation products in or near the passband

Spurious responses outside the passband consist mainly of harmonics of the desired output frequencies, products of frequency synthesis, and broadband noise from lower level stages amplified by the power amplifiers. The most direct method of measurement of this type of undesired response is the receiver/signal generator substitution method shown in the block diagram of figure 8-2. A portion of the transmitter output is sampled to provide approximately 1 or 2 volts r-f at desired signal frequencies through a 50 ohm attenuator to a 50 ohm load. When measurements are to be made, the transmitter is operated to provide carrier only or one sideband of modulation by a single tone of known frequency. The receiver is tuned to the transmitter output frequency and the 50 ohm attenuator is adjusted to obtain a convenient reference point on the receiver level indicator. The signal generator is then substituted for the transmitter and tuned for maximum indication on the receiver level indicator. Then the signal generator output, the 50 ohm attenuator, or both are adjusted to obtain the same indication on the receiver level indicator as was obtained with the transmitter. The reading on the signal generator level indicator, corrected to compensate for the attenuator setting, is the amplitude of the signal across the 50 ohm load which was equivalent to that obtained from the transmitter, and is the reference or zero decibel indication for measurements of spurious products. The transmitter is then reconnected to the attenuator and operated exactly as before, but the receiver is

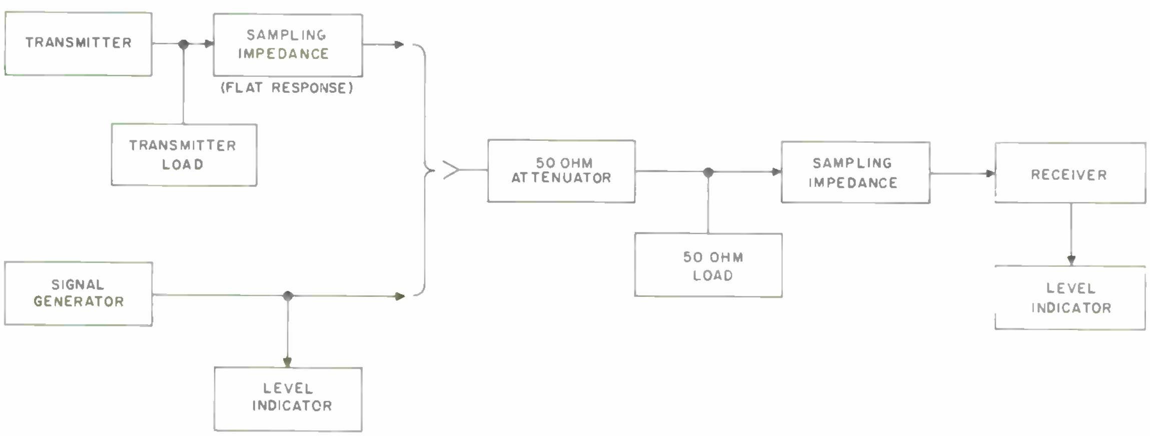


Figure 8-2. Receiver/Signal Generator Substitution, Block Diagram

tuned to a known point in its frequency range where a harmonic would appear, or a search for a spurious response is made. After the receiver is tuned for maximum level indication at a spurious response, the attenuator is adjusted to obtain a convenient reference point on the receiver level indicator. Once again the signal generator is substituted for the transmitter and tuned for maximum indication on the receiver level indicator at the frequency where the spurious response was found. The signal generator and attenuator are adjusted to obtain the same level indication at the receiver as produced by the spurious response and the voltage output of the signal generator, corrected to include the attenuator setting, provides a second reading of amplitude across the 50 ohm load. Comparison of the two readings gives the relative amplitude of the undesired response with respect to the desired signal. The accuracy of these measurements is determined by the accuracy with which the amplitude indication at the receiver level indicator is duplicated when the signal generator is substituted for the transmitter. If the receiver and signal generator are suitably isolated from any interconnecting paths such as the power line, are suitably shielded from each other, and the transmitter is shut down completely when measurements are made with the signal generator, it is possible to obtain reliable and repeatable measurements about 120 db below the desired output of the transmitter.

Since it is most practical to generate single sideband by initially generating the sideband frequencies at an intermediate frequency and subsequently heterodyning these signals to the desired r-f output frequency, products of output frequency synthesis may appear at the output of the transmitter. These products may be the actual heterodyning frequencies used to translate the intermediate frequency single-sideband signal to the r-f output frequency, or they may be mixer products of this process. Normal equipment specifications for these products are that they shall be from 70 to 80 db below the desired output of the transmitter. Low order harmonics of the output frequency are often specified 50 to 60 db below the desired output.

Since the Q of the output circuits of a reasonably efficient r-f power amplifier will be in the vicinity of 10 to 12, broadband noise generated in or amplified by the r-f power amplifier stages will not be affected appreciably by the selectivity of the output tank circuits. In a linear power amplifier with three or four stages, this noise will be due primarily to thermal and shot noise in the lower level stages. The effects of this noise may be observed on a spectrum analyzer or on a highly selective receiver. It is not normal for this noise to interfere with communication on the channel of the transmitter causing the noise, however, if the noise is particularly severe, its broadband nature may cause it to interfere with adjacent channels several channel widths removed.

The second type of undesired power output includes those spurious responses inside or very near the pass-bands of frequencies including the intelligence to be transmitted. These in-band spurious products are caused by:

(1) Intermodulation distortion resulting from operation of mixers or amplifiers beyond their capabilities.

(2) Amplitude or angle modulation resulting from imperfect stabilization of the oscillator or synthesizer from which translating frequencies are derived.

In addition, the following characteristics are important to a good single-sideband signal:

(1) Suppression of opposite sideband

(2) Suppression of carrier

(3) Minimum compression of desired signal due to power amplifier loading

Two equal amplitude audio tones have become a standard test signal for distortion measurements because:

(1) One signal is insufficient to produce intermodulation.

(2) More than two signals result in so many intermodulation products that analysis is impractical.

(3) Tones of equal amplitude place more demanding requirements on the system than it is likely to encounter in normal use.

Any two tones will serve for this test but with many frequency relationships, intermodulation products and harmonics tend to merge making identification of these products impossible. A 3 to 5 frequency ratio is the simplest that will alleviate this problem. Tones having a more complex ratio may produce products with frequency relationships more suitable to certain tests, but these products will be more difficult to identify than those of the simpler ratio. The following chart shows the relationships between products produced by distortion of the upper sideband of 300 kc modulated by 3 kc and 5 kc tones.

TABLE 8-2. SINGLE-SIDEBAND DISTORTION PRODUCTS

Frequency (kc)	Order of Products							
		7	6	5	4	3	2	1
295								OSB
296								
297		IM						OSB
298								
299				IM				
300								Car
301				AIM		IM		
302			AIM				AD	
303		AIM						3 kc DT
304					AIM			
305		AIM						5 kc DT
306			AIM				AH	
307				AIM		IM		
308							AS	
309				IM		AH		
310			AIM				AH	
311		IM						
312					AIM AH			
313		AIM						
314			AIM					
315				AH		AH		
316					AIM			

Legend:

A = Audio
AD = Audio Difference
AH = Audio Harmonic
AIM = Audio Intermodulation
AS = Audio Sum
Car = Carrier
DT = Desired Tone
IM = Intermodulation
OSB = Opposite Sideband

The idealized spectrum analyzer pattern for a two-tone single-sideband signal will consist of three discreet frequencies as illustrated in figure 8-3.

