BASIC ELECTRONICS

BUREAU OF NAVAL PERSONNEL

NAVY TRAINING COURSE NAVPERS 10087-A

PREFACE

Basic Electronics, is written for men of the U. S. Navy and Naval Reserve whose duties require them to have a knowledge of the fundamentals of electronics. Electronics is defined as the science and technology that is concerned with devices involving the emission, behavior, and effect of electrons in vacuums, gases, and semiconductors. Technically speaking, electronics is a broad term extending into many fields of endeavor. Today, electronics projects itself into Navy life at every turn. It aims guns, drops bombs, navigates ships, and helps control engineering plants. It is therefore important to become well informed in basic electronics in order to be able to qualify for any of the many applicable rates or ratings.

As one of the Navy Training Courses, this book was prepared by the U. S. Navy Training Publications Center, Washington, D. C., for the Bureau of Naval Personnel.

Printed 1955

Revised 1962

THE UNITED STATES NAVY

GUARDIAN OF OUR COUNTRY

The United States Navy is responsible for maintaining control of the sea and is a ready force on watch at home and overseas, capable of strong action to preserve the peace or of instant offensive action to win in war.

It is upon the maintenance of this control that our country's glorious future depends; the United States Navy exists to make it so.

WE SERVE WITH HONOR

Tradition, valor, and victory are the Navy's heritage from the past. To these may be added dedication, discipline, and vigilance as the watchwords of the present and the future.

At home or on distant stations we serve with pride, confident in the respect of our country, our shipmates, and our families.

Our responsibilities sober us; our adversities strengthen us.

Service to God and Country is our special privilege. We serve with honor.

THE FUTURE OF THE NAVY

The Navy will always employ new weapons, new techniques, and greater power to protect and defend the United States on the sea, under the sea, and in the air.

Now and in the future, control of the sea gives the United States her greatest advantage for the maintenance of peace and for victory in war.

Mobility, surprise, dispersal, and offensive power are the keynotes of the new Navy. The roots of the Navy lie in a strong belief in the future, in continued dedication to our tasks, and in reflection on our heritage from the past.

Never have our opportunities and our responsibilities been greater.

CONTENTS

ACTIVE DUTY ADVANCEMENT REQUIREMENTS

<table>
<tr><th>REQUIREMENTS *</th><th>E1 to E2</th><th>E2 to E3</th><th>E3 to E4</th><th>E4 to E5</th><th>E5 to E6</th><th>E6 to E7</th><th>† E7 to E8</th><th>‡ E8 to E9</th></tr>
<tr><td>SERVICE</td><td>4 mos. service—or completion of recruit training.</td><td>6 mos. as E-2.</td><td>6 mos. as E-3.</td><td>12 mos. as E-4.</td><td>24 mos. as E-5.</td><td>36 mos. as E-6.</td><td rowspan="2">48 mos. as E-7. 8 of 11 years total service must be enlisted. Must be permanent appointment.</td><td rowspan="2">24 mos. as E-8. 10 of 13 years total service must be enlisted.</td></tr>
<tr><td>SCHOOL</td><td>Recruit Training.</td><td></td><td>Class A for PR3, DT3, PT3. ǂ AME 3</td><td></td><td></td><td>Class B for AGCA, MUCA, MNCA.</td></tr>
<tr><td>PRACTICAL FACTORS</td><td>Locally prepared check-offs.</td><td colspan="7">Records of Practical Factors, NavPers 760, must be completed for E-3 and all PO advancements.</td></tr>
<tr><td>PERFORMANCE TEST</td><td colspan="2"></td><td colspan="4">Specified ratings must complete applicable performance tests before taking examinations.</td><td></td><td></td></tr>
<tr><td>ENLISTED PERFORMANCE EVALUATION</td><td colspan="2">As used by CO when approving advancement.</td><td colspan="6">Counts toward performance factor credit in advancement multiple.</td></tr>
<tr><td>EXAMINATIONS</td><td colspan="2">Locally prepared tests.</td><td colspan="4">Service-wide examinations required for all PO advancements.</td><td colspan="2">Service-wide, selection board, and physical.</td></tr>
<tr><td>NAVY TRAINING COURSE (INCLUDING MILITARY REQUIREMENTS)</td><td></td><td colspan="5">Required for E-3 and all PO advancements unless waived because of school completion, but need not be repeated if identical course has already been completed. See NavPers 10052 (current edition).</td><td colspan="2">Correspondence courses and recommended reading. See NavPers 10052 (current edition).</td></tr>
<tr><td rowspan="2">AUTHORIZATION</td><td colspan="2">Commanding Officer</td><td colspan="3">U.S. Naval Examining Center</td><td colspan="3">Bureau of Naval Personnel</td></tr>
<tr><td colspan="6">TARS are advanced to fill vacancies and must be approved by CNARESTRA.</td><td colspan="2"></td></tr>
</table>

* All advancements require commanding officer's recommendation.
† 2 years obligated service required.
‡ 3 years obligated service required.
ǂ Effective 1 Jan. 1963.

INACTIVE DUTY ADVANCEMENT REQUIREMENTS

<table>
<tr><th colspan="2">REQUIREMENTS *</th><th>E1 to E2</th><th>E2 to E3</th><th>E3 to E4</th><th>E4 to E5</th><th>E5 to E6</th><th>E6 to E7</th><th>E8</th><th>E9</th></tr>
<tr><td></td><td>FOR THESE DRILLS PER YEAR</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td rowspan="3">TOTAL TIME IN GRADE</td><td>48</td><td>6 mos.</td><td>6 mos.</td><td>15 mos.</td><td>18 mos.</td><td>24 mos.</td><td>36 mos.</td><td>48 mos.</td><td>24 mos.</td></tr>
<tr><td>24</td><td>9 mos.</td><td>9 mos.</td><td>15 mos.</td><td>18 mos.</td><td>24 mos.</td><td>36 mos.</td><td>48 mos.</td><td>24 mos.</td></tr>
<tr><td>NON-DRILLING</td><td>12 mos.</td><td>24 mos.</td><td>24 mos.</td><td>36 mos.</td><td>48 mos.</td><td>48 mos.</td><td></td><td></td></tr>
<tr><td rowspan="2">DRILLS ATTENDED IN GRADE †</td><td>48</td><td>18</td><td>18</td><td>45</td><td>54</td><td>72</td><td>108</td><td>144</td><td>72</td></tr>
<tr><td>24</td><td>16</td><td>16</td><td>27</td><td>32</td><td>42</td><td>64</td><td>85</td><td>32</td></tr>
<tr><td rowspan="3">TOTAL TRAINING DUTY IN GRADE †</td><td>48</td><td>14 days</td><td>14 days</td><td>14 days</td><td>14 days</td><td>28 days</td><td>42 days</td><td>56 days</td><td>28 days</td></tr>
<tr><td>24</td><td>14 days</td><td>14 days</td><td>14 days</td><td>14 days</td><td>28 days</td><td>42 days</td><td>56 days</td><td>28 days</td></tr>
<tr><td>NON-DRILLING</td><td>None</td><td>None</td><td>14 days</td><td>14 days</td><td>28 days</td><td>28 days</td><td></td><td></td></tr>
<tr><td colspan="2">PERFORMANCE TESTS</td><td colspan="2"></td><td colspan="6">Specified ratings must complete applicable performance tests before taking examination.</td></tr>
<tr><td colspan="2">PRACTICAL FACTORS (INCLUDING MILITARY REQUIREMENTS)</td><td colspan="8">Record of Practical Factors, NavPers 760, must be completed for all advancements.</td></tr>
<tr><td colspan="2">NAVY TRAINING COURSE (INCLUDING MILITARY REQUIREMENTS)</td><td colspan="8">Completion of applicable course or courses must be entered in service record.</td></tr>
<tr><td colspan="2">EXAMINATION</td><td colspan="6">Standard exams are used where available, otherwise locally prepared exams are used.</td><td colspan="2">Standard EXAM, Selection Board, and Physical.</td></tr>
<tr><td colspan="2">AUTHORIZATION</td><td colspan="5">District commandant or CNARESTRA</td><td colspan="3">Bureau of Naval Personnel</td></tr>
</table>

* Recommendation by commanding officer required for all advancements.
† Active duty periods may be substituted for drills and training duty.

CHAPTER 1

OPERATING PRINCIPLES OF THE ELECTRON TUBE

INTRODUCTION

Basic Electronics, NavPers 10087-A, presents many of the basic concepts in the field of electronics. Emphasis is placed primarily on the theory of operation of typical electronic components and circuits that have frequent application in naval electronic equipments. The description of specific equipments is left to the rating texts.

This training course is intended as a basic reference for all personnel of the Navy whose duties require them to have a knowledge of the fundamentals of electronics. However, it is not intended that each rate of each rating concerned must study every chapter in the book. The chapters on amplifiers, for example, includes an introductory chapter which will be useful to all who are concerned with amplifiers, plus two other chapters which cover d-c amplifiers and audio-power amplifiers which will be necessary only for those who service and maintain those types of equipment. For example the AT, ET, and FT ratings will use the entire training course whereas the IC, EM, and AE will use only the pertinent chapters. A detailed study guide will be found in Training Publications for Advancement in Rating, NavPers 10052.

In general, irrespective of the field, electrical networks comprise not more than four fundamental qualities—resistance, inductance, capacitance, and control devices such as vacuum tubes. *Basic Electronics* discusses the action of circuits and components in terms of these fundamental concepts and applies Ohm's law to the solution of related problems.

This training course introduces the subject of electronics as it is applied to electron tubes and power supplies. It then applies the concept of resonance, which is a basic quality of tuned circuits. Elementary mathematics, including algebra, geometry, and trigonometry is used to illustrate circuit behavior. Algebraic derivations are provided for those equations that require explanation. It should be understood, however, that the formula derivations are included only to strengthen the background of understanding of why particular components behave as they do under different circuit conditions. Those who find the formula derivations too difficult to follow should not be discouraged. If the reader acquires an understanding of what the characteristics of particular circuits are and how the circuits behave, he is getting the main points. In other words, he can study most parts of this text with or without the mathematical derivations of formulas; he will acquire more understanding of the fundamentals of electronics if he uses both text and formulas. Throughout the text, emphasis is placed on circuit behavior.

A knowledge of the principles of basic electricity is especially important to an understanding of basic electronics, and the student is urged to familiarize himself with these principles before attempting to read *Basic Electronics*.

This training course begins with a discussion of electron tubes, followed by their action in power supplies, amplifiers, and oscillators. Chapter 3 discusses tuned circuits as applied to band-pass and band-stop filters and to inductively coupled circuits.

The central portion of the text includes a discussion of modulation, detection, transmitters, receivers, antennas, and radio wave propagation.

The text concludes with a discussion of electronic test equipment, an introduction to radar, introduction to transistors, and digital computers.

The student is urged to study this training course thoughtfully and deliberately with pencil and paper at hand, and refrain from skimming the text. It is hoped that the rhyme, "This is the age of the half-read page," will not apply to those who read *Basic Electronics*.

The electron tube is considered primarily responsible for the rapid evolution of electronics to its present stage. It is one of the basic components of almost every piece of electronic equipment. Without the discovery and development of the electron tube, elaborate yet compact equipments, such as radio, radar, and sonar would not have been possible. It should, therefore, be apparent to the student that, in order to have a clear concept of electronic theory and the operating principles of electronic equipment, electron tube principles are of utmost importance.

The electron tube is made up of a highly evacuated glass or metal shell, which encloses several elements. The elements consist of the (cathode) emitter, the plate and sometimes one or more grids. Another element of importance in many tubes is the heater, sometimes called filament, which serves to heat the emitter.

Electron tubes are of many types and designations and perform many functions. They can be made to (1) convert currents and voltages from one waveform to another, (2) amplify weak signals with minimum distortion, and (3) generate frequencies much higher than any conventional a-c generator.

DIODES

The simple 2-electrode tube contains a heated cathode and a cold plate. DI is a prefix signifying two. ODE is the suffix as in electrode, cathode, and anode. The plate collects electrons when the cathode is heated in a vacuum, and a positive potential exists on the plate with respect to the cathode.

CONSTRUCTION

The original diode was constructed by Thomas Edison, inventor of the incandescent electric lamp, shortly after his invention of the lamp itself. He added a metal plate inside his evacuated lamp and provided an external terminal from it for use as an electrode. Then he used his heated filament as another electrode and arranged that diode as we shall describe for figure 1-2. A modern version of Edison's diode is shown in figure 1-1,A, with its two elements indicated as plate and filament, respectively.

Another modern version of a diode is shown in figure 1-1,B. Its filament serves only as a heater.

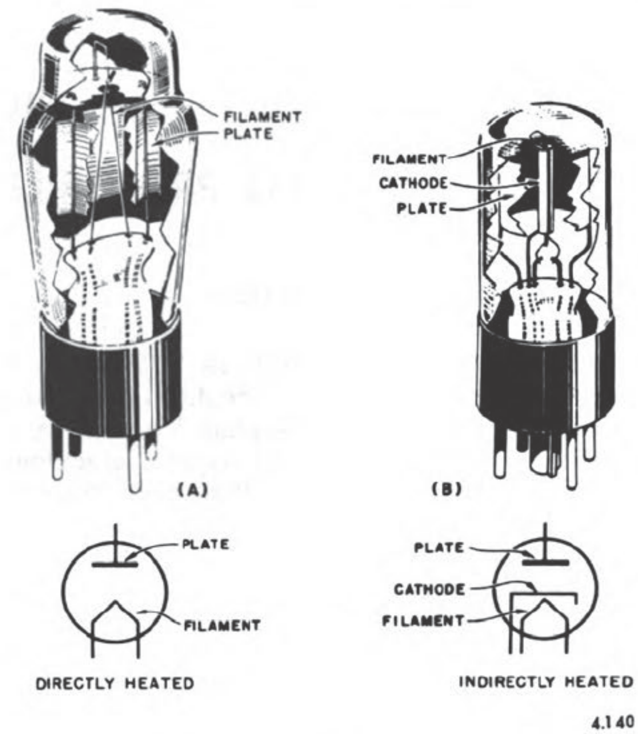


Figure 1-1.—Cutaway of 2-element tubes.

In an electrical circuit the two electrodes of a diode act in the manner of a valve (actually called a valve by the British instead of a tube). Let us look for this property as we examine figure 1-2.

OPERATION

The behavior of a diode is observed after connecting the plate and emitter elements in series with a battery and milliammeter, as shown in figure 1-2, carefully observing polarity changes of that battery when used in arrangements A and B, respectively. The emitter is brought up to normal temperature by applying rated voltage across the heater terminals. If the battery is connected so that the plate is positive with respect to the emitter (fig. 1-2,A), the meter will indicate a current flow. This phenomenon, the emission of electrons from hot bodies, first observed by Edison in 1883, is known as the Edison effect. However, if the battery is reconnected (fig. 1-2,B) so that the plate is negative with respect to the emitter, the meter will indicate no plate current flow. In chapter 3 you will learn how the battery in figure 1-2 is replaced with an alternating voltage

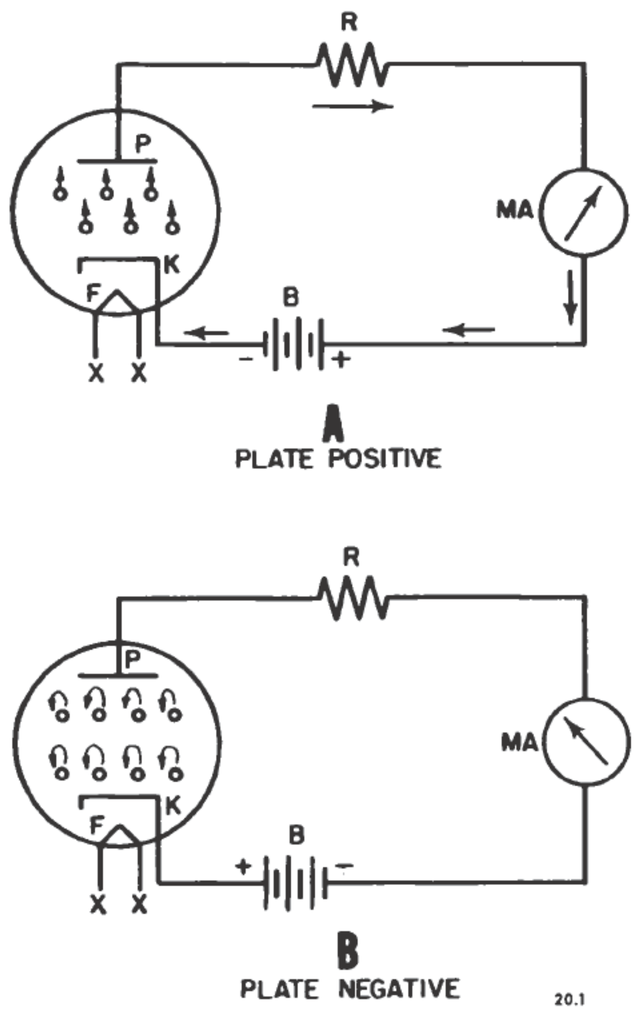


Figure 1-2.—Action of diode.

source to use this one-way conduction of the diode for converting alternating current into direct current.

Let us return to figure 1-2,B, and learn why the meter fails to show a flow of current. The total number of electrons emitted by the hot electrode at a given operating temperature is always the same regardless of the plate voltage. This same condition exists regardless of plate polarity because the electrons fly into the space surrounding the emitter to produce a cluster or cloud, which is in turbulence (great agitation). This cloud constitutes a negative space charge that constantly tends to repel the electrons toward, and into, the emitter as fast as they are being emitted. The negative charge on the plate of figure 1-2,B, only repels the nearby electrons within the cloud, but the action is so effective that none of the electrons reaches the plate (regardless of amount of voltage) as long as the plate remains negative.

Now let us return to figure 1-2,A. With low values of positive plate voltage, only those electrons of the space-charge cloud that are nearest to the plate are attracted to it, and the plate current is low. As the plate voltage is increased (the cathode temperature remaining constant), greater numbers of electrons are attracted to the plate and, correspondingly, fewer of those being emitted are repelled back into the cathode.

Eventually a plate voltage value (saturation voltage) is reached at which all the electrons being emitted are in transit to the plate, and none are repelled back into the cathode. The corresponding value of current is called the saturation current. Any further increase in plate voltage can cause no further increase in plate current flowing through the tube.

The relation between the plate current in a diode and the plate potential for different cathode temperatures for oxide-coated, tungsten, and thoriated-tungsten cathodes is shown in figure 1-3. At high plate voltages the flow of plate current is practically independent of plate voltage but is a function of the cathode temperature. However, at lower values of plate voltage the plate current is controlled by the voltage between the plate and cathode and is substantially independent of the cathode temperature.

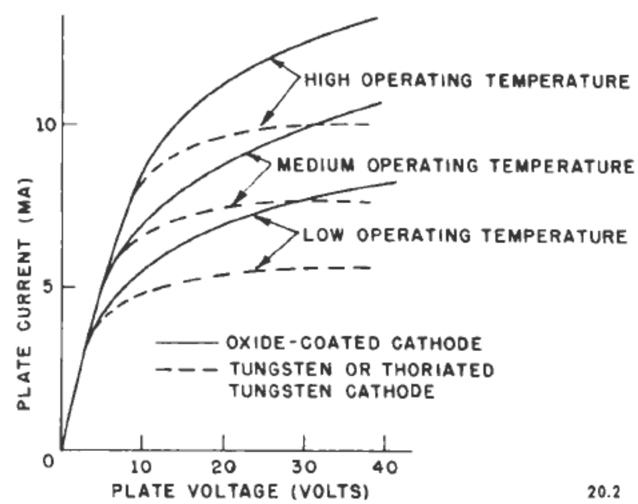


Figure 1-3.—Diode plate-current plate-voltage characteristic curves for various operating temperatures.

In other words, with a fixed plate voltage, electron emission and plate current will increase with cathode temperature until at some value of temperature the plate current is limited by the space charge. Thus, more electrons are being emitted by the cathode than are being attracted by the plate. Continued increase of cathode temperature fails to produce any further increase in plate current. The temperature at which the plate current stops increasing is called the saturation temperature.

The dotted portion of the characteristic curves is representative of tungsten and thoriated-tungsten emitters, and the solid curves are typical of oxide-coated emitters. It is unlikely that the plate current in a tube employing an oxide-coated emitter will ever be entirely independent of the plate voltage. Before the plate voltage could be increased sufficiently to produce emission saturation it is probable that the cathode would be damaged seriously.

TYPES

Diodes that have been discussed thus far are of the high-vacuum type. There are other types of diodes that contain gas at a relatively low pressure. For example, hot-cathode mercury-vapor rectifier tubes are used to provide plate power for transmitters. In other applications cold-cathode diodes containing gas at low pressure are used as voltage regulators, relaxation oscillators, and transmit-receive switching devices.

The original use of the word, diode, was restricted to vacuum tubes. Scientific research has extended our knowledge about other products that have identical properties of those earlier diodes, although they are not vacuum tubes. In the next chapter you will learn about semiconductor diodes that are identified by their predominating constituent as silicon diode, germanium diode, and selenium diode. The modern definition of diode, therefore, omits the earlier restriction that confined it to the vacuum tube.

USES

Since current can flow in only one direction through a diode its basic use is as a rectifier. If the battery in figure 1-2 is replaced with an alternating voltage source, current will flow through the load resistor in the plate lead only on alternate half cycles—when the plate is positive with respect to the cathode. This unidirectional characteristic of the diode is also used when the tube is employed as a detector.

A partial listing of other important uses of diodes includes clamper circuits, clippers, frequency converters, d-c restorers, frequency multipliers, harmonic generators, limiters, logic circuits in electronic computers, noise limiters, photo diodes, thermistors, voltage regulators, and volume expanders.

TYPES OF EMITTERS

Only a few substances can be heated to the high temperatures that are required to produce satisfactory thermionic emission without melting. Tungsten, thoriated-tungsten, and oxide-coated emitters are the only types that are commonly used in electron tubes.

TUNGSTEN EMITTERS

Tungsten has a great durability as an emitter but requires a large amount of heating power and a high operating temperature for satisfactory emission. Tungsten cathodes are used primarily in high-power electron tubes like those in high-power radio transmitting equipment.

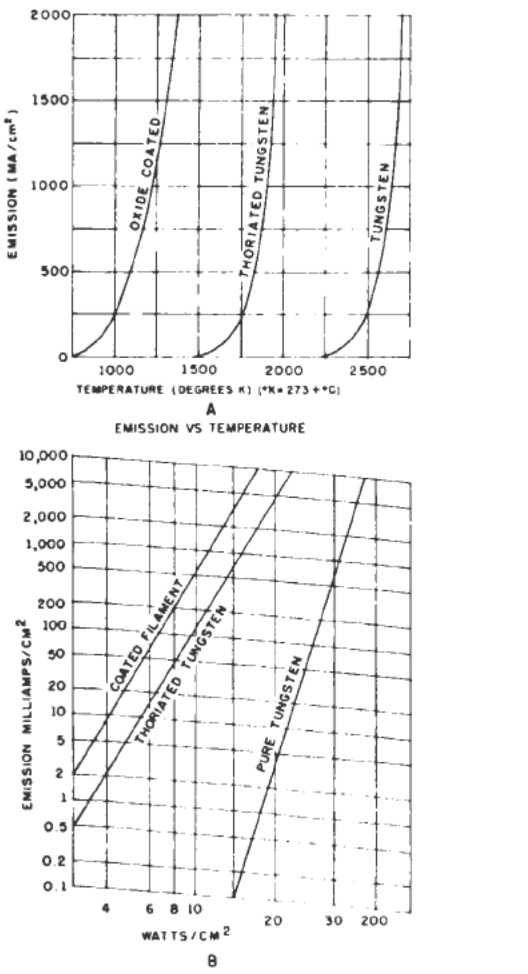
THORIATED-TUNGSTEN EMITTERS

A thoriated-tungsten emitter has a thin layer of thorium on the surface of the tungsten. The layer of thorium is monomolecular—that is, only 1 molecule thick. Thoriated-tungsten cathodes have greater electron emission at a lower operating temperuture than a cathode of pure tungsten and are normally used in tubes that are operated at plate voltages of 500 to 5,000 volts. Tubes such as the 860 and 861 use this type of emitter. These tubes and others like them are used extensively in low-power radio transmitters.

OXIDE-COATED EMITTERS

Oxide-coated emitters consist of metal, such as nickel, coated with a mixture of barium and strontium oxides, over which is formed a monomolecular layer of metallic barium and strontium. This is the most efficient type of emitter. It operates at a lower temperature than tungsten or thoriated-tungsten and therefore requires less power, resulting in a longer life at a higher emission efficiency. It is used in almost all types of receiving tubes.

The graphs in figure 1-4,A, show electron emission as a function of cathode temperature for the three types of emitter materials discussed in this chapter. The temperature at which emission becomes appreciable is called the normal operating temperature. The emitter comprises the cathode of the electron tube. The emission efficiency of the three types of emitter materials is shown in figure 1-4,B.

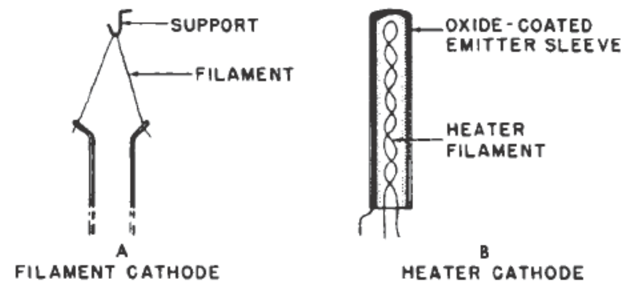


20.3

Figure 1-4.–Emission vs temperature curves for three types of emitters.

HEATING THE EMITTER

The electron-emitting cathodes of electron tubes are heated in two ways—directly, and indirectly. A directly heated emitter receives its heat by the passage of a current through the filament itself which serves as the cathode. An indirectly heated cathode comprises a metal sleeve that surrounds the filament but is electrically insulated from it. The sleeve serves as the cathode emitter and receives its heat mostly by radiation. Both types are shown in figure 1-5.



13.21

Figure 1-5.—Methods of heating the cathodes of electron tubes.

Directly heated cathodes are generally employed in portable equipment that is supplied from batteries. The d-c filaments of these tubes are so constructed that the drain on the filament battery will be low. Indirectly heated cathodes would require too much power for heating purposes. Because the filament current is steady, the heating is uniform; and the filament cross section is relatively small compared with a-c filaments. Directly heated a-c filaments require relatively large cross sections to reduce the temperature variations that occur at twice the power frequency. When a-c power is available, it is common practice in receiving equipment to employ indirectly heated cathodes. The cathode in this type of tube is isolated from the a-c- heater supply and therefore hum occurring at the power frequency (or at twice the power frequency) is largely eliminated.

PHYSICAL CHARACTERISTICS OF ELECTRON-TUBE MATERIALS

The outer walls of an electron tube are constructed either of thin glass or metal. The larger the tube, the thicker the glass must be because of the greater weight of the atmosphere to be sustained on the walls of the tube.

Evacuation of air from a tube is required for two reasons—to prevent destruction of the cathode and heating element by oxidation or burning, and to allow the flow of current from cathode to plate without colliding with gas particles. The lightest gas particle is approximately 1,800 times as heavy as an electron; thus a gas molecule would divert an electron upon impact and make the current flow erratic.

High vacuum is produced by burning a small amount of magnesium or barium, known as a "getter," inside the tube after it has been sealed off from the outside air and after most of the air has been removed with high-vacuum pumps. The getter is ignited by means of a high-frequency coil which is placed around the electron tube. The high-frequency field induces eddy currents in the metal within the tube. These eddy currents heat up a metal cap which contains a small charge of gun powder. This heat fires the magnesium getter and combines with gas left in the tube to form a silvery deposit on the inner walls. This deposit of magnesium carbonate occupies much less space than did the gas, hence the degree of vacuum is increased.

The external leads from the tube are electrically welded to the tube elements and brought out at the bottom through a special glass-metal fusion to make the envelope airtight. In metal tubes a glass button is used at the base to afford electrical insulation. The materials selected for the external leads have nearly the same coefficient of expansion as that of glass. Thus during heating and cooling periods, the glass expands and contracts the same amount as the metal and the vacuum seals are maintained.

Metal tubes are an outgrowth of the competitive field in tube manufacturing. They are designed to act as a shielded unit (the same as a glass tube with an external shield placed over it). The shield on a tube acts primarily to prevent the introduction of stray fields within the envelope where induced voltages might be amplified many times, thus causing distortion in the output stages. There are a few circuits in electronic equipments where metal and glass tubes cannot be interchanged. Before making an interchange, the technician should always investigate.

The spacing of the electrodes in a tube is dependent on many factors but the two most important are frequency utilization and interelectrode voltages.

The anode (plate) is made of materials that will not emit electrons by thermionic means at normal tube operating temperatures. Metals used as plates include iron, nickel, carbon, and tantalum. The plate is mounted externally with respect to the cathode. It is electrically insulated from the cathode and usually surrounds it in order to receive all of the cathode field of emission. The plate usually has a dark surface to radiate the heat caused by the plate current.

Electron tubes are identified by a number or a combination of numbers and letters. So many different types of tubes have been introduced that it has become impossible to adhere rigidly to the system as it was originally set up. However, some of the ideas contained in the original system have been followed for many years. In the old system, the type number is divided into four parts. First, a number consisting of one or more digits designates the filament or heater voltage. Second, one or more letters designate the type or function of the tube. Third, a number designates the number of useful elements in the tube. Fourth, one or more letters designate the size or construction. For example, the type 6SK7 electron tube is a variable-mu pentode. The number 6 designates a filament voltage of 6.3 volts; the letter S denotes a single-ended construction (no grid cap) and inter-lead shields; the letter K designates variable-mu characteristics; the number 7 designates seven connections to elements of the tube; the GT designates that it is a glass tube.

TYPES OF EMISSION

Electrons flow within a conductor when a potential difference is applied across the terminals of the conductor. These electrons break away from the outer shells of their parent atoms and move with a rapid vibratory motion, the velocity of which increases with temperature. At ordinary temperatures the particles do not leave the surface of the conductor because their velocity is not great enough to overcome the attractive forces within the conductor.

To escape from a metallic surface, electrons must do work to overcome the forces of attraction which are always present. This amount of work is called the work function of the material. Increasing the heat intensity of a metallic emitter increases the kinetic energy of the so-called free electrons in the material.

THERMIONIC EMISSION

Thermionic emission is the process by which electrons gain enough energy by means of heat to escape from the surface of the emitter. Thermionic emission is the type of emission most frequently employed in electron tubes.

PHOTOELECTRIC EMISSION

An emission of electrons can also be caused by light striking the surface of certain materials. This type of emission is called photoelectric emission. The energy of the light rays striking the substance is imparted to electrons on the surface. If the energy acquired by the electrons is sufficient, the force thus acquired will overcome the attractive forces at the surface and the electrons will escape from the substance. The velocity at which the electrons are emitted is directly proportional to the light frequency of the radiant energy striking the material; therefore, the higher the light frequency (shorter the wavelength) the greater is the velocity of emitted electrons. The number of electrons emitted is directly proportional to the intensity of the light. Materials that are particularly sensitive to light are zinc, potassium, and the other alkali metals. Two of the principal uses of photoelectric emission are photoelectric cells and television camera, or iconoscope, tubes.

SECONDARY EMISSION

Emission of electrons from a body caused by the impact of other electrons striking its surface is called secondary emission. If a stream of electrons flowing at a high velocity strikes a material, the force may be great enough to dislodge other electrons on the surface. Secondary emission is not commonly used as a source of electrons. However, it does occur spontaneously in tubes and must be controlled. This problem is discussed later in this chapter.

TRIODES

CONSTRUCTION

The triode, or 3-element electron tube, is similar in construction to the diode, except that a grid of fine wire is added between the cathode and the plate. The addition of the grid gives to the tube its most useful function—the ability to amplify. It is common practice to make the grid in the form of a spiral helix of circular or elliptical cross section with the cathode at the center. Other arrangements, however, may be used provided the essential requirement of being able to control the flow of plate current is met. The space between the meshes is sufficiently large not to block the flow of electrons from cathode to plate. On the other hand, the grid mesh is sufficiently small and close enough to the cathode to control effectively the flow of plate current when the proper voltage is applied between the grid and cathode. The grid is called the control grid (G_1) to distinguish it from other grids that are used in multi-element tubes.

The construction features of a typical triode are shown in figure 1-6. Electrical connections to the grid and plate are made through the base pins and support wires. The cathode sleeve is insulated from the filament and is connected by means of a short lead to one of the base pins. The grid is seen to be much closer to the cathode than to the plate.

OPERATION

Plate current in a given diode depends on the plate voltage and the cathode temperature. Plate current in a triode depends not only on these factors, but also on the grid-to-cathode voltage. A small change in grid voltage causes a relatively large change in plate current. The effective grid control of plate current is caused by its close proximity to the cathode and its placement in a region of the heaviest negative space charge. A small change in grid voltage will produce the same variation in plate current that is produced by a much larger variation in plate voltage. If the grid-to-cathode voltage is increased sufficiently and the grid is negative with respect to the cathode, plate current will stop flowing. The smallest voltage between grid and cathode, with the grid end negative, that will cut off the flow of plate current is called the cutoff bias.

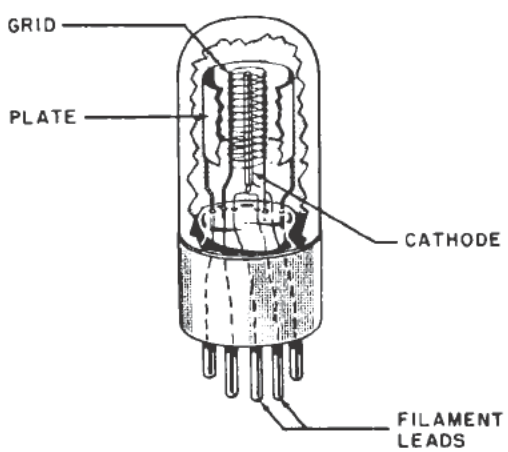


Figure 1-6.—Representative triode.

Figure 1-7 indicates the effect on plate current of making the control-grid voltage progressively less negative with respect to the cathode. When the negative bias is high (cutoff or higher), as in figure 1-7,A, no plate current flows because the negative charge on the grid is sufficient to repel the electrons back toward the cathode. As the bias is reduced (fig. 1-7,B) more electrons pass through the grid spaces on their way from the cathode to the plate. With zero bias (fig. 1-7,C) the grid has little or no control on the electron flow to the plate, and the triode operates much the same as a diode. As long as the grid is negative with respect to the cathode, no grid current flows and no power is consumed in the grid circuit.

If the grid is made positive with respect to the cathode the electrons in the space charge are accelerated toward the plate. Some of them, however, will be attracted to the grid, and grid current will flow, the amount depending on the magnitude of the positive voltage on the grid. Power will then be dissipated in the grid circuit. When this power dissipation is undesirable, the grid bias is increased to the point where the peak positive a-c signal voltage will not cause the grid to be positive with respect to the cathode and no grid current will flow.

AMPLIFICATION

The grid may be considered as an electronic control valve that regulates the flow of electrons through the tube and through the load in the plate circuit. Thus an a-c signal of sine waveform appearing in series with the grid bias causes the plate current to vary in the same manner. The variations in plate current through the plate load are accompanied by corresponding variations in plate voltage. These plate voltage and current variations constitute the output signal of the stage. A relatively small variation in grid

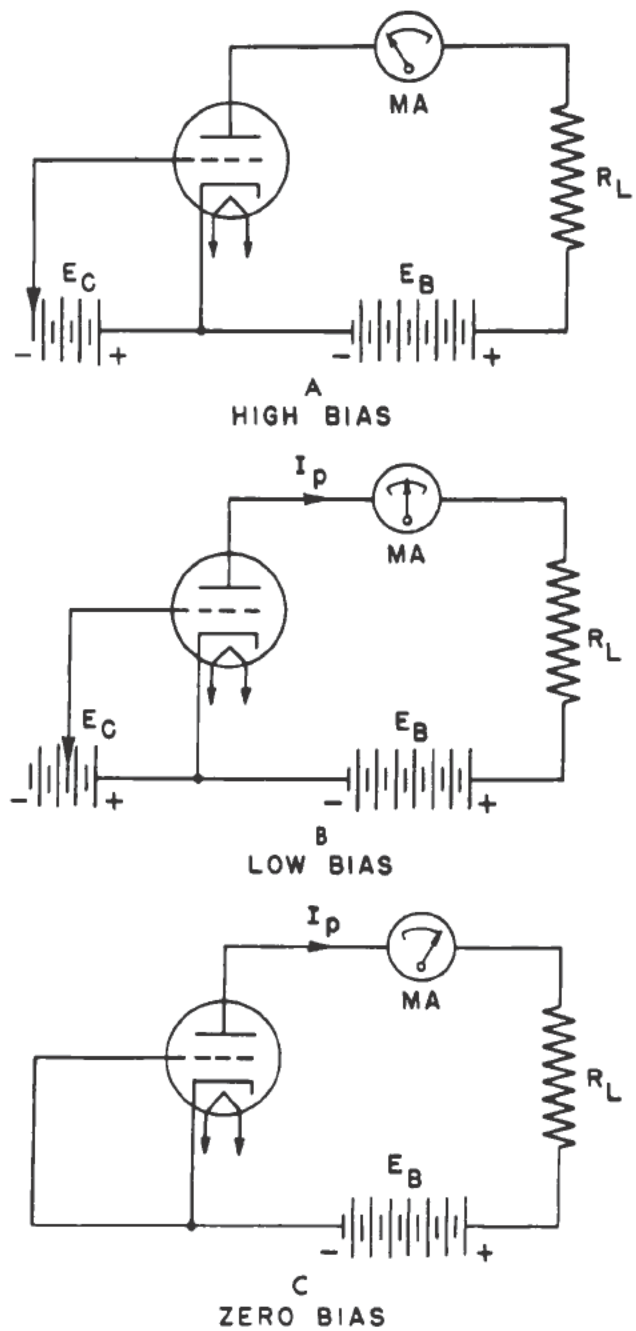


Figure 1-7.—Effect of control-grid voltage on plate current.

input signal is accompanied by a relatively large variation in output signal. Thus the grid signal is said to be amplified in the plate circuit.

TUBE CHARACTERISTICS

The characteristics of electron tubes with cathode, grid, and plate elements involve the relation between grid voltage, plate voltage, and their resulting plate current. Their respective symbols are

e_g = grid voltage
e_p = plate voltage
i_p = plate current

Notice that the symbol, e, represents any chosen voltage value, and, i, indicates any value of current. The subscript symbols, g and p, identify the grid circuit and plate circuit, respectively. If the Greek letter Δ (delta) is used to symbolize the phrase, a small change in, then:

Δe_g = a small change in grid-circuit voltage
Δe_p = a small change in plate-circuit voltage
Δi_p = a small change in plate-circuit current

The above values are always subject to changes. Therefore, they are classed as variable characteristics. However, the relationships between them often produce unchanging (that is, constant) characteristics.

The three important constant characteristics of a triode (as well as other tubes to be described later) are amplification factor, designated μ; a-c plate resistance, designated r_p; and transconductance, designated g_m.

The parameters are determined from the linear portions of a characteristic curve unless there is a specific demand for producing a variable parameter as, for example, in "variable-mu tubes" where a variable amplification factor is desired. Then the nonlinear characteristic portion of the curve is employed. Variable-mu tubes are explained later in this chapter.

Linear and nonlinear characteristic curves can be of either the static or the dynamic type. Both static and dynamic characteristic curves exist for each electron tube. They differ in shape as well as in the actual values they represent. A simple explanation of the difference between these two types of curves is that in static characteristics the values are obtained with different d-c potentials applied between grid, cathode, and plate, and the results are not typical of actual circuit operation. The dynamic characteristics are the values obtained with both a-c and d-c components present as in actual operation. The static characteristics provide an understanding of how the tube itself operates and are discussed in this chapter.

The first characteristic, which is a measure of the voltage amplification of which a tube is capable, is known as the amplification factor, designated μ (pronounced mu). It is the ratio of the increase in plate voltage to the increase in grid voltage required to produce the same change in plate current. An i_p-e_g characteristic for a triode is shown in figure 1-8. A value of plate voltage, e_p, is selected and the grid voltage, e_g, is adjusted to operate the tube at point A on the curve (with a 6-volt e_g, let us say for example, while the e_p is 200 volts).

Next, the value of e_g is changed a specific amount (either increased or decreased, but for our example let us choose a decrease from -6 volts to -4 volts, thus giving Δe_g a value of 2 volts, and the plate current has increased from point A to point B. The resulting Δi_p is shown in figure 1-8.

The plate current can be reduced to its original value by reducing the plate voltage, e_p, from 200 volts to 160 volts (point C) thus giving Δe_p a value of 40 volts.

The obvious conclusion is now reached that Δe_g (a value of 2 volts) produces the same effect as if it were Δe_p (a value of 40 volts).

The amplification factor is determined by the ratio of Δe_p to Δe_g, and is expressed as

$$\mu = -\frac{\Delta e_p}{\Delta e_g} \qquad (i_p \text{ constant})$$

The minus sign indicates that the changes in plate and grid voltages are in opposite directions. Triodes have practical amplification factors of from 3 to 100.

A second important characteristic is the variational, or a-c plate resistance, designated r_p. It is the ratio, for a constant grid voltage, of a small plate voltage change, Δe_p, to the resulting small plate current change, Δi_p. It is expressed in ohms when Δe_p is in volts and Δi_p is in amperes. Three i_p-e_p characteristic curves for a triode are shown in figure

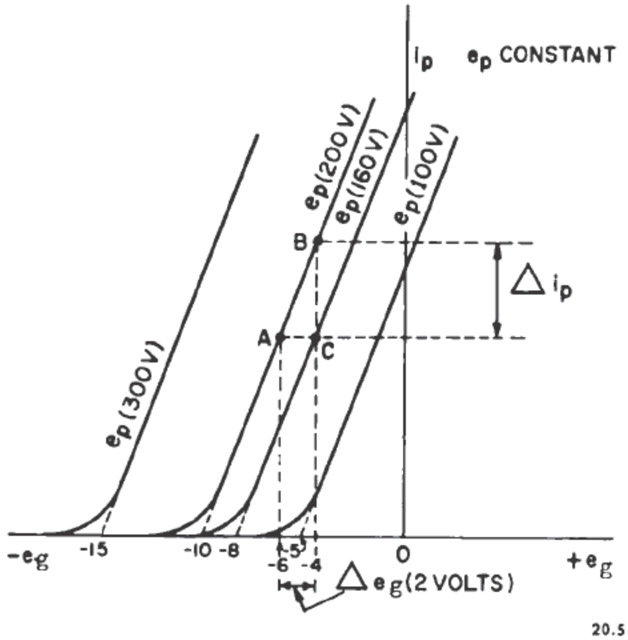


Figure 1-8.—Triode i_p-e_g curve.

1-9. The middle curve is arbitrarily chosen and the grid bias of -2 volts held constant as the plate voltage is adjusted to operate the tube at point A. The plate voltage is increased an amount Δ e_p so that the tube now operates at point B. The ratio of this small increase in plate voltage, Δ e_p, to the small increase in plate current, Δ i_p, which it produces is a measure of the variational, or a-c, plate resistance. Thus,

$$r_p = \frac{\Delta e_p}{\Delta i_p} \qquad (e_g \text{ constant})$$

A third characteristic used in describing the properties of electron tubes is the grid-plate transconductance, designated g_m. It is defined as the ratio, with plate voltage held constant, of a small change in plate current to the small change in grid voltage that causes the change in plate current. It is usually expressed in micromhos. The mho is the unit of conductivity and is the reciprocal of the ohm, or $\frac{I}{E}$. The word, mho, is ohm, spelled backward.

We recall that figure 1-8 shows the i_p-e_g characteristics for a triode. In a middle curve of this figure the plate voltage is held constant at 200 volts and the grid voltage is adjusted so that the tube is operated at point A. The grid voltage is reduced an amount, Δe_g, and the tube then operates at point B. The ratio of the small change in plate current, Δ i_p, to the small change in grid voltage, Δ e_g, indicates the transconductance—that is,

$$g_m = \frac{\Delta i_p}{\Delta e_g} \qquad (e_p \text{ constant})$$

If i_p is expressed in amperes and e_g in volts, g_m must be multiplied by 1,000,000 to express the result in micromhos.

These tube characteristics are interrelated and depend primarily upon the tube structure. This relation is defined by the expression

$$\mu = g_m r_p,$$

where g_m is in mhos and r_p is in ohms.

DISTORTION

Figure 1-10 illustrates the effects on the output current curve of shifting the bias from a value that allows the tube to operate on the straight portion of the i_p-e_g characteristic curve to a value that forces it to operate on the nonlinear portion of the curve. When the operating bias is at point A (-3 volts), the grid-voltage variation is within the limits of

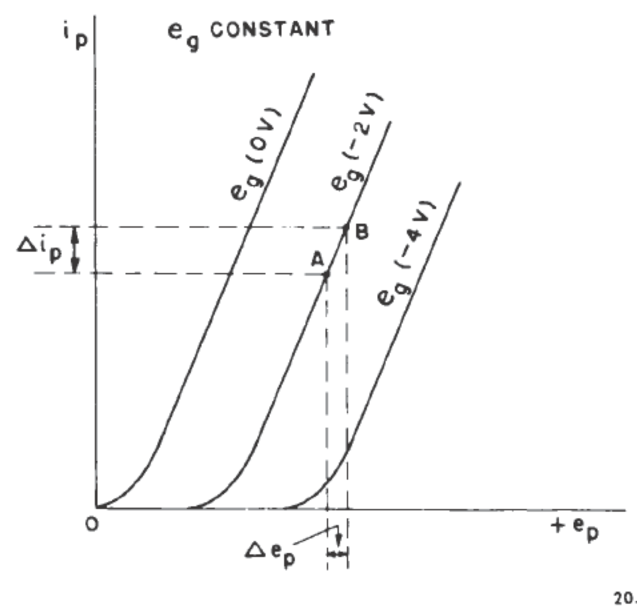


Figure 1-9.—Triode i_p-e_p curve.

the straight-line portion of the characteristic curve and the plate current faithfully reproduces the grid-voltage waveform. However, if the fixed bias is increased to point B (-7 volts), the amplitude of the output waveform is considerably distorted. The extent of this distortion depends upon the actual biasing point of the tube and the extent of the grid-voltage swing.

The point on the zero axis intersected by the characteristic curve (point C in figure 1-10) is commonly known as the cutoff point. An amplifier biased to cutoff functions much as a diode rectifier because only alternate half cycles are reproduced in the output circuit. When an amplifier is biased well beyond cutoff and is driven with an excessively large input grid voltage, only that part of the grid-voltage waveform extending into the operating region of the characteristic curve is reproduced in the output. The input signal is thus distorted in the output because only a small portion is amplified. Other forms of the distortion are treated in the chapters on electron-tube amplifiers.

INTERELECTRODE CAPACITANCE

Capacitance exists between any two metal surfaces separated by a dielectric. The amount of capacitance depends upon the area of the metal surfaces, the distance between them, and the type of dielectric. The electrodes of an electron tube produce a similar characteristic, known as interelectrode capacitance, which is illustrated schematically in figure 1-11. The capacitances that exist in a triode are the grid-to-cathode capacitance, the grid-to-plate capacitance, and the plate-to-cathode capacitance.

The shunting effect of the interelectrode capacitance of a tube is increased when the electrodes are connected to a circuit having grid, plate, and cathode leads of appreciable length. The capacitance is increased because of the increase in area afforded by the conducting surfaces comprising the circuit wiring, tube bases, sockets, and so forth.

At low and medium frequencies the interelectrode capacitances, as well as the distributed capacitances due to circuit wiring, have only a slight shunting effect because the reactance at these frequencies is high compared with that of other circuit components.

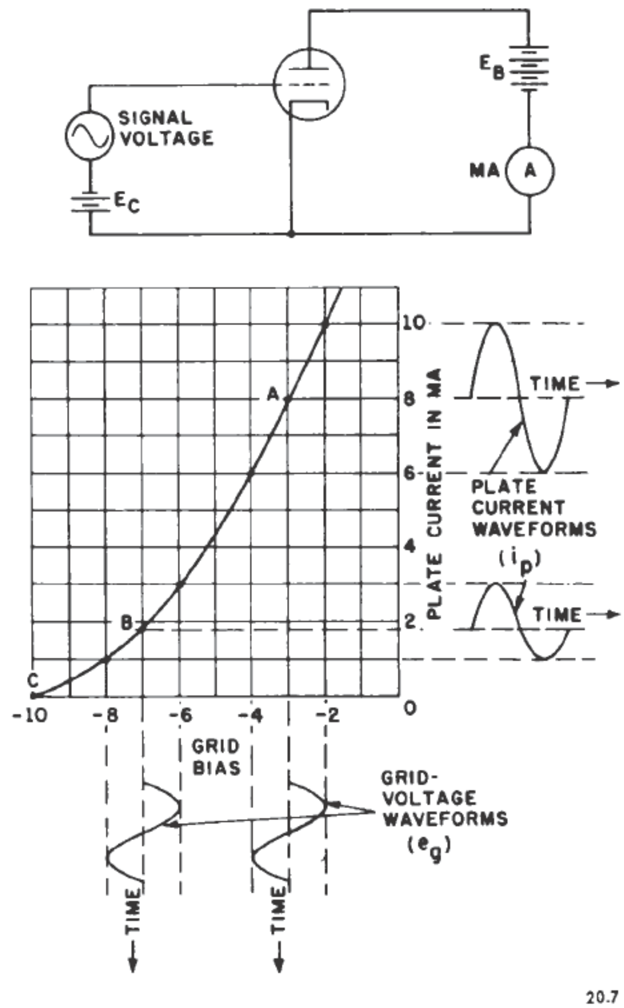


Figure 1-10.—Nonlinear operation due to excessive bias.

At high frequencies the interelectrode and distributed capacitances cause appreciable shunting effect because of the reduced reactance offered at these frequencies. Also the grid-to-plate capacitance can feed back some of the plate-signal voltage in the proper phase with respect to the grid-signal voltage to cause undesired oscillations. The effect of this interelectrode capacitive feedback can be neutralized by introducing, by means of a capacitor, a voltage of equal magnitude and opposite polarity from the plate to the grid circuit. Such an external capacitor is called a neutralizing capacitor. It is usually variable to permit adjustment for precise cancellation of the objectionable internal feedback voltage.

At ultrahigh frequencies (u-h-f) interelectrode capacitance becomes very objectionable

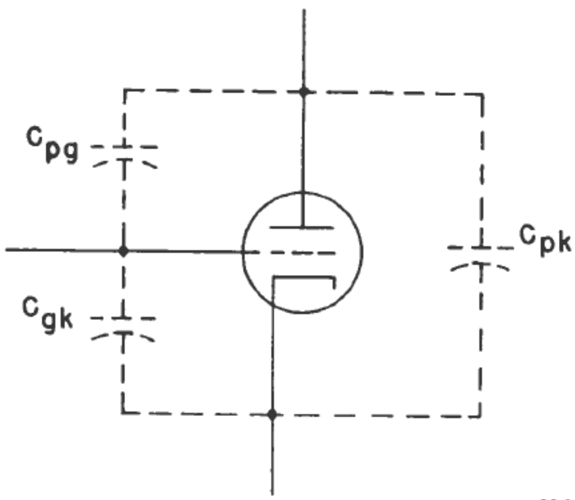


Figure 1-11.—Schematic representation interelectrode capacitance.

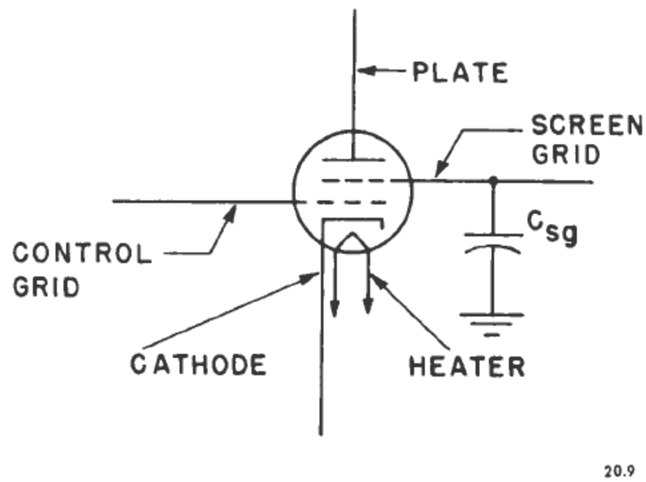


Figure 1-12.—Schematic diagram of a tetrode.

and prevents the use of ordinary electron tubes. Special u-h-f tubes are used at such operating frequencies. These are characterized by tube elements having very small physical dimensions and closely spaced electrodes that often do not terminate in conventional tube bases.

MULTIELEMENT TUBES

Many desirable characteristics may be attained in electron tubes by the use of more than one grid. Some common types include tetrodes, which contain 4 electrodes and pentodes, which contain 5 electrodes. Others containing as many as 8 electrodes are available for certain applications.

TETRODES

The relatively large values of interelectrode capacitances of the triode, particularly the plate-to-grid capacitance, impose a serious limitation on the tube as an amplifier at high frequencies. To reduce the plate-to-grid capacitance, a second grid called a screen grid (G_2) is inserted between the grid and plate of the tube, as shown in figure 1-12.

Because the screen grid is shunted by a screen bypass capacitor, C_{sg}, having a low reactance at the signal frequency, it acts as a shield or screen between the plate and control grid. It effectively reduces the interelectrode capacitance coupling between the plate and control grid circuits. The screen is supplied with a potential somewhat less positive than the plate. The positive voltage on the screen grid accelerates the electrons moving from the cathode. Some of these electrons strike the screen and produce a screen current which generally serves no useful purpose. The larger portion, however, passes through the open-mesh screen grid to the plate.

Because of the presence of the screen grid, a variation in the plate voltage has little effect on the flow of plate current. The control grid, on the other hand, retains its control as in the triode. The tetrode has high-plate resistance and an amplification factor ranging up to 800. The high amplification factor is brought about by the close proximity of the control grid to the cathode and the electrical isolation of the plate from the control grid. The transconductance of tetrodes is also relatively high compared with that of triodes.

A typical family of i_p-e_p characteristic curves of a tetrode is shown in figure 1-13,A.

The negative slope of the plate characteristic at plate voltages lower than the screen voltage (90 v) is the result of secondary emission from the plate. This condition results from the fact that with the screen voltage fixed, the velocity with which the electrons strike the plate increases with plate voltage. When the electrons strike the plate with sufficient force, other loosely held electrons are knocked out of the plate material into the space between the plate and the screen. Because the screen is

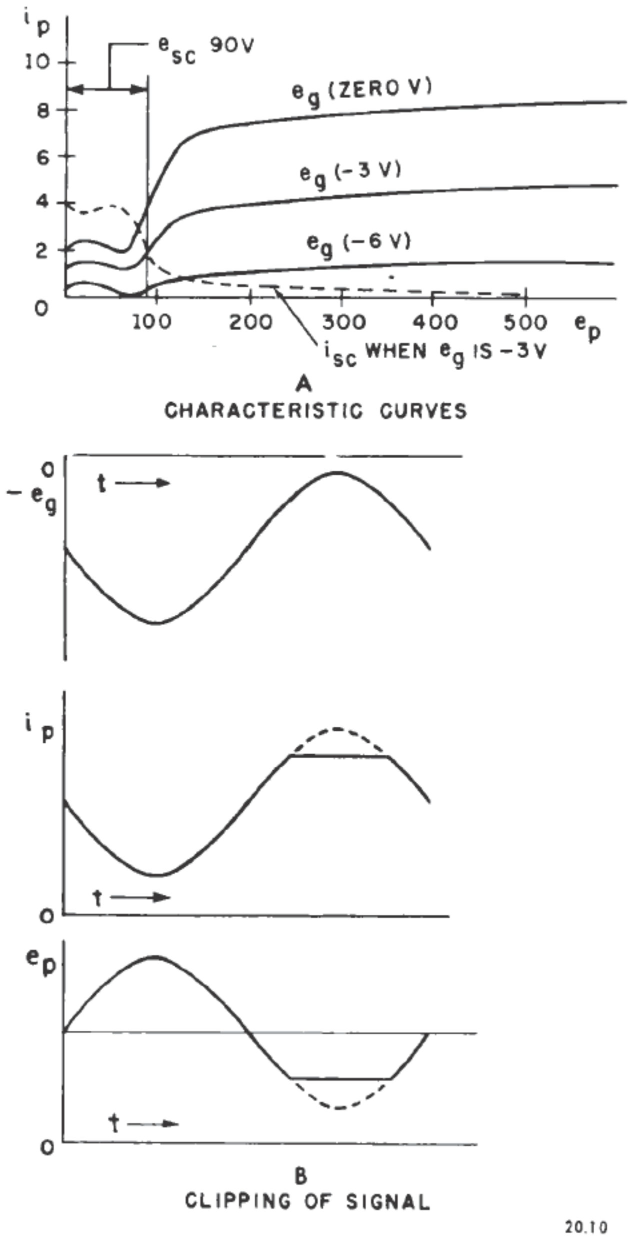


Figure 1-13.—Representative i_p-e_p tetrode characteristic curves.

at a higher positive potential than the plate, these secondary electrons are attracted to the screen. The flow of these electrons to the screen is in the opposite direction to the normal flow from cathode to plate and the plate current is decreased. This reduction in plate current continues until the potential of the plate approaches the screen-grid potential. Further increase in plate voltage causes the secondary electrons to be pulled back to the plate and the plate current again increases.

The action in the region where plate current decreases as plate voltage increases is called negative resistance. This action is opposite to that encountered in a normal resistor. When the tetrode is used as an amplifier, plate voltage should not fall below the screen voltage. If plate voltage falls below that of screen, plate current will fail to follow the grid-signal waveform and the output-signal plate-voltage variation is clipped as shown in figure 1-13,B. This distortion may be eliminated by reducing the amplitude of the grid signal or increasing the B-supply voltage. However, the relatively large screen current and the effects of secondary emission from the plate limit the usefulness of the tetrode as an r-f voltage amplifier.

PENTODES

The effects of secondary emission in the tetrode may be eliminated by the addition of a third grid. The pentode (5-element tube) includes a suppressor grid inserted between the screen grid and the plate for the purpose of preventing the screen from attracting secondary electrons from the plate. The 5 elements are cathode, control grid (G_1), screen grid (G_2), suppressor grid (G_3), and plate. Figure 1-14 is a schematic diagram of a pentode.

In the pentode, the suppressor grid (usually internally connected to the cathode) serves to repel or suppress secondary electrons from the plate. It also serves to slow down the primary electrons from the cathode as they approach the suppressor. These actions do not interfere with the flow of electrons from cathode to plate but serve to prevent any interchange of secondary

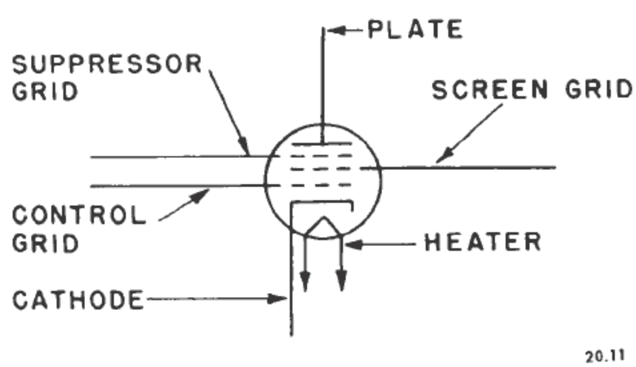


Figure 1-14.—Schematic diagram of a pentode.

electrons between screen and plate. The suppressor thus eliminates the negative resistance effect which appears in the tetrode in the region where plate voltage falls below that of the screen. Thus plate current rises smoothly from zero up to its saturation point as plate voltage is increased uniformly with grid voltage held constant. Typical pentode i_p-e_p characteristic curves are shown in figure 1-15.

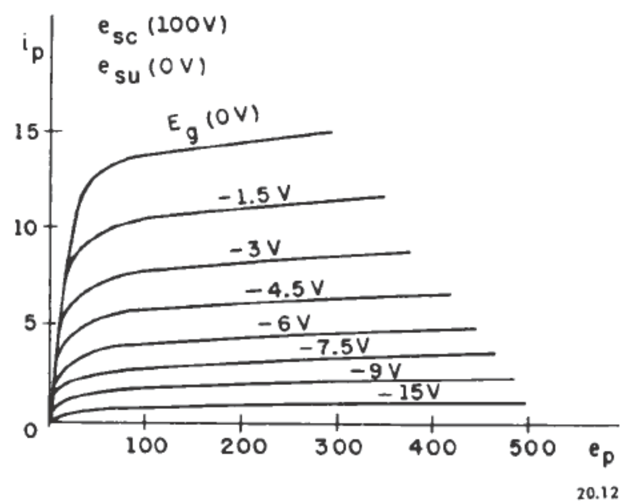


Figure 1-15.—Pentode i_p-e_p characteristic curves.

Pentodes produce an increased voltage-output signal for a given input-grid voltage compared with triodes. The amplification factor of pentodes ranges up to 1,500. The plate resistance and transconductance are both high. In the r-f pentode the chief purpose of the screen grid is to eliminate the effects of interelectrode capacitance coupling between control grid and plate circuits. In the power pentode, at audio frequencies, the screen permits the output signal plate voltage variation to be relatively large without the degenerative action occurring in the triode. Plate current is substantially independent of plate voltage in the power pentode since the screen voltage is the principal factor influencing plate current. With the addition of the suppressor the allowable output voltage variation is larger than that of the tetrode, and the distortion effects shown in the tetrode of figure 1-13,B, are eliminated. Thus an audio-frequency power pentode has an allowable output voltage variation in which the plate voltage can fall a large amount below that of the screen voltage on the positive half cycle of input signal without clipping the plate signal current. Thus the ratio of output power to grid driving voltage is relatively large.

BEAM-POWER TUBES

A beam-power tube is so named because it is constructed so that the electrons flow in concentrated beams from the cathode through the grids to the plate. The principal differences in construction between the beam-power tube and a normal tetrode and pentode are that in the beam-power tube the spaces between the turns of the grids are lined up, two beam-forming plates are added, and the spacing between the screen and plate is usually greater.

The internal structure of a beam-power tube is shown in figure 1-16. Because the spaces between the grids are lined up, fewer electrons strike the screen grid; therefore, screen grid current is lower and plate current higher than in other pentodes. Furthermore, when no actual suppressor is used, the beam-forming plates at cathode potential produce the desired beam effect. Secondary emission from the plate is then reduced because of the space charge between the screen grid and plate.

The space charge results from the slowing down of the electrons as they pass from the high-potential screen to the lower potential plate. The space charge thus formed in front of the plate is sufficient to repel back to the plate secondary electrons emitted as a result of the impact of primary electrons. This action also

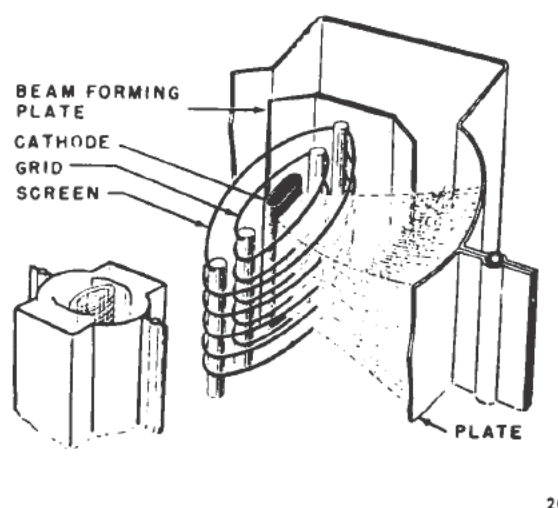


Figure 1-16.—Internal structure of a beam-power tube.

increases plate current and reduces the screen current.

A beam-power tube that is operated at the same plate and screen voltages as a normal tetrode provides more power output for a given signal voltage with no increase in internal tube capacitances.

VARIABLE-MU TUBES

The amplification of a tube may be controlled by varying the bias voltage applied to the grid, but normally the range of this control is limited by the cutoff bias and the permissible distortion. In receivers employing automatic volume control (a-v-c) in the r-f amplifier section, the amplification is varied over a wide range so that strong or weak signals may be accommodated. To permit this increased range of volume control, the variable-mu tube was developed. This tube is also known as the supercontrol or remote cutoff type.

The only difference in construction between variable-mu tubes and normal, or sharp cutoff tubes is in the spacing between the turns of the control grid. In sharp cutoff tubes the turns of the grid wires are equally spaced, while in remote cutoff types the grid turns are closely spaced at the ends and widely spaced in the center. The construction of variable-mu tubes is shown in figure 1-17,A.

With a small bias voltage, electrons flow through all the spaces of the grid and the amplification factor is relatively large because of the close spacing of the end turns of the control grid. As the bias is increased, the electron flow is cut off through the narrow spaces at the ends of the grid structure. However, they are still able to pass through the relatively large spaces at the center of the grid. The increased bias causes a decrease in the amplification due to the coarser turns in the central portion of the grid. A much greater value of bias is required to cut off the plate-current flow in this type of tube. The remote-cutoff tube is so named because the cutoff bias value is greater than (more remote from) the value required to cut off plate-current flow in tubes of evenly spaced turns.

Figure 1-17,B, shows the i_p-e_g curves for both a conventional sharp cutoff tube and a variable-mu or remote-cutoff tube. The cutoff bias for the normal tube is -5 volts, and because the slope is almost constant any change in bias produces little change in amplification. Contrasted with this characteristic, the curve for the variable-mu tube has a pronounced change in slope as the grid bias is increased from -10 volts to -15 volts and a small value of plate current is still flowing at a bias of -25 volts. The changing slope of this curve indicates a variation of amplification with bias. Thus, if a variable-mu tube is used with a bias source that varies with the signal strength, the output signal can be made substantially independent of the input signal strength. Automatic-gain-control circuits employing variable-mu tubes are discussed in connection with "Receivers" in chapter 12.

MULTIGRID TUBES

Electron tubes may be constructed with 4, 5, or 6 grids (fig. 1-18) in order to obtain certain characteristics. The grids may be used to influence the plate-current flow by introducing additional signal voltages having different frequencies, as in pentagrid converters or pentagrid mixers used in superheterodyne receivers. These applications are treated in chapter 12.

MULTIUNIT TUBES

To reduce the number of tubes in radio circuits, the electrodes of two or more tubes frequently are placed within one envelope. Multiunit tubes generally are identified according to the way the individual types contained in the envelope would be identified if they were made as separate units. Thus, a multiunit tube may be identified as a duplex-diode, a diode-pentode, a diode-triode-pentode, a pentagrid converter, and so forth. A number of multiunit tubes are shown in figure 1-19. Other tubes that are not shown include the double-triode and the triple-triode types.

TUBES OPERATING AT ULTRAHIGH FREQUENCIES

As the operating frequency is increased, the capacitive reactance between electrodes in electron tubes decreases—that is,

$$X_C = \frac{1}{2\pi fC}.$$

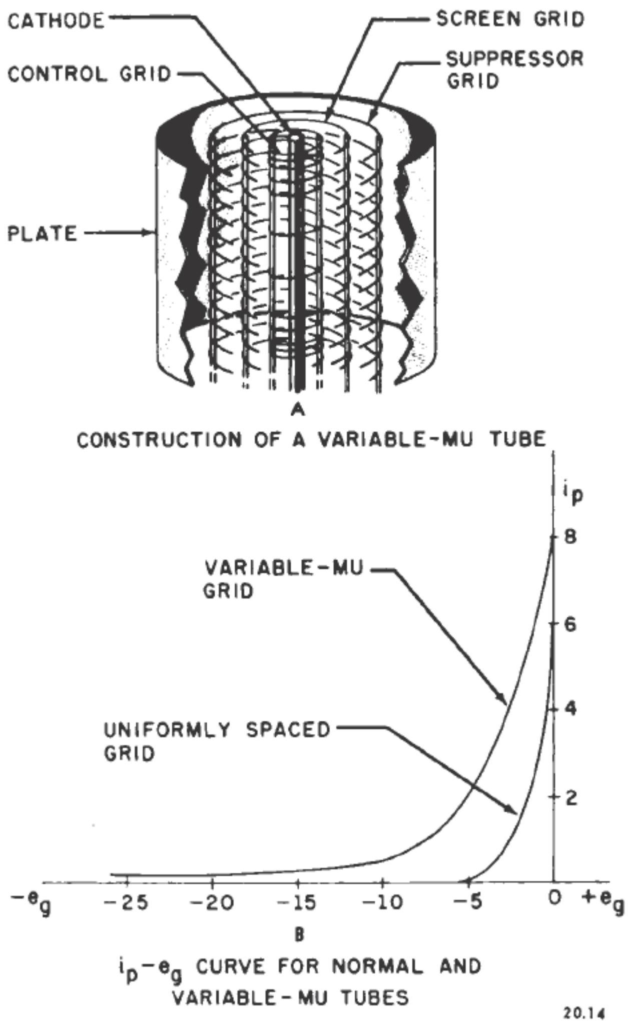


Figure 1-17.—Construction of variable-mu tubes and i_p-e_g curves.

At frequencies higher than 100 mc, the interelectrode capacitance of an ordinary electron tube provides a low-impedance path which shunts the external circuit. Also at these frequencies the electron transit time between cathode and plate becomes appreciable. The transit time is about one-thousandth of a microsecond. As insignificant as this interval of time may seem, it nevertheless approaches and sometimes equals the time of one cycle of the applied signal and thus causes an undesirable shift in phase.

ORDINARY TUBES

A small number of ordinary tubes can be operated at frequencies higher than 100 mc under

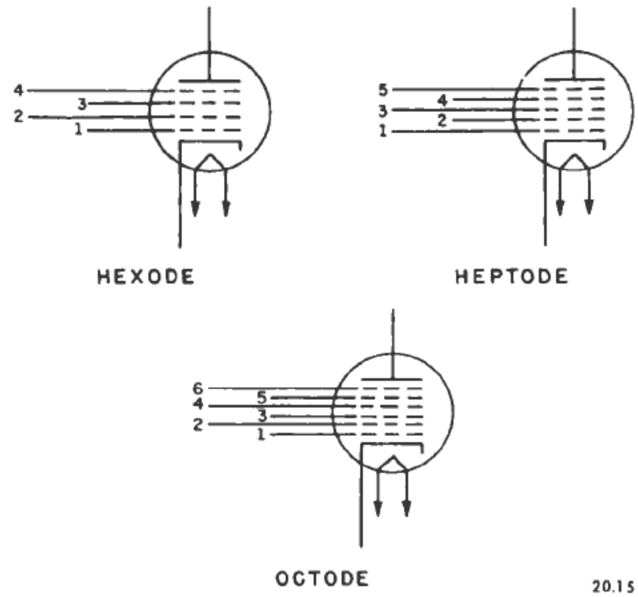


Figure 1-18.—Schematic diagrams of multigrid tubes.

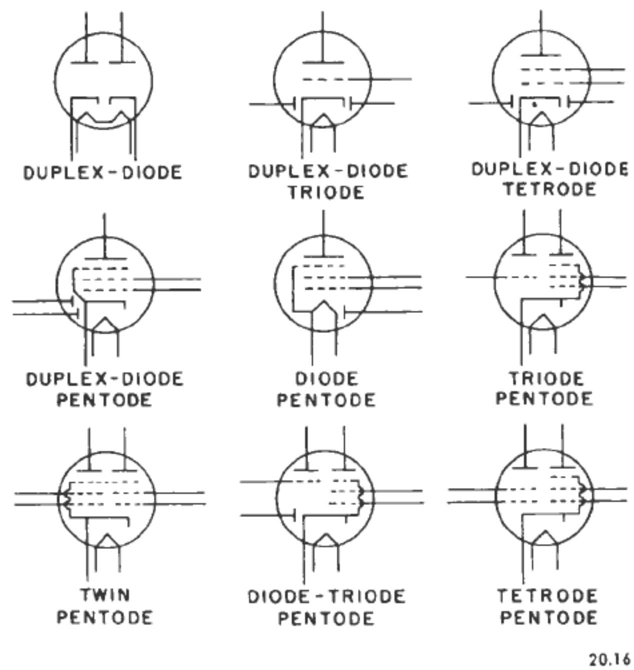


Figure 1-19.—Schematic diagrams of multiunit tubes.

certain critical operating conditions. The most suitable tubes of this type are triodes having low-interelectrode capacitance, close spacing of the electrodes to reduce transit time, a high-amplification factor, and a fairly low plate resistance. Because some of these requirements are conflicting, a compromise has to be made and tubes that strike a medium between the conflicting values are generally selected.

SPECIAL ULTRAHIGH-FREQUENCY TUBES

The objectionable features of ordinary electron tubes are minimized considerably in the construction of special u-h-f tubes. These tubes have very small electrodes placed close together, and often have no socket base. By a proportionate reduction in all physical dimensions of a tube, the interelectrode capacitances are decreased without affecting the amplification factor or the transconductance. The electron transit time is likewise reduced.

Acorn electron tubes (fig. 1-20) have been developed especially for u-h-f operation and are available as diodes, triodes, and r-f pentodes. These tubes are very small physically and have closely spaced electrodes and no base. The tube connections are brought out to short wire pins sealed in the glass envelope. Such tubes are not used extensively because of limited power capabilities.

An enlarged version of the acorn tube, known as the doorknob tube, can be operated at a considerably higher power level and at frequencies as high as 600 mc.

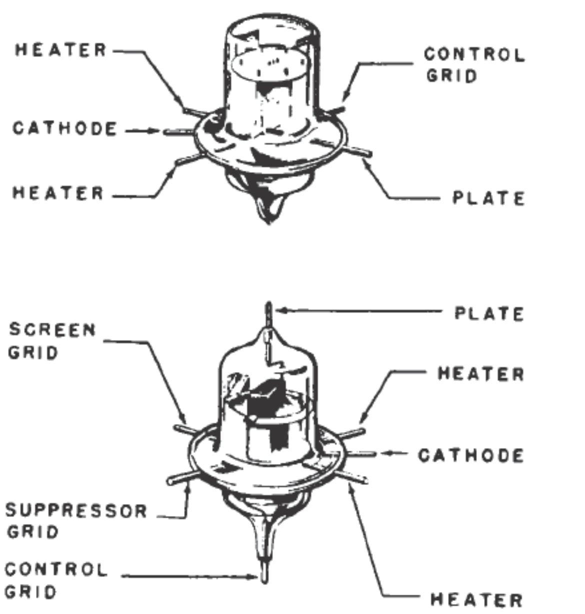


Figure 1-20.—Acorn electron tubes.

GAS-FILLED TUBES

In the manufacture of high-vacuum tubes, as much of the air as possible is removed from the envelope. In some cases low-vacuum tubes are designed purposely to contain a specific gas in place of air—usually nitrogen, neon, argon, or mercury vapor.

In a high-vacuum triode, the grid retains complete control of the current flowing in the tube. However, in a gas-filled tube the grid loses control when the gas ionizes, because of a sheath of positive ions that surround the grid. The plate current then rises rapidly to its full value. In this respect the gas tube acts like a snap-action switch. When the plate voltage falls below the deionization potential, the gas is deionized.

The gas-filled tube normally has a higher plate current rating than a high-vacuum tube of the same physical dimensions. When ionization occurs, the tube presents a lower impedance to the external circuit. Several types of gas-filled tubes are represented in figure 1-21. The small dot within the circle indicates that the tube is gas-filled.

ELECTRICAL CONDUCTION IN GAS TUBES

In a gas-filled tube, such as the diode of figure 1-21,A, the electron stream from the hot cathode encounters gas molecules on its way to the plate. When an electron collides with a gas molecule the energy transmitted by the collision may cause the molecule to release an electron. This second electron may then join the original stream of electrons and thus be capable of liberating other electrons through collision with other gas molecules. This process which is cumulative is a form of ionization. The molecule that has lost an electron is called an ion and bears a positive charge. The tube in its ionized condition contains molecules, ions, and free electrons within the envelope. The positive gas ions are relatively large and in the vicinity of the cathode they neutralize a portion of the space charge. Thus electrons flow from cathode to plate with less opposition than in a high-vacuum tube.

The heavier positive ions are attracted toward the negative cathode and while moving toward it they attract additional electrons from the space charge.

The energy needed to dislodge electrons from their atomic orbits and to produce the ionization is supplied by the source which supplies the voltage between the plate and cathode. There is a certain voltage value for a particular gas-filled tube at which ionization begins. When ionization occurs large currents flow at relatively low voltage across the tube. The voltage at which ionization commences is known as ionization potential, striking potential, or firing point.

After ionization has started, the action maintains itself at a voltage considerably lower than the firing point. However, a minimum voltage is needed to maintain ionization. If the voltage across the tube falls below this minimum value, the gas deionizes and conduction stops. The voltage at which current ceases to flow is known as the deionizing potential or the extinction potential. The tube may therefore be used as an electronic switch that closes at a certain voltage and permits current to flow and then opens at some lower voltage and thus blocks the flow of current. Such a tube has almost infinite resistance before ionization and very low resistance after ionization.

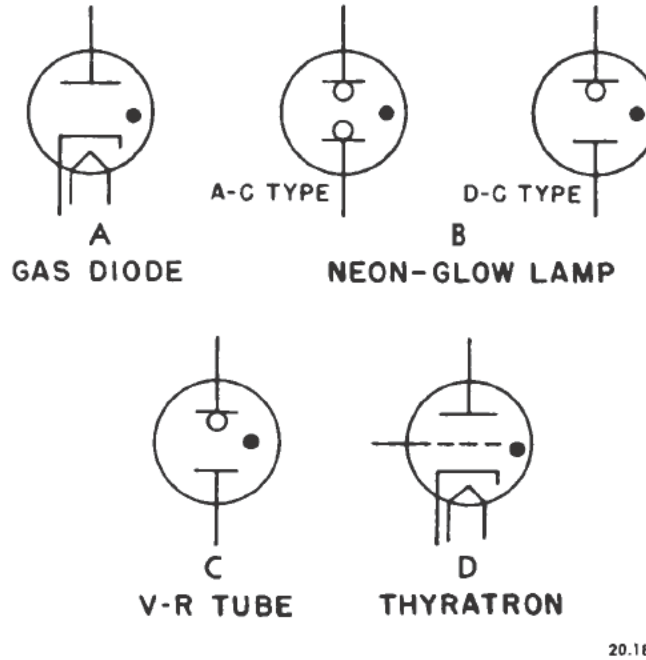


Figure 1-21.—Schematic diagrams of representative gas-filled tubes.

LIMITATIONS IN THE USE OF GAS TUBES

One limitation in the use of gas in electron tubes is the possibility that the tube will permit current to flow in the reverse direction (arcback) when the plate has a high negative (inverse) voltage with respect to the cathode. The peak inverse voltage rating varies inversely with the temperature and pressure of the gas.

A second limitation is the possibility that the cathode may be destroyed by positive-ion bombardment as the plate voltage is increased to a high value. Because the mass of the ion is very much greater than that of the electron, the result of its impact on the cathode may be serious, especially if the cathode is of the oxide-coated type. If the plate voltage is raised to a sufficiently high value, double ionization (two electrons dislodged from the gas molecule) occurs and the resultant increased velocity of the ions caused by the increased positive charge may quickly damage the emitter surface. The solution is to keep the plate voltage below the double ionization potential or to use a more rugged emitter which unfortunately will also have a higher work function.

Another limitation at high-operating frequencies is the possibility that arcback will occur because too many ions remain between the plate and cathode on the negative half cycle. This condition results from the fact that at high frequencies there is not sufficient time for the ions to be neutralized by the electrons before the full reverse voltage is applied. Because arcback causes the tube to offer a low resistance on both halves of the cycle, the power dissipated is increased and the tube will probably be destroyed. At high operating frequencies arcback may occur at a fairly low voltage and hence the tube is said to have a low inverse voltage rating.