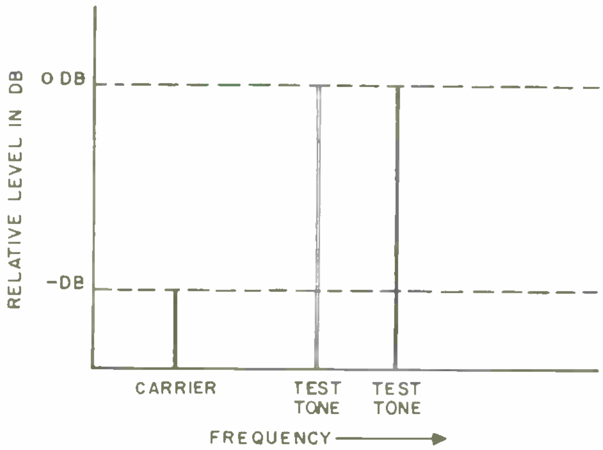


Figure 8-3. Idealized Spectrum Analyzer Pattern

These are the frequencies of each of the two audio test tones translated to the desired r-f output frequency and the carrier, which should be suppressed to the required level. The amplitudes of all undesired products and the carrier are measured in terms of decibels below either of the two equal amplitude test tones. Practical circuits always have some degree of intermodulation distortion which appears in the form of new discreet frequencies above and below the two test tones as illustrated in figure 8-4. The spacing

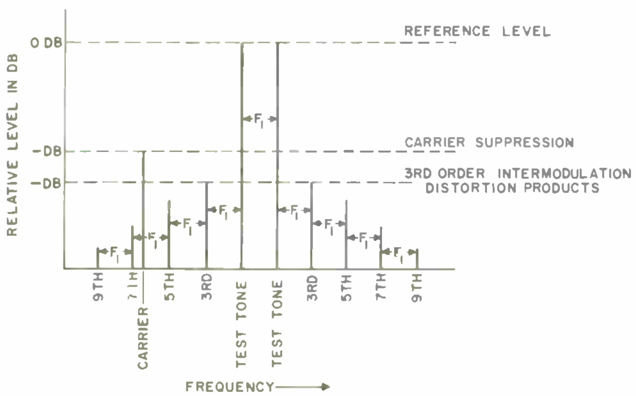


Figure 8-4. Practical Spectrum Analyzer Pattern

between each tone and the adjacent intermodulation products and the spacing between subsequent intermodulation products are equal to the spacing, F_1, between the two tones. The intermodulation products first adjacent to the desired tones are the third-order intermodulation distortion products; the next pair of products are the fifth-order intermodulation distortion products spaced equally outside the third-order intermodulation products; the next pair are the seventh, then the ninth and so on. The order of a distortion product is the sum of the coefficients in the frequency expression. For example, the third-order intermodulation products will be two times the frequency of one desired tone minus the frequency of the opposite tone. The fifth-order product is three times the frequency of one tone minus two times the frequency of the other. The odd-order products fall in or near the desired transmission band and are, therefore, the most objectionable because once generated they cannot be eliminated by either the transmitter or receiver. The signal to distortion ratio is the ratio of either of the two desired test tones to the largest undesired product expressed in decibels. A signal to distortion ratio of 40 db is usually acceptable for high-frequency communication systems when the equipment is tested on a two-tone basis. Unless unusual cancellation exists in the power amplifier, the third-order intermodulation products will be largest and the higher order products will be progressively smaller.

Over-all distortion resulting from several cascaded stages of amplifiers, modulators or mixers may be computed if the distortion of each stage is known. It is useful to note that each "stage" may be a "black box" actually composed of multiple stages, and one of the black boxes may be the distortion analysis equipment itself.

To obtain the over-all distortion in a system composed of several cascaded stages the following formula applies:

$$db_t = 10\left\{10 - \log\left[\log^{-1}\left(10 - \frac{db_1}{10}\right) + \log^{-1}\left(10 - \frac{db_2}{10}\right) + \ldots\ldots + \log^{-1}\left(10 - \frac{db_n}{10}\right)\right]\right\}$$

where db_t = Total or over-all signal to distortion ratio in db

db_1 = Signal to distortion ratio of 1st stage in db

db_2 = Signal to distortion ratio of 2nd stage in db

db_n = Signal to distortion ratio of nth stage in db

log = logarithm with a base of 10

$\log^{-1}$ = antilogarithm with a base of 10

This equation yields the following typical results from two stages:

Difference between signal to distortion ratio between the two stages in db	Amount by which the overall signal to distortion ratio is degraded beyond that of the poorer stage in db
0	3.0
3	1.8
6	1.0
10	0.4

Any of the oscillators in a transmitter, receiver, or in an analyzer may have amplitude or angle modulation caused by such deficiencies as power supply ripple, alternating currents in tube heaters, mechanical vibration, or strong electric or magnetic fields in the vicinity of the oscillators or their control devices. This incidental modulation causes new sidebands to be produced by the transmitter. These may be observed on the spectrum analyzer display as responses symmetrically located on either side of all desired tones. Each distortion product will also exhibit these sidebands. When the oscillator that is used in the frequency scheme of a transmitter is modulated, the sidebands produced thereby are often unequal in amplitude because of simultaneous angle and amplitude modulation. An analysis of this phenomenon is summarized in the article entitled "Linearity Testing Techniques for Sideband Equipment" by Icenbice and Fellhauer in The Proceedings of the IRE, December, 1956. Phase modulation distorts the amplitude symmetry of the two sidebands produced by a single sine wave simultaneously angle and amplitude modulating a carrier or other desired output signal by subtracting a component from one sideband and adding it to the other.

5. SPECTRUM ANALYZER 478R-1

The most informative and universal method of measuring in-band spurious is by means of a spectrum analyzer such as the Collins 478R-1 Spectrum Analyzer. In this analyzer, a complete picture of the spectrum in the vicinity of the intelligence passband is plotted directly on an oscilloscope screen or may be recorded by means of a two-axis recorder. The problem of construction of this equipment was primarily to reduce intermodulation within the analyzer to a level appreciably below that to be measured. Although the analyzer is large and complex, the signal under test passes through only two tubes before detection by the narrow-band selective amplifier. These two tubes are both mixers, and much of the remainder of the equipment is devoted to insuring that the injection or translating frequencies applied to these mixers are sufficiently free from noise, distortion, and incidental angle modulation.

The basic circuit of the Spectrum Analyzer is that of a wave analyzer for measuring frequencies generated by signals passing through an amplifier or mixer or other system with an unknown amplitude transfer characteristic. Figure 8-5 is a block diagram of the 478R-1. Additional selectivity, variable sweep width, and other features permit accurate and simultaneous measurements of level in decibels versus frequency of distortion, hum, noise, and other spurious products in a direct plot on the analyzer screen or on a two-axis recorder. Included in the analyzer is a two-tone audio generator consisting of two audio oscillators and a filter-mixer panel. This portion of the analyzer generates a two-tone audio test signal for use as an audio input for intermodulation distortion measurements of the equipment under test.

The two-tone mixer panel satisfies several requirements for minimizing harmonic and intermodulation products in the two-tone audio test signal. One such requirement is sufficient isolation between the two audio signal generators to reduce intermodulation distortion in the output tubes of one generator because of coupling to the output of the other. This isolation is provided by pads in the output of each generator ahead of the mixing circuit. Second and higher order harmonics in the output of the generators are attenuated by plug-in low-pass filters selected to have cutoff frequencies between the fundamental frequencies and the second harmonic frequencies of their respective generators. Second-order intermodulation distortion which appears to be third-order intermodulation results from direct mixing of the second harmonic output of one audio generator with the fundamental of the other. This effect also is minimized by the low-pass filters. The difficulty of mixing the output of two signal generators so that they do not modulate each other is illustrated in figure 8-6. Both circuits A and B have the same output level but circuit A has approximately 50 db more isolation between the oscillators. The level of third-order intermodulation products is approximately 20 db higher in circuit B than in circuit A. While the attenuation of the audio output signal from oscillator no. 1 is affected primarily by the output series resistor and the 560 ohm load, the attenuation between audio oscillator no. 1 and no. 2 is affected by the output series resistor of oscillator no. 2 and the internal generator impedance of oscillator no. 2 as well as the attenuation between oscillator no. 1 and the output.