GAS DIODES

The neon-glow lamp or neon bulb (fig. 1-21,B) is a cold-cathode gas-filled diode. The cathode may have the same shape and size as the plate so that the tube can conduct in either direction depending only on the applied potential, or the structures of the cathode and plate may be such as to permit conduction in only one direction (fig. 1-21,B). Because the cathode is not heated in this type of tube no electrons are emitted to help in the ionization process.

Therefore the firing potential for a neon-glow tube is higher than that for a tube in which a hot cathode is used, and the neon-glow tube is somewhat erratic in that the firing potential varies during the operation. The passage of current through the tube is indicated by a glow. The color of the glow depends on the gases that may be mixed with the neon. The glow is on the negative electrode or cathode. When an alternating voltage is applied both electrodes are alternately surrounded with a glow discharge.

A neon-glow tube placed in an r-f field of sufficient strength to ionize the gas in the tube will indicate the presence of such a field by glowing. A glow tube may also be used as a voltage regulator (chapter 3). Additional uses of glow tubes are as a source of light, as a part of a relaxation oscillator, as a rectifier, and to control circuit continuity in noise limiters.

Hot cathode, mercury vapor diodes are specially designed to serve as rectifiers. Tubes of this type can pass much higher currents than high-vacuum tubes because the ionization of the mercury vapor partially dispels the cathode space charge. Mercury vapor is formed in these tubes when the small amount of liquid mercury enclosed in the envelope is vaporized by the hot cathode. These tubes are not capable of supplying their rated output until the mercury is completely vaporized. The relatively high voltage existing between the plate and cathode, before the tube begins to conduct load current, causes a large increase in the electron velocity. These high-velocity electrons cause the gas ions to acquire a higher positive charge and thus to bombard the cathode with a greater impact that is high enough to disintegrate the emitter surface if the action is allowed to continue for even a short period. Therefore sufficient time must be allowed for the tube to become heated before the plate voltage is applied.

THYRATRONS

A gas-filled triode (fig. 1-21,D) or tetrode in which a grid is used to control the firing potential is called a thyratron. The grid in this tube functions somewhat the same as that of an ordinary electron tube, but the resultant control action is entirely different. Figure 1-22 shows the grid control characteristics of a typical thyratron.

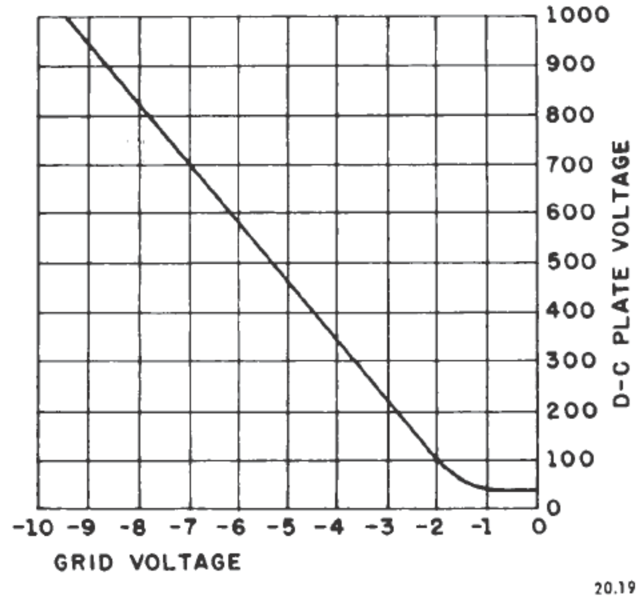


Figure 1-22.—Grid control characteristic of a representative thyratron.

Thus at a given plate voltage, for example 800 volts, the bias would have to be reduced from -10 volts to -8 volts before the tube would begin to conduct. Likewise, at a plate voltage of 300 volts the tube would begin to conduct at a grid potential of approximately -4 volts. When conduction starts, the grid loses control over the plate current and is no longer effective as a control element. To stop plate current flow, the plate voltage must be reduced below the ionizing potential. The grid operates in this manner because when conduction starts, positive ions are formed as a result of collisions and some of these ions are attracted to the negative grid. A positive-ion sheath is formed around the grid, thus destroying its effectiveness as a control element. Other positive ions move toward the cathode and neutralize the space charge. These two actions account for the fact that once current flow starts, the grid loses control and the current rises rapidly to a large value.

Thyratrons have many practical applications in relay and trigger circuits.

CATHODE-RAY TUBES

Cathode-ray tubes are electron tubes of a special construction that permit the visual observation of current and voltage waveforms. A discussion of their construction and operation is included in the latter part of this training course.

CHAPTER 2

INTRODUCTION TO TRANSISTORS

SEMICONDUCTORS

The transistor is a relatively new form of electronic device. It can perform many of the functions of an electron tube and in addition can do some things better and more efficiently. Electron tubes depend on the flow of electrons through a vacuum, a gas, or a vapor; whereas, transistors make use of the flow of electrons through a special type of solid, known as a SEMICONDUCTOR.

A semiconductor is a material having a conductivity lower than that of a conductor but higher than that of an insulator. For example, a germanium crystal in its purest state acts as an insulator, but when a small amount of impurity such as arsenic is added to it, the crystal becomes a satisfactory conductor.

The semiconductors used in transistors include germanium and silicon. A new concept is needed to understand the somewhat unexpected performance of these materials.

A brief review of the atomic structure of crystalline matter will follow in order to help the reader understand transistor conduction.

ATOMIC STRUCTURE

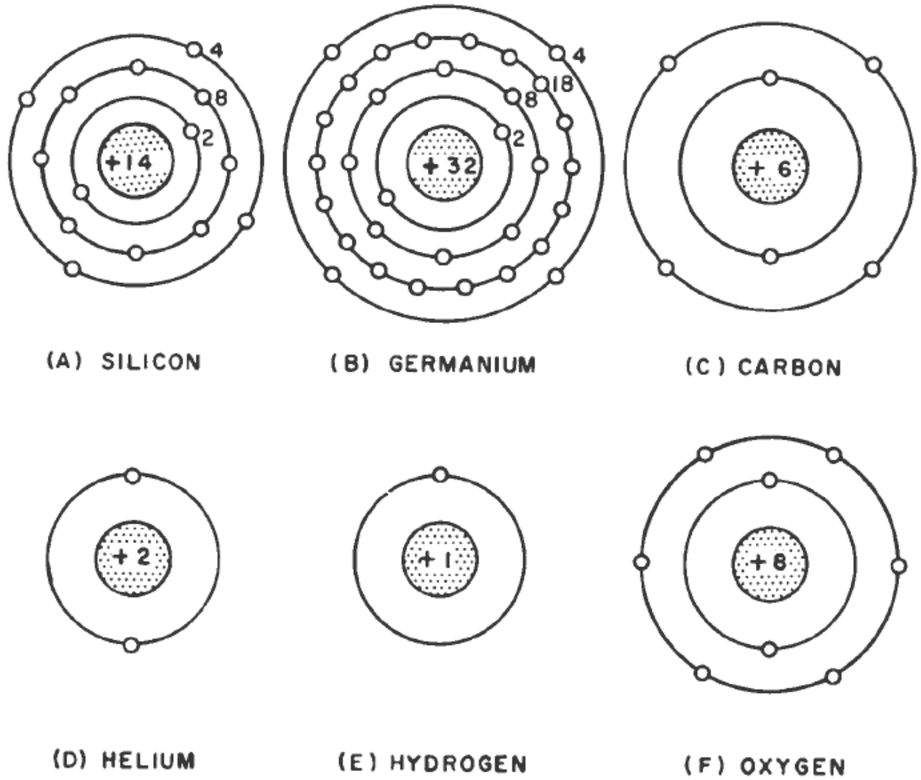
An atom (in its normal state) consists of a positive nucleus containing a fixed number of protons (indicated by the atomic number) surrounded by an equal number of electrons in various shells. Each shell can contain not more than a fixed number of electrons. When a shell contains this fixed maximum number of electrons, it is said to be complete. In many atoms the inner shells are complete; whereas, one of the outer shells is incomplete. The innermost (complete) shells, together with the core, form a stable "ionic core" with a net positive charge. The core can be considered completely inactive as far as chemical reactions and electrical phenomena are concerned. Only the outermost shell of electrons determines the chemical and electrical characteristics of the substance.

The orbits of the planetary electrons are grouped in shells with a specified number of electrons in each shell. The shells are numbered 1, 2, 3, and so forth, starting with the shell nearest the nucleus and proceeding to the outermost shell. Shell number 1 contains a maximum of 2 electrons; shell number 2, 8 electrons; shell number 3, 18 electrons; and so on. The maximum number of electrons permitted in each of the first four shells (counting from the inside out) is equal to the square of the shell number multiplied by 2. For example, the maximum number of electrons permitted in shell number 4 is $4^2 x 2 = 32$. As mentioned before the outer shell may not always contain the full complement of electrons.

For example, silicon (fig. 2-1, A) has an atomic number of 14, indicating 14 protons in the nucleus, and 14 planetary electrons distributed in 3 concentric shells around it. Germanium (fig. 2-1, B) has an atomic number of 32 indicating 32 protons in the nucleus and 32 planetary electrons. However, in both silicon and germanium the outermost shells are incomplete, each having only 4 electrons instead of the maximum permitted, 18 and 32 respectively.

For these atoms and others that have 3 or more shells, it has been found that if the outermost shell contains 8 electrons it can be considered to be complete and the atom will be stable. Therefore the outermost shell of silicon or germanium requires only 4 electrons in addition to the 4 already there, to become stable.

The 4 electrons in the outer shell of the germanium atom are the only electrons that can be influenced by external forces; the others are bound so closely to the nucleus that they cannot be removed. Thus the germanium atom may be represented by 2 concentric circles; the inner circle showing a net charge of +4 units and the outer circle containing the 4 planetary electrons that can be influenced by external forces.



13.1

Figure 2-1.—Arrangement of planetary electrons in atoms of various elements.

Similarly, the carbon atom (fig. 2-1, C) has 4 electrons in the outer shell that can be influenced by external forces; the 2 electrons in shall number 1 are bound tightly by the nucleus and cannot be removed. Thus, carbon (shown later in the diamond lattice) can be represented as 2 concentric circles; the inner circle showing a net positive charge of 4 units and the outer circle showing the 4 planetary electrons that can be influenced by external forces.

VALENCE BONDS

The valence of an element is a measure of its ability to combine with other elements. It depends on the number of planetary electrons in the outer shell of the atom. For example, the helium atom (fig. 2-1, D) has 2 planetary electrons in the outer shell, and in this case the shell contains the full complement of electrons. The valence of helium is zero. Helium does not combine with other atoms and is said to be inert.

On the other hand hydrogen (fig. 2-1, E) has a valence of 1 and will combine readily with other atoms. The reason is that the outer shell of the hydrogen atom has only 1 planetary electron and requires 1 more electron to be complete. Oxygen (fig. 2-1, F) has a valence of 2 because the outer shell has only 6 planetary electrons; it needs 2 more to make up the full complement of 8.

Two atoms of hydrogen combine with 1 atom of oxygen to form a molecule of water, figure 2-2, A. The electrons in the outer shells are shared by the oxygen (0) atom and the hydrogen atoms so as to make the outer shells of all 3 atoms complete. All atoms try to complete their outer shells, and when they do, a stable condition exists.

When hydrogen and oxygen combine to form water the force that holds them together is called an ionic valence bond or an electrovalence bond. An atom of oxygen has a combining power of valence of 2 because there are 2 vacancies (2 more electrons needed) in the outer shell. When water is formed, 2 hydrogen atoms contribute their planetary electrons to the outer shell of the oxygen atom so as to

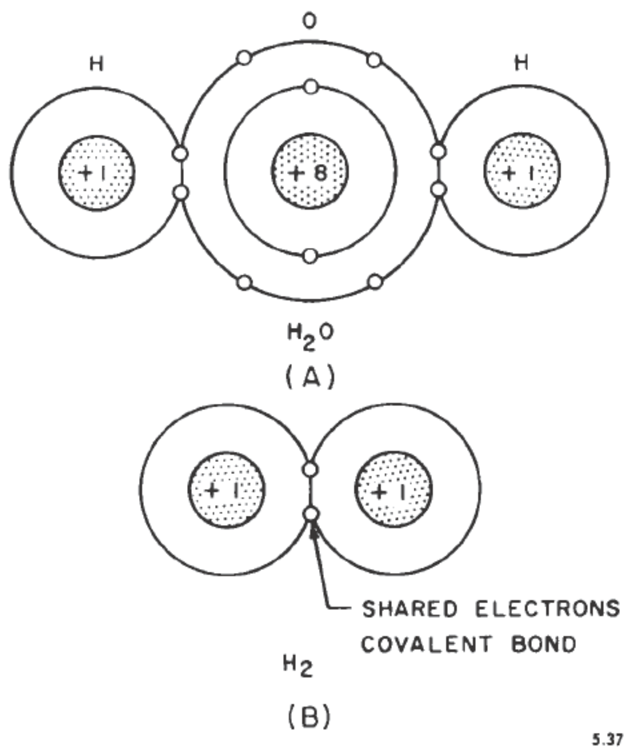


Figure 2-2.—Atoms combining to form molecules.

complete the outer shell (a total of 8 electrons) and at the same time the outer shell of each hydrogen atom is completed (2 electrons).

Hydrogen always tries to complete its outer shell (fig. 2-2, B) by appearing in molecular form (2 atoms joining together to share each others planetary electrons). This action produces a stable molecule. The 2 hydrogen atoms are held together by a covalent bond.

CONDUCTORS AND INSULATORS

Substances that produce the free motion of a large number of electrons are called conductors; those that do not are called insulators (see Basic Electricity, NavPers 10086). In insulators the electrons are bound more closely to the nucleus. It is the degree of difficulty in dislodging the planetary electrons from the outermost shell of the atom that determines whether the element is a conductor, an insulator, or a semiconductor.

Copper is a conductor. The copper atom has 29 protons in the nucleus and 29 planetary electrons revolving in orbits within 4 shells around the nucleus. The first shell contains 2 electrons; the second, 8; the third, 18; and the fourth or outermost shell, 1 electron. The maximum number permitted in the fourth shell is 2×4^2 or 32. Thus, the single electron in the outermost shell of the copper atom is not very closely bound to the nucleus; it can be moved easily. In a copper conductor containing billions of atoms, it is easy to produce an electron flow of billions of electrons without encountering very much resistance. An atom of an insulator contains a nucleus and 2 or more shells, with each shell completely filled with its quota of electrons. Thus, if the nucleus contains a net positive charge of 10 units, the first shell will contain 2 electrons and the second, 8 electrons. It is very difficult to get one of these electrons out of an atom and this material is an insulator or nonconductor.

The important difference between conductors and insulators is that in a conductor there are 1 or 2 electrons in the outer shell that are not tightly bound to the nucleus, whereas, in the insulator the outer shell is filled or almost filled and the electrons are tightly bound to the nucleus.

As mentioned before, a material that is classified as a semiconductor has characteristics between those of a conductor and those of an insulator. The electrons in the outer shell of the atoms of a semiconductor can be removed when some form of energy is applied to the material. The energy may be in the form of heat, light, or an electric field. Then the material will act like a conductor.

The carbon atom (fig. 2-1, C) may be redrawn so as to show only those electrons that may be influenced by external forces; that is, those electrons in the outermost shell. The inner shell and the nucleus may be drawn as an inner circle showing a net charge of 6-2 or +4 units. The outer shell is represented as an outer concentric circle having 4 electrons spaced equally around it.

In order to fill its outer shell with the maximum number of electrons permitted (8), the carbon atoms in a diamond lattice (a form of crystalline carbon) arrange themselves as shown in figure 2-3. This arrangement is called a crystal. In the diamond crystal the 4 electrons in the outer shell of 1 atom are bound as closely to that atom as the 4 electrons in the outer shell of another atom are bound to it. Thus 1 atom cannot pull electrons away from another atom; instead, adjacent atoms will share the

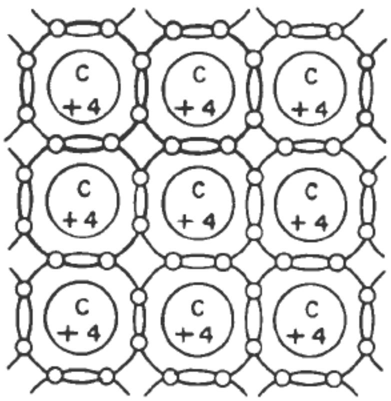


Figure 2-3.—Arrangement of carbon atoms in diamond lattice.

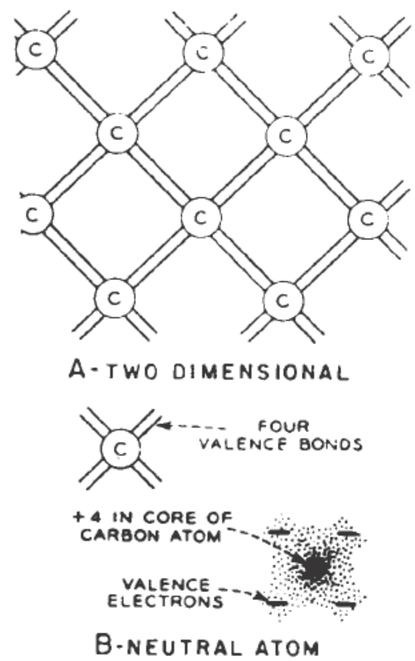


Figure 2-4.—Simplified diamond lattice.

outer electrons in such a manner that the outer shells are filled to their quota of 8 electrons. Each pair of atoms shares 2 electrons, 1 from each atom. This pair of electrons (as previously stated for the hydrogen molecule) is called a covalent bond. Thus, to try to fill its outer shell with 8 electrons each carbon atom will establish covalent bonds with 4 other carbon atoms. Each carbon atom shares each of its 4 valence electrons in the outer shell with 1 other atom, and in return shares an electron in the outer shell of the second atom.

The elements carbon, silicon, and germanium, have the common property of being tetravalent; that is, 4 electrons in the outer shell are able to respond to external forces (enter into chemical reactions). The 4 valence electrons form bonds with 4 electrons from adjacent atoms to form a crystal structure. The atoms arrange themselves in a definite pattern (found in the crystalline forms of carbon, silicon, and germanium). The position of an atom in the crystal is referred to as a lattice site (location).

A 3-dimensional diamond lattice is drawn in 2-dimensional simplified form in figure 2-4, A. The 4 valence bonds are shown as 4 pairs of lines extending between circles. The circles represent the lattice sites; each pair of parallel lines represents a covalent bond. Every atom has 4 covalent bonds.

Figure 2-4, B, illustrates the fact that each atom with its own share of 4 valence electrons is electrically neutral. At the same time, every electron is tightly bound and unable to leave the lattice structure. This explains why the diamond is a good insulator unless something is done to disturb the valence-bond structure.

As previously described, some of the outermost electrons form VALENCE BONDS between adjacent atoms. These bonds, which hold the atom rigidly in place in solid crystalline substances, like the diamond lattice, are stable for certain given numbers of electrons in the outermost shell. For example, the bond is especially stable when it contains precisely 2 electrons. The bond is weakened when 1 electron is removed; it is not strengthened much when a third electron is added.

Conductivity can be produced in a diamond in several ways, all of which involve destroying the perfection of the valence-bond structure. One way is to bombard the diamond with high-energy radiation (an electric field) which breaks the bonds and causes electrons to be ejected from them. The bonds in silicon and germanium are weaker than diamond bonds and can be broken by heating. The increase in thermal agitation causes the bonds to break.

TWO KINDS OF CURRENT

If a sample of germanium or silicon with broken bonds is subjected to a voltage, two kinds of current will flow. The ejected electrons move through the crystal from the negative terminal to the positive terminal, thus constituting ordinary conduction (electron) current. The second current (hole current), however, results from the motion of electrons from one valence bond breaking away to fill up the hole caused by the absence of an electron from an adjacent bond. Thus, holes seem to move in a direction opposite to that of electron flow. Because the hole is a region of net positive charge, the apparent motion is like the flow of particles having a positive charge. An anology of hole motion is the movement of balls through a tube (fig. 2-5). When ball number 1 is removed from the tube, a space is left. This space is then filled by ball number 2. Ball number 3 then moves into the space left by ball number 2. This action continues until all the balls have moved one space to the left at which time there is a space left by ball number 8 at the right-hand end of the tube.

The motion of the space is similar to hole motion in the covalent bond structure of some semiconductors. The hole motion depends on the movement or shifting of valence electrons in the covalent bonds. The same electrical effect is produced whether electrons move in one direction or holes move in the opposite direction.

These two currents are called ELECTRON CURRENT and HOLE CURRENT. If a suitable

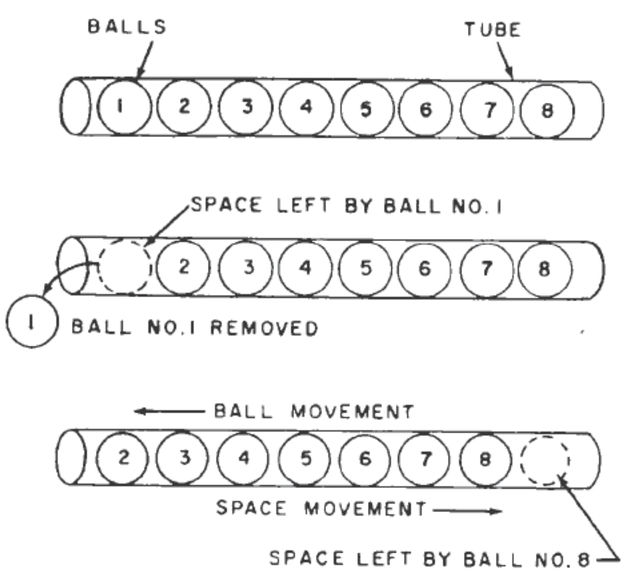


Figure 2-5.—Analogy of hole movement. 20.21

"impurity" is added to the semiconductor, the resulting mixture can be made to have either (1) an excess of electrons, thus causing more electron current; or (2) an excess of holes, thus causing more hole current.

IMPURITY DONORS AND ACCEPTORS

In the pure form, germanium and silicon crystals are of no use as a semiconductor device. However, as stated at the beginning of this chapter, when a certain amount of impurity is added, the crystal can be made to conduct a current. In order to accomplish this result the quality and quantity of the impurity must be carefully controlled. The added impurities will create either an excess or a deficiency of electrons depending on the kind of impurity added.

The impurities that are important in semiconductor materials are those impurities that align themselves in the regular lattice structure whether they have 1 valence electron too many, or 1 valence electron too few. The first type loses its extra electron easily and in so doing increases the conductivity of the material by contributing a free electron. This type of impurity has 5 valence electrons and is called a pentavalent impurity. Arsenic, antimony, bismuth, and phosphorous are pentavalent impurities. Because these materials give up or donate 1 electron to the material they are called donor impurities.

The second type of impurity tends to compensate for its deficiency of 1 valence electron by acquiring an electron from its neighbor. Impurities of this type in the lattice structure have only 3 electrons and are called trivalent impurities. Aluminum, indium, gallium, and boron are trivalent impurities. Because these materials accept 1 electron from the material they are called acceptor impurities.

Semiconductors that have no impurities are called intrinsic semiconductors. Semiconductors that have either acceptor or donor impurities are called extrinsic semiconductors.

N-Type Germanium

When a pentavalent impurity like arsenic is added to germanium it will form covalent bonds with the germanium atoms. Figure 2-6,A, illustrates an arsenic atom (As) in a germanium

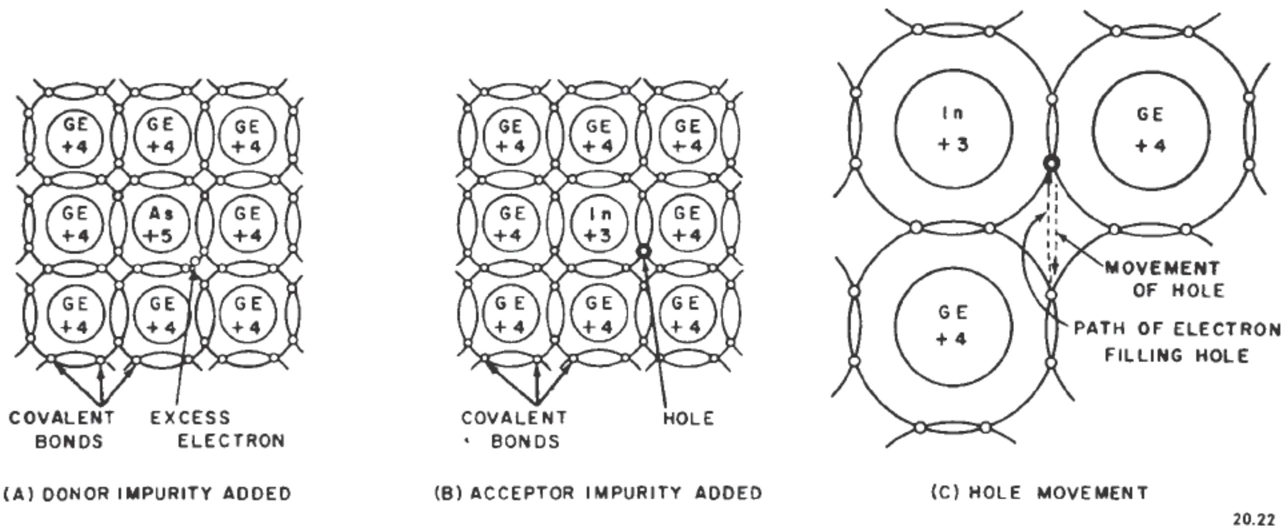


Figure 2-6.—Germanium lattice with impurities added.

lattice structure. The arsenic atom has 5 valence electrons in its outer shell but uses only 4 of them to form covalent bonds with the germanium atoms, leaving 1 electron relatively free in the crystal structure. Because this type of material conducts by electron movement it is called a negative-carrier type or N-type semiconductor. Pure germanium may be converted into an N-type semiconductor by "doping" it with any element containing 5 electrons in its outer shell. Other pentavalent elements which may be used in place of arsenic as "dopants" are phosphorous, antimony, and bismuth. The amount of the impurity added is very small; it is of the order of 1 atom of impurity in 10 million atoms of germanium.

P-Type Germanium

A trivalent impurity element can also be added to pure germanium to "dope" the material. In this case the impurity has 1 less electron than it needs to establish covalent bonds with 4 neighboring atoms. Thus in 1 covalent bond there will be only 1 electron instead of 2. This arrangement leaves a hole in that covalent bond.

Figure 2-6,B, shows the germanium lattice structure with the addition of an indium atom (symbol In). The indium atom has 1 electron less than it needs to form covalent bonds with the 4 neighboring atoms and thus creates a hole in the structure. Gallium and boron also exhibit these characteristics. The holes are present only if a trivalent impurity is used. Note that a hole carrier is not created by the removal of an electron from a neutral atom, but is created when a trivalent impurity enters into covalent bonds with a tetravalent (4 valence electrons) crystal structure. Because this semiconductor material conducts by the movement of holes which are positive charges, it is called a positive carrier-type or P-type semiconductor. When an electron fills a hole (fig. 2-6,C) the hole appears to move to the spot previously occupied by the electron.

Charges in N- and P-Type Materials

When a donor material such as arsenic is added to germanium, the fifth electron in the outer ring of the arsenic atom does not become a part of a covalent bond. This extra electron (when acted on by some force) may move away from the arsenic atom to one of the nearby germanium atoms in the N-type material.

The arsenic atom has a positive charge of 5 units on the inner circle, as shown in figure 2-6,A, and when the electron moves away from the arsenic atom there will be only 4 electrons to neutralize the positive charge and as a result there will be a region of positive charge around the arsenic atom. Similarly, the excess electron that has moved into the germanium atom outer shell makes a total of 5 electrons instead of 4 electrons for that atom of germanium. Thus, there is a region of negative charge around this atom.

Although there is a region of positive charge around the arsenic atom, after the electron has moved away, and a region of negative charge around the germanium atom with the extra electron, the total charge on the N-type crystal remains the same. In other words the total charge is zero. There are exactly enough electrons to neutralize the positive charges on the nuclei of all the atoms in the crystal. However, because some of the electrons may move about in the crystal, there will be regions in the crystal where there are negative charges and other regions where there will be positive charges, even though the net charge on the crystal is zero.

In a P-type material having an impurity such as indium added to it, a similar situation may exist. Indium has only 3 electrons in its outer ring. Three electrons are all that are needed to neutralize the net positive charge of 3 units on the inner circle (fig. 2-6, B). However, with only 3 electrons in the outer shell, there is a hole in one of the covalent bonds formed between the indium atom and the 4 adjacent germanium atoms. If an electron moves in to fill this hole (fig. 2-6, C) there is one more electron in the indium atom than is needed to neutralize the positive charge of 3 units. Thus there will be a region of negative charge around the indium atom.

Similarily, if one of the germanium atoms gives up an electron to fill the hole in the covalent bond, the germanium atom will be short an electron and there will be a region of positive charge around this atom. While the giving up of an electron by a germanium atom and the acquisition of an electron by the indium atom charges (ionizes) both atoms involved, the net charge on the P-type crystal is still zero. There is simply one atom that is short an electron and another atom that has one too many. The crystal itself does not acquire any charge.

These ionized atoms produced in both N- and P-type germanium are not concentrated in any one part of the crystal, but instead are spread uniformly throughout the crystal. If any region within the crystal were to have a very large number of positively charged atoms, these atoms would attract free electrons from other parts of the crystal to neutralize part of the charged atoms so that the charge would spread uniformly throughout the crystal. Similarly, if a large number of atoms within a small region had an excess of electrons, these electrons would repel each other and spread throughout the crystal.

Both holes and electrons are involved in conduction. The holes are called positive carriers and the electrons negative carriers. The one present in the greatest quantity is called the majority carrier; the other is called the minority carrier. In N-type material the electrons are the majority carriers and holes the minority carriers. In P-type material the holes are the majority carriers and electrons the minority carriers.

As mentioned before, pure semiconductor material such as germanium is a poor conductor. In fact it is an insulator if it is protected from all outside sources of energy. However, even at room temperature there is enough heat present in the germanium to produce some electron movement. An electron moving out of a covalent bond leaves a hole in the bond. The hole will attract an electron from a nearby atom, producing a hole in that atom. Thus both the holes and the electrons appear to move. The holes are positive carriers, and the electrons negative carriers. This formation of hole-electron pairs is undesirable in transistors (the reason is given later) and steps are taken to keep it as low as possible.

Conduction in germanium due to the formation of hole-electron pairs is called intrinsic conduction. It occurs even though no voltage is applied across the crystal. It is a random movement (diffusion). Holes and electrons may move in any direction. Intrinsic conduction is kept at a minimum by holding the operating temperature down and shielding the semiconductor from light, and other forms of electromagnetic radiation.

CURRENT FLOW IN N-TYPE MATERIAL

Current flow through an N-type material is illustrated in figure 2-7. Conduction in this type of semiconductor is similar to conduction in a copper conductor. That is, the application of voltage across the material will cause the loosely bound electron to be released from the impurity atom and move toward the positive potential point.

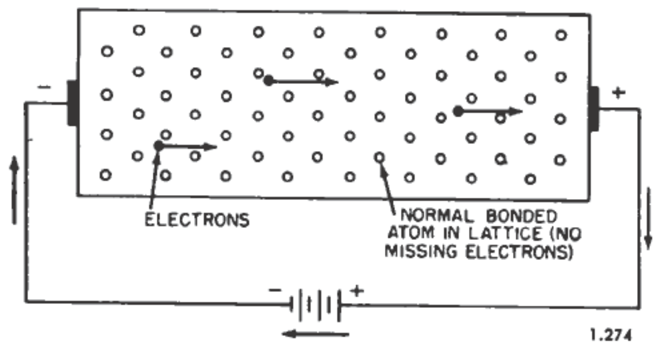


Figure 2-7.—Current flow in N-type material.

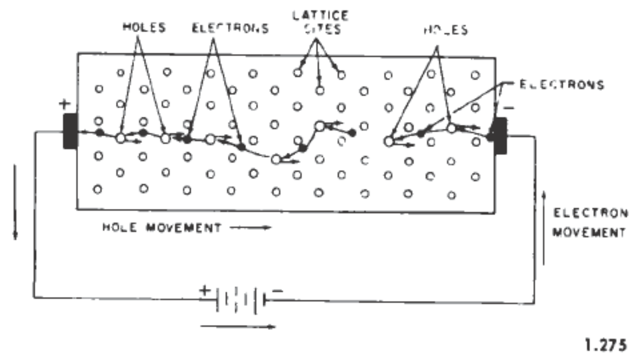


Figure 2-8.—Current flow in P-type material.

However, certain differences exist between the N-type semiconductor and a copper conductor. For example, the semiconductor resistance decreases with temperature increase, because more carriers are made available at higher temperatures. Increasing the temperature releases electrons from more of the impurity atoms in the lattice causing increased conductivity (decreased resistance). In the copper conductor, increasing the temperature does not increase the number of carriers but increases the thermal agitation or vibration of the structure so as to impede the current flow further (increase the resistance).

CURRENT FLOW IN P-TYPE MATERIAL

Current flow through a P-type material is illustrated in figure 2-8. Conduction in this material is by positive carriers (holes). In order that the hole appears to move, an electron in a nearby lattice site must shift to the position where the hole existed originally. Thus the hole moves from the positive terminal to the negative terminal. Electrons from the negative terminal cancel holes in the vicinity of the terminal while at the positive terminal, electrons are being removed from the covalent bonds, thus creating new holes. The new holes then move toward the negative terminal (the electrons shifting to the positive terminal) and are cancelled by more electrons emitted from the negative terminal. This process continues as a steady stream of holes (hole current) moving toward the negative terminal.

In both N-type and P-type materials, current flow in the external circuit is out of the negative terminal of the battery and into the positive terminal.

SEMICONDUCTOR DIODES

PN JUNCTION

A semiconductor diode is made by taking a single crystal (for example, germanium) and adding a donor impurity to one region and an acceptor impurity to the other. This gives a single crystal with an N section and a P section. Where the two sections meet is a junction. Contacts are fastened to the two ends of the crystal. A simple PN junction or junction diode is illustrated in figure 2-9, A.

One portion of the crystal is P-type material; this is the portion containing the acceptor impurity. The other portion is N-type material. This region contains the donor impurity. The end contacts are large surfaces that make a good connection with the crystal. If the connections were not good there might be rectifying properties where they come in contact with the crystal.

In the PN junction illustrated in figure 2-9, B, the P material is shown at the left and the N material is shown at the right. The P material contains acceptor impurity atoms. These atoms are represented as negative charges enclosed in circles. As mentioned before, these atoms take on electrons from the pure crystal, leaving holes as the current carriers. The holes are represented in the P material as small circles interspersed between the acceptor atoms. The N material contains donor atoms that give up electrons when they become a part of the crystal lattice. The donor atoms are represented as positive charges enclosed in circles. The free electrons in the N material are represented as dots interspersed between the donor atoms.

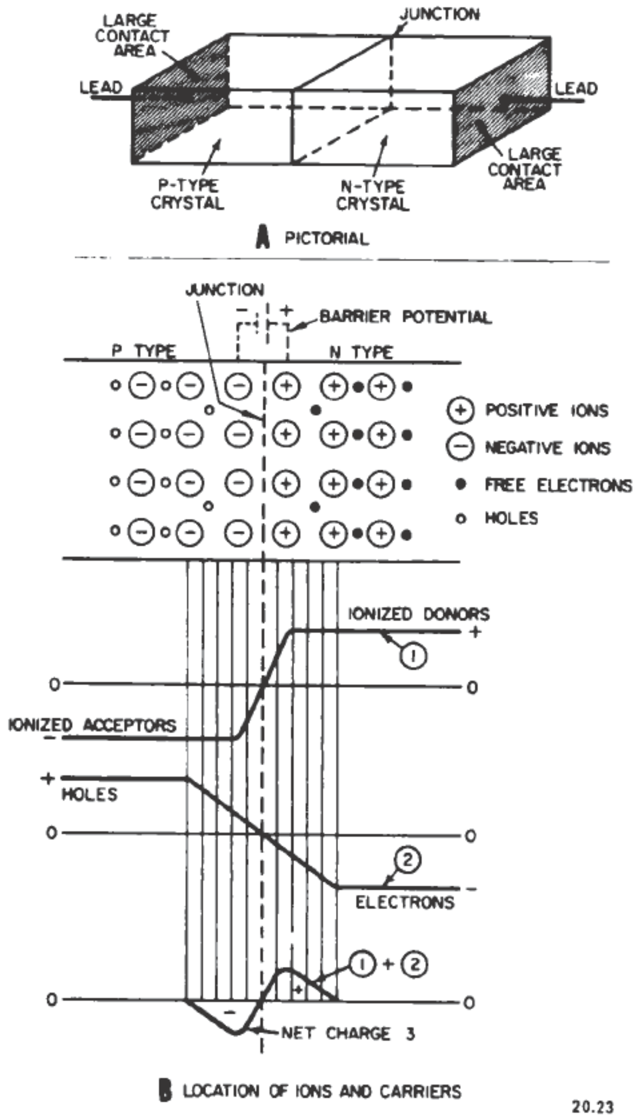


Figure 2-9.—PN junction.

Depletion Layer

When the junction is formed, free electrons in the N-type region diffuse across the junction and fill holes near the junction in the P region. Holes diffuse across the junction from the P region to the N region and capture free electrons near the junction in the N region.

When an electron leaves the donor atom in the N region and moves over to the P region the atom has fewer electrons than it needs to neutralize the positive charge on the nucleus and it becomes charged (ionized). It has one extra positive charge equal to the negative charge of the electron which it lost.

Similarly, when a hole leaves an acceptor atom in the P region, the atom takes on a negative charge, because the hole has been filled by an electron, and the atom has one more electron than it needs to neutralize the charge on its nucleus.

These charged atoms, or ions as they are called, are fixed in place in the crystal lattice structure, and cannot move. Thus, they make up a layer of fixed charges on the two sides of the junction. On the N side of the junction there is a layer of positively charged ions; on the P side of the junction there is a layer of negatively charged ions.

Notice in figure 2-9, B, that there is a barrier of negative ions on the P side of the junction. This negative barrier will repel electrons from the immediate vicinity of the junction, and will prevent the diffusion of any more electrons from the N side over into the P side of the crystal. Similarly on the N side of the junction there is a barrier of positive ions which will repel holes away from the immediate vicinity of the P side of the junction and prevent the diffusion of any additional holes across the junction from the P material into the N-type material.

The two layers of ionized atoms form a barrier to any further diffusion across the junction. Because the charges at the junction force the majority carriers away from the junction, the barrier is known as the depletion layer. It is also known as the barrier layer, or barrier potential

The charge on the impurity atoms is distributed across the PN junction as shown in figure 2-9, B, curve 1. In the P region the ionized acceptors have a negative charge, and in the N region the ionized donor atoms have a positive charge. At the junction the charge is zero.

However, in the P region there are holes which have a positive charge and in the N region there are free electrons which have a negative charge. The distribution of holes and free electrons is shown by curve 2 (fig. 2-9, B).

The potentials at the junction have driven the holes away from the junction in the P region and the electrons away from the junction in the N region so that the charges in the P region and in the N region are moved further apart. Thus, the slope of curve 2 is more gradual that that of curve 1. The charge at the junction is zero, but the rise on either side is more gradual than that of curve 1. Moving further into the P region the charge becomes positive due to the holes; moving into the N side of the crystal the charge is negative due to the electrons.

The net charges on the crystal in the P region is equal to the difference between the charge on the ionized acceptor atoms and the charge on the holes. The net charge on the crystal in the N region is equal to the difference between the charge of the ionized donor atoms and the electrons. These charges cancel except in the immediate region of the junction. This arrangement is indicated by curve 3 which is the sum of curves 1 and 2 (fig. 2-9, B).

In the area near the junction there is a negative charge in the P region and a positive charge in the N region. As stated earlier, they act as a barrier to prevent further diffusion of holes from the P region into the N region and the diffusion of electrons from the N region into the P region. This potential barrier is a potential difference (or voltage) across the junction and is of the order of a few tenths of a volt. It may be represented as a dotted battery with the negative terminal connected to the P-material and the positive terminal connected to the N-type material (top of fig. 2-9, B).

This barrier potential is like the plate-cathode voltage of a diode. If the plate is made positive with respect to the heated cathode the diode can be made to conduct a current. If the plate is made negative with respect to the cathode the diode will block the flow of current. Thus the diode tube is a rectifier. The semiconductor diode is also a rectifier.

Effect of Intrinsic Conduction

A limitation imposed on the junction diode is the result of hole-electron pairs that are being formed at random within the crystal due to energy imparted to the crystal by heat, light and electromagnetic radiation. Away from the depletion layer these carrier pairs will recombine without materially affecting the carrier concentration in the crystal. In other words the holes will remain the majority carriers in the P material and electrons will remain the majority carriers in the N material.

However, as previously mentioned, both holes and electrons are involved in conduction at all times. There are minority carriers in both regions; holes in the N material and electrons in the P material. The holes produced in the N material near the junction are attracted by the negative ions on the P side of the junction and pass across the junction. These holes will tend to neutralize the negative ions on the P side of the junction. Similarly free electrons produced on the P side of the junction will pass across the junction and neutralize positive ions on the N side of the junction. This action is an example of intrinsic conduction which is undesirable.

This flow of minority carriers weakens the potential barrier around the atoms that they neutralize. When this happens, majority carriers are able to cross the junction at the location of the neutral atom. This means that holes from the P-material will cross over to the N-material and electrons from the N-material will cross over to the P-material.

This action results in both holes and electrons crossing the junction in both directions. These motions cancel each other and the net movement contributes nothing toward the net charge or current flow through the junction. Because of intrinsic conduction the junction is no longer a rectifier when an external voltage is applied across it. It is analogous to an electron-tube diode in which not only the cathode emits electrons but the plate is heated to the point where it also will emit enough electrons to break down the rectifying properties of the diode.

BIASED JUNCTIONS

If a battery is connected across the PN junction the battery potential will bias the junction. If the battery is connected so that its voltage opposes the barrier potential across the junction it will aid current flow through the junction and the junction is said to be biased in the forward direction (low resistance). If the battery is connected across the junction so that its voltage aids the barrier potential across the junction it will oppose current flow through the junction and the junction is said to be reversed-biased or biased in the reverse direction. This is the direction of high resistance.

Forward Bias

The forward bias connection is illustrated in figure 2-10, A. Here the positive terminal of the bias battery is connected to the negative side of the barrier potential (P-type side of the junction) and the negative terminal of the battery is connected to the positive side of the barrier potential (N-type side of the junction).

The positive terminal of the battery connected to the end of the P-type germanium, repels holes toward the junction and attracts electrons from the negative ions near it. The combination of holes moving towards the junction to neutralize charged negative ions on the P side of the junction, and electrons taken from the negatively charged acceptor atoms, tends to neutralize the negative charge of the barrier potential on the P side of the junction.

On the N side of the crystal the negative terminal of the battery repels electrons towards the junction. These electrons tend to neutralize the positive charge on the donor atoms at the N side of the junction. At the same time the negative terminal of the battery attached to the N side of the crystal attracts holes away from the charged positive ions (donor atoms) on the N side of the junction. Both of these actions tend to neutralize the positive charge on the donor atoms at the junction, thereby reducing the barrier potential.

The effect of the battery bias voltage in the forward direction is to reduce the barrier potential across the junction and to allow majority carriers to cross the junction. Thus, more electrons flow from the N-type material across the junction. At the same time more holes travel from the P-type germanium across the junction where they combine with the electrons from the N-type material.

At the same time that the hole and electron movement is going on in the crystal, electrons are moving from the negative terminal of the bias battery in the external circuit to the N-type terminal, and electrons are moving from the P-type terminal in the external circuit to the positive terminal of the battery.

It is important to remember that in the forward biased junction condition, conduction is by the majority carriers (holes in the P-type material and electrons in the N-type material). Increasing the battery voltage will increase the number of majority carriers arriving at the junction and the current flow increases. The only limit to current flow is the resistance of the material on the two sides of the junction. If the battery voltage is increased to the point where the barrier potential across the junction is completely neutralized heavy current will flow and the junction may be damaged from the resulting heat. Therefore, the voltage of the bias battery is limited to a relatively small voltage.

Reverse Bias

With reverse bias applied to the junction diode the negative terminal of the battery is

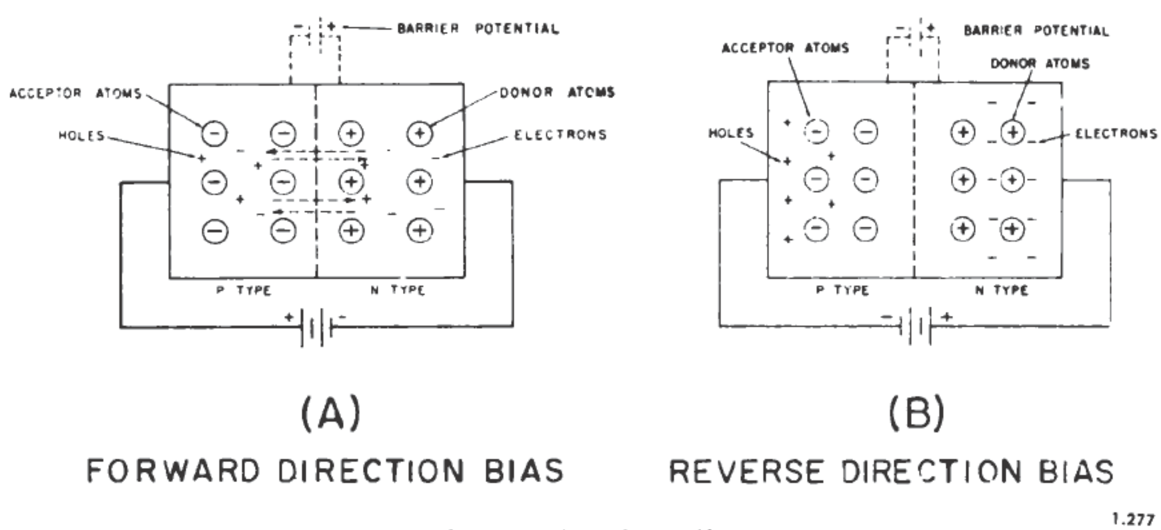


Figure 2-10.—Biased junctions.

connected to the P-type section and the positive terminal connected to the N-type section. The negative terminal attracts holes away from the junction and depletes the holes in the P material. At the same time the positive terminal of the battery attracts electrons away from the junction and increases the shortage of electrons on the N side of the junction. This action increases the barrier potential across the junction because there are fewer holes on the P side of the junction to neutralize the negative ions, and fewer electrons on the N side to neutralize the positive ions formed on this side of the junction. The increase in barrier potential helps to prevent current flow across the junction by majority carriers.

The current flow across the barrier is not zero, however, because of minority carriers crossing the junction. Holes forming in the N side of the depletion layer are attracted by the negative potential applied to the end of the P-type section; electrons breaking loose from their outer shells in the atoms of the P material are attracted by the positive voltage applied to the end of the N-type section of the germanium.

This situation was described as intrinsic conduction or hole-electron pairs continually forming at random within the crystal (before any bias is applied) due to the energy of the crystal. With no bias applied to the crystal, minority carriers neutralize ions near the junction and allow majority carriers to cross the junction. With bias applied, the minority carriers are attracted away from the junction by the potential applied across the crystal so that all of the minority carriers do not remain near the junction to neutralize charged atoms. Hence the minority carriers no longer allow the passage of an equal number of majority carriers in the opposite direction. The flow of minority carriers across the junction is not fully offset by a flow of majority carriers in the opposite direction. Therefore, there is a small current flow across the junction caused by the minority carriers crossing the junction. This current flow is small and nearly constant at normal operating voltages.

Reverse bias (also called backward bias) applied across a junction diode increases the barrier potential, making it more difficult for majority carriers to cross the junction. However, some minority carriers will still cross the junction with the result that there will be a small current. This action is indicated in the static curve for a germanium crystal diode (fig. 2-11, A). The forward portion of the curve indicates that the diode conducts easily when the potential across the junction is in the direction of forward bias (P side positive and N side negative). The diode conducts poorly in the high resistance direction (backward bias, P side negative and N side positive). For this condition the holes and electrons are drawn away from the junction, causing an increase in the barrier potential. However, if the backward bias is increased beyond a critical value (about 70 volts for some typical designs), the reverse current increases rapidly due to avalanche breakdown.

Avalanche breakdown occurs when the applied voltage is sufficiently large to cause the covalent bond structure to break down. At this point a sharp rise in reverse current occurs. The acceleration of the few holes and electrons continues to such a point that they have violent collisions with the valence bond electrons of the germanium crystal atoms releasing more and more carriers. The maximum reverse voltage of the semiconductor diode corresponds to the peak inverse voltage of an electron-tube diode. Static characteristics for several semiconductor diodes are illustrated in figure 2-11, B.

Semiconductor diodes operated in the reverse voltage breakdown region (lower left portion of static characteristics) are used as voltage regulators. Various types of voltage regulators are described in chapter 3 of this training course.

COMPARISON WITH ELECTRON-TUBE DIODE

Although junction diodes and electron-tube diodes operate differently they are used to perform the same jobs, and so may be compared. When there is no voltage applied across a junction diode from an external source, the net current flow across the junction is zero (fig. 2-11).

When no voltage is applied between the plate and cathode of an electron-tube diode, and the cathode is heated, some of the electrons leave the cathode with sufficient speed to travel across the space between the cathode and plate and reach the plate. Thus a small current flows even though no external voltage is applied across the plate-cathode circuit.

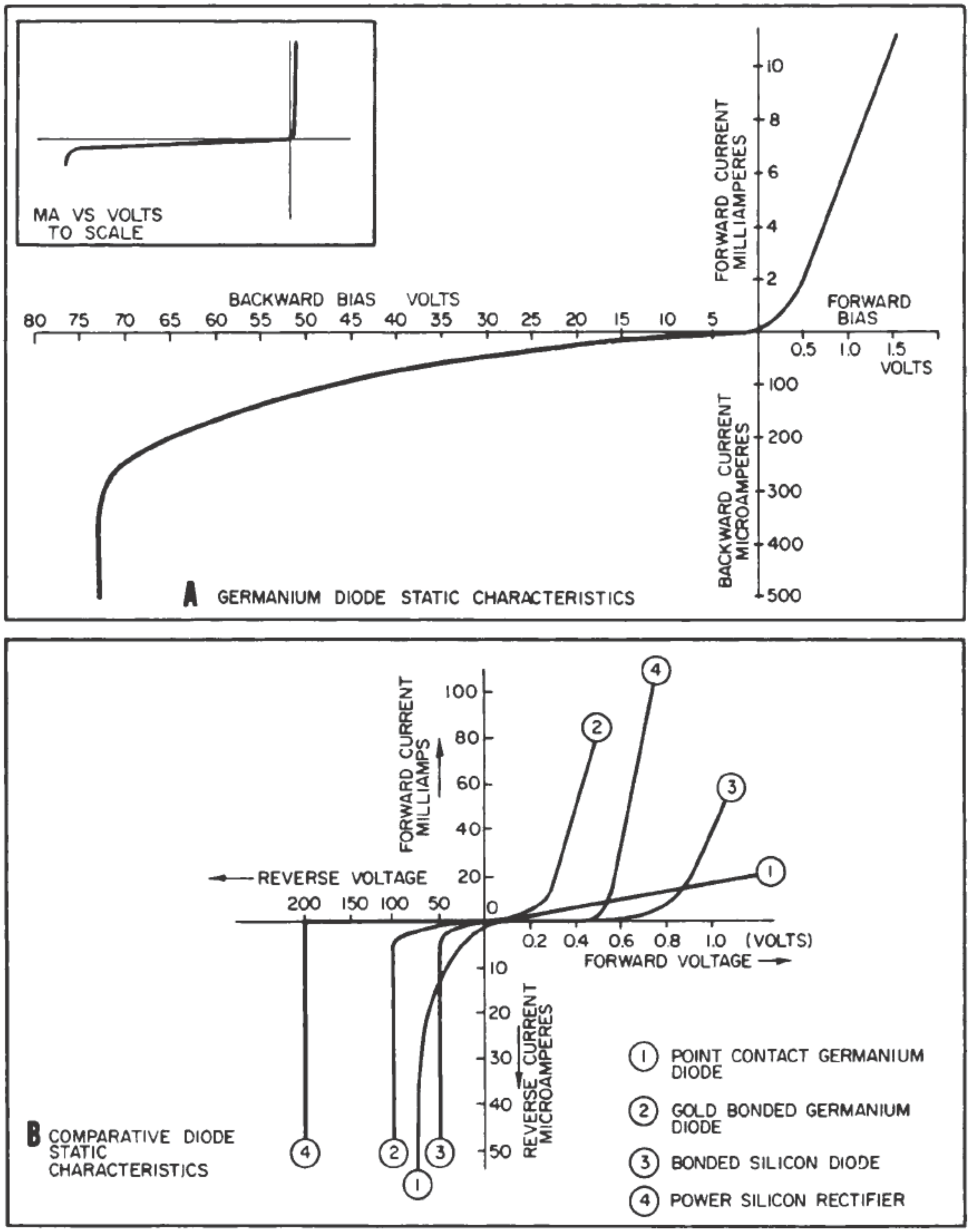


Figure 2-11.—Static characteristics of solid state diodes.

Placing a voltage across the plate-cathode circuit of an electron-tube diode, with the plate connected to the positive terminal of the battery and the cathode connected to the negative terminal, is similar to placing a forward bias connection across a junction diode. The forward bias battery positive terminal is connected to the P side of the junction, and the negative terminal of the battery is connected to the N side of the junction. For both tube and junction diodes there is a current flow. Both are similar.

When a reverse bias voltage is applied across the electron-tube diode (plate negative, cathode positive) no current will flow through the tube. Electrons leaving the cathode are repelled by the negative potential on the plate so that no plate current will flow. However, in a junction diode, there will be some reverse current with reverse bias due to conduction by minority carriers. The current flow will remain a small value unless the reverse bias voltage is increased to a large amount, as previously described in connection with avalanche breakdown.

POINT CONTACT DIODE

Another type of semiconductor diode is the point contact diode (fig. 2-12). A schematic drawing is shown in figure 2-12, A. Unlike the junction diode, the point contact diode depends on the pressure or contact between a point and a semiconductor crystal for its proper operation.

One section consists of a small rectangular crystal of N-type germanium. A fine beryllium-copper, phosphor-bronze, or tungsten wire called the catwhisker presses against the crystal and forms the other part of the diode. The reason for using a fine pointed wire instead of a flat metal plate is to produce a high intensity

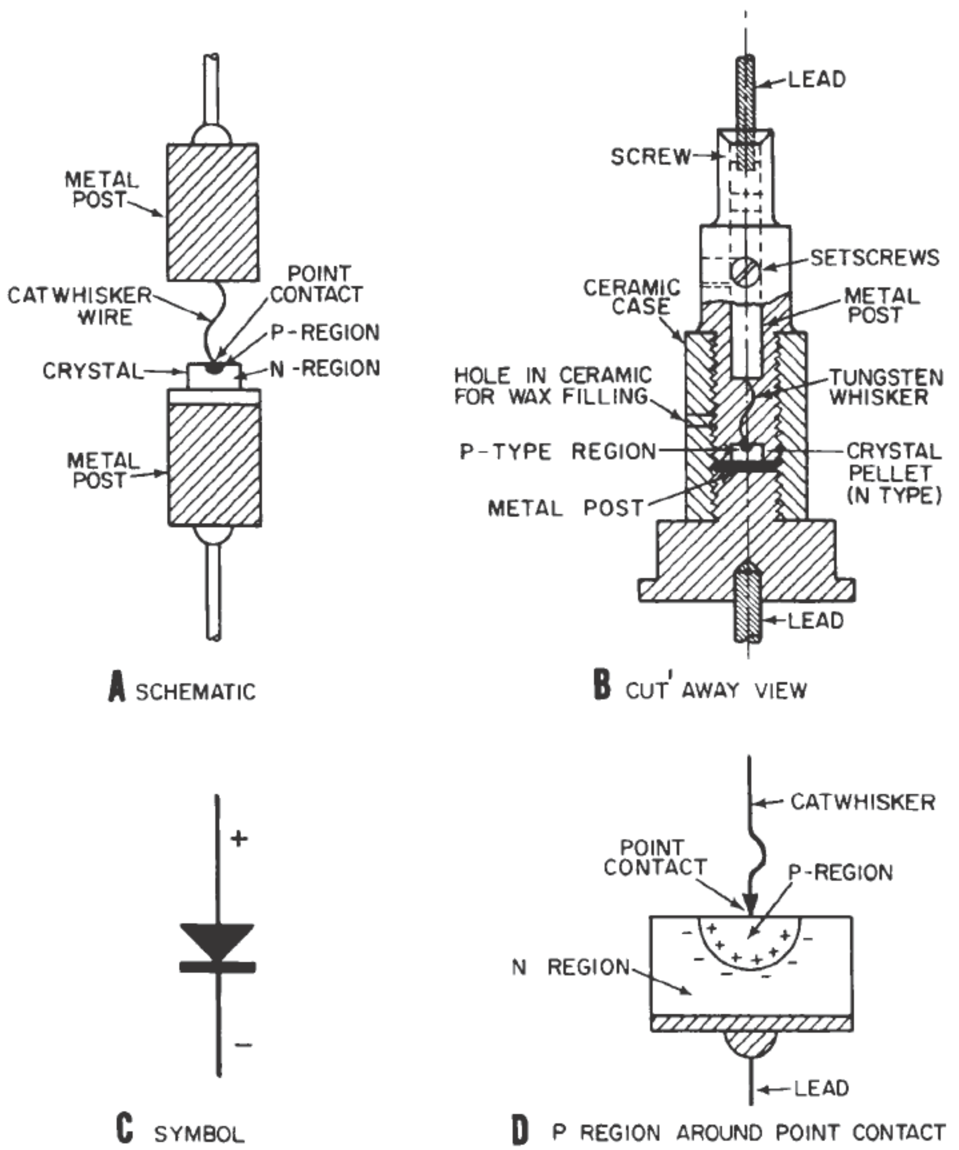


Figure 2-12.—Point contact crystal diode. 20.25

electric field at the point contact without using too large an external voltage source. As previously mentioned, it is not possible to apply large voltages across semiconductors because of excessive heating.

The opposite end of the catwhisker is one terminal of the diode. It has a low resistance contact to the external circuit. Figure 2-12, B, illustrates a cutaway view of the point contact diode and the low resistance path to the external circuit. A flat metal plate on which the crystal is mounted forms the lower contact of the diode with the external circuit. Both contacts with the external circuit are low resistance contacts. The conventional symbol for the crystal diode is shown in figure 2-12, C. The arrow points in the direction of conventional current flow; electron flow is in the opposite direction.

During the manufacturing process of the point contact diode, a relatively large current is passed from the catwhisker to the germanium crystal. The result of this large current is the formation of a small region of P material around the crystal in the vicinity of the point contact, as shown in figure 2-12, D. Thus, there is a PN junction formed which behaves in the same way as the PN junction previously described.

The characteristics of the point contact diode under forward and reverse bias are somewhat different from those of the junction diode. With forward bias the resistance of the point contact diode is higher than that of the junction diode. With reverse bias the current flow through a point contact diode is not as independent of the voltage applied to the crystal as it is in the junction diode. The point contact diode has an advantage over the junction diode in that the capacitance between the catwhisker and the crystal is less than the capacitance between the two sides of the junction diode. As such, the capacitive reactance existing across the point contact diode is higher and the capacitive current that will flow in the circuit at high frequencies is smaller and creates less disturbance.

MAXIMUM RATINGS

Semiconductor diodes are specified according to the applications for which they are to be used. Two maximum ratings are usually given: the maximum reverse voltage, and the maximum value of the average forward current. The maximum permitted average forward current depends upon the internal heating and the maximum allowable temperature. For germanium the maximum allowable temperature is 100° C: for silicon it is 225° C. The maximum forward current in the current that will cause the junction temperature to rise to the specified upper limit. The surrounding temperature is taken as 25° C which is average room temperature.

Maximum allowable reverse voltage is also related to allowable termperature rise. A certain value of reverse current is taken as a reference and the measured reverse voltage required to sustain this current is termed the maximum value. For general purpose silicon diodes, the reference current is often taken as 100 microamperes; for germanium diodes 1.5 milliamperes is frequently specified.

Crystal diodes frequently contain heat radiators often called "heat sinks" to prevent the temperature rise from exceeding the maximum permissible value. These radiators surround the crystal diode with sufficient metal surface to conduct the heat away from the junction rapidly enough to prevent excessive temperature rise even though the power loss is high or the room temperature is above normal.

DIODE TYPES AND USES

There are many different types of crystal diodes in use today. They vary in size from tiny ones hardly bigger than a pin head (used in subminiature circuitry such as computers and satellites) to large 500 ampere rectifiers used in power supply applications. Figure 2-13 illustrates some of the various types that are presently being manufactured. The elements germanium and silicon are used most often in these devices. Certain compounds such as gallium arsenide, indium antimonide, and silicon carbide are being developed for special applications such as high temperature operation. Substances most often used as impurity agents (dopants) are aluminum, arsenic, gallium, and indium.

SEMICONDUCTOR TRIODES

Although semiconductor diodes can permit more current to flow in one direction than the other (ability to rectify) they cannot amplify a signal. Three element semiconductors (like the three element electron tube) are needed in order to amplify a signal.

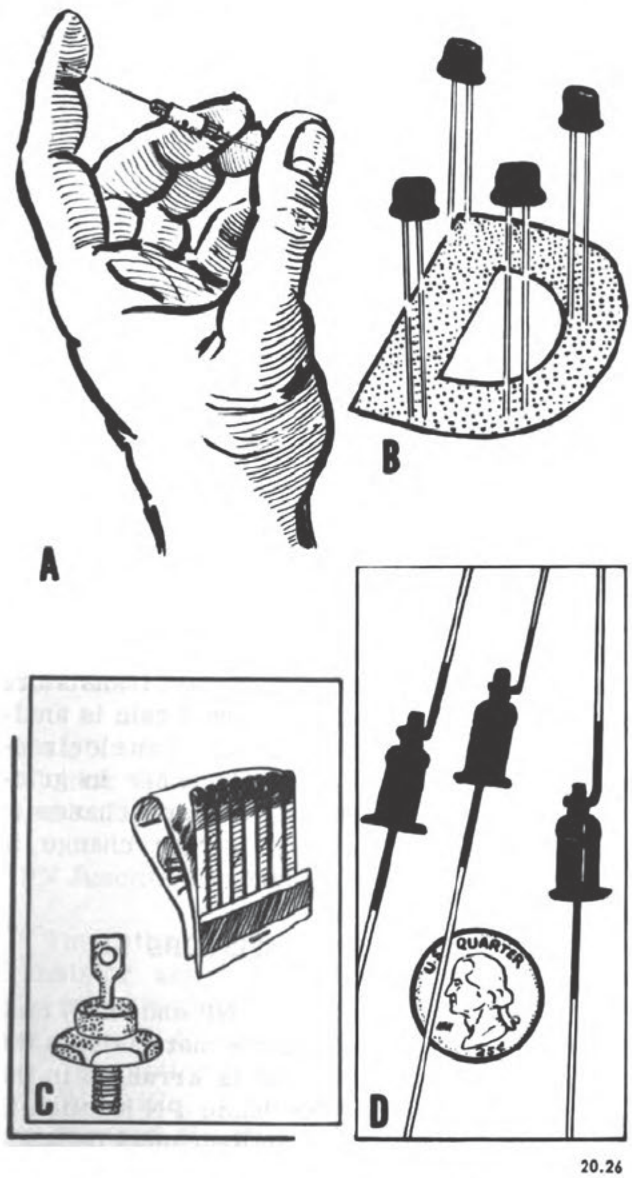


Figure 2-13.—Types of diodes.

Semiconductors that can amplify a signal are called transistors. There are many different types of transistors with individual characteristics, but the theory of operation is basic to all of them.

The point contact transistor was developed first. The junction transistor followed later. Transistors are classed as PNP or NPN according to the arrangement of the impurities in the crystal. Both point contact transistors and junction transistors contain two of the basic PN junctions previously described.

POINT CONTACT TRANSISTOR

The point contact transistor is similar to the point contact diode except that it has two catwhiskers instead of one. The two catwhiskers are placed with their point contacts very close together (about 0.002 inch). The diameter of the contacts is about 0.005 inch. The contacts are arranged to provide a spring-like pressure on the flat surface of the crystal.

The crystal may be either N-type or P-type germanium. During the manufacturing process a large current is passed from the catwhiskers into the crystal. This action forms a small region around the point contacts of P-type material (fig. 2-14, A) when the crystal is N-type germanium, or N-type material (fig. 2-14, B) when the crystal is P-type germanium. The same action occurs in the manufacturing process of the point contact diode previously described. Both point contact transistors contain two PN

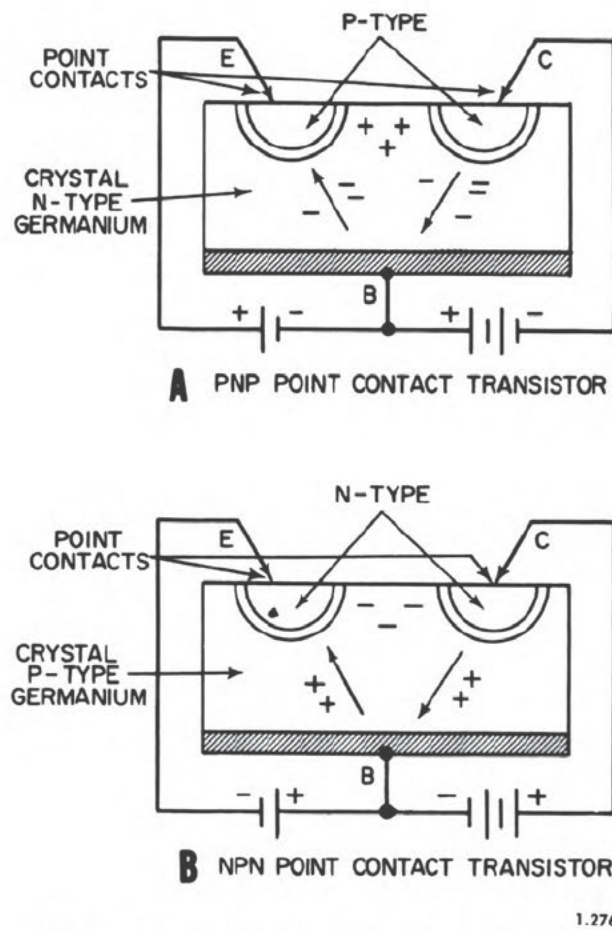


Figure 2-14.—Point contact transistor.

junctions indicated as semicircles in the vicinity of the catwhiskers. The base, B, forms a common connection between the two junctions and the external circuit. One catwhisker is called the emitter, E; the other is called the collector, C. Thus, one PN junction is between the emitter and the base. The other PN junction is between the collector and the base. The point contact transistor is called either a PNP (fig. 2-14,A) or an NPN (fig. 2-14, B) point contact transistor.

The emitter-base junction in both transistors is forward biased, and the collector-base junction reverse biased. The effect of forward and reverse bias on the barrier potential across a PN junction was described in detail in connection with semiconductor diodes.

The effect of forward bias across the PN junction between the emitter and the base in the point contact transistor is to lower the barrier potential and to increase the flow of majority carriers across the junction. The emitter is so named because its function is to emit majority carriers across the junction.

Consider the action in the PNP point contact transistor of figure 2-14, A. The positive terminal of the left-hand battery is connected to the P material in contact with the emitter. The negative terminal of the battery is connected to the N material in contact with the base of the crystal. In accordance with the previously described action of forward bias across a PN junction, holes are emitted from the P material across the PN junction as electrons are emitted from the N material. Some of these holes will enter the N material as minority carriers. They have an important effect on the reverse biased collector-base PN junction. The holes will attract electrons assisting them across the collector-base PN junction increasing the collector current proportionately.

Hole Injection

The movement of holes from the emitter region into the collector region is known as HOLE INJECTION. As holes are injected into the collector region of the N material, they exert an attractive force on the electrons, assisting them across the PN barrier around the collector, and increasing the collector current. The effect is to neutralize the barrier potential in proportion to the hole injection so that more electrons can flow across the barrier. The action is analogous to that of a space charge or cloud of positive ions in a conducting gas tube. The positive ions neutralize a portion of the negative space charge around the cathode, thereby increasing plate current. In the transistor, the holes injected from the emitter attract additional electrons across the collector PN barrier and increase the flow of collector current.

Current Gain

In the point contact transistor (either PNP or NPN) the total collector current is greater than the emitter current. This action is equivalent to current gain. Current gain is the ratio of the change in collector current resulting from a given change in emitter current for a constant collector-base voltage. The symbol for current gain is α ce and is read alpha c-e, meaning the ratio of collector to emitter current change. The current gain for point contact transistors is of the order of 2 or 3. Current gain is analogous to the amplification factor of an electron-tube triode in which a small change in grid-cathode voltage produces the same change in plate current that a much larger change in plate-cathode voltage produces.

JUNCTION TRANSISTORS

The junction transistor (PNP and NPN) employs the same semiconductor materials as the point-contact transistor but is arranged in the form of a sandwich. The basic PN junction is used. However, such a junction cannot be made satisfactorily by simply bringing two surfaces together mechanically because of the difficulty in achieving the necessary smoothness and cleanliness of the surface to be joined. Instead, the junction is formed by a diffusion or alloy process.

PNP Junction Transistors

Current flow in the external circuits of figure 2-15 consists of electron movement in a counterclockwise direction in both the emitter and collector circuits. Current flow within the P material of the transistor consists of the movement of holes through the P material from the positive terminal to the negative terminal. Current flow in the N material of the transistor

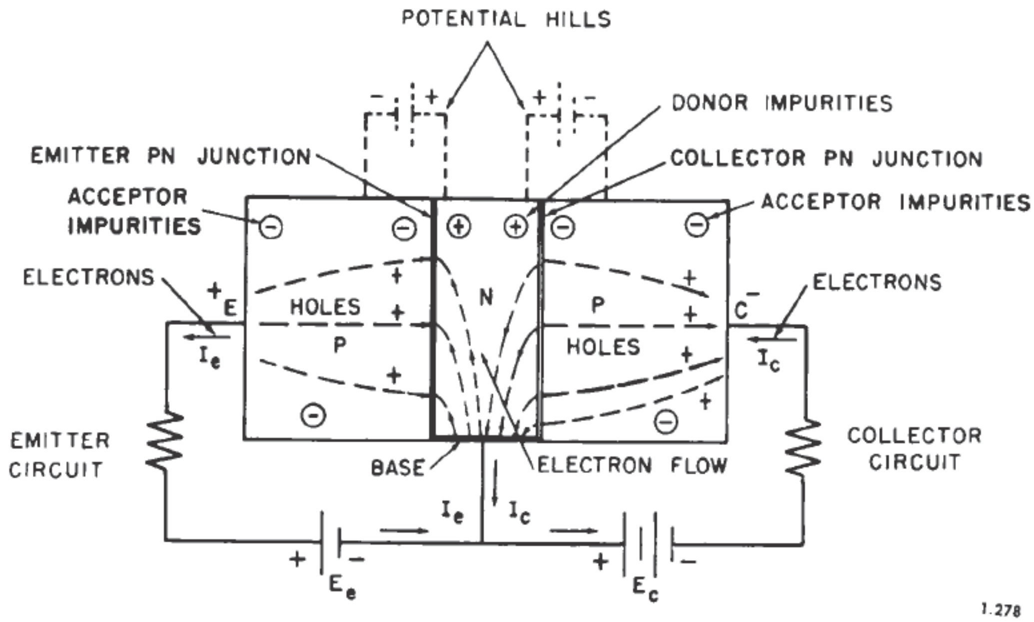


Figure 2-15.—Current flow in PNP junction transistors.

consists of the movement of electrons from the collector PN junction to the base and from the base to the emitter PN junction.

NPN Junction Transistors

The paths for current flow in an NPN junction transistor are indicated in figure 2-16. The barrier potentials across the two junction planes are opposite to those of the PNP junction transistor. The dotted battery symbols represent these potential hills. The emitter circuit is biased in the forward direction and the collector circuit is biased in the reverse, or high-resistance, direction.

Electrons flow in the external circuits of the emitter and the collector in a clockwise direction. The emitter bias neutralizes the barrier potential across the emitter PN junction, allowing electrons to be injected from the emitter through the barriers into the collector P region. The majority carriers in the P material are holes.

As electrons move across the narrow P region of the NPN junction transistor, a certain amount of recombination of electrons with holes will occur. This action somewhat reduces the available current gain of the transistor by causing less increase in collector current for a given increase in emitter current.

The voltage and power gains of the NPN junction transistor are of about the same magnitude as those of the PNP junction transistors. Like those of the grounded base PNP transistors previously described, these gains are principally because of the high resistance of the collector circuit compared to the low resistance in the emitter circuit.

Comparison with Electron-Tube Triode

The three elements in a triode electron tube and the corresponding elements in the PNP and NPN junction transistors are illustrated in figure 2-17.

The transistors have a collector terminal that corresponds to the triode plate, a base terminal that corresponds to the triode grid, and an emitter terminal that corresponds to the triode cathode. In the PNP transistor the collector collects holes. In the NPN transistor the collector collects electrons. The corresponding symbols for the PNP and NPN junction transistors are shown at the right of the schematic drawings in figure 2-17.

The cathode of the electron-tube triode is common to the grid input and plate output circuits. The emitter is common to the base input and collector output circuits. The grid-bias battery in the electron-tube triode corresponds

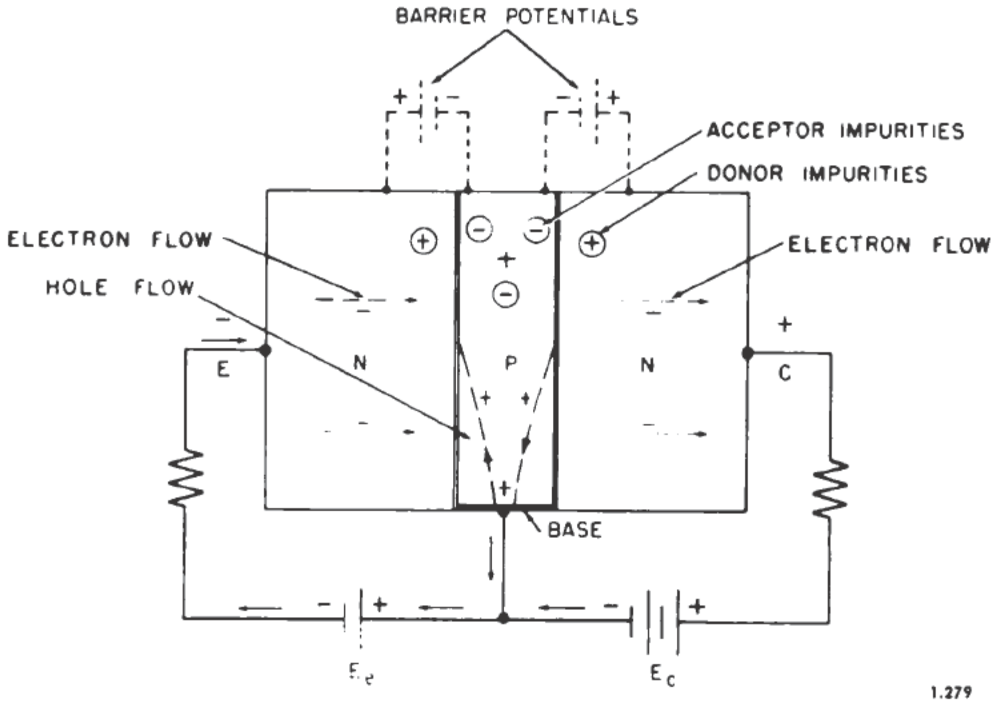


Figure 2-16.—NPN junction transistor.

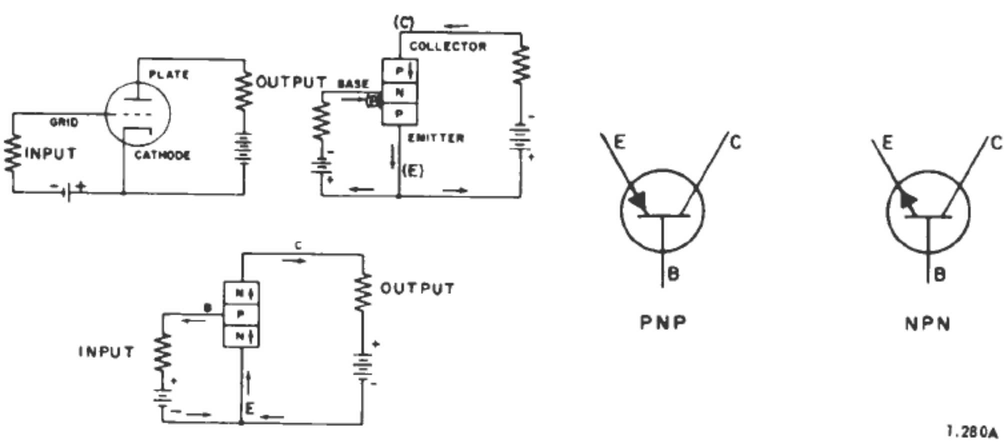
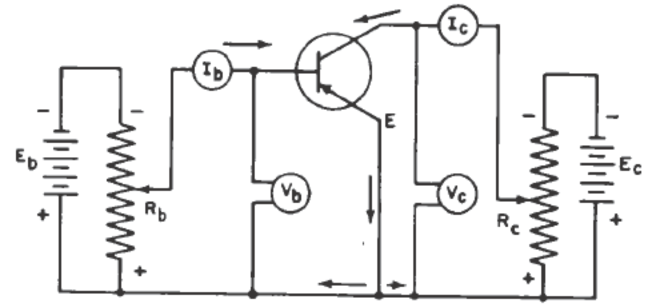


Figure 2-17.—Corresponding elements in triode and transistor.

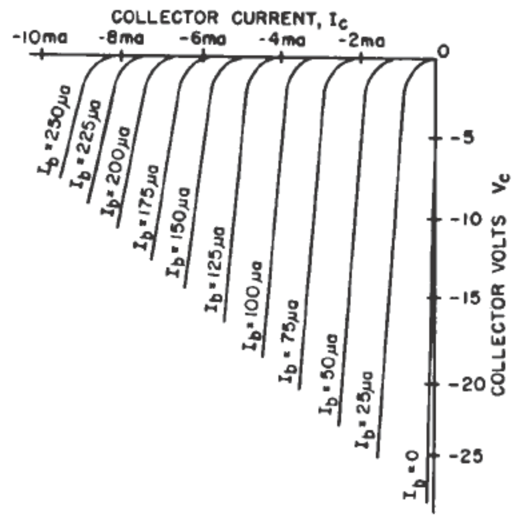
to the base-emitter bias battery in the transistor input circuit. Signal voltages are developed across the resistors. This action is described later in this training course under amplifiers.

For a grounded-emitter PNP-junction transistor, the relation between collector current and collector voltage for different values of emitter-base current, I_b is illustrated in figure 2-18. The circuit for determining the relation between these values is illustrated in figure 2-18,A. Battery E_b and potentiometer R_b bias the base-emitter input circuit in the forward, or low-resistance, direction for a PNP junction transistor. Moving the potentiometer arm of R_b toward the negative end of R_b will increase the forward bias on the input circuit and therefore will increase I_b and I_c. Moving the arm of potentiometer R_c toward the negative end of R_c will increase the collector voltage, V_c.