During spectrum analysis of a transmitter, products observed by the analyzer can be more readily identified by removing one or the other of the audio tones and observing the effect on the intermodulation products of interest. The ON-OFF switches in the two-tone audio generator are provided with dummy loads in the OFF position to preserve the impedance termination on the audio mixing circuit and thereby prevent a change in the amplitude of the remaining tone when

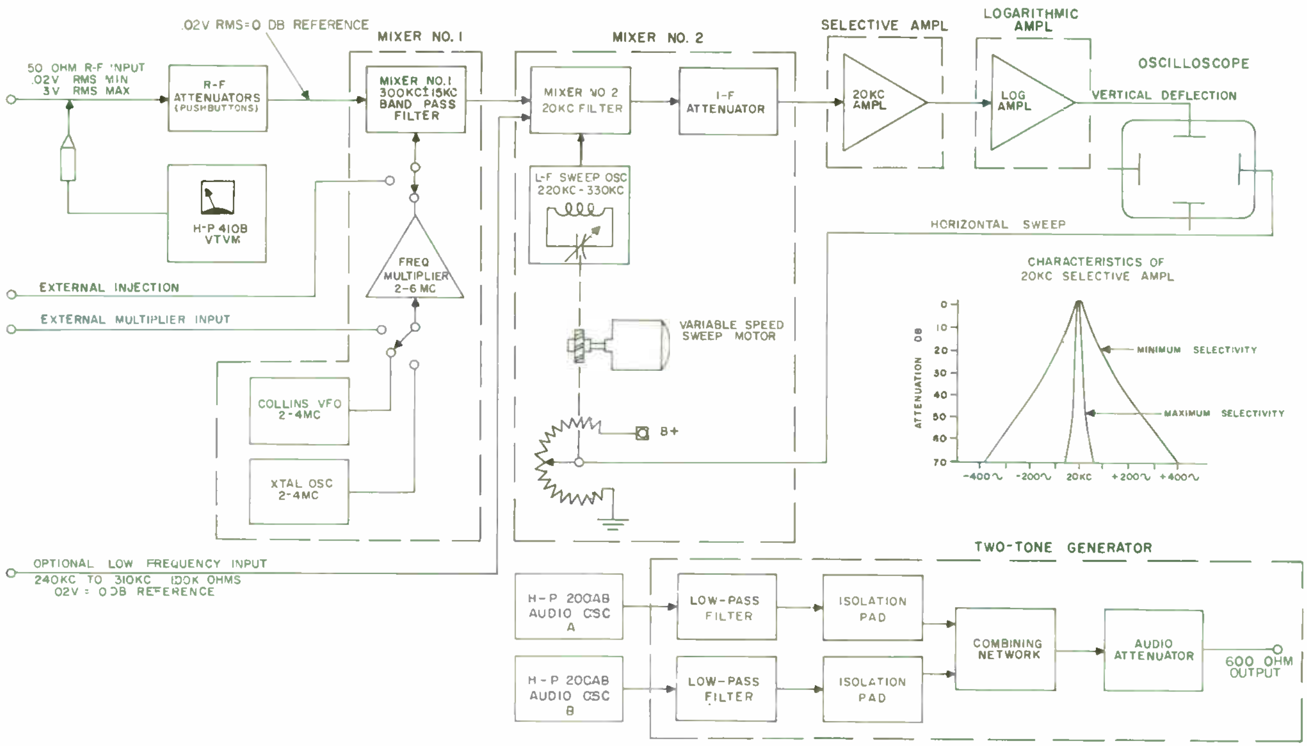


Figure 8-5. Spectrum Analyzer 478R-1, Block Diagram

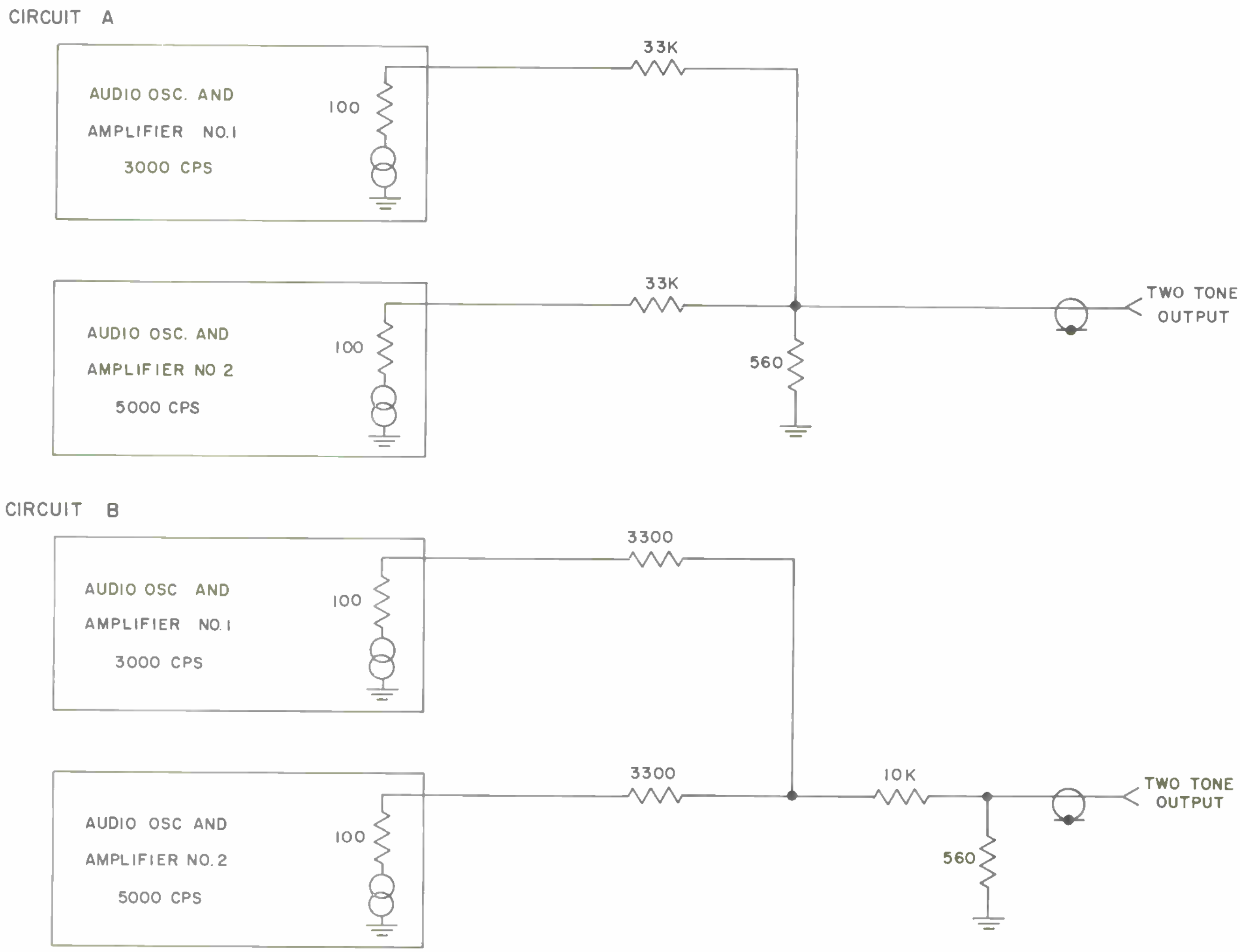
one tone is switched off. The audio filters may be switched out of the circuit to allow the use of audio frequencies beyond the range of the filters, and a set of decade attenuators is provided to enable rapid and accurate testing of equipment with different amplitudes of audio input.

The dynamic range of the analyzer is 70 to 80 db, displayed on one scale to an accuracy of ±1 db. Continuous metering circuits are provided on the front panel to insure correct mixer injection level. The analyzer will accept a frequency spectrum with a center frequency from 1.7 mc to 64.3 mc and from 240 kc to 310 kc without additional coils or test equipment. The spectrum display is on a 17 in. cathode ray tube. Signal to be analyzed is fed into the precision attenuator panel where it may be monitored by the internal vacuum-tube voltmeter, Hewlett-Packard 410B. The attenuator is adjusted to the proper level and the attenuated signal is applied to mixer no. 1 which converts the signal to the 300 kc i-f. The output of mixer no. 1 passes through a 300 ±15 kc band-pass filter to mixer no. 2 where it is converted to 20 kc. Mixer no. 2 can accept directly any frequency from equipment under test between 240 kc and 310 kc. The tuning capacitor of the injection oscillator for mixer no. 2 is rotated by a variable speed motor to sweep the frequency of this oscillator through the required range.

Sweep widths of 4 kc, 8 kc, and 16 kc are available. The output of mixer no. 2 is fed through a precision attenuator with 0.1 db steps to a narrow-band 20 kc selective amplifier. The half bandwidth of the selective amplifier at 40 db below maximum response is variable from 30 to 145 cps by a control on the front panel. The output of the 20 kc selective amplifier is fed to a logarithmic amplifier. The output of this amplifier is a d-c voltage that is a logarithmic function of the input over a 70 db dynamic range. The calibrated attenuator between mixer no. 2 and the 20 kc selective amplifier provides for checking the linearity of the logarithmic amplifier and the oscilloscope to insure an accuracy of ±1 db throughout the 70 db dynamic range of the analyzer. The varying d-c output from the logarithmic amplifier is applied directly to the vertical deflection amplifier of the oscilloscope or to an external recorder. Synchronized horizontal sweep voltage is provided by a potentiometer ganged to the oscillator sweep tuning capacitor.

The signal path in the analyzer includes only three nonlinear devices ahead of the 20 kc selective amplifier after which nonlinearity causes no further intermodulation. The first nonlinear device in the signal path is the Hewlett-Packard 410B vtvm probe, but the loading of this high-impedance probe on the 50 ohm circuit is so slight that negligible intermodulation distortion results. Mixer no. 1 and mixer no. 2 consist of only one tube each and their operating characteristics have been very carefully selected to minimize intermodulation distortion. Microammeters on the mixer front panels provide continuous monitoring of injection grid current to insure that the mixers are always operated under optimum injection level conditions.

An ideal panoramic display of a constant carrier with no modulation would appear as a single line at right angles to the frequency axis. However, in practical equipment this display is a single plot of the selectivity of the analyzer. If the selectivity of the analyzer is changed, the displayed shape of the same carrier under test will change to the new shape of the selectivity curve of the test equipment. Signals under test which have sidebands or intermodulation products to be observed by the analyzer will produce individual responses corresponding to each of these sidebands, or products, together with the desired responses themselves and each response will be



CIRCUIT	FREQ (CPS)	3RD ORDER LEVEL (DB)
A	1000	> 75
	7000	75
B	1000	52
	7000	58

Figure 8-6. Two-Tone Generator, Source Isolation

basically the shape of the analyzer selectivity curve, each with it own maximum amplitude. The maximum response from each discreet frequency is the required measurement. When these discreet frequencies or sidebands are spaced only a few cps apart, such as may be encountered with hum modulation, their corresponding responses on the analyzer screen tend to merge into each other. The responses, for example, of hum modulation will appear on the skirt of the response to the carrier for that modulation. The ability to separate such discreet frequencies is known as the resolving power of the analyzer. Maximum resolving power is attained when the equipment is operated with minimum sweep width, minimum sweep speed, and maximum selectivity. Since this mode of operation reduces the speed with which data may be obtained, provision is made for varying all three parameters so that data requiring less resolving power may be obtained more rapidly. With maximum selectivity, the speed of the sweep and the sweep width may be adjusted so that the frequency is swept through the response frequency of the analyzer so rapidly that the effective Q or selectivity of the analyzer will not allow the signal to build up to its peak amplitude before the sweep has passed this frequency. This error is always present with any reasonable amount of selectivity, but the effect will be negligible if the sweep width and sweep speed used are commensurate with the selectivity to which the analyzer is adjusted. The easiest check to insure a safe sweep speed is merely to reduce speed about one half and note whether the amplitude increases. If the peak amplitude increased more than 1 db, the original speed was too fast.

When single-sideband suppressed carrier equipment is under test, it is helpful to take the intermodulation distortion test data in steps of about 3 db. This is accomplished by inserting 3 db of attenuation in the output of the two-tone mixer by means of the audio attenuators and removing 3 db of attenuation from the r-f input to the analyzer by means of the r-f attenuators. This preserves the same amplitude of desired tones on the analyzer thus allowing observation to be concentrated on the changes in the intermodulation products. In this way, the point on the distortion versus signal curve, at which the equipment under test should be operated, can be established rapidly. The intermodulation products will increase relative to the desired output signals at an ever increasing rate, as the rated signal level of balanced modulators, mixers, or amplifier stages in the r-f equipment is approached. Near the overload point, intermodulation products commonly increase at a rate 2 or 3 db faster than the desired output signals. Beyond this point the rate of increase of the ratio of intermodulation distortion products to desired output signals becomes much more rapid. If the audio input signal level is reduced well below the normal level, the rate of increase of intermodulation distortion products' amplitude will be reduced until at very low levels the intermodulation distortion products will change imperceptibly with respect to the desired output signals.

6. ANGLE MODULATION MEASUREMENTS

In order to make use of the resolution available from the selectivity of the 478R-1 Spectrum Analyzer in measurements of hum phase modulation sidebands, special precautions were taken to minimize hum phase modulation on the injection oscillators and in the signal path through the analyzer. Electronically regulated power supplies were used to reduce hum ripple on the plate voltages to less than 1 millivolt, wherever possible cathodes were operated at ground potential, and plate circuits were either shunt fed or a pair of tuned circuits were coupled to insure that the grids of the following stages were grounded with respect to hum frequencies. In critical circuits, such as in the variable frequency oscillator, where the preceding methods were not applicable, filaments were supplied with direct current, circuits and components were selected to minimize microphonic pickup, and the Sola regulating transformer was housed in a special case to reduce magnetic fields at harmonics of the 60 cps line frequency.

When angle modulation is analyzed on a spectrum basis, only a simple amplitude detecting rectifier circuit is required in addition to the analyzer selectivity. Since the selectivity of the analyzer can slope detect angle modulation, it is necessary to integrate the output of the detector rectifier because the slope detection is a function of the analyzer selectivity and will not truly represent the spectrum of the equipment under test. On the 478R-1 analyzer, an external plug-in capacitor of several microfarads may be placed across the oscilloscope deflection input terminals to perform this integration and eliminate the slope detection.

A sinusoidally modulated FM wave has a spectrum which contains not just two side frequencies as in AM, but an infinite number of side frequencies spaced equally from the carrier by intervals equal to the modulating frequency. When the angle modulation level is very low, the amplitudes of the higher order sidebands drop very rapidly. When the modulation level is high the amplitude of the carrier may be lower than some of the sidebands, and the sidebands will extend over a much larger band of frequencies. A qualitative check of the effect of low-level incidental angle modulation on a carrier may be obtained by slope detection in a receiver which has a relatively high degree of selectivity, such as the 51J or R390. The receiver is tuned to one side of the carrier signal so that the S meter indicates 1/2 or less of the maximum deflection obtained when peaked exactly on the signal. If no hum or other tone is heard in the receiver as the avc allows the sensitivity of the receiver to increase and the noise to rise, it may be assumed that closely spaced discreet angle modulation spectra are below the noise level.