Collector current is measured along the X axis, and collector voltage is measured along the Y axis in the collector characteristic curves (fig. 2-18,B). These curves represent the relation between collector current and collector voltage for various values of emitter-base current. The curves indicate that collector current (I_c) depends principally upon the magnitude of the emitter (base) current (I_b) and that, for a given value of emitter current, wide variations in collector voltage will have little effect on the magnitude of the collector current.



A PNP TRANSISTOR GROUNDED EMITTER CIRCUIT



B TRANSISTOR CHARACTERISTICS 1.282

Figure 2-18.—Grounded emitter PNP junctions transistor static characteristics.

CHAPTER 3

POWER SUPPLIES FOR ELECTRONIC EQUIPMENTS

INTRODUCTION

The electron tubes in the various electronic equipments used in the Navy require voltages of various values for their filament, grid, screen, and plate circuits. It is the function of a power supply to supply these voltages at the necessary current ratings. Except for filament power, which may be alternating current, the output from a power supply must be nearly pure direct current, and the voltage must be of the correct value for the circuits of the equipment being used. Transmitters require more power than receivers, and consequently transmitter power sources must supply higher voltages with greater current ratings than receiver power sources.

Power used to heat the filaments of tubes is sometimes called the A supply and is normally furnished at a low voltage, with a relatively high current drain. In portable and mobile equipments filament power is furnished by batteries, generators, or dynamotors. In permanent land-based or shipboard installations, filament power is obtained from the standard a-c line via step-down transformer.

The plate and screen power supply, normally called the B supply, furnishes power at a relatively high voltage and low current. In portable equipment the plate voltage may be supplied by batteries or a hand generator. Mobile sets may employ dynamotors or vibrators operated from batteries to generate the high voltage necessary for the plates and screens. Permanent installations ordinarily use a transformer-rectifier-filter system for the high-voltage plate and screen supply.

When a separate grid-bias voltage is used, it is sometimes called the C supply. In many instances grid-bias voltage is obtained by some means of self-bias. However, for large power-amplifier tubes, or where self-bias is not satisfactory, a separate transformer-rectifier-filter system or a d-c generator may be employed.

Radio power supplies may be conveniently divided into three general classes—battery, alternating current, and electromechanical systems. However, in this chapter primary consideration is given to obtaining the necessary d-c potentials from an a-c source via a transformer-rectifier-filter system. Dynamotors and vibrator power supplies are also treated briefly in this chapter; but batteries and generators are included in basic electricity texts and hence are not discussed in this chapter.

B-VOLTAGE SUPPLY

The B supply for Navy electronic equipment is generally obtained via a transformer-rectifier-filter system from the ship's a-c bus. The voltage and phase considerations are dependent upon the type of ship. On ships having direct current as the power source, a motor-generator is generally utilized to supply alternating current to the electronic equipment. There are many types of low-power emergency communications equipment that use storage batteries or dry batteries for power sources. Such equipments use vibrator-type rectifier circuits or dynamotors to convert the low d-c voltage into a d-c voltage sufficiently high for the plates and screen grids.

Airborne equipments use dynamotors for the high-voltage B supply. Inverters are also used in airborne equipments. An inverter utilizes a low-voltage d-c supply to operate a d-c motor which in turn drives an a-c generator. The a-c generator supplies power generally at 400 or 800 cycles. Electromechanical systems such as these are treated later in this chapter.

If it is assumed that a suitable a-c supply is available, the problems involved in obtaining a suitable high d-c potential are as follows: (1) if the voltage is not sufficiently high, a step-up transformer must be provided; (2) the voltage must be rectified—that is, changed into pulsating direct voltage; (3) the ripples must be removed; and (4) some form of voltage regulation must be employed. Most of the remainder of this chapter is devoted to these problems.

RECTIFIERS FOR POWER SUPPLIES

The majority of rectifier systems employing electron tubes utilize either high-vacuum or gas-filled tubes. The high-vacuum diode is the most widely used in low-current applications. The hot-cathode mercury-vapor tube is widely used in high-current applications. The presence of mercury vapor in the tube envelope reduces the vacuum and results in low internal resistance, thus allowing a large amount of current to be drawn.

On the other hand, dry-disk rectifiers, such as selenium and copper-oxide rectifiers, do not employ electron tubes. Selenium rectifiers are sometimes used as plate-supply rectifiers in small radio receivers. Copper-oxide rectifiers have miscellaneous applications, for example, they may be used as instrument rectifiers or bias-supply rectifiers. Both of these rectifiers have the advantage of not requiring heater current or warmup time.

High-Vacuum Rectifiers

A diode acts as a rectifier because it passes current in only one direction - from cathode to plate. Conduction takes place only when the plate is positive with respect to the cathode. The characteristics of diodes are discussed in chapter 1. The important characteristics of the high-vacuum rectifier tube are its maximum peak plate current and its maximum inverse peak plate voltage ratings.

The peak plate current is limited by the number of electrons emitted by the cathode and therefore is dependent upon cathode construction. It is evident that current in a single-diode rectifier never flows for more than one-half of each a-c cycle. At the power frequency, the d-c output current as indicated by a meter is less than one-half the peak plate current.

The peak inverse plate voltage is the peak negative voltage that is applied to the plate during the portion of the cycle when the tube is not conducting. The d-c output voltage and peak inverse voltage vary with the type of circuit. In general, the peak inverse voltage is equal to or twice the peak value of the d-c output voltage.

An important factor in circuit design utilizing high-vacuum tubes is the voltage drop across the tube during the conducting half cycle. In most high-vacuum rectifiers this voltage drop is relatively large compared to the drop in gas-filled tubes and is the limiting factor in the design of high-current electron-tube rectifiers. The drop increases with an increase in current and results in poor regulation for high-current applications. High-vacuum rectifiers are used almost universally in receiver power supplies and also in many high-voltage low-current applications.

The types of rectifiers used in receivers are generally provided with directly heated, oxide-coated filaments. Most of these rectifiers consist of two units in a single envelope because they are generally used in pairs for full-wave operation (described later).

In addition to low-power applications, high-vacuum rectifiers for high-voltage applications have been designed to withstand a peak inverse voltage of 100,000 volts. Commercial units have been built to provide peak plate currents as high as 7.5 amperes. High-vacuum rectifiers are seldom used for high-voltage high-current applications except where inherent ruggedness outweighs the other disadvantages. The mercury-vapor rectifier is used in the majority of medium- and high-power equipments.

RECTIFIER CIRCUITS

Half-Wave Rectifier

A half-wave rectifier is a device by means of which alternating current is changed into pulsating direct current by permitting current to flow through the device only during one-half of each cycle.

In a diode, electrons are attracted to the plate when it is more positive than the cathode. When the plate becomes negative with respect to the cathode, electrons are repelled by it and no electron stream can flow in the tube. Therefore, a single diode may be used as a half-wave rectifier because electrons can flow in the tube during only the half of the cycle when the plate is positive relative to the cathode.

Figure 3-1,A, shows a simple half-wave rectifier circuit. The primary winding of the transformer is shown connected to an a-c input source. The principal action of the transformer is to increase the voltage from 115 volts to a higher value in the main secondary winding. The action of the small secondary winding at the top of figure 3-1,A, steps down the 115 volts to a suitably lower voltage. It supplies heater current to the filament of the rectifier tube. The actual connections of this winding are not shown in the figure, but are indicated by the symbol, XX.

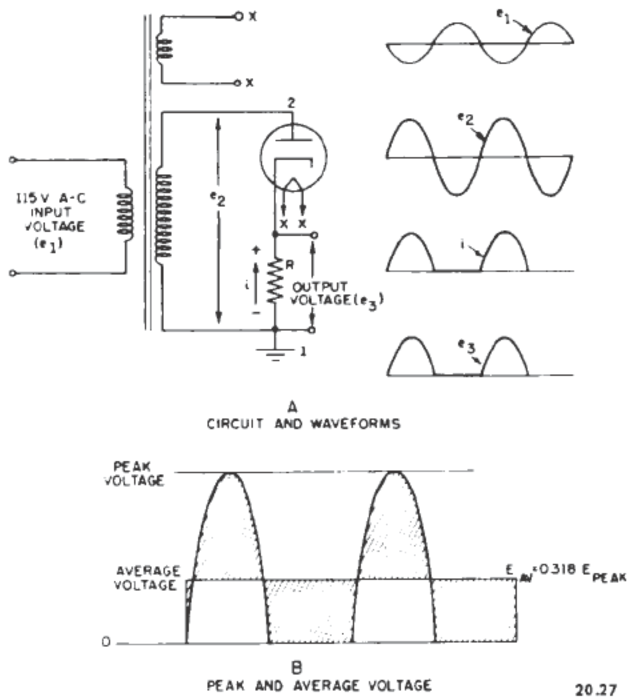


Figure 3-1.—Simple half-wave rectifier circuit and waveforms.

The upper end of the high-voltage secondary winding is connected to the plate of the diode and the other end is connected to a junction with ground and a load resistance represented by the resistor, R. The load resistor is connected to the cathode and is in series with the tube.

The operation of the circuit is illustrated by the waveforms at the right of figure 3-1. The alternations of the input voltage, e_1, are reproduced by the transformer with an increase in voltage, e_2, in the secondary winding. The waveforms indicate a 180° difference in phase between e_1 and e_2. This difference is characteristic of induced voltages. The induced secondary voltage, e_2, is impressed across the diode and its series load resistance. This voltage causes current i to flow through the diode and its series resistor on the positive half cycles (when the plate is positive). The resultant output voltage, e_3, across the load has a pulsating waveform, as shown in the figure. This pulsating waveform is called a RIPPLE VOLTAGE.

When point 2 is positive and point 1 is negative, electrons flow from the ground junction (1), through the load resistor (R), to the cathode, to the plate, and thus to the upper terminal (2) of the transformer. The secondary winding thus acts as the immediate source of voltage for the current flow.

Because current flows in only one direction (point 1 to point 2) through the diode and its load, the polarity of the voltage across the load resistance is always as shown. The fact that the positive point of the load resistance is connected to the cathode of the tube may be confusing when the cathode is thought of as being the negative element of the tube. However, the only requirement necessary for conduction is that the plate be more positive than the cathode. The cathode can be positive with respect to other points in the circuit. The voltage drop across the tube is usually quite small when compared with that across the load resistance.

The load is connected to the cathode, as shown, rather than in the plate circuit to enable the use of a common ground for the negative side of the load and transformer winding, with ground continuity maintained throughout the entire input cycle.

The half-wave rectifier utilizes the transformer during only one-half of the cycle, and therefore for a given size of transformer less power can be developed if the transformer were utilized on both halves of the cycle. In other words, if any considerable amount of power is to be developed in the load the half-wave transformer must be relatively large compared with that it would have to be if both halves of the cycle were utilized. This disadvantage limits the use of the half-wave rectifier to applications that require a very small current drain. The half-wave rectifier is widely used for commercial a-c d-c radio receivers and for the accelerating voltage supplies of oscilloscopes.

For a half-wave rectifier (assuming half sine waves, as shown) the rms voltage, E, in the tube is

$$E = \frac{e_{max} \times 0.707}{2}$$

and the average voltage (fig. 3-1, B) is

$$E_{av} = \frac{e_{max} \times 0.636}{2}$$

The rms and average output currents are determined in the same way.

Because the d-c load current flows through the transformer secondary in only one direction, there is a tendency for the molecules in the iron core to become oriented in one direction. This effect is called D-C CORE SATURATION and reduces the effective inductance of the transformer. The net effective inductance with the d-c core saturation effect present is known as TRANSFORMER INCREMENTAL INDUCTANCE. Thus, the transformer incremental inductance is reduced with increasing d-c load current. The resultant effect is to decrease the primary counter emf to a greater degree and thus increase the load component of primary current correspondingly. Therefore, the efficiency of the transformer is reduced and the regulation is impaired. The output is far from being a continuous d-c voltage and current and for these reasons the half-wave rectifier circuit is seldom used for high current loads.

Full-Wave Rectifier

A full-wave rectifier is a device that has two or more elements so arranged that the current output flows in the same direction during each half-cycle of the a-c supply.

Full-wave rectification may be accomplished by using two diodes in the same envelope (a dual diode) with a common cathode connected to one end of the load resistance, as shown in figure 3-2,A. The other end of the load resistor is connected to the center tap, C, of the transformer secondary. The two halves of the secondary winding, AC and BC, may be a center-tapped winding as shown, or they may be separate windings. In either case the load circuit is returned to a point midway in potential between A and B so that the load current is divided equally between the two tubes.

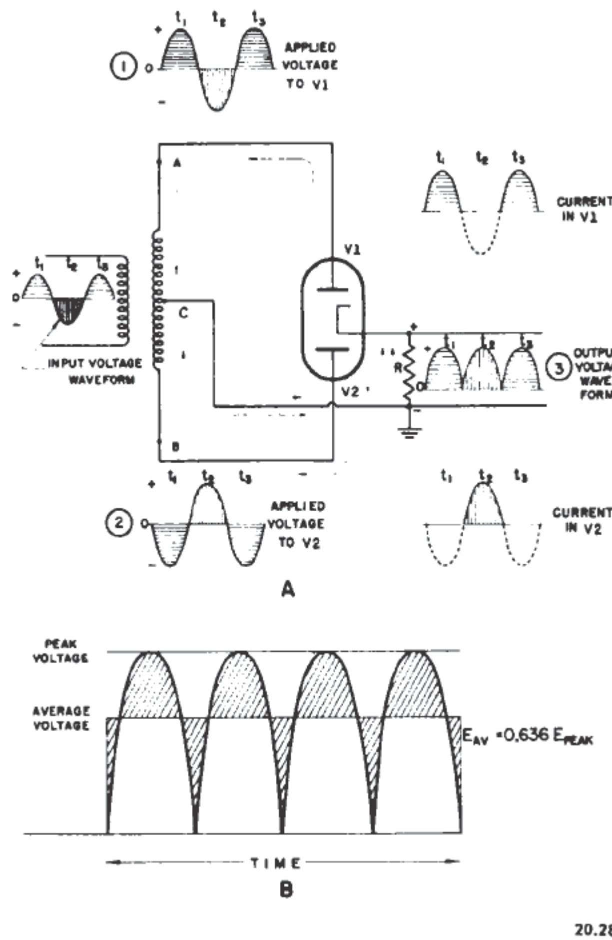


Figure 3-2.—Simple full-wave circuit and waveforms.

That part of the secondary winding between A and C may be considered a voltage source that produces a voltage of the shape shown at (1) in figure 3-2,A. This voltage is impressed across the plate to cathode of V1 in series with load resistor R. During the half-cycle, marked t_1, the plate of V1 is positive with respect to its cathode. Therefore, electrons flow in the direction indicated by the solid arrows. This flow of electrons from ground up through load resistor R makes the cathode positive with respect to ground. Thus the load voltage is developed across R between cathode and ground. During this same half-cycle the voltage across BC is negative, as shown at (2) in figure 3-2,A, and the plate of V2 is negative with respect to the cathode. Thus during the time V1 conducts, V2 is nonconducting. A half-cycle later, during interval t_2, the polarity of the voltages on the plates of the two tubes is reversed. V2 now conducts, and V1 is nonconducting. The electron flow through V2 is in the direction indicated by the dotted arrows. This current also flows from ground up through R and makes the cathode positive with respect to ground. Thus another half-cycle of load voltage is developed across R. A study of the figure shows that only one section of the twin diode is conducting at any given instant.

Because there are two pulsations of current in the output for each cycle of the applied alternating voltage, the full-wave rectifier utilizes the power-supply transformer for a greater percentage of the input cycle. Hence, the full-wave rectifier is more efficient than the half-wave rectifier, has less ripple effect, and may be used for a much wider variety of applications.

For a full-wave rectifier (having sine waves, as shown) the rms output voltage, E, across the load is

$$E = e_{max} \times 0.707$$

and the average voltage (shown in fig. 3-8,B) is

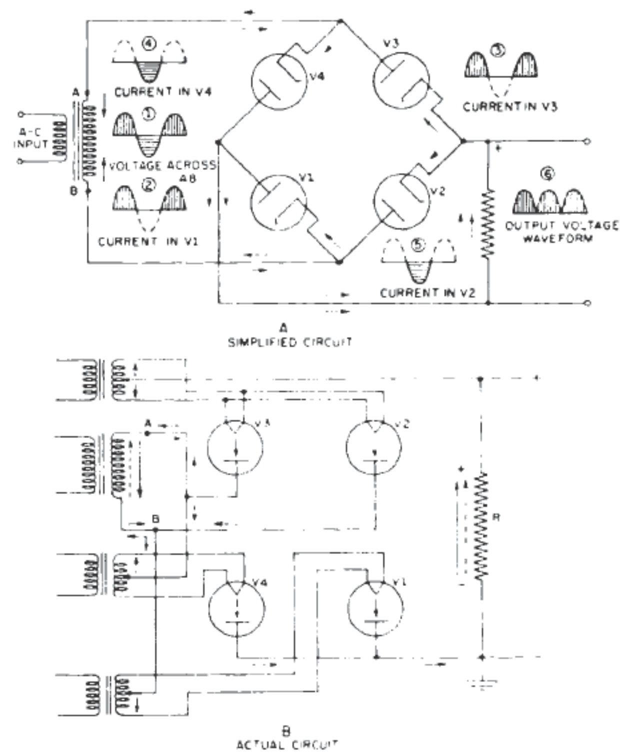
$$E_{av} = e_{max} \times 0.636$$

The rms and average output currents are similarly determined. In the case of both the half-wave and the full-wave rectifier, no filtering is assumed in computing the rms and average values.

The d-c load current flows equally through the two halves of the transformer secondary and in opposite directions. Thus the ampere turns are equal and act in opposite directions so that there is no tendency to orient the molecules of the iron core in any one direction. Therefore the transformer inductance is not reduced as it is in the half-wave rectifier and both the voltage regulation and efficiency are improved. The full-wave rectifier is used widely in radio transmitters and receivers.

Bridge Rectifier

If four rectifiers are connected as shown in figure 3-3, A, the circuit is called a BRIDGE RECTIFIER. The input (waveform ①) to such a circuit is applied to diagonally opposite corners of the network, and the output is taken from the remaining two corners.



20.29

Figure 3-3.—Bridge-rectifier circuit.

During one-half cycle of the applied alternating voltage, point A becomes positive with respect to point B by the amount of the voltage induced in the secondary of the transformer. During this time, the voltage across AB may be considered to be impressed across a load consisting of V1, load resistor R, and V3 in series. The voltage applied across these tubes makes their plates more positive than their cathodes, and electrons flow in the path indicated by solid arrows. The waveform of this current is shown at ② and ③. One-half cycle later, V1 and V3 are nonconducting, and an electron stream flows through V4, R, and V2 in the direction indicated by the dotted arrows. The waveform of this current is shown at ④ and ⑤. The current through the load, R, is always in the same direction. As this current flows through R it develops a voltage having the waveform shown at ⑥. The bridge rectifier is a full-wave rectifier because current flows in the load during both halves of each cycle of applied alternating voltage.

One advantage of a bridge rectifier over a conventional full-wave rectifier is that with a given transformer the bridge circuit produces a voltage output nearly twice that of the full-wave circuit. This increase in voltage may be illustrated by assigning values to some of the components in figure 3-2, A, and 3-3. Assume that the same transformer is used in both figures. Note that the center tap is not used in the bridge circuit. The peak voltage developed between A and B is assumed to be 1,000 volts in both figures. In the full-wave circuit in figure 3-2, A, the peak voltage from the center tap, C, to either A or B is 500 volts. Because only V1 or V2 is conducting at any instant, the maximum voltage that can be rectified at any instant is 500 volts. Therefore, the maximum voltage that can be developed across the load resistor, R, is 500 volts, less the small voltage drop across the tube that is conducting. In the bridge circuit of figure 3-9, however, the maximum voltage that can be rectified is the full voltage of the secondary of the transformer, or in this case 1,000 volts. Therefore, the voltage that can be developed across the load resistor, R, is 1,000 volts less the voltage drop across the two tubes that are conducting. Thus, the full-wave bridge circuit produces a higher output voltage than the conventional full-wave rectifier does with the same transformer.

A second advantage of the bridge circuit is that the peak inverse voltage across a tube is only half the peak inverse voltage impressed on a tube in a conventional full-wave rectifier that is designed for the same output voltage. For example, if the two circuits are to produce the same output voltage (1,000 v), the transformer secondary in the full-wave rectifier (fig. 3-2, A) has a 2,000-volt peak developed across it, while that for the bridge rectifier (fig. 3-3) has only a 1,000-volt peak. When V1 is figure 3-2, A, is not conducting, its plate is made negative relative to its cathode by a maximum voltage of 2,000 volts. The same is true for V2. This negative voltage is called the PEAK INVERSE VOLTAGE (PIV) and if greater than the maximum PIV rating of the tube, breakdown within the tube will occur. In figure 3-3, however, when tubes V1 and V3 or V2 and V4 are not conducting, the peak inverse voltage for any one tube is then 1,000 volts, which is half of the peak inverse voltage across either tube in figure 3-2, A.

The bridge rectifier circuit has a disadvantage in that three filament transformers are required for the tubes. In the actual circuit in figure 3-3, B, the filament transformer connections of V2 and V3 are operated at the same relative potential and therefore may be connected to the same filament winding.

The filaments of V1 and V4, however, are returned to opposite ends of the high-voltage secondary of the power-supply transformer and therefore operate at the full potential difference that exists across the load. Thus, if the filaments of V1 and V4 were supplied by a single filament transformer winding, the common connection would short-circuit the load. Therefore, the filament windings of V1 and V4 must be insulated from each other to withstand the full output voltage across the load. If the lower end of the load, R, is grounded, filament windings for V1 and V4 operate alternately at the full difference of potential of the high-voltage secondary with respect to ground. Thus V1 and V4 must be well insulated from ground also.

This disadvantage does not apply to the rectifier elements of the dry-disk type which are often used in bridge circuits.

Mercury-Vapor Rectifiers

The introduction of mercury-vapor tubes to the electronics industry was one of the greatest single contributions to the development of high-powered electronic equipment. As a help in understanding the importance of the mercury-vapor tube, consider the foremost disadvantage of high-vacuum rectifiers. As stated previously, the voltage drop across a high-vacuum tube varies with the load current; and when the current varies widely, the regulation is poor. The high-vacuum rectifier used on heavy loads has a relatively high loss and low efficiency. In some high-power applications a water cooling system is employed to carry the heat from the tube elements. The power loss in a high-vacuum tube is usually of the order of 15 percent of the input power to the rectifier. In comparison, the power loss in a mercury-vapor rectifier is only about 1.5 percent of the total input.

The greater efficiency of the mercury-vapor rectifier is a result of the low-voltage drop across the tube. In normal operation this voltage drop rarely exceeds 15 volts, even when the tube is operating at very high values of load current. The filament of the high-vacuum rectifier is surrounded by a space charge which acts as a shield to impede electron flow between cathode and plate.

In a mercury-vapor rectifier, a small amount of mercury is introduced into the tube envelope. Because of the low pressure within the tube, the mercury vaporizes completely as the unit reaches normal operating temperature. If a positive potential is applied to the anode of the tube, electrons are emitted from the filament and move toward the anode. Because the tube envelope is filled with mercury vapor, collisions occur between the moving electrons and the atoms of mercury.

Each collision knocks an electron of a mercury atom away from the influence of the nucleus. The atom, now minus an electron, becomes a positive ion. This ion is then drawn toward the negative filament and is promptly neutralized by one of the electrons forming the space charge around that element. Ionization of mercury vapor occurs when the potential gradient between the plate and cathode is 10.4 volts. Any further increase in plate voltage will ionize more atoms—each in turn neutralizing a space electron—until the drop reaches 15 volts, and at that time the tube will no longer show a plate-voltage rise for a proportional rise in current.

Figure 3-4 shows graphically the i_p-e_p relation during ionization. This figure indicates the most useful characteristics of a mercury-vapor tube—THE VOLTAGE DROP ACROSS THE TUBE REMAINS AT A CONSTANT VALUE OF 15 VOLTS REGARDLESS OF THE CURRENT FLOWING THROUGH THE TUBE, provided the rated tube current is not exceeded. On overload the voltage drop increases to some extent. When the tube drop exceeds 22 volts the filament may be damaged by excessive bombardment of positive ions. At potentials below 22 volts, this bombardment is insufficient to cause damage to the filament.

Another important characteristic of the mercury-vapor tube is its maximum inverse-voltage rating. This is the sparking voltage through the mercury vapor in a direction opposite to that of normal flow and is always less than it would be if the vapor were not present. A mercury-vapor tube always has a lower flashback voltage than a high-vacuum tube of similar construction. Nevertheless,

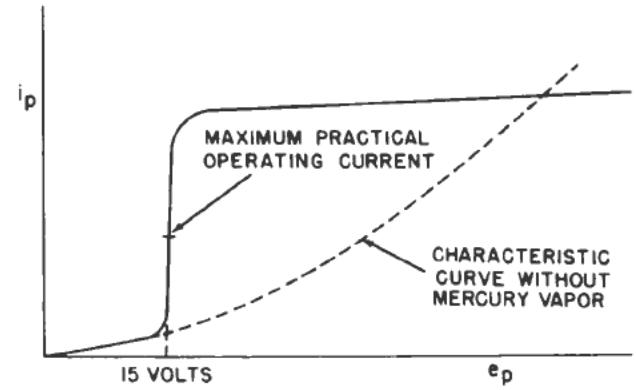


Figure 3-4.—Current vs voltage relation in a mercury-vapor diode.

mercury-vapor tubes with high inverse-voltage ratings have been developed. For example, a mercury-vapor diode having an output of 10 amperes and a maximum safe peak inverse voltage of 22,000 volts is used extensively in broadcast-transmitter power supplies.

In the practical operation of mercury-vapor rectifiers the vapor must reach its proper operating temperature before plate voltage is applied. If this precaution is not taken, the high voltage drop across the tube causes secondary emission from plate to cathode, and arcback occurs. In high-vacuum rectifiers the only factor considered in cathode heating is the emission of electrons. The cathode of the mercury-vapor tube must not only emit free electrons for conduction, but must also heat the surrounding space in order for the mercury-vapor temperature to be in the range of 20° to 60° centigrade. The cathode construction indicated in figure 3-5 facilitates this heating.

The heat given off by the inner turns of the spiral filament is absorbed by the outer turns. Radiation from the outer surface is reduced by a polished shield surrounding the filament. The plate (anode) is a metal cup fitting over the cathode. This arrangement reduces the tendency to arc back. It also shields the plate-cathode region from external electric fields. The graph shown in figure 3-6 plots flashback voltage and tube drop against operating temperature (C°). Below the indicated operating range of temperatures, the tube drop is excessive with inherent danger of positive-ion bombardment. At high operating temperatures, flashback potentials drop to an intolerably low value.

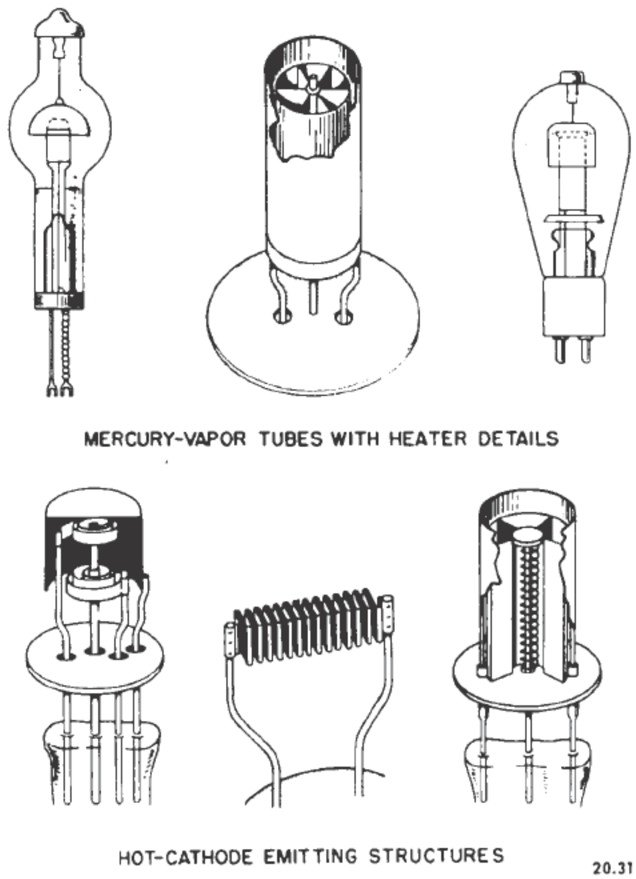


Figure 3-5.—Mercury-vapor diodes with hot-cathode emitting structure.

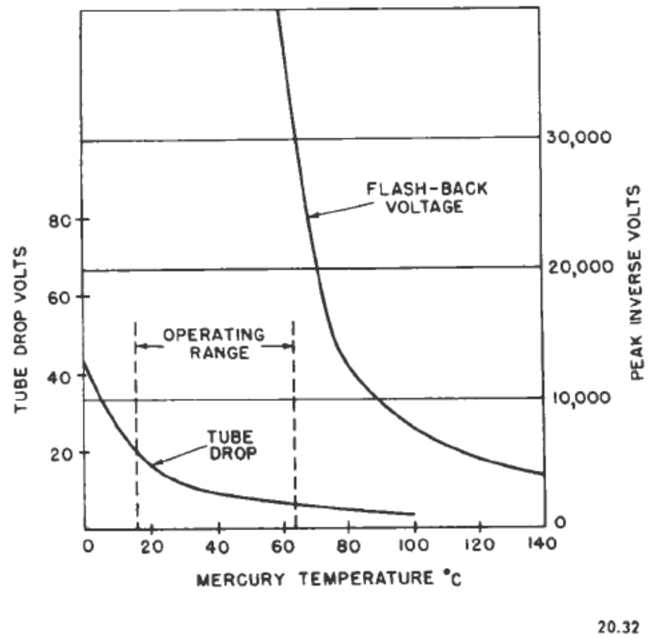


Figure 3-6.—Peak inverse voltage vs temperature.

Metallic Rectifiers

When dissimilar metals are in contact, electrons flow in greater numbers in one direction across the area of contact than in the other direction. Metallic rectifiers operate on this principle. The two combinations of substances that are most widely used for dry-disk rectifiers are: (1) A thin film of copper oxide (cuprous oxide) and copper, and (2) selenium on either iron or aluminum and an alloy. Metallic rectifier units are represented by the symbol in figure 3-7,A. The arrowhead in the symbol points in the opposite direction to the electron flow.

Figure 3-7,B, indicates how a metallic rectifier may be used in place of a diode rectifier. The waveforms indicate that much more current flows in one direction than in the other. Although the copper-oxide rectifier is shown, the selenium rectifier may be used.

Copper-Oxide Rectifiers

In the copper-oxide rectifier shown in figure 3-8,A, the oxide is formed on the copper by partial oxidation of the copper by a high temperature. In this type of rectifier the electrons flow more readily from the copper to the oxide than from the oxide to the copper. External electrical connection may be made by connecting terminal lugs between the left pressure plate and the copper and between the right pressure plate and the lead washer.

For the rectifier to function properly, the oxide coating must be very thin. Thus, each individual unit can stand only a low inverse voltage. Rectifiers designed for moderate and high-power applications consist of many of these individual units mounted in series on a single support. The lead washer enables uniform pressure to be applied to the units so that the internal resistance may be reduced. When the units are connected in series, they normally present a relatively high resistance to the current flow. The resulting heat developed in the resistance must be removed if the rectifier is to operate satisfactorily. Many commercial rectifiers have copper fins between each unit to dissipate the excess heat. The useful life of the unit is extended by keeping

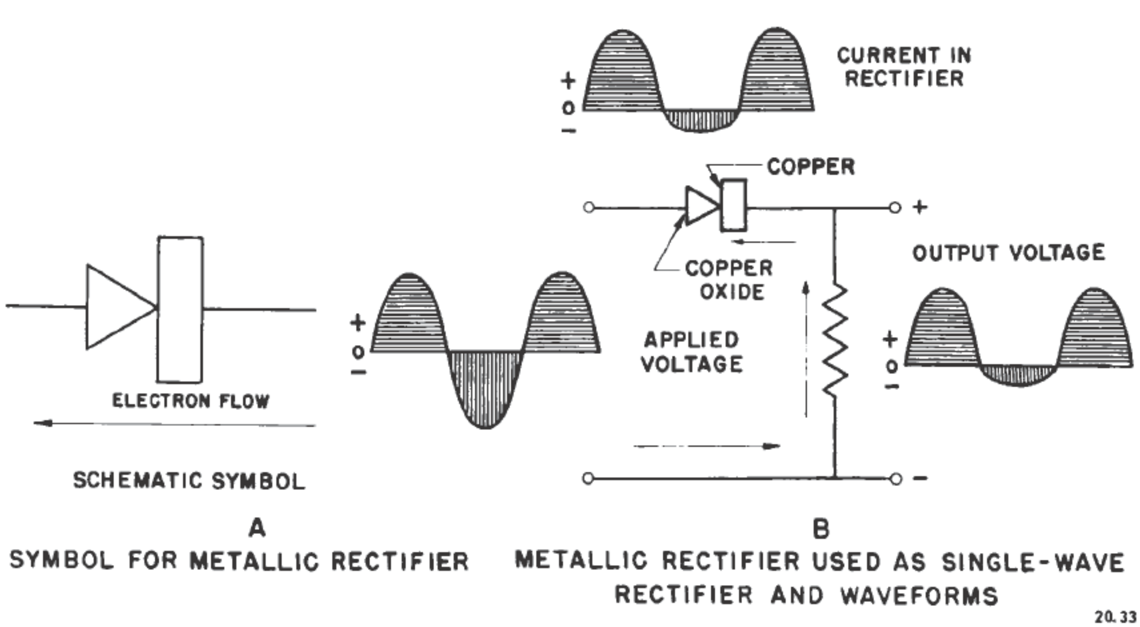


Figure 3-7.—Metallic rectifier symbol and waveforms.

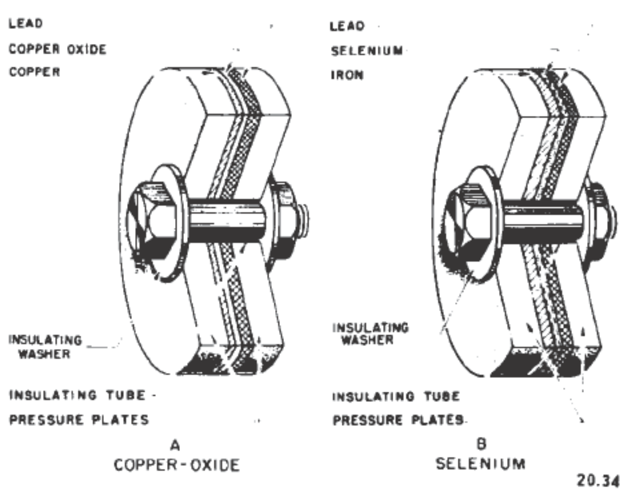


Figure 3-8.—Metallic rectifier construction.

the temperature low (below 140° F). The efficiency of this type of rectifier is generally between 60 and 70 percent.

Selenium Rectifiers

Selenium rectifiers function in much the same manner as copper-oxide rectifiers. A selenium rectifier is shown in figure 3-8,B. Such a rectifier is made up of an iron disk coated with a thin layer of selenium. In this type of rectifier the electrons flow from the selenium to the iron.

Commercial selenium rectifier units are designed to pass 50 milliamperes per square centimeter of plate area. This type of rectifier may be operated at a somewhat higher temperature than a copper-oxide rectifier of similar rating. The efficiency is between 65 and 85 percent, depending on the circuit and the loading. As in the case of the copper-oxide rectifier, any practical number of units may be bolted together in series to increase the voltage rating. Larger element disks and the necessary cooling fins may be used for higher current ratings. Also forced-air cooling may be used.

Metallic rectifiers may be used not only as half-wave rectifiers, as shown in figure 3-7, but also in full-wave and bridge circuits. In each of these applications the action of the metallic rectifier is similar to that of a diode.

Metallic rectifiers may be used in battery chargers, instrument rectifiers, and many other applications including welding and electroplating. Commercial radios also frequently use selenium rectifiers in the high-voltage power supply, as do other electronic equipments.

Silicon Diodes

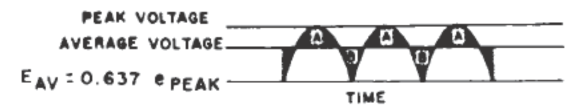
These and other types of metallic rectifiers are composed of materials called semiconductors. They are also known as "solid state" rectifiers and "semiconductor diodes." Germanium and silicon are used extensively. Crystals of these materials are grown from a "melt" which includes small quantities of impurity substances such as gallium or indium (P material), and arsenic or phosphorous (N material). The crystal structure of the resulting metallic "chips" or "wafers" permits current flow in one direction only. According to the type of semiconductor material and the impurity substance the structure of the crystals is such as to provide either an excess of electrons (N-type crystals) or a deficiency of electrons (P-type crystals) for current conduction. Junctions of these materials comprise the semiconductor diode.

Many types of semiconductor diodes are available. They vary in size from tiny ones hardly bigger than a pinhead used in subminiature circuitry, such as computers, to large 500-ampere diode rectifiers used in power supplies. A silicon diode rectifier about an inch long and an inch in diameter (with heat radiator) will supply a d-c current of 50 amperes (peak) and has a peak inverse voltage rating of 60 volts. Semiconductor action is described in more detail in the chapter "Introduction to Transistors."

FILTER CIRCUITS

The preceding paragraphs have discussed methods of converting alternating current into pulsating direct current. Most electronic equipments require a smooth d-c supply, approaching the ripple-free output of a battery. Conversion of pulsating direct current to pure direct current is accomplished by the use of properly designed filters.

The unfiltered output of a full-wave rectifier is shown in figure 3-9. The polarity of the output voltage does not reverse, but its magnitude fluctuates about an average value as the successive pulses of energy are delivered to the load. In figure 3-9, the average voltage is shown as the line that divides the waveform so that area A equals area B. The fluctuation of voltage above and below this average value is called RIPPLE. The frequency of the main component of the ripple for the full-wave rectifier shown in figure 3-9 is twice the frequency of the voltage that is being rectified. In the case of the half-wave rectifier the ripple has the same frequency as the input alternating voltage. Thus, if the input voltage is obtained from a 60-cps source, the main component of ripple in the output of a half-wave rectifier is 60 cps, and in the full-wave rectifier it is 120 cps.



20.35

Figure 3-9.—Unfiltered output voltage of a full-wave rectifier.

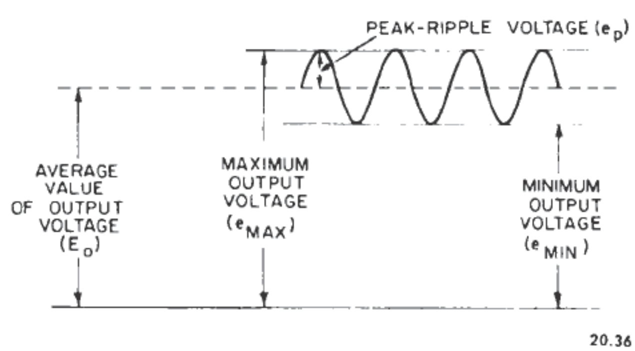
The output of any rectifier is composed of a direct voltage and an alternating or ripple voltage. For most applications, the ripple voltage must be reduced to a very low amplitude. The amount of ripple that can be tolerated varies with different applications of electron tubes.

The PERCENTAGE OF RIPPLE is 100 times the ratio of the rms value of the ripple voltage at the output of a rectifier filter to the average value, E_o, of the total output voltage. Figure 3-10 indicates graphically how the percentage of ripple may be determined. It is assumed that the ripple voltage is of sine waveform. The formula for determining the percentage of ripple is

$$\text{percentage of ripple} = \frac{E_{rms}}{E_o} \times 100$$

where E_{rms} = 0.707 of e_p, and e_p is the peak value of the ripple voltage.

A circuit that eliminates the ripple voltage from the rectifier output is called a FILTER. Filter systems in general are composed of a combination of capacitors, inductors, and in some cases resistors.



20.36

Figure 3-10.—Percentage of ripple.

Capacitance Input Filter

Ripple voltage exists because energy is supplied in pulses to the load by the rectifier. The fluctuations can be reduced considerably if some energy can be stored in a capacitor while the rectifier is delivering its pulse and allowed to discharge from the capacitor between pulses.

Figure 3-11, A, shows the output of a half-wave rectifier. This pulsating voltage is applied across a filter capacitor (C in fig. 3-11, B) to supply the load, R. Because the rate of charge of C is limited only by the reactance of the transformer secondary and the plate resistance of the tube in the rectifier, the voltage across the capacitor can rise nearly as fast as the half-sine-wave voltage output from the rectifier. In other words, the RC charge time is relatively short. The capacitor, C, therefore, is charged to the peak voltage of the rectifier within a few cycles. The charge on the capacitor represents a storage of energy. When the rectifier output drops to zero, the voltage across the capacitor does not fall immediately. Instead, the energy stored in the capacitor is discharged through the load during the time that the rectifier is not supplying energy (when the anode is negative). The voltage across the capacitor (and the load) falls off very slowly if it is assumed that a large capacitance and a relatively large value of load resistance are employed. In other words, the RC discharge time is relatively long. The amplitude of the ripple therefore is greatly decreased, as may be seen in figure 3-11, C.

Figure 3-11, D, shows the input voltage to the filter when a full-wave rectifier is used, and figure 3-11, E, shows the resulting output-voltage waveform.

After the capacitor has been charged (with either half-wave or full-wave input), the rectifier does not begin to pass current until the output voltage of the rectifier exceeds the voltage across the capacitor. Thus, in figures 3-11, C and E, current begins to flow in the rectifier when the rectifier output reaches a voltage equal to the capacitor voltage. This occurs at some time, t_1, when the rectifier output voltage has a magnitude E_1. Current continues to flow in the rectifier until slightly after the peak of the half-sine wave, at time t_2. At this time the sine-wave voltage is falling faster than the capacitor can discharge. A short pulse of current, beginning at t_1 and ending at t_2, is therefore supplied to the capacitor by the power source.

The average voltage of the rectifier output is shown in figures 3-11, A and D. Because the capacitor absorbs energy during the pulse and delivers this energy to the load between pulses, the output voltage can never fall to zero. Hence, the average voltage of the filtered output (fig. 3-11, C and E) is greater than that of the unfiltered input (fig. 3-11, A and D). However, if the resistance of the load is small, a heavy current is drawn by the load and the average or direct voltage falls. For this reason, the simple capacitor filter is not used with rectifiers that must supply a large load current. Also the input

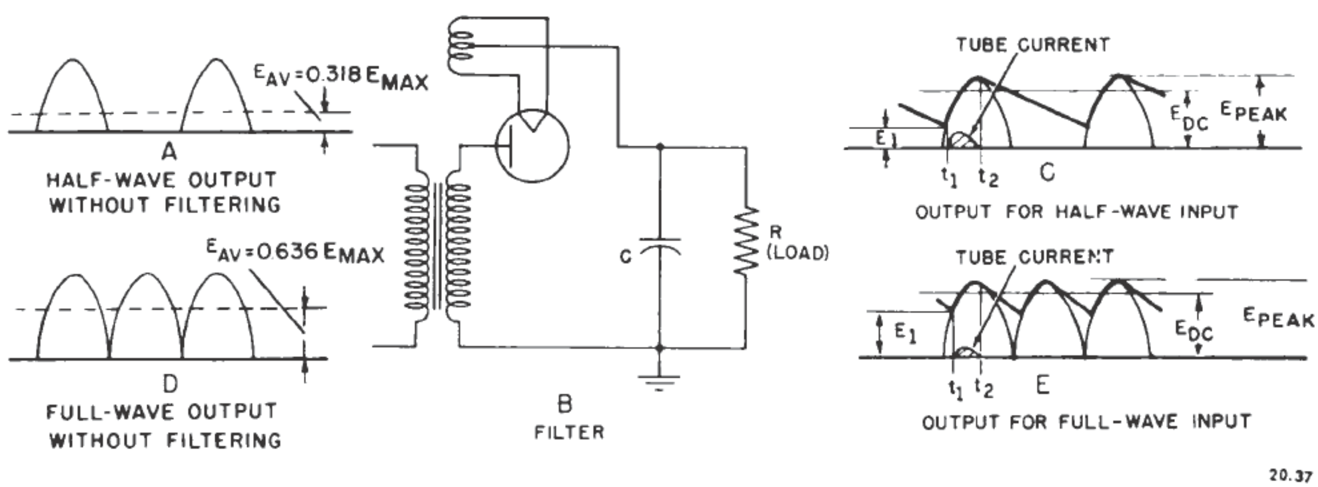


Figure 3-11.—Capacitance-type filter and waveforms.

capacitor acts like a short circuit across the rectifier while the capacitor is charging. Because of this high peaked load on the rectifier tubes, the capacitor input filter is seldom used with gas tubes in high-current installations.

Inductance Filter

Because an inductor resists changes in the magnitude of the current flowing through it, an inductor can be placed in series with the rectifier output to help prevent abrupt changes in the magnitude of the current. An inductance-type filter together with its input and output waveforms is shown in figure 3-12. The input waveforms from a half-wave and a full-wave rectifier are shown respectively in figures 3-12, A and B. Figure 3-12, C, shows the inductance-type filter, and figure 3-12, D and E, shows the output current for the half-wave and full-wave input respectively. When no inductor is used in series with R, the output current waveforms are indicated by dotted lines. The solid lines indicate the output-current waveforms when an inductor is used. The use of an inductor prevents the current from building up or dying down quickly. If the inductance is made large enough, the current becomes nearly constant.

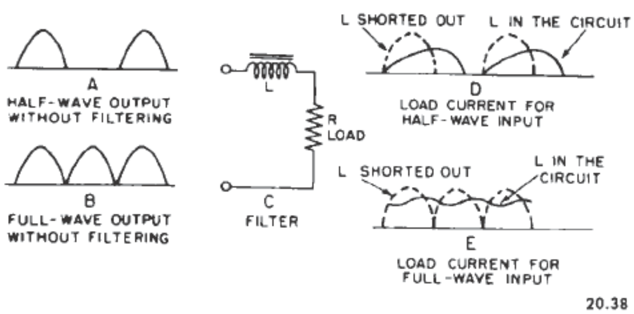


Figure 3-12.—Inductance-type filter and waveforms.

The inductance prevents the current from ever reaching the peak value that is reached without the inductance. Consequently, the output voltage never reaches the peak value of the applied sine wave. Thus, a rectifier whose output is filtered by an inductor cannot produce as high a voltage as can one whose output is filtered by a capacitor. However, this disadvantage is partly compensated because the inductance filter permits a larger current drain without a serious change in output voltage.

Pi-Section Filter

The ripple voltage present in a rectifier output cannot be eliminated adequately in many cases by either the simple capacitor or inductor filter. Filters that are much more effective can be made if both inductors and capacitors are used. The function of the capacitor is to store and release energy, while the inductors simultaneously tend to prevent change in the magnitude of the current. The result of these two actions is to remove the ripple from the rectifier output and to produce a voltage having a nearly constant magnitude.

Figure 3-13 shows a circuit diagram of an inductance-capacitance filter used primarily with receiver power supplies and other low-current power supplies.

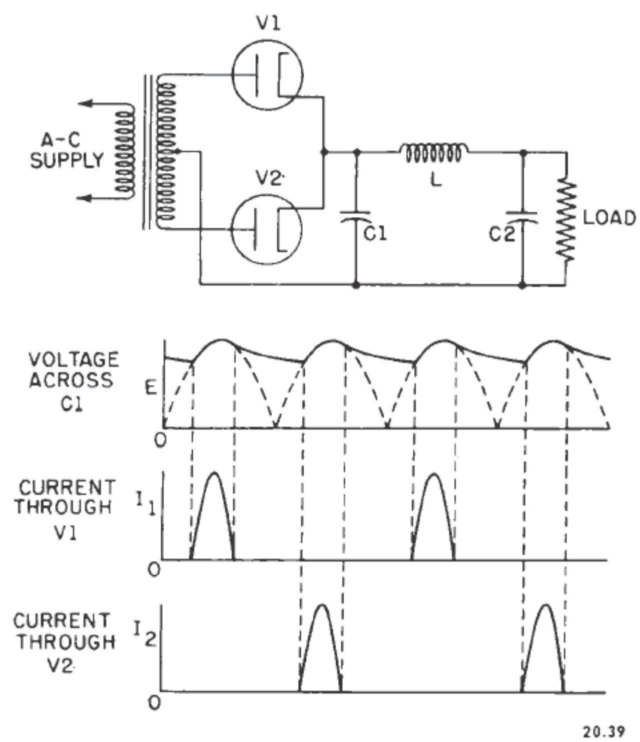


Figure 3-13.—Waveforms of current and voltage in rectifier with pi-section filter.

This type of filter is given the name PI-SECTION because the configuration of the schematic diagram resembles the Greek letter, π.

It is also called a CAPACITOR INPUT FILTER. With this type of filter the output waveform closely approximates that of pure direct current. The first (input) capacitor acts to bypass the greatest portion of the ripple component to ground. In all filters the major portion of the filtering action is accomplished in this first component. The series choke in the pi-section filter serves to maintain the current at a nearly constant level during the charging and discharging cycles of the input capacitor.

At the bottom of figure 3-13 are shown the waveforms of current through V1 and V2 and the voltage across C1. The final capacitor, C2, acts to bypass residual fluctuations existing after filtering by the input capacitor and inductor. The current flow through the rectifier tubes is a series of sharp peaked pulses, because the input capacitor acts like a short circuit across the rectifier while the capacitor is charging. Because of this high peaked load on the rectifier tubes, the pi-section filter is used only in low-current installations such as radio receivers.

L-Section Filter

A second type of filter used primarily in high-current applications is the L-section filter, so named because of its resemblance to an inverted "L." A schematic diagram of this type of filter is shown in figure 3-14. The components perform the same functions as in the pi-section filter except that the inductor, or choke, input reduces the voltage output of the filter. This filter is also called a CHOKE INPUT FILTER. The input choke allows a continuous flow of current from the rectifier tubes rather than the pulsating current flow demanded by the capacitor input filter. The L-section filter is seldom used with half-wave rectifiers because there is no device to maintain current flow between half cycles.

Because of the uniform flow of current, the L-section filter has applications in most high-power circuits, and is used with mercury-vapor rectifiers. It has the additional advantage of better voltage regulation. The inductive reactance of the choke reduces the ripple voltage without reducing the d-c output voltage.

Because the L-section filter is widely used in naval electronic equipment, the factors affecting the design of these filters are important

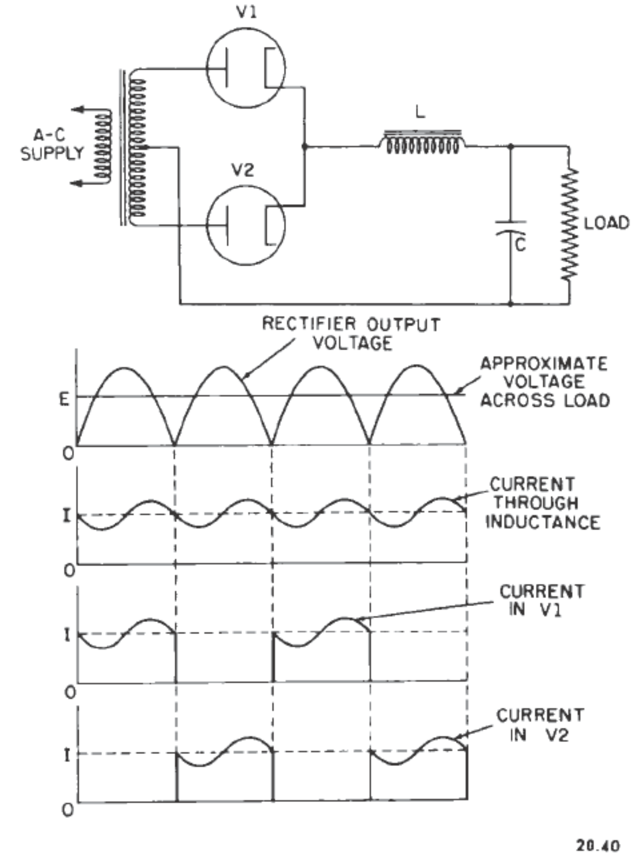


Figure 3-14.—Current and voltage waveforms in full-wave rectifier with L-section filter.

to the electronics technician. As has been mentioned, this type of filter is generally used where high currents are required. In this respect, its advantage lies in the fact that it allows each rectifier tube to operate at a relatively constant level of current flow during its half-cycle of operation. This type of operation allows a rectifier to supply the maximum current to the load that it is capable of delivering. A disadvantage of the L-section filter is that instead of delivering a voltage equal to the peak value of the transformer secondary, it supplies a voltage equal to the average of the a-c voltage delivered to the rectifier.

Two L-section filters are sometimes used in series to obtain a higher degree of filtering action.

VOLTAGE REGULATION

The output voltage developed by any source of power tends to decrease when current is drawn from the source. The amount of change in the output voltage is usually expressed by a quantity called the PERCENTAGE OF VOLTAGE REGULATION.

The formula for the percentage of regulation is

$$\text{percentage of voltage regulation} = \frac{E_{NL} - E_{FL}}{E_{FL}} \times 100$$

where E_{NL} is the no-load voltage, and E_{FL} is the output voltage when full-load current is flowing out of the supply. For example, assume the no-load voltage of a certain power supply to be 300 volts and the voltage at the output terminals to be 250 volts when the load resistance is applied and load current begins to flow. Substituting these values in this formula gives

$$\text{percentage of voltage regulation} = \frac{300-250}{250} \times 100 = 20 \text{ percent}$$

The difference between the no-load voltage and the full-load voltage is caused by the flow of load current through the internal resistance of the power supply. The IR drop caused by the load current within the supply circuit is subtracted from the voltage available for the load resistance at the output terminals. A perfect power supply would have zero internal impedance and the percentage of regulation would be zero. Such a supply would provide the same voltage under full-load that it develops with no-load current flowing. In general, the lower the percentage of regulation, the better is the power supply in furnishing direct voltage and direct current for electronic equipment.

The regulation of the choke-input filter circuit is superior to that of the capacitor-input circuit as long as current is flowing in the filter choke. In this condition the output voltage changes very little when the load current changes in value. If, however, the load current should become zero, the choke coil can no longer prevent the first capacitor in the filter from charging to a value equal to the peak value of the applied voltage.

If the load current is a low value, or if it varies between a low value and zero, the regulation of the circuit is poorer than when larger currents are being drawn by the load. In order to improve the regulation of the choke-input filter, a resistor is often connected across the output terminals so that at least a minimum current will always flow through the choke.

VOLTAGE REGULATORS

Most electronic gear used in the Navy can operate satisfactorily with a certain amount of variation in the supply voltage without suffering severe operational deficiency. However, some circuits are very critical and even a slight deviation from the normal supply voltage will cause unsatisfactory operation. These circuits require the use of some type of voltage-regulating device.

A voltage-regulating device may be inserted in the circuit at one of two points—either between the rectifier and its load or at the power source that supplies electrical energy to the rectifier. The regulators that are used within a power supply are generally electronic and those affecting the power source itself (for example, the generator) are generally mechanical. Mechanical voltage regulators are treated in basic electricity texts.

Fundamental Voltage Regulator

The regulator that is used to stabilize the output voltage of a rectifier usually takes the form of a variable resistance in series with the output. This variable resistance and the load resistance form a voltage divider. The variable element is controlled so that the voltage across the load is held constant.

Figure 3-15 shows a simple circuit that demonstrates this principle. The variable resistor, R, and the resistance of the load comprise a voltage divider that is connected across the rectifier output terminals. All the load current passes through R and causes a voltage drop across it. If the rectifier output voltages rises, the voltage across the load rises in proportion. To counteract this rise, the resistance of R is increased (manually) so that a greater proportion of the available voltage appears across R. The voltage across the load therefore is held constant if the resistance of R is increased sufficiently to neutralize the increase of the rectifier output. If the resistance of the load increases, a greater fraction of the available voltage appears across the load. Therefore, the resistance of R must be increased in order to hold the voltage across the load constant.

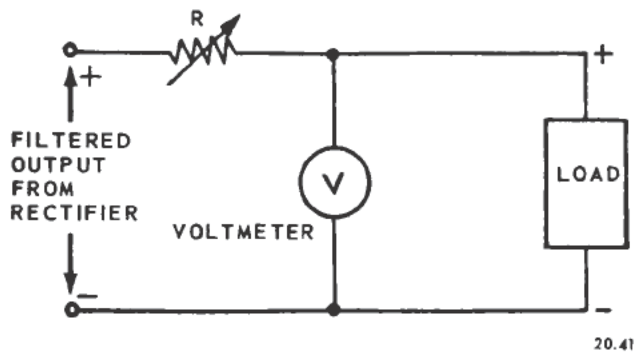


Figure 3-15.—Fundamental voltage regulator.

In the system shown in figure 3-15, the resistor, R, must be varied manually in order to keep the voltmeter reading constant. If the voltmeter reading increases, R must be increased; if the voltmeter reading decreases, R must be decreased. This same type of action must take place in all of the voltage regulators that are to be discussed, but the action is automatic. The more complicated circuits that follow are more desirable than the simpler circuits because they are more accurate and can respond more quickly.

All the voltage regulators discussed in this chapter are essentially voltage dividers. The variable voltage drop may be supplied in many ways, but the action of most of the circuits may be explained in terms of the fundamental circuit shown in figure 3-15.

Amperite Voltage Regulator

A regulator tube that consists of an iron wire enclosed in a hydrogen-filled envelope is called an AMPERITE TUBE or BALLAST TUBE.

An amperite regulator circuit is shown in figure 3-16. The resistance of the iron wire in the ballast tube varies as the current through it changes. If the output voltage tends to increase, more current flows through the ballast tube. The resistance of the tube then increases and more of the voltage drop takes place across the tube. Therefore, the voltage across the load remains nearly constant.

The amperite regulator does not regulate the voltage if the load changes. If the load increases, more current is drawn from the power supply and the load voltage falls. In addition, the greater current drawn causes the resistance of the amperite to increase, and the load voltage is made even lower by this additional drop.

Although the ballast tube may be used to compensate for line voltage variations it is generally inserted in series with several additional elements through which it is desired to maintain a constant current. In such applications, the resistance of the ballast tube changes to counteract the effect of changing voltage across the circuit.

Glow-Tube Voltage Regulator

In a glow-discharge tube, such as the neon glow tube, the voltage across the tube remains constant over a fairly wide range of current through the tube. This property exists because the degree of ionization of the gas in the tube varies with the amount of current that the tube conducts. When a large current is passed, the gas is very highly ionized and the internal impedance of the tube is low. When a small current is passed, the gas is lightly ionized and the internal impedance of the tube is high. Over the operating range of the tube, the product of the current through the tube and the internal impedance of the tube is practically constant.

A simple glow-tube regulator is shown in figure 3-17, A. The load current and the current that flows in the neon glow tube both pass through the series resistor, R. If the supply voltage drops, the voltage across the neon tube tends to drop. Therefore, the gas in the neon tube deionizes slightly and less current passes through the tube. The current through R is decreased by the amount of the current decrease in the tube. Because the current through R is less, the voltage drop across R is less. If the resistor is of the proper value relative to the load and to the glow tube that is used, the voltage across the load is held fairly constant. In any case, the value of R must not be so large that the neon tube fails to ionize.

Glow tubes are designed to operate at various useful values of voltage. These values are usually indicated in the tube-type number.

When a regulated voltage in excess of the maximum rating of one glow tube is required, two or more tubes may be connected in series, as in figure 3-17, B. This arrangement permits several regulated voltages with small current drains to be obtained from a single rectifier power supply.

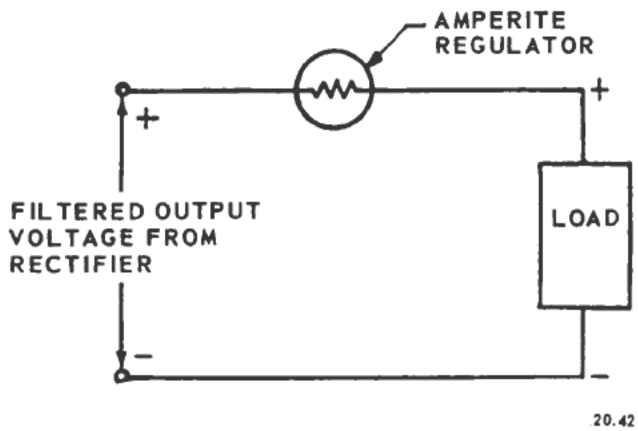


Figure 3-16.—Amperite regulator circuit.

Crystal Diode Voltage Regulator

A specially-manufactured crystal (semiconductor) diode similar to the diode described in chapter 2 of this training course will under certain conditions exhibit a practically constant voltage with wide changes in current (see figure 2-11, B). The crystal diode contains a PN junction. With small reverse bias across the PN junction the barrier potential is increased. Only a small leakage current will flow because the current carriers in both P and N sides of the junction are attracted away from the junction. This action leaves a space charge depletion layer adjacent to the junction.

Increasing the reverse voltage across the junction increases the velocity of the minority carriers in this region. Some of these carriers collide with covalent bond electrons releasing them as carriers. This action has a cumulative effect called avalanche ionization. It comprises a rapidly rising reverse current that, unless checked by a series limiting resistor, may destroy the semiconductor. The reverse voltage at which avalanche effect occurs is called the reverse breakdown voltage and is abbreviated BV_R.

The symbol B is used in the crystal diode CR (fig. 3-17, C) to indicate operation in the reverse breakdown voltage region. When operated in this way the voltage across the load is held constant because the load is in parallel with the crystal diode, and the reverse breakdown voltage of the diode is approximately constant for wide changes in diode current. Semiconductor diodes operated in this region as voltage regulators are often called Zener diodes because their action was first described as a "field emission" effect by Zener.

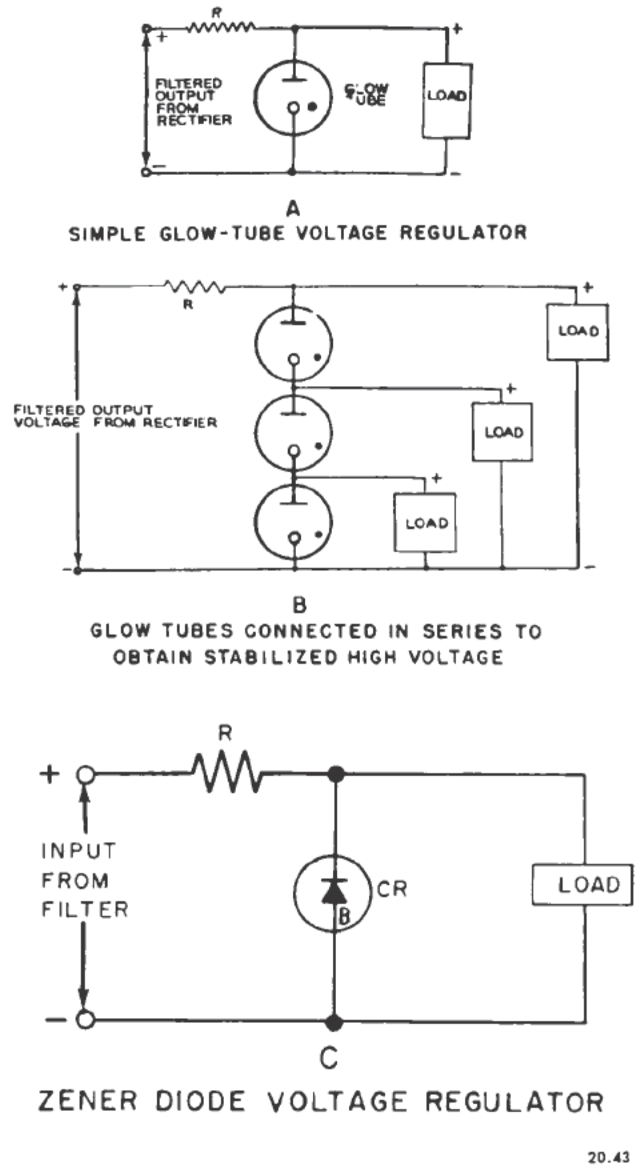


Figure 3-17.—Glow-tube regulators.

If the supply voltage decreases (fig. 3-17, C) the reverse voltage across the semiconductor diode CR will tend to decrease. Thus the speed of the current carriers in the crystal will decrease and the number of collisions with valence electrons will decrease with the result that the reverse current through the crystal will decrease. The current through R is decreased by the amount of the current decrease through crystal CR. Because the current through R is decreased, the voltage drop across R is proportionately less. If R has the right value relative to the load and to CR the voltage across the load will remain fairly constant. However, R must not be so large that the avalanche effect will not take place in the crystal.

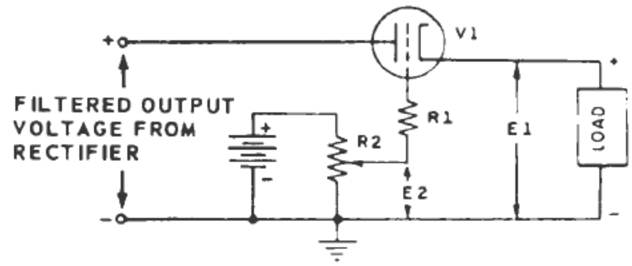
Zener diodes are designed to operate at various voltages. When a regulated voltage in excess of the rating of one Zener diode is required, two or more diodes may be connected in series in the same manner as for glow tube voltage regulators (fig. 3-17, B). As previously stated, this arrangement permits several regulated voltages with small current drain to be obtained from a single rectifier power supply.

Electron-Tube Voltage Regulator

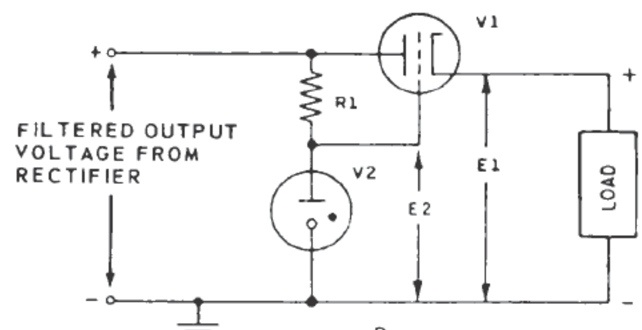
An electron tube may be considered as a variable resistance. When the tube is passing a direct current, this resistance is simply the plate-to-cathode voltage divided by the current through the tube and is called the d-c plate resistance, r_p. For a given plate voltage, the value of r_p depends upon the tube current, and the tube current depends upon the grid bias.

The variable resistor, R, of figure 3-15 can be replaced by an electron tube (fig. 3-18, A), because the electron tube has a variable resistance. The effective resistance of V1 is established initially by the bias on the tube. Assume that the voltage across the load is at the desired value. Then the cathode is positive with respect to ground by some voltage, E_1. The grid can be made positive relative to ground by a voltage, E_2, that is less than E_1. The potentiometer, R2, is adjusted until the bias (grid-to-cathode voltage), which is E_1-E_2, is sufficient to allow V1 to pass a current equal to the desired load current. With this bias, the resistance of V1 is established at the proper value to reduce the rectifier output voltage to the desired load voltage.

If the rectifier output voltage increases, the voltage at the cathode of V1 tends to increase. As E_1 increases, the negative bias on the tube increases and the effective plate resistance of the tube becomes greater. Consequently, the voltage drop across V1 increases with the rise in input voltage. If the circuit is properly designed, the increased voltage drop across V1 is approximately equal to the increase in voltage at the input to the regulator. Thus the load voltage remains essentially constant.



A
ELECTRON-TUBE VOLTAGE REGULATOR
EMPLOYING A BATTERY FOR THE
FIXED BIAS



B
ELECTRON-TUBE VOLTAGE REGULATOR
EMPLOYING A GLOW TUBE FOR THE
FIXED BIAS

20.44

Figure 3-18.—Electron-tube voltage regulator.

The resistor, R1, is used to limit the grid current. This is necessary in this particular circuit because the battery is not disconnected when the power is turned off. However, the battery can be eliminated from the circuit by the use of a glow tube, V2, in figure 3-18, B, to supply a fixed bias for the grid of the tube. The action of the circuit is the same as the action of the circuit in figure 3-18, A.

The output voltage of the simple voltage regulators shown in figure 3-18 cannot remain absolutely constant. As the rectifier output voltage increases, the voltages on the cathode of V1 must rise slightly if the regulator is to function.

The voltage regulators shown in figure 3-18 compensate not only for changes in the output voltage from the rectifier, but also for changes in the load. For example, in figure 3-18, B, if the load resistance decreases, the load current will increase. The load voltage will tend to fall because of the increased drop across V1. The decrease in load voltage is accompanied by a decrease in bias voltage on V1. The bias voltage on V1 is equal to E_1-E_2. Thus, the effective resistance of V1 is reduced at the same time the load current is increased. The IR drop across V1 increases only a slight amount because R decreases about as much as I increases. Therefore, the tendency for the load voltage to drop when the load is increased is checked by the decrease in effective resistance of the series triode.

Improved Voltage Regulator

A very stable voltage regulator (more stable than those shown in fig. 3-18) can be designed by taking advantage of the high amplification possible with a pentode. A voltage regulator employing this type of tube is shown in figure 3-19. It produces an output that is independent of fluctuations in the a-c supply and changes in load over a wide range.

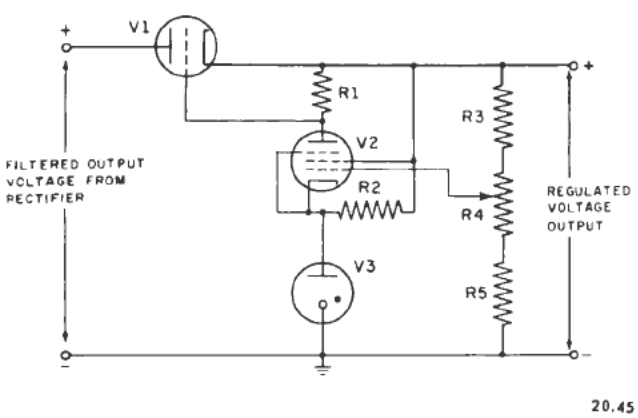


Figure 3-19.—Improved voltage regulator.

The output voltage of this regulator is developed across bleeder resistors R3, R4, and R5 in parallel with the resistance of the load. These resistors make up the resistance of one part of the total voltage divider. The other resistance, through which all of the load current must flow, is the cathode-to-plate effective resistance of V1. The other elements of the circuit are used to control the resistance of V1 and thereby to maintain a constant voltage across the load.

The potential of the cathode of V2 is held at a constant positive value with respect to ground by the glow tube, V3. In other words, current flowing from ground through V3 causes an IR drop across V3 that maintains the cathode of V2 positive with respect to ground. The grid potential of V2 is a voltage selected by potentiometer R4. This potentiometer is set so that the grid-to-ground voltage is less positive than the cathode-to-ground voltage by an amount (the bias) that causes V2 to pass a certain plate current. In other words, the IR drop between the moving contact on R4 and ground is less than the IR drop across V3 by an amount that is equal to the bias on V2. The plate current of V2 flows through R1 and causes a drop across it. The magnitude of the voltage across R1 is the bias on tube V1. Therefore, the adjustment of potentiometer R4 establishes the normal resistance of V1. This adjustment is used to set the value of load voltage that the regulator is to maintain.

If the load voltage tends to rise, whether from a decrease in load current or from an increase in the input voltage, the voltage between the moving contact on R4 and ground will increase. The difference in this voltage and the fixed voltage across V3 decreases. These two voltages are in opposition, and the voltage between the moving contact on R4 and ground is less than the fixed voltage across V3. Thus, the grid bias of V2 decreases, and the plate current of V2 increases through R1. The increase in voltage across R1 increases the effective resistance of V1. If the load voltage tends to rise because of an increase in input voltage, this increase is accompanied by an increase in voltage across V1 and the rise in load voltage is checked. If the rise in load voltage is caused by a decrease in load current, this rise is checked because the IR drop across V1 remains constant, because the decrease in I is accompanied by an equal increase in R.

A pentode is used for V2 because of the high amplification possible with this type of tube. The use of such a tube makes the output voltage much more constant because small variations of load voltage are amplified sufficiently to cause operation of the circuit.

The anode of the glow tube, V3, is connected to the cathode of V2 and to the plus terminal of the regulated voltage output through resistor R2. It is necessary to connect the glow tube to the B^+ in this way in order to cause the gas in this tube to ionize when the power supply is first turned on.

All the load current must pass through V1, therefore, this tube must be capable of passing a large current. In some regulators a single tube does not have sufficient capacity to pass the required current. In such cases, several identical tubes may be connected in parallel.

The type of regulator shown in figure 3-19 is used very widely to stabilize the output voltage of rectifier power supplies. Because of its excellent sensitivity to small changes in input voltage, this regulator is very effective in removing ripple from the output of rectifier power supplies. The regulator, then, serves also to filter the output of a rectifier, although the conventional filter systems usually are used in connection with a regulator.

Figure 3-20 is a complete rectifier-power-supply circuit, showing the power transformer, the full-wave rectifier tube, the filter circuit, the voltage regulator, and the voltage divider. This figure summarizes much of the foregoing power-supply discussion.

VOLTAGE DIVIDERS AND BLEEDERS

A resistor is frequently placed across the output terminals of a rectifier power supply to bleed off the charge on the filter capacitor when the rectifier is turned off, or to apply a fixed load to the filter and thus improve the voltage regulation of the power supply. In the latter case, the resistor is designed to draw at least 10 percent of the full-load current in order for the change in power supply current to be less for a given change in load and thus reduce the magnitude of the variation in output voltage. In both cases the resistor is called a BLEEDER RESISTOR. If leads are connected to the resistor at various points to provide a variety of voltages that are less than the total output voltage, the resistor is called a VOLTAGE DIVIDER.

Voltage Divider Circuits

A resistor that is used as a load resistor may also serve as a divider and as a bleeder. A simple voltage divider composed of three similar resistors in series is shown in figure 3-21. As long as no load is drawn from any terminal except the top, or line, terminal, the voltages across the resistors will divide in proportion to the resistance of each as shown in figure 3-21.

It is common practice to ground one side of voltage dividers. Therefore, ground potential is normally used as a reference for measurement of voltages as indicated at point D in figure 3-22, A. If a rectifier and its filter are connected so that no parts of the system are grounded, the divider can be grounded at any point without affecting the operation of the rectifier, provided the insulation of all parts is sufficient to withstand the voltage involved. Thus in figure 3-22, B, point C is grounded, and point D becomes negative with respect to ground. Such a circuit may be used to furnish both plate and bias voltages from the same power supply. In figure 3-22, C, point A is grounded, and all voltages along the divider are negative with respect to ground. Note, however, that point A will always be positive with respect to points, B, C, and D as long as the electron flow is maintained from D to A as shown in figure 3-22, C.

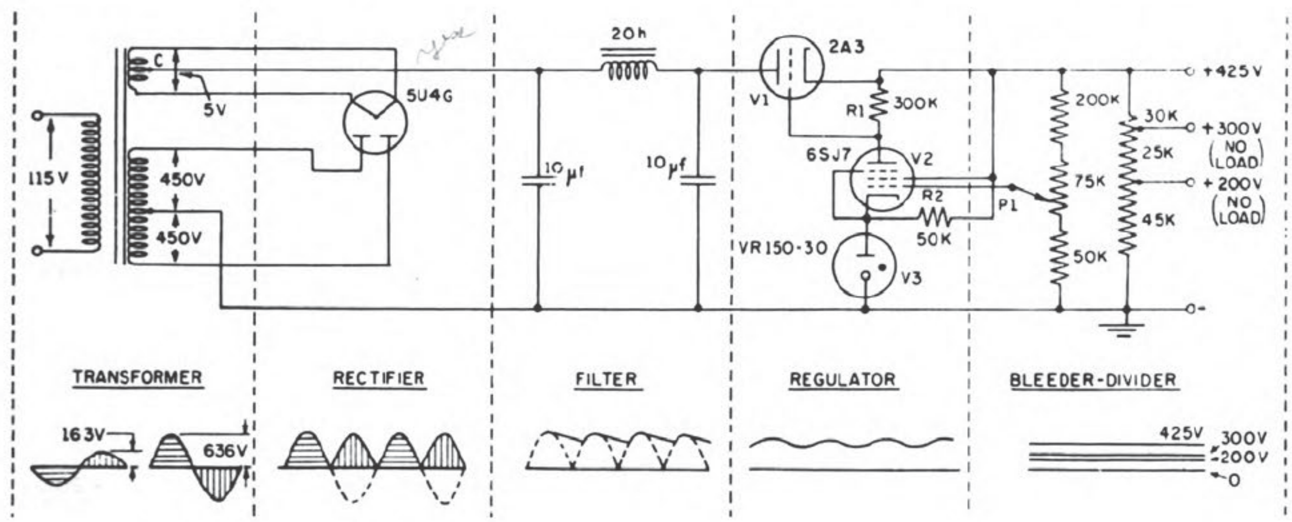
In the voltage divider circuits (fig. 3-22) it has been assumed that no load was attached to the divider except across terminals A and D, and that voltages could be measured without drawing appreciable current. As soon as a load is connected across the divider at any intermediate terminals, however, the voltage division shown in figure 3-22 is no longer correct. The resistance of the attached load forms a parallel circuit with that part of the divider across which it is placed, and therefore there is a change in the resistance of that part of the divider in relation to the total resistance between terminals A and D.

For example, in figure 3-23, a load of 150,000 ohms (150 k-ohms) is placed across BD, and a load of 50 k-ohms is placed across CD. The resistance between C and D is first determined by Ohm's law for parallel resistance—

$$R_{CD} = \frac{50 \times 50}{50 + 50} \quad 25 \text{ k-ohms}$$

To this resistance is added the series resistance (50 k-ohms) between B and C, making a total of 75 k-ohms. The resistance across BD is then found by the parallel resistance rule. Thus,

$$R_{BD} = \frac{75 \times 150}{75 + 150} \quad 50 \text{ k-ohms}$$



Low voltage is stepped up by the transformer from 115 volts to 900 volts. Center tap provides a dividing point so that 450 volts are applied to each section of the 5U4G rectifier. The ends of the transformer alternately become positive and negative.

Center tap C on heater winding is used to force plate current to divide equally in each filament lead. If there is no center tap, a voltage divider of two equal 50 ohm resistors may be put across the secondary to produce the same effect.

Alternately positive and negative voltage is applied to the plates of the rectifier.

The two plates conduct alternately as each plate is made positive in turn by the secondary of the transformer. Pulses of current flow from the filament line to each plate in turn. The plates alternately become positive and negative with the applied a. c., but the filament line will show a one-directional flow.

Capacitors charge when the rectifier conducts, and they discharge through the bleeder resistor when the tube is not conducting.

Choke builds up a magnetic field when the tube draws current. The field collapses as current decreases, tending to keep a constant current flowing in in the same direction through the bleeder resistor and the load.

Capacitor input (illustrated) gives higher voltage output with low current loads.

Choke input gives steadier output with less ripple under load conditions.

If the load draws more current or if the a-c input voltage falls, the terminal voltage of the power supply falls.

Resistor R1, tube V2, and gas-tube V3 are in series across the rectifier terminals. V3 holds the cathode of V2 at a constant positive potential with respect to ground, and setting of P1 determines bias on V2. A fall in terminal voltage causes more negative bias on V2, less current through V2, hence, less current through R1. Less IR drop across R1 causes less negative bias on V1. V1, then acts as a lower value resistor, and terminal voltage decrease is checked.

As a bleeder, the resistor is for safety to discharge the capacitors when power is removed.

As a load resistor, it acts as a stabilizer to protect the voltage regulator at no load, and to improve the regulation.