7. COMPRESSION MEASUREMENT

An additional characteristic of power amplifiers known as compression is often measured as an indication of capability of the power amplifier and its power supply. Compression of the output signal may result from less than optimum d-c regulation of the power supply for the power amplifier plate, screen and bias voltages and may serve as an indication of intermodulation in the power amplifier when subjected to close spaced tones. Since the power amplifiers are normally operated in class AB_1, or some other mode of class B operation with respect to plate current, the load on the power supply varies with the instantaneous amplitude of the signal envelope. Compression results when, in the presence of one signal which does not utilize full peak envelope power capability, a second signal is applied which approaches full peak envelope power. The amplitude of the first signal is compressed by an amount which is a function of the variation in the power supply output voltage as a result of the additional loading demanded by the second signal. Measurements of compression must be conducted with selective equipment which is capable of observing the amplitude of one continuous signal as a second signal is varied in amplitude. Such measurements are usually obtained with spectrum analysis equipment by observation of a continuous desired signal 10 to 20 db below peak envelope power while a second signal which demands approximate peak envelope power is switched on and off. The effect on the fixed amplitude tone is plotted in terms of decibels versus the number of decibels by which peak envelope power is approached or exceeded. The 478R-1 analyzer is particularly adaptable to this type of measurement in that its decibel scale may be expanded 10 to 1 so that each inch of oscilloscope scale equals 1 db, and the accuracy of the measurement is then ±.1 db. Such measurements may be conducted using the same tones as are used with the standard two-tone test signal and stopping the sweep motor so that the amplitude of one tone is continuously monitored while the other tone is switched on and off. Intermodulation distortion may be produced by the r-f power amplifier when operating with close spaced tones if the low-frequency a-c impedance of the power supply is too high. The screen voltage supplies for pentode power amplifiers often present stringent requirements because such amplifiers are sensitive to screen voltage changes.

8. INTERMODULATION MEASUREMENTS WITH BUILT-IN MONITOR

Intermodulation distortion in a transmitter may be tested by means of a monitor which uses the same frequency scheme as the transmitter, but operating in reverse to translate the r-f output signal back to audio or some convenient fixed intermediate frequency. In this manner, spectrum analysis can be made and compared with analysis of the original audio signals applied to the transmitter. The monitor itself must be carefully designed so that its intermodulation is lower than that of the transmitter to be tested. For example, if the level of intermodulation distortion products are 40 db below desired output signals, and the intermodulation within the test equipment is 46 db below desired output, then the resulting intermodulation distortion measurement will be in error by approximately 1 db. Therefore, the result of the measurement will indicate intermodulation products 39 db below desired output rather than the actual 40 db figure. These relationships represent normal conditions but do not guarantee this result for every situation.

Measurements made by means of such a monitor, however, will not show incidental angle modulation since use of the same translating frequency sources for both the transmitter under test and the monitor will tend to cancel the effect of such incidental angle modulation. Separate translating frequencies for the test equipment must be used to measure spurious sidebands caused by incidental angle modulation.

9. LINEARITY MEASUREMENT WITH NOISE LOADING

Intermodulation distortion measuring equipment using two tones is very versatile for identification of linearity characteristics. However, measuring equipment using noise as the input signal has the advantage that the test signal more nearly simulates the complex signal typical of voice or multiple tone modulation.

If band limited noise is introduced into a system under test, linearity of the system may be partially described in terms of the noise outside the original band limits. If the output of a random noise generator is fed into a band-pass filter which equally passes all noise frequencies in the passband to be tested except for a small portion at the upper and lower extremes, the noise loads all but a few cycles of the transmission band to any degree of modulation desired. At the receiving end of the system, three band-pass filters with equal bandwidth and insertion loss are used for measurement purposes. One such filter is selected near the center of the transmitted noise passband, and a true rms noise voltage from this filter is used as a reference signal level. The other two filters pass the distortion sidebands just outside the intended noise passband. The output of these two filters are measured separately with the true rms voltmeter, and these levels in decibels below the reference voltage represent the intermodulation distortion generated at the low- and high-frequency ends of the loaded passband. Noise generators for this purpose are commercially available and the band-pass filters may be selected to satisfy the requirements of the equipment under test.

As indicated in figure 8-7, a transmitter loaded with noise signals over a discreet bandwidth, B, will have all third-order intermodulation products appearing in a band equal to three times the desired bandwidth and with the same center frequency. All fifth-order

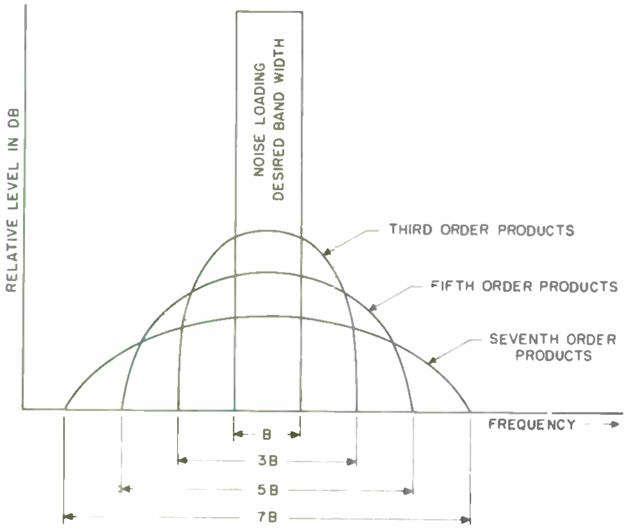


Figure 8-7. Intermodulation Products in a Noise Loaded System

intermodulation products will fall inside a bandwidth five times the width of the desired band and having the same center frequency. All seventh-order products will fall inside a band having seven times the width of the desired band and having the same center frequency and so on. Since the amplitude of the intermodulation distortion products are usually in approximate inverse proportion to the order of the product, the shape of the curves describing the amplitudes of the products may be predicted. If a two-tone test signal is employed, the discreet frequency relationships between the desired signals and the intermodulation products will be known or may be computed. When these are superimposed upon the curves of the intermodulation products for a noise loaded system (figure 8-6), an approximate plot of the entire spectrum resulting from the two-tone test signal may be predicted.

The Collins 478R-2 Baseband Spectrum Analyzer may be used in the same manner to test intermodulation distortion in a noise loaded system or in one employing multiple discreet frequencies. Only one filter is necessary at the receiving end of the system and a sweeping frequency scheme is employed to allow a panoramic plot on an oscilloscope or on a recorder of the responses throughout and beyond the system bandwidth. This equipment permits simultaneous observation of a range of audio and video signals from 3 kc to 2 megacycles. The plot on the oscilloscope or recorder is in terms of decibels plotted on the Y axis against frequency in kilocycles on the X axis, as in the 478R-1 Spectrum Analyzer. The sweep width is variable from 0 to 70 kc, and the main tuning dial is detented to position the center frequency at 50 kc intervals over the 3 kc to 2 mc range so that the complete frequency range may be examined accurately and rapidly in 50 kc increments. This equipment is particularly suited for portable and field use when used with a portable two-axis recorder.

10. DELAY DISTORTION

A transmitter which has delay distortion but negligible nonlinear distortion will not cause the production of new output frequencies. The amplitude of the output components are not affected by delay distortion; therefore, the existence of delay distortion within the transmitter will not influence the results of measuring nonlinear distortion by either multisignal loading or noise loading methods if the delay distortion does not vary with time. Phase always varies with frequency in a reactive network, but phase distortion is not necessarily produced. Instruments for the direct measurement of phase of audio frequencies are commercially available. When a plot of the phase measurements against frequency is differentiated with respect to frequency, the derivative is the envelope delay. Phase delay is defined as the ratio of the phase with respect to frequency and approaches the value of the envelope delay expression, namely, the derivative of phase with respect to frequency, when the phase-frequency plot approaches a straight line. The perfect system is never available in practice, so the phase delay is never exactly equal to the envelope delay. Delay distortion measurements may be obtained by passing a modulated signal through a network and measuring the resultant modulation envelope phase shift caused by the network under test. Time delay may then be computed by using $T_d = \Theta/360f_m$, where T_d is the delay in seconds; Θ is the phase shift of the modulation envelope in degrees, and f_m the frequency of modulation in cps. Systems in which delay distortion can seriously affect or completely destroy the useful characteristics of the desired signal must be tested using equipment designed for this particular purpose, the common method being the measurement of envelope phase shift.

11. FIELD TEST SET FOR INTERMODULATION DISTORTION MEASUREMENTS

A smaller portable spectrum analyzer would be useful for field test purposes. Such a unit could serve as a transmitter monitor as well as an intermodulation distortion analyzer. The following measurements could be provided by this equipment:

(1) Linearity or intermodulation distortion in a transmitter, receiver, or audio amplifier.

(2) Carrier leak or suppression

(3) Alignment checks

(4) Low-level r-f voltage measurements

(5) Transmitter monitoring

The basic technique of such a distortion analysis field test set would be to translate the r-f signal through low-distortion mixers to audio and separate the signals and distortion products with audio filters. The relative amplitudes of the intermodulation products with respect to desired signals would be obtained from attenuator readings and would be limited to about a 50 db range. When measuring intermodulation distortion in a transmitter or in a receiver, a study of relative magnitudes of the intermodulation distortion products, combined with familiarity with the frequency scheme, will usually isolate a malfunction with respect to the r-f, i-f, or audio section of the equipment under test. Such an analyzer would not indicate r-f signal levels directly in volts; however, it would indicate whether the same readings prevail that existed during a previous test. This function would be especially useful for trimming tuned circuits that operate at levels below normal vtvm sensitivities. The monitor portion of the test unit would allow aural checks with the equipment in operation to determine if speech or multiple-tone data circuits sound normal. When the equipment is used for monitoring purposes either or both sidebands, as transmitted, would appear in the audio output without separation.

The field analyzer would consist of three basic units:

(1) An audio two-tone test signal generator

(2) An audio distortion analysis filter

(3) An r-f to audio converter

For monitoring purposes, a fourth unit consisting of an audio amplifier with provisions for headphones or loudspeaker could be used.

a. AUDIO TWO-TONE SIGNAL GENERATOR

The audio two-tone signal generator (figure 8-8) would consist of two oscillators whose frequencies would be controlled by audio resonators such as are provided in Kineplex* equipment. Each oscillator