A voltage divider meets the requirements of a load resistor and a bleeder, but in addition has taps placed at intervals for voltage at less than the maximum.

It is usually grounded at the lower end but may be grounded at any higher point to get a negative output.

20.46

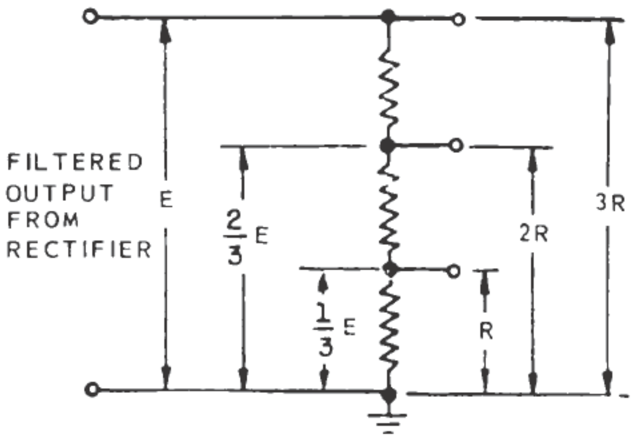
Figure 3-20.—Complete rectifier-filter-regulator-divider circuit.

The total resistance between A and D before the main load is applied is the resistance between B and D plus the resistance between A and B, or 50+50 = 100 k-ohms.

The total current, I, drawn through the divider and its two loads is then the total voltage divided by the resistance between A and D. Therefore,

$$I = \frac{300}{100{,}000} \quad 0.003 \text{ ampere}$$

A current of 0.003 ampere flowing through R1 produces an IR drop of 50,000x0.003 = 150 volts.



20.47

Figure 3-21.—Simple voltage divider.

Therefore, when load 1 and load 2 are connected as shown, the voltage across R1 increases from 100 to 150 volts. The voltage across load 1 is

$$300-150 = 150 \text{ volts.}$$

The current through load 1 is

$$\frac{150}{150{,}000} = 0.001 \text{ ampere,}$$

and the current through R2 is

$$0.003-0.001 = 0.002 \text{ ampere.}$$

The 0.002 ampere flowing through R2 produces an IR drop of

$$50{,}000 \times 0.002 = 100 \text{ volts.}$$

Thus, the voltage remaining to be applied across CD is

$$150-100 = 50 \text{ volts.}$$

The current in load 2 is

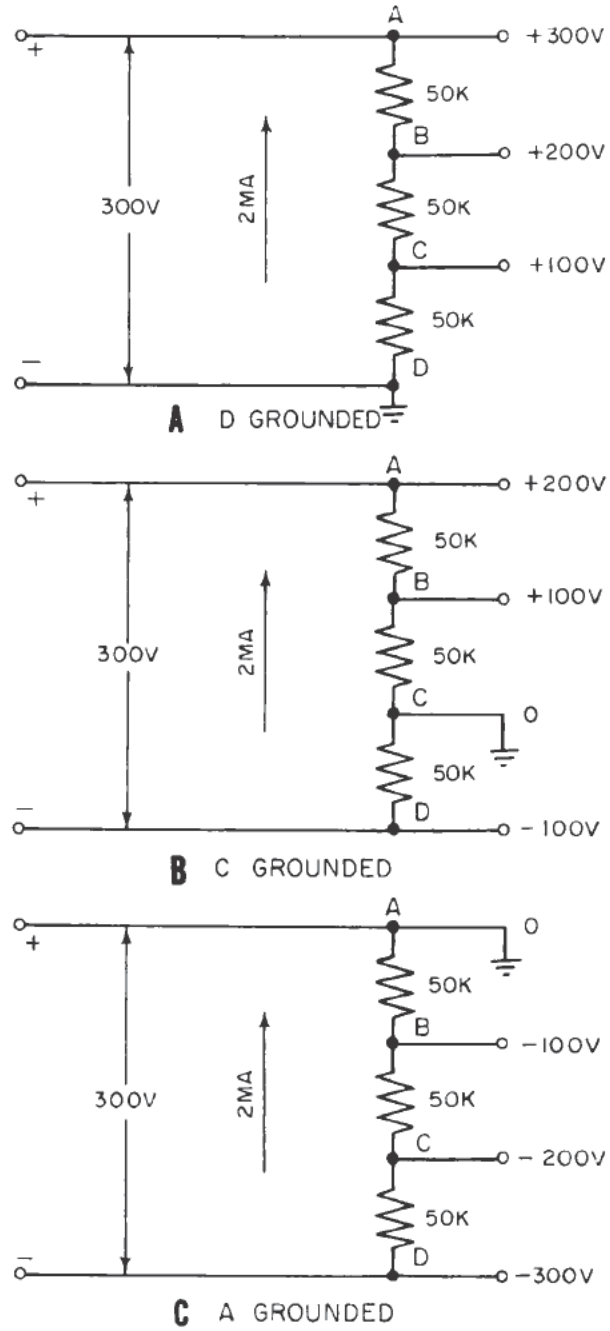
$$\frac{50}{50{,}000} = 0.001 \text{ ampere.}$$

leaving 0.001 ampere to flow through R3. As a check, the IR drop across R3 can be found as

$$50{,}000 \times 0.001 = 50 \text{ volts.}$$

Because this voltage is the same as that previously determined across CD, the value of current is correct.

Instead of a voltage of 200 volts between point B and ground, and 100 volts between point C and ground, as in figure 3-22, A, the voltage now is 150 volts at B and 50 volts at C, when the load values are as indicated in figure 3-23.



20.48

Figure 3-22.—Effect of moving the ground point on a voltage divider.

Other load values will give correspondingly different values of voltage at B and C. Thus, the voltage appearing across the intermediate terminals of a voltage divider divides proportionately to the values of the divider resistors only as long as no appreciable load current is drawn from these terminals. Under loaded

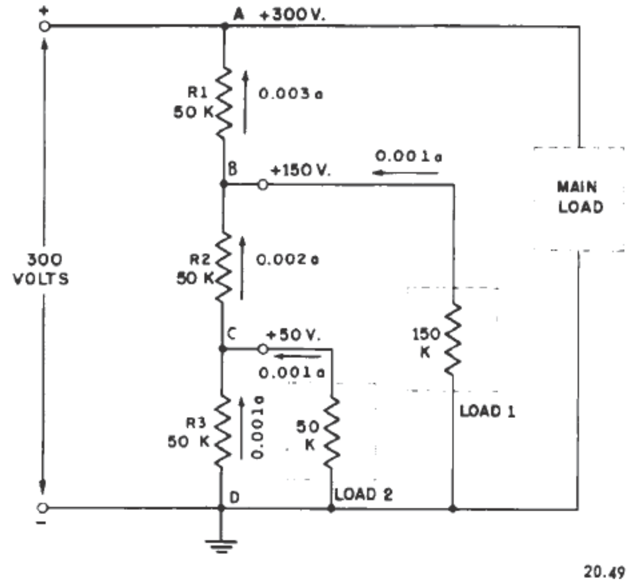


Figure 3-23.—Effects of loads on voltage division.

conditions the voltages at these terminals will have various values, depending upon the resistance of the loads. A voltage divider must therefore be designed for the particular load conditions under which it is to operate.

VOLTAGE-MULTIPLYING CIRCUITS

HALF-WAVE DOUBLER

If a rectifier unit (a metallic type in this case) and a capacitor are connected in series across a source of alternating current, as shown in figure 3-24, the voltage will be doubled. The arrow in the doubler circuits in this chapter acts as the anode, and electron flow is in the opposite direction of this arrow. When the input voltage is as indicated in figure 3-24, A, the capacitor charges to the peak value of the line voltage. On the next half cycle the condition shown in figure 3-24, B, results. The full peak voltage across the capacitor is retained, but because of the polarity reversal of the source, the rectifier no longer conducts and as a result the voltage across the capacitor adds to that of the source. Therefore, the total voltage across the rectifier has a peak value twice that of the source. Thus, the output voltage varies between zero and twice the peak input voltage during each cycle.

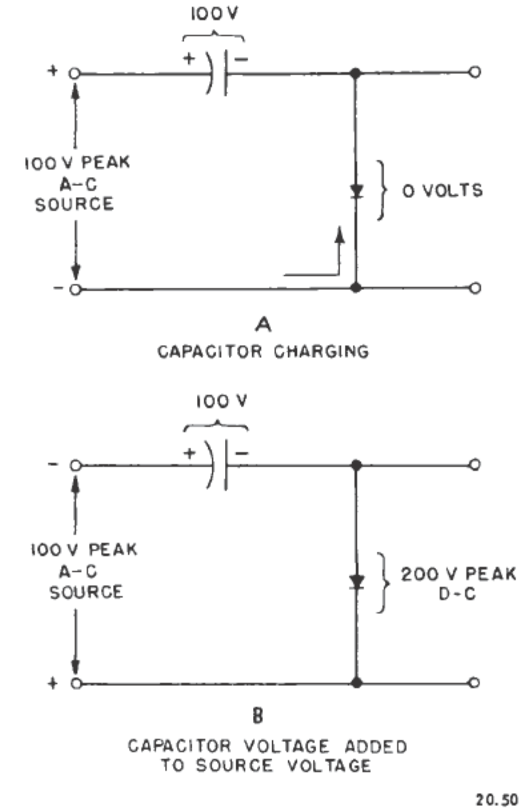


Figure 3-24.—Analysis of voltage doubler.

The output voltage can be maintained over the entire cycle if a second rectifier unit and capacitor are added, as shown in figure 3-25. The second capacitor charges twice the peak input voltage when rectifier D2 conducts and holds its charge during the time D2 is nonconducting.

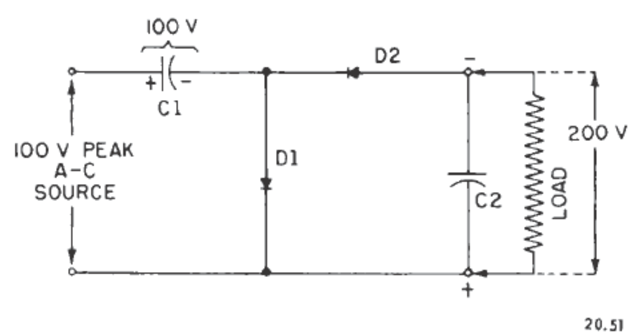


Figure 3-25.—Half-wave voltage doubler.

Capacitor C2 cannot, however, maintain the full output voltage over the complete cycle if there is any appreciable load. This limitation results from the fact that when rectifier D2 is nonconducting no current is drawn from the input circuit. Thus, C2 supplies the load current during discharge, and the output voltage falls proportionately. Because current is drawn from the source for only one-half of a cycle, this circuit is called a HALF-WAVE VOLTAGE DOUBLER. The half-wave characteristic and the size of capacitor C2 limit the use of this circuit to applications requiring only a small output current.

FULL WAVE DOUBLER

A doubler that operates as a full-wave rectifier is shown in figure 3-26. Actually, such a connection is equivalent to connecting a pair of half-wave doublers across the source so that the direction of their conducting paths is opposite. As a full-wave rectifier, this circuit draws current from the voltage source during both halves of the input cycle. For one half cycle, C1 charges to the source voltage through rectifier D1, and for the next half cycle C2 charges to the source voltage through rectifier D2. The voltages impressed across C1 and C2 will therefore combine in series across the load to give the polarity and current flow indicated in the figure.

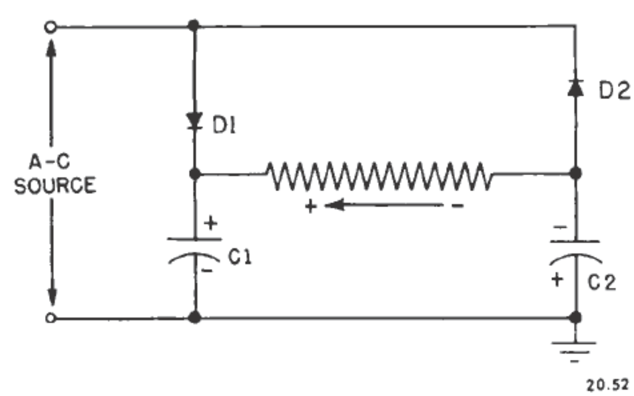


Figure 3-26.—Full-wave voltage doubler.

VOLTAGE MULTIPLIERS

The process of increasing the voltage can be performed at higher levels of multiplication such as tripling and quadrupling. Theoretically, the voltage could be multipled an infinite number of times by this process. Practical considerations, however, generally limit the multiplication to four or five times. Figure 3-27 shows a schematic diagram in which an output voltage equal to seven times the peak of the input voltage is developed. The number of sections in this type of circuit could be extended to give high output voltages, but with each additional section the voltage regulation is adversely affected. The circuit acts as a half-wave multiplier and thus large capacitors must be employed to maintain current flow during alternate half cycles.

The operation of this circuit may be explained as follows: When the upper terminal is negative, electrons flow through all of the diodes, charging C1, C3, C5, and C7 to the peak voltage of the source. When the upper terminal is positive, the charges stored on these capacitors act in series with the input voltage to charge C2, C4, and C6 to twice the value of the input voltage. This reasoning may be continued through the first seven half cycles, and at this time the voltage across C7 will have been built up to seven times the input voltage.

Contrary to what may be expected upon first inspection of this circuit, the inverse peak voltage across any one of the rectifiers does not increase with the number of stages. Actually the peak inverse voltage across each rectifier, regardless of its position in the circuit, is twice the peak value of the input voltage.

Voltage multipliers are not widely used in naval electronic equipment. When high voltages are required, designers prefer the more conservative and dependable transformer and high-vacuum rectifier type of supply. When multipliers are utilized, the disk-type rectifier is well suited to the application because it requires no filament supply. If electron-tube rectifiers were used, the simplicity of the multiplier would be defeated by the necessity for providing a separate filament transformer for each stage of multiplication.

GRID-BIAS VOLTAGES

GRID BIAS FROM B SUPPLY

In modern electronic equipment the grid-bias voltage is frequently derived from the plate supply voltage. Three methods are shown in figure 3-28.

The methods shown in figure 3-28,A and B, are commonly used in low-power amplifiers. The prime consideration is that the bypass

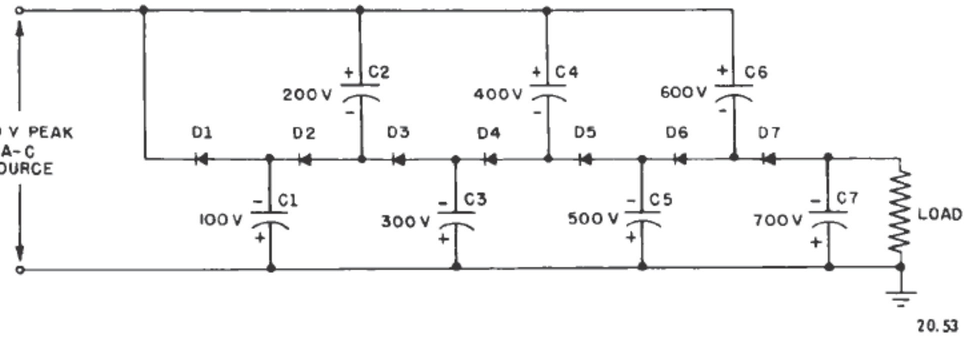


Figure 3-27.—Voltage multiplier.

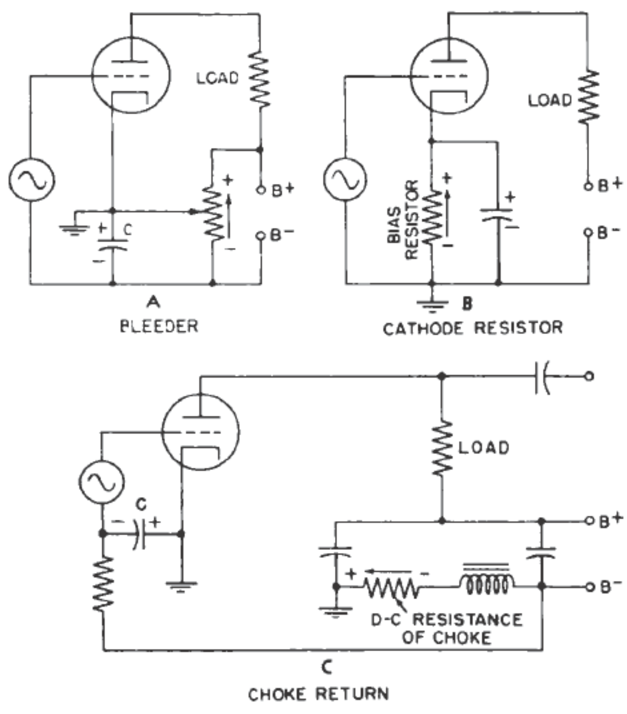


Figure 3-28.—Methods of obtaining grid-bias voltage from the plate power supply.

capacitor be large enough to maintain a steady voltage across the resistor at the lowest operating frequency of the amplifier. In other words, the impedance of the capacitor should be very low, compared with the resistance of the bias resistor, at the lowest frequency to be passed by the amplifier. These arrangements will obviously reduce the effective plate-to-cathode voltage by the absolute value of the bias voltage.

A brief consideration of current flow through the bleeder and through the tube in figure 3-28,A, may be of value in understanding how the bias is established. Bleeder current flows from B- to B+, and the no-signal (bias) voltage between grid and cathode is determined by the position of the movable contact. The grid is thus negative with respect to the cathode. The d-c component of plate current also flows through the lower portion of the bleeder. Capacitor C offers less opposition to the signal component of the plate current than does the lower portion of the bleeder. Therefore, the signal component flows from B- through C to the cathode, and back to B+. Because the signal component does not flow through the bleeder the bias voltage is maintained at a steady value.

In figure 3-28,B, the bias is developed by the flow of the d-c component of the plate current through the bias resistor. The a-c component is passed around the bias resistor by the bypass capacitor. Thus, a steady negative bias is established between grid and cathode.

The method shown in figure 3-28,C, is used frequently in power amplifiers. This arrangement makes use of the resistance in the power-supply filter choke to supply the voltage drop necessary for grid bias. In this circuit the grid is connected to B-, and the signal component of the plate current flows from B-, through the low reactance of C, to the cathode, and back to B +via the tube and load resistor.

Batteries are sometimes used for bias supplies when absolute stability of the bias voltage is necessary. Under these conditions the battery supplies no grid current and its effective life is its "shelf life." A more popular type of bias battery now in use is composed of tiny individual dry cells capable of being clamped together in a special type of holder resembling a fuse clamp. Any number of these dry cells can be placed in series to obtain bias voltages in multiples of 1.5 volts. These cells have extremely long shelf life.

Very large, high-power amplifiers frequently have separate bias supplies. They may be d-c generators or rectifier-filter systems. If d-c generators are used a filter must be placed in the output to eliminate any commutator ripple that may be present. If a filter were not used, the ripple, no matter how small, would be amplified by the tube and cause distortion of the output signal.

FIXED BIAS VOLTAGE SUPPLY

Some electronic equipments require a fixed bias. Transmitters, especially those that handle considerable amounts of power, need a fixed negative bias in order to protect the tubes and circuits, should other systems of bias fail. In the absence of a fixed bias a failure of the normal bias would cause a large increase in current through the tube and the associated equipment, and damage to the equipment might result.

One of the circuits that may be used to produce a fixed bias voltage is shown in figure 3-29. A conventional full-wave rectifier together with its filter circuit is shown at the top of the figure. This rectifier furnishes the high d-c voltages for the plates and screen grids of the tubes. The fixed bias supply is shown enclosed in the dotted lines. The lower half of the high-voltage secondary is used, together with a half-wave rectifier diode, V2, and filters R1, R2, C1, and C2, to produce the fixed bias across R3.

When the lower end of the high-voltage secondary (fig. 3-29) is negative, electrons flow from the cathode to the plate of V2 and down through R3 to ground and thus back to the center tap on the secondary; thus the output voltage is negative with respect to ground in contrast with the B-supply circuit.

The tap on R3 permits the bias to be adjusted to the correct value. Bias resistor R3 generally has a relatively low value of resistance to minimize voltage variations due to grid current in class-C amplifiers, overdriven class-A amplifiers, or class-AB_2 amplifiers. If grid current should flow in one of the stages connected to the fixed bias supply, the voltage across the bias resistor (in the absence of capacitor C3) would increase and the bias would be increased accordingly.

When less bias voltage is needed than is developed in the circuit shown in figure 3-29, the 6-volt filament transformer supply circuit shown in figure 3-30,A, may be used instead.

The operation of the circuit in figure 3-30,A, may be explained by the use of the simplified circuit shown in figure 3-30,B. Assume that at

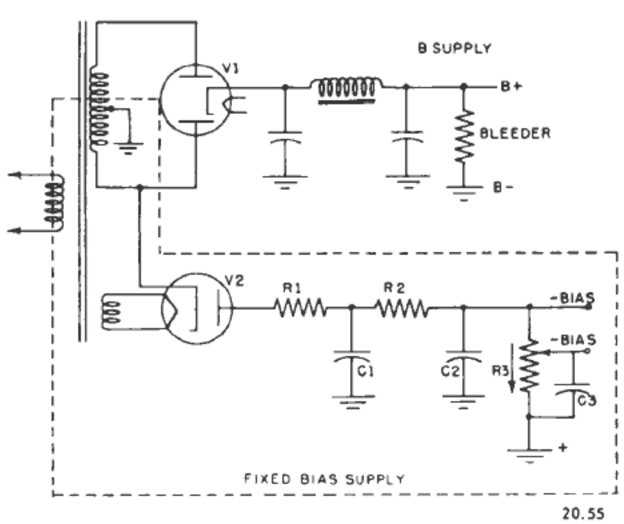


Figure 3-29.—Fixed bias supply circuit.

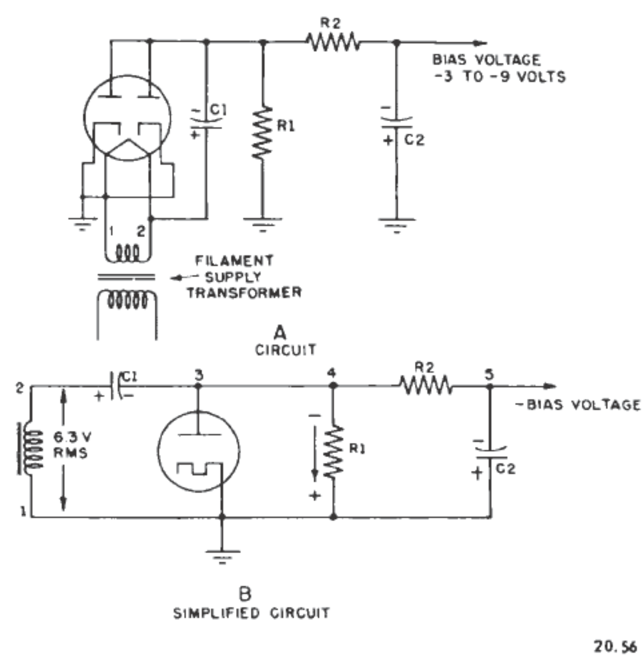


Figure 3-30.—Shunt-fed rectifier bias supply.

a given instant point 1 is negative and point 2 is positive. Electrons flow from point 1 through the tube and charge the right-hand plate of C1 negative. During this conducting time R1 is short-circuited by the tube. On the other half cycle when point 1 is positive and point 2 is negative, electrons cannot flow through the tube. Instead, C1, which has charged to approximately the full peak voltage of the supply, now partially discharges through R1, thus developing the bias voltage. The time constant of C1R1 is relatively long, so that only a small amount of charge leaks off C1 during the nonconducting period. R2 and C2 serve as filters, so that the voltage appearing between point 5 and ground is relatively free of ripple.

ELECTROMECHANICAL SYSTEMS

DYNAMOTORS

The basic electrical power source in many aircraft is a 24-volt storage battery and an engine-driven generator. The generator charges the battery and supplies engine ignition, aircraft lights, and other electrical loads. In addition, aircraft communications equipment generally has incorporated within it another rotating machine called a DYNAMOTOR.

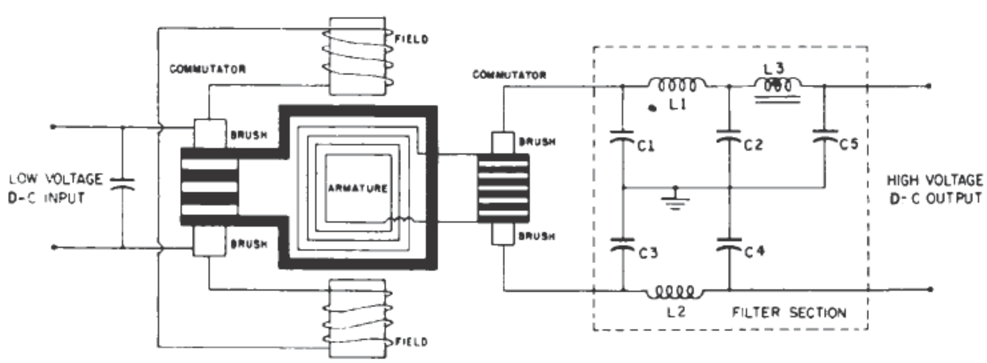
The dynamotor performs the dual functions of motor and generator, changing the relatively low voltage of the 24-volt power supply into a much higher value for the plates and screens of electron tubes. The dynamotor usually employs two windings on a single armature. The two windings occupy the same set of slots and terminate in two or more separate commutators. The armature rotates in a single field frame with a conventional field winding to provide the excitation for both motor and generator. The motor armature winding is connected to the 24-volt power source and develops driving torque which rotates the armature as a motor. The generator winding is connected to the plates and screens of the electron tubes of the associated equipment and generates the relatively high voltage for these loads.

A functional diagram of a typical dynamotor is illustrated in figure 3-31. The heavy lines represent the motor circuit. Relatively high current from the low-voltage source flows through the motor winding of the armature. The field is also energized from the low-voltage source. The interaction of the large current in the motor armature conductors with the field causes the armature to rotate.

The high-voltage winding, represented by the finer lines between the field, is wound in the same armature and rotates with the motor winding. When turning, the high-voltage winding cuts the lines of force of the common field and generates a voltage which is developed across the brushes on the high-voltage commutator. The greater the number of turns in the high-voltage armature winding, the greater will be the voltage output.

Because the armature and field windings in the diagram of figure 3-31 are connected in parallel this is called a SHUNT-WOUND (shunt-connected) MOTOR. The desirable characteristic of this type of motor is that the speed remains fairly constant with changes in the load placed upon it.

The high current required by the motor necessitates a correspondingly large size in the motor components such as the commutator, brushes, and armature wire as compared to



20.57

Figure 3-31.—Functional diagram of a dynamotor.

those components of the generator. The motor commutator is larger in diameter but has fewer segments than that of the generator. Because more turns in the generator armature winding produce a higher output voltage, there are a greater number of turns in that winding and the wire size is correspondingly reduced.

Filters are placed at the high-voltage output terminals to filter out high-frequency currents produced by sparking between the brushes and the rotating commutator segments, thereby eliminating any possible interference that the sparking may cause. The filter consists of a combination of chokes and capacitors such as shown in the typical filter section at the right of figure 3-31. The purpose of the chokes, L1 and L2, is to present a high impedance to the high-frequency currents so that they will not be present in the output. The low impedance of the capacitors, C1, C2, C3, and C4, bypasses high-frequency currents to ground.

Additional audio filtering is provided to eliminate commutator ripple which compares with the ripple found in the output of conventional a-c rectifiers. This audio filtering consists usually of a series inductor of comparatively high value and a shunt capacitor. It is represented in the figure by the iron-core choke, L3, and capacitor C5. The capacitor across the low-voltage input leads reduces sparking between the brushes and commutator at the input end of the dynamotor.

VIBRATORS

The vibrator is another type of voltage supply used to obtain a high a-c or d-c voltage from a comparatively low d-c source. It has certain advantages over the dynamotor type of power supply. For example, the vibrator is lighter and less expensive than the dynamotor; it is also more efficient. However, the vibrator can be used only when a limited amount of high-voltage current is needed. Also, its life is relatively short, and it tends to produce radio interference (hash). It is extensively used in the "power packs" of lightweight mobile equipment. Actually, neither the dynamotor nor the vibrator are power supplies as such, being only the means by which low-voltage direct current is converted to high-voltage direct current.

A simple vibrator power supply is shown in figure 3-32. It is nothing more than a simple interrupter, similar in many respects to a buzzer or doorbell. Pulsating direct current is used to energize the primary winding of a transformer which in turn induces an a-c voltage in the secondary. The turns ratio of the transformer windings are chosen to give the desired output voltage.

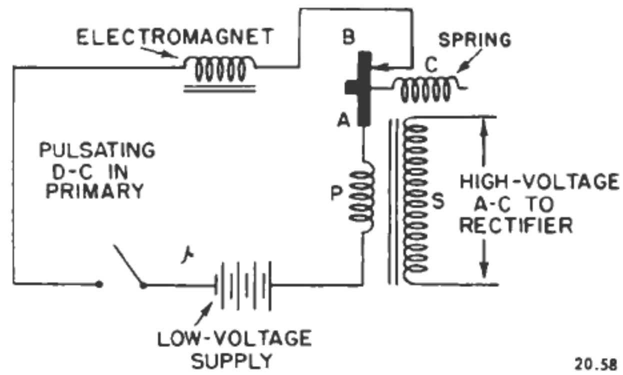


Figure 3-32.—Basic vibrator power-supply system.

When the switch is closed in the circuit of figure 3-32, current flows from the battery through the electromagnet, and then through contact B, armature A, primary winding P, and back to the battery. In passing through the electromagnet the current sets up a magnetic field that attracts the armature. As the armature moves it breaks the circuit at contact B. As soon as the circuit is broken, the electromagnet no longer attracts the armature, thus allowing spring C to pull it back to the starting position. At the starting position, contact B again closes the circuit and the process is repeated. In this way a pulsating direct current that induces a high voltage in the secondary winding flows through the primary of the transformer.

The output voltage of the secondary is applied to a conventional rectifier and filter network, which converts the alternating current back into direct current, but at a higher voltage.

Two typical vibrator power-supply systems are shown in figure 3-33. In figure 3-33,A, is shown the nonsynchronous type of vibrator power supply. This type of power supply requires a separate rectifier and filter. Either cold-cathode or high-vacuum rectifiers are used with nonsynchronous vibrators.

The operation of this type of vibrator is much the same as that of the basic vibrator shown in figure 3-32. When the battery switch is closed, current flows through the lower half of the primary, the electromagnet, and back to the battery, producing an expanding magnetic field. As the armature is drawn down, the electromagnet is temporarily short-circuited, and loses its magnetism. The armature is released and makes contact with terminal 2. Current now flows through the upper half of the primary and back to the battery. At the same time this is occurring, the field previously established by the current in the lower half winding is collapsing. The effect of the simultaneous expansion of one field and collapsing of the other field is to increase the magnitude of the induced voltage in the secondary.

The synchronous vibrator, shown in figure 3-33,B, does not require a rectifier tube. This type of vibrator is called a synchronous vibrator because two additional contact connected to the ends of the secondary winding are so synchronized with the contact in the primary circuit that rectification takes place.

Contacts 3 and 4 perform the function of rectification. Insofar as the primary is concerned, the action is similar to that of the nonsynchronous vibrator. If at a given instant point 5 is negative and the armature is touching points 2 and 3, electrons will flow from point 5, to point 3, to ground; and return to the center tap on the secondary by way of the load. A half cycle later, point 6 is negative and the armature is touching points 4 and 1. Electrons then flow from point 6 through 4 to ground, and return to the center tap via the load. Thus, current always flows through the load in the same direction—that is, from ground to the center tap.

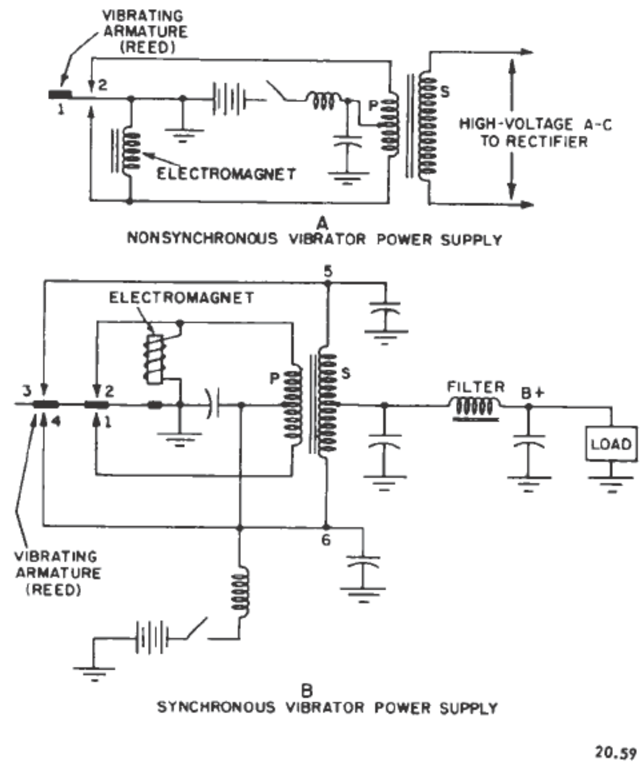


Figure 3-33.—Typical vibrator power-supply systems.

INVERTERS

In many naval aircraft the primary source of power that is available for use in the electronic equipment is the 24-volt d-c supply that is used to run dynamotors and to supply low-voltage filament and heater power of low-power transmitters and receivers. In addition, 115-volt alternating current is often derived from the low-voltage direct-current source by the use of a device called an INVERTER. The inverter is a rotating type of machine that takes an input of 24 volts d-c and changes it into 115 volts a-c at about 800 cps.

The 115-volt power is then used to provide the input for rectifier power supplies in the high-power electronic equipment, such as radar units.

Electromechanical systems that change alternating current to direct current are commonly called CONVERTERS, and the electromechanical systems that change direct current to alternating current are commonly called INVERTERS.

Because alternating current is used for some of the instruments on aircraft, and also for control equipment, radar, radio, fluorescent lighting, and the like, some means of producing alternating current from the aircraft direct-current system is necessary. Although lighter equipment may be used in certain instances, the motor-generator (d-c motor, a-c generator) set is more flexible and in general more satisfactory from the point of view of frequency and voltage control.

Common frequencies of aircraft inverters are 400 and 800 cycles per second. The higher frequencies permit smaller inductive components in the equipment and thus the weight may be kept to a minimum.

Both d-c motors and a-c generators are treated in texts on basic electricity and hence will not be treated here.

CHAPTER 4

TUNED CIRCUITS

INTRODUCTION

A tuned electrical circuit has capacitance, inductance, and resistance in series or in parallel. When the circuit is energized at a particular frequency, known as the resonant frequency, an interchange of energy occurs between the coil and capacitor. This interchange of energy tends to build up in amplitude far above the amount delivered by the energizing source. This is known as a resonant condition. At resonance, the inductor stores energy during the half cycle that the capacitor discharges and returns the energy during the next half cycle to recharge the capacitor. Because the circuit resistance acting in series with the inductor and capacitor is maintained low, large amounts of energy may be exchanged at the resonant frequency, with minimum loss of energy in the circuit. The small energy loss incurred in the circuit is replenished from the source feeding the circuit.

At resonance, the time needed to charge the capacitor must be equal to the time needed to discharge the coil, otherwise the charge and discharge will be out of step and cancellation will result.

In a series-tuned circuit, the impedance in ohms across the terminals of such a circuit is

$$Z = \sqrt{R^2 + (X_L - X_C)^2}.$$

Since the inductive reactance, X_L, and the capacitive reactance, X_C, are equal and opposite in polarity at the resonant frequency, they balance each other and the actual total reactance is reduced to zero. Because $X_L - X_C = 0$, the total impedance of the circuit at the resonant frequency is equal to the resistance of the circuit, or $Z=R$; at resonance, the maximum amount of current will flow in the circuit.

The most important characteristic of a SERIES-TUNED CIRCUIT is that AT RESONANCE THE CIRCUIT IMPEDANCE IS A MINIMUM.

At frequencies below resonance the series circuit acts like a capacitor plus a resistor, and accordingly the circuit current is reduced. At frequencies above resonance, the circuit acts like an inductor plus a resistor, and the current is likewise reduced.

At resonance, the voltage across the capacitor and the voltage across the inductor are equal in magnitude. Because they are 180° out of phase with each other their vector sum is zero and the source voltage appears across the circuit resistance. Because the series resistance is low, the source voltage is small in relation to the voltage across the coil and capacitor. Under these circumstances the voltage appearing across either the inductor or the capacitor may be much higher than the input voltage.

A parallel-tuned circuit consists of a combination of inductance and capacitance connected in parallel. A small value of resistance (representing the inherent resistance of the two components) may be considered as acting in series with the inductance and capacitance (fig. 4-1, A).

The inductive branch containing losses may be represented as two parallel branches, one having an equivalent inductance and no loss, and the other having an equivalent shunt resistance with the same loss as in the original series resistance. The total current of the original inductive circuit with loss may be regarded as being made up of two components 90° out of phase. One component (I_{en}) called the energy component flows in the equivalent shunt resistance and is in phase with the applied voltage.

The other component (I_{nonen}) lags the applied voltage by 90° and is called the nonenergy component of current. It flows in the pure inductive branch with no losses. The nonenergy component of current is so named because it is a component of reactive power which is returned to the circuit twice every cycle of applied voltage.

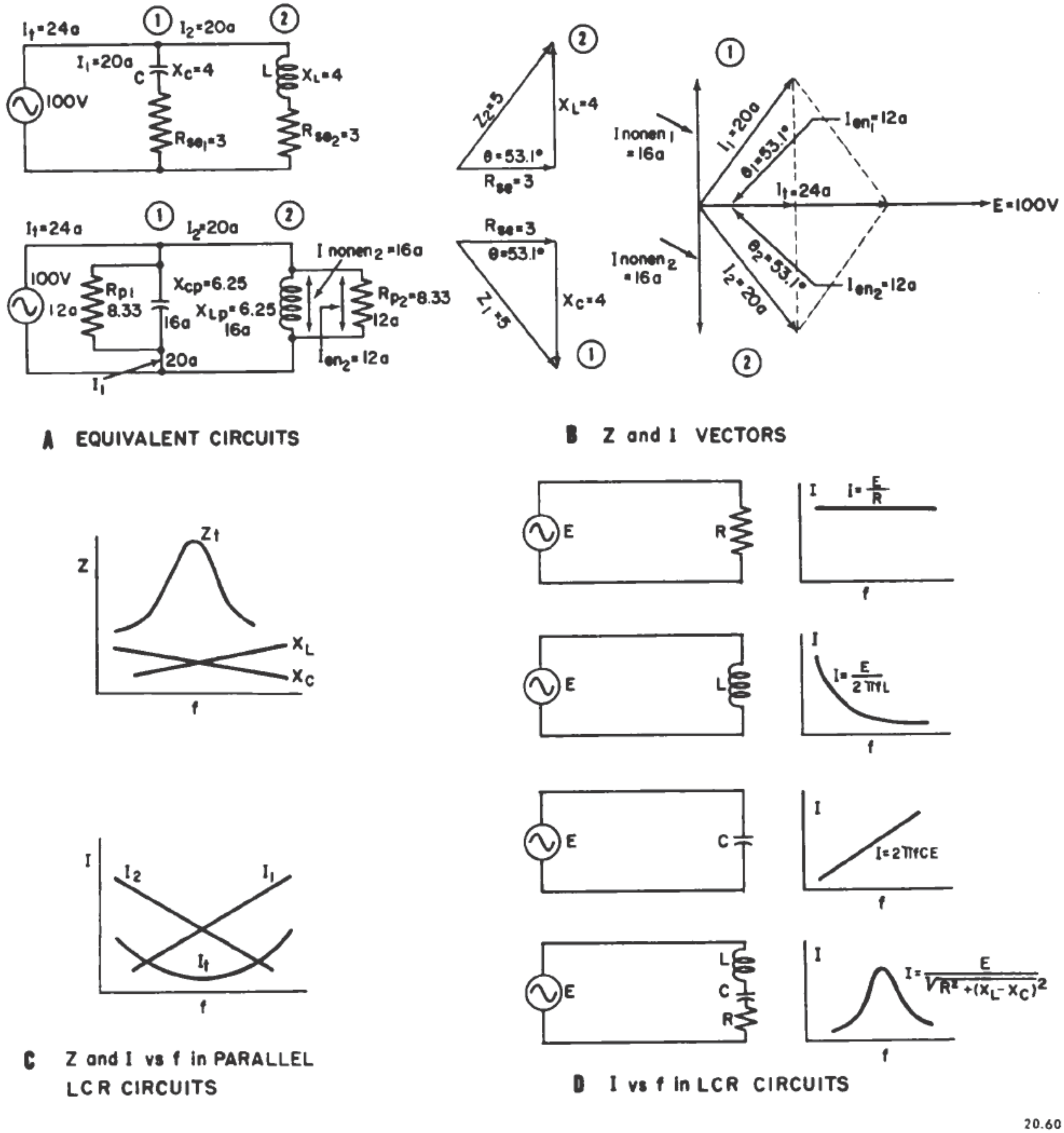


Figure 4-1.—Parallel-tuned circuit.

The vector sum of the two components of current is equal to the total current of the original inductive branch containing losses and lags the applied voltage by angle θ where $\cos \theta$ is the power factor of the original inductive branch.

In the example of figure 4-1, A, the current in branch 2 of the original inductive branch with losses is

$$I_2 = \frac{E_2}{Z_2} = \frac{100}{5} = 20 \text{ amperes}$$

The energy component of current in R_{p2} of the equivalent parallel circuit of branch 2 is

$$I_{2en} = 20\ (\cos 53.1^\circ = 0.6) = 12 \text{ amperes.}$$

The nonenergy component of current in X_{LP2} of the equivalent parallel circuit of branch 2 is

$$I_{2nonen} = 20\ (\sin 53.1^\circ = 0.8) = 16 \text{ amperes}$$

The reactance X_{LP2} of the equivalent parallel reactance of branch 2 is

$$X_{LP2} = \frac{E}{I_{2nonen}} = \frac{100}{16} = 6.25 \text{ ohms}$$

The resistance R_{P2} of the equivalent parallel resistance of branch 2 is

$$R_{P2} = \frac{E}{I_{2en}} = \frac{100}{12} = 8.33 \text{ ohms}$$

The combined impedance of the equivalent parallel circuit of branch 2 comprising X_{LP2} and R_{P2} is equal to the impedance of the original branch 2 comprising X_{LS} and R_S, because they both have the same total current and voltage.

The same approach is made to the current flowing in the capacitive branch containing some resistance, except that in the latter case the nonenergy component of the current leads instead of lags the applied voltage by 90°.

At resonance, the nonenergy component of the lagging current flowing through the inductive branch is exactly equal to the nonenergy component of the leading current flowing through the capacitive branch. Under these circumstances the current flowing in either of the branches may be much greater than the line, or input, current. Because these currents are 180° out of phase, they neutralize. The small in-phase current that now flows in the line is due to the inherent resistance of the circuit. Therefore, at resonance the impedance offered by the parallel circuit is a MAXIMUM and is purely RESISTIVE.

The variation of Z and I with frequency is illustrated in figure 4-1, C.

The most important characteristic of a PARALLEL-TUNED CIRCUIT is that AT RESONANCE THE CIRCUIT IMPEDANCE IS A MAXIMUM.

At frequencies below resonance, the current through the inductive branch is large and lags the applied voltage by approximately 90°. At the same time a smaller component of current, which leads the applied voltage by approximately 90°, flows through the capacitive branch. Above resonance, the opposite conditions prevail.

If the ratio of the reactance to the inherent resistance of both the inductor and the capacitor is high, the circuit will be in resonance at the same frequency irrespective of whether the components are connected in series or in parallel.

In radio receivers, tuned circuits are used both for the selection of the desired frequency and for the rejection of undesired frequencies. The relative ability of a receiver to select the desired signal while rejecting all others is called SELECTIVITY.

In radio transmitters the entire process of radio-frequency power generation and amplification depends on the proper functioning of tuned circuits.

Test instruments such as signal generators, oscillators, and frequency meters, as well as other electronic devices such as television transmitters and receivers, and radar and sonar equipments, employ many tuned circuits.

The variation of current with frequency in R, L, C, and RCL series circuits is summarized in figure 4-1, D. In constant potential circuits where the resistance is constant, the effective current is also constant and independent of frequency. If the circuit contains only inductance (L), the effective current decreases as the frequency increases; in a capacitive circuit the current increases directly with the frequency. In a series circuit containing C, L, and R, current rises to a maximum at a resonance and decreases both above and below resonance.

Before tuned circuits can be analyzed, an elementary understanding of vectors and vector algebra is required. Accordingly, a brief review of vectors as they are expressed both in the rectangular and the polar form follows.

EXPRESSING VECTORS ALGEBRAICALLY

Many common physical quantities such as temperature, the speed of a moving object, or the displacement of a ship can be expressed as a certain number of units. These units define only the magnitude and give no indications of the direction in which the quantity acts. Such quantities are called SCALAR quantities. If both the magnitude and the direction in which the quantity acts are indicated, it is called a VECTOR quantity and may be represented by a vector. For example, a vector representing the speed and heading of a ship having a speed of 10 knots and a heading of 45° (northeast) is a straight line extending upward and to the right. The length of the line is proportional to the speed of 10 knots. The angle that the line makes with the vertical (north at the top) is 45° clockwise from the vertical.

Electrical vectors are commonly used to represent a-c currents and voltages and their phase relations. The length of the vector represents the magnitude of the quantity involved, and the direction of the vector, with respect to a reference axis, represents the lapse in time between the positive maximum values of current and voltage.

Impedance triangles, the sides of which represent vector quantities, are also used to represent the resistance and reactance components of a-c circuits. These are right triangles having a base equal to the resistive component, an altitude equal to the reactive component, and a hypotenuse equal to the combined impedance. The angle between the combined impedance and the resistive component (hypotenuse and base) is equal to the phase angle between the voltage across the impedance and the current flowing through it.

In this chapter it is necessary to determine circuit impedances by the addition, subtraction, multiplication, and division of vector quantities. When it is inconvenient to express the quantity by simple algebra, a system of complex notation is used.

OPERATOR J

In calculations in electronics it is often necessary to perform operations involving the square root of a negative number—for example, $\sqrt{-9}$, $\sqrt{-5}$, and $\sqrt{-x}$. Because no number when multiplied by itself will produce a negative result, the roots of numbers such as the foregoing cannot be extracted. It therefore becomes necessary to introduce a new type of notation to indicate the square root of a negative number. These numbers are called IMAGINARY NUMBERS to distinguish them from the so-called REAL NUMBERS. Actually, the numbers that we call imaginary in the mathematical sense are real in the physical sense. The term is merely one of convenience, as will be pointed out in the succeeding paragraphs.

In algebra, the foregoing quantities are treated as $\sqrt{-1}\sqrt{9}$, or $\sqrt{-1x3}$; $\sqrt{-1}\sqrt{5}$; and $\sqrt{-1}\sqrt{x}$. The term, $\sqrt{-1}$, is expressed as i (for imaginary) in mathematics books, but when working with electrical circuits it is convenient to use the term j (called the J OPERATOR), because i is used to indicate the instantaneous value of the circuit current.

GRAPHICAL REPRESENTATION

In order to present a quantity graphically, some system of coordinates must be employed. Quantities involving the j operator may be conveniently expressed by the use of RECTANGULAR COORDINATES, as shown in figure 4-2. In order to specify a vector in terms of its X and Y components, some means must be employed to distinguish between X-axis and Y-axis projections. Because the +Y-axis projection is +90° from the +Y-axis projection, a convenient operator is one that will, when applied to a vector, rotate it without altering the magnitude of the vector. Let +j be such an operator that produces 90° COUNTERCLOCKWISE rotation of any vector to which it is applied as a multiplying factor. Also, let -j be such an operator that produces 90° CLOCKWISE rotation of any vector to which it is applied as a multiplying factor.

Successive applications of the operator +j to a vector will produce successive 90° steps rotation of the vector in the counterclockwise direction without affecting the magnitude of the vector. Likewise, successive applications of the operator -j will produce successive 90° steps of rotation in the clockwise direction. This rotation is shown in table 4-1.

In the four quadrants (upper right, upper left, lower left, and lower right) the signs indicate the direction of the vertical (J) component. The + sign indicates a vertically upward direction from the X axis and the - sign indicates a vertically downward direction from the X axis.

Consider the following example: The number, +4, in figure 4-2,A, indicates that 4 units are measured from the origin along the X axis in the positive direction. A +j operator placed before the 4 indicates that the number is to be rotated 90° counterclockwise and will now be measured along the Y axis in a positive direction. Likewise, a -j operator placed before the 4 indicates

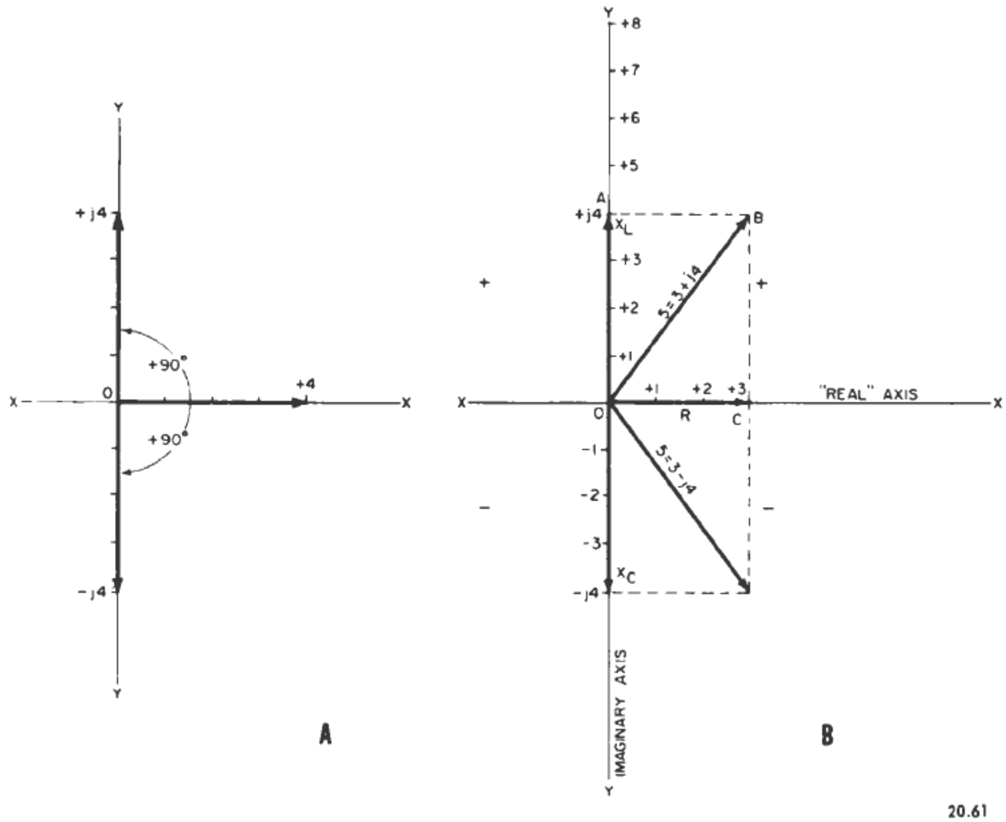


Figure 4-2.—Coordinates showing real and imaginary axes.

Table 4-1.—Relation of Operator J to Vector Rotation.

Operator	Mathematical equivalent	Direction of rotation	Degree of rotation
j	$\sqrt{-1}$	Counter-clockwise	90
j^2	-1	do	180
j^3	$-\sqrt{-1}$	do	270
j^4	1	do	360
$-j$	$-\sqrt{-1}$	Clockwise	-90
$(-j)^2$. . .	-1	do	-180
$(-j)^3$. . .	$\sqrt{-1}$	do	-270
$(-j)^4$. . .	1	do	-360

that the number is to be rotated 90° clockwise, and will now be measured along the Y axis in the negative direction.

It may be recalled that inductive reactance, X_L, is indicated as lying along the Y axis in the positive direction, and capacitive reactance, X_C, is indicated as lying along the Y axis in the negative direction; resistance in each case is measured along the X axis in the positive direction. Therefore, $+j$ has a direct association with X_L in that both are measured in the same direction along the Y axis, and $-j$ similarly has a direct association with X_C.

The function of the j operator may be shown as follows: The expression, 4 ohms, indicates that pure resistance is involved. In order to indicate that the 4 ohms represent capacitive reactance or inductive reactance a special symbol is needed. The use of the j operator gives a clear indication of the type of reactance. For example, if the j operator is not used, the 4 ohms is resistive. If $+j$ is used ($+j4$), the 4 ohms is inductive reactance. If $-j$ is used

(-j4), the 4 ohms is capacitive reactance (fig. 4-2,B).

The so-called COMPLEX NUMBER contains the "real" and the "imaginary" terms connected by a plus or a minus sign. Thus, 3+j4 and 3-j4 are complex numbers. This means that the 3 and the 4 in each instance are to be added vectorially, and the +j and -j indicates the direction of rotation of the vector following it. The real number in these examples is 3 and could be represented by a line drawn three units out from the origin on the positive X (resistance) axis. The imaginary number, + j4, could likewise be represented by a line extended 4 units from the origin on the positive Y, or X_L, axis; and -j4 could be represented by a line extended 4 units from the origin on the negative Y, or X_C, axis. The IMAGINARY, or QUADRATURE, quantities (for example, the X_L and X_C values) are always assumed to be drawn along the Y axis, and the REAL quantities (for example, the R values) are always assumed to be drawn along the X axis.

ADDITION AND SUBTRACTION OF COMPLEX NUMBERS

Values that are at right angles to each other cannot be added or subtracted in the usual sense of the word. Their sum or difference can only be indicated, as is done in the case of binomials (an expression involving two terms). Thus, assume that it is desired to add 3+j4 to 3-j4.

$$\begin{array}{r} 3+j4 \\ \underline{3-j4} \\ 6-\ 0 \end{array}$$

The imaginary term disappears, and only the real term, 6, remains. If 3+j4 is added to 3+j4, the sum is the complex quantity, 6+j8.

One complex expression may also be subtracted from another complex expression in the same manner that binomials are treated. For example, 3-j2 may be subtracted from 3+j4 as

$$\begin{array}{rr} & 3+j4 \\ (-) & \underline{3-j2} \\ & 0+j6 \end{array}$$

The real term disappears, and the result is 6 units measured upward from the origin on the Y axis. If 3-j2 is subtracted from 6+j4, the difference is the complex quantity, 3+j6.

MULTIPLICATION AND DIVISION OF COMPLEX NUMBERS

Complex numbers are multiplied the same way that binomials are multiplied—for example, if 3-j2 is multiplied by 6+j3

$$\begin{array}{r} 3-j2 \\ \underline{6+j3} \\ 18-j12 \\ +j9-j^26 \\ 18-j3-j^26 \end{array}$$

Because $j^2=-1$, the product becomes 18-j3-(-1)6, or 24-j3.

Complex numbers may be divided in the same way that binomials containing a radical in the denominator are divided. The denominator is rationalized (multiplied by its conjugate-a term that is the same as the denominator except that it has the opposite algebraic sign before the j term), and the quotient is expressed as a term having only a real number as the divisor. For example, if 4+j3 is divided by 2-j2,

$$\frac{4+j3}{2-j2} = \frac{(4+j3)(2+j2)}{(2-j2)(2+j2)} = \frac{2+j14}{8}$$

$$= \frac{1+j7}{4} = 0.25+j1.75$$

RECTANGULAR AND POLAR FORMS

Sometimes it is more convenient to employ polar coordinates than rectangular coordinates. In RECTANGULAR FORM the vector is described in terms of the two sides of a right triangle, the hypotenuse of which is the vector. Thus, in figure 4-3, vector OB is described in rectangular form by the complex number 3+j4. In other words, the end of the vector, OB, is 3 units along the + X axis and 4 units along the +Y axis, and its length is 5 units.

The vector, OB, may also be described if its length and the angle, θ, are given. When a vector is described by means of its magnitude and the angle it makes with the reference line it is expressed in the POLAR FORM. In this instance the length is 5 units and the angle, θ, is approximately 53°. The vector, OB, may then be expressed in the polar form as $5\angle+53°$. If the rectangular form is 3-j4, the polar form is $5\angle-53°$.

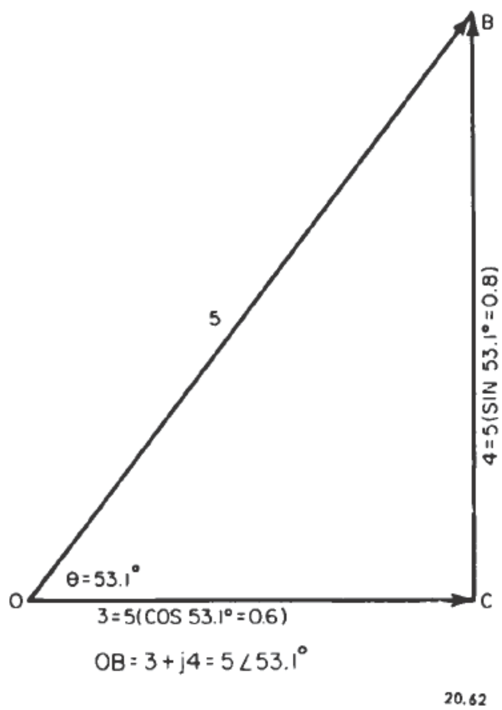


Figure 4-3.—Rectangular and polar forms

The plus sign is shown with positive angles in this chapter in order to emphasize positive angles as contrasted with negative angles. The negative sign preceding the angle indicates clockwise rotation of the vector from the zero position.

CONVERTING FROM ONE FORM TO THE OTHER

Assume that the rectangular form is expressed by the complex number, 3+j4. The angle, θ, and the actual length of the vector, OB, are not given. The length, OB, can be determined by use of the Pythagorean theorem ($OB = \sqrt{3^2+4^2}$), but it is usually simpler to determine first the angle, θ, by finding the angle whose tangent is $\frac{4}{3} = 1.33$. The angle is 53.1° from a table of trigonometric functions. From the same table, sin 53.1° = 0.8. Since $\sin \theta = \frac{BC}{OB}$, it follows that

$$OB = \frac{BC}{\sin 53.1^\circ} = \frac{4}{0.8} = 5;$$

and the vector may be expressed in the polar form as $5 \angle +53.1^\circ$.

If the vector is originally expressed in the polar form as $5 \angle +53.1^\circ$, it may be converted to the rectangular form by the use of cos 53.1° and sin 53.1°. In this instance the vector is 5 units in length and makes an angle of approximately 53.1° with the +X axis. Thus,

$$\sin 53.1^\circ = \frac{BC}{5},$$

or

$$BC = 5 \sin 53.1^\circ = 5 \times 0.8 = 4; \cos 53.1^\circ = \frac{OC}{5},$$

or $$OC = 5 \cos 53.1^\circ = 5 \times 0.6 = 3.$$

Therefore with BC and OC known, the vector may be expressed as the complex number 3+j4 (fig. 4-2,B).

The polar form may be converted to the rectangular form more concisely in the following manner:

$$\begin{aligned} 5 \angle +53^\circ &= 5 \cos 53.1^\circ + j5 \sin 53.1^\circ \\ &= 5 \times 0.6 + j5 \times 0.8 \\ &= 3+j4 \end{aligned}$$

ADDITION AND SUBTRACTION OF POLAR VECTORS

Unless polar vectors are parallel to each other they cannot be added or subtracted algebraically. Therefore, the polar form is converted first to the rectangular form. Then the real components are added algebraically, and likewise, the imaginary components are added algebraically. Finally the result may be converted back to the polar form. Vector summation is indicated by the symbol ⊕.

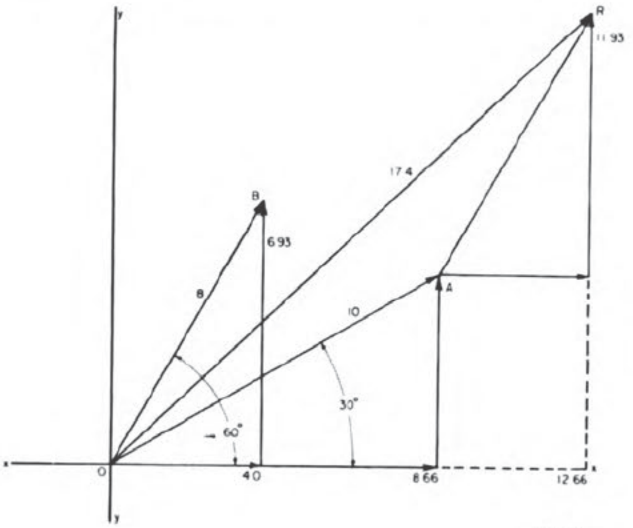
As an example, find the resultant vector, OR (fig. 4-4) of vectors OA and OB when $OA = 10 \angle 30^\circ$ and $OB = 8 \angle 60^\circ$. OR=OA⊕ OB
Converting to rectangular form:

$$\begin{aligned} OA &= 10 \cos 30^\circ + j10 \sin 30^\circ = 8.66 + j5.0 \\ OB &= 8 \cos 60^\circ + j\,8 \sin 60^\circ = 4.0 + j6.93 \end{aligned}$$

Adding, like components, OR=12.66+j11.93.

Converting to polar form, $OR = \sqrt{12.66^2 + 11.93^2}$ $= 17.4$ and $\tan \theta = \frac{11.93}{12.66} = 0.943$

from which $\theta = 43.4^\circ$.



20.63

Figure 4-4.—Addition of vectors.

MULTIPLICATION AND DIVISION OF POLAR VECTORS

The method of multiplying and dividing complex numbers by treating them as binomials and rationalizing the denominators may be simplified considerably by first converting the vectors into polar form and then proceeding to combine them in the following manner:

To obtain the product of two vectors, multiply the numbers representing the vectors in polar form and add their corresponding angles algebraically. The resultant vector is in polar form. Thus,

$$(5 \angle +53^\circ)\ (5 \angle -53^\circ) = 25 \angle 0^\circ.$$

To obtain the quotient of two vectors, divide the numerator by the denominator as in ordinary division, then subtract algebraically the angle of the denominator from the angle in the numerator. The resultant vector is in polar form. Thus,

$$\frac{10 \angle +25^\circ}{5 \angle -20^\circ} = 2 \angle +45^\circ.$$

SERIES RESONANCE

SERIES-RESONANT CIRCUIT

A series-resonant circuit is composed of a capacitor, an inductor, and a resistor, as shown in figure 4-5,A. The circuit losses that occur in the capacitor, the inductor, and the connecting leads are assumed to occur in the resistor, R, in the circuit.

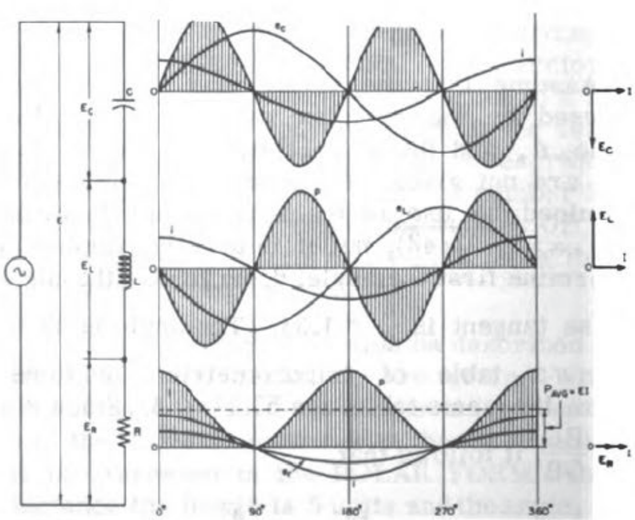
CONDITIONS REQUIRED FOR SERIES RESONANCE

In order for the series circuit (fig. 4-5,A) to be in resonance, the frequency of the applied voltage must be such that $X_L = X_C$.

When a series circuit contains resistance, inductive reactance, and capacitive reactance, the total impedance for any frequency is given as

$$Z=R+j(X_L-X_C).$$

Because X_L increases and X_C decreases with an increase in frequency, at a certain frequency (the resonant frequency) X_L will equal X_C, they will cancel, the j term will drop out, and Z will equal R. Therefore, at the resonant frequency, the power factor is unity. Furthermore, because the total impedance is now only the resistance, R, of the circuit, the circuit current is a maximum. In other words, at resonance the generator is looking into a pure resistance.



20.64

Figure 4-5.—Series resonance.

The resonant frequency of the series circuit is established as follows: At resonance

$$X_L = X_C \tag{4-1}$$

in which

$$X_L = 2\pi fL$$

and

$$X_C = \frac{1}{2\pi fC}.$$

Therefore, substituting the proper values in equation (4-1) gives

$$2\pi fL = \frac{1}{2\pi fC}. \tag{4-2}$$

Transposing (4-2)

$$f^2 = \frac{1}{4\pi 2LC}; \tag{4-3}$$

and solving for f in (4-3),

$$f = \frac{1}{2\pi\sqrt{LC}} \tag{4-4}$$

where f is in cycles per second, L is in henrys, and C is in farads.

At frequencies below resonance, X_C is greater than X_L and the circuit contains resistance and capacitive reactance; at frequencies above resonance, X_L is greater than X_C and the circuit contains resistance and inductive reactance. At resonance, the current is limited only by the relatively low value of resistance.

Because the circuit (fig. 4-5,A) is a series circuit, the same current flows in all parts of the circuit, and therefore the voltage across the capacitor is equal to the voltage across the inductor, because X_L is equal to X_C. These voltages (fig. 4-5,C), however, are 180° out of phase, since the voltage across a capacitor lags the current through it by approximately 90° and the voltage across the inductor leads the current through it by approximately 90°. The total value of the input voltage, E, then appears across R and is shown as E_R in phase with the current, I.

Assume that at a given instant corresponding to angle 0°, the current through the circuit is a maximum as indicated in figure 4-5,B. During the first quarter cycle (from 0° to 90°) the circuit current falls from maximum to zero. The capacitor is receiving a charge, as is indicated by the rising voltage, e_c, across it. The product of the instantaneous values of e_c and i for this interval indicates a positive power curve. The shaded area under this curve represents the energy stored in the capacitor during the time it is receiving a charge.

During the first quarter of a cycle (0° to 90°), when the capacitor is receiving a charge, the magnetic field about the inductor is collapsing because the circuit current is falling and the inductor acts like a source of power that supplies the charging energy to the capacitor. The voltage, e_L, across the coil, is opposite in phase to the voltage building up across the capacitor and is shown below the line. Therefore, the product of the instantaneous values of the current and voltage across the inductor indicates a negative power curve for the coil between 0° and 90°.

During the second quarter cycle (90° to 180°) the capacitor discharges from maximum to zero, as indicated by the capacitor voltage curve, e_c, and the coil reverses its function and acts like a load on the capacitor. Thus, the capacitor now acts as a source of power. The product of a negative current and a positive voltage (e_c) indicates a negative power curve for the capacitor for this interval. During the same quarter cycle the current is rising through the inductor (in the opposite direction) and energy is being stored in the magnetic field. The product of the negative current and negative voltage, e_L, for the second quarter cycle indicates a positive power curve for the inductor.

A similar interchange of energy between the capacitor and inductor takes place in the third and fourth quarter cycles. Therefore, the average power supplied to the inductor and capacitor by an external source is essentially zero. All circuit losses are assumed to be in the resistor, R. The voltage across the resistor and the current through it are in phase. The product of the voltage and current curves associated with the resistor indicates a power curve that has its axis displaced above the X axis. The displacement is proportional to the true average power which is equal to the product, EI (where E and I are effective values). Whatever power is dissipated in R is supplied by the source.

A simple mechanical analogy of the series resonant circuit is helpful in describing the interchange of energy between the inductance and the capacitance and the relatively large increase in voltages appearing across them when the circuit is resonant.

The analogy consists of a weight (about 4 oz) attached to a string of thin rubber bands (about six). With one end of the rubber bands suspended from the fingers the weight will bounce up and down with a very slight motion of the hand.

The weight represents inductance in the series circuit. Its inertia is like the inductance of the coil. The rubber bands represent capacitance in the series circuit. Their elasticity resembles the capacitance of the capacitor. The amplitude of motion of the hand represents the magnitude of the input voltage. The frequency of motion of the hand represents the frequency of the input voltage.

When the hand is moved slowly up and down, the weight will follow the motion with little or no change in the length of the rubber bands. This action corresponds to that in a series L-C circuit when the frequency of the applied voltage is considerably below resonance. Little or no change in the length of the rubber bands represents reduction of current. The capacitive reactance of the capacitor is large and restricts the current. The weight follows the motion of the hand. The inductive reactance of the coil is low (below resonance) and only the capacitive reactance restricts the current.

Conversely, when the motion of the hand is made very rapid (above resonance) the weight stands still and the rubber bands elongate and shorten in step with the motion of the input. Little or no change in the position of the weight represents a reduction of current. The inductive reactance of the coil is large and restricts the current. The rubber bands follow the motion of the hand. The capacitive reactance of the capacitor is low and only the inductive reactance restricts the current.

Thus both below and above resonance, current is reduced. When the rate of motion of the hand is adjusted to the resonant frequency of the system, the weight will bounce through a much larger range of motion than that of the input. The large amplitude of motion of the weight and rubber bands represents a large increase in current (maximum) and voltage (maximum) developed across both the coil and the capacitor. The weight is pulled up as the rubber bands shorten so that the direction of the two amplitudes is opposite. This action corresponds to the 180° phase shift between the voltage across the coil and the voltage across the capacitor.

The analogy of resonance can be summed up as follows. Maximum elongation of the rubber band represents maximum energy stored in the capacitor. The rubber band shortening and thus lifting the weight represents the capacitor discharging its energy to the inductor. The weight at maximum elevation represents the maximum energy stored in the inductor's magnetic field. The falling weight stretching the rubber band represents the inductor discharging its energy to the capacitor. Thus the interchange of energy between the capacitor and the inductor is repeated.

The ratio of the amplitude of motion of the weight to the input amplitude represents the voltage gain of the series resonant circuit. The interchange of the energy between the weight and rubber bands keeps the system moving with very little input energy. The ratio of the energy stored periodically in the weight (or rubber bands) to the input energy is called the quality, or Q, of the system.

QUALITY, OR Q

The ratio of the energy stored in an inductor during the time the magnetic field is being established to the losses in the inductor during the same time is called the **QUALITY**, or **Q**, of the inductor; it is also called the **FIGURE OF MERIT** of the inductor. This ratio is

$$\frac{I^2X_Lt}{I^2Rt} = \frac{X_L}{R} = Q.$$

In terms of the mechanical analogy of the weight and rubber bands the ratio corresponds to the ratio of energy stored in the weight (during the time it is moving upward against gravity) to the input energy supplied by hand. If the oscillating weight is immersed in a liquid, it will be damped and the amplitude of oscillation will be reduced. More input energy will be required to maintain a given amplitude of oscillation than in air. This action corresponds to an increase in the effective resistance of the series resonant circuit. If the input voltage is not increased, the current and voltage drops across the coil and the capacitors will be reduced.

The Q of the inductor is therefore equal to the ratio of the inductive reactance to the effective resistance in series with it, and it approaches a high value as R approaches a low value. Thus, the more efficient the inductor, the lower the losses in it and the higher is the Q.

In terms of the impedance triangle (fig. 4-6,A)

$$Q = \frac{X_L}{R} = \tan\theta,$$

where θ is the phase angle between the hypotenuse, Z, and the base, R. As θ approaches 90°, $\tan\theta$ approaches infinity, and the coil losses approach zero.

Similarly, in a capacitor the Q is a measure of the ratio of the energy stored to the energy dissipated in heat within the capacitor for equal intervals of time. This ratio is

$$\frac{I^2X_Ct}{I^2Rt} = \frac{X_C}{R} = Q,$$

where R is the effective resistance in series with the capacitive reactance, X_C (fig. 4-6,B). The effective resistance is low with respect to the capacitive reactance, and is such that when multiplied by the square of the effective capacitor current equals the true power dissipated in heat within the capacitor.

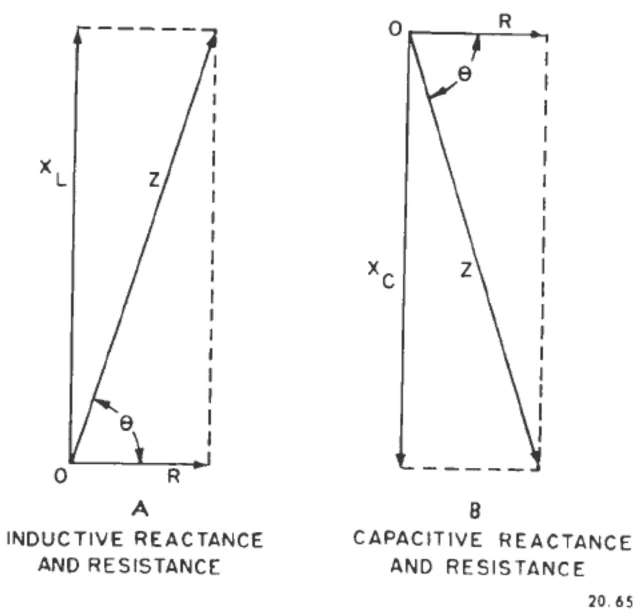


Figure 4-6.—Impedance triangles.

In terms of the mechanical analogy of the weight and rubber bands, the capacitor losses correspond to those in the string of rubber bands. If the rubber bands are stiff and not very elastic, the amplitude of motion of the weight will be reduced. This action corresponds to an increase in the effective resistance of the capacitor.

Since most of the losses in a solid-dielectric capacitor occur within the dielectric rather than in the plates, the Q of low-dielectric-loss capacitors is high. The losses of an air-dielectric capacitor are negligible, and thus the Q of such a capacitor may have a very high value.

The Q of a circuit, like the series-resonant circuit of figure 4-5,A, is the ratio of the energy stored to the energy lost in equal intervals of time. The expression becomes

$$Q = \frac{I^2X_Lt}{I^2Rt} = \frac{I^2X_Ct}{I^2Rt} = \frac{X_L}{R} = \frac{X_C}{R},$$

where R represents the total effective series resistance of the entire circuit. If the capacitor has negligible losses, the circuit Q becomes equivalent to the Q of the coil. The circuit Q may be maintained satisfactorily high by keeping the circuit resistance to a minimum.

In figure 4-5,A, the voltage across L is

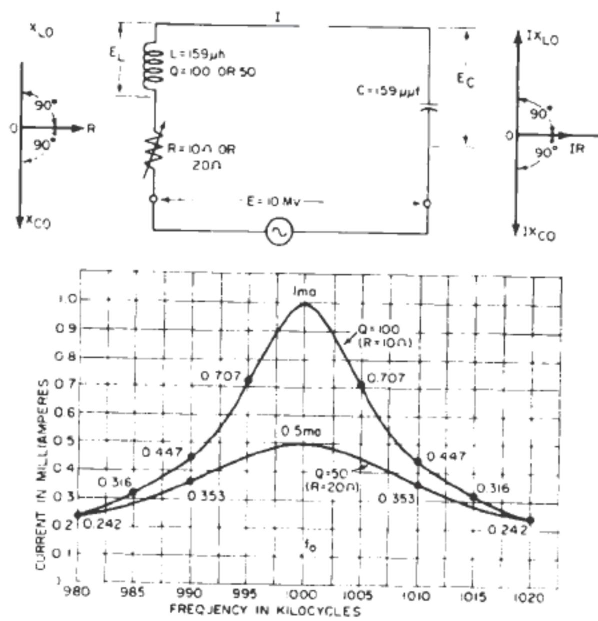
$$E_L = IX_L = \frac{EX_L}{R} = QE,$$

where $I = \frac{E}{R}$ and $Q = \frac{X_L}{R}$.

The Q of the circuit is the ratio of the voltage across either the inductor or capacitor to that across the effective series resistance. In other words, the voltage gain of the series-resonant circuit depends on the circuit Q—that is

$$V.G. = \frac{E_L}{E} = \frac{IX_L}{IR} = \frac{X_L}{R} = \frac{IX_C}{IR} = \frac{X_C}{R} = Q.$$

Figure 4-7 shows the relation between the effective current and frequency in the vicinity of resonance for a series circuit containing a 159-μh coil, a 159-$\mu\mu$f capacitor, and an effective series resistance of either 10 ohms, or 20 ohms.



20.66

Figure 4-7.—Resonance curves of a series L-C-R circuit.

The resonant frequency, f_o, is

$$f_o = \frac{1}{2\pi\sqrt{LC}}$$

$$f_o = \frac{1}{6.28\sqrt{159 \times 10^{-6} \times 159 \times 10^{-12}}}$$

$$= 1 \times 10^6 \text{ cycles or 1000 kc.}$$

The reactances and impedance at resonance may likewise be determined. Thus,

$$X_{L_O} = 2\pi fL = 6.28 \times 10^6 \times 159 \times 10^{-6}$$

$$= 1{,}000 \; +90^\circ \text{ ohms}$$

where X_{L_O} is the inductive reactance at resonance. The +90° angle indicates that the IX_{L_O} and X_{L_O} vectors are plotted vertically upward because the current vector is horizontal and extends to the right. The current vector thus lags the voltage IX_{L_O} across the coil by 90° (counterclockwise rotation is positive, figure 4-7).

$$X_{C_O} = \frac{1}{2\pi fC} = \frac{1}{6.28 \times 10^6 \times 159 \times 10^{-12}}$$

$$= 1{,}000\angle{-90^\circ} \text{ ohms}$$

where X_{C_O} is the capacitive reactance at resonance. The -90° angle indicates that the vectors X_{C_O} and IX_{C_O} are plotted vertically downward because the current vector is the horizontal reference vector extending to the right and leads the voltage drop, IX_{C_O}, across the capacitor by 90° (fig. 4-7). Note that current is a common factor to both voltage and impedance vectors.

When R = 10 ohms,

$$Z_O = R + jX_L - jX_C = 10 + j1{,}000 - j1{,}000$$

$$= 10\angle 0^\circ \text{ ohms.}$$

If the applied voltage is assumed to be 10 millivolts (mv) at a frequency of 1000 kc, the circuit current is

$$I = \frac{E}{Z} = \frac{0.01}{10} = 0.001, \text{ or 1 ma.}$$

At the resonant frequency, the voltage across the inductor is

$$E_L = IX_L = 0.001 \times 1{,}000 = 1 \text{ v,}$$

and the voltage across the capacitor is the same, except it is 180° out of phase with the voltage across the coil. The losses in the coil and capacitor are assumed to be lumped in the effective series resistance. The circuit Q is

$$Q = \frac{X_L}{R} = \frac{1{,}000}{10} = 100.$$

The voltage gain at resonance is

$$\frac{E_L}{E} = \frac{IX_{L_O}}{IR} = \frac{0.001 \times 1000}{0.001 \times 10} = \frac{1.00}{0.01} = 100.$$

The resonance curves of current vs frequency are symmetrical about a vertical line extending through the points of maximum current and resonant frequency (fig. 4-7). This symmetry is the result of an equality of the quantity X_C-X_L below resonance a given amount and the quantity X_L-X_C above resonance by the same amount. The amount of deviation may be expressed in terms of the circuit Q and the resonant frequency as demonstrated in the following discussion.

The shape of the resonance curve may be approximated in the vicinity of resonance by applying the following rules that can be derived from the resonant circuit equations. (The derivation is not given here because of its length.)