* Registered in U.S. Patent Office

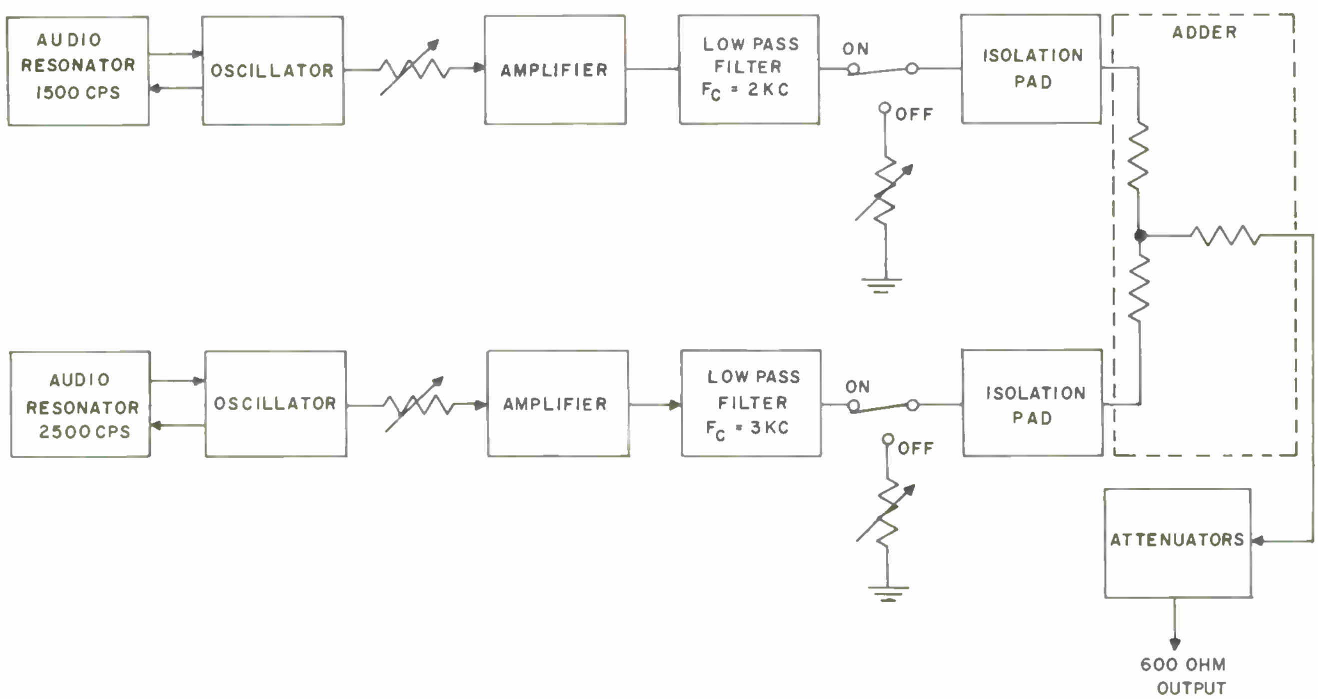


Figure 8-8. Audio Two-Tone Test Generator, Block Diagram

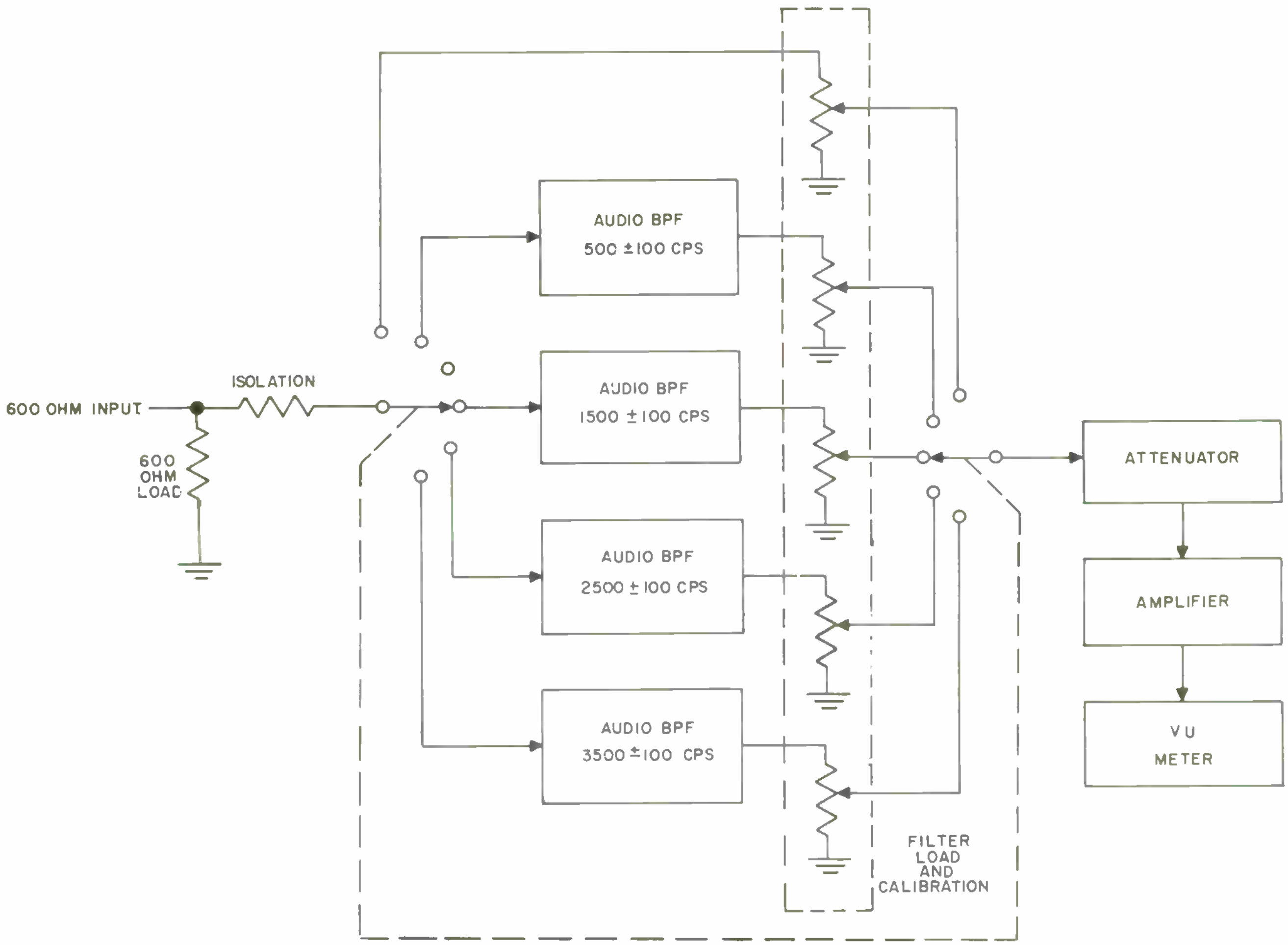


Figure 8-9. Audio Distortion Analysis Filter, Block Diagram

would be provided with a level control, its own amplifier, and a low-pass filter with the cutoff between the fundamental and second harmonic of its respective oscillator. On - off switches would be provided for each tone to enable identification of intermodulation distortion products. Isolation pads in each signal path plus the additional isolation afforded by the resistive adding network and isolation due to the presence of the low-pass filters would minimize intermodulation between the two audio amplifiers to a practical level. One and ten decibels per step attenuators in the combined output signal path would provide for intermodulation distortion measurements yielding a curve of intermodulation distortion versus audio input amplitude. The output impedance of the test generator would be 600 ohms, with an output signal level variable by means of the attenuators from approximately 3 volts per tone to 100 db below this level, approximately 30 microvolts.

b. DISTORTION ANALYSIS FILTER

Although audio analysis may be performed with commercial wave analyzers, the process is slow and commercial units for this purpose are expensive and impractical for field use. The audio distortion analysis filter (figure 8-9) for this field test set would have a 600 ohm input termination, suitable isolation of the internal audio band-pass filters from the 600 ohm input, and a selector switch for inserting any one of four audio band-pass filters or an unfiltered circuit in the path to the indicating meter. Each band-pass filter would be wide enough to pass only one of the discreet audio frequencies of interest, allowing for approximately 50 cps error in tuning of the vfo in the r-f to audio converter. Two of the filters normally would pass the desired two-tone test signals, and the other two filters would pass the third-order intermodulation distortion products resulting from the two desired test tones. The unfiltered circuit would provide measurement of signal levels under normal operating conditions of the equipment monitored. Fifth-order products in a transmitter could be measured by tuning the r-f to audio converter to one side so that the desired two tones fall on the frequencies of the first and second filters. Then the third-order intermodulation distortion product would appear in the third filter and the fifth-order intermodulation product in the fourth filter. Seventh-order products could be measured by side stepping the tuning of the r-f to audio converter one additional slot so that the higher of the two desired test tones falls on the frequency of the first filter. Then the third, fifth, and seventh-order products would fall on the second, third, and fourth filters, respectively. Provision would be made for compensating for the insertion loss of the various filters individually, and a second gang on the selector switch would connect the attenuator, amplifier, and level indicating meter to the filter circuit to which the input was switched. With a 10 db per step attenuator and a level indicating meter calibrated in decibels over a 1 to 10 db range, signals could be measured with an accuracy of ±1 db, while the dynamic range requirements on the meter amplifier would be only 10 db and, since it would pass only one frequency at a time, its distortion requirements would be negligible.

c. R-F TO AUDIO CONVERTER

The r-f to audio converter (figure 8-10) would consist of three mixers. Only fixed tuned low-pass and band-pass filters would be used in the signal path of the converter to eliminate the necessity for tracking tuned circuits. The first mixer input might be switched to either a high-impedance input through a potentiometer or to a 50 ohm calibrated r-f attenuator. A limited range of input level control would be obtained thereby to allow setting input signals at optimum mixer operating levels. There would be no internal gain controls. If required, additional attenuation could be obtained before the signal is applied to the converter by means of sampling impedances in the external isolating circuits. Transmitters and other equipments under test should have suitable test points provided for monitor pickup.

The first mixer of the converter would obtain translating signals from a multiple crystal oscillator and multiplier to heterodyne the r-f input to a range of 1.7 to 4.3 mc. Since the output of the vfo need not be