1. If the frequency of the applied voltage is decreased by an amount $\frac{1}{2Q}$ times the resonant frequency, f_o, the current in the tuned circuit decreases to 0.707 of its value at the resonant frequency and leads the applied voltage by 45°. Thus, the input frequency is decreased an amount equal to

$$\frac{1}{2Q} \times f_o = \frac{1}{2 \times 100} \times 1{,}000 = 5 \text{ kc},$$

and the new frequency is therefore 995 kc. At a frequency of 995 kc,

$$X_L = 2\pi f L = 6.28 \times 0.995 \times 10^6 \times 159 \times 10^{-6}$$
$$= 995\angle{+90^\circ} \text{ ohms},$$

and

$$X_C = \frac{1}{2\pi f C} = \frac{1}{6.28 \times 0.995 \times 10^6 \times 159 \times 10^{-12}}$$
$$= 1{,}005\angle{-90^\circ} \text{ ohms}.$$

The circuit impedance at 995 kc is

$$Z = 10 + j995 - j1{,}005 = 10 - j10 = 14.14\angle{-45^\circ} \text{ ohms}.$$

The circuit current at 995 kc is

$$I = \frac{E}{Z} = \frac{0.010}{14.41} = 0.000707, \text{ or } 0.707 \text{ ma}.$$

At this frequency, the voltage across the coil, or the capacitor, is reduced to approximately 70 percent of its value at resonance because the current is reduced to this amount, and the reactance change is very small. The voltage across the coil is

$$E_L = IX_L = 0.707 \times 995 = 705 \text{ mv}.$$

2. If the frequency of the applied voltage is decreased by an amount $\frac{1}{Q}$ times the resonant frequency, the current decreases to 0.447 of its value at resonance and leads the applied voltage by 63.4°. Thus, in the example of figure 4-4

$$\frac{1}{Q} \times f_o = \frac{1}{100} \times 1{,}000 = 10 \text{ kc},$$

and 1,000-10=990 kc.

The inductive reactance at 990 kc is

$$X_L = 2\pi f L = 6.28 \times 0.990 \times 10^6 \times 159 \times 10^{-6}$$
$$= 990\angle{+90^\circ} \text{ ohms},$$

and the capacitive reactance is

$$X_C = \frac{1}{2\pi f C} = \frac{1}{6.28 \times 0.990 \times 10^6 \times 159 \times 10^{-12}}$$
$$= 1{,}010\angle{-90^\circ} \text{ ohms}.$$

The impedance of the series circuit at 990 kc is

$$Z = R + jX_L - jX_C = 10 + j990 - j1{,}010$$
$$= 10 - j20 = 22.4\angle{-63.4^\circ} \text{ ohms}.$$

At this frequency the circuit current is

$$I = \frac{E}{Z} = \frac{0.010}{22.4} = 0.000447, \text{ or } 0.447 \text{ ma},$$

and the voltage across the coil is

$$E_L = IX_L = 0.447 \times 990 = 444 \text{ mv}.$$

Corresponding increases in the frequency of the applied voltage above the resonant frequency will produce the same reductions in circuit current and voltage across the reactive portions of the circuit. In this case, however, the circuit current lags the applied voltage instead of leading it. Thus, the resonance curve is symmetrical about the resonant frequency in the vicinity of resonance.

Influence of Q on Voltage Gain

If the circuit resistance is increased to 20 ohms, the Q is reduced to 50 and the resonance curve is flattened, as shown by the lower curve in figure 4-7. The series-resonant circuit amplifies the applied voltage at the resonant frequency. If the circuit losses are low the circuit Q will be high and the voltage amplification relatively large. For resonant circuits involving iron-core coils the Q may range from 20 to 100; for silver-plated resonant cavities at very high frequencies, the Q may range as high as 30,000. In practice, because nearly all of the resistance of a circuit is in the coil, the ratio of the inductive reactance to the resistance is especially important. The higher the Q of the coil, the better is the coil and the more effective is the series-resonant circuit that utilizes it.

Reduction in Voltage across C and L near Resonance

If the circuit Q is low, the amplification at resonance is relatively small and the circuit does not discriminate sharply between the resonant frequency and the frequencies on either side of resonance, as is shown by the lower curve in figure 4-7. The range of frequencies included between the two frequencies at which the current drops to 70 percent of its value at resonance is called the bandwidth for 70-percent response. A measure of the bandwidth for 70-percent response is $\frac{f_0}{Q}$. If the circuit Q is 100, the bandwidth for 70-percent response is

$$\frac{1,000}{100} = 10 \text{ kc.}$$

Thus, if the frequency of the applied voltage is reduced from 1,000 kc to 995 kc or increased to 1,005 kc, the circuit current is reduced to 70 percent of its value at resonance. Likewise, the voltage across L or C is reduced to approximately 70 percent of its value at resonance. For the lower curve, representing a circuit having a Q of 50, the bandwidth is 20 kc.

APPLICATIONS OF SERIES-RESONANT CIRCUITS

Series-resonant circuits are used largely as filters (to be treated later in this chapter) for audio and radio frequencies. With proportionately larger componet values the series circuit may be used as a power-supply filter. For example, assume that a d-c generator has a ripple frequency of 500 cps. A series-resonant circuit tuned to 500 cps may be connected across the terminals of the generator and thus effectively short-circuit the ripple voltage. The coil and capacitor insulation must be able to withstand the relatively high a-c voltages caused by the series-resonant action.

The series-tuned circuit may also be used to give an indication of frequency if the capacitor is calibrated for the appropriate frequency range. The capacitor and the inductor are connected in series with a current-indicating device across the source of the unknown frequency. At resonance the current, as indicated by the device, will be a maximum.

PARALLEL RESONANCE

PARALLEL-RESONANT CIRCUIT

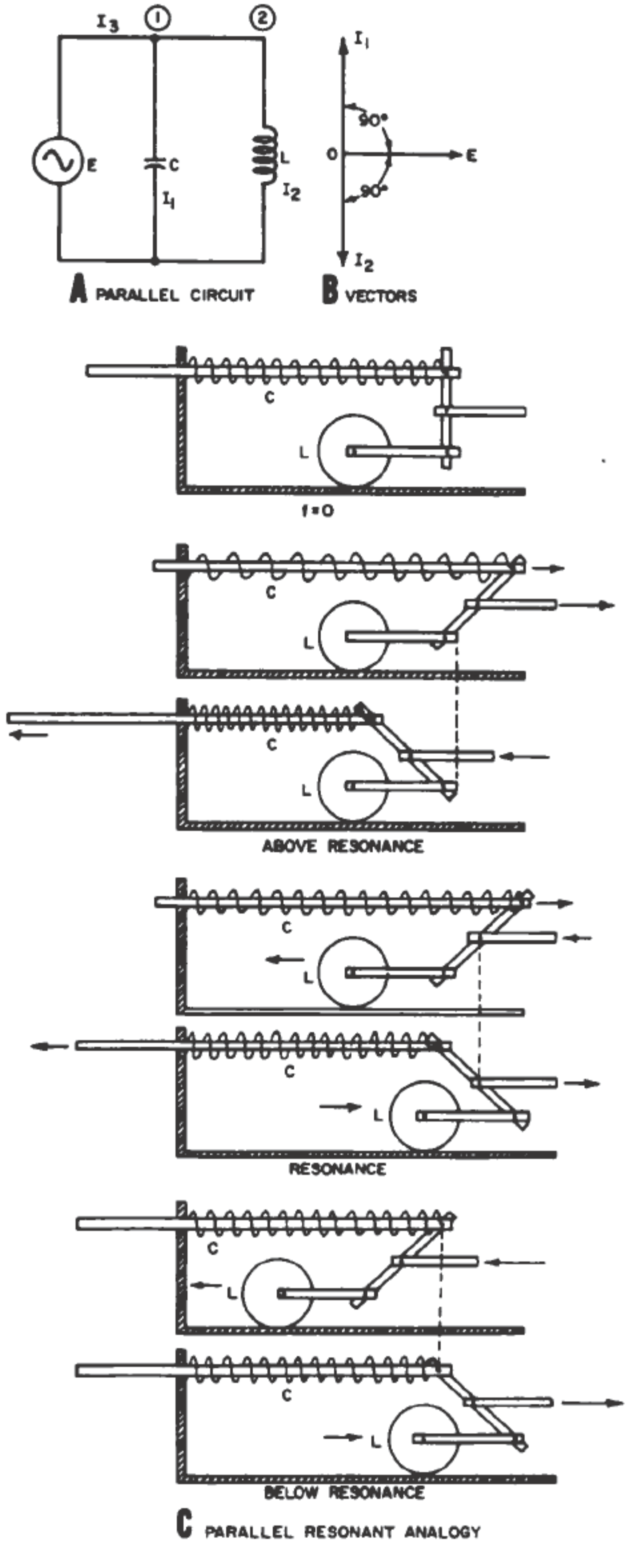
A parallel circuit containing no losses has capacitance C in one branch and inductance L in the other (fig. 4-8,A). The current, I_1, in the capacitive branch leads the voltage, E, by 90°, and the current, I_2, in the inductive branch lags E by 90° (fig. 4-8,B). At resonance, $I_1=I_2$, and the vector sum of these currents is equal to the line current, I_3. Since I_1 is 180° out of phase with I_2, the vector sum of I_1 and I_2 is zero, and the line current, I_3, is negligible.

A mechanical analogy for the two-branch parallel resonant circuit is illustrated in figure 4-8,C. The wheel has inertia (like inductance) and the spring is elastic (like capacitance). Above resonance the spring moves back and forth (large capacitive current) while the wheel moves hardly at all (low inductive current). At resonance both the wheel and the spring move through a wide range (large tank current) while the input moves only a little (small line current). Below resonance the wheel moves back and forth (large inductive current) while the spring remains almost stationary (low capacitive current).

A parallel-resonant circuit with losses consists of a combination of inductance, resistance, and capacitance in two parallel branches (fig. 4-9,A). Because the losses of the circuit are generally associated with the inductor (wire), this branch includes a series resistor, R, in which all the losses are lumped. The other branch consists of a capacitor having negligible loss. At resonance, the same interchange of energy occurs between the capacitor and the inductor that occurs in the series-resonant circuit. The circuit impedance vs frequency is shown in figure 4-9,B, and the vector diagram is shown in figure 4-9,C.

CONDITIONS REQUIRED FOR PARALLEL RESONANCE

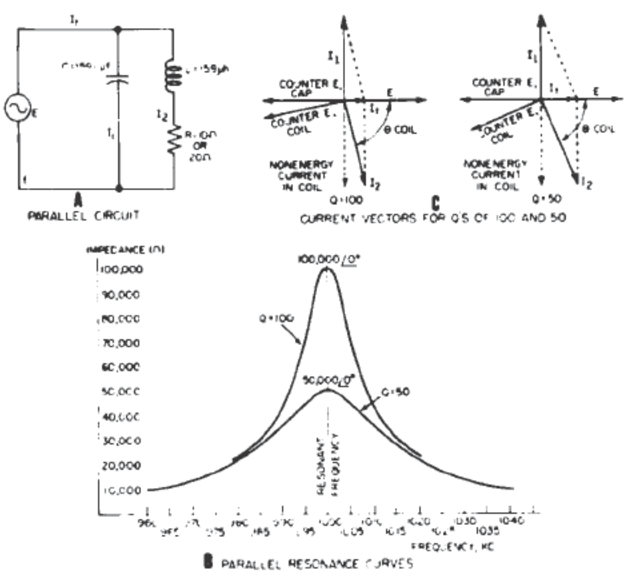
At resonance the current, I_1, in the capacitive branch (fig. 4-8) is equal to the current, I_2, in the inductive branch. These currents are nonenergy currents; they are 90° out of phase with the applied voltage (one leads 90° and the other lags 90°), and they flow in circuits where the true power is zero, and the reactive power is periodically exchanged between the branches.



20.67

Figure 4-8.—Parallel resonance and mechanical analogy.

In the inductive branch containing losses (fig. 4-9) the current may be thought of as being made up of two components-one in phase with the applied voltage (an energy component) and the other lagging the voltage 90° (a non-energy component). In this circuit, at resonance, the nonenergy component of the current is equal to the nonenergy current, I_1, in the capacitive branch. The vector sum of these two currents is zero because they are equal in magnitude and 180° out of phase. The line current, I_t, represents the relatively small value of energy current that flows in the inductive branch (fig. 4-9,C). Thus, the parallel-resonant circuit has a high input impedance, and the line current is in phase with the applied voltage, which is the condition of unity power factor.



20.68

Figure 4-9.—Parallel resonance.

FORMULAS FOR f_o AND Z_o

For all practical purposes the resonant frequency of a parallel circuit having a Q of 10 or higher may be expressed as

$$f_o = \frac{1}{2\pi\sqrt{LC}},$$

where f_o is in cycles, L is in henrys, and C is in farads. This is the same as the formula for the series-resonant circuit, because here too there is an (approximate) equality between the inductive and capacitive reactances.

The parallel circuit is resonant when the nonenergy components of the currents in the two branches are equal, and the power factor of the circuit is unity. In the zero loss capacitive branch the impedance is equal to X_C, and the capacitive current is $\frac{E}{X_C}$.

In the inductive branch the impedance is approximately equal to X_L if the losses are low and the Q of the coil is 10 or higher. In this case, the nonenergy components of the coil current is almost equal to the total coil current. Thus, at resonance

$$I_1 = I_{COIL\ NONEN}$$

and

$$\frac{E}{X_C} = \frac{E}{X_L}$$

from which

$$X_L = X_C.$$

If Q is less than 10, Z_{COIL} is greater than X_L, and the total coil current is greater than the nonenergy component of the coil current. In this case X_L does not equal X_C at the frequency for which the line current is minimum and the combined impedance is maximum. However, for most cases Q (which is equal to tan θ) is more than 10, and the above approximation is adequate.

The combined impedance, Z_o, at resonance for the two-branch parallel circuit is equal to the product of the individual branch impedances divided by their sum. This calculation is similar to that for the combined resistance of two resistors in parallel, except that impedance vector products are substituted for the product of the resistors, and impedance vector sums are substituted for the resistor sums. The symbol for vector summation is $\oplus$. Vector products and sums are described earlier in this chapter.

$$Z_o = \frac{Z_1 Z_2}{Z_2 \oplus Z_1} \qquad (4\text{-}5)$$

In figure 4-9,A, the following approximations are made:

$$Z_1 = -jX_C,\ Z_2 = R + jX_L \approx jX_L,\ X_L = X_C$$

and

$$Z_1 Z_2 = (-jX_C)(jX_L) = X_C X_L. \qquad (4\text{-}6)$$

Substituting (4-6) in (4-5),

$$Z_o = \frac{X_C X_L}{R + jX_L - jX_C} = \frac{X_C X_L}{R} = \frac{X_L^2}{R} \qquad (4\text{-}7)$$

Note that

$$X_C = \frac{1}{2\pi fC} \text{ and } X_L = 2\pi fL. \qquad (4\text{-}8)$$

Substituting (4-8) in (4-7) and simplifying,

$$Z_o = \frac{\frac{1}{2\pi fC}\ \frac{2\pi fL}{1}}{R} = \frac{L}{CR}. \qquad (4\text{-}9)$$

Thus, the $\frac{L}{C}$ ratio (R being constant) is a factor that determined the magnitude of the impedance of the parallel circuit at the resonant frequency. The impedance-frequency curves (fig. 4-9,B) for a parallel circuit have the same shape as the current-frequency curves for a series circuit. Note that the impedance across the terminals of the parallel circuit is maximum at resonance, whereas it is minimum for the series circuit. As in the series circuit, the resonance curves are sharper when the internal resistances are smaller, and the Q's are higher.

For example, the two-branch parallel-tuned circuit (fig. 4-9,A) has a capacitor of 159 $\mu\mu$f with negligible losses in parallel with a 159-μh coil having an effective resistance of 10 ohms. The resonant frequency is

$$f_o = \frac{1}{2\pi\sqrt{LC}} = \frac{1}{6.28\sqrt{159\text{x}10^{-6}\text{x}159\text{x}10^{-12}}}$$
$$= 10^6 \text{ cycles, or 1000 kc.}$$

The impedance at resonance is

$$Z_o = \frac{L}{CR} = \frac{159\text{x}10^{-6}}{159\text{x}10^{-12}\text{x}10} = 100{,}000 \text{ ohms.}$$

The coil reactance at the resonant frequency is

$$X_L = 2\pi fL = 6.28\text{x}10^6\text{x}159\text{x}10^{-6} = 1{,}000 \text{ ohms.}$$

The coil Q is

$$\frac{X_L}{R_1} = \frac{1{,}000}{10} = 100.$$

Thus, as the parallel circuit strikes resonance, there is a rise in the combined impedance equal to Q times the coil reactance. The lower the coil resistance, the higher will be the coil Q and the combined impedance at the resonant frequency. The vector diagram for this example is shown at the left in figure 4-9,C.

In the mechanical analogy for the parallel resonant circuit (fig. 4-8,C) the Q of the system is the ratio of the amplitude of motion of pivot P2 to that of pivot P. This ratio corresponds to the ratio of the current in the coil to the current in the line. If the losses are high the line current will be relatively large, and the circuit Q will be low. Thus, if the flywheel rolls on a soft surface, the input will have to increase to maintain the same movement of the flywheel as would occur if it rolls on a hard flat surface. The lower the losses in the parallel resonant circuit, the higher will be the ratio of coil current to line current, and the sharper will be the rise in the impedance as resonance is approached.

PARALLEL IMPEDANCE NEAR RESONANCE

The parallel impedance, Z_t, of a parallel circuit at any frequency is

$$Z_t = \frac{Z_1 Z_2}{Z_1 \oplus Z_2},$$

where Z_1 is the impedance of one branch and Z_2 is the impedance of the other branch.

Below resonance—for example, at 995 kc—the impedance of branch 1 is

$$\frac{1}{2\pi fC} = \frac{1}{6.28 \times 995 \times 10^3 \times 159 \times 10^{-12}}$$
$$= 1{,}005\angle -90^\circ,$$

and the impedance of branch 2 is

$$R + j2\pi fL = 10 + j6.28 \times 995 \times 10^3 \times 159 \times 10^{-6}$$
$$= 10 + j995, \text{ or } 995\angle +90^\circ.$$

The parallel impedance is

$$Z_t = \frac{(1{,}005\angle -90^\circ)(995\ \ +90^\circ)}{0 - j1{,}005 + 10 + j995}$$
$$= \frac{1{,}000{,}000\angle 0^\circ}{10 - j10}$$
$$= \frac{1{,}000{,}000\angle 0^\circ}{14.14\angle -45^\circ}$$
$$= 70{,}700\angle +45^\circ.$$

In this example, the frequency deviation is $\frac{f_o}{2Q}$ or $\frac{1{,}000}{2 \times 100} = 5$ kc. When the frequency is deviated, an amount equal to $\frac{f_o}{2Q}$, the impedance falls to 70 percent of its value at resonance, and the phase angle increases from 0° to 45°. Above resonance, the angle is negative; below resonance it is positive. In other words, above the resonant frequency, the circuit acts like capacitance in series with resistance; below resonance it acts like inductance in series with resistance.

A summary of the characteristics of series and parallel resonant circuits is given in table 4-2.

LOADING THE PARALLEL-RESONANT CIRCUIT

Increasing the resistance (from 10 to 20 ohms) in series with the coil lowers the coil Q from 100 to 50. Since there are negligible losses in the capacitor the circuit Q is halved and the total impedance of the parallel circuit at the resonant frequency is (from equation 4-9)

$$Z_o = \frac{L}{CR_2} = \frac{159 \times 10^{-6}}{159 \times 10^{-12} \times 20}$$

$$= \frac{1{,}000{,}000}{20} = 50{,}000 \text{ ohms.}$$

Thus, the parallel impedance at resonance varies inversely with the resistance in the coil branch. This series resistance may represent the load on the parallel circuit. Hence, an increase in series resistance in the coil circuit may represent an increased load and a decrease in the total impedance of the parallel circuit. The vector diagram for this example is shown at the right in figure 4-9,C.

The Q of a parallel-tuned circuit at resonance may be defined as the ratio of either the current in the capacitive or inductive branch, I_1 or I_2 to the line current I_t, or

$$Q = \frac{I_1}{I_t} = \frac{I_2}{I_t}$$

Table 4-2.—Characteristics of Series and Parallel Resonant Circuits.

Quantity	Series Circuit	Parallel Circuit
At resonance: Reactance $(X_L - X_C)$	Zero; because $X_L = X_C$	Zero; because nonenergy currents are equal
Resonant frequency	$\frac{1}{2\pi\sqrt{LC}}$	$\frac{1}{2\pi\sqrt{LC}}$
Impedance	Minimum; Z=R	Maximum; $Z = \frac{L}{CR}$
I_{LINE}	Maximum value	Minimum value
I_L	I_{LINE}	$Q \times I_{LINE}$
I_C	I_{LINE}	$Q \times I_{LINE}$
E_L	$Q \times E_{LINE}$	E_{LINE}
E_C	$Q \times E_{LINE}$	E_{LINE}
Phase angle between E_{LINE} and I_{LINE}	0°	0°
Angle between E_L and E_C	180°	0°
Angle between I_L and I_C	0°	180°
Desired value of Q	10 or more	10 or more
Desired value of R	Low	Low
Highest selectivity	High Q, low R, high $\frac{L}{C}$	High Q, low R
When f is greater than f_o: Reactance	Inductive	Capacitive
Phase angle between I_{LINE} and E_{LINE}	Lagging current	Leading current
When f is less than f_o: Reactance	Capacitive	Inductive
Phase angle between I_{LINE} and E_{LINE}	Leading current	Lagging current

When a load is inductively coupled to the tuned circuit, the load, in effect, adds resistance within the tuned circuit. This action is described later in the section on coupled circuits. The impedance of the circuit is thereby reduced, and the line current is increased, thus lowering the circuit, Q. If a load is connected in shunt with the tuned circuit, the line current is increased and the circuit Q is lowered.

As long as the circuit Q is maintained above 10, the increased load that the increased series resistance represents does not materially affect the phase angle between line current and line voltage and does not change resonant frequency. For values of Q above 10, θ_{COIL} varies only slightly. Thus, when Q varies from 10 to infinity, tan θ varies from 10 to infinity, and θ_{COIL} varies from 84.3° to 90° or 5.7°. For values of Q below 10, θ_{COIL} varies widely. Thus, as Q varies from 10 to 0, tan θ_{COIL} varies from 10 to 0, and θ_{COIL} varies from 84.3° to 0°. This wide variation makes an appreciable difference in the phase angle between line current and voltage, and lowers the resonant frequency.

The increase in line current shown at the right in figure 4-9,C, as a result of the increase in resistance of the coil branch from 10 ohms to 20 ohms, is the result of the slight decrease in phase angle θ for the coil and the corresponding slight shift in the phase angle between the counter emf of the coil and the applied voltage.

The parallel-resonant circuit is often called a TANK CIRCUIT because it acts like a storage tank when used in some electron-tube circuits. The inertia effect of the inductor gives the tank a FLYWHEEL effect that permits the alternating current to build up in the tank. The relatively large current in the tank is equal to the circuit Q times the line current; the amplification of current is like the gain in momentum of a flywheel as it is being accelerated. The high input impedance of the parallel-resonant circuit is the result of the relatively large inductive emf of the inductor and the capacitive emf of the capacitor, both in approximate phase opposition with respect to the source voltage (fig. 4-9,C).

The parallel circuit is frequency sensitive to a varying degree, depending on the Q of the circuit. In other words, the change in impedance with a change in frequency is more pronounced in a high Q parallel circuit than in a low Q circuit. Below resonance, the lower impedance of the inductive branch causes the line current to increase and to lag the applied voltage. Conversely, above resonance the lowered impedance of the capacitive branch causes the line current to again increase and to lead the applied voltage. At resonance, the impedance is high and resistive. Generally the tank circuit is supplied by a relatively high-impedance source compared with series-resonant circuit sources. As the frequency of the source voltage is varied from below to above the resonant frequency, the voltage rises across the tank at the resonant frequency, and the line current falls as the tank current rises. The rise in voltage across the tank as resonance is approached is due to the decrease in line current and internal voltage loss of the source.

Frequently the tank circuit is used to couple energy into a load by utilizing the inductor as the primary of a transformer with the secondary connected to the load. When the secondary is tuned to the resonant frequency of the tank, the secondary current becomes a maximum. The field of the secondary current cuts the primary inductor and induces a counter emf in that coil. As mentioned previously, this action is equivalent to adding effective resistance in series with the inductive branch. Thus, coupling a load to the tank through mutual inductance lowers the parallel impedance and increases the line current. Coupling the load in this manner tends to slightly detune the tank so that it is generally necessary to return by adjusting the capacitor. Loading the tank lowers the circuit Q, the parallel impedance, the tank current, and the voltage across the tank. At the same time the line current is increased.

APPLICATIONS OF PARALLEL-RESONANT CIRCUITS

The parallel-resonant circuit is one of the most important circuits used in electronic transmitters, receivers, and frequency-measuring equipment.

The i-f transformers of radio and television receivers employ parallel-tuned circuits. These are transformers used at the input and output of each intermediate frequency amplifier stage for coupling purposes and to provide selectivity. They are enclosed in a shield can provided with openings at the top through which screwdriver adjustments may be made when the set is being aligned.

Parallel-tuned circuits are also used in the driver and power stages of transmitters, as well as in the oscillator stages of transmitters, receivers, and frequency-measuring equipment.

Various types of filter circuits employ parallel-tuned circuits as well as series-tuned circuits.

TUNED CIRCUITS AS FILTERS

Tuned circuits are employed as filters for the passage or rejection of specific frequencies. Band-pass filters and band-rejection filters are examples of this type. Tuned circuits have certain characteristics that make them ideal for certain types of filters, especially where high selectivity is desired. A series-tuned circuit offers a low impedance to currents of the particular frequency to which it is tuned and a relatively high impedance to currents of all other frequencies. A parallel-tuned circuit, on the other hand, offers a very high impedance to currents of its natural, or resonant, frequency and a relatively low impedance to others.

BAND-PASS FILTERS

A band-pass filter is designed to pass currents of frequencies within a continuous band, limited by an upper and lower cutoff frequency, and substantially to reduce, or attenuate, all frequencies above and below that band. A simple band-pass filter is shown in figure 4-10,A.

The curves of current vs frequency are shown in figure 4-10,B. The high Q circuit gives a steeper current curve; the low Q circuit gives a much flatter current curve.

The series-and parallel-tuned circuits are tuned to the center frequency of the band to be passed by the filter. The parallel-tuned circuit offers a high impedance to the frequencies within this band, while the series-tuned circuit offers very little impedance. Thus, the desired frequencies within the band will travel on to the load without being affected; but the currents of unwanted frequencies—that is, frequencies outside the desired band will meet with a high series impedance and a low shunt impedance so that they are in a greatly attenuated form at the load.

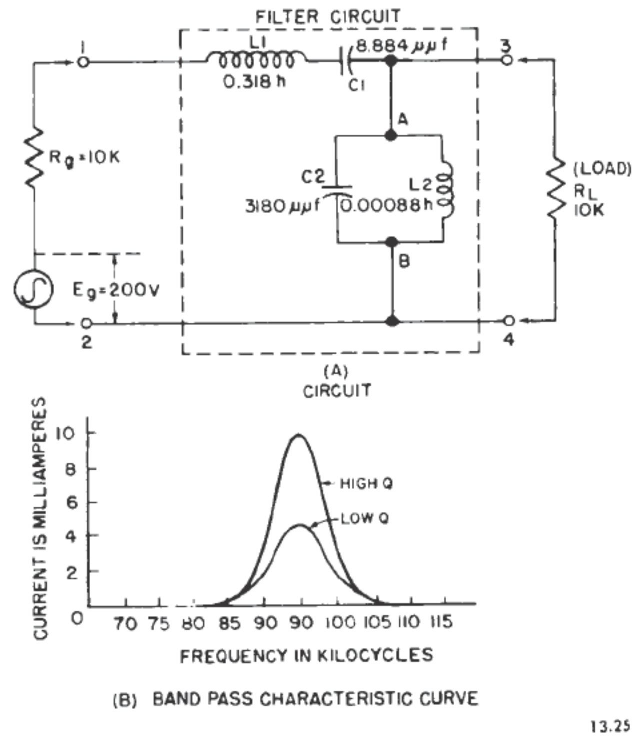


Figure 4-10.—Band-pass filter.

There are many circuit arrangements for both band-pass and band-elimination filters. However, for the purpose of a brief analysis the band-pass circuit shown in figure 4-10 will be considered. Let it be assumed that a band of frequencies extending from 90 to 100 kc is to be passed by the filter. For an input and output resistance of 10,000 ohms, the values of inductance and capacitance are as indicated in the figure. The formulas by which these values are obtained may be found in handbooks on the subject.

The resonant frequency of the series circuit, L1C1, is

$$f_o = \frac{1}{2\pi\sqrt{LC}}$$

$$f_o = \frac{1}{6.28\sqrt{0.318 \times 8.884 \times 10^{-12}}}$$

$$= \frac{0.159 \times 10^{+6}}{\sqrt{2.81}} = \frac{159{,}000}{1.68} = 95{,}000 \text{ cps},$$

and for the parallel circuit, L2C2, is

$$f_o = \frac{0.159}{\sqrt{3{,}180 \times 10^{-12} \times 8.8 \times 10^{-4}}} = 95{,}000 \text{ cps.}$$

Thus, both circuits are resonant at the center frequency of the band-pass filter; the upper limit of which is 100 kc and the lower limit 90 kc.

At resonance the impedances of L1 and C1 cancel and maximum current flows through the load, R_L; also, the parallel circuit, C2L2, offers almost infinite impedance and may be considered an open circuit. The inherent resistances associated with the filter components are neglected. Thus, at resonance, with an assumed source voltage of 200 volts and a total impedance of 20,000 ohms (10,000 ohms at the source and 10,000 ohms at the load) the current through the load is approximately

$$I = \frac{E}{R} = \frac{200}{20{,}000} = 0.01 \text{ a, or } 10 \text{ ma.}$$

Below resonance—for example, at 90 kc-the impedances of L1C1 and C2L2 are such that with the assumed source voltage of 200 volts only about 4 milliamperes flow through the load. The calculations are as follows:

at 90 kc:

$$X_{L1} = 2\pi f L_1 = 6.28 \times 90 \times 10^3 \times 0.318$$
$$= 180{,}000 \text{ ohms}$$

$$X_{C1} = \frac{1}{2\pi f C_1} = \frac{10^{12}}{6.28 \times 90 \times 10^3 \times 8.884}$$
$$= 200{,}000 \text{ ohms}$$

$$X_{C1} - X_{L1} = j20{,}000 \text{ ohms}$$

$$X_{L2} = 2\pi f L_2 = 6.28 \times 90 \times 10^3 \times 0.88 \times 10^{-3}$$
$$= 497 \text{ ohms} \angle 90^\circ$$

$$X_{C2} = \frac{1}{2\pi f C_2} = \frac{10^{12}}{6.28 \times 90 \times 10^3 \times 3180}$$
$$= 556 \text{ ohms} \angle -90^\circ$$

The impedance offered to the flow of current entering and leaving terminals AB is equal to the product of L2C2 impedances divided by their sum.

$$Z_{AB} = \frac{X_{L2} X_{C2}}{+jX_{L2} - jX_{C2}} = \frac{(497\angle 90^\circ)(556\angle -90^\circ)}{+j497 - j556}$$

$$= \frac{(497 \times 556)\angle 0^\circ}{59\angle -90^\circ} = 4700\angle 90^\circ \text{ ohms.}$$

The parallel impedance, Z_{AB34}, of R_{LOAD} and Z_{AB} is equal to their product divided by their sum.

$$Z_{AB34} = \frac{Z_{AB} R_L}{R_L + Z_{AB}} = \frac{(4700\angle 90^\circ)(10{,}000\angle 0^\circ)}{10{,}000 + j4700}$$

$$\frac{(4700 \times 10{,}000)\angle 90^\circ}{11080\angle 25.4^\circ} = 4240\angle 64.6^\circ$$

$$= 1830 + j3830 \text{ ohms.}$$

The total impedance Z_t of the circuit is equal to the vector sum of R_g, X_{LI}, X_{C1}, and Z_{AB34}.

$$Z_t = 10{,}000 + j180{,}000 - j200{,}000 + 1830 + j3830$$

$$= 11{,}830 - j16{,}170$$

$$= 20{,}000\angle -53.7^\circ \text{ ohms}$$

$$I_t = \frac{E_g}{Z_t} \quad \frac{200}{20{,}000} \cdot 0.010 \text{ a}$$

$$E_{34} = I_t Z_{AB34}$$

$$= 0.010 \times 4240$$

$$= 42.4 \text{ v}$$

$$I_{LOAD} = \frac{E_{AB34}}{R_{LOAD}}$$

$$= \frac{42.4}{10{,}000} \quad 0.00424 \text{ a or } 4.24 \text{ ma.}$$

Further below resonance the current through the load is even less; and at 75 kc the load current drops to approximately 0.0024 milliampers.

The same relative decrease in current occurs through the load with a corresponding increase in frequency. Figure 4-10,B, is a graph of the load current vs frequency characteristic of the filter shown in figure 4-10,A.

The current through the 10 k-ohm load (fig. 4-10,A) for an arbitrarily chosen frequency of 92.5 kc may be determined as follows:

$$I_L = \frac{E_{3,4}}{R_L}, \qquad (4\text{-}10)$$

where I_L is the current through the load, $E_{3,4}$ is the voltage across the load, and R_L is the load resistance.

$$E_{3,4} = I_t Z_{AB34} \qquad (4\text{-}11)$$

where I_t is the total current supplied by the generator and Z_{AB34} is the combined impedance of the parallel circuit, C2L2, and the load, R_L.

$$I_t = \frac{E_g}{Z_t} \qquad (4\text{-}12)$$

where E_g is the voltage of the generator, and Z_t is the total impedance of the circuit.

Z_t may be determined by solving for the parallel impedance of C2 and L2, combining it with R_L, and then combining this impedance ($Z_{3,4}$) with the impedance of C1, L1, and R_g. First the parallel impedance, Z_{AB}, of C2 and L2 is equal to their product divided by their sum. At 92.5 kc

$$X_{L2} = 2\pi f L_2 = 6.28 \times 92.5 \times 10^3 \times 0.88 \times 10^{-3}$$

$$= 510 \angle 90^\circ \text{ ohms}$$

$$X_{C2} = \frac{1}{6.28 f C_2} = \frac{10^{12}}{6.28 \times 92.5 \times 10^3 \times 3180}$$

$$= 541 \angle -90^\circ \text{ ohms}$$

$$Z_{AB} = \frac{(X_{L2})(X_{C2})}{+jX_{L2} - jX_{C2}} = \frac{(510 \angle 90^\circ)(541 \angle -90^\circ)}{+j510 - j541}$$

$$= \frac{(510 \times 541) \angle 0^\circ}{-j31} \quad 8{,}900 \angle 90^\circ$$

The parallel impedance, Z_{AB34}, of Z_{AB} and R_L is equal to their product divided by their sum.

$$Z_{AB34} = \frac{Z_{AB} R_L}{R_L \oplus Z_{AB}} = \frac{(8{,}900 \angle 90^\circ)(10{,}000 \angle 0^\circ)}{10{,}000 + j8{,}900}$$

$$= \frac{(8{,}900 \times 10{,}000) \angle 90^\circ}{13{,}370 \angle 41.7^\circ}$$

$$= 6{,}670 \angle +48.3^\circ = 4{,}450 + j4{,}980$$

The total impedance, Z_t, of the circuit is equal to the vector sum of R_g, X_{L1}, X_{C1}, and Z_{AB34}.

At 92,5 kc

$$X_{L1} = 2\pi f L_1 = 6.28 \times 92.5 \times 10^3 \times 0.318 = 185{,}000$$

ohms, and

$$X_{C1} = \frac{1}{6.28 f C_1} = \frac{10^{12}}{6.28 \times 92.5 \times 10^3 \times 8.884}$$

$$= 194{,}000 \text{ ohms.}$$

$$Z_t = 10{,}000 + j185{,}000 - j194{,}000 + 4{,}450 + j4{,}980$$

$$= 4{,}450 + j4{,}980 + 10{,}000 - j9{,}000$$

$$= 14{,}450 - j4{,}020.$$

Expressed in the polar form,

$$Z_t = 15{,}000 \angle -15.6^\circ.$$

From equation (4-12), the total circuit current is

$$I_t = \frac{E_g}{Z_1} = \frac{200}{15{,}000} \quad 0.01335 \text{ a;}$$

from equation (4-11), the voltage across the load is

$$E_{3,4} = I_t Z_{AB34} = 0.01335 \times 6{,}670 = 89.0 \text{ v;}$$

and from equation (4-10), the load current is

$$I_L = \frac{E_{3,4}}{R_L} = \frac{89}{10,000} = 0.0089 \text{ amperes or } 8.9 \text{ ma.}$$

BAND-ELIMINATION FILTERS

A band-elimination filter (or band-suppression filter) is designed to suppress current of all frequencies within a continuous band, limited by the lower and upper cut-off frequencies, and to pass all frequencies below or above that band. A simple band-suppression filter is shown in figure 4-11,A. This type of filter is just the opposite of the band-pass filter; currents of frequencies within the band are greatly attenuated or weakened. The series- and parallel-tuned circuits are tuned to the center of the band to be eliminated. The parallel-tuned circuit in series with the source offers a high impedance to this band of frequencies, and the series-tuned circuit in shunt with the load offers very low impedance; therefore, the signals within the elimination band are both blocked and diverted from the load. All other currents—that is, currents at all frequencies outside the band-pass through the parallel circuit with very little opposition are unaffected by the series-tuned circuit since it acts as an open circuit at these frequencies.

Thus, for all frequencies outside the stop band the total circuit impedance is the sum of R_L and R_g or 20 k ohms (approximately). The total current is

$$I = \frac{E}{R_g + R_L} = \frac{200}{20,000} = 0.010 \text{ a or } 10 \text{ ma.}$$

The voltage across the load is

$$IR_L = 0.010 \times 10,000 = 100 \text{ v,}$$

and the load current is

$$I = \frac{E_L}{R_L} = \frac{100}{10,000} = 0.010 \text{ a or } 10 \text{ ma.}$$

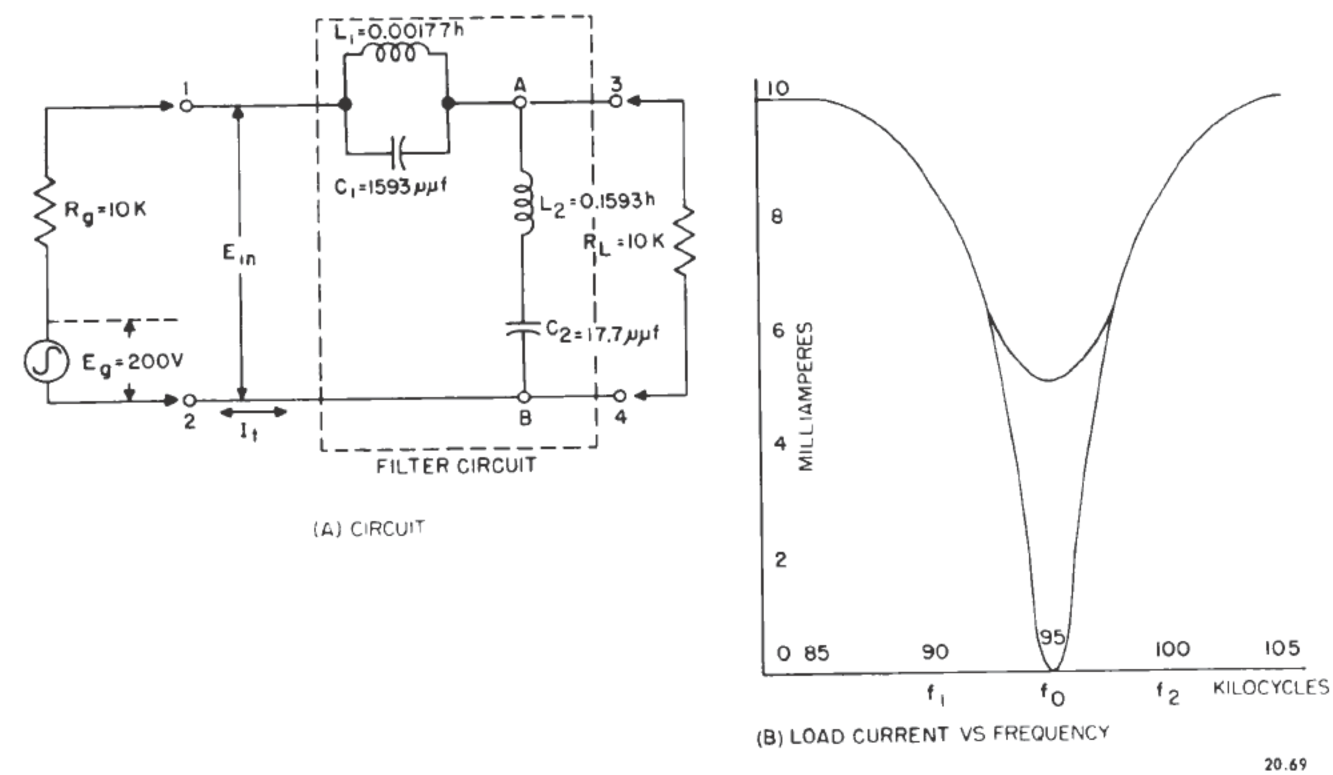


Figure 4-11.—Band-elimination filter.

Assume that a band of frequencies extending from 90 kc to 100 kc is to be suppressed by the filter. For an input and output resistance of 10,000 ohms, the values of inductance and capacitance are as indicated in the figure. The formulas by which these values are determined may be found in radio engineering handbooks.

In the example of figure 4-11 the following handbook formulas are used to find the magnitude of L1, L2, C1, and C2.

$$L_1 = \frac{R(2\pi f_2 - 2\pi f_1)}{(2\pi f_1)(2\pi f_2)} \quad \text{henry} \qquad (4\text{-}13)$$

$$C_1 = \frac{10^{12}}{R(2\pi f_2 - 2\pi f_1)} \quad \text{micromicrofarad} \qquad (4\text{-}14)$$

$$L_2 = \frac{R}{2\pi f_2 - 2\pi f_1} \quad \text{henry} \qquad (4\text{-}15)$$

$$C_2 = \frac{(2\pi f_2 - 2\pi f_1)10^{12}}{(2\pi f_1)(2\pi f_2)R} \quad \text{micromicrofarad} \qquad (4\text{-}16)$$

Substituting the following values in 4-13 through 4-16,

$$f_2 = 10\text{x}10^4 \text{ cps}$$

$$f_1 = 9\text{x}10^4 \text{ cps}$$

$$R = R_g = R_L = 10^4 \text{ ohms}$$

$$L_1 = \frac{10^4(2\pi)(10\text{x}10^4 - 9\text{x}10^4)}{(2\pi)(2\pi)(10\text{x}10^4\text{x}9\text{x}10^4)} = 0.00177 \text{ h}$$

$$C_1 = \frac{10^{12}}{10^4(2\pi)(10\text{x}10^4 - 9\text{x}10^4)} = 1{,}590 \quad \mu\mu\text{f}$$

$$L_2 = \frac{10^4}{2\pi(10\text{x}10^4 - 9\text{x}10^4)} = 0.1593 \text{ h}$$

$$C_2 = \frac{2\pi(10\text{x}10^4 - 9\text{x}10^4)10^{12}}{2\pi\text{x}2\pi\text{x}10\text{x}10^4\text{x}9\text{x}10^4\text{x}10^4} = 17.7 \quad \mu\mu\text{f}$$

These formulas are not derived in this text because of their length.

At resonance (95 kc) the parallel circuit, L1C1, offers maximum impedance and may be considered as almost an open circuit. At the same frequency the series circuit, L2C2, in effect short-circuits the load, so that minimum current will flow through the load at the resonant frequency. As in the case of the band-pass circuit, the inherent series resistance of the two tuned circuits is small and can be neglected. The calculations for the magnitude of the load current at 96 kc are given in the following paragraphs:

$$f_o = 95 \text{ kc or } 95\text{x}10^3 \text{ cps}$$

$$L1 = 0.00177\text{h or } 1.77\text{x}10^{-3} \text{ h}$$

$$X_{L1} = 2\pi f_o L_1 = 6.28\text{x}95\text{x}10^3\text{x}1.77\text{x}10^{-3}$$
$$= 1{,}058 \text{ ohms}$$

$$C1 = 1{,}593 \ \mu\mu\text{f or } 1{,}593\text{x}10^{-12} \text{ farad}$$

$$X_{C1} = \frac{1}{2\pi f_o C_1} = \frac{10^{12}}{6.28\text{x}95\text{x}10^3\text{x}1{,}593}$$
$$= 1{,}050 \text{ ohms}$$

$$L2 = 0.1593\text{h or } 159.3\text{x}10^{-3}\text{h}$$

$$X_{L2} = 2\pi f_o L_2 = 6.28\text{x}95\text{x}10^3\text{x}159.3\text{x}10^{-3}$$
$$= 95{,}000 \text{ ohms}$$

$$C2 = 17.7 \ \mu\mu\text{f or } 17.7\text{x}10^{-12} \text{ farad}$$

$$X_{C2} = \frac{1}{2\pi f_o C_2} = \frac{10^{12}}{6.28\text{x}95\text{x}10^3\text{x}17.7}$$
$$= 94{,}700 \text{ ohms}$$

The parallel impedance, Z_{1A}, of L1C1 is equal to the vector product divided by the vector sum of the impedance of the two branches.

$$Z_{1A} = \frac{X_{L1} X_{C1}}{X_{L1} \oplus X_{C1}} = \frac{(1058\angle 90°)(1{,}050\angle -90°)}{+j1{,}058 - j1{,}050}$$

$$= \frac{1{,}110{,}000\angle 0°}{8\angle 90°} \quad 139{,}000\angle -90° \text{ ohms.}$$

The series impedance Z_{AB} of L2C2 is the vector sum of the impedance of L2 and C2.

$X_{L2} = +j95{,}000$ ohms

$X_{C2} = j94{,}700$ ohms (slide rule calculations are approximate)

$$Z_{AB} = X_{L2} \oplus X_{C2} = +j95{,}000 - j94{,}700$$

$$= 300\angle 90^\circ \text{ or } +j300 \text{ ohms.}$$

The parallel impedance Z_{AB34} of Z_{AB} and R_L is their vector product divided by their vector sum.

$$Z_{AB34} = \frac{Z_{AB}R_L}{R_L \oplus Z_{AB}} = \frac{(300\angle 90^\circ)(10{,}000\angle 0^\circ)}{10{,}000+j300}$$

$$= 300\angle 90^\circ$$

The total impedance Z_t of the entire circuit is equal to the vector sum of R_g, Z_{1A}, Z_{AB34}.

$$Z_t = R_g \oplus Z_{1A} \oplus Z_{AB34}$$

$$= 10{,}000 + 139{,}000\angle -90^\circ \;\; 300\angle 90^\circ$$

$$= 10{,}000 - j139{,}000 + j300$$

$$= 10{,}000 - j138{,}700$$

$$= 139{,}000\angle -85.8^\circ \text{ ohms.}$$

The phase angles are neglected in the following calculations in order to simplify the results. The total circuit current, I_t, at 95 kc is

$$I_t = \frac{E}{Z_t} = \frac{200}{139{,}000} \quad 0.00144 \text{ or } 1.44 \text{ ma.}$$

The load voltage E_{34} is equal to the product of the total current and the parallel impedance Z_{AB34}. Thus,

$$E_{34} = I_t Z_{AB34}$$

$$= 0.00144 \times 300$$

$$= 0.432 \text{ v.}$$

The load current is equal to the load voltage, E_{34}, divided by R_L.

$$I_L = \frac{E_{34}}{R_L} = \frac{0.432}{10{,}000} \quad 0.0000432 \text{ or } 0.0432 \text{ ma.}$$

From these calculations it is seen that the load current at the center frequency of 95 kc is reduced to a small fraction of the normal load current of 10 ma outside the band. The percent of normal current is $\frac{0.0432}{10} \times 100$ 0.432 percent.

Below resonance—for example, at 90 kc-the impedances of L1C1 and L2C2 are such that with an assumed voltage of 200 volts about 8.9 milliamperes flow through the load, R_L. At 85 kc (10 kc below resonance) the current through the load is increased to approximately 9.9 milliamperes.

The band-suppression characteristic is symmetrical about the resonant frequency, and the same relative increase in current with increase in frequency may be assumed. Figure 4-11,B, is an indication of the current-frequency characteristic of the filter shown in figure 4-11,A. The current flows through the load, R_L.

The calculations for the load current at frequencies other than resonance are carried out in the same manner as are those for resonance. The important points to remember in carrying out these calculations are that inductive reactance varies directly with the frequency, and capacitive reactance varies inversely with the frequency.

The lower side of the stop band is 90 kc, and the upper side is 100 kc. The calculations for load current at 90 kc are given in the following paragraphs:

$$f_1 = 90 \text{ kc or } 90{,}000 \text{ cycles; } L_1 = 1.77\times10^{-3}\text{h;}$$

$$C_1 = 1{,}593\ \mu\mu\text{f or } 1{,}593\times10^{-12} \text{ farads}$$

$$X_{L1} = 2\pi f_1 L_1 = 6.28\times90\times10^3\times1.77\times10^{-3}$$

$$= 1{,}000 \text{ ohms}$$

$$X_{C1} = \frac{1}{2\pi f_1 C_1} = \frac{10^{12}}{6.28 x 90 x 10^3 x 1,593}$$
$$= 1,110 \text{ ohms}$$

$L2 = 0.1593h$; $C2 = 17.7\ \mu\mu f$ or $17.7 x 10^{-12}$ farads

$$X_{L2} = 2\pi f_1 L_2 = 6.28 x 90 x 10^3 x 0.1593$$
$$= 90,000 \text{ ohms}$$

$$X_{C2} = \frac{1}{2\pi f_1 C_2} = \frac{10^{12}}{6.28 x 90 x 10^3 x 17.7}$$
$$= 100,000 \text{ ohms.}$$

The parallel impedance Z_{1A} of L1C1 at 90 kc is the vector product divided by the vector sum of the impedances of the two branches. Thus,

$$Z_{1A} = \frac{X_{L1} X_{C1}}{X_{L1} \oplus X_{C1}} = \frac{(1,000\angle 90^\circ)(1,110\angle -90^\circ)}{+j1,000 - j1,110}$$

$$= \frac{(1,000 x 1,110)\angle 0^\circ}{110\angle -90^\circ} \quad 10,000\angle 90^\circ$$ (Slide calculations are approximate.)

The series impedance, Z_{AB}, of L2C2 at 90 kc is equal to the vector sum of the impedance of L2 and C2. Thus,

$$Z_{AB} = +jX_{L2} - jX_{C2} = +j90,000 - j100,000$$

$$= j10,000 \text{ or } 10,000\angle -90^\circ.$$

The parallel impedance, Z_{AB34}, of Z_{AB} and the load, R_L, is equal to their vector product divided by their vector sum. Thus,

$$Z_{AB34} = \frac{Z_{AB} R_L}{R_L \oplus Z_{AB}} = \frac{(10,000\angle -90^\circ)(10,000\angle 0^\circ)}{10,000 - j10,000}$$

$$= \frac{10^8 \angle -90^\circ}{14,140\ \ -45^\circ} \quad 7,070\angle -45^\circ$$

$$= 5,000 - j5,000.$$

The total circuit impedance, Z_t, is equal to the vector sum of R_g, Z_{1A}, and Z_{AB34}. Thus,

$$Z_t = R_g \oplus Z_{1A} \oplus Z_{AB34}$$

$$= 10,000 + j10,000 + 5,000 - j5,000$$

$$= 15,000 + 5,000$$

$$= 15,850\angle 18.4^\circ.$$

The total circuit current at 90 kc is

$$I_t = \frac{E}{Z_t} = \frac{200\angle 0^\circ}{15,850\angle 18.4^\circ} = 0.0126 \angle -18.4^\circ \text{ a.}$$

The load voltage, E_{34}, is equal to the product of the total current and the parallel impedance Z_{AB34}.

$$E_L = I_t Z_{AB34}$$

$$= 0.0126 x 7,070$$

$$= 89.4$$

The load current, I_L, is equal to the load voltage divided by R_L.

$$I_L = \frac{E_L}{R_L} = \frac{89.4}{10,000} = 0.00894 \text{ a or } 8.94 \text{ ma.}$$

From these calculations it is seen that the load current has increased to a value that is within a few percent of the normal load current of 10 ma that will flow outside the band. The percent of load current that will flow at 90 kc is equal to $\frac{8.94}{10}$ x 100 or 89.4 percent.

The current curve is symmetrical about the resonant frequency of 95 kc. At 90 kc and 100 kc the load current is 8.94 ma, and at 85 kc and 105 kc the load current is approximately 9.94 ma.

The Q of the coils and capacitors has the same effect on the shape of the current curve of the band-stop filter as it has on the band-pass filter previously described. Thus, for low Q coils and capacitors the curve will not dip as far, and the slope of the sides will not be as steep as for high Q coils and capacitors.

WAVE TRAPS

Wave traps, sometimes used in the antenna circuits of radio receivers, are forms of band-elimination filters. The two general types of wave traps are the parallel-tuned filter and the series-tuned filter. These traps are used to prevent interference, for example, from a nearby station that is strong enough to be heard over the entire frequency band of the receiver. The trap reduces the signal strength from the unwanted station so that it will not be heard except when the receiver is tuned to that station.

The parallel circuit, in series with the antenna in figure 4-12,A, is tuned to resonance at the frequency of the undesired signal.

The parallel wave trap presents a high impedance to currents of this unwanted frequency and allows currents of other frequencies to enter the receiver with only slight attenuation.

The series circuit, connected as shown in figure 4-12,B, is tuned to resonance at the frequency of the undesired signal. The impedance of the series circuit, C1L1, at resonance is low. Hence, these unwanted frequencies will be bypassed to ground around the receiver input transformer primary, L3. The desired frequencies will be essentially unaffected because either L, or C, act as a high impedance when not in resonance.

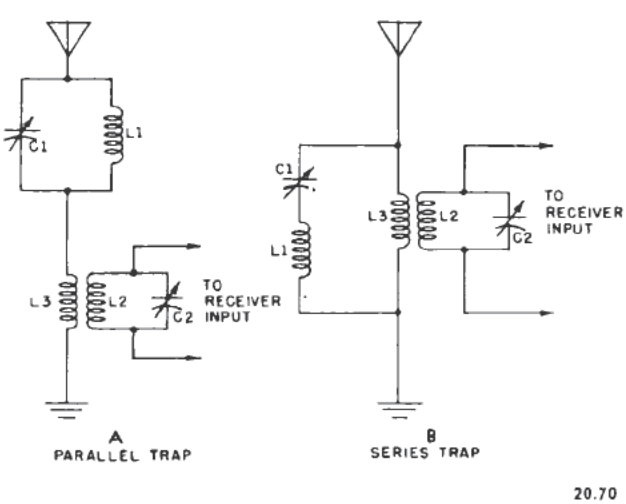


Figure 4-12.—Wave traps.

INDUCTIVELY COUPLED TUNED CIRCUITS

When two separate circuits (fig. 4-13,A) are so positioned that energy from one circuit is coupled to the other circuit by transformer action, they are said to be inductively coupled. Mutual inductance is the common property of the two circuits that determines for a particular frequency and current in one of the circuits the amount of mutually induced voltage in the other circuit. Mutual inductance may be expressed in henries and is designated by the letter, M. It is described in *Basic Electricity*, NavPers 10086-A.

The following relation exists between the current and induced voltage in an inductively coupled circuit (fig. 4-13,A).

$$E_s = 2\pi f M I_p$$

Where E_s is the secondary voltage, I_p the primary current, M the mutual inductance, and f the frequency. This relation is similar to that for self-induced voltage E_{IND} described in *Basic Electricity*, NavPers 10086-A, as

$$E_{IND} = 2\pi f L I$$

where E_{IND} is the self induced voltage, I is the current, L is the self inductance, and f is the frequency.

Although problems involving inductively coupled circuits may be somewhat complex, they may be simplified if the following assumptions are made: (1) The effect of the presence of the coupled secondary on the primary is the same as if an impedance, Z_C, in figure 4-13,B (called the coupled impedance) had been added in series with the primary. Z_C can be calculated in terms of the mutual inductance, the frequency, and the secondary impedance as follows:

$$Z_c = \frac{(2\pi f M)^2}{Z_s}$$

Where M is the mutual inductance, Z_s (a vector, figure 4-13,C) is the impedance of the secondary (this formula will be derived shortly). (2) The secondary voltage, E_s, induced by the primary current, I_p, has a value of $2\pi f M I_p$ and lags the primary current by 90° (fig. 4-13,D). (3) The

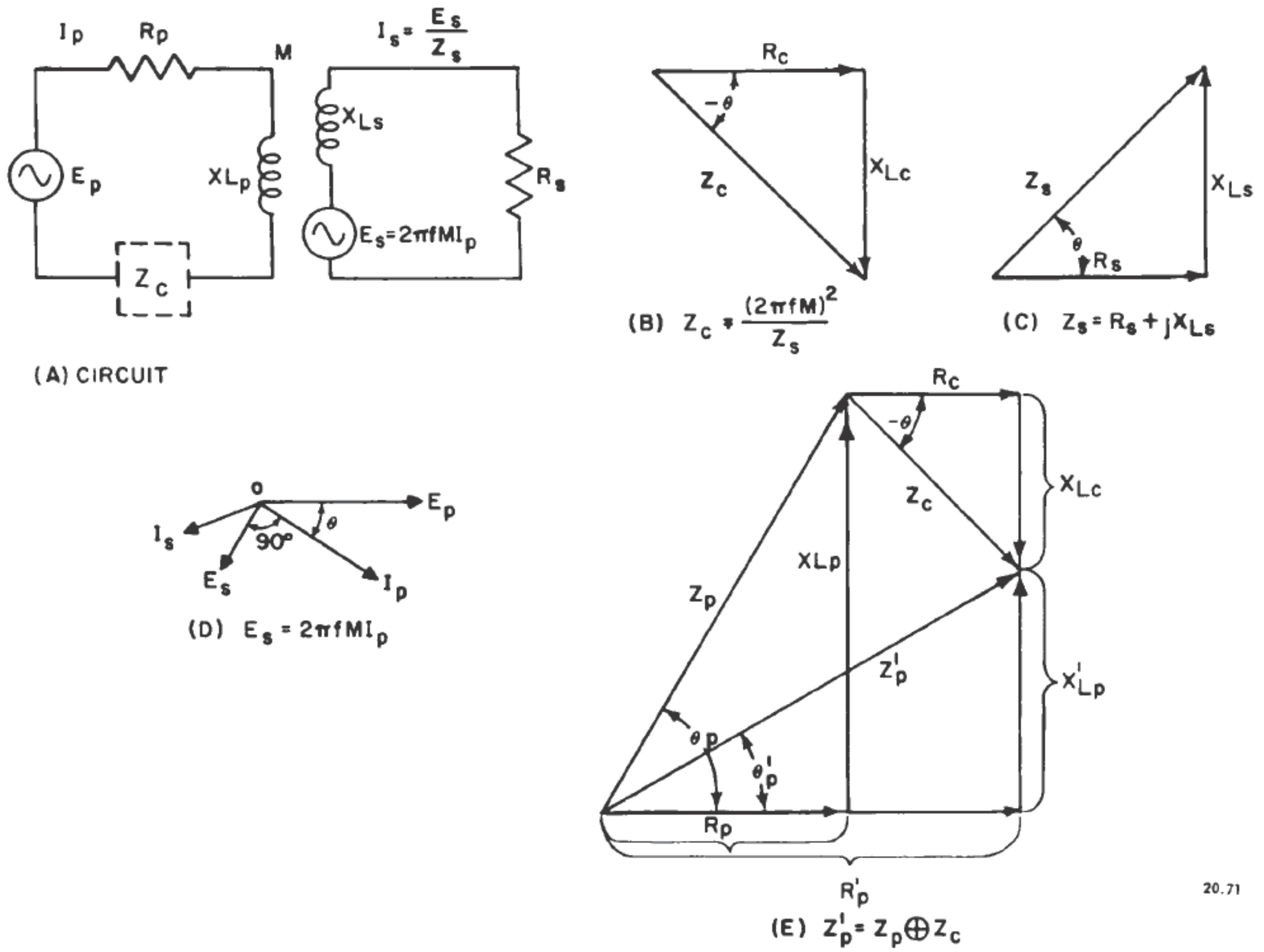


Figure 4-13.—Coupled impedance.

secondary current, I_S, is that value of current that would flow if the primary were removed and the induced secondary voltage were applied in series with the secondary coil (fig. 4-13,A). The coupled impedance, Z_C, is a vector quantity. It has the same phase angle as the secondary impedance, Z_S, but is of opposite sign (fig. 4-13,B and C). The total impedance Z_p' of the primary is the vector sum of the primary impedance, Z_p, with the secondary removed, and the coupled impedance, Z_C. Thus, $Z_p' = Z_p \oplus Z_C$ (fig. 4-13,E).

The coupled impedance is derived from the relation that a voltage, E_{MP}, is induced in the primary by the secondary current, I_s (just as a voltage is induced in the secondary by the primary current). The ratio of the voltage induced in the primary (by the presence of the secondary) to the primary current is called the coupled impedance. Thus,

$$Z_C = \frac{E_{MP}}{I_p} \tag{4-17}$$

The mutual voltage induced in the primary by the presence of the secondary is equal to the product of the secondary current, the mutual inductance, the frequency, and 2π. This product is similar to that for the self-induced voltage, E_{IND} in a coil, except that M is substituted for L. Thus,

$$E_{MP} = 2\pi fMI_S. \tag{4-18}$$

Substituting 4-18 in 4-17,

$$Z_C = \frac{2\pi f M I_s}{I_p} \qquad (4\text{-}19)$$

Applying Ohm's law to the secondary,

$$I_s = \frac{E_s}{Z_s}, \qquad (4\text{-}20)$$

and substituting 4-20 in 4-19,

$$Z_C = \frac{2\pi f M \frac{E_s}{Z_s}}{I_p} \qquad (4\text{-}21)$$

The mutually induced secondary voltage is equal to the product of the primary current, the mutual inductance, the frequency, and 2π. This product is similar to that for the mutually induced primary voltage due to the presence of the secondary, as expressed in equation 4-18, except that E_{MP} becomes E_s and I_s becomes I_p. Thus,

$$E_s = 2\pi f M I_p. \qquad (4\text{-}22)$$

Substituting 4-22 in 4-21 and simplifying,

$$Z_C = \frac{(2\pi f M)(2\pi f M I_p)}{Z_s I_p}$$

$$= \frac{(2\pi f M)^2}{Z_s} \qquad (4\text{-}23)$$

The primary current, secondary voltage, and secondary current (fig. 4-13) may be determined by Ohm's law as applied to a-c circuits. Thus the primary current, I_p, may be determined as

$$I_p = \frac{E}{Z_p \oplus \frac{(2\pi f M)^2}{Z_s}},$$

where E is the voltage applied to the primary. The secondary voltage has been stated as $2\pi f m I_p$. The secondary current, I_s, is determined as

$$I_s = \frac{2\pi f M I_p \angle +90}{Z_s}.$$

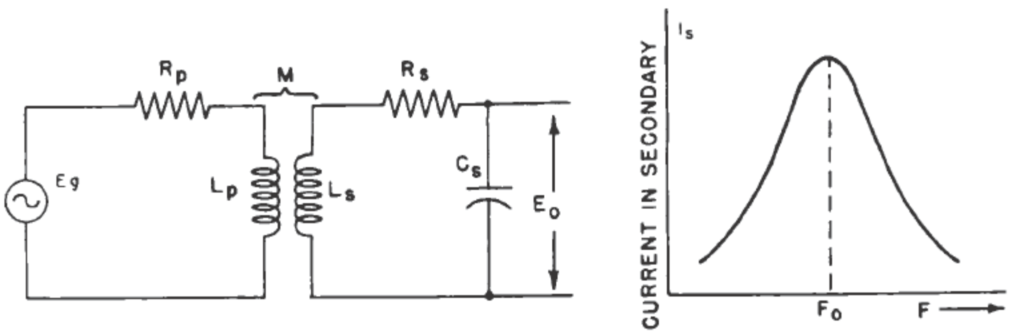
From the expression for coupled impedance $\frac{(2\pi f M)^2}{Z_s}$ some of the characteristics of a coupled circuit can be determined. For example, if M is large and Z_s is small, the coupled impedance will be large and the primary current may be reduced as a result of the increased impedance when the coupling occurs. The voltage induced in the secondary and the secondary current, will be correspondingly affected. If, on the other hand M is small and Z_s is large, the effect of the presence of the secondary on the primary is slight and little change occurs in primary current with increased coupling. A special case in which the secondary is tuned is of importance because it is widely used in r-f voltage amplifiers in radio receivers.

UNTUNED-PRIMARY TUNED-SECONDARY CIRCUIT

A simplified untuned-primary tuned-secondary circuit is shown in figure 4-14,A.

Ordinarily the voltage source connected to the primary circuit, has a resistance, R_g acting in series with the primary, L_p, and the coupled impedance.

If the coupling is small and the source resistance is large with respect to the coupled impedance (as in the case of some amplifiers), the primary current is substantially independent of the resonant condition of the secondary. The curve of secondary current vs frequency (fig. 4-14,B) then has the same general shape as that of the ordinary series-tuned circuit of figure 4-7. If a low-resistance source is used in place of the high-resistance source the coupled impedance becomes an appreciable part of the primary impedance. At resonance the secondary impedance is low and resistive, and the coupled impedance causes a dip in primary current. At frequencies away from resonance the secondary impedance increases, and the coupled impedance is less, causing an increase in primary current. This increases the voltage induced in the secondary by transformer action at these off-frequency



A
UNTUNED-PRIMARY TUNED-SECONDARY CIRCUIT

B
CURRENT IN SECONDARY VS FREQUENCY

20.72

Figure 4-14.—Untuned-primary tuned-secondary circuit and response curve.

points and prevents the secondary current from falling off as rapidly as it would if the secondary induced voltage were constant. Thus, the curve is broader and indicates a lower effective Q for the series-tuned circuit than would exist if the source resistance were relatively large and the primary current and secondary mutually induced voltage were independent of the resonant condition of the secondary.

The formula of the coupled impedance, which is responsible for the shape of the characteristic curve of an untuned-secondary circuit, is as previously stated.

$$\text{Coupled impedance} = \frac{(2\pi fM)^2}{Z_s}$$

$$= \frac{(2\pi fM)^2}{R_s \pi j2\pi fL_s - j\frac{1}{2\pi fC_s}}$$

where R_S is the secondary resistance, $+j2\pi fL_S$, the secondary inductive reactance, and $-j\frac{1}{2\pi fC_S}$, the secondary capacitive reactance.

In the vicinity of resonance, $(2\pi fM)^2$ varies only slightly. The denominator represents the series impedance of the secondary. The coupled impedance formula, $\frac{(2\pi fM)^2}{Z_s}$, is similar in form to that of the previously derived impedance formula (4-7) of a parallel resonant circuit, $\frac{(2\pi fL)^2}{R_s}$. Thus, the coupled impedance due to the tuned secondary varies with frequency according to the same mathematical law as the impedance of a parallel circuit varies with frequency. In the case of coupled impedance, however, the magnitude of the curve depends on the mutual inductive reactance, $2\pi fM$, instead of the inductive reactance, $2\pi fL$.

TUNED-PRIMARY TUNED-SECONDARY CIRCUIT

A simplified tuned-primary tuned-secondary circuit and current vs frequency response curves are shown in figure 4-15.

As indicated by the response curves (fig. 4-15, B and C), this type of circuit has a band-pass characteristic that depends in part on the coefficient of coupling k and in part on the circuit Q's.

As in the band-pass filter previously described, essentially uniform amplification of a relatively narrow band of frequencies may be achieved, and amplification of frequencies outside this band may be sharply reduced.

Because of resistance in the circuit the slope of the response curve is not perfectly vertical. The circuit cannot completely discriminate against frequencies just outside the desired channel without also attenuating to some extent the frequencies at the upper and lower limits of the pass band. However, double-tuned circuits approach an ideal band-pass characteristic much more closely than do single-tuned circuits, which have rounded response curves.

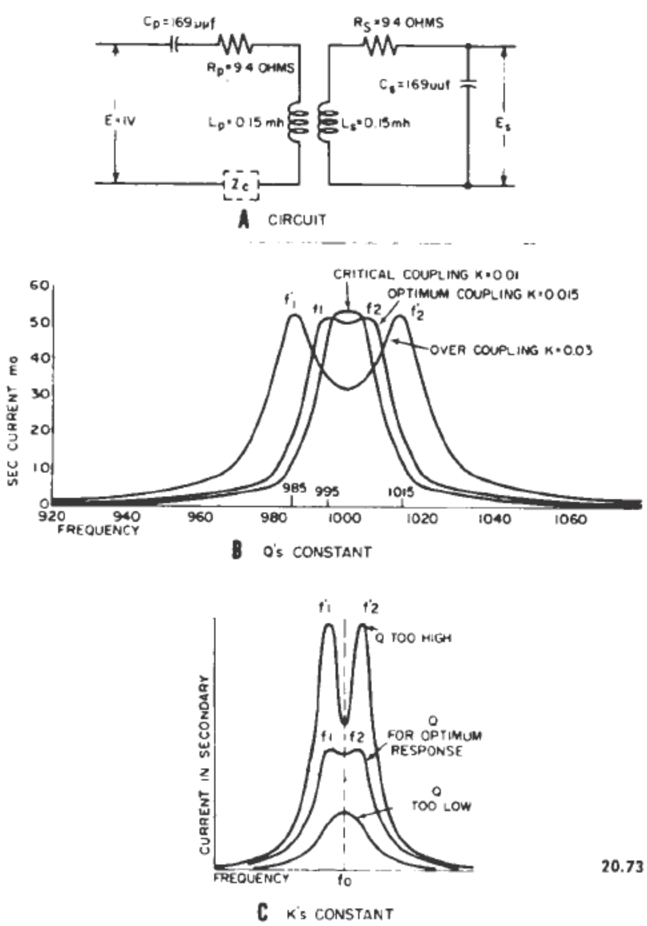


Figure 4-15.—Tuned-primary tuned-secondary circuit and response curve.

The double-tuned circuits develop a resonant buildup of current in both primary and secondary in the vicinity of resonance. This action maintains the secondary induced voltage and current at a higher value than if the secondary alone were tuned. The explanation of this effect is given in more detail later in this chapter and again in chapter 6.

When the coefficient of coupling (fig. 4-15,B) is low, the response is sharply peaked at the resonant frequency and the pass band is very narrow. This action is the result of a low resistance high Q secondary and the reduction in influence of the secondary on the primary.

The coupled impedance is low and the primary current and secondary induced voltage are essentially independent of the magnitude of the secondary current.

As the coupling is increased to the critical value, maximum current flows in the secondary, and the output voltage across the secondary is also at its maximum. At this point (critical coupling), the energy transferred from the primary to the secondary is maximum. For this condition the coupled impedance (resistance) at the resonant frequency is equal to the primary impedance (resistance). Expressed as an equation,

$$\frac{(2\pi fM)^2}{R_s} = R_p \tag{4-24}$$

$$(2\pi fM)^2 = R_p R_s. \tag{4-25}$$

Note that

$$M = k\sqrt{L_p L_s}. \tag{4-26}$$

Substituting (4-26) in (4-25),

$$(2\pi fk\sqrt{L_p L_s})^2 = R_p R_s.$$

Transposing and simplifying,

$$(2\pi fL_p)(2\pi fL_s)k^2 = R_p R_s$$

$$\frac{(2\pi fL_p)(2\pi fL_s)}{R_p R_s} = \frac{1}{k^2} \tag{4-27}$$

Note that

$$\frac{2\pi fL_p}{R_p} = Q_p \text{ and } \frac{2\pi fL_s}{R_s} = Q_s. \tag{4-28}$$

Substituting (4-28) in (4-27),

$$Q_p Q_s = \frac{1}{k^2} \tag{4-29}$$

from which

$$k = \frac{1}{\sqrt{Q_p Q_s}}, \tag{4-30}$$

and if the Q's are equal,

$$k = \frac{1}{Q}. \quad (4\text{-}31)$$

The pass band is still relatively narrow.

If the coupling is further increased until the OPTIMUM value is reached, the gain is still relatively high; but the band pass has been increased and the output is essentially uniform. At this point (optimum coupling),

$$k = \frac{1.5}{\sqrt{Q_p Q_s}}, \text{ (approximate)},$$

and if the Q's are equal

$$k = \frac{1.5}{Q}.$$

As the coupling is again increased, the humps at f_1 and f_2 are well defined, and the gain at resonance is considerably reduced. Although the pass band is now much wider, the gain throughout the band is not sufficiently uniform.

The two humps, f_1 and f_2, in the curve are due to the reactance, $\frac{(2\pi f M)^2}{Z_s}$, that is coupled into the primary as the coupling is increased. Below resonance this reactance is inductive, and above resonance it is capacitive, For the same frequencies, the coupled reactance has the opposite sign to that of the primary, and the impedance of the primary is therefore reduced. Accordingly, there is an increase in primary current at frequencies slightly off resonance; and this results in increased current in the secondary, and also an increase in voltage at output.

The frequencies at the two humps, f_1 and f_2, which define the practical lower and upper limits of the pass band, are determined by the following equations:

$$f_1 = \frac{f_o}{\sqrt{1+k}}$$

$$f_2 = \frac{f_o}{\sqrt{1-k}}$$

Figure 4-15,C, shows the effects of varying the Q while maintaining a constant coefficient of coupling. Actually, the desired response curve could be achieved by the proper manipulation of both k and Q because they are interrelated.

From the foregoing equations it is seen that in order to have a wide pass band, k must be large, and the circuit Q's must be small. However, the proper relation between k and Q is essential if the desired bandwidth and the desired response within the band are to be maintained.

The following example is an exercise in coupled circuit calculations presented at this time to increase your understanding of coupled circuit theory. The example is based on the values given in the tuned-primary tuned-secondary transformer coupled circuit in figure 4-15,A. Table 4-3 lists the nomenclature used in the calculations, and table 4-4 lists the calculated values of resistance, reactance, impedance, and current for certain frequencies and coupling coefficients.

The resonant frequency of the tuned circuits may be verified by substituting the given values of L and C in the formula,

$$f_o = \frac{1}{2\pi\sqrt{LC}}$$

where

$$L = 0.15\text{mh or } 0.15\text{x}10^{-3}\text{h}$$

and

$$C = 169\mu\mu\text{ f or } 169.0\text{x}10^{-12} \text{ farads.}$$

$$f_o = \frac{1}{2\text{x}3.14\sqrt{0.15\text{x}10^{-3}\text{x}169\text{x}10^{-12}}}$$

$$= \frac{1}{6.28\sqrt{1.50\text{x}10^{-4}\text{x}169\text{x}10^{-12}}}$$

$$= \frac{1}{6.28\sqrt{254\text{x}10^{-16}}}$$

$$= \frac{0.159\text{x}10^{8}}{0.159\text{x}10^{2}}$$

$$= 10^{6} \text{ cps or 1,000 kc.}$$

The inductive reactance of the primary and secondary coils at the resonant frequency of 1000 kc is found by substituting the given values of

Table 4-3.—Nomenclature used in the Example of Figure 4-15,A.

Input volts, E = 1 volt	Sec. cap. react. (ohms), $X_{Cs} = \frac{1}{2\pi fC_s}$
Pri. ind., $L_p = 0.15\text{x}10^{-3}$ henry	Mutual ind. (henries), $M = k\sqrt{L_pL_s}$
Sec. ind., $L_s = 0.15\text{x}10^{-3}$ henry	Mut. ind. react. (ohms), $X_M = 2\pi fM$
Pri. cap., $C_p = 169\text{x}10^{-12}$ farad	Pri. resist. (ohms), $R_p = \frac{X_{Lp}}{Q_p}$
Sec. cap., $C_s = 169\text{x}10^{-12}$ farad	Sec. resist. (ohms, $R_s = \frac{X_{Ls}}{Q_s}$
Pri. quality, $Q_p = 100$	Pri. imped. (without sec), $Z_p = R_p+jX_{Lp}-jX_{Cp}$
Sec. quality, $Q_s = 100$	Sec. imped. (ohms), $Z_s = R_s+jX_{Ls}-jX_{Cs}$
Resonant freq., $f_o = 10^6$ cps or 1000 kc	Coupled imped. (ohms), $Z_c = \frac{X_M{}^2}{Z_s}$
Coefficient of coupling, k = 0.01; 0.015; 0.03	Total pri. imped. (ohms), $Z_{pt} = Z_p \oplus Z_c$
Pri. ind. reactance, $X_{Lp} = 2\pi fL_p$	Pri. current (amperes), $I_p = \frac{E}{Z_{pt}}$
Sec. reactance, $X_{Ls} = 2\pi fL_s$	Sec. induced volts, $E_s = X_MI_p$
Pri. capacitive reactance, $X_C = \frac{1}{2\pi fC_p}$	Sec. current (amperes), $I_s = \frac{E_s}{Z_s}$

frequency and inductance in the formula for inductive reactance as follows:

$$X_{Lp} = X_{Ls} = 2\pi fL_s = 6.28\text{x}10^6\text{x}0.15\text{x}10^{-3}$$
$$944 \text{ ohms.}$$

The capacitive reactance of the primary and secondary capacitors is found by substituting the given values of frequency and capacitance in the formula for capacitive reactance as follows:

$$X_{Cp} = X_{Cs} = \frac{1}{2\pi fC_s} = \frac{1}{6.28\text{x}10^6\text{x}169\text{x}10^{-12}}$$
$$= 944 \text{ ohms.}$$

The resistance of the primary and secondary circuits is determined by substituting the given value of Q and the calculated value of inductive reactance in the formula $Q = \frac{X_L}{R}$, and transposing for R as follows:

$$X_L = 944 \text{ ohms and } Q = 100$$

$$R_p = R_s = \frac{X_{Ls}}{Q_s} = \frac{944}{100} = 9.44 \text{ ohms.}$$

The mutual inductance M which exists between the two coils when the coefficient of coupling k is equal to 0.01 is determined by substituting the specified value of k and the given values of primary and secondary inductance in the formula for mutual inductance as follows:

$$L_p = L_s = 0.15\text{x}10^{-3}\text{ h}; k \quad 0.01$$

$$M = k\sqrt{L_p L_s}$$

$$= 0.01\sqrt{0.15\text{x}10^{-3}\text{x}0.15\text{x}10^{-3}}$$

$$= 0.0015\text{x}10^{-3}\text{ h.}$$

The inductive (mutual) reactance, X_M, is determined by substituting the calculated value of M and the given value of resonant frequency in the formula for inductive (mutual) reactance X_M as follows:

$$M = 0.0015\text{x}10^{-3}\text{ h}; f = 10^6\text{ cps}$$

$$X_M = 2\pi fM = 2\text{ x}3.14\text{x}10^6\text{x}0.0015\text{x}10^{-3}$$

$$= 9.44\text{ ohms.}$$

Table 4-4.—Summary of values in example of figure 4-15 A.

Given	k	0.01	0.015		0.03		
	f	$1\text{x}10^6$ cps (1000 kc)		995 kc	1000 kc	985 kc	1015 kc
Find	X_{L_p} ohms	+j944		+j937.6	+j944	+j928	+j956
	X_{C_p} ohms	-j944		-j947	-j944	-j956	-j928
	R_p ohms	9.44					
	Z_p ohms	9.44∠0°		9.4-j9.4 13.3∠-45°	9.44∠0°	9.44-j28 29.5∠-71.4°	9.44+j28 29.5∠71.4°
	X_{L_s} ohms	+j944		+j937.6	+j944	+j928	+j956
	X_{C_s} ohms	-j944		-j947	-j944	-j956	-j928
	R_s ohms	9.44					
	Z_s ohms	9.44∠0°		9.44-j9.4 13.3∠-45°	9.44∠0°	9.44-j28 29.5∠-71.4°	9.44+j28 29.5∠71.4°
	M henry	$1.5\text{x}10^{-6}$	$2.25\text{x}10^{-6}$		$4.5\text{x}10^{-6}$		
	X_M ohms	9.44	14.14	14.08	28.28	27.8	28.7
	Z_C ohms	9.44∠0°	21.2∠0°	10.5+j10.5 14.88∠45°	84.7∠0°	8.32+j24.9 26.2∠71.4°	8.9-j26.4 27.8∠-71.4
	Z_{pt} ohms	18.88∠0°	30.6∠0°	19.9-j1.1 20∠-3°	94.1∠0°	17.82-j3.1 18∠-9.9°	18.3+j1.6 18.37∠5°
	I_p amperes	0.0530	0.0327	0.05	0.0106	0.0555	0.0545
	E_s volts	0.0500	0.463	0.704	0.3	1.54	1.56
	I_s amperes	0.0530	0.049	0.0528	0.0318	0.0522	0.053

The secondary impedance Z_s at the resonant frequency of 1000 kc is equal to the vector sum of X_{L_s}, X_{C_s}, and R_s. Expressed in vector algebra form

$$Z_s = R_s \; jX_{L_s} - jX_{C_s}.$$

Substituting the calculated values of $R_s = 9.44$, $S_{L_s} = 944$, and $X_{C_s} = 944$,

$$Z_s = 9.44 + j944 - j944$$

$$= 9.44 \angle 0^\circ \text{ ohms (resistive).}$$

The coupled impedance, Z_C, is found by substituting the calculated value of X_M at the resonant frequency and the calculated value of secondary impedance at the resonant frequency in the formula for coupled impedance as follows:

$X_M = 9.4_24$ ohms; $Z_s = 9.44$ ohms.

$$Z_C = \frac{X_M}{Z_s} \; \frac{(9.44)^2}{9.44} = 9.44 \text{ ohms (resistive).}$$

The total primary impedance, Z_{pt}, at the resonant frequency is the vector sum of the primary impedance, Z_p (without the secondary) and the coupled impedance, Z_C. Note that $Z_p = Z_s = 9.44$ ohms since $L_p = L_s$ and $Q_p = Q_s$.

$$Z_{pt} = Z_p \oplus Z_C$$

$$= 9.44 \;\; 9.44 = 18.88 \text{ ohms (resistive).}$$

The primary current, I_p, is equal to the input voltage, E, divided by the total impedance of the primary circuit,

$$I_p = \frac{E}{Z_{pt}} = \frac{1}{18.88} \;\; 0.0530 \text{ a.}$$

The secondary mutually induced voltage, E_s, is determined by substituting the calculated values of primary current and mutual inductive reactance in the formula for mutually induced voltage as follows:

$X_M = 9.44$ ohms; $I_p = 0.053$ a

$$E_s = X_M I_p = 9.44 \times 0.0530 = 0.500 \text{ v.}$$

The secondary current at the resonant frequency is equal to the calculated value of the secondary voltage divided by the calculated value of secondary impedance as follows:

$E_s = 0.500$ v; $Z_s = 9.44$ ohms (resistive)

$$I_s = \frac{E_s}{Z_s} = \frac{0.500}{9.44} = 0.0530 \text{ a.}$$

The calculated values of X_{L_p}, X_{L_s}, X_{C_p}, X_{C_s}, R_p, R_s, Z_p, Z_s, Z_{pt}, I_p, E_s, and I_s are given in the first row of table 4-2.

The frequencies corresponding to the peaks of secondary current in the vicinity of resonance are determined for two specified coefficients of coupling. First, for $k = 0.015$ (approximate optimum coupling) the lower frequency, f_1, is found by substituting the specified value of $k = 0.015$, and the resonant frequency, $f_o = 10^6$ cps in the following formula:

$$f_1 = \frac{f_o}{\sqrt{1+k}} = \frac{10^6}{\sqrt{1+0.015}} = \frac{1 \times 10^6}{1.007}$$

$$= 0.995 \times 10^6 \text{ cps or 995 kc}$$

The upper frequency, f_2, is found by substituting $k = 0.015$ and $f_o = 10^6$ cps in the following formula:

$$f_2 = \frac{f_o}{\sqrt{1-k}} = \frac{10^6}{\sqrt{1-0.015}} = \frac{10^6}{\sqrt{0.985}} = \frac{10^6}{0.994}$$

$$= 1.005 \times 10^6 \text{ cps or 1,005 kc.}$$

For the second specified value of $k = 0.03$ (over coupling) the lower frequency, f_1', and the upper frequency, f_2', are found as follows:

$$f_1' = \frac{f_o}{\sqrt{1+k}} = \frac{10^6}{1+0.03} = \frac{10^6}{1.015}$$

$$= 0.985 \times 10^6 \text{cps or 985 kc}$$

and

$$f_2' = \frac{f_o}{\sqrt{1-k}} = \frac{10^6}{\sqrt{1-0.03}} = \frac{10^6}{\sqrt{0.97}} = \frac{10^6}{\sqrt{0.985}}$$

$$= 1.015 \times 10^6 \text{ cps or 1,015 kc.}$$

The calculations for secondary current at the resonant frequency when the coefficient of coupling is $k = 0.015$ are similar to those previously given for $k = 0.01$. The mutual inductance increases from 1.5×10^{-6} to 2.25×10^{-6} henry; the inductive (mutual) reactance X_M increases from 9.44 ohms to 14.14 ohms; the coupled impedance Z_C increases from 9.44 ohms to 21.2 ohms; the primary current I_p, decreases from 0.0530 ampere to 0.0327 ampere; and the secondary current decreases from 0.0530 ampere to 0.049 ampere.

The calculations when the frequency is decreased from 1000 kc to 995 kc show corresponding changes in reactance and coupled impedance as follows. The primary and secondary inductive reactances are

$$X_{L_p} = X_{L_s} = 2\pi f L_s = 6.28 \times 0.995 \times 10^6 \times 0.15 \times 10^{-3}$$

$$= 937.6 \text{ ohms.}$$

The primary and secondary capacitive reactances are

$$X_{C_p} = X_{C_s} = \frac{1}{2\pi f C_s} = \frac{1}{6.28 \times 0.995 \times 10^6 \times 169 \times 10^{-12}}$$

$$= 947 \text{ ohms.}$$

The secondary impedance is

$$Z_s = R_s + jX_{L_s} - jX_{C_s}$$

$$= 9.44 + j937.6 - j947$$

$$= 9.44 - j9.40$$

$$= 13.3\angle -44.9^\circ \text{ or } 13.3\angle -45^\circ.$$

The mutual inductance when $k = 0.015$ is

$$M = k\sqrt{L_p L_s} = 0.015\sqrt{0.15 \times 10^{-3} \times 0.15 \times 10^{-3}}$$

$$= 0.015 \times 15 \times 10^{-3} = 0.00225 \times 10^{-3} \text{ h.}$$

and the mutual inductive reactance at 995 kc is

$$X_M = 2\pi f M = 6.28 \times 0.995 \times 10^6 \times 0.00225 \times 10^{-3}$$

$$= 14.08 \text{ ohms.}$$

The coupled impedance at 995 kc and $k = 0.015$ is

$$Z_C = \frac{(X_M)^2}{Z_s} = \frac{(14.08)^2}{13.3\angle -45^\circ} = 14.88\angle 45^\circ \text{ ohms}$$

$$= 10.51 + j10.51 \text{ ohms,}$$

and the total primary impedance is

$$Z_{pt} = R_p + jX_{L_p} - jX_{C_p} \oplus Z_C$$

$$= 9.44 \;\; j937.6 - j947 + 10.51 + j10.51$$

$$= 19.95 - j1.11$$

$$= 20\angle -3.2^\circ.$$

The primary current at 995 kc and $k = 0.015$ is

$$I_p = \frac{E}{Z_{pt}} = \frac{1}{20} = 0.05 \text{ a,}$$

and the secondary voltage is

$$E_s = X_M I_p = 14.08 \times 0.05 = 0.704 \text{ v;}$$

the secondary current is

$$I_s = \frac{E_s}{Z_s} = \frac{0.704}{13.3} = 0.0528 \text{ a.}$$

The mutual inductance when $k = 0.03$ is

$$M = k\sqrt{L_p L_s} = 0.03\sqrt{0.15 \times 10^{-3} \times 0.15 \times 10^{-3}}$$

$$= 0.0045 \times 10^{-3} \text{ h.}$$

The mutual inductive reactance at 1000 kc is

$$X_M = 2\pi f M = 6.28 \times 10^6 \times 0.0045 \times 10^{-3}$$

$$= 28.28 \text{ ohms.}$$

The primary and secondary impedance from previous calculations are $Z_p = Z_s = 9.44$ ohms.

The coupled impedance at 1000kc and k = 0.03 is

$$Z_C = \frac{(X_M)^2}{Z_s} = \frac{(28.28)^2}{9.44} = 84.7\angle 0^\circ \text{ ohms},$$

and the total primary impedance is

$$Z_{pt} = Z_p \oplus Z_C$$
$$= 9.44 + 84.7$$
$$= 94.14\angle 0^\circ \text{ ohms.}$$

The primary current at 1000 kc when k = 0.03 is

$$I_p = \frac{E}{Z_{pt}} = \frac{1}{94.1} = 0.1062 \text{ ampere;}$$

the secondary voltage is

$$E_s = X_M I_p$$
$$= 28.28 \times 0.01062$$
$$= 0.3 \text{ volt;}$$

and the secondary current is

$$I_s = \frac{E_s}{Z_s} = \frac{0.3}{9.44} = 0.0318 \text{ a.}$$

The secondary current at the lower frequency $f_1' = 985$ kc when k = 0.03 is found as follows: The primary and secondary inductive reactances at 985 kc are

$$X_{L_p} = X_{L_s} = 2\pi f L_s = 6.28 \times 0.985 \times 10^6 \times 0.15 \times 10^{-3}$$
$$= 928 \text{ ohms.}$$

The primary and secondary capacitive reactances at 985 kc are

$$X_{C_p} = X_{C_s} = \frac{1}{2\pi f C_s} = \frac{1}{6.28 \times 0.985 \times 10^6 \times 169 \times 10^{-12}}$$
$$= 956 \text{ ohms.}$$

The secondary impedance at 985 kc is

$$Z_s = R_s + jX_{L_s} - jX_{C_s}$$
$$= 9.55 + j928 - j956$$
$$= 9.44 - j28 = 29.5\angle -71.4^\circ.$$

The mutual inductance for k = 0.03 is from a previous calculation, M = 0.0045×10^{-3} henry. The mutual inductive reactance at 985 kc is

$$X_M = 2\pi f M = 6.28 \times 0.985 \times 10^6 \times 0.0045 \times 10^{-3}$$
$$= 27.8 \text{ ohms.}$$

The coupled impedance at 985 kc and k = 0.03 is

$$Z_C = \frac{(X_M)^2}{Z_s} = \frac{(27.8)^2}{29.5\angle -71.4^\circ} = 26.2\angle 71.4^\circ$$
$$= 8.32\ j24.9 \text{ ohms.}$$

The total primary impedance at 985 kc and k = 0.03 is

$$Z_{pt} = Z_p \oplus Z_C$$
$$= R_p + jX_{L_p} - jX_{C_p} \oplus Z_C$$
$$= 9.44 + j928 - j956 + 8.32 + j24.9$$
$$= 17.76 - j3.1 = 18\angle -9.9^\circ.$$

The primary current at 985 kc and k = 0.03 is

$$I_p = \frac{E}{Z_{pt}} = \frac{1}{18} = 0.0555\text{a;}$$

the secondary induced voltage is

$$E_s = X_M I_p = 27.8 \times 0.055 = 1.54 \text{ volts;}$$

and the secondary current is

$$I_s = \frac{E_s}{Z_s} = \frac{1.54}{29.5} = 0.0522 \text{ a.}$$

The secondary current at the upper frequency $f_2' = 1015$ kc when k = 0.03 is found as follows:

The primary and secondary inductive reactances are

$$X_{L_p} = X_{L_s} = 2\pi f L_s = 6.28 \times 1.015 \times 10^6 \times 0.15 \times 10^{-3}$$

$$= 956 \text{ ohms.}$$

The primary and secondary capacitive reactances are

$$X_{C_p} = X_{C_s} = \frac{1}{2\pi f C_s} = \frac{1}{6.28 \times 1.015 \times 10^6 \times 169 \times 10^{-12}}$$

$$= 928 \text{ ohms.}$$

The secondary impedance is

$$Z_s = R_s + jX_{L_s} - jX_{C_s}$$

$$= 9.44 \; j956 - j928$$

$$= 9.44 \; j28 = 29.5 \angle 71.4^\circ \text{ ohms.}$$

The mutual inductance when k = 0.03 is from previous calculations, M = 0.0045×10^{-3} henry.

The mutual inductive reactance at 1015 kc is

$$X_M = 2\pi f M = 6.28 \times 1.015 \times 10^6 \times 0.0045 \times 10^{-3}$$

$$= 28.7 \text{ ohms.}$$

The coupled impedance when k = 0.03 and f_2 = 1015 kc is

$$Z_C = \frac{(X_M)^2}{Z_s} = \frac{(28.7)^2}{29.5} = 27.8 \angle -71.4^\circ \text{ ohms}$$

$$= 8.9 - j26.4 \text{ ohms.}$$

The total primary impedance is

$$Z_{pt} = Z_p \oplus Z_C = R_p + jX_{L_p} - jX_{C_p} \oplus Z_C$$

$$= 9.44 \; j956 - j928 + 8.9 - j26.4$$

$$= 18.3 + j1.6 \text{ ohms}$$

$$= 18.37 \angle 5^\circ \text{ ohms.}$$

The primary current is

$$I_p = \frac{E}{Z_{pt}} = \frac{1}{18.37} = 0.0545 \text{ a;}$$

the secondary voltage is

$$E_s = X_M I_p = 28.7 \times 0.0545 = 1.56 \text{ v;}$$

and the secondary current is

$$I_s = \frac{E_s}{Z_s} = \frac{1.56}{29.5} = 0.053 \text{ a.}$$

The application of inductively coupled circuits in electron-tube amplifiers, together with an amplifier voltage-gain problem is treated in chapter 6.

CHAPTER 5

INTRODUCTION TO ELECTRON-TUBE AMPLIFIERS

BASIC TRIODE AMPLIFIER

The most important function of an electron tube is its ability to amplify or increase the amplitude of the input signal. An electron-tube amplifier consists of one or more tubes and associated circuit elements necessary for its operation and is used to increase the voltage, current, or power of a signal For example, a minute amount of power at the input of a broadcast receiver is amplified by a number of amplifier stages in the receiver to the level necessary to operate a loudspeaker. The amount of amplification or gain that results is dependent primarily upon the number of amplifier stages used. GAIN is defined as the ratio of output to input. The greater the number of stages the greater the overall gain will be. One stage may have a larger gain (gain per stage) than another. In many circuits the gain is dependent largely upon the amplification afforded by the electron tube. In other circuits the gain may be due to transformer action or the resonant qualities of a circuit. The amplification of the tube itself is expressed as an amplification factor, μ (defined in chapter 1).

A signal voltage, e_s, of sine waveform applied to the control grid of a tube (fig. 5-1,A) results in plate current variations through the load impedance and voltage variations between plate and ground, as shown in figure 5-1,B. The voltages and currents are made up of a d-c component that exists when no signal is present and an a-c component that exists in addition to the d-c component when a signal is applied to the grid. In most cases the a-c component is of chief interest although the d-c component determines the portion of the tube characteristic in which the operation occurs. The a-c components of plate voltage and current constitute the useful output of the tube.

TRIODE CHARACTERISTICS

The action of the triode amplifier may be explained by studying the i_p-e_p curves for a triode, figure 5-2. These curves are approximately parallel straight lines. The dotted portions are projected to the X axis to obtain the values of plate voltage for which the grid voltage of any given curve will reduce the plate current to zero. These values are approximations, based on the assumption that the i_p-e_p curves are parallel straight lines.

In the simplified triode circuit (fig. 5-2,A) there is no a-c signal present and the plate load resistor R_L of figure 5-1,A has been omitted. Grid bias E_c and plate voltage E_b are adjustable. However, the grid is always maintained negative with respect to the cathode and the plate is always positive with respect to the cathode. Thus the grid always repels electrons and the plate always attracts them. (Like charges repel, unlike charges attract.)

The no-signal condition is represented at "O" (fig. 5-2,B). The grid bias is -6 volts, the plate current is 0.008 ampere and the plate voltage is 250 volts. The triode plate resistance is from the triangle (fig. 5-2,B)

$$r_p = \frac{\Delta ep}{\Delta ip} = \frac{300 - 250}{0.012 - 0.008} = \frac{50}{0.004} = 12{,}500 \text{ ohms}$$

The triode amplification factor is (from the same triangle)

$$\mu = \frac{\Delta ep}{\Delta ip} = \frac{300 - 250}{8 - 6} = \frac{50}{2} = 25$$

Assuming the $i_p e_p$ curves are parallel straight lines the equation of each of these curves is

$$e_p = -\mu e_g + i_p r_p \qquad (5\text{-}1)$$

where e_p is the plate voltage, μ the amplification factor of the triode, e_g the instantaneous grid voltage, i_p the plate current, and r_p the plate resistance. The equation states that the voltage

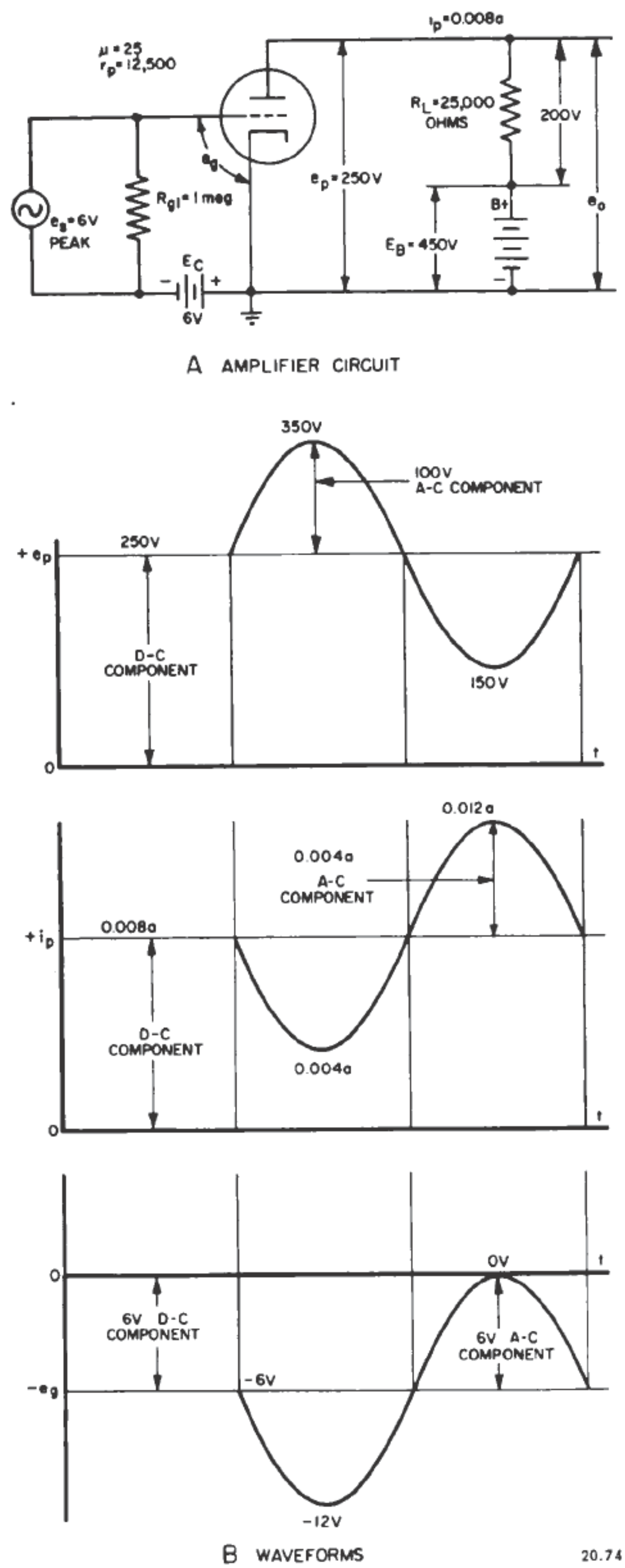


Figure 5-1.—Triode amplifier and waveforms.

between plate and cathode is equal to the sum of the i_pr_p drop across the plate resistance and another voltage $-\mu e_g$.

The term $-\mu e_g$ is equivalent to counter voltage acting in the plate circuit against the plate current flow. The plate current is the movement of electrons from cathode to plate; electrons are negative, the grid is negative, and like charges repel. Thus the counter voltage is a real voltage acting in the plate circuit against the flow of plate current and is equal to the product of the triode amplification factor and the instantaneous grid to cathode voltage.

In the example of figure 5-2, $\mu = 25$, $r_p = 12{,}500$ ohms, $i_p = 0.008$ ampere and $e_g = -6$ volts. Substituting these values in equation 5-1 and solving for plate voltage,

$$e_p = -25\ (-6) + 0.008 \times 12{,}500 = 250 \text{ volts.}$$

Note the counter voltage is 150 volts and the i_pr_p drop in the plate resistance is 100 volts; the sum of the two voltages comprises the plate-cathode voltage of 250 volts.

PROJECTED CUTOFF

From equation 5-1, when i_p is reduced to zero,

$e_p = -\mu e_g$ and transposing for e_g,

$$e_g = -\frac{e_p}{\mu} \qquad (5\text{-}2)$$

This value of grid bias is called projected cutoff or cutoff bias. In the example under consideration the projected cutoff (co) bias required to reduce plate current to zero when the plate voltage is 250 volts and the amplification factor is 25, is

$$co = -\frac{e_p}{\mu} = \frac{-250}{25} = -10 \text{ volts}$$

The cutoff bias increases directly with plate voltage as shown in figure 5-2,B. When e_p is 300 volts, the cutoff bias is -12 volts; when e_p is 400 volts, the cutoff bias is -16 volts, and when e_p is 500 volts the cutoff bias is -20 volts.

CIRCUIT ANALYSIS

Now study the circuit if figure 5-1,A, to see how the triode operates when a plate load resistor R_L is inserted in series with the

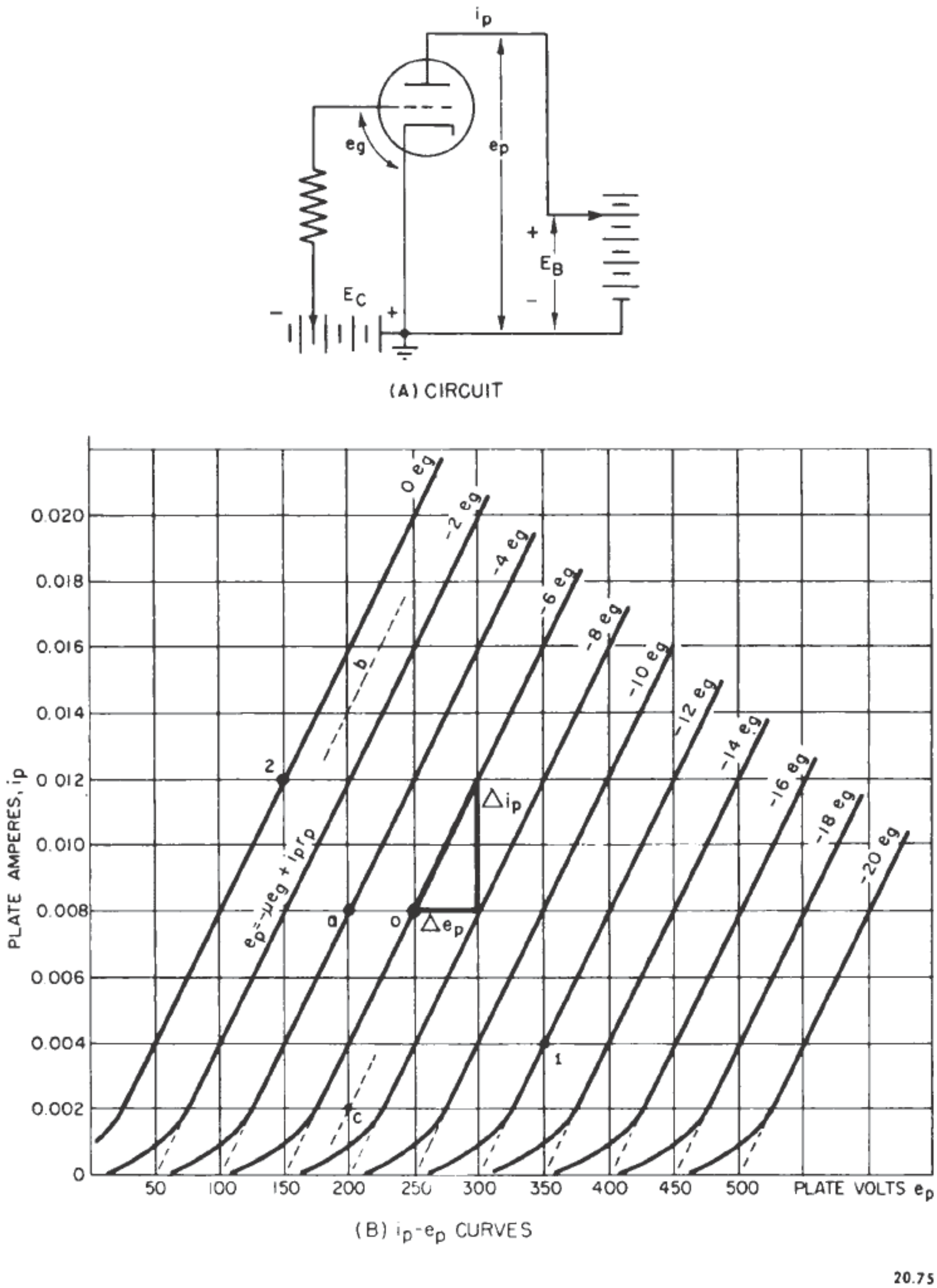


Figure 5-2.—Triode characteristics.

plate lead and a sine waveform signal e_s is applied across the grid leak resistor R_{g1}.

First consider the plate circuit. The algebraic sum of the voltages around the plate circuit is zero. In other words the B supply voltage is equal to the sum of the plate voltage and the voltage across the plate load resistor R_L. Expressed as an equation,

$$E_B = e_p + i_p R_L \tag{5-3}$$

Substituting the value of e_p from (5-1) in 5-3)

$$E_B = -\mu e_g + i_p r_p + i_p R_L \tag{5-4}$$

Substituting the values for the no-signal condition point "O" figure 5-2, B (μ is 25, e_g is -6 volts, i_p is 0.008 ampere, r_p is 12,500 ohms, R_L is 25,000 ohms) and solving for E_B,

$$E_B = -25(-6) + 0.008 \times 12{,}500 + 0.008 \times 25{,}000$$
$$= 150 + 100 + 200$$
$$= 450 \text{ volts.}$$

The B supply voltage provides 450 volts which is distributed around the plate circuit with 250 volts between plate and ground and 200 volts across R_L when the plate current is 0.008 ampers and the grid voltage is -6 volts This condition exists as the a-c signal is about to increase from the -6 volt level in a negative direction (fig. 5-1,B).

Now assume that the grid input signal increases to a peak of 6 volts with the grid end of the source NEGATIVE (fig. 5-1,A). the instantaneous grid-cathode voltage will be

$$e_g = E_c + e_s \text{ or } -6 -6 \text{ or } -12 \text{ volts.}$$

Transpose equation (5-4) and solve for plate current as follows:

$$i_p = \frac{E_B + \mu e_g}{r_p + R_L} \qquad (5\text{-}5)$$

Substituting the following values in equation (5-5):

E_B = 450 volts, μ = 25, e_g = -12 volts,

r_p = 12,500 ohms and R_L = 25, 000 ohns,

$$i_p = \frac{450 + 25\,(-12)}{12{,}500 + 25{,}000}$$
$$= \frac{450 - 300}{37{,}500}$$
$$= 0.004 \text{ ampere}$$

With a plate current of 0.004 ampere (fig. 5-1,A) the drop across R_L is 0.004 x 25,000 or 100 volts and the plate voltage increases from 250 volts to 450-100 or 350 volts (point 1 fig. 5-2,B). This swing in plate voltage comprises the output signal that occurs as the input signal, e_s, to the grid changes from zero to -6 volts. Note the peak a-c signal output component is 350-250 or 100 volts.

Now consider the action of the triode amplifier when the a-c signal input voltage has reversed its polarity and increases to a peak of 6 volts with the grid end of the source POSITIVE. This is the second half cycle of e_s. Now the instantaneous grid-cathode voltage is 6-6 or zero volts. Substituting the following values in equation (5-5):

e_g = 0 volts, μ = 25, E_B = 450 volts,

r_p = 12,500 ohms and R_L = 25,000 ohms,

$$i_p = \frac{E_B + \mu e_g}{r_p + R_L}$$
$$= \frac{450 + 25 \times 0}{12{,}500 + 25{,}000}$$
$$= 0.012 \text{ ampere}$$

With a plate current of 0.012 ampere the voltage across R_L is 0.012 x 25,000 = 300 volts and the plate voltage decreases to 450 - 300 or 150 volts (point 2, fig. 5-2,B). This swing in plate voltage from 250 volts to 150 volts comprises the output voltage that occurs as the input signal to the grid changes from zero to +6 volts. Note the peak a-c signal output component is 250 - 150 or 100 volts.

Summarizing the action, the a-c signal acting in series with the bias voltage (fig. 5-1,B) swings the grid voltage from -6 volts to -12 volts back to -6 volts to zero volts, returning to -6 volts to complete one input cycle. During this time the plate current varies from 0.008 ampere to 0.004 ampere, to 0.008 ampere to 0.012 ampere returning to 0.008 ampere to complete the cycle. During the same time the plate voltage swings from 250 volts to 350 volts to 250 volts to 150 volts and back to 250 volts to complete the cycle. As stated earlier, the plate voltage variation comprises the output signal of the triode amplifier.

VOLTAGE GAIN

The voltage gain of the amplifier is the ratio of the a-c output voltage to the a-c input voltage. In this example the a-c output voltage is 100 volts (peak) and the a-c input voltage is 6 volts (peak). The voltage gain is $\frac{100}{6}$ or 16.67.

A simplified equivalent circuit for the triode amplifier eliminates the d-c voltages and shows only the a-c components. (See fig. 5-3.) The circuit contains a voltage μe_s acting in series with the plate resistance r_p and load resistance R_L. In this example μ is 25 and e_s is 6 volts (peak a-c signal input to the grid); the product μe_s is 25 x 6 or 150 volts (peak) acting in series with 12,500 ohms plate resistance and 25,000 ohms, plate load resistance, or a total of 37,500 ohms. The peak signal current is

$$i_p = \frac{\mu e_s}{r_p + R_L} = \frac{25 \times 6}{12{,}500 + 25{,}000}$$

$$= \frac{150}{37{,}500} = 0.004 \text{ a}$$

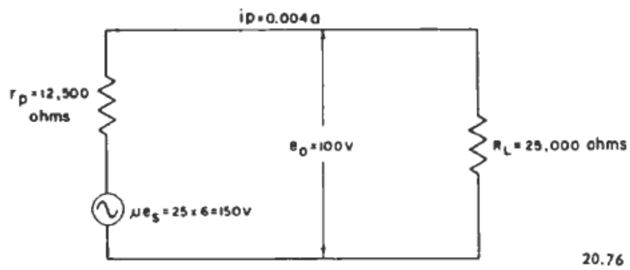


Figure 5-3.—Equivalent circuit of a triode amplifier.

The output voltage is the voltage across R_L and is equal to i_pR_L or 0.004 x 25,000 = 100 volts.

The voltage gain is $\frac{e_o}{e_s} = \frac{100}{6} = 16.67$

CATHODE BIAS

Point "O" (fig. 5-2,B) has been called the no-signal point for the triode amplifier. It is also called the operating point. In the example of figure 5-1,A, the bias voltage between grid and cathode is provided by the C battery.

This voltage may also be provided by the B battery. The arrangement consists of inserting a resistor in series with the cathode lead (fig. 5-4) and shunting it with a relatively large capacitor to stabilize the voltage across it when the plate-cathode current varies with the signal. This action established the operating bias for the triode. The grid is returned to the lower end of the cathode resistor through the grid leak resistor across which the a-c signal is developed.

A 750 ohm resistor connected between cathode and ground, and carrying a no-signal current of 0.008 ampere will have a voltage drop across it equal to 0.008 x 750 or 6 volts. The B supply voltage is increased from 450 to 456 volts to provide the 6 volt grid bias and maintain the same no-signal plate voltage of 250 volts.

GRID RESISTOR

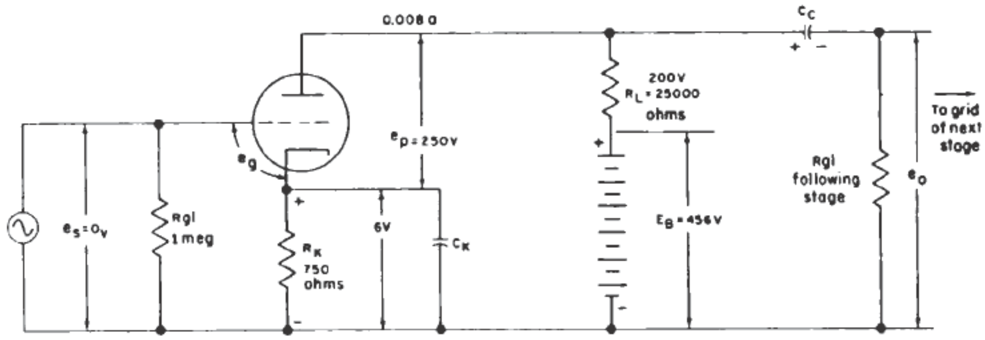
The grid (leak) resistor R_{g1}, figure 5-4 conducts the bias voltage developed across R_k to the grid. The charge on the grid is actually a static potential since no grid current flows. The magnitude of the resistance of R_{g1} should be large so as to minimize a-c signal currents and associated power loss and heating. Values of the order of 1 to 10 megohms are specified by the manufacturer. The upper limit is imposed to protect the tube in case it becomes gassy (loses its vacuum).

Gas particles (due to loss of vacuum) in the triode become ionized by electron bombardment from the cathode. As positive ions they surround the grid and attract electrons from the grid itself, causing a negative (reverse) grid current. This reverse grid current flows through R_{g1} in a direction to subtract the voltage across R_{g1} from that across R_k, thereby reducing the grid cathode bias voltage to such a low value that plate current may burn up the triode.

For example, if 0.6 microampere of reverse grid current flows through a 10 megohm grid leak resistor, the voltage drop across R_{g1} is 0.6 x 10 = 6 volts and this voltage will attempt to neutralize the effect of the 6 volt drop across R_k (fig. 5-4). A lower value of grid leak resistance, for example, 1 megohm would have only 0.6 x 1 = 0.6 volt in opposition to the 6 volt bias across R_k and so the effect on plate current would be much less pronounced.

COUPLING CAPACITOR

The coupling capacitor C_c (fig. 5-4) must have a high d-c resistance (low leakage) because it insulates the grid circuit of one stage from the plate circuit of the stage that is coupled to it. The d-c resistance of a coupling capacitor



20.77

Figure 5-4.—Triode amplifier with cathode bias and R-C coupled output.

should be at least 50 megohms. If the leakage becomes appreciable direct current will flow through R_{g1} of the following stage and the resulting voltage will impress a positive bias on the grid of the tube in that stage. This condition is especially important if an R-C coupled amplifier has good low frequency response (described later) because the grid leak resistor has relatively high resistance and the coupling capacitor has relatively high capacitance.

In practical applications, input and output coupling circuits must always be used with the electron tube. If resistance-capacitance coupling is used the voltage gain of the stage will be less than the amplification factor of the electron tube because of losses in the coupling elements. If transformer coupling is used, the voltage gain may be greater or less than μ depending on whether the coupling transformer has a step-up or step-down turns ratio.

In order to obtain certain waveform characteristics, amplifiers are sometimes purposely designed to distort the signal. When voltage or power is to be amplified without appreciably changing the shape of the wave, as in high-fidelity amplifiers, it is generally necessary to sacrifice some of the gain that the stage would normally have if this condition were not imposed.

Amplifiers may be classified in a number of ways according to use, bias, frequency response, or resonant quality of the load.

CLASSIFICATION OF AMPLIFIERS

ACCORDING TO USE

When classified according to use or type of service, amplifiers fall into two general type groups—VOLTAGE AMPLIFIERS and POWER AMPLIFIERS.

Voltage Amplifiers

Voltage amplifiers are so designed that signals of relatively small amplitude applied between the grid and the cathode of the tube will produce large values of amplified signal voltage across the load in the plate circuit. In order to produce the largest possible amplified signal voltage across the plate load (which may be a resistor, an inductor, or a combination of both) this value of impedance must be as large as practicable.

The GAIN of a voltage amplifier is the ratio of the a-c output voltage to the a-c input voltage. This type of amplifier is commonly used in radio receivers to increase the r-f or i-f signal to the proper level to operate the detector. It is used also to amplify the a-f output of the detector stage. In phone transmitters, voltage amplifiers are used to increase the output of the microphone to the proper level to be applied to the modulator.

Power Amplifiers

Power amplifiers are designed to deliver a large amount of power to the load in the plate circuit. Since power, in general, is equal to the voltages times the current, a power amplifier must develop across its load sufficient voltage to cause rated current to flow. The POWER AMPLIFICATION of such a circuit is the ratio of the output power to the input grid driving power.

The load impedance of a power amplifier is selected to give either maximum plate efficiency or maximum power output for a certain minimum level of distortion. PLATE EFFICIENCY is the ratio of useful output power (a-c voltage component times a-c current component times cos θ) to d-c input power to the plate (plate current times plate voltage).

POWER SENSITIVITY, another term used with power amplifiers is the ratio of the power output in watts to the grid signal voltage causing it, (when no grid current flows). If grid current flows, the term usually means the ratio of plate power output to grid power input. Pentodes require less driving power than triodes for the same power output and thus have a higher power sensitivity.

Power amplifiers are commonly used as the output stage of radio receivers. They are used also in transmitters to increase the power of the modulated carrier to the desired level before it is fed to the antenna.

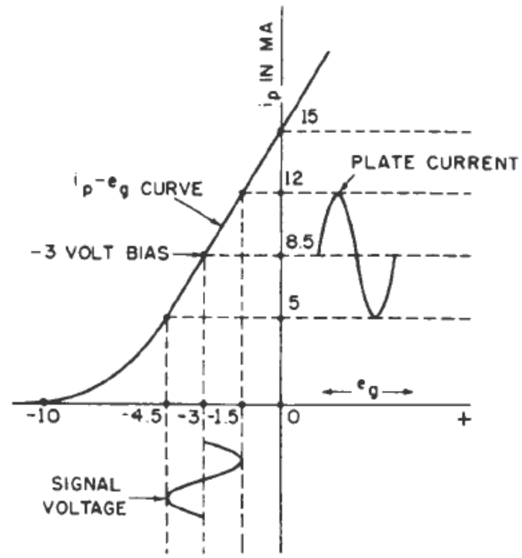


Figure 5-5.—Class A operation. 20.78

CLASSIFICATION ACCORDING TO BIAS

Amplifiers may be classified also according to the conditions under which the tube operates—that is, according to the portion of the a-c signal voltage cycle during which the plate current flows as controlled by the bias on the grid. The four classes of amplifier operation according to bias are A, B, AB, and C.

Class-A Amplifiers

Class A amplifiers are biased so that, with normal input signal, plate current flows during the entire input cycle, and the amplification is essentially linear, as indicated in figure 5-5. Grid current does not flow in most class A amplifiers.

To show that grid current does not flow during any part of the input cycle, the subscript "1" may be added to the letter or letters of the class identification. The subscript "2" may be used to indicate that grid current flows during some parts of the input cycle. Thus, if the grid is not driven positive at any time in the class A cycle no grid current will flow and the amplifier is designated class A_1.

The principal characteristics of class A amplifiers are minimum distortion, low power output for a given tube (relative to class B and class C amplifiers), high power amplification, and relatively low plate efficiency (20 to 35 percent). This type of amplifier finds wide use in various audio systems where low distortion is important.

Class-B Amplifiers

Class B amplifiers are biased so that no plate current flows when no signal is applied to the grid. Plate current then flows for approximately one-half of each cycle of grid signal voltage (fig. 5-6). Such amplifiers are characterized by medium power output, medium plate efficiency (50 to 60 percent), and moderate power amplification.

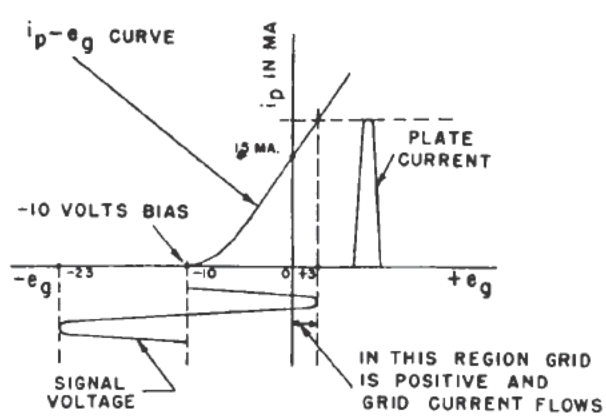


Figure 5-6.—Class B operation. 20.79

Single-ended (singletube) class B amplifiers are used in r-f amplifier stages having a parallel-tuned circuit as the plate load.

Class-AB Amplifiers

Class AB amplifiers have grid biases and input-signal voltages of such values that plate current flows for appreciably more than half the input cycle but for less than the entire cycle, as indicated in figure 5-7. Class AB operation is essentially a compromise between the low distortion of the class A amplifier and the higher efficiency of the class B amplifier.

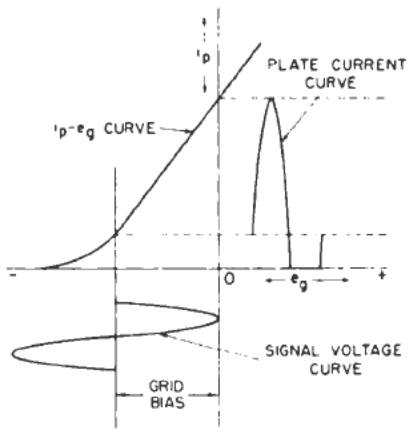


Figure 5-7.—Class AB operation.

If the input signal drives the grid positive with respect to the cathode, grid current will flow during the positive peaks and the amplifier is designated as a class AB_2 amplifier. Although a class AB_2 amplifier delivers slightly more power to its load, the class AB_1 amplifier has the advantage of presenting to its driver a constant impedance. In contrast with this effect the amplifier that draws grid current over a portion of its input cycle presents a changing impedance to its driver at the point where grid current starts to flow. Thus, before grid current the impedance may be relatively high, and during the part of the input cycle when grid current flows the impedance falls to a relatively low value. The driver that supplies this kind of load must be designed to supply undistorted power to the load during these periodic intervals of low impedance.

Class-C Amplifiers

Class C amplifiers have a bias that is appreciably greater than cutoff; consequently plate current flows for appreciably less than half of each cycle of the applied grid signal voltage (fig. 5-8). This class of amplifier has a relatively high plate efficiency (70 to 75 percent), high power output, and low power amplification.

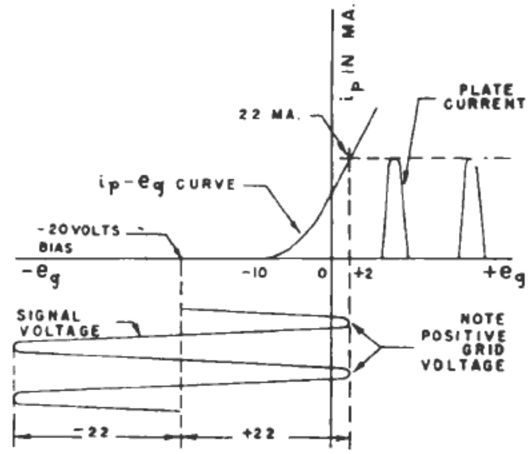


Figure 5-8.—Class C operation.

Class C amplifiers are not used as audio amplifiers, but they are used as r-f power amplifiers in transmitters. If power is delivered to a tuned load the load will present a high impedance at the resonant frequency and low impedance at other frequencies. If the load is tuned to the same frequency as that which is applied to the grid it will offer optimum loading at this frequency. Low impedance will be offered to the harmomics (multiples) of the frequency applied to the grid, and hence these undesirable components will be eliminated.

CLASSIFICATION ACCORDING TO FREQUENCY

Amplifiers may be classified according to the frequency range over which they are to operate. In general, amplifiers operating in these ranges are known as direct-current (d-c); audio-frequency (a-f); intermediate-frequency (i-f); radio-frequency (r-f); and video-frequency (v-f), or pulse, amplifiers.

When the signal current is in but one direction a d-c amplifier must be used. In order to overcome certain problems inherent in such an amplifier the circuits must be balanced and stablized by means of resistors.

Audio-frequency amplifiers operating in the range from 30 to 15,000 cycles per second may be transformer-coupled, impedance-coupled, or resistance-coupled.

I-f and r-f amplifiers are ordinarily designed for tuned-circuit coupling, although in actual operation they may resemble either transformer-coupled or impedance-coupled circuits.

Video-frequency amplifiers, which operate in a range extending from the lower audio frequencies to perhaps 5,000,000 cycles per second, commonly use resistance-coupled amplifiers in which the coupling resistance is made low enough to produce the necessary high-frequency response. However, in actual radar and television applications the resistance-coupled amplifier must be modified to make the response essentially flat over a wide range of frequencies. In addition, the circuits must be modified to keep time-delay distortion within a certain minimum value at the high- and low-frequency ends of the spectrum.

CLASSIFICATION ACCORDING TO CIRCUIT CONFIGURATION

Grounded-Cathode Amplifier

Amplifiers may be classified according to the connection of the tube elements in the circuit. Conventional electron-tube amplifier circuits return the cathode either to ground through a cathode resistor or, if separate bias is provided, to ground directly. Both of these circuit configurations are classed as grounded-cathode types. In the former, the cathode is positive with respect to ground by the amount of the grid bias and the cathode bypass capacitor holds the cathode at ground potential with respect to the signal component. A grounded-cathode amplifier circuit is shown in figure 5-9,A, for a separate heater cathode and in figure 5-9,B, for a direct heater cathode. In both of these circuits the interelectrode capacitance between plate and grid introduces feedback at high frequencies, and unstable amplifier operation results.

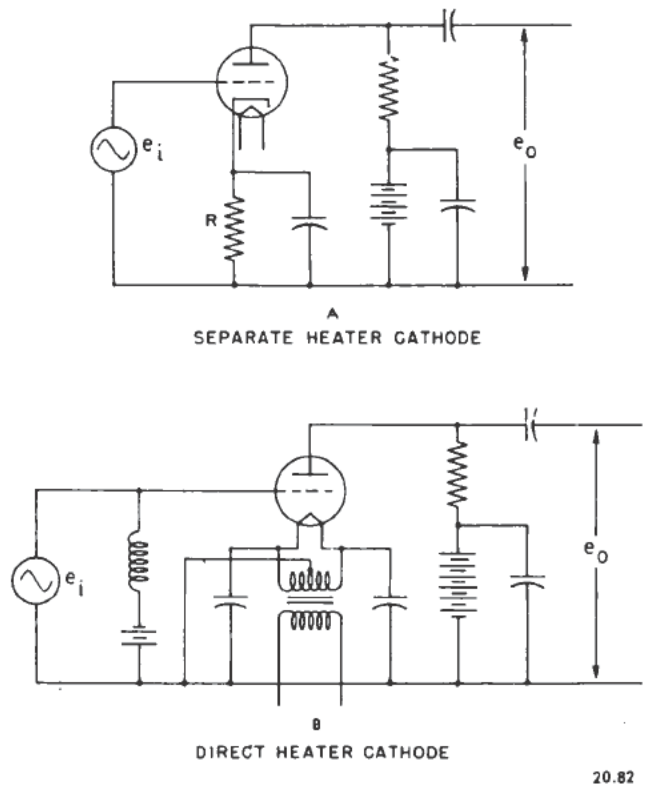


Figure 5-9.—Grounded-cathode amplifier circuits.

Grounded-Grid Amplifier

For very high frequencies the grounded-grid amplifier shown in figure 5-10 removes the feedback coupling between grid and plate and places the grid-plate interelectrode capacitance effectively in parallel with the load. Grounded-grid amplifiers are used as r-f amplifiers in the lower radar frequencies and in television circuits in the v-h-f and u-h-f bands. The input signal is introduced into the cathode circuit in series with the grid bias and varies the grid-to-cathode voltage in the normal manner.

The output signal is taken between the plate and ground. The plate current (including the a-c component) flows through the signal source which is in series with the cathode circuit. The signal source has appreciable impedance and the plate current through it is accompanied by a voltage drop across it which acts between the cathode and grid. The action is degenerative and lowers the gain of the amplifier compared to the gain of the grounded-cathode type. Some of the power in the load is supplied by the signal source since the load and source are in

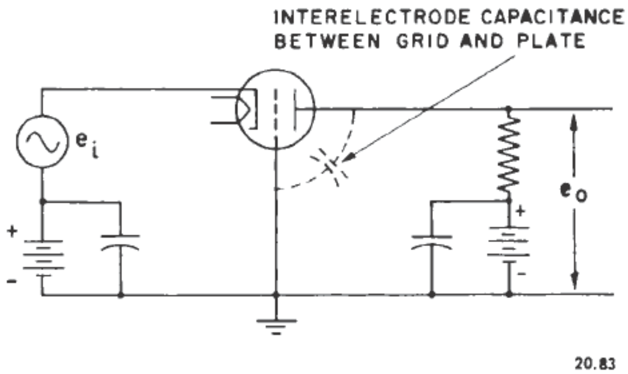


Figure 5-10.—Grounded-grid amplifier circuit.

series with the plate- to-cathode resistance of the tube. The source is thus required to furnish a considerable amount of power.

Grounded-Plate Amplifier

The grounded-plate amplifier (V1 in fig. 5-11) is another circuit configuration that may be used in u-h-f amplifiers. This amplifier circuit has a lower signal-to-noise ratio than the grounded-cathode, grounded-grid circuits and also poorer stability since the grid is not used as a shield between cathode and plate. However, the input capacitance is slightly less than that of the grounded-grid circuit. If the grounded-plate amplifier is used to drive a grounded-grid amplifier, V2, at very high frequencies the lower induced grid noise gives this circuit configuration a slight advantage over that of the grounded-cathode or grounded-grid type. In the cascode amplifier, described in the following section, a grounded plate amplifier is used to drive a grounded grid amplifier.

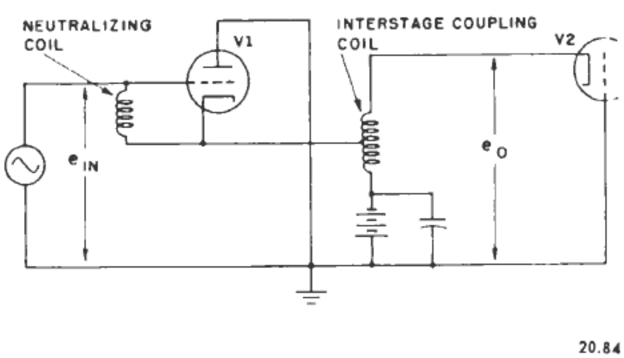


Figure 5-11.—Grounded-plate amplifier.

Cascode Amplifier

A popular r-f amplifier, the cascode amplifier, combines the good features of both the triode and the pentode. A simplified circuit of one type of cascode amplifier is shown in figure 5-12. The V1 plate voltage is held fixed while the plate current is permitted to vary. This action is like a pentode with the advantage that no screen current is required and less noise is introduced. Triodes have less noise than pentodes because they have fewer internal elements.

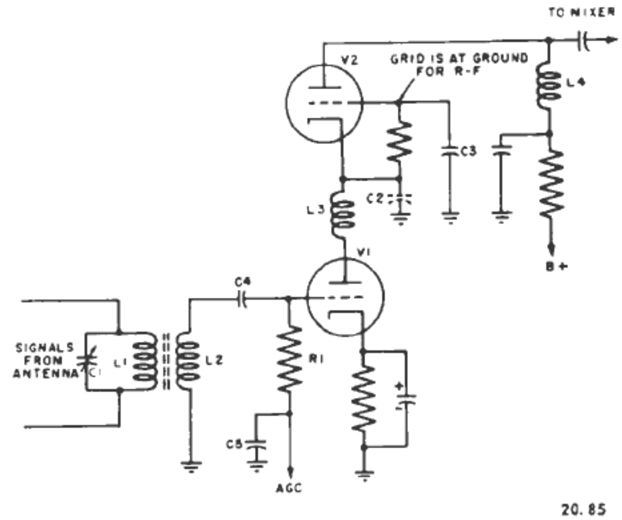


Figure 5-12.—Simplified circuit of a cascode amplifier.

Signals from the antenna are fed through tuned circuit C1L1 to L2. The frequency response is improved by having low X_C ohms in the coupling capacitor C4 and the r-f bypass capacitor C5 compared with high resistive ohms in the grid resistor R1. Neutralization (to prevent V1 from developing self oscillations) is not necessary for operation on the low band because of the compact arrangement of the circuit and the tube connections. Likewise, neutralization is unnecessary for operation on the high band because L3 and C2 (the distributed capacitance to ground) form a series resonant circuit (resonant at the center of the high band) from the plate of V1 to ground. The signal voltage across L3 varies the potential of the cathode of V2 with respect to ground; and, because the grid is grounded, the signal is applied effectively between the cathode and the grid of V2. This arrangement, in effect, nullifies the shunting distributed capacitance insofar as injecting undesired noise or degeneration into the grid cricuit of V1 and V2 is concerned.

CLASSIFICATION ACCORDING TO RESONANT QUALITY OF LOAD

Amplifiers are also classified according to whether they are TUNED or UNTUNED—that is, according to whether they amplify a restricted range or wide range of frequencies respectively.

Tuned amplifiers may be further sub-divided into NARROW-BAND and WIDE-BAND amplifiers. Whether a band of frequencies is considered narrow or wide depends on the ratio of the bandwidth to the center frequency, expressed as a percentage of the center frequency. An example of a narrow-band amplifier is the i-f amplifier in a broadcast radio receiver. The range is about 10 kc with a center frequency of 450 kc. The bandwidth in this example is 2.2 percent of the center frequency. Examples of wider-band amplifiers are the i-f stages or radar and television receivers which may have a range of about 4 mc at a center frequency of about 30 mc. In this case the bandwidth is 13.3 percent of the center frequency.

Untuned amplifiers are not tuned to any specific band of frequencies. The circuit components, however, may limit the range of frequencies that the circuit can handle. All audio amplifiers come under this classification.

DISTORTION IN AMPLIFIERS

The output of an ideal amplifier is identical with the input in all respects except for an increase in amplitude. This statement does not include, of course, the various wave-shaping and special-purpose amplifiers. A practical amplifier, however, falls short of this ideal. Not all frequency components present in the input may be amplified equally; the amplitude of the output voltage may not be proportional to the amplitude of the input voltage, and thus new frequencies will be introduced; or the relative phases from those of the input. These deviations from the ideal are known as FREQUENCY DISTORTION, AMPLITUDE (or nonlinear) DISTORTION, and PHASE (or delay) DISTORTION, respectively.

To achieve the special waveforms necessary in certain radar, television, or test circuits distortion is deliberately introduced by the amplifier or an associated circuit. In certain other circuits, however, less distortion of all three types is permitted than would be tolerated in the case of broadcast radio amplifiers.

FREQUENCY DISTORTION

When some frequency components of a signal are amplified more than others or when some frequencies are excluded, the result is frequency distortion. Essentially, this type of distortion results from bandwidth restrictions imposed by the various amplifier circuit components. For example if a coupling circuit does not pass the third or higher harmonics that are present in the input, the circuit introduces frequency distortion.

For purposes of comparison, figure 5-13,A, shows the input and output of a two-stage amplifier that has introduced frequency distortion. The input e_{in}, contains the fundamental and the third harmonic; but the output, e_{out}, contains only the fundamental since the amplifier is unable to pass the third harmonic. Frequency distortion may occur at low frequencies if the coupling capacitor between the stages is so small that it presents a high series impedance to the low-frequency components of a signal. Distortion may also occur at high frequencies because of the shunting effects of the distributed capacitance in the circuit.

Low- and high-frequency compensation—that is, boosting the response of the amplifier at the low- and high-frequency ends of the desired band—is discussed in chapter 6 under "Video Amplifiers."

PHASE DISTORTION

Most coupling circuits shift the phase of a sine wave, but this shift has no effect on the shape of the output. However, when more complex waveforms are amplified, each component frequency that makes up the overall waveform may have its phase shifted by an amount that depends on its frequency. Thus, the output is

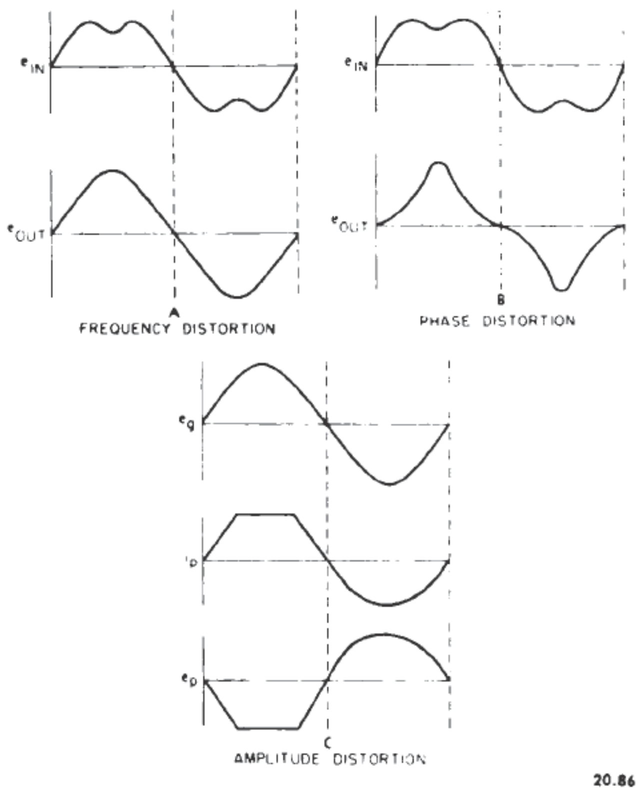


Figure 5-13.—Types of amplifier distortion.

not a faithful reproduction of the input waveform.

Figure 5-13,B, shows the input and output waveforms of a two-stage amplifier that has introduced phase distortion. The input signal, e_{in}, consists of a fundamental and a third harmonic. Although the amplitudes of both components have been increased by identical ratios, the output, e_{out}, is considerably different from the input because the phase of the third harmonic has been shifted with respect to the fundamental.

Basically, phase distortion is present whenever the component frequencies in the input of an amplifier are not all passed through the amplifier in the same length of time. Phase or time delay distortion is not important in the amplification or reproduction of sound because the ear is unable to detect relative phase shifts of the individual components. Such distortion, however, is important in radar, television, and measuring equipment where the waveform must be accurately maintained during amplification. Phase distortion may be reduced by varying the amount or type of coupling. In video amplifiers special coupling circuits are used to reduce this distortion.

AMPLITUDE DISTORTION

If a signal is amplified by an electron tube that is not operating on the linear portion of its characteristic curve, amplitude (nonlinear) distortion will occur. In the nonlinear region a change in grid voltage does not result in a change in plate current that is directly proportional to the change in grid voltage. For example, if a tube is overdriven by applying a grid signal that drives the tube beyond the linear portion of the characteristic curve (nonlinear distortion) and also to the point where the grid draws current (grid-limiting distortion) the resulting signal is distorted in amplitude, as shown in figures 5-13,C, and 5-14. This type of distortion is to be expected, since for a portion of the negative half of the grid signal swing the tube operates on a nonlinear portion of the characteristic curve, and for a portion of the positive swing the grid draws current.

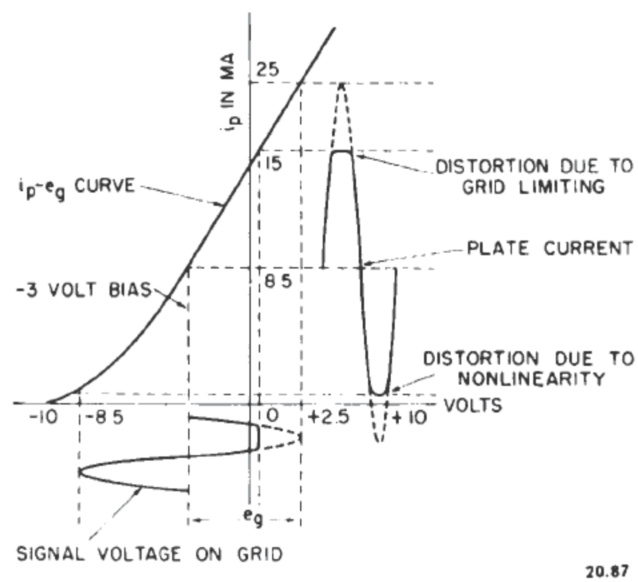


Figure 5-14.—Distortion in a class A amplifier due to excessive signal voltage.

Beyond the linear portion of the curve, a further increase in negative grid potential will not cause a proportionate reduction in plate current. On the other half cycle, the positive swing of grid voltage (beyond the point where the grid draws current) is limited by the loss in voltage within the source impedance and no further increase in plate current can occur. The result of this nonlinearity is the production of harmonics that were not present in the input of the amplifier. This concept can be better understood if it is recalled that any complex periodic waveform may be considered as being composed of a number of sine waves of different frequencies and amplitudes. The sine wave that has the same frequency as the complex periodic wave is called the FUNDAMENTAL. These frequencies higher than the fundamental are called HARMONICS. Thus, from the complex waveform is obtained a number of harmonics plus the fundamental frequency.

At the higher frequencies, harmonics may be reduced by the use of a parallel resonant circuit as a plate load, by link coupling, or by filtering. At the audio frequency, however, there is an overlap of frequencies and filtering is not practicable. The best solution is to operate the tube on the straight portion of the characteristic curve for class A operation, or to operate it in a push-pull arrangement for class B operation.

Complex waveforms are necessary in certain television, radar, and test instrument circuits. These waveforms include square, sawtooth, and peaked waves. In each of these waveforms the distortions are deliberately introduced. The circuits for producing these distorted waveforms and an analysis of nonsinusoidal waves and transients are treated in a later chapter of this course.

MISCELLANEOUS DISTORTION

HUM is a type of distortion particularly objectionable in audio- and video-frequency amplifiers. It may be caused by alternating current in the filaments or heaters of the amplifier tubes by stray electromagnetic or electrostatic fields, or by insufficient filtering of the power supply. A center-tapped resistor across the filament terminals to which the grid return is connected may reduce hum at the power frequency. On the other hand, the elimination of hum due to cyclic variations in filament temperature, at twice the power frequency, is largely a design problem. The elimination of hum in heater-type tubes is also largely a design problem, although, as in the case of filament-type cathodes, the a-c leads may be twisted together and placed in positions that cause the least magnitude of induced voltages in the signal circuits.

It is important to return the grid and plate circuits of a tube to a point in the exact electrical center of the filament circuit, as shown in figure 5-15,A. In this figure the resistance of the resistor is high with respect to the resistance of the filament, so that most of the current flows through the filament. Tapping the resistor at its electrical center is essentially the same as tapping the filament at its electrical center.

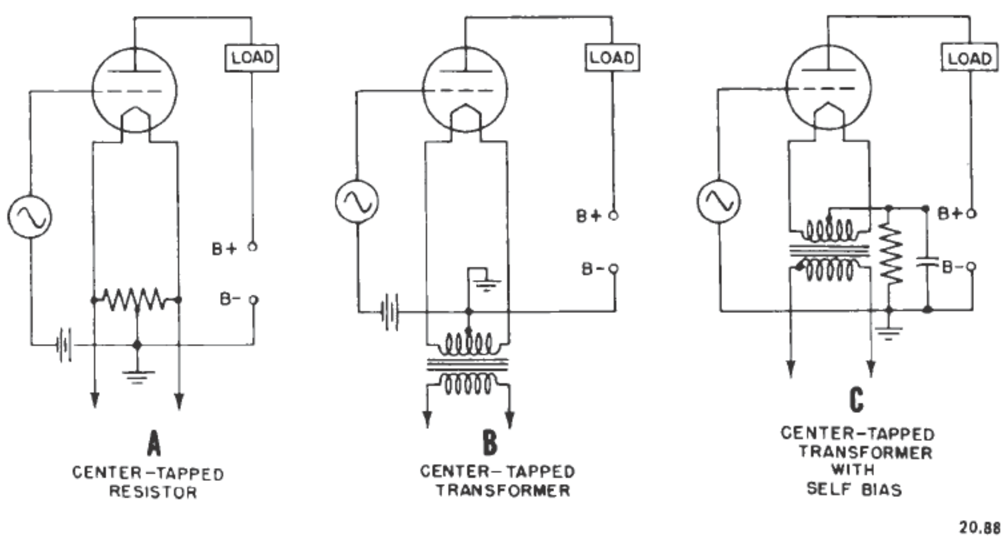


Figure 5-15.—Filament connections for reducing hum.

In figure 5-15,B, the grid return path may be traced through the two halves of the center-tapped filament winding simultaneously outward from the center tap to the two ends of the winding. At any instant the a-c filament voltage makes the grid more negative with respect to one end of the filament and less negative with respect to the other. The increases and decreases from one end of the filament to the other tend to cancel each other and therefore the bias (cathode-to-grid potential) remains the same. In figure 5-15,C, the same action takes place except that cathode bias is developed across the resistor connected between center tap and ground.

The magnetic field set up by the current that passes through the heater wires may induce a power-frequency hum in other tube elements by electromagnetic induction. The effect of this coupling can be materially reduced by using twisted filament leads. The current in each wire flows in opposite directions and the magnetic flux fields tend to cancel each other.

Even though the effects of the fundamental frequency of the heater current are essentially eliminated by center-tapping the filament, another serious problem presents itself—the output of the tube will contain an a-c component whose frequency is twice that of the fundamental. This condition results from the very nature of alternating current. The current reaches a peak value twice during each cycle—once during each direction of current flow. Therefore, twice during each cycle the filament is heated to maximum temperature and the maximum number of electrons is emitted. In order to minimize this effect, large heavy cathodes are used to provide enough thermal inertia to prevent the temperature of the filament from changing significantly with voltage alternations.

The undesired effects produced by stray fields may be reduced by proper placement of transformers; proper shielding of transformers, leads and tubes; and arrangement of circuits and components so that there will always be a low impedance bypass to ground to the undesired currents.

MICROPHONIC EFFECTS are the result of slight vibrations in the tube elements. Variations in plate current due to these vibrations are amplified in each succeeding stage and appear in the output of audio and video amplifiers. These slight displacements of the tube elements may be caused either by physical vibration of the chassis or by the sound vibrations emitted by the speaker.

The obvious remedy is to employ some method that will insulate the tube or tubes from the vibrating source. Some tubes, however, are less susceptible to microphonics than others, and occasionally simply replacing a tube will cure the trouble.

NOISE in audio and video amplifiers may be caused by faulty contacts, faulty components such as resistors or capacitors or THERMAL-AGITATION NOISE. Thermal-agitation noise occurs because all electrical conductors contain electrons moving at random Some of these electrons move at random even if there is an impressed voltage across the conductor. By chance at any given instant, more of these electrons move in one direction than in another. When amplified, the accompanying voltage results in thermal-agitation noise.

Also inherent in electron tubes are other noises such as SHOT EFFECT, which results from a variation in the rate of electron emission from a cathode; GAS NOISE, which results from a variation in the rate of production of ions; and SECONDARY EMISSION NOISE, which results from a variation in the rate of production of secondary electrons. There are also other variations that produce noise in the output of a receiver.

In the final analysis tube noise is the limiting factor that determines the ultimate sensitivity of an amplifier.

COUPLING METHODS

A single stage of voltage or power amplification normally is not sufficient for radio or radar applications. To obtain the necessary gain, several stages must often be connected together. The output of one stage then becomes the input of the next throughout the series of stages and this arrangement is called a CASCADE AMPLIFIER.

A cascade amplifier is designated according to the method used to couple one amplifier stage to the next. There are a number of methods each having certain advantages and disadvantages and the choice for a particular application depends on the needs of the circuit. The basic methods are: (10) resistance-capacitance coupling; (2) impedance coupling; (3) transformer coupling; and (4) direct coupling.

Before considering the details of each coupling method it is desirable to establish the equivalent circuit of an electron-tube amplifier. The characteristics of an amplifier are determined more readily by replacing the tube with its equivalent circuit and analyzing this circuit.

RESISTANCE-CAPACITANCE COUPLING

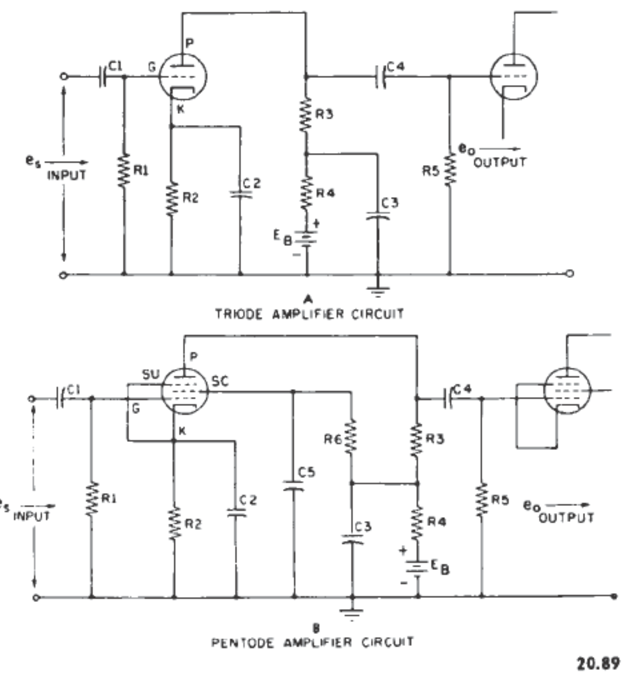
One of the most widely used methods of connecting amplifier stages is R-C coupling. Amplifiers coupled in this manner are relatively inexpensive lack heavy components have good fidelity over a comparatively wide frequency range, are relatively free from undesirable induced currents from a-c heater leads, and are especially suitable for use with pentodes and high-mu triodes.

A resistance-capacitance coupled amplifier (generally shortened to resistance-coupled amplifier) can be designed to have good response for almost any desired frequency range. For instance it can be designed to give fairly uniform amplification of all frequencies in the range from 100 to 20,000 cps. Slight modification of the circuits can extend the frequency to cover the wide band required in video amplifiers. However, extension of the range can be obtained only at the cost of reduced amplification over the entire range. Thus the R-C method of coupling amplifiers gives a good frequency response with minimum distortion, but it also gives low amplification.

Typical resistance-coupled amplifiers are shown in figure 5-16, together with the names of the various circuit elements.

In the triode shown in figure 5-16,A, the d-c grid circuit includes G, R1, R2, and K; and the a-c grid circuit includes G, R1, C2, and K. In the pentode in figure 5-16,B, the d-c screen circuit includes SC, R6, R4, E_b, R2, and K; and the a-c circuit includes SC, C5, C2, and K. In each case the d-c plate circuit includes P, R3, R4, E_b, R2, and K; and the a-c plate circuit includes P, R3, C3, C2, and K.

In order that the output voltage may be large, the load resistor should have as high a value as practicable. However, the higher this value becomes, the greater is the voltage drop across



R1—Grid-leak resistor.
R2—Cathode bias resistor.
R3—Plate load resistor.
R4—Plate decoupling resistor.
R5—Second-stage grid resistor.
R6—Screen dropping resistor.
C1—Input coupling.
C2—Cathode bypass capacitor.
C3—Plate supply bypass capacitor.
C4—Output coupling capacitor.
C5—Screen bypass capacitor.

Figure 5-16.—Typical resistance-coupled amplifiers.

it and the lower is the voltage remaining between the plate and cathode of the tube. To obtain the required effective plate voltage, the voltage drop across the load resistor is subtracted from the plate supply voltage. Thus there is a practical limit to the size of the plate load resistor if the plate is to be supplied with its rated voltage. If a larger plate resistor is necessary, the only alternative is to increase the plate supply voltage. There is, of course, a practical limit to the amount that the plate voltage may be increased. An example of the d-c voltage distribution around the plate circuit is shown in figure 5-17. The plate current is 6 ma and the voltage across the 30 k-ohm load resistor is 6 x 30 or 180 volts. The voltage drop across the 500-ohm cathode resistor is 0.006 X 500 or 3 volts which provides the grid bias for the tube. The plate-cathode voltage is the B-supply voltage less the drop through R_L and R_K, or 300 - 180 - 3 = 117 volts.

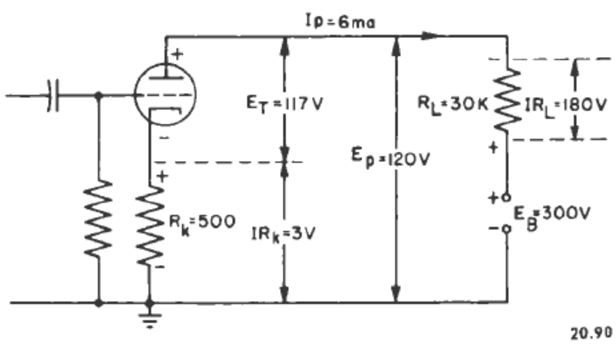
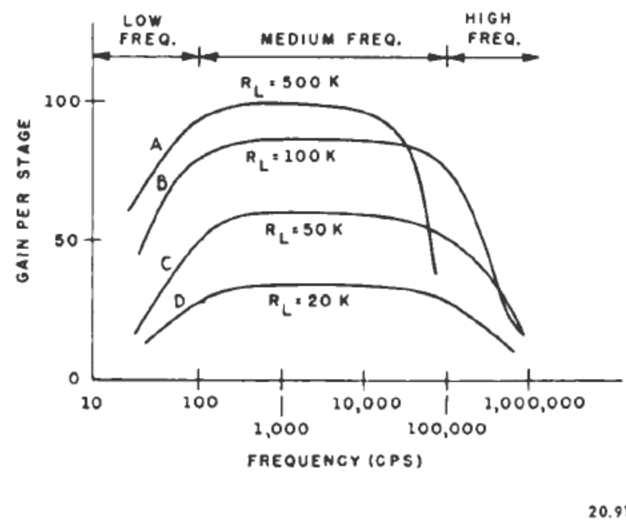


Figure 5-17.—Plate-circuit voltage distribution.

The screen resistor, R6, in figure 5-16,B, has the necessary voltage drop across it so that when this drop is subtracted from the B-supply voltage, the rated screen voltage will remain. The value of the cathode resistor, R2, is determined by the grid bias required and the no-signal plate and screen currents. For the range of frequencies to be amplified, the cathode bypass capacitor, C2, has a low reactance to the a-c component of plate current in comparison with the resistance of R2. The decoupling (or filter) circuit, C3R4, tends to prevent the a-c component of plate current from flowing through the B supply because R4 offers a high series resistance and C3 offers a low shunt reactance to the a-c signal component.

Typical frequency response curves for an R-C coupled audio amplifier are shown in figure 5-18. The response is measured in terms of the voltage gain of the amplifier over a range of frequencies. The voltage gain is the ratio of e_o to e_s. The gain falls off at very low frequencies because of the increase in the capacitive reactance of the interstage coupling capacitor, C_c (fig. 5-19,A). This capacitor acts in series between the source and the load and has developed across it an increasing part of the signal voltage as the frequency is decreased.



CURVE	R_L OHMS	I_p MA	G_m μMHOS	R_p MEGOHMS
A	500,000	0.3	414	5
B	100,000	1.5	1100	2
C	50,000	3.0	1400	1
D	20,000	3.0	1400	1

Figure 5-18.—Gain vs frequency of an R-C coupled amplifier for various plate loads.

The reduction in gain at the higher frequencies is caused by the shunting effect across load resistor R_L of the output capacitance, C_o, of one stage, the input capacitance (C_i) of the next stage, and the distributed capacitance (C_d) of the coupling network. The combined effect of these capacitances is to increase the part of the total signal voltage that is developed across the internal resistance, r_p, of the generator (fig. 5-19,B) and to decrease the part that appears as output voltage, e_{out}.

Middle-Frequency Gain

The middle-frequency gain is flat and in the example in figure 5-18 is assumed to extend approximately from 100 to 200,000 cps. The equivalent circuits shown in figure 5-20 represent the active circuit components and their connections for the middle-frequency range. Figure 5-20,A, illustrates the constant-voltage generator form and figure 5-20,B, the constant-current generator form. The reactance of the coupling capacitor, C_c, is low at the middle

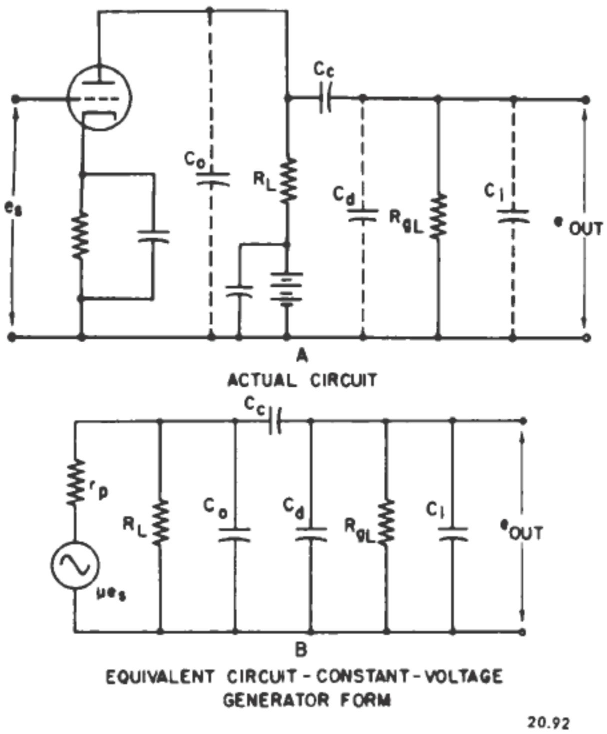


Figure 5-19.—Single-stage resistance-coupled amplifier and equivalent circuit.

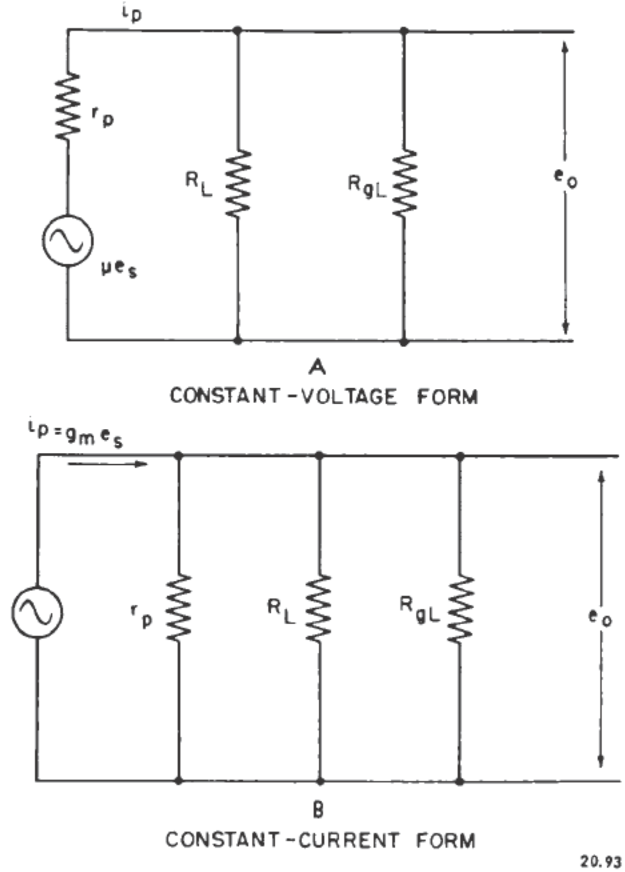


Figure 5-20.—Middle-frequency equivalent circuits.

frequencies and thus is omitted between R_L and R_{g1}. The reactances of the shunting capacitances, C_o, C_i, and C_d, are high at the middle frequencies and so they too are omitted from the equivalent circuit. Thus the equivalent circuits are reduced to include only the generator, r_p, R_L, and R_{g1}. The amplification is independent of frequency, and a flat response may be expected.