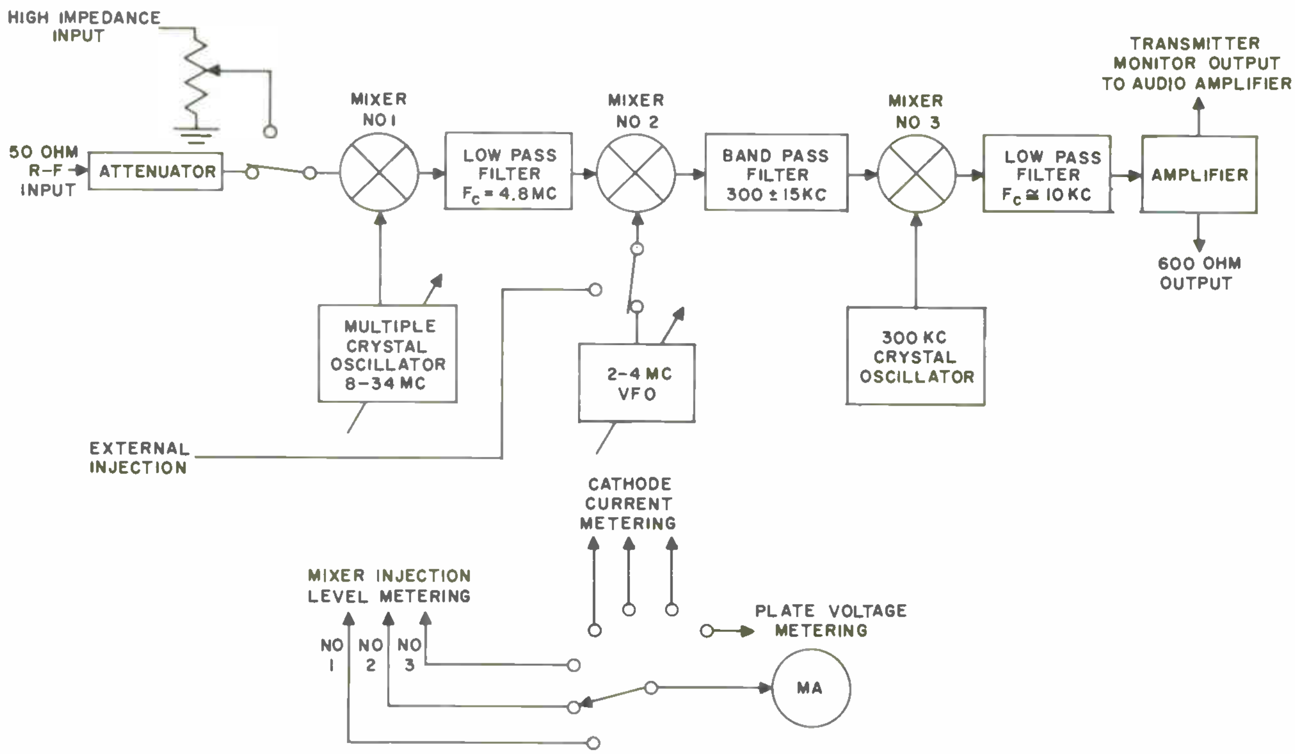


Figure 8-10. R-F to Audio Converter, Block Diagram

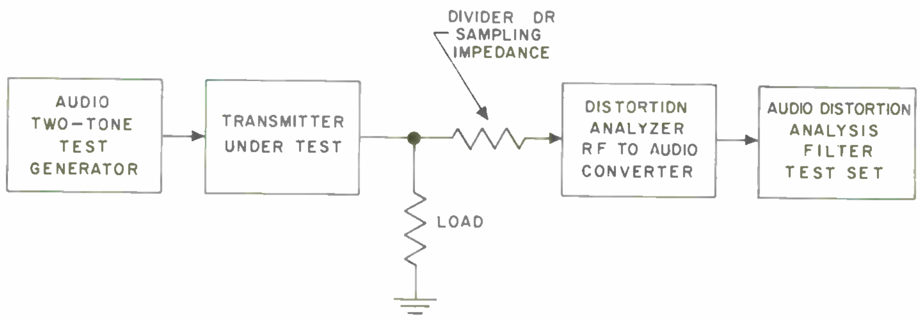


Figure 8-11. Test of Intermodulation in Transmitter

multiplied for this purpose, the stability and ease of tuning in the converter would be improved. The multiple crystal oscillator and multipliers would be provided with suitable tuned circuits for rejecting the undesired multiples where necessary. A level adjustment would be required for each range. All crystals, coils, and level controls would be selected by a range position switch. Since the output of mixer no. 1 would be filtered by a low-pass filter, signals in the range 1.7 to 4.3 mc could be applied through mixer no. 1 as an amplifier directly to mixer no. 2 without heterodyning. The second mixer would obtain a heterodyning signal from a 2 to 4 mc ten-turn variable frequency oscillator which could have a vernier control to facilitate fine tuning. If desired, the injection for mixer no. 2 could be applied externally to allow detection of signals at frequencies below 1.7 mc. The output of mixer no. 2 would be filtered by a band-pass filter approximately 30 kc wide and centered at 300 kc. Mixer no. 3 would heterodyne the signal to audio with injection obtained from a 300 kc crystal oscillator which could be provided with a trimmer for very fine tuning. The output of mixer no. 3 would be filtered to pass only the audio range after which the signal would be amplified by a low-distortion audio amplifier and made available in the form of a 600 ohm output capable of driving the audio distortion analysis filter. This 600 ohm output could be bridged with a high-impedance input audio amplifier driving headphones or a loudspeaker. If a monitor audio amplifier and speaker is used, and if the low-frequency response is adequate, the monitor would indicate the presence of hum in the transmitter. The unfiltered circuit in the distortion analysis filter could be used as an aid in rough tuning the r-f to audio converter in the absence of an audio monitor, or for measurements of the amplitudes of single signals which do not fall in the range of the filters but can be identified by audio monitoring. The latter application could implement frequency response measurements. Metering of the mixer injection levels would be provided to insure optimum mixer conversion efficiency with minimum intermodulation distortion. In addition, metering of mixer and audio amplifier cathode current and plate voltage would be provided. As in the 478R-1 Spectrum Analyzer, no preselectivity would be required and tuning of only the 2 to 4 mc vfo and selection of the proper crystal oscillator circuit for mixer no. 1 would suffice to translate the r-f signal to audio for these measurements.

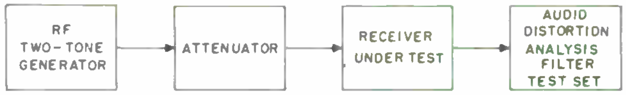


Figure 8-12. Test of Intermodulation in Receiver

The block diagrams of figures 8-11 through 8-13 demonstrate the use of the several sections of an analyzer of this type for tests of intermodulation distortion in a transmitter, a receiver or in an audio amplifier. An entire communications system could

Figure 8-13. Test of Intermodulation in Audio Amplifier

be tested or the equipment could be used as a transmitter monitor as indicated in figures 8-14 and 8-15. The analyzer would be designed so that it could be used to check itself as shown in figures 8-16 and 8-17. This system would be capable of measuring intermodulation distortion products 6 db or more higher in amplitude than the indicated measurements of these products when the analyzer is used to test itself. This condition would provide an accuracy of approximately 1 db in tests of other equipment.

12. SUMMARY

The demand for more communication channels in the high-frequency band and the large volume of information that must be carried on each channel has led to a search for a more effective and efficient method of communication. Technological advances in frequency control have made possible the use of single-sideband techniques which eliminate dependence upon a carrier for automatic frequency control at the receiver. However, to assure full utilization of the advantages of these techniques, test equipment must be provided which is capable of making measurements with a much higher degree of resolution than has been customary.

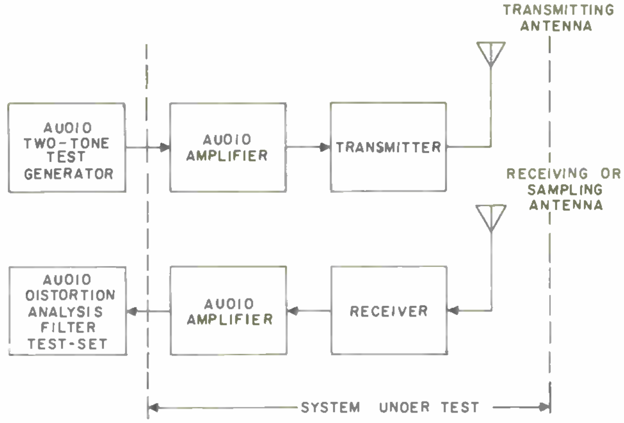


Figure 8-14. System Test for Intermodulation Distortion

Precision frequency counters provide a convenient method of measuring frequency accuracy while spectrum analysis measures the degree of linearity and frequency stability directly in terms of bandwidth requirements. Laboratory test equipment capable of the required resolution has been built and used in the development of SSB equipment. The design and construction of precision test equipment for field use is receiving added attention as the use of SSB equipment becomes more widespread.

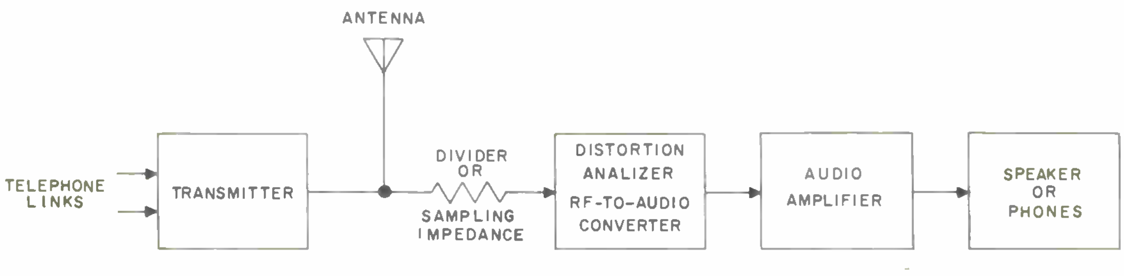


Figure 8-15. Transmitter Monitor, Block Diagram

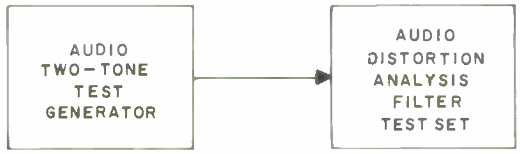


Figure 8-16. Test of Two-Tone Audio Generator

Figure 8-17. Test of Intermodulation in R-F to Audio Converter

COLLINS