In the constant-voltage generator form (fig. 5-20,A), the a-c component of plate current is

$$i_p = \frac{\mu e_s}{r_p + R_{eq}}$$

where R_{eq} is the combined resistance of R_L and R_{g1} in parallel. The output voltage across R_{eq} is

$$e_o = i_p R_{eq}$$

$$= \frac{\mu e_s R_{eq}}{r_p + R_{eq}}$$

The voltage gain is

$$\frac{e_o}{e_s} = \frac{\mu R_{eq}}{r_p + R_{eq}}$$

For example, if the triode has an amplification factor of 20, a plate resistance of 10 k-ohms, a plate load resistance of 50 k-ohms, and a following stage grid-leak resistance of 100 k-ohms, the middle-frequency voltage gain is

$$\frac{e_o}{e_s} = \frac{20 \text{ X} \dfrac{50 \text{ X } 100}{50 + 100}}{10 + \dfrac{50\text{X}100}{50+100}} = \frac{20 \text{ X } 33.3}{10 + 33.3} = 15.4$$

In the constant-current generator form (fig. 5-20,B), the output voltage appearing across r_p, R_L, and R_{g1} in parallel is

$$e_o = i_p \frac{1}{\dfrac{1}{r_p} + \dfrac{1}{R_L} + \dfrac{1}{R_{g1}}}$$

Substituting

$$i_p = \frac{\mu e_s}{r_p}$$

and $$g_m = \frac{\mu}{r_p}$$

$$e_o = g_m e_s \frac{1}{\frac{1}{r_p} + \frac{1}{R_L} + \frac{1}{R_{gl}}}$$

The voltage gain is

$$\frac{e_o}{e_s} = g_m \frac{1}{\frac{1}{r_p} + \frac{1}{R_L} + \frac{1}{R_{gl}}}$$

For example, if an R-C coupled pentode has a plate resistance of 1 megohm, a plate-load resistance of 0.125 megohm, a following stage grid-leak resistance of 0.25 megohm, and a transconductance of 900 micromhos, the middle-frequency gain is

$$\frac{e_o}{e_s} = 900 \times 10^{-6} \times \frac{1}{\frac{1}{1 \times 10^6} + \frac{1}{0.125 \times 10^6} + \frac{1}{0.25 \times 10^6}}$$

$$= 900 \times 10^{-6} \times \frac{10^6}{1+8+4}$$

$$= 69.4$$

The voltage gain of 69.4 is less than 10 percent of the amplification factor of the pentode because the combined resistance across which the output voltage appears is less than 10 percent of the pentode plate resistance.

Low-Frequency Limit

In the lower range of frequencies amplified by an R-C coupled amplifier, C_o, C_d, C_i, are unimportant and are omitted from the equivalent circuit since the X_c ohms are high and are in parallel with R_L and R_{gl}. The reactance of coupling capacitor C_c, however, becomes increasingly important at the low frequencies and cannot be neglected. A low-frequency equivalent circuit is shown in figure 5-21.

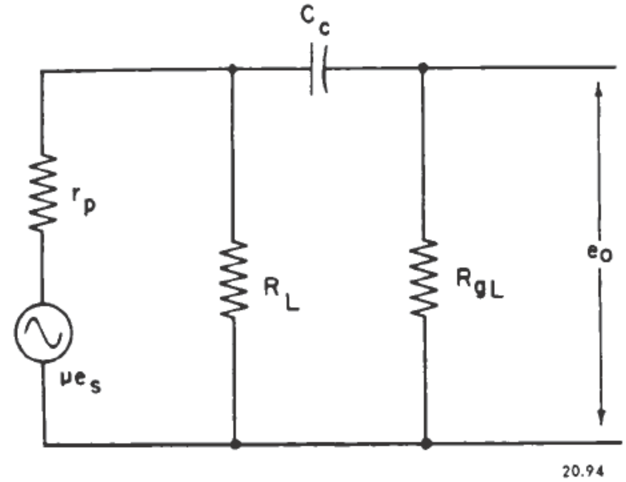


Figure 5-21.—Low-frequency equivalent circuit.

Since X_c varies inversely with frequency more of the total voltage, μe_s, appears across C_c and less across R_{gl} as the frequency decreases. The approximate frequency at which the output voltage falls to 70 percent of its (middle frequency) value is the frequency at which X_c of the coupling capacitor is equal to R_{gl}. The approximation is applicable to triode amplifiers. The low-frequency limit is derived as follows,

$$X_c = R_{gl}$$

$$\frac{1}{2\pi f C_c} = R_{gl}$$

$$f = \frac{1}{2\pi C_c R_{gl}}$$

For example, an R-C coupled triode amplifier has a coupling capacitor of 0.04 f and a following-stage grid-leak resistor of 100,000 ohms. The low-frequency limit is

$$f = \frac{1}{2\pi \times 0.04 \times 10^{-6} \times 10^{-5}}$$

$$= 39.8 \text{ cps}$$

The low frequency limit of approximately 40 cps is the frequency at which the output voltage is 70 percent of its middle-frequency value.

Increasing the size of C_c lowers the frequency response (increases the number of low-frequencies) of the amplifier but there is a practical limit. This limit is caused by a type of regeneration that may occur between several R-C coupled stages supplied by a common plate and screen power source. Such regeneration is called MOTOR BOATING and occurs when the B-supply source impedance is relatively high compared with the X_c ohms of the interstage coupling capacitors. This action is described in Chapter 8 in connection with relaxation oscillators.

High-Frequency Limit

In the high-frequency range the shunting capacitances, C_o, C_d, and C_i, of the general equivalent circuit (fig. 5-19,B,) become significant. These capacitances limit the output voltage at high frequencies. In figure 5-22 these parallel capacitances are combined and designated "C_s." The high-frequency limit (arbitrary) is the frequency at which the output voltage falls to 70 percent of its value at the middle frequencies. This limit occurs at the frequency at which the X_c ohms of the shunting capacitance, C_s, is equal to the combined resistance of r_p, R_L, and R_{gl} in parallel. Thus

$$\frac{1}{2\pi f C_s} = R_{eq}$$

and

$$f = \frac{1}{2\pi C_s R_{eq}}$$

For example, an R-C coupled pentode amplifier has a plate resistance of 1 megohm, a load resistance of 0.125 megohm, a grid-leak resistance of 0.25 megohm, and a shunting capacitance of 100 $\mu\mu$f. The combined resistance of r_p, R_L, and R_{gl} in parallel is

$$R_{eg} \text{ (in megohms)} = \frac{1}{\frac{1}{1} + \frac{1}{0.125} + \frac{1}{0.25}}$$

$$= \frac{1}{1+8+4}$$

$$= \frac{1}{13} = 0.077 \text{ megohm}$$

The high-frequency limit is

$$f = \frac{1}{6.28 X 100 X 10^{-12} X 0.077 X 10^6}$$

$$= 20{,}700 \text{ cps}$$

The high-frequency limit of approximately 21,000 cps is the frequency at which the output voltage falls to 70 percent of its middle-frequency value. The upper-frequency limit may be extended by using tubes having low interelectrode capacitances. The upper limit may also be extended by reducing R_L, but at the expense of mid-frequency gain.

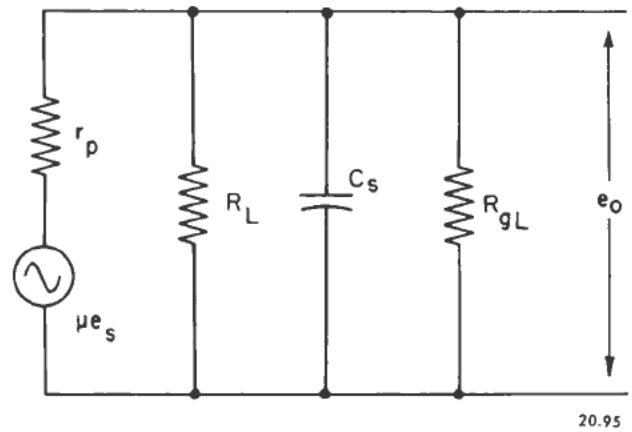


Figure 5-22.—High-frequency equivalent circuit.

Wide-band R-C coupled amplifiers are characterized by many stages having relatively low gain per stage, large coupling capacitors between stages, and low-shunting capacitances.

IMPEDANCE COUPLING

Impedance or inductance-coupling coupling is obtained by replacing the load resistor, R_L, of a normal R-C coupled amplifier with an inductance, L, as shown in figure 5-23. To obtain as much amplification as possible, particularly at the lower frequencies, the inductance is made as large as practicable. To avoid undesirable magnetic coupling a closed-shell type of inductor is used. Because of the low d-c resistance of the inductor, less d-c voltage appears across it. Thus the tube can operate at a higher plate voltage.

The degree of amplification is not uniform as it is with the R-C coupled amplifier because the load impedance, Z_L, varies with the frequency—that is

$$Z_L = R + j2\pi f L$$

Since the output voltage appears across Z_L, the voltage gain increases with the frequency up to the point where the shunting capacitance limits it. The shunting capacitance includes not only the interelectrode and distributed wiring capacitances found in R-C coupled amplifiers but also the distributed capacitance associated with the

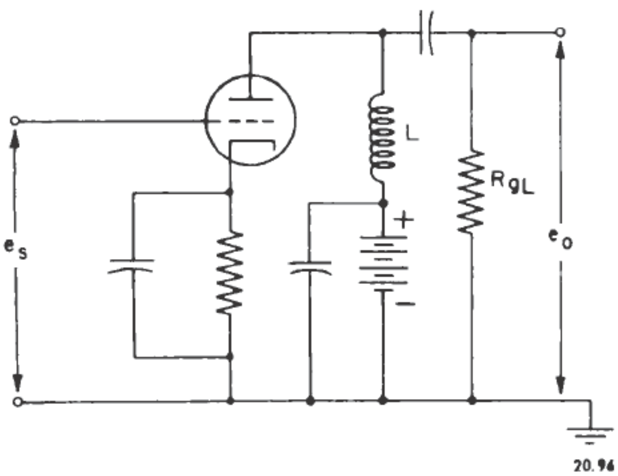


Figure 5-23.—Impedance-coupled amplifier.

turns of the inductor. The distributed capacitance between the turns of the coil greatly increases the capacitance to ground and plays a major part in limiting the use of this coupling at the higher frequencies.

TRANSFORMER COUPLING

A transformer-coupled stage of amplification (fig. 5-24) has certain advantages over other types of coupling. The voltage amplification of the stage may exceed the amplification of the tube if the transformer has a step-up turns ratio. Direct-current isolation of the grid of the next tube is provided without the need for a blocking capacitor; and the d-c voltage drop across the coupling resistor, which is necessary when R-C coupling is used, is avoided. This type of coupling is also used to couple a high-impedance source to a low-impedance load, or vice versa by choosing a suitable turns ratio. Also it may be used as a simple means of providing phase inversion for a push-pull amplifier without the use of special phase inverting circuits.

Transformer coupling has the disadvantages of greater cost, greater space requirement, the necessity for greater shielding, and the possibility of poorer frequency response at the higher and lower frequencies. The voltage gain as a function of frequency throughout the range in question is shown in figure 5-25. The curve shows that the transformer-coupled voltage amplifier has a relatively high gain and uniform frequency response over the middle range of audio frequencies, but poor response for both low and high audio frequencies.

Like those of resistance-coupled amplifiers, complete equivalent circuits of transformer-coupled amplifiers are complex networks. An analysis of them can be considerably simplified by considering one at a time the equivalent circuits for the low, middle, and high frequencies.

Middle-Frequency Gain

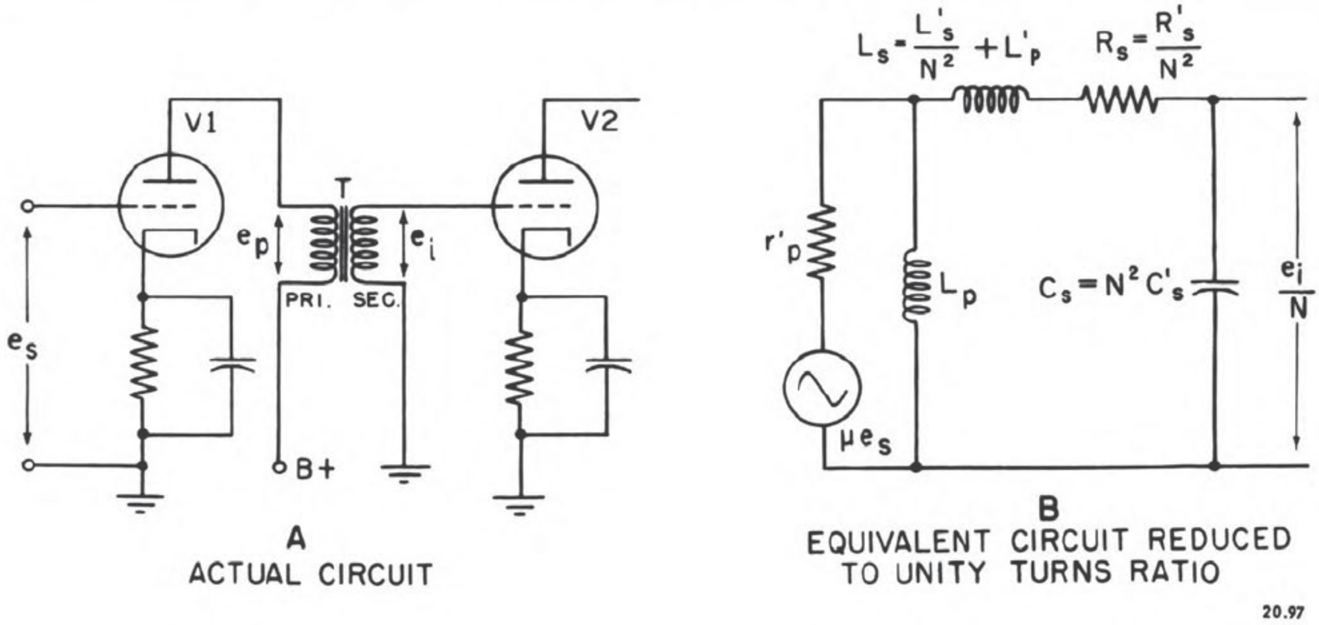
The primary of transformer T (fig. 5-24,A) is connected in the plate circuit of V1 and the secondary is connected between the grid and cathode of V2. An input signal, e_s, applied between the grid and cathode of V1, appears as an amplified plate signal, e_p, across the transformer primary.

At the middle frequencies the reactances of the transformer primary inductance, L_p, and secondary distributed capacitance, C_s (fig. 5-24,B), are sufficiently high to be considered as open circuits and there is no loss of signal voltage in the plate resistance of V1. Thus μe_s is equal to e_p. The secondary output voltage, e_i, applied to the input of V2, is equal to Ne_p or μNe_s, where N is the secondary-to-primary transformer turns ratio.

The simplified equivalent circuit applicable to the middle frequencies is shown in figure 5-26. In this figure the primary inductance and secondary distributed capacitance have been omitted and resistor r'_p includes the plate resistance and the primary winding resistance in series. The transformer leakage reactances are negligible and the output voltage, $\frac{e_i}{N}$, is equal to μe_s since there is no drop through r'_p. The middle-frequency gain is equal to $\frac{e_i}{e_s}$, or $\frac{\mu Ne_s}{e_s} = \mu N$. The circuit applies to a class A voltage amplifier in which no grid current flows during any part of the input cycle.

Low-Frequency Limit

At the lower frequencies the shunting effect of the interelectrode and distributed capacitances is even less than it was at the middle frequencies (X_c varies inversely with the frequency), and is omitted in the equivalent circuit shown in figure 5-27. The reactance, $2\pi fL_p$, of the transformer primary is reduced at low frequencies.



e_s —input signal.
e_p —primary signal voltage.
e_i —input voltage to second tube.
r_p —plate resistance of tube & resistance of primary windings.
μ—Amplification factor of tube.
L_p —incremental inductance of primary.
L'_p—primary leakage inductance.
L'_s—secondary leakage inductance.
R'_s—resistance of secondary windings.
C'_s—distributed capacitances of secondary circuit.
N —secondary-to-primary turns ratio of transformer. (unity turns ratio assumed in the equivalent circuit).
r'_p—combined plate resistance and primary resistance.

Figure 5-24.—Single-stage transformer-coupled voltage amplifier and equivalent circuit.

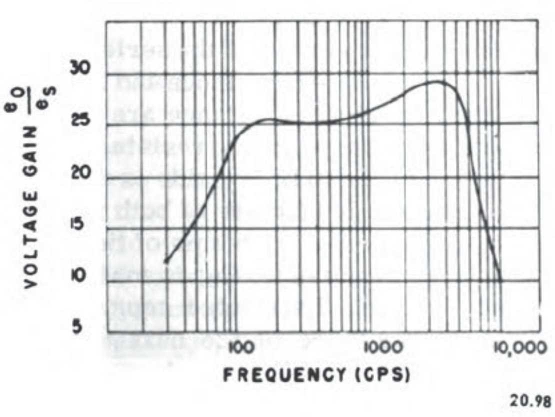


Figure 5-25.—Voltage gain vs frequency of a transformer-coupled voltage amplifier.

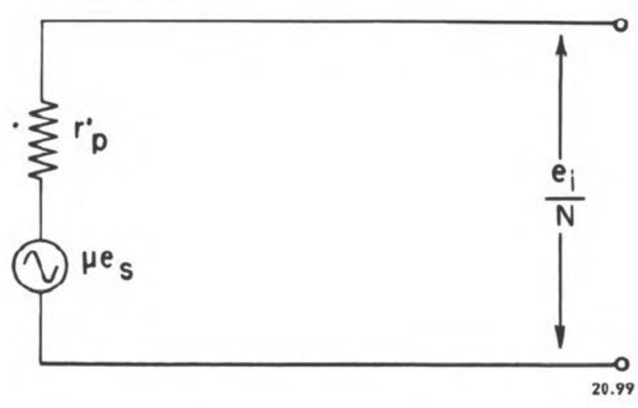


Figure 5-26.—Equivalent circuit of transformer-coupled voltage amplifier for mid-frequency operation.

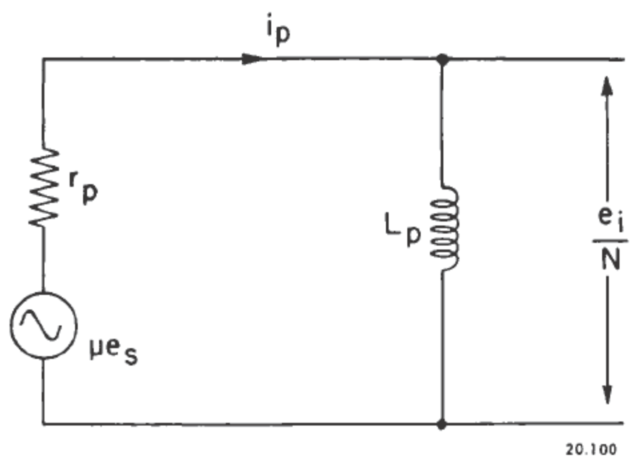


Figure 5-27.—Equivalent circuit of transformer-coupled voltage amplifier for low-frequency operation.

The low frequency at which the gain falls to 70 percent of the middle frequency gain is the frequency at which $2\pi fL_p$ is equal to r'_p. The primary coil resistance is small with respect to the plate resistance, r_p, and can be neglected. Thus $r'_p = r_p$. Also,

$$2\pi fL_p = r_p$$

and

$$f = \frac{r_p}{2\pi L_p}$$

For example, an audio voltage amplifier triode has an r_p of 16,000 ohms and a μ of 8.5. The coupling transformer has a step-up voltage ratio of $\frac{3}{1}$ and a primary inductance of 40 henrys. The low frequency limit is

$$f = \frac{16{,}000}{2\pi X40} = 63.7 \text{ cps}$$

The middle-frequency gain is μN, or $8.5X3=25.5$. Thus the low-frequency limit of 63.7 cps is the frequency at which the gain falls to 70 percent of 25.5, or about 17.8 (fig. 5-25).

The decrease in the reactance of the transformer primary inductance causes a falling off in gain at the lower frequencies. The falling off begins at higher frequencies when high-mu tubes having high r_p are used. The larger the transformer primary incremental inductance (inductance with partial d-c core saturation) the better is the low-frequency response.

High-Frequency Limit

At the high-frequency end of the band the reactance of the primary inductance is high and is neglected in the equivalent circuit (fig. 5-28). The effect of the shunting capacitance, C_S, is appreciable and the output voltage appears across it. The leakage inductance of the transformer acts in series with C_S, r_p, and the winding resistances.

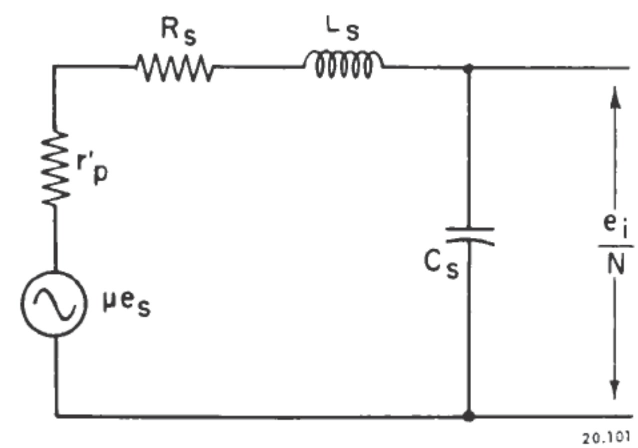


Figure 5-28.—Equivalent circuit of transformer-coupled voltage amplifier for high-frequency operation.

These factors are included in the equivalent circuit for high-frequency operation which is similar to a series L-C-R circuit having a low Q operating in the vicinity of the series-resonant frequency. The plate resistance and the primary transformer winding resistance are included in r'_p and the secondary winding resistance expressed in terms of the primary side is equal to R_S. The total leakage inductance of both windings is equal to L_S expressed in terms of the primary. The shunting capacitance, C_S, is made up of the secondary-winding distributed capacitance and the input capacitance of the next tube, and is expressed in terms of the transformer primary.

Although the Q of this circuit is relatively low, the series resonant effect is sufficiently pronounced to increase the gain at and near resonance unless precautions are taken to prevent this effect. Above resonance the gain falls off rapidly. Greatest uniformity of gain occurs in the high-frequency end of the band when the Q of the series-resonant circuit is approximately 0.85.

The high-frequency limit is in the vicinity of the series-resonant frequency, which is determined by the transformer leakage inductance and the effective shunting capacitance.

For example, the high-frequency end of the band of a transformer-coupled amplifier in which the leakage inductance of the transformer is 0.4 henry and the shunting capacitance is 1,500 micromicrofarads is

$$f_o = \frac{1}{2\pi\sqrt{L_s C_s}}$$

$$= \frac{1}{6.28\sqrt{0.4X1,500x10^{-12}}}$$

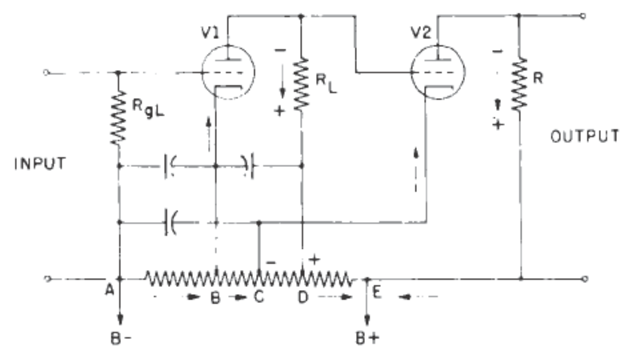
$$= 6,460 \text{ cps}$$

Above this frequency the gain falls off rapidly because of the decrease in reactance of the shunting capacitance across which the output voltage is developed and also because of the increase in reactance of the leakage inductance in series with the output. Good high-frequency response is obtained by using a transformer having a small distributed capacitance. This effect is obtained by using a low secondary-to-primary turns ratio which means a limited voltage gain, and a small transformer which means a poor low-frequency response. Thus transformer coupling is a compromise between high gain and good high-and low-frequency response.

DIRECT COUPLING

In each of the coupling circuits that have been considered so far, the coupling device isolates the d-c voltage in the plate circuit of one tube from the grid circuits of the next tube; but they are designed to transfer the a-c component with minimum attenuation.

In a direct-coupled amplifier, on the other hand, the plate of one tube is connected directly to the grid of the next tube without going through a capacitor, a transformer, or any similar coupling device. This arrangement presents a problem of voltage distribution. Since the plate of a tube must have a positive voltage with respect to its cathode, and the grid of the next tube must have a negative voltage with respect to its cathode, it follows that the two cathodes cannot operate at the same potential. Proper voltage distribution is obtained by a voltage divider, as shown at A,B,C,D, and E in figure 5-29. In this amplifier the plate of V1 is connected directly to the grid of V2. The grid of V1 is returned to point A through R_{g1}. The cathode of V1 is returned to point B and the cathodes cannot operate at the same potential. Proper voltage distribution is obtained by a voltage divider, as shown at A,B,C,D, and E in figure 5-29. In this amplifier the plate of V1 is connected directly to the grid of V2. The grid of V1 is returned to point A through R_{g1}. The cathode of V1 is returned to point B and the grid bias for V1 is developed by the voltage drop between points A and B of the voltage divider. The plate of V1 is connected through its plate load resistor, R_L, to point D on the divider. R_L also serves as the grid resistor for V2.



20.102

Figure 5-29.—Direct-coupled amplifier.

Since the plate current from V1 flows through R_L, a certain amount of the supply voltage appears across R_L. The amount of voltage developed across R_L must be allowed for in choosing point D on the divider. Point D is so located that approximately half of the available voltage is applied to the plate of V1. The plate of V2 is connected through a suitable output load, R, to point E, the most positive point on the divider. Since the voltage drop across R_L may place too high a negative bias on the grid of V2, it may be necessary to connect the cathode of V2 at point C, which is negative with respect to point D, in order to lower the bias on the grid of V2 (since the voltages across R_L and CD are in opposition). Point C, together with the value of R, determines the proper voltage for V2.

The entire circuit is a complex resistance network that must be adjusted carefully to obtain the proper plate and grid voltages for both tubes. If more than two stages are used in this type of amplifier, it is difficult to achieve stable operation. Any small changes in the voltages of the first tube will be amplified and will thus make it difficult to maintain proper bias on the final tube connected into the circuit. Because of the instability thus encountered, direct-coupled amplifiers are practically always limited to two stages. Furthermore, the power supply must be twice that required for one stage.

One method of supplying the range of voltage needed is to use a power supply which provides approximately equal amounts of both positive and negative voltages with respect to ground. This allows cascading without necessitating excessively high plate supply voltages. In figure 5-29, either point C or point D might properly be tied to ground potential.

When the tube voltages are properly adjusted to give class A operation, the circuit serves as a distortionless amplifier whose response is uniform over a wide frequency range. This type of amplifier is especially effective at the lower frequencies because the impedance of the coupling elements does not vary with the frequency. Thus a direct-coupled amplifier may be used to amplify very low frequency variations in voltage. Also, because the response is practically instantaneous, this type of coupling is useful for amplifying pulse signals where all distortion caused by the coupling elements must be avoided.

CHAPTER 6

ELECTRON-TUBE AMPLIFIER CIRCUITS

DIRECT-CURRENT AMPLIFIERS

OPERATION

Direct-current amplifiers are used to amplify changes in direct current or voltage as well as very-low-frequency voltages. One of its most important uses is that of a d-c electron-tube voltmeter (to be described later).

The simplest form of d-c amplifier consists of a single tube with a grid resistor across the input and a load connected in the plate circuit. The load may be an electromechanical device such as a meter, relay, or counter; or the output may be used to control the gain of an amplifier or the frequency of an oscillator, as in the automatic-frequency-control circuits of microwave receivers.

The d-c voltage change to be amplified is applied directly to the grid of the amplifier tube. Therefore, direct coupling (fig. 6-1,A) is required in the input circuit. A capacitor-input circuit (fig. 6-1, B) is also shown to indicate how the capacitor blocks the d-c signal.

In the capacitor-input circuit of figure 6-1,B, graphs of the signal voltage, grid voltage, and plate current are shown above the circuit. The applied d-c voltage charges the capacitor, and momentarily the voltage drop across R_g equals the applied voltage; and this voltage then appears between the grid and cathode of the tube. However, when the capacitor is charged up to the value of the d-c input voltage, the current stops flowing through R_g and the grid reverts to its original value, that of the

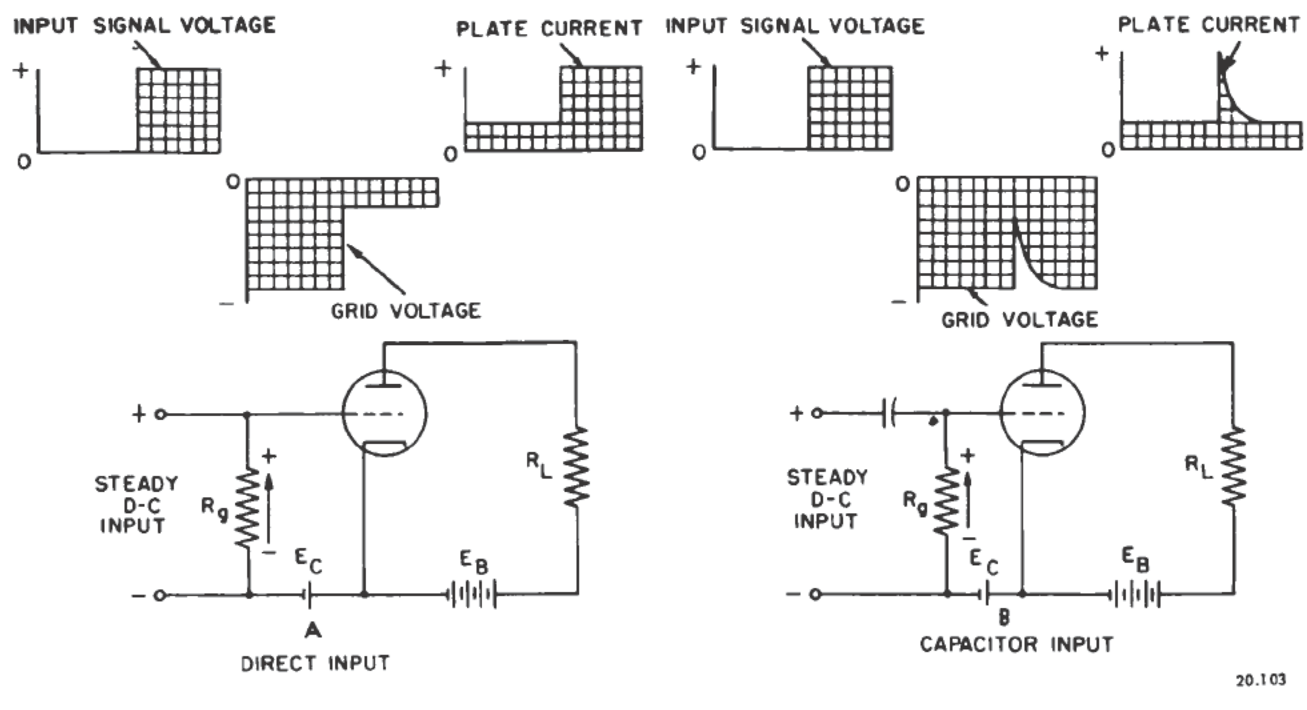


Figure 6-1.—Comparison of direct input and capacitor input to d-c amplifier.

bias voltage. Thus, except for the original surge of plate current, which occurs when the capacitor is charging, there is no increase in voltage across R_L and hence no amplification.

In the direct-coupled input circuit of figure 6-1,A, the graphs of input signal, grid voltage, and plate current are shown above the circuit. The input signal is like that in figure 6-1,B, but here the similarity ends. With no input signal the negative bias voltage is present on the grid of the tube and a steady value of plate current flows. This action causes a fixed voltage drop across R_L. When a direct voltage of the polarity indicated is applied across the input terminals, there is no blocking action by a capacitor as in the previous case. Instead, the applied signal continues as a steady voltage drop across R_g canceling a portion of the negative bias. The net bias then drops to the new value indicated in the grid-voltage graph. This reduction in grid bias causes a greater current flow in the plate circuit, and thus a greater drop appears across R_L. Thus, the increase in plate current is sustained as long as the input signal voltage exists at the corresponding level that caused the plate current to increase.

USE

As previously mentioned, one of the most important applications of a d-c amplifier is its use as a d-c electron-tube voltmeter. A typical circuit, equivalent to the one in figure 6-1,A, is shown in figure 6-2.

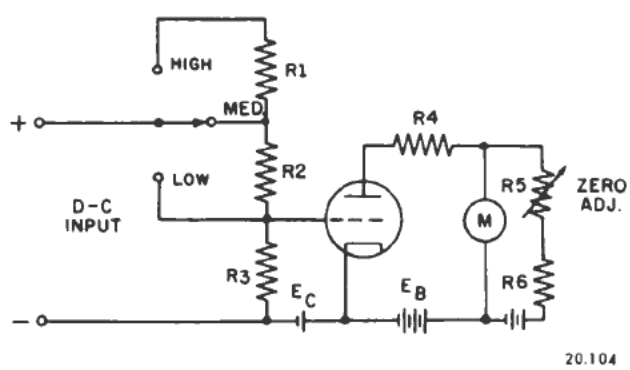


Figure 6-2.—D-c amplifier used as an electron-tube voltmeter.

The d-c voltage to be measured is applied via the range switch to one of the three taps on the voltage divider which is made up of R1, R2, and R3. The purpose of resistor R4 is to prevent damage to the tube in the event that too high a voltage is applied to the input. In the plate circuit an additional d-c voltage and two resistors, R5 and R6, are used to balance the meter current to zero when no voltage is applied to the input. One of the resistors, R5, is variable and permits zero adjustment of the meter under no-signal conditions by bucking out the voltage applied across the meter by E_B.

If a d-c voltage is then applied to the input, current flows through the meter. This current is proportional to the applied voltage and its value can be read directly on the calibrated scales (one for each range) of the meter. Various modifications permit this basic circuit to be used for measuring other quantities, or even to be incorporated in a multitester.

Another important application of the d-c amplifier is shown in figure 6-3. Here the amplifier is employed in two of the legs of a bridge circuit that is used to show the exact point of balance between two d-c voltages. If the tubes are properly matched and if there is no input signal between terminals A and B or B and C, no current flow is indicated through the meter since the IR drops across R1 and R2 are equal. When unequal signals of the polarity indicated are applied between A and B and between B and C, the grid-to-cathode voltages of the two tubes are unbalanced. This action unbalances the plate currents and the IR drops across R1 and R2. Thus the bridge becomes unbalanced and the two d-c

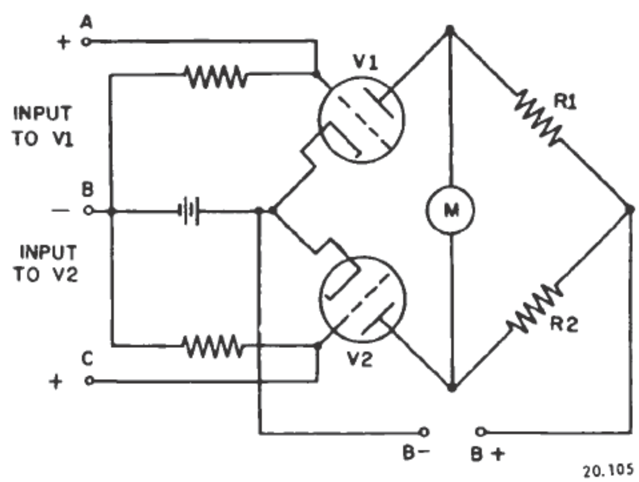


Figure 6-3.—Balanced d-c amplifier.

voltages may be compared if they are applied simultaneously with the polarities indicated.

The amount of current indicated by the meter is proportional to the difference between the two applied voltages. Since the meter has its zero position in the middle of the scale, the direction that the needle swings indicates which voltage is greater. This type of amplifier circuit has many radio and radar applications.

FEEDBACK AMPLIFIERS

As the term implies, a voltage feedback amplifier transfers a voltage from the output of the amplifier back to its input. If the signal is fed back in phase with the input signal it is called POSITIVE, DIRECT, or REGENERATIVE feedback because it adds to the voltage of the input. If the signal fed back to the input is 180° out of phase with the applied signal is called NEGATIVE, INVERSE, or DEGENERATIVE feedback because it subtracts from the input voltage. Positive feedback is used in oscillators. Negative feedback is used in amplifiers to prevent self-oscillation and to provide a more uniform output with changes in tube characteristics.

PRINCIPLE OF THE FEEDBACK AMPLIFIER

The principle of the feedback amplifier may be understood in part from a consideration of figure 6-4. A signal voltage, e_s, is applied to the input terminals, as shown in the figure. Let a portion, βe_o, of the output voltage, e_o, be fed back in series with e_s in such a way that the signal, e_{gk}, appearing between the grid and cathode is of the form

$$e_{gk} = e_s + \beta e_o \qquad (6\text{-}1)$$

where β is a fractional part of e_o.

Since the normal gain, A, of the amplifier is defined as

$$A = \frac{e_o}{e_{gk}}$$

then by transposing we have

$$e_o = Ae_{gk} \qquad (6\text{-}2)$$

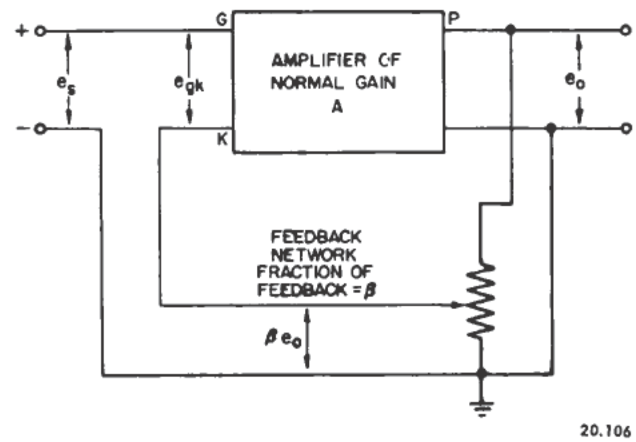


Figure 6-4.—Feedback employing series injection.

Substituting the value of e_{gk} from equation (6-1) in equation (6-2) we have

$$e_o = A(e_s + \beta e_o)$$

and solving for e_o,

$$e_o = Ae_s + \beta Ae_o$$

$$e_o - \beta Ae_o = Ae_s$$

$$e_o(1-\beta A) = Ae_s$$

$$e_o = \frac{Ae_s}{1-\beta A} \qquad (6\text{-}3)$$

NEGATIVE FEEDBACK

If βA is greater than 1, the quantity $1-\beta A$ is negative and the resulting amplification is less than it would be without feedback. Thus the amplification is said to be negative or degenerative.

The resulting gain, A_r (with feedback considered), is

$$A_r = \frac{e_o}{e_s}$$

Substituting equation (6-3) in the above expression,

$$A_r = \frac{\frac{Ae_s}{1-\beta A}}{e_s} = \frac{A}{1-\beta A}$$

The resulting amplifier gain is expressed in terms of the gain without feedback, A, and the fraction of the output, β, fed back to the input. A_r, A, and β may be complex quantities. Generally the feedback factor, βA, is so much larger than 1 that the resulting amplification, A_r, for all practical purposes may be expressed as

$$A_r = -\frac{1}{\beta}$$

Advantages

Negative feedback may be used to reduce the nonlinear distortion—that is, to make the output waveform more nearly similar to the input waveform by reducing nonlinearities that are introduced within the amplifier tube itself. This use may be understood by the following considerations:

The input signal applied to the grid of an electron-tube amplifier is amplified by an amount determined by the μ of the tube, but any nonlinearities introduced within the tube are not amplified. If a portion, βA, of the output is fed back 180° out of phase with the input, the distortion component of this feedback voltage will be amplified along with the input signal.

The amplified distortion component will tend to cancel the distortion component introduced within the tube, and the output may be practically free of nonlinear distortion. It is necessary that the distortion occur in the plate circuit of the stage across which negative feedback is to be applied, in order to separate the distortion from the desired signal.

However, the overall gain of the desired signal will also be reduced; but increasing the number of stages compensates for this reduction. Distortion caused by the flow of grid current cannot be corrected by negative feedback because this distortion occurs at the source and not within the amplifier tube.

Noise introduced within an amplifier may be reduced by negative feedback in the same manner that nonlinear distortion is reduced; and the same limitations apply—that is, for feedback to be effective, the noise to be canceled out must be generated in a tube around which the feedback is applied. Thus, thermal agitation, induced hum, microphonics, and shot effects introduced in the early stages of a receiver cannot be reduced by negative feedback unless the feedback is applied at these stages.

Negative feedback in these stages would not be practical because the amplification, particularly at the high frequencies encountered in most radar receivers, is low, and negative feedback would reduce it even more. Additional stages, each with its own circuit noises, would have to be added to make up for the reduced gain. Negative feedback is very effective, however, in reducing noises, particularly hum introduced in the high-level (high-power) stages of an amplifier. If that part of the output voltage fed back to the input is obtained by a resistance network, the resulting amplification is essentially independent of frequency.

When it is desired to have the amplification vary in some specific manner with respect to frequency, the negative feedback network through which β is obtained may be designed to attenuate the frequencies desired in the output of the amplifier. For example, if the high frequencies are to be amplified more than the low frequencies, the high frequencies must be attenuated in the feedback network more than the low frequencies. Only the low frequencies will be fed back to the grid in phase opposition to the input signal and will therefore be reduced in the output.

The overall gain of a negative feedback amplifier can be made substantially independent of the magnitude of the load impedance provided the load impedance does not interfere with the feedback signal. This may be understood if it is assumed, for example, that as the effective load resistance is reduced, the a-c component of the plate voltage tends to decrease. Accordingly, there is less negative feedback and the amplitude of the grid signal is increased. Thus, the increased signal offsets the tendency of the output voltage to drop and the overall gain is approximately constant. On the other hand, if the effective resistance of the load is increased, the a-c component of plate voltage will tend to increase and the negative feedback will increase. Thus, the amplitude of the grid signal is decreased and the tendency for the output voltage to rise is checked. Again, the overall gain is held approximately constant.

Since the gain is proportional only to the feedback factor, the gain is independent also of such factors as variations in supply voltages or the aging of tubes.

In negative-feedback, amplifiers nonlinear distortion, noise originating within the tube, and frequency distortion may be reduced. In other words, amplitude and phase characteristics can be corrected by negative feedback. Likewise, the effects of variations in loads and plate voltage supply, as well as the effects of tube aging, may be effectively counteracted. The price paid for these advantages is a reduction in gain—that is, an increase in the number of stages.

Methods of Obtaining Negative Feedback

In a practical amplifier, negative feedback may be obtained in a number of ways. It may involve one or two stages, and in rare instances more than two stages. Also, it may use voltage feedback, current feedback, or a combination of voltage and current (compound) feedback.

Voltage feedback is obtained by a voltage divider network, C1R1R2, as shown in figure 6-5. In this circuit a part of the output voltage appearing across the primary of the output transformer, T2, is fed back through the coupling network, C1R1, in series with the secondary of the input transformer, T1, as the voltage drop across R2.

To analyze the action, assume that the input signal causes the grid voltage to swing in a positive direction. The plate voltage falls as plate current increases. During this time, capacitor C1 discharges through R1 and R2. The voltage drop across R2 has a polarity that makes the top end of R2 negative with respect to ground. Thus, the voltage across R2, is in phase opposition with respect to the secondary voltage of T1. The grid-cathode signal is therefore equal to the difference in the secondary voltage of T1 and the feedback voltage across R2.

Another method of obtaining negative feedback is illustrated in figure 6-6. This method employs current feedback. Here the cathode resistor bypass capacitor has been omitted. The degenerative action may be analyzed as follows: Assume that the input signal swings the grid voltage in a positive direction. The increase in plate current causes an increase in the voltage drop across R_k. Since R_k is not bypassed, plate circuit signal currents flowing through R_k will add to the bias produced by the no-signal component. The grid-to-cathode voltage on the positive half cycle is equal to the difference in the input and the drop across R_k. The magnitude of the grid voltage swing in a positive direction is not as great as it would be without feedback because the drop across R_k is increased.

Similarly on the negative half cycle the input signal swings the grid voltage in a negative direction and plate current decreases. The decrease in current through R_k causes a a decrease in the voltage across R_k. During

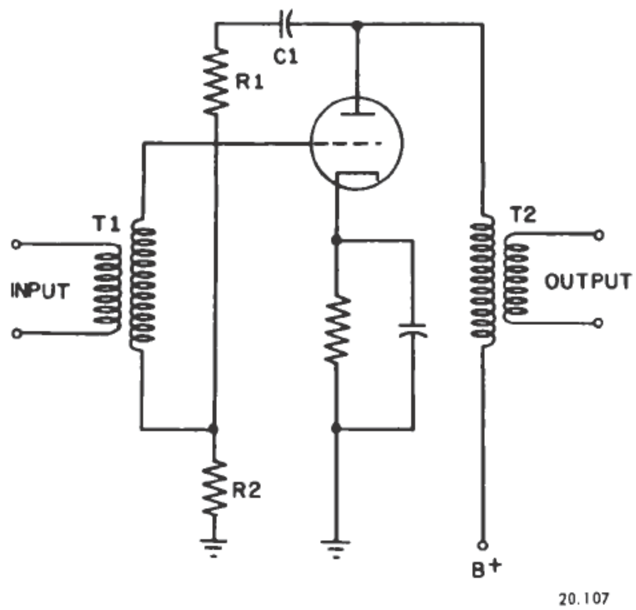


Figure 6-5.—Degenerative amplifier employing voltage feedback.

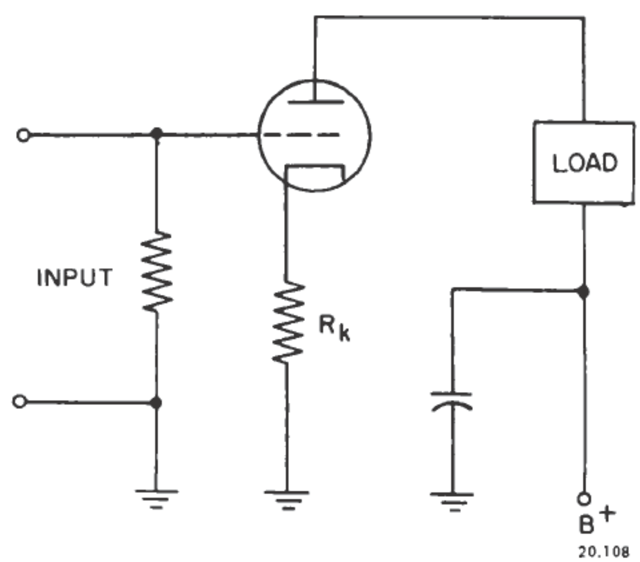


Figure 6-6.—Degenerative amplifier employing current feedback.

this half cycle the grid-to-cathode voltage is equal to the sum of the input voltage and the drop across R_k. The magnitude of the negative swing of grid voltage is less than it would be without feedback because the drop across R_k is less.

The fact that an output voltage, opposite in phase to the input voltage, may be developed across an unbypassed cathode resistor is used in designing cathode followers and phase inverters. These circuits are considered later in this chapter.

If proper phase relations are established, negative feedback involving more than one stage may be used. Figure 6-7 shows a 2-stage negative feedback amplifier using voltage feedback. In this case, special attention must be paid to the phase relations throughout the circuit.

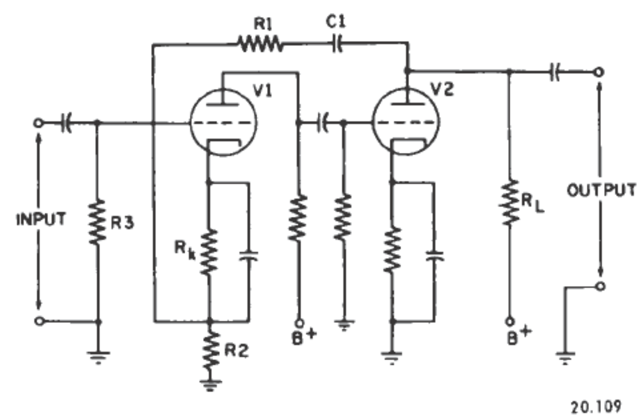


Figure 6-7.—Degenerative 2-stage amplifier employing voltage feedback.

Assume that at a given instant the input voltage makes the grid of V1 less negative. Plate current then increases in V1 and the plate voltage decreases, causing the grid of V2 to become more negative. At the same time the plate of V2 becomes more positive because of the reduction in plate current. This increase in potential causes an increase in the charge of C1. The charging current flows from ground up through R2 and R1 to the left plate of C1, making the top end of R2 more positive with respect to ground. The increase in voltage across R2 acts in series with the input and the bias across R_k to reduce the magnitude of the positive-going signal impressed on the grid. In short, the grid input signal is reduced by the amount of the feedback voltage because these two voltages are 180° out of phase.

Various combinations of voltage and current feedback circuits may be employed to satisfy specific requirements. Thus, a compound feedback using both current and voltage feedback may be used in a single stage, or current feedback may be used in one stage of a 2-stage amplifier section, and in addition, voltage feedback may be used between the stages.

POSITIVE FEEDBACK

If the quantity $1-\beta A$ is less than 1, the gain of the amplifier is increased over what it would be without feedback, and the amplifier is said to be positive or regenerative. Under these conditions the response curve is sharpened and the gain is increased, but the frequency range of uniform response is reduced. Thus, positive feedback affords both an increase in gain and an increase in selectivity.

The increase in gain, however, is accompanied by an exaggeration of any undesirable distortion or noise that was introduced within the amplifier itself. For this reason positive feedback is not used if a distortionless output is required. If the feedback factor, βA, is increased until it is equal to 1, the quantity $1-\beta A$ reduces to 0; and the resulting gain theoretically becomes infinite, or at least large enough to sustain oscillations. Under this condition no input voltage is required to obtain an output voltage and the amplifier becomes an oscillator. This aspect of feedback is discussed in chapter 8.

TUNED AMPLIFIERS

A part of the coupling circuit of a tuned amplifier is a parallel resonant circuit. These circuits are used because they offer high impedance at the desired frequency and low impedance at other frequencies, permitting the amplification of a relatively narrow band of frequencies. The final benefit of all sharply tuned circuits is the placement of many narrow bands (within the space previously held by a single-but-wider band of frequencies) which are available by selection for several users. In addition, the limitations imposed on untuned amplifiers by interelectrode and distributed capacitances are used to an advantage because these capacitances become a part of the tuned circuit. The amplifiers may be single- or double-tuned, depending on whether the plate circuit only or both the plate and the grid circuits contain parallel-tuned circuits.

The three basic tuned-amplifier circuits are shown in figure 6-8. Each circuit has specific

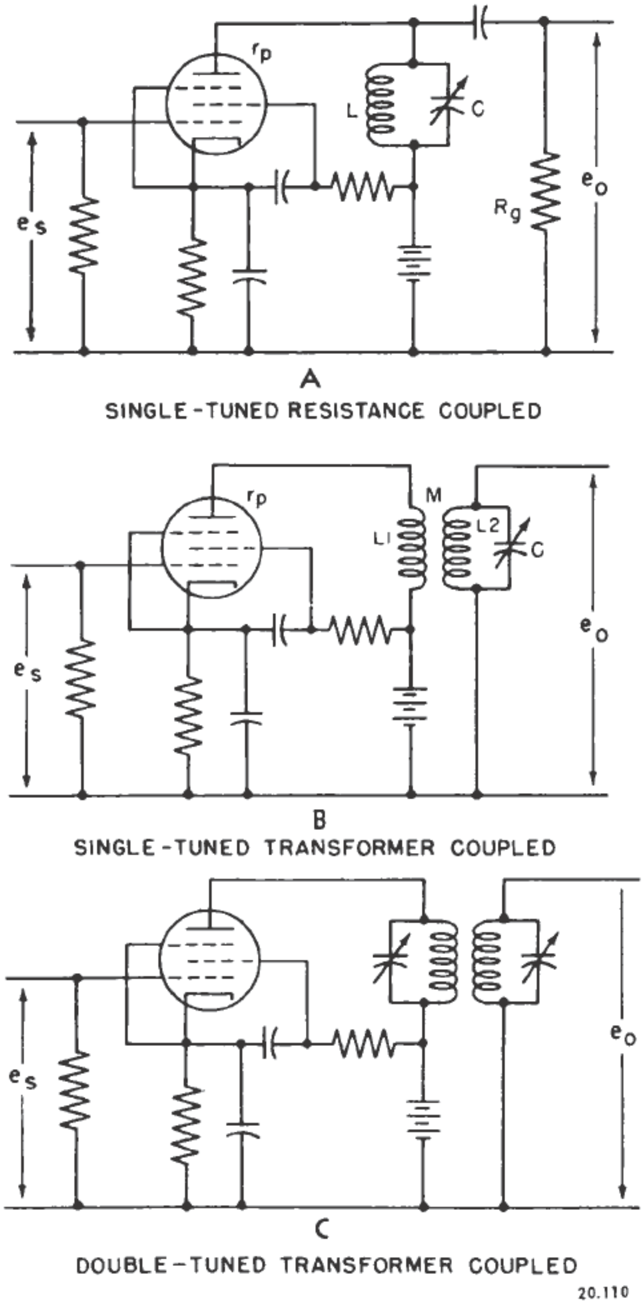


Figure 6-8.—Basic tuned-amplifier circuits.

applications. For example: the single-tuned resistance-coupled amplifier (fig. 6-8,A) might be used with various modifications to deliver energy from a class-C power amplifier; the single-tuned transformer-coupled amplifier circuit (fig. 6-8,B) might be used as a class-A r-f voltage amplifier (preselector) ahead of the first detector in a superheterodyne receiver; and the double-tuned transformer-coupled amplifier circuit (fig. 6-8,C) might be used as an intermediate frequency stage of amplification between the first and second detectors in the superheterodyne receiver.

A better understanding of how these circuits operate may be had if the equivalent circuit of each basic amplifier is presented along with a brief analysis of the circuit.

ANALYSIS OF SINGLE-TUNED R-C COUPLING

The simplified equivalent circuit of figure 6-8,A, is shown in figure 6-9. In this figure, capacitor C includes the capacitance of the tank tuning capacitor, stray capacitances, and the output and input tube capacitances. The capacitance of the series coupling capacitor is large and its impedance is negligible, therefore it has been omitted from the equivalent circuit. The tank coil, L, resonates with C at the desired frequency. The plate resistance, r_p, of the amplifier stage and the grid leak resistance, R_g, of the next stage are in parallel

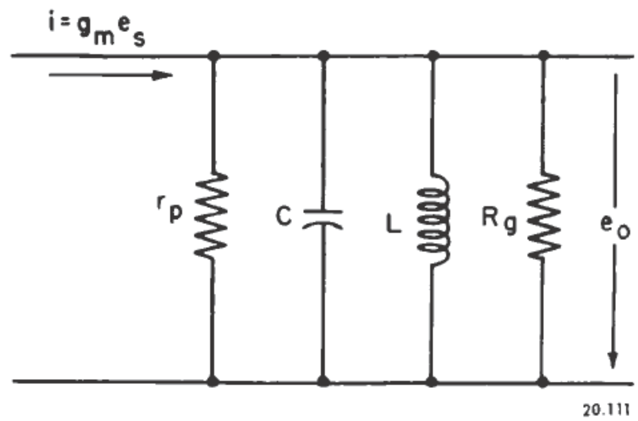


Figure 6-9.—Equivalent circuit of a single-tuned amplifier—constant-current generator form.

with the tank. The amplification or voltage gain (A) is

$$A = \frac{e_o}{e_s} = \frac{iZ_L}{e_s} = \frac{g_m e_s Z_L}{e_s} = g_m Z_L$$

where Z_L is the combined impedance of the parallel circuit.

At resonance, the impedance of the parallel resonant tank, CL, is large and resistive and is equal to Q times the reactance of either C or L. Therefore,

$$R_o = X_L Q = \omega LQ = 2\pi f_o LQ$$

where Q is the effective value of the tank circuit alone (r_p and R_g omitted), $\omega = 2\pi f_o$ and f_o is the resonant frequency.

The total impedance of the circuit, including the impedance of the resonant tank, CL, in parallel with r_p and R_g is

$$Z_L = \frac{1}{\frac{1}{r_p} + \frac{1}{2\pi f_o LQ} + \frac{1}{R_g}}$$

multiplying numerator and denominator by $2\pi f_o LQ$

$$= \frac{2\pi f_o LQ}{1 + \frac{2\pi f_o LQ}{r_p} + \frac{2\pi f_o LQ}{R_g}}$$

The amplification at resonance becomes

$$A = \frac{g_m 2\pi f_o LQ}{1 + \frac{2\pi f_o LQ}{r_p} + \frac{2\pi f_o LQ}{R_g}}$$

Normally, for the type of tube generally used, R_g is high with respect to the impedance ωLQ, of the resonant tank, and for pentodes, r_p is very high.

The approximate amplification may be expressed as

$$A \cong g_m 2\pi f_o LQ$$

ANALYSIS OF SINGLE-TUNED TRANSFORMER COUPLING

The derivation of an expression for the voltage gain of the single-tuned transformer-coupled amplifier may be more easily understood if the following considerations are stated first: the voltage induced in the secondary, L2 (fig. 6-10) by the primary current, i_p, is equal to $2\pi f_o M i_p$, where M is the mutual inductance between the primary and the secondary. This induced voltage is multiplied by a factor, Q, to obtain the output voltage across C because of the resonant voltage rise in the tuned circuit.

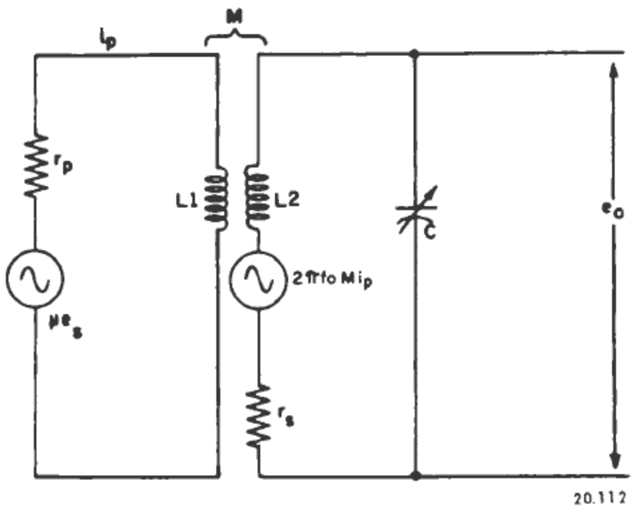


Figure 6-10.—Equivalent circuit of a single-tuned transformer-coupled amplifier—constant-voltage generator form.

The output is now established as

$$e_o = Q 2\pi f_o M i_p$$

Because r_p is large with respect to $\omega L1$ and the effect of the presence of the tuned secondary on the primary is slight (low coupling), the primary current is dependent upon μe_s and r_p.

The expression for primary current then becomes

$$i_p = \frac{\mu e_s}{r_p}$$

and, therefore, substituting for i_p (in the equation for e_o given above) we get:

$$e_o = \frac{Q 2\pi f_o M \mu e_s}{r_p}$$

Since $\frac{\mu}{r_p}$ is equal to g_m,

$$e_o = Q2\pi f_o M g_m e_s$$

and

$$\text{gain} = \frac{e_o}{e_s} = \frac{Q2\pi f_o M g_m e_s}{e_s} = Q2\pi f_o M g_m$$

Thus, a high-gain amplifier has a low-loss tuned circuit and requires a tube having a high mutual conductance.

ANALYSIS OF DOUBLE-TUNED TRANSFORMER COUPLING

As indicated in figure 6-11, the double-tuned transformer-coupled amplifier has a band-pass characteristic which depends in part on the degree of coupling and in part on the circuit Q's.

Under proper operating conditions, essentially uniform amplification of a relatively narrow band of frequencies may be achieved and amplification of frequencies outside this band may be sharply reduced.

Since the slope of the response curve is not perfectly vertical, the circuit cannot completely discriminate against frequencies just outside the desired channel without also attenuating to some extent the frequencies at the upper and lower limits of the pass band. However, double-tuned amplifiers approach an ideal band-pass characteristic much more closely than single-tuned amplifiers, which have rounded response curves.

To find the gain of the double-tuned transformer-coupled circuit the concept of coupled impedance is used. As previously described, the effect of the presence of the tuned secondary on the primary is the same as if an impedance $\frac{(2\pi f_o M)^2}{Z_s}$ had been added in series with the primary. This idea is used in the example in figure 6-12,A.

In this example the following values are given:

The operating frequency, f_o, is 260 kc.

The input signal voltage, e_g, is 9.5×10^{-5} volts (rms).

The amplification factor, μ, is 1280.

The mutual conductance, gm, is 1600 micromhos.

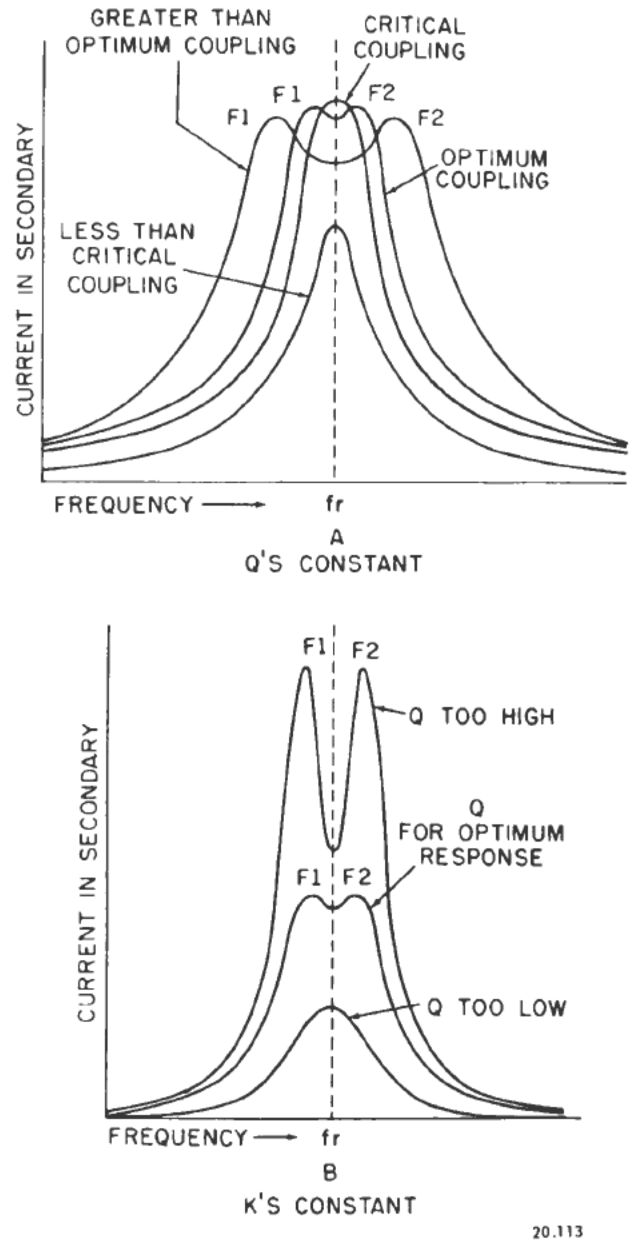


Figure 6-11.—Response curves for a double-tuned transformer-coupled amplifier.

The plate resistance, r_p, is 800,000 ohms.

$L_1 = L_2 = 4 \times 10^{-3}$ henry inductance.

$Q_p = Q_s = 50$ quality.

$k = 0.03$ coefficient of coupling.

$C_1 = C_2 = 94$ micromicrofarads.

Question: Find the output voltage at 260 kc and the voltage gain of the stage.

The problem is simplified by substituting an equivalent series circuit (fig. 6-12,B) for the original circuit (fig. 6-12,A). After the

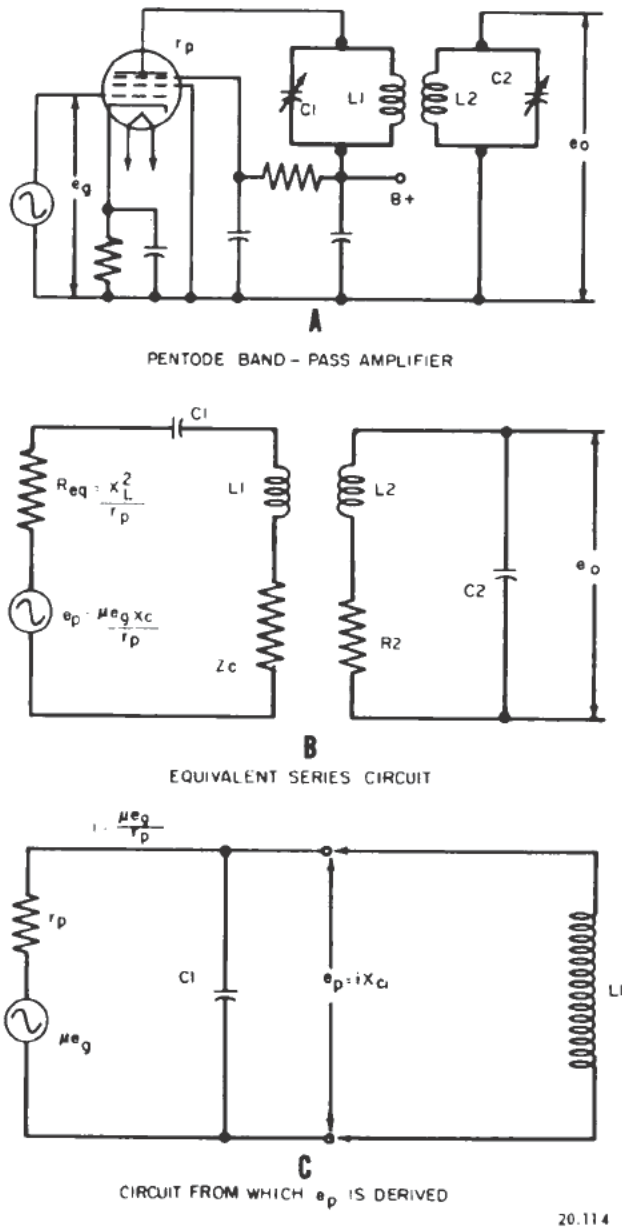


Figure 6-12.—Circuits used in calculating the voltage gain of a pentode i-f band-pass amplifier.

equivalent values of this circuit are obtained, the calculations for output voltage (across C_2) and voltage induced in the secondary by the action of the primary current are similar to those given in chapter 4 on tuned circuits.

In the original circuit (fig. 6-12,A), the pentode plate resistance acts in shunt with the tank, L_1 and C_1, and loads it. The effect is the same as if an equivalent series resistance $\frac{X_L{}^2}{r_p}$ had been added in series with the tank and the shunt resistance removed (derived below). The equivalent series resistance, R_{eq}, produces the same effect on the primary circuit impedance (fig. 6-12,B) as the plate resistance, r_p, produces in figure 6-12,A. The relation between equivalent series and shunt resistances was derived in the training course *Basic Electricity*, and for circuits in which $\tan\theta = 10$ or more the relation is sufficiently accurate for slide rule calculations. The derivation is given again, see figure 6-13.

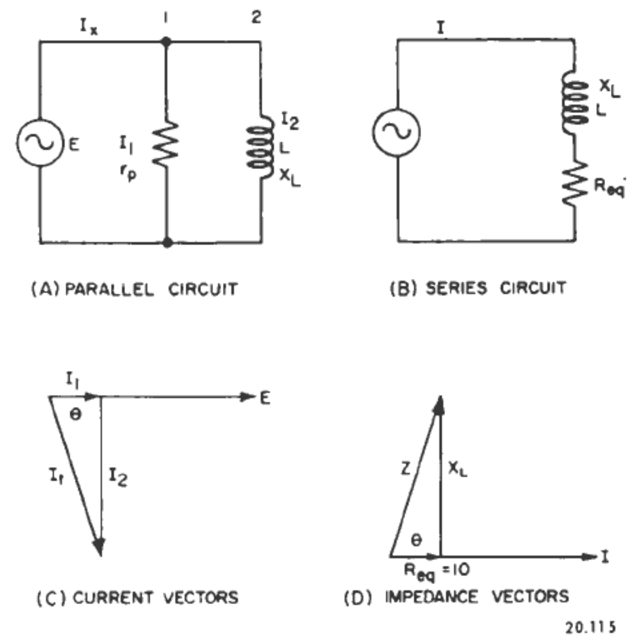


Figure 6-13.—Equivalent parallel and series circuits.

The parallel circuit (fig. 6-13,A) and the series circuit (fig. 6-13,B) are equivalent circuits. Both circuits present the same impedance to the source voltage E. The total current in the parallel circuit is equal to the total current in the series circuit. Both currents lag E by the same phase angle θ . The current vectors for the parallel circuit (fig. 6-13,C) form a current triangle which is similar to the impedance triangle formed by the impedance vectors for the series circuit (fig. 6-13,D).

In the parallel circuit triangle (fig. 6-13,C)

$$\cos\theta = \frac{I_1}{I_t} = \frac{I_1}{\sqrt{I_1{}^2 + I_2{}^2}}$$

The quantity under the radical is the sum of the squares of the currents in the two branches. Since I_1 is small in relation to I_2 it is possible to neglect I_1 under the radical sign without introducing appreciable error. For example if I_1 is 1 ampere and I_2 is 10 amperes the total current is $\sqrt{1^2 + 10^2} = \sqrt{101}$ or 10.05 the error is 0.05 if the current in branch 1 is neglected. An error of 0.05 in 10.05 is an error of 5 parts in about 1,000 or 0.5 percent. It is not possible to neglect I in the numerator because it is the only factor present and its use is required to find cos θ. In this example: $\tan\theta = 10$ (where $\tan\theta = \frac{I_2}{I}$) and θ is 84.3°. If θ increases beyond 84.3° toward 90°, $\tan\theta$ increases above 10 toward a large value and the error decreases rapidly. Thus, the total current is approximately

$$I_t = \sqrt{I_2^2} = I_2$$

and

$$\text{Cos}\ \theta = \frac{I_1}{I_2} \qquad (6\text{-}4)$$

The current in branch 1 is

$$I_1 = \frac{E}{r_p} \qquad (6\text{-}5)$$

The current in branch 2 is

$$I_2 = \frac{E}{X_L} \qquad (6\text{-}6)$$

(The coil contains negligible resistance and the impedance of the coil is equal to the coil reactance, X_L)

Substituting equations 6-5 and 6-6 in equation 6-4,

$$\text{Cos}\ \theta = \frac{\frac{E}{r_p}}{\frac{E}{X_L}} = \frac{X_L}{r_p} \qquad (6\text{-}7)$$

In the equivalent series circuit triangle (fig. 6-13,D)

$$\text{Cos} = \frac{R_{eq}}{Z} = \frac{R_{eq}}{\sqrt{R_{eq}^2 + X_L^2}}$$

The quantity under the radical is equal to the sum of the squares of R_{eq} and X_L. Since R_{eq} is small in relation to X_L the same reasoning can be applied here as for the current triangle (fig. 6-13,C). When $\tan\theta$ is 10 or more, X_L is 10 or more times as large as R_{eq} (fig. 6-13,D) and the impedance of the equivalent series circuit is approximately

$$Z = \sqrt{X_L^2} = X_L$$

and

$$\cos\theta = \frac{R_{eq}}{X_L} \qquad (6\text{-}8)$$

Because the parallel and series circuits are equivalent, $\cos\theta$ in equation 6-7 is equal to $\cos\theta$ in equation 6-8 and the two expressions may be equated as follows:

$$\frac{X_L}{r_p} = \frac{R_{eq}}{X_L} \qquad (6\text{-}9)$$

Transposing equation 6-9 and solving for the equivalent series resistance

$$R_{eq} = \frac{X_L^2}{r_p} \qquad (6\text{-}10)$$

It should be remembered that this expression is a close approximation and sufficiently accurate for slide rule calculations.

In the example in figure 6-12, the pentode plate resistance is the equivalent shunt resistance, r_p, acting in parallel with the coil, L_1. In this example the frequency of the applied voltage is 260,000 cycles and the inductance of L_1 is 0.004 henry. At this frequency the inductive reactance of the coil is

$$X_L = 2\pi f_o L = 6.28 \times 260 \times 10^3 \times 4 \times 10^{-3}$$
$$= 6530 \text{ ohms}$$

The pentode plate resistance, r_p, is 800,000 ohms.

Substituting, $X_L = 6530$ and $r_p = 800{,}000$ in equation 6-10

The equivalent series resistance is

$$R_{eq} = \frac{6530^2}{800{,}000} = 53.2 \text{ ohms}$$

Thus the effect of the pentode plate resistance of 800,000 ohms in parallel with L_1 is the same as though this resistance had been removed and an equivalent series resistance of 53.2 ohms had been added in series with L_1.

The equivalent voltage, e_p, acting in the equivalent series circuit (fig. 6-12,B) is derived from the circuit of figure 6-12,C. In this circuit L_1 is disconnected from C_1 and the voltage across C_1 calculated. This voltage is the equivalent voltage e_p. It is equal to the iX_c drop across C_1.

The current, i, depends on the input voltage e_g and the circuit impedance. The voltage e_g acting in the grid circuit (fig. 6-12,A) is equivalent to μe_g acting in the plate circuit (fig. 6-12,C). In this example

$$\mu e_g = 1280 \times 9.5 \times 10^{-5} = 0.122 \text{ volt}$$

The capacitive reactance of C_1 is equal to the inductive reactance of L_1 at the resonant frequency.

As a check

$$X_c = \frac{1}{2\pi f_o C_1} = \frac{1}{6.28 \times 260 \times 10^3 \times 94 \times 10^{-12}}$$

$$= 6530 \text{ ohms}$$

In the circuit of figure 6-12,C, the X_c ohms of C_1 are relatively small compared with the pentode plate resistance r_p and can be neglected when calculating the current, i.

$$i = \frac{\mu e_g}{r_p} \qquad (6\text{-}11)$$

Substituting, $\mu e_g = 0.122$ and $r_p = 800{,}000$ in equation 6-11

$$i = \frac{0.122}{800{,}000} = 0.153 \times 10^{-6} \text{ ampere}$$

The equivalent voltage, e_p, is

$$e_p = iX_c \qquad (6\text{-}12)$$

Substituting, $i = 0.153 \times 10^{-6}$ and $X_c = 6530$ in equation 6-12

$$e_p = 0.153 \times 10^{-6} \times 6530 = 1 \times 10^{-3} \text{ volt}$$

Thus an input of 1×10^{-3} volt acts as the generator voltage in the primary circuit (fig. 6-12,B). At the resonant frequency of 260 kc the X_L ohms of L_1 cancel the X_c ohms of C_1 and the total impedance of the primary consists of three quantities, all resistive. These quantities are: (1) R_{eq} (previously described); (2) the resistance associated with the coil; and (3) the coupled impedance due to the tuned secondary. The procedure is to find the total primary impedance, the primary current, the secondary voltage, the secondary current, the output voltage, and finally the voltage gain of the stage.

The mutual inductance of L_1 and L_2 is

$$M = K\sqrt{L_1 L_2} \qquad (6\text{-}13)$$

Substituting, $K = 0.03$ and $L_1 = L_2 = 4 \times 10^{-3}$ in equation 6-13

$$M = 0.03\sqrt{4 \times 10^{-3} \times 4 \times 10^{-3}}$$

$$= 0.13 \times 10^{-3} \text{ henry}$$

The mutual inductive reactance is

$$X_M = 2\pi f_o M \qquad (6\text{-}14)$$

Substituting, $f_o = 260 \times 10^{-3}$ and $M = 0.12 \times 10^{-3}$ in equation 6-14

$$X_M = 6.28 \times 260 \times 10^3 \times 0.12 \times 10^{-3}$$

$$= 196 \text{ ohms}$$

The inductive reactance of each coil as previously calculated is 6,530 ohms.

The capacitive reactance of C_2 is equal to that of C_1. As previously calculated X_c is 6,530 ohms.

At resonance the X_c ohms of C_2 cancel the X_L ohms of L_2 and the series impedance of the secondary circuit is the resistance of the coil. The losses of the capacitor are considered negligible. The resistance of the coil L_2 is

$$R = \frac{X_L}{Q} \qquad (6\text{-}15)$$

Substituting, $X_L = 6530$ and $Q = 50$ in equation 6-15.

$$R = \frac{6530}{50} = 130.6 \text{ ohms}$$

Thus the impedance, Z_s, of the secondary circuit at resonance is 130.6 ohms.

The coupled impedance is

$$Z_c = \frac{X_M{}^2}{Z_s} \tag{6-16}$$

Substituting, X_M = 196 and Z_s = 130.6 in equation 6-16

$$Z_c = \frac{(196)^2}{130.6} = 295 \text{ ohms}$$

The total primary impedance, Z_p as previously described is the sum of R_{eq}, the resistance R_1 associated with coil L_1 and the coupled impedance Z_c. Expressed as an equation

$$Z'_p = R_{eq} + R_1 + Z_c \tag{6-17}$$

Substituting, R_{eq} = 53.2, R_1 = 130.6 and Z_c = 295 in equation 6-17

$$Z'_p = 53.2+130.6+295 = 478.8 \text{ ohms}$$

The primary current, i_p, is equal to the ratio of the applied voltage, e_p, to the primary impedance Z'_p. Expressed as an equation

$$i_p = \frac{e_p}{Z'_p} \tag{6-18}$$

Substituting, $e_p = 1x10^{-3}$ and Z'_p = 478.8 in equation 6-18

$$i_p = \frac{1x10^{-3}}{478.8} = 2.09x10^{-6} \text{ ampere}$$

The voltage e_s induced in the secondary by the primary current i_p is

$$e_s = X_M i_p \tag{6-19}$$

Substituting, X_M = 196 and $i_p = 2.09x10^{-6}$ in equation 6-19

$$e_s = 196x2.09x10^{-6} = 410x10^{-6} \text{ volt}$$

The secondary current i_s is equal to the ratio of secondary voltage to secondary impedance.

Expressed as an equation

$$i_s = \frac{e_s}{Z_s} \tag{6-20}$$

Substituting, $e_s = 410x10^{-6}$ and Z_s = 130.6 in equation 6-20.

$$i_s = \frac{410x10^{-6}}{130.6} = 3.14x10^{-6} \text{ ampere}$$

The output voltage e_o is equal to the product of the secondary current, i_s, and the capacitive reactance X_c of C_2.

Expressed as an equation

$$e_o = i_s X_{c2} \tag{6-21}$$

Substituting, $i_s = 3.14x10^{-6}$ and X_{c2} = 6530 in equation 6-21

$$e_o = 3.14x10^{-6}x6530 = 2.05x10^{-2} \text{ volt}$$

The voltage gain of the stage is

$$\text{v.g.} = \frac{e_o}{e_g} \tag{6-22}$$

Substituting, $e_o = 2.05x10^{-2}$ and $e_g = 9.5x10^{-5}$ in equation 6-22

$$\text{v.g.} = \frac{2.05x10^{-2}}{9.5x10^{-5}} = 216$$

Thus a pentode amplifier having an amplification factor of 1,280 provides a voltage gain of approximately 216 at an operating frequency of 260 kc.

A better understanding of the gain at resonance, as well as the response throughout the pass band of a double-tuned transformer-coupled stage, may be gained from a further consideration of the curves of secondary current versus frequency, as shown in figure 6-11.

When the coefficient of coupling in figure 6-11,A, is low, the response is sharply peaked at the resonant frequency and the pass band is very narrow. As the coupling is increased to the critical value, maximum current flows in the secondary, and the output voltage across the secondary is also at its maximum. At this point (critical coupling)

$$k = \frac{1}{\sqrt{Q_p Q_s}}$$

and if the Q's are equal

$$k = \frac{1}{Q}$$

The pass band is still relatively narrow and would attenuate the side band frequencies farthest removed from the resonant frequency.

If the coupling is further increased until the optimum value is reached the gain is still relatively high; but the pass band has been increased and the response is essentially uniform. At this point (optimum coupling)

$$k = \frac{1.5}{\sqrt{Q_p Q_s}}$$

and if the Q's are equal

$$k = \frac{1.5}{Q}$$

As the coupling is again increased the humps at F_1 and F_2 are well defined and the gain at resonance is considerably reduced. Although the pass band is now much wider, the gain throughout the band is not sufficiently uniform.

The two humps in the curve are due to the reactance that is coupled into the primary on each side of resonance as the coupling is increased. Below resonance this reactance is inductive, and above resonance it is capacitive. For the same frequency the coupled reactance has the opposite sign to that of the primary and the impedance of the primary is therefore reduced. Accordingly, there is an increase in primary current at frequencies slightly off resonance, and a corresponding increase in secondary induced voltage and current at these frequencies.

The width of the pass band may be as important as the response within the band. The approximate width is

$$\text{width of pass band} = kf_o$$

where f_o is the resonant frequency which is the center of the pass band.

The frequencies at the two humps, F_1 and F_2, which define the practical lower and upper limits of the pass band are determined by the following equations:

$$F_1 = \frac{f_o}{\sqrt{1+k}}$$

$$F_2 = \frac{f_o}{\sqrt{1-k}}$$

Figure 6-11,B, shows the effects of varying the Q while maintaining a constant coefficient of coupling. Actually, the desired response curve could be achieved by the proper manipulation of both k and Q because they are interrelated.

From the foregoing equations it is seen that in order for the pass band to be wide, k must be large and the circuit Q's small. However, the proper relation between k and the Q's is essential if both the desired bandwidth and the desired response within the bandwidth are to be maintained.

To increase k, the coils are brought closer together. To lower the Q's, the coils are shunted with suitable resistors. The lower the shunting resistance, the lower will be the circuit Q.

MODIFICATIONS FOR HIGH-FREQUENCY OPERATION

When the frequency to be amplified is very high, conventional r-f and i-f amplifiers require special circuit modifications. The circuit of a high-frequency tuned amplifier used at radar frequencies and requiring such modification is shown in figure 6-14. Tuning is accomplished by a small coil which resonates with its own distributed capacitance and the interelectrode capacitance of the tube. It is tuned by either a brass slug which acts as a short-circuited turn or by a powdered iron core. Both methods vary the effective inductance of the coil and thus the resonant frequency of the amplifier stage.

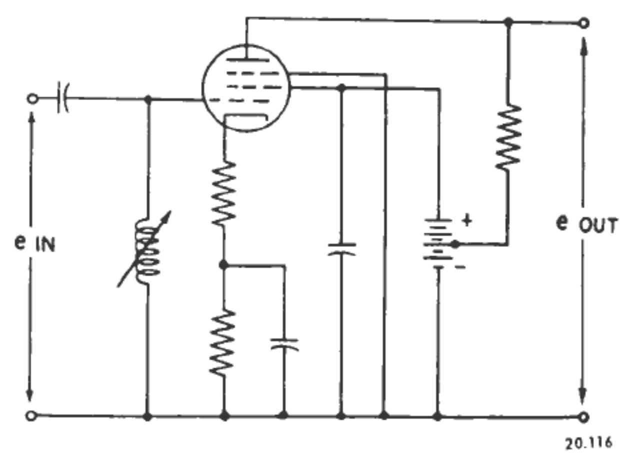


Figure 6-14.—Tuned amplifier for operation at radar i-f frequencies (60mc).

Other features of this amplifier may be summarized as follows: The interelectrode capacitance of the tube must be low so that the coil will be large enough to be tuned conveniently. Since a wide band of frequencies is to be amplified, the load resistance must be low also. The tube must have a high transconductance.

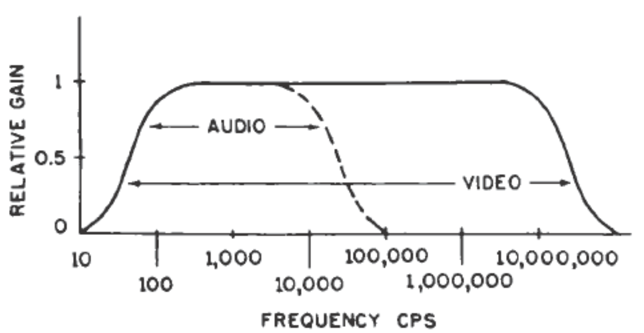
The input capacitive reactance of a tube, such as the one in figure 6-14 varies inversely with the frequency of the a-c component of grid voltage, thus causing frequency distortion. Therefore, a part of the cathode resistance is unbypassed so that the correct amount of degeneration to overcome this distortion will be introduced.

Because of the low plate load, the plate supply voltage is also low, in this case lower even than the screen-grid voltage. The suppressor is returned to ground rather than to the cathode to increase the negative potential of the suppressor with respect to the plate.

Double tuning is seldom used because of the restrictions imposed by this type of amplifier on bandwidth.

VIDEO AMPLIFIERS

Video amplifiers such as those used in modern television sets are designed to give essentially uniform amplification of all frequencies from 30 cps to over 4 mc. Audio-frequency amplifiers, on the other hand, are considered good if they have a relatively flat response between 30 and 15,000 cps. Figure 6-15 compares the response of the two types of amplifiers.



20.117

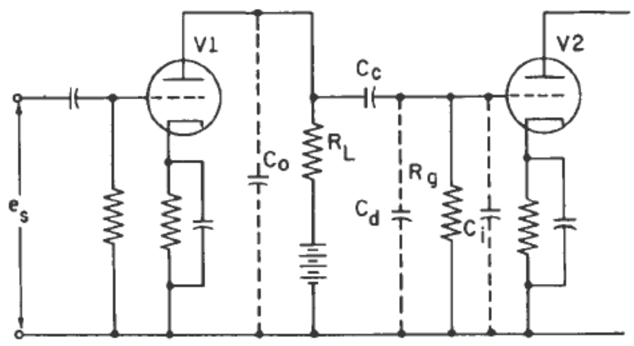
Figure 6-15.—Relative frequency response of audio and video amplifiers.

It has been shown that resistance coupling gives the best response over a wide range even at audio frequencies. This is especially significant in video amplifiers; however, to extend the range at both the low- and high-frequency ends, special compensation is necessary in video amplifiers.

RESISTANCE-CAPACITANCE COUPLED CIRCUITS

The R-C coupled amplifier circuit shown in figure 6-16 indicates the effects that must be overcome if the range is to be extended on both ends of the frequency spectrum.

The high-frequency response is limited by the interelectrode output capacitance, C_o, the distributed wiring capacitance, C_d, and the input interelectrode capacitance, C_i. These three capacitances, acting in parallel, shunt the load, R_L, and reduce the output at the high frequencies.



20.118

Figure 6-16.—R-C coupled amplifier.

The low-frequency response is limited by the reactance of the coupling capacitor, C_c. Thus at the lower frequencies the divider action of C_cR_g reduces the voltage available between the grid and cathode of V2.

It is obvious that the interelectrode capacitances of the tubes and the distributed capacitances of the wiring must be kept as low as practicable. Keeping the distributed capacitances low requires careful placement of the tubes in order to keep the leads as short as possible.

HIGH-FREQUENCY COMPENSATION

There are various methods of extending the range of a video amplifier at the high-frequency end of the range, but perhaps the simplest and most effective is the shunt-peaked method, shown in figure 6-17,A. As mentioned previously, the gain at high frequencies is reduced because the load is shunted by C_o, C_d, and C_i. These same values can be made to extend the range if a small inductor, L1, is inserted in series with the load resistor, R_L, to form a parallel resonant circuit. If the value of L1 is properly chosen so that the circuit will be in resonance at the point where the response curve begins to fall appreciably, the range can be extended. The value of L1 is critical. If the value is not correct, the amplification may be increased before the point at which the response curve begins to fall, with the result that frequency distortion ensues.

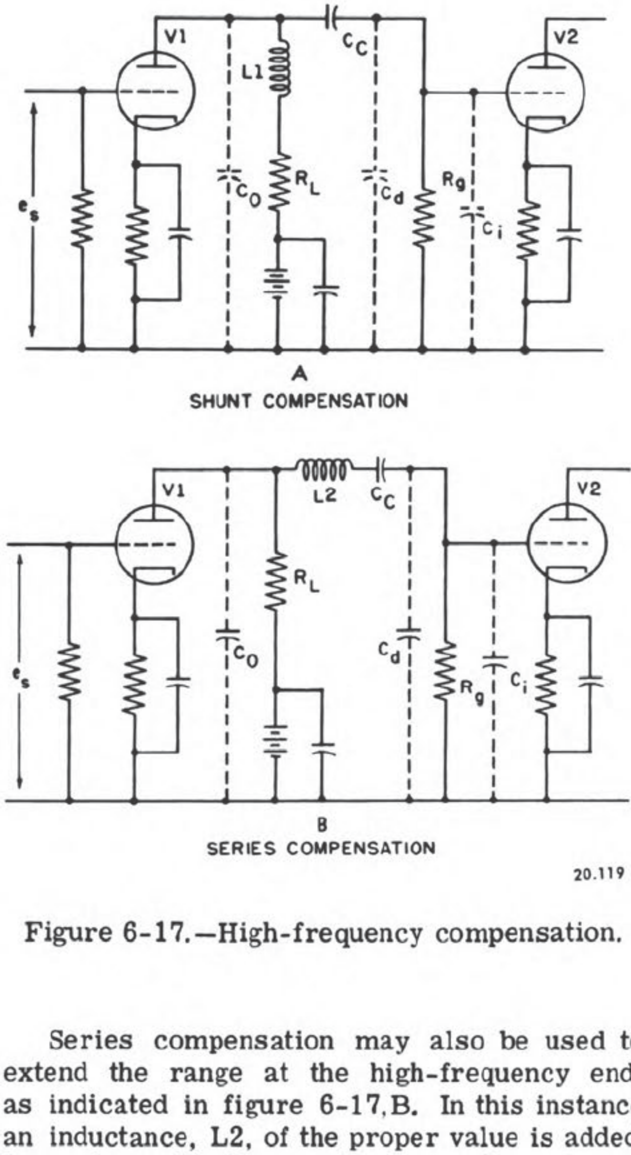


Figure 6-17.—High-frequency compensation.

Series compensation may also be used to extend the range at the high-frequency end, as indicated in figure 6-17,B. In this instance an inductance, L2, of the proper value is added in series with the coupling capacitor, C_c, so that a series-resonant circuit is formed with the parallel combination of C_d and C_i. At resonance, increased current will flow through these capacitances and larger output voltage will be applied between the grid and cathode of V2.

The high-frequency peaking effect of shunt compensation in addition to the increased gain of series compensation may be obtained if both of these methods of compensation are used in the same coupling circuit. There are other factors, however, such as the transient response, which have to be considered in a network such as this.

LOW-FREQUENCY COMPENSATION

At low frequencies, the distributed and interelectrode capacitances may be neglected but the reactance of the coupling capacitor becomes increasingly important. Since the reactance of this capacitor is

$$X_C = \frac{1}{2\pi f C_c}$$

it becomes appreciable at low frequencies. Consider a voltage divider made up of C_c and R_g (fig. 6-18) in which the reactance of C_c is large, as it would be at low frequencies. More of the voltage would then appear across C_c and less would appear across R_g. Since in some practical applications the voltage gain must be maintained within 70 percent of the mid-frequency gain, this loss at low frequencies could not be tolerated.

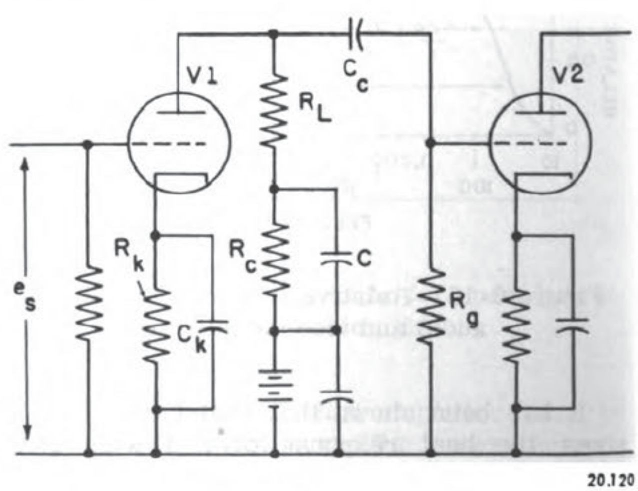


Figure 6-18.—Low-frequency compensation.

The capacitance of C_c could be increased, but such a procedure would increase the stray capacitance and thus cut down the high-frequency gain.

Another factor that is perhaps more important than the loss of gain is the large shift in phase that occurs at these frequencies. If 10 stages of video amplification are used, a phase shift (lead) of about 2° is all that can be permitted for each stage.

The phase shift can be reduced by employing a large value of coupling capacitance and the largest permissible grid-leak resistance; but both of these expedients have their disadvantages also.

Amplifiers that do not have as critical requirements as video amplifiers may be made to operate satisfactorily by using large cathode and screen bypass capacitors and a coupling capacitor as large as practicable. Video amplifiers, however, require special compensation at the low frequencies.

Both the loss in gain and the increase in phase shift may be corrected by dividing the load resistance into two parts and bypassing one part with a capacitor. A circuit employing this method of low-frequency compensation is shown in figure 6-18.

The load resistance is made up of two parts, R_L and R_c, of which R_c is bypassed by C. At the higher frequencies the load is effectively R_l because R_c is bypassed by C which has a low reactance at these frequencies. At low frequencies, however, C offers a high impedance and the load is effectively $R_L + R_c$. This effective load increase causes a greater proportion of the plate signal to appear as output and thus counteracts the normal drop at the low frequencies. Care must be used in selecting the values of R_c and C so that uniform gain can be extended into the lower frequencies beyond the point where the gain begins to fall off. Distortion will occur if the gain takes place before the curve begins to fall off.

Another component influencing low-frequency gain is the cathode-bypass capacitor. In order to prevent degeneration from occurring at the lower frequencies because of inadequate shunting, the capacitance of the cathode-bypass capacitor must be great enough to offer low impedance (with respect to the cathode resistor) at the lowest frequency to be amplified. Thus in video amplifiers the capacitances of cathode-bypass capacitors are many times those of comparable audio amplifiers.

CATHODE FOLLOWERS

To achieve uniform response over a wide frequency range it has been shown that an amplifier should have a low effective input capacitance and a low effective load impedance. The overall response may also be improved by the use of degenerative feedback. The cathode follower possesses these qualities, and in addition it may be used to match the impedance of one circuit to that of another.

The cathode follower is a single-stage class A degenerative amplifier the output of which appears across the unbypassed cathode resistor. The high input impedance (no grid current) and the low output impedance make it particularly useful for matching a high-impedance source to a low-impedance load. Thus, the cathode follower might be used between a pulse-generating stage and a transmission line whose effective shunt capacitance might be great enough to cause objectionable effects. More power, of course, can be delivered when the source is matched to the load. For example, a conventional amplifier having high output impedance would supply less power to a low-impedance coaxial line than would a cathode follower having an output impedance that corresponds to the load impedance.

The advantages obtained by the use of a cathode follower can be had only at the price of a voltage gain that is less than unity; however, the circuit is capable of producing power gain.

As the name implies, the output voltage FOLLOWS the input voltage—that is, it has not only the same waveform but also the same instantaneous polarity (phase).

CIRCUIT OPERATION

A conventional cathode follower is shown in figure 6-19,A. Under no-signal conditions a certain amount of plate current flows through R_k, and this flow establishes the normal bias. When a positive-going signal is applied to the grid, the plate current increases. This increase causes an increase in the voltage drop across R_k, giving the cathode a higher positive potential with respect to ground than it had under the no-signal condition. When a negative-going signal is applied to the grid, the opposite effect occurs. Thus, the output polarity FOLLOWS the polarity of the voltage applied between grid and ground.

Since R_k is not bypassed, degeneration occurs both on the positive half cycle when plate current through R_k increases the bias and on the negative half cycle when plate current through R_k decreases the bias. During the positive half cycle, the increase in bias subtracts from the input signal and reduces the amplitude of the grid-to-cathode voltage. Also during the negative half cycle, the bias adds to the input signal and the accompanying decrease in bias again reduces the amplitude of the grid-to-cathode voltage. Thus, in both half cycles the peak value of the a-c component of plate current is decreased and the output voltage is correspondingly reduced below the value it would have had if degeneration were not present.

VOLTAGE GAIN

Voltage gain in the cathode follower is less than unity. The theory of operation of a representative cathode follower is analyzed in the example in figure 6-19 and the voltage gain calculated. The same i_p-e_p characteristic curves for the triode amplifier in the example in figure 5-2 (see preceding chapter) are used to establish the plate current, plate voltage, and grid bias relationships for this example.

The triode is biased for class A operation. The no-signal plate current is $i_p = 0.008$ ampere; the no-signal plate voltage is $e_p = 200$ volts; and the grid bias e_c is -4 volts (point a, fig. 6-19,C). Since the bias of -4 volts is developed across the cathode resistor R_k and there is no plate load resistor, the plate supply voltage is $200 + 4 = 204$ volts. The resistance of the cathode resistor is $R_k = \frac{e_k}{i_p} = \frac{4}{0.008} = 500$ ohms.

The input signal e_s (fig. 6-18,A) has a sine waveform and a peak value of 6.12 volts. At time t_1 the plate current increases to the peak positive value. The value is calculated by applying equation 5-4 from the preceding chapter and modifying it for the example in figure 6-19.

Restated, the equation becomes

$$E_B = -\mu e_g + i_p r_p + i_p R_k \qquad (6\text{-}23)$$

Where E_B is the B supply voltage
μ is the amplification factor of the triode
i_p is the plate current
r_p is the plate resistance
R_k is the cathode (load) resistance and
e_g is the grid-cathode voltage.

In the cathode follower circuit of figure 6-19 there is no plate load resistor and

$$e_g = e_s - i_p R_k \qquad (6\text{-}24)$$

Where e_s is the input signal voltage (peak). and $i_p R_k$ is the voltage across the cathode load resistor.

Substituting e_g from equation 6-24 in equation 6-23

$$E_B = -\mu(e_s - i_p R_k) + i_p r_p + i_p R_k \qquad (6\text{-}25)$$

Transposing equation (6-25) and solving for plate current

$$i_p = \frac{E_B + \mu e_s}{\mu R_k + R_k + r_p} \qquad (6\text{-}26)$$

Substituting, $E_B = 204$, $\mu = 25$, $R_k = 500$, $r_p = 12500$, and $e_s = 6.12$ in equation 6-26, and solving for i_p

$$i_p = \frac{204 + 25 \times 6.12}{25 \times 500 + 500 + 12{,}500} = 0.014 \text{ ampere}$$

At time t_1 the peak positive plate current is 0.014 ampere, point b, figure 6-19,C.

At the same instant the grid cathode voltage, e_g is calculated by substituting $e_s = 6.12$, $i_p = 0.014$ and $R_k = 500$ in equation 6-24 and solving for e_g as follows:

$$e_g = 6.12 - 0.014 \times 500 = -0.88 \text{ volt}$$

In figure 6-19,C, the i_p-e_p curve corresponding to $e_g = -0.88$ volt is estimated to be approximately half way between the 0-e_g curve and the -2-e_g curve. The dotted portion is shown extending through point b.

At time t_2, figure 6-19,A, e_s returns to zero and the plate current returns to 0.008 ampere. The change in current is $\Delta i_p = 0.014 - 0.008 = 0.006$ ampere. The peak (positive) output voltage, e_o, is

$$e_o = \Delta i_p R_k \qquad (6\text{-}27)$$

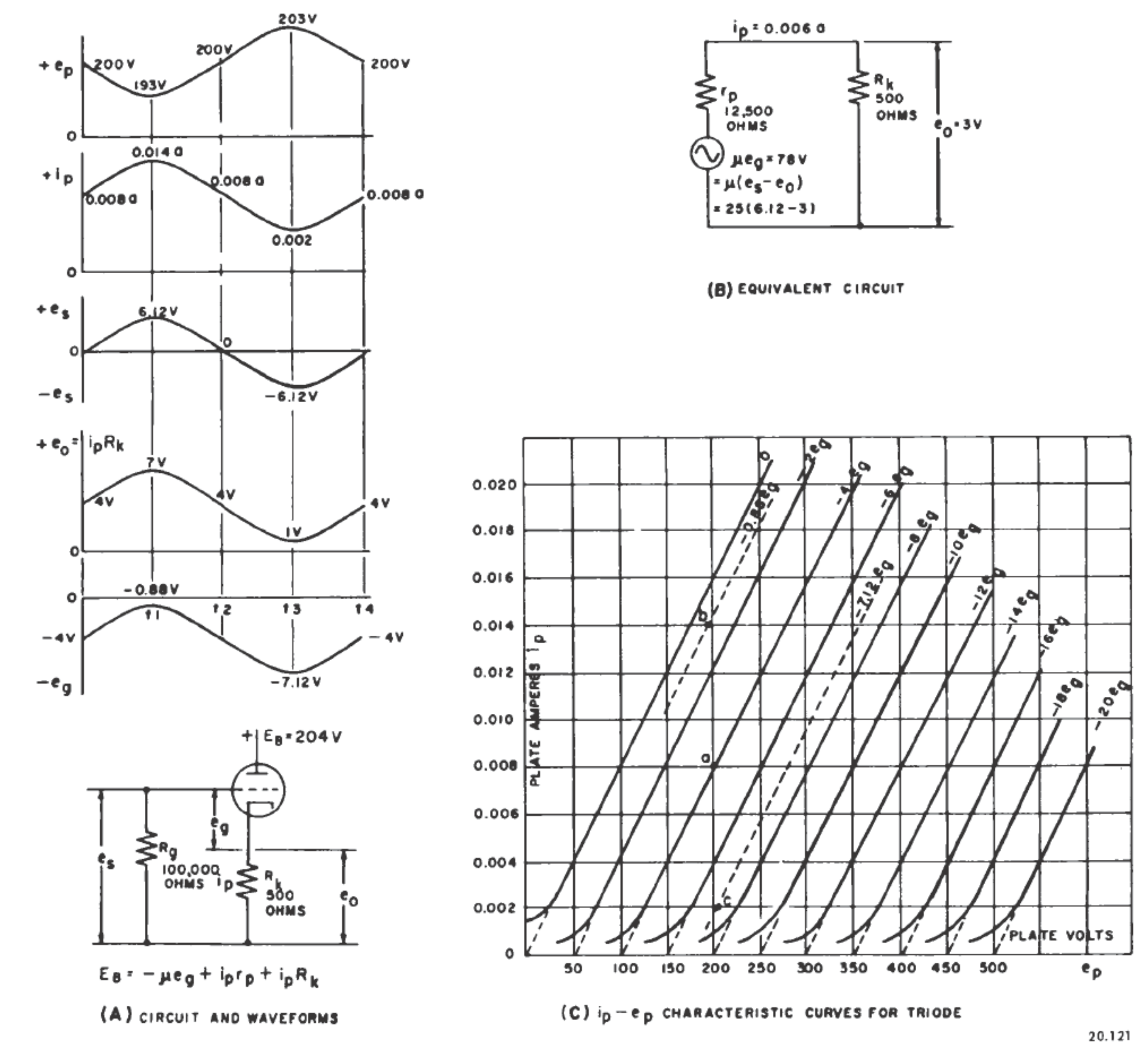


Figure 6-19.—Cathode follower analysis.

Substituting, i_p = 0.006 and R_k = 500 in equation 6-27 and the output voltage at time t_2 is

$$e_o = 0.006 \times 500 = 3 \text{ volts (peak)}$$

At time t_3, e_s increases to the peak (negative value) of -6.12 volts. Substituting, e_s = -6.12 in equation 6-26 and the other values as in the preceding calculation for the peak positive plate current, and solving for plate current,

$$i_p = \frac{204 - 25 \times 6.12}{25 \times 500 + 500 + 12500} = \frac{51}{25500} = 0.002 \text{ ampere}$$

The minimum plate current at time t_3 is 0.002 ampere, point c, figure 6-19,C.

At the same instant (time t_3) the grid-cathode voltage e_g is found by substituting $e_s = -6.12$, i_p 0.002, and $R_k = 500$ in equation 6-24 and solving for e_g as follows:

$$e_g = 6.12 - 0.002 \times 500 = -7.12 \text{ volts}$$

In figure 6-19,C, the i_p-e_p curve corresponding to $e_g = -7.12$ volts is estimated to be approximately half way between the -6-e_g curve and the -8-e_g curve. The dotted portion is shown extending through point c.

At instant t_4, e_s returns to zero and i_p returns to 0.008 ampere to complete the cycle. The change in current is $\Delta i_p = 0.008 - 0.002 = 0.006$ ampere. The peak output voltage (negative swing) is found by substituting $\Delta i_p = 0.006$ and $R_k = 500$ in equation 6-27 as follows:

$$e_o = 0.006 \times 500 = 3 \text{ volts (peak)}$$

The voltage gain of the cathode follower is

$$\text{v.g.} = \frac{e_o}{e_s} = \frac{3}{6.12} = 0.5 \text{ (approximately)}$$

A voltage gain of 0.5 is representative for this cathode follower.

The equivalent circuit, figure 6-19,B, is similar to the equivalent circuit described for a triode amplifier in the preceding chapter (fig. 5-3), with some modifications. All d-c voltages are eliminated; only the a-c components are shown. The circuit contains a voltage, μe_g, acting in series with the plate resistance, r_p and cathode load resistor R_k.

The output voltage, e_o, is equal to the $i_p R_k$ voltage across the cathode load resistor. From the original circuit, figure 6-19,A,

$$e_g = e_s - e_o = e_s - i_p R_k$$

Applying ohms law to the circuit (fig 6-19,B)

$$i_p = \frac{\mu e_g}{r_p + R_k} \qquad (6\text{-}28)$$

Substituting $e_g = e_s - i_p R_k$ in equation (6-28) and solving for i_p

$$i_p = \frac{\mu e_s}{r_p + R_k + \mu R_k} \qquad (6\text{-}29)$$

In this example substitute $\mu = 25$, $e_s = 6.12$ (peak), $r_p = 12{,}500$, and $R_k = 500$ in equation 6-29 and solve for i_p

$$i_p = \frac{25 \times 6.12}{12500 + 500 + 25 \times 500}$$

$$= \frac{153}{25500} = 0.006 \text{ ampere}$$

The output voltage is $i_p R_k = 0.006 \times 500 = 3$ volts and the voltage gain is $\frac{e_o}{e_s} = \frac{3}{6.12} = 0.5$ (approx)

INPUT IMPEDANCE

The input impedance of a cathode follower is high, and the effective input capacitance is low compared with that of a conventional amplifier. Both of these effects result from the degenerative action that occurs across the unbypassed cathode resistance.

Under no-signal conditions the grid is negative with respect to the cathode. When a positive-going signal is applied to the grid the bias is increased, because of degenerative action, to such an extent that no grid current will flow. The result is the same as if the input impedance had been increased. On the other half cycle when a negative-going signal is applied to the grid, the bias is decreased but no grid current can flow and the input impedance remains high.

The reduced input capacitance results from the fact that degeneration reduces the amplitude of the a-c component of the grid-to-cathode voltage, or in effect increases the input impedance, and thus causes less current to flow through the tube capacitances.

Because of the constant high impedance presented by the input to the cathode follower, it presents negligible loading to the circuit that drives it. In the example in figure 6-19, the input impedance is $R_g = 100{,}000$ ohms.

OUTPUT IMPEDANCE

The output impedance of the cathode follower is relatively low. In the example in figure 6-19, the output impedance is approximately 500 ohms. This means that if R_k is removed it will be necessary to substitute a 500 ohm load in place of R_k if we wish to maintain the same values of current and voltage as were used in the example.

The output impedance is also approximately equal to r_p/μ. In this example $r_p = 12,500$ ohms and $\mu = 25$; the output impedance is approximately $\frac{12,500}{25} = 500$ ohms.

If the load to be supplied by the cathode follower is less than 500 ohms (in this example) a series resistor may be inserted to make up the difference. For example, if the load is to be 400 ohms, a 100 ohms resistor inserted in series with the load will make the total of 500 ohms needed. If the load is larger than 500 ohms, for example 1,000 ohms, a 1,000 ohm resistor in shunt with the 1,000 ohm load will make the total parallel resistance equal to the required 500 ohms.

The output impedance is generally resistive and the approximations of the preceding example are valid when μ is much greater than unity. (In the example μ is 25.) The output impedance is low in relation to the input impedance and there is a minimum of amplitude distortion of the output signal even though current is drawn from the output terminals.

DISTORTION CAUSED BY LIMITING

Under normal operating conditions the output of a cathode-follower amplifier is practically free of amplitude distortion. However, if the input signal swings the grid voltage too far negative or positive, the output waveform will be limited or distorted in amplitude with respect to the input waveform. Beyond a certain negative value of grid voltage the plate current will be cut off and any further increase in negative grid potential will cause no corresponding change in plate current.

If the signal swings the grid voltage in a positive direction far enough for the grid to draw current, the loss in voltage in the driving source limits the output signal and distortion again occurs.

The cathode-follower amplifier may be modified, as in figure 6-20, to adjust the grid bias to the correct value if the cathode resistance is greater than the value required to give the correct grid bias and if limiting occurs only on the negative peaks of the input signal. In this modified circuit, the grid resistor, R_g, is connected to a point above ground on the cathode resistors R_{k1} and R_{k2}. This point is determined by the input voltage level. Thus the grid bias is reduced by an amount equal to the drop across R_{k2}.

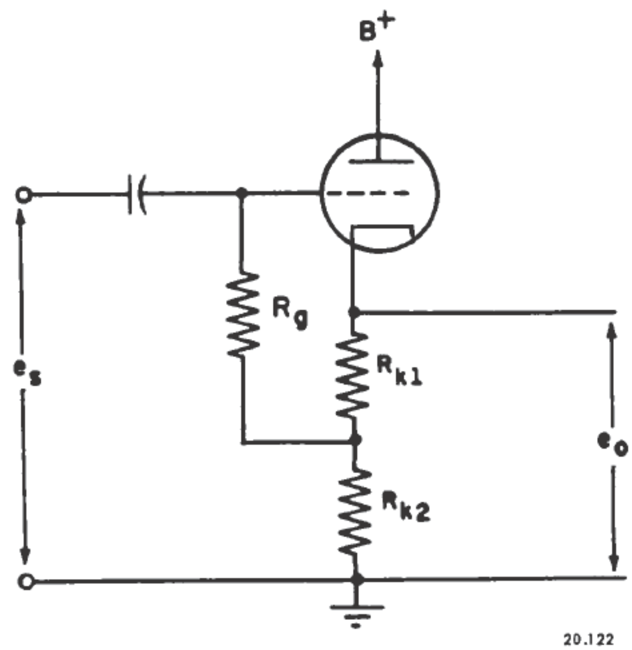


Figure 6-20.—Cathode follower modified to prevent limiting.

ADVANTAGES OF CATHODE FOLLOWERS

As previously stated one of the principal advantages of a cathode follower is that it can be used to match a high impedance to a low impedance. Thus it can take the voltage developed across a high impedance and supply a low impedance load with only a slightly less voltage but with a correspondingly large increase in current. One or more of the circuit elements of a cathode follower may be varied to achieve a more precise impedance match if the match is critical.

When tubes having a high mutual conductance are used, the low value of output impedance extends the amplification into the upper range of frequencies because the shunting effects of interelectrode and distributed capacitances are proportionately smaller. The low-frequency response is improved by allowing the d-c component of cathode current to flow in the load, thus avoiding the use of a series blocking capacitor.

The degenerative effect caused by the unbypassed cathode resistor increases the input impedance. Thus less shunting effect is offered to the previous stage, and a better overall frequency response is produced.

As stated before, the input and output voltages have the same instantaneous polarity. When pulses are used it may be necessary to feed a positive- or a negative-going pulse to a load without polarity inversion. The cathode follower could thus serve two purposes—to prevent polarity inversion and to afford an impedance match.

Circuit stability is also improved, as in regular amplifiers, by degenerative feedback. Specifically, amplitude distortion occuring within the tube, the effect of plate-supply voltage variations, aging of tubes, production of harmonics, and other undesirable effects that occur within the stage are counteracted by this type of circuit.

However, these advantages are achieved at the expense of an overall reduction in voltage gain. Normally, the voltage gain is slightly less than unity, but the circuit is capable of producing a gain in power.

PHASE INVERTERS

Since phase is generally associated with time, it is somewhat of a misnomer to apply this term to a device that simply changes a positive-going signal to a negative-going signal or vice versa. In the case of a sine-wave signal, however, the effect is the same as if a 180° phase shift had occured.

Paraphase amplifiers (phase splitters) produce, from a single input waveform, two output waveforms that have exactly opposite instantaneous polarities. If these two waveforms were produced as the result of a single sine-wave input they might be considered 180° out of phase, one waveform having been displaced 180° along the time axis.

One type of phase inverter is the transformer, with which the instantaneous polarity of the load may be reversed with respect to the source by reversing either the connections of the secondary leads to the load or the primary leads to the source. A conventional electron-tube amplifier (untuned and R-C coupled) also produces an output of opposite polarity to the input; and if no gain is desired, various methods may be employed to produce unity gain. Either single- or two-tube amplifiers may be used to convert one input waveform into two output waveforms of opposite polarity. Such amplifiers are called PHASE SPLITTERS or PARAPHASE amplifiers.

TRANSFORMER PHASE INVERTER

In operation, all transformers produce across the secondary an induced emf that is opposed to the change in flux producing it. The instantaneous polarity of the actual output voltage across a load depends on how the leads from the secondary are connected.

Figure 6-21 indicates phase inversion of square waves and sine waves. With square waves the polarity has simply been inverted. This is also true for sine waves but in this case it may be more convenient to refer to the inversion as a 180° phase shift—in effect, the same result as if the waveform had been moved along the time axis 180°. If no change in voltage is desired, a 1-to-1 turns ratio is employed.

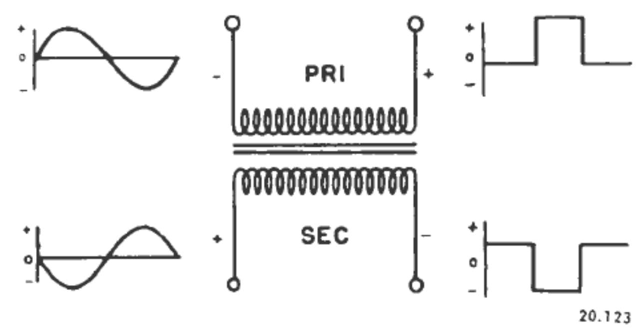


Figure 6-21.—Transformer phase inversion.

A transformer with a center-tapped secondary or with a center-tapped resistor shunting the secondary, is used in Class-B push-pull circuits to supply voltages of opposite instantaneous polarity to the grids of the tubes, as shown in figure 6-22.

If at a given instant the polarity of point X goes negative with respect to the grounded center tap, the polarity of point Y will go positive with respect to the center tap. Thus a negative potential is applied between the grid and ground of V2 and at the same time a positive potential is applied between the grid and ground of V3. This condition is necessary for the proper operation of a push-pull amplifier. Of course, the transformer must be tapped at the electrical center; otherwise the combined signal present in the output transformer will not be symmetrical with respect to the tap.

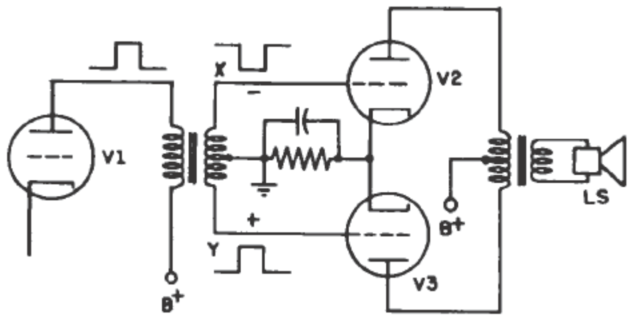


Figure 6-22.—Center-tapped transformer driving a push-pull amplifier.

This type of transformer phase inverter has limited application because of distortions and losses inherent in transformers. For example, the loss in voltage through leakage reactance is greater for higher frequencies than it is for lower frequencies. The shunting capacitance effect and hysteresis losses also increase with frequency. Since in many circuits harmonics must be transmitted unattenuated and undistorted, the transformer phase inverter is generally replaced with a circuit that performs phase inversion without the use of transformers. The paraphase amplifier is such a circuit.

ELECTRON-TUBE PHASE INVERTER

As mentioned before, every electron tube used as a conventional amplifier introduces polarity inversion—that is, a negative-going signal between grid and ground causes a positive-going signal to be produced across the plate load. If there is to be polarity inversion with no gain in amplitude, some method must be employed to reduce the normal gain to unity. One method of reducing the normal gain is through the use of degenerative feedback. Degenerative feedback is readily obtained by omitting the cathode bypass capacitor.

Another method of reducing the gain is to employ a voltage divider in the input circuit. For example, if the normal gain of the tube is 100 the grid is tapped down on the divider so that one-hundredth of the available voltage is applied between grid and ground. If harmonics are to be included some method must be employed to reduce the input shunting effects of capacitance.

SINGLE-TUBE PARAPHASE AMPLIFIER

One of the simplest forms of single-tube paraphase amplifiers is shown in figure 6-23. The values of resistors R2 and R3 are the same. Therefore the voltage drop across both of them is the same, since the same plate current flows through both. The instantaneous polarities, however, are exactly opposite because at the instant a positive-going signal is applied to the grid, point X becomes less positive with respect to ground and point Y becomes more positive. These signals, with the polarities indicated in the figure, are impressed across load resistors R4 and R5 through blocking capacitors C3 and C4. C2 is the plate supply bypass capacitor.

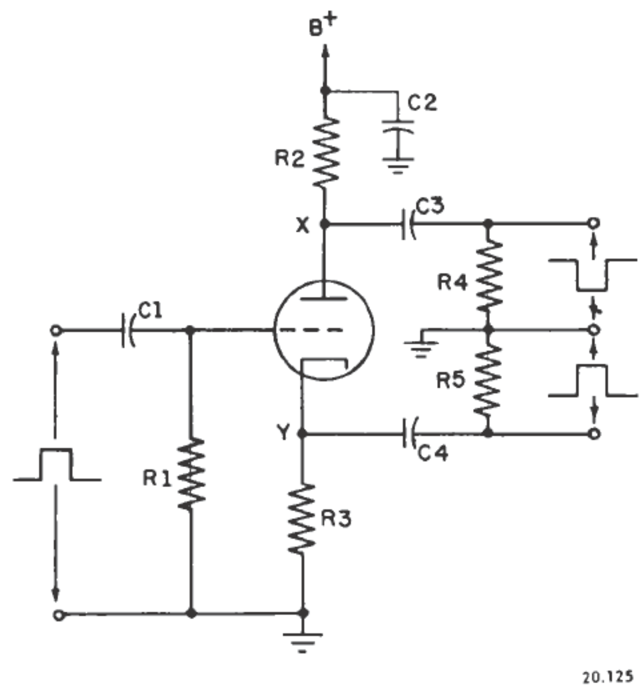


Figure 6-23.—Single-tuned paraphase amplifier.

In actual practice this basic type of single-tube paraphase amplifier may be modified to avoid some of the degenerative action due to the unbypassed cathode resistor or it may be compensated to permit a better frequency response.

TWO-TUBE PARAPHASE AMPLIFIER

A 2-tube paraphase amplifier utilizes ① tube as a regular amplifier and a second tube as a phase inverter, or these functions may be performed by 2 sections of the same tube. The combination is frequently referred to as a PHASE INVERTER.

One of the simpler forms of 2-tube paraphase amplifiers is shown in figure 6-24. V1 operates as a conventional amplifier having normal gain, and V2 operates as a phase inverter, the input of which is reduced to the same value as the input of V1. Thus V2 amplifies the signal as much as V1 and the output is essentially symmetrical about the zero-voltage reference line.

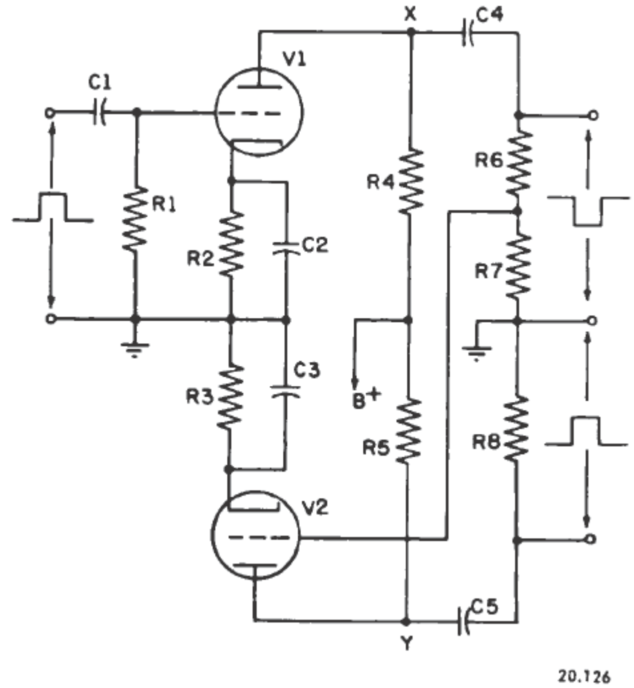


Figure 6-24.—Two-tube paraphase amplifier.

A positive-going signal on the grid of V1 causes an increase in plate current and a reduction in positive plate potential at point X. This reduction in positive potential is transmitted as a negative-going signal through coupling capacitor C4 to resistors R6 and R7. The grid input to V2 is tapped down on resistors R6 and R7 to feed the proper magnitude of negative-going signal to V2. For example, if V1 and its associated circuit has a voltage gain of 50, the resistance of R7 should be one-fiftieth of the total value of R6 plus R7. At the instant a positive-going signal is applied to the grid of V1 a negative-going signal is thus applied to the grid of V2. The positive potential at point Y is increased, and a positive-going signal is applied to resistor R8, through coupling capacitor C5. At the same time the negative-going signal appears across resistors R6 and R7.

If the operating conditions of the two tubes are carefully chosen and the circuits are properly adjusted, the two amplified output signals should be essentially undistorted and of opposite instantaneous polarity. In actual practice this method presents some difficulty because the adjustments are critical. However, it is widely used as a means of driving class-A push-pull audio power amplifiers.

CHAPTER 7

AUDIO POWER AMPLIFIERS

Audio amplifiers are classified into two groups: voltage amplifiers, and power amplifiers.

GENERAL

The primary function of the common VOLTAGE amplifier is to increase the voltage of a signal to a higher value without distorting the waveform. Under ideal conditions no appreciable power is consumed from the preceding stage, and no appreciable power is supplied to the succeeding stage. In general, the output voltage is proportional to the product of the input voltage and the mu of the tube.

The primary function of a POWER amplifier is to deliver power to a load, and any increase in voltage is of secondary importance. Because the power amplifier requires a relatively high signal-voltage input, the power amplifier must usually be preceded by one or more voltage amplifiers to raise the voltage to the proper level to operate the power stage.

Various types of tubes may be used as audio power amplifiers, including triodes (or tubes operated as triodes) and pentodes. The tubes may be operated singly as class-A amplifiers, or in pairs as in push-pull stages in which the tubes are operated as class-A, class-AB, or class-B amplifiers.

In general, audio power amplifiers have low amplification factors, low plate resistance, and high plate current. In order to obtain low plate resistance, the space between the plate and cathode is made smaller in a power tube than it is in a voltage amplifier tube. Also, in a power tube the area of the plate is made larger and the cathode is designed to supply a larger number of electrons. The grid must not block too many of the electrons flowing to the plate; accordingly, the grid meshes are widely separated, and the amplification factor is thereby reduced.

Pentodes used as power amplifiers have higher amplification factors than triodes, but the plate resistance is proportionately higher. As will be seen later, increasing the amplification factor of the tube used will increase the power output of an amplifier stage, while increasing the plate resistance will cause a decrease in output power.

Power amplifiers have numerous applications; the most familiar perhaps is the output stage of a radio receiver. Power is needed to operate the loudspeaker; therefore, the last audio stage is operated as a power amplifier.

CLASS-A TRIODE AMPLIFIERS

Class-A amplifiers are operated so that plate current flows during the entire input-voltage cycle. If the correct operating point, load impedance, and input voltage are chosen, the output waveform will be essentially the same as the input in all respects except for the amplitude.

Many limitations are imposed on class-A amplifiers if they are to be operated with minimum allowable distortion. The curvature of the lower portion of the i_p-e_g characteristic curve places a practical limit on the minimum current that may flow in the plate circuit. This limitation, in turn, places a limit on the negative swing of grid-signal voltage. If class-A_1 operation is assumed, the positive swing of the grid-signal voltage is limited by the magnitude of the bias in order that grid current will not flow. This limitation, however, is not serious.

The efficiency of class-A amplifiers is limited to a low value (15 to 25 percent for class-A_1 operation) because appreciable d-c plate current flows during the entire grid-voltage cycle. Although its efficiency is low, the class-A amplifier has remarkably high fidelity if the proper operating conditions are chosen.

LOAD LINE

One of the simplest methods of determining the output voltage and current components under a variety of operating conditions is by the use of a load line, as shown in figure 7-1, A. This line is a graph of the equation

$$e_p = E_B - i_p R_L$$

where e_p is the instantaneous plate-to-cathode potential, E_B the plate supply voltage, and i_pR_L the voltage drop across load resistor R_L.

The I_p-E_p curves across which the load line is plotted are known as STATIC CHARACTERISTIC CURVES because grid voltage changes and the accompanying plate current changes occur at constant plate potential. For example, if the grid bias is decreased from -35 to -30 volts at a constant plate potential of 250 volts the plate current will increase from 30 ma to 40 ma, or an increase of 10 ma.

In contrast to this static action, DYNAMIC CHARACTERISTIC I_p-E_p CURVES take into account the change in plate voltage that always occurs with a change in plate current when a load resistor is connected in series with the plate. The load line makes possible the calculation of the dynamic characteristic. Thus, if the grid bias is decreased from -35 volts (point B on the load line) to -30 volts, the plate current will increase from 30 ma to 33 ma, or an increase of 3 ma.

Hence, in the examples given, the plate current increases 10 ma using the static curves, and only 3 ma using the dynamic characteristic because in the first case the plate voltage remains constant at 250 volts, whereas in the second case the plate voltage decreases from 250 volts to 235 volts because of the increased voltage drop in load resistor R_L.

Load resistor R_L seldom exists as a real resistor but is used to represent the equivalent plate-load resistance in order to simplify the actual load circuit. For example, the actual load circuit might consist of a step-down transformer and its associated loudspeaker, in which case the actual B-supply voltage would be only large enough to supply the plate voltage plus the d-c voltage drop in the transformer primary. However, the load line is based upon the theoretical B-supply voltage that would be required if R_L were actually a load resistor.

To account for the difference in the a-c and d-c resistance components of the load when such differences exist, a second load line called the a-c load line may be drawn through the operating point B. This line will have a slope of $-\frac{1}{R_L}$ where R_L now represents the a-c impedance of the load.

If a branch composed of C and R_g is added, as shown in figure 7-1, D, and the d-c load remains at 5000 ohms, the effective a-c impedance of the total load will be less than 5000 ohms. If the reactance of C is negligible, the a-c load impedance will be $\frac{5000}{2}$ or 2500 ohms. The a-c load line will then have a slope of $-\frac{1}{2500}$ and will pass through point B as indicated by the line having the steeper slope.

The slope of the d-c load line is $-\frac{1}{5000}$ or $\frac{.080}{400}$. The slope of the a-c load line is $-\frac{1}{2500}$ or $\frac{.080}{200}$, and is represented as a dotted line between points Y and X' to establish the slope. The a-c load line is then drawn through operating point B as a line parallel to the line YX'.

The terminal points Y and X on the Y axis and the X axis, respectively, are easily established. Thus, if the electron tube resistance could be reduced to zero, the plate current would become a maximum value and the B-supply voltage would appear across the load resistance. The plate current would be

$$i_p = \frac{E_B}{R_L}$$

This value of i_p determines the Y-axis terminal of the load line, indicated in figure 7-1, A, as point Y.

When no plate current flows, the full value of the plate supply voltage is applied between the plate and cathode because there is no voltage drop across R_L. Thus,

$$e_p = E_B$$

This value of e_p determines the X-axis terminal of the load line and is indicated as point X.

When a number of i_p-e_p curves for various values of grid bias are included on the graph, the plate current and voltage corresponding to a

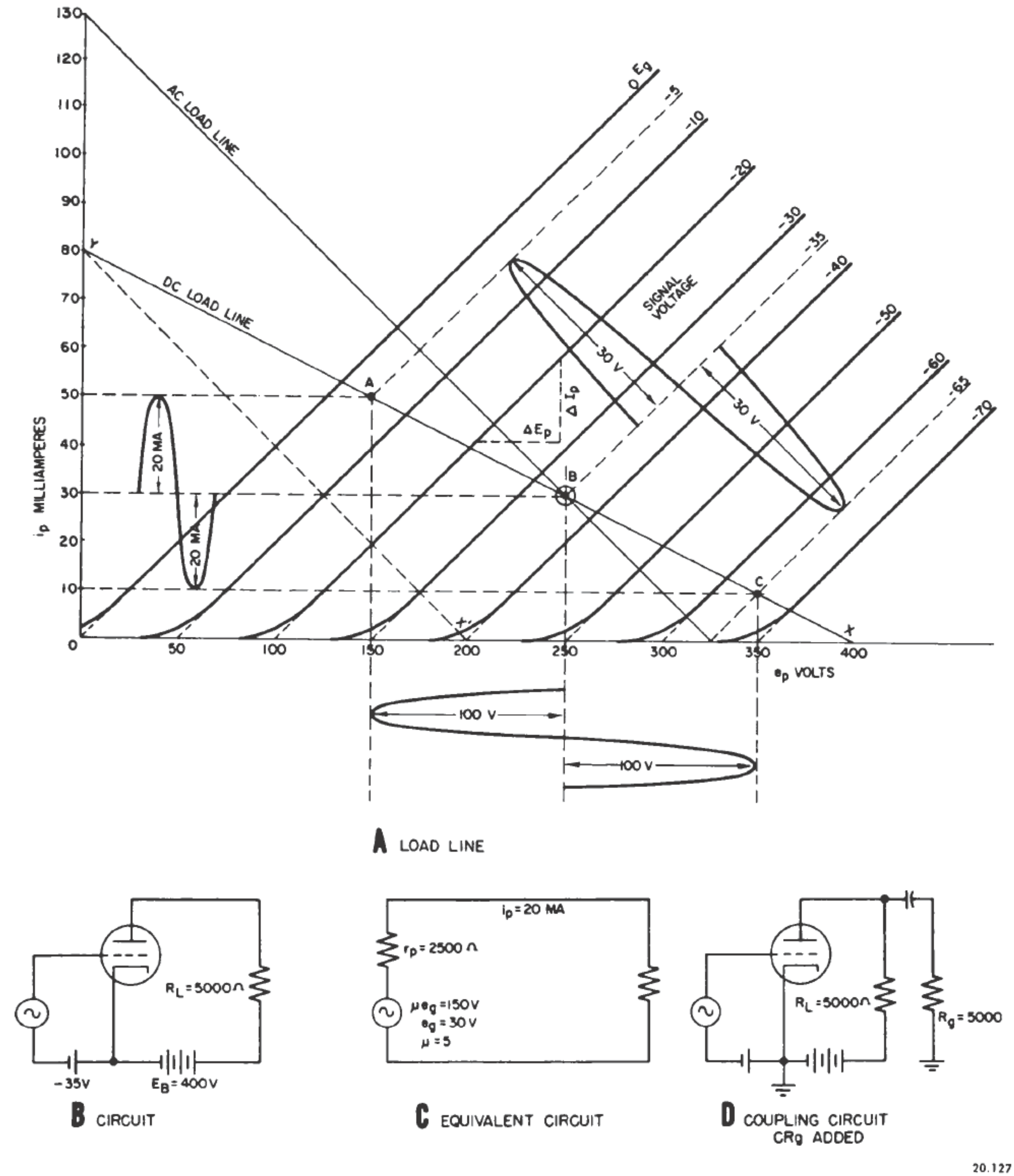


Figure 7-1.—Load line and circuits for a triode amplifier.

given bias can be determined at a glance. Thus in figure 7-1, A, point A represents a grid bias of -5 volts. A line drawn horizontally from this point to the Y axis indicates a plate current of approximately 50 ma. Likewise, a line drawn vertically downward to the X axis from point A indicates a corresponding plate voltage of approximately 150 volts.

A line drawn horizontally from point B (the operating point where the i_p-e_p curve for a grid bias of -35 volts intersects the load line) to the Y axis indicates a plate current of 30 ma. The corresponding plate voltage is approximately 250 volts. The same procedure is followed in determining the plate current and voltage for point C or any other given value of grid bias.

Point B is called the OPERATING POINT because it represents the grid bias when no signal is applied. It is therefore the point about which the grid voltage will vary when a signal is applied. The instantaneous values of plate current and voltage may be considered as being determined by the position of an imaginary point oscillating about B along the load line. The limits of the oscillations are determined by the point at which the grid begins to draw current on the positive swing and by the point at which the load line extends into the nonlinear portion of the plate-current curves on the negative swing. In establishing these limits it is of course assumed that the source is capable of supplying the input voltage swing between these limits.

If, as in figure 7-1, A, the grid voltage is assumed to have a maximum positive swing of 30 volts, then at the positive peak of the swing the net voltage impressed on the grid is 30-35 = -5 volts. The plate current at this instant reaches its maximum value of 50 ma, and the plate voltage reaches its minimum value of 150 volts. At the negative peak of the input signal the total voltage between grid and cathode is -30-35 = -65 volts. The plate current at this instant reaches its minimum value of 10 ma and the plate voltage reaches its maximum value of 350 volts.

From the maximum to minimum values the peak a-c plate voltage is

$$e_p = \frac{E_{max}-E_{min}}{2} = \frac{350-150}{2} = 100 \text{ v}$$

The peak a-c plate current is

$$i_p = \frac{I_{max}-I_{min}}{2} = \frac{50-10}{2} = 20 \text{ ma}$$

By the use of the I_p-E_p curves of figure 7-1, A, the static characteristics of the triode may be determined. The plate resistance, for example, is the ratio of the change in plate voltage to the corresponding change in plate current for a constant bias--that is

$$r_p = \frac{\Delta E_p}{\Delta I_p} = \frac{250-200}{60-40} = \frac{50}{20} = 2.5 \text{ k-ohms}$$

The mu of the triode is the ratio of the change in plate voltage necessary to produce a certain change in plate current, to the change in grid voltage necessary to produce the same change in plate current. Thus,

$$\mu = \frac{\Delta E_p}{\Delta E_g} = \frac{250-200}{30-20} = \frac{50}{10} = 5$$

The mutual conductance, g_m, of the triode is the ratio of the change in plate current to the change in grid voltage producing it. If the change in plate current is expressed in microamperes and the change in grid signal is expressed in volts, the transconductance will be expressed in micromhos. Thus,

$$g_m = \frac{\Delta I_p}{\Delta E_g} = \frac{(60-40)10^3}{30-20} = \frac{20 \times 10^3}{10} =$$

2,000 micromhos

In terms of μ and r_p the mutual conductance may be expressed as

$$g_m = \frac{\mu}{r_p} = \frac{5.0}{2,500} = 0.002 \text{ mhos, or } 2,000$$

micromhos

The voltage gain of an amplifier stage is the ratio of the signal-voltage output to the signal-voltage input—

$$\text{voltage gain} = \frac{e_o}{e_g}$$

Under the conditions established in figure 7-1, A, the voltage gain is

$$V.G. = \frac{e_o}{e_g} = \frac{100}{30} = 3.33$$

From the equivalent circuit of figure 7-1, C, μ is 5, r_p is 2.5 k-ohms, and R_L is 5 k-ohms. Expressed in terms of these values, the voltage gain is

$$V.G. = \frac{\mu R_L}{r_p + R_L} = \frac{5x5}{2.5+5} = 3.33$$

POWER OUTPUT

The power output of the amplifier of figure 7-1, B, may be calculated from the I_p-E_p curves of figure 7-1, A. The plate voltage swings between 150 v and 350 v as the plate current swings from 50 ma to 10 ma. The peak a-c plate current is $\frac{50-10}{2} = 20$ ma. The maximum a-c signal voltage is $\frac{350-150}{2} = 100$ volts. The power output of the triode amplifier is

$$P_o = \frac{E_{max} I_{max}}{2} = \frac{100x0.02}{2} = 1 \text{ watt}$$

The power output may also be calculated from the equivalent circuit of figure 7-1, C, as the power delivered to the load resistor, R_L.

The voltage output of an amplifier stage is proportional to μe_g and the power output is proportional to $(\mu e_g)^2$.

Because μe_g acts in series with r_p and R_L in the equivalent circuit, the a-c component of plate current is

$$i_p = \frac{\mu e_g}{r_p + R_L}$$

The output voltage appears across R_L as $i_p R_L$. The symbols are identified as follows:

e_o = instantaneous values of output voltage
E_o = rms values of output voltage
i_p = instantaneous values of plate current
I_p = rms values of plate current

Thus

$$e_o = i_p R_L = \frac{\mu e_g R_L}{r_p + R_L}$$

The output power in watts, when I_p and $E_o\theta$ are effective (rms) values, becomes

$$P_o = I_p E_o = \frac{(\mu E_g)^2 R_L}{(r_p + R_L)^2} \qquad (7\text{-}1)$$

In the example in figure 7-1, C, R_L is 5 k-ohms, r_p is 2.5 k-ohms, E_g is 21.2 volts (rms), and μ is 5. The power output is

$$P_o = \frac{(5x21.2)^2 x5x10^3}{(2.5x10^3+5x10^3)^2} = 1 \text{ watt}$$

Because maximum transfer of energy occurs when $R_L = r_p$, the power output for this condition becomes

$$P_o = \frac{(\mu E_g)^2 r_p}{(2r_p)^2} = \frac{(\mu E_g)^2}{4r_p} \qquad (7\text{-}2)$$

In the example in figure 7-1, C, if R_L is changed to 2.5 k-ohms, the power output becomes

$$P_o = \frac{(5x21.2)^2}{4x2.5x10^3} = 1.12 \text{ watts}$$

The problem of distortion is present in the power amplifier the same as it is in the voltage amplifier and here again a balance must be struck between maximum power output and minimum distortion. Because the human ear is not particularly sensitive to distortion below about 5 percent, this amount may be allowed in the output circuit. When the term "undistorted" is used in the following considerations, distortion up to 5 percent is allowed. Experiments show that when $R_L = 2r_p$ the most noticeable distortion (that due to the second harmonic) is reduced to less than 5 percent. The reduction in power output, when $R_L = 2r_p$ as compared with the power output when $R_L = r_p$, is only about 11 percent, as shown by the example in figure 7-1, C.

Increasing the load resistance in the plate circuit of an electron-tube amplifier tends to reduce the slope of the i_p-e_g characteristic curve, as shown in figure 7-2. The curves are flattened because the higher the load resistance the lower is the voltage that is available at the plate, and consequently the lower becomes the plate current.

In general, the curve will have an appreciable bend if the load impedance (unity power factor) is equal to the plate resistance of the power-amplifier tube. Unfortunately this value of load impedance would permit maximum transfer of power.

As mentioned previously, it has been found experimentally that MAXIMUM UNDISTORTED POWER OUTPUT may be achieved when the load impedance is approximately twice the plate resistance of the tube and the plate current variations are at the maximum permissible value for class-A operation.

The equation for maximum undistorted power output when the load impedance, $R_L = 2r_p$, is established as follows: In equation (7-1), if R_L is replaced by $2r_p$ and E_g is expressed in effective (rms) volts,

$$P_o = \frac{(\mu E_g)^2\, 2r_p}{(r_p + 2r_p)^2} = \frac{(\mu E_g)^2}{4.5 r_p}$$

A comparison of equations (7-2) and (7-3) verifies the reduction in power output as a result of making the load impedance twice the plate resistance to be only about 11 percent. This reduction is in agreement with the example in figure 7-1, C.

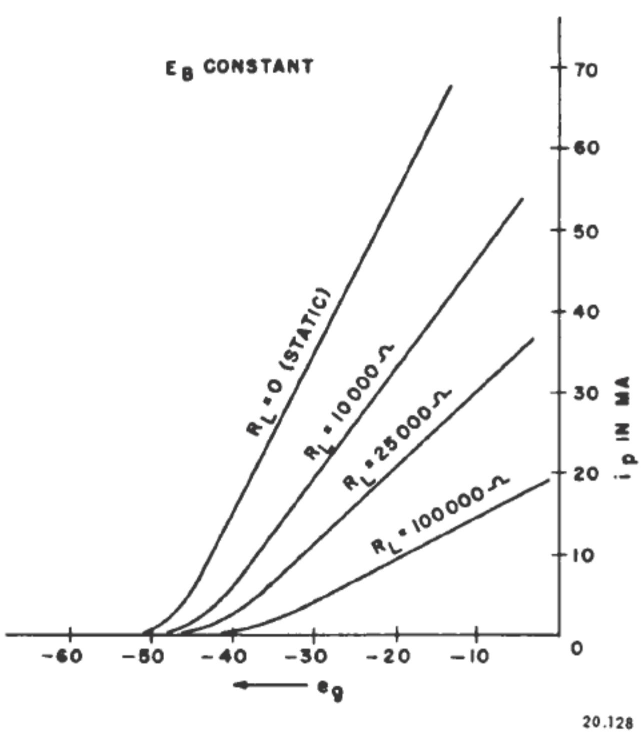


Figure 7-2.—i_p-e_g curves for a given plate voltage under various load conditions.

SECOND-HARMONIC DISTORTION

A careful examination of figures 7-2 and 7-3 will reveal some of the problems involved in designing triode power amplifiers and also how the concept of second-harmonic addition may be used to describe distortion.

If the bias is adjusted so that only the straight portion of the i_p-e_g characteristic curve is used, the signal-voltage swing will be greatly reduced. Because the power output varies as the square of the input-signal voltage, there is a practical limit to the amount of restriction that can be imposed. The load impedance may be further increased, but here again the current variation in the plate circuit will be reduced accordingly, and so will the power. If the bias is reduced, grid clipping may occur and second harmonics, as well as other harmonics, may be introduced. Harmonics are also introduced if the bias is so high that the negative half of the plate-current swing is reduced. If the signal voltage is too high, both grid clipping and operation beyond the bend in the lower portion of the i_p-e_g characteristic curve may occur.

When certain types of distortion are referred to as second- or third-harmonic distortion or perhaps as even- or odd-harmonic distortion, these harmonics are not necessarily actually produced in the tube. Rather the EFFECT on the output signal is the same as if these harmonics had been introduced. This concept is useful in explaining many types of distortion caused by electron tubes.

A consideration of the distortion of the plate-current curve in figure 7-3 reveals the presence of a second-harmonic component (deliberately exaggerated in the figure). In this instance the positive half of the (solid-line) plate-current

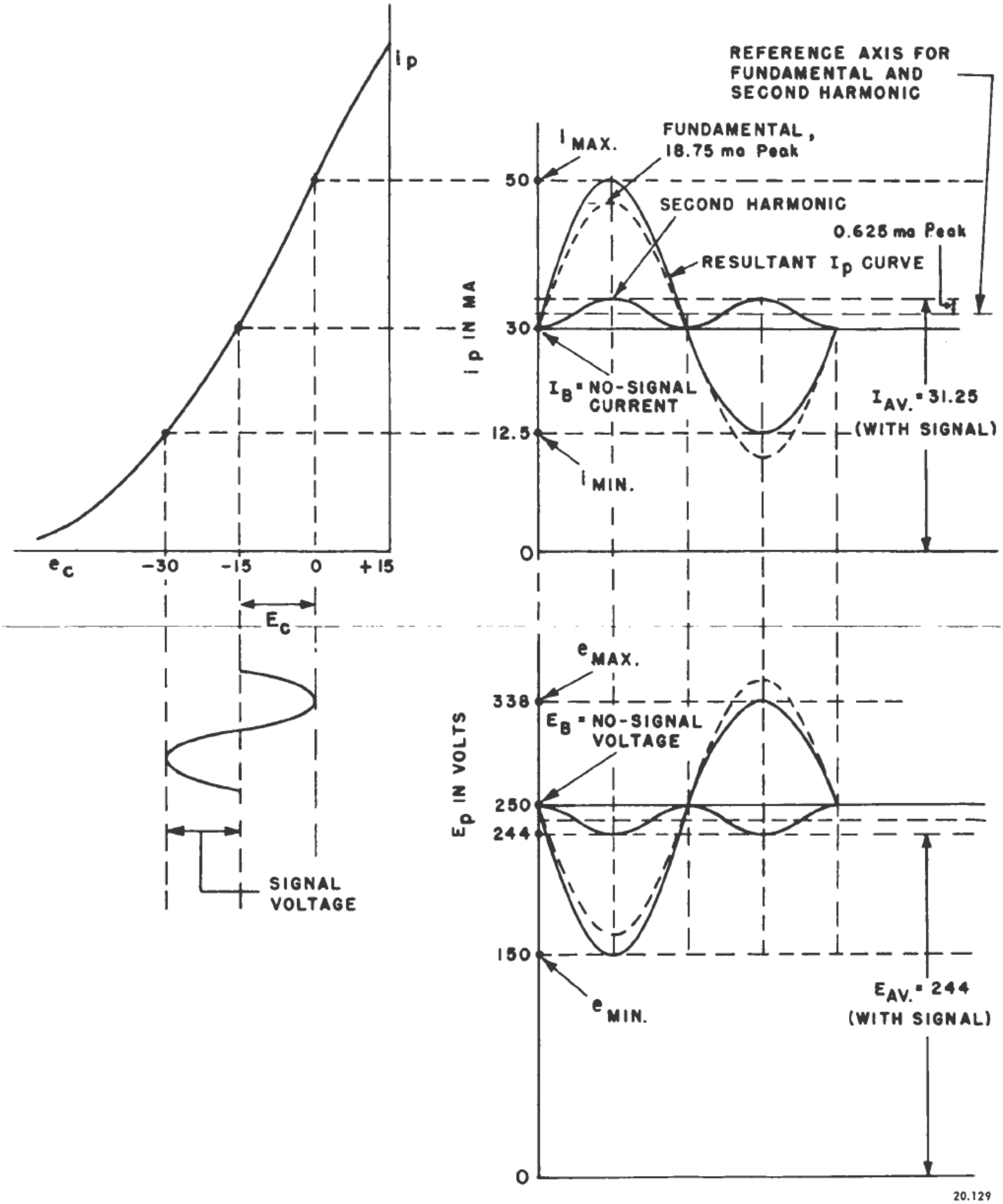


Figure 7-3.—Second-harmonic distortion in a class-A power amplifier.

curve is increased beyond the boundaries of the accurate sine curve (broken-line curve) and the negative half is decreased.

The lack of symmetry in the (solid-line) plate-current curve may be adequately explained if a second-harmonic signal is assumed to exist, as shown above the no-signal current line.

It is helpful to make a brief mathematical analysis of the contributions made by the second-harmonic component, and by the fundamental, to the resultant plate current.

The AVERAGE PLATE CURRENT, I_{av}, may be established by inspection or verified by formula

$$I_{av} = \frac{I_{max}+I_{min}}{2} = \frac{50+12.5}{2} = 31.25 \text{ ma}$$

The MAXIMUM PLATE-CURRENT SWING of the second-harmonic component, I_2, is

$$I_{av}-I_B = 31.25-30 = 1.25 \text{ ma}$$

where I_B is equal to the no-signal value of plate current. The maximum value of the second-harmonic current is

$$I_{2max} = \frac{1.25}{2} = 0.625$$

Also, the maximum plate-current swing of the fundamental is

$$I_1 = I_{max}-I_{min} = 50-12.5 = 37.5$$

and the maximum value of the fundamental is therefore $\frac{37.5}{2} = 18.75$ ma.

The equation for the AVERAGE VALUE of the second-harmonic component of plate current is

$$I_2 = \frac{I_{max}+I_{min}-2I_B}{4}$$

With the aid of the voltage curve, the contribution of the two components to the plate-voltage variations may be determined in a similar manner.

The average power output, P_o, is

$$P_o = \frac{(I_{max}-I_{min})(E_{max}-E_{min})}{8}$$

$$= \frac{(0.05-0.0125)(338-150)}{8} = 0.881 \text{ watts}$$

The percentage of distortion (P.D.) due to the second-harmonic current component is determined by dividing the second-harmonic component by the fundamental component and multiplying the result by 100. Thus,

$$\text{P.D.} = \frac{I_2}{I_1} \times 100 = \frac{0.625}{18.75} \times 100 = 3.3 \text{ percent}$$

The general equation for determining the percentage of distortion due to the second harmonic is

$$\text{P.D.} = \frac{I_{max}+I_{min}-2I_B}{2(I_{max}-I_{min})} \times 100$$

OUTPUT TRANSFORMER

The output transformer used with audio power amplifiers serves as an IMPEDANCE-MATCHING device. Because the plate resistance of a power-amplifier tube may range from perhaps 1,000 ohms to more than 20,000 ohms; and since the impedance of the loudspeaker voice coil may range down to 4 ohms, the output transformer has a step-down turns ratio to provide the correct ratio of primary voltage and current to secondary voltage and current.

Impedance matching by means of a transformer is illustrated in figure 7-4. It is recalled that the output voltage of a transformer varies directly as the turns ratio, and that the output current varies inversely as the turns ratio—that is,

$$\frac{E_p}{E_s} = \frac{N_p}{N_s}$$

and

$$\frac{I_s}{I_p} = \frac{N_p}{N_s}$$

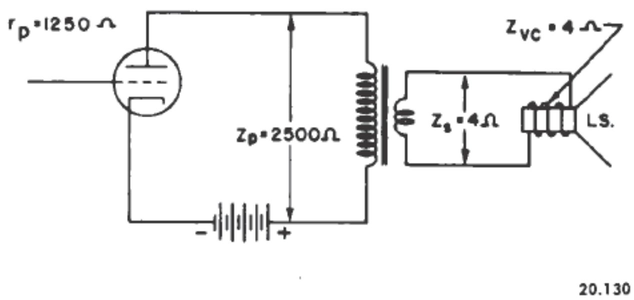


Figure 7-4.—Output transformer used as an impedance-matching device.

If the 2 left sides and the 2 right sides of these preceding two equations are multiplied together, the impedance may be determined. Thus,

$$\frac{E_p}{I_p} \times \frac{I_s}{E_s} = \frac{N_p^{\,2}}{N_s^{\,2}}$$

The primary impedance of a matching transformer is defined as the ratio of rated primary voltage to rated primary current. Similarly, the secondary impedance is the ratio of rated secondary volts to rated secondary amperes. If Z_p is substituted for $\frac{E_p}{I_p}$ and $\frac{1}{Z_s}$ is substituted for $\frac{I_s}{E_s}$, the impedance equation becomes

$$\frac{Z_p}{Z_s} = \left(\frac{N_p}{N_s}\right)^2$$

Thus, the ratio of the two impedances that a transformer can match is equal to the turns ratio squared.

As a practical illustration, find the turns ratio needed for the transformer shown in figure 7-4. Recall the need for $R_L = 2r_p$. Because the plate resistance is 1,250 ohms, the primary impedance is given as twice this value, to permit maximum undistorted power output. The power fed to the 4-ohm voice coil, however, will be reduced unless the proper impedance is afforded by the transformer. The turns ratio that will satisfy this condition is readily determined by extracting the square root of both sides of the last equation. Thus,

$$\frac{N_p}{N_s} = \sqrt{\frac{2,500}{4}} = 25$$

The amount of power that can be handled by an output transformer is determined by the current and voltage RATINGS of the windings and the allowable losses. The primary frequently contains a d-c component that limits the incremental inductance and frequency response. The equation for the induced voltage, E (rms), of a transformer winding is

$$E = 4.44\ fNBA\ 10^{-8}$$

where f is the frequency in cycles per second, N the number of turns in the winding, B the flux density of the core, and A the cross-sectional area of the core. In a given transformer the induced voltage is proportional to the product of the frequency and the flux density. At low frequencies the flux density is high and more distortion is introduced because of the saturation of the iron. Such saturation is frequently tolerated to avoid using more iron which increases size and cost of the transformer. The maximum allowable flux density is determined by the maximum allowable distortion.

The output transformer causes a reduction in the output of a power amplifier at both the high and low frequencies as illustrated in the frequency response curve (fig. 7-5). The reduced output at the low frequencies results from the shunting action of the transformer

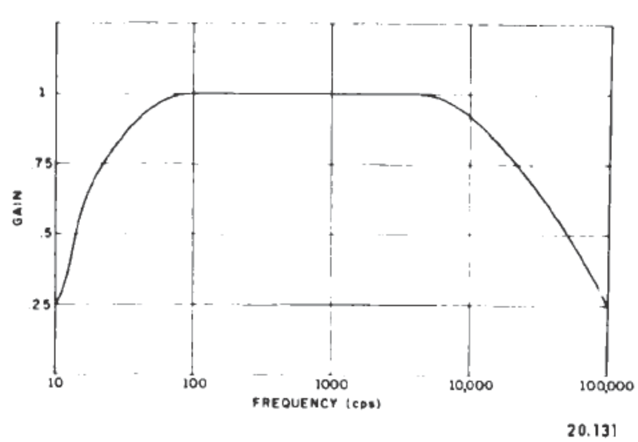


Figure 7-5.—Frequency response of an output transformer.

primary inductance on the load, as indicated in figure 7-6, A.

The following symbols are used in figure 7-6:

Symbol	Definition
μ	Amplification factor of tube
E_g	Input voltage (rms)
r_p	Plate resistance
R_1	Resistance of primary winding
L_1	Leakage inductance of primary
L_p	Incremental primary inductance
N	Primary-to-secondary turns ratio
L_2	Leakage inductance of secondary
R_2	Resistance of secondary winding
R_L	Load resistance
E_L	Voltage (rms) developed across the load
k	Coefficient of coupling

The middle-frequency gain is independent of frequency, as indicated by the absence of reactance in the equivalent circuit (fig. 7-6, B). The reactance of the primary inductance is large enough for its shunting effect to be

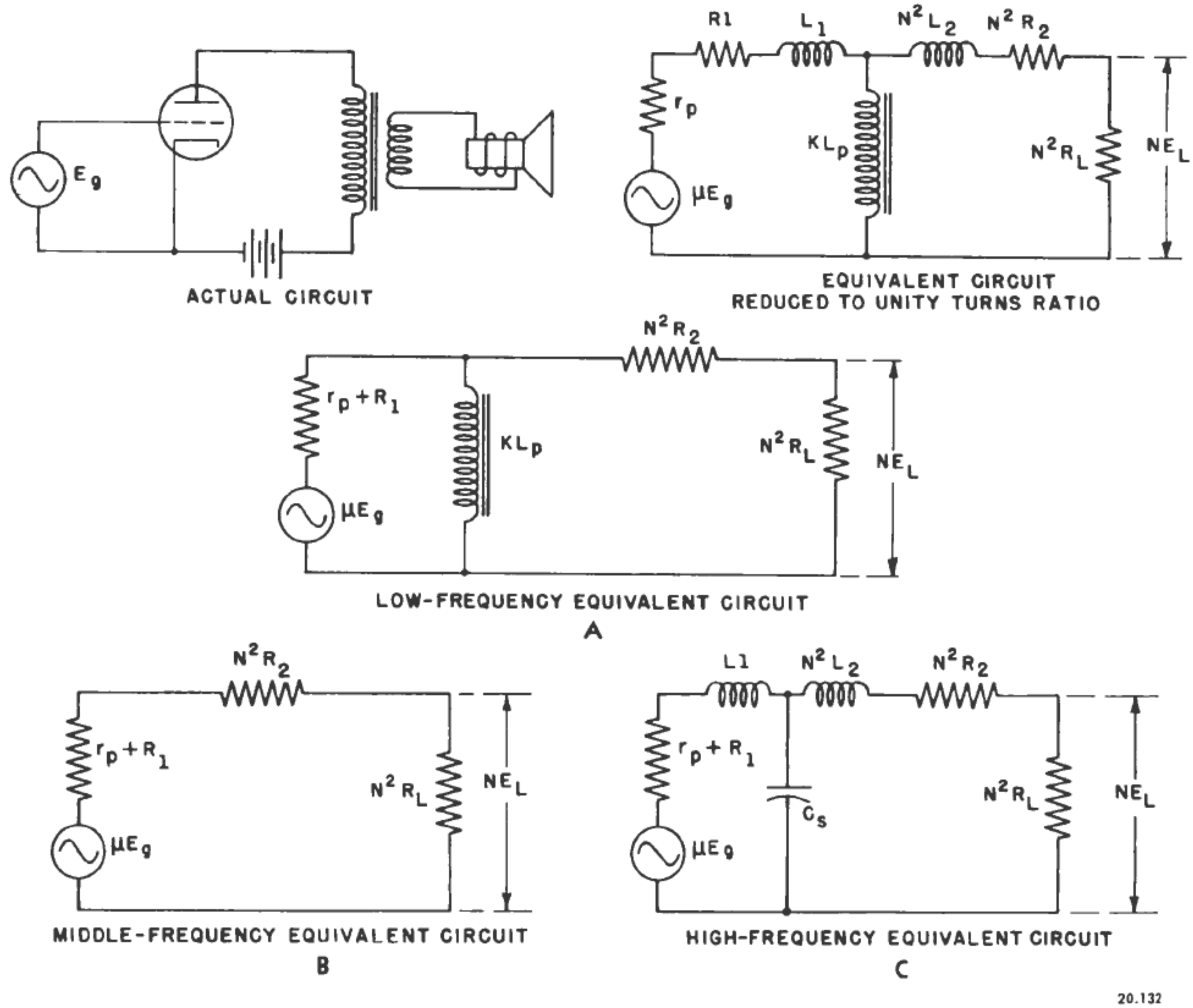


Figure 7-6.—Circuit analysis of an output transformer.

disregarded and the leakage reactances are low enough to be neglected.

The reduced output at the high frequencies (fig. 7-6, C) results from the loss in voltage through the leakage reactances as a result of (1) load current, and (2) capacitive current due to shunting capacitance.

In order to extend the flat portion of the frequency-response curve into the low-frequency region, the transformer should have a high primary inductance. In order to extend the flat portion into the high-frequency region, the leakage inductance values, L_1 and L_2, should be low. With a given triode and transformer, increasing the load resistance, R_L, improves the high-frequency response without disturbing the low-frequency response very much.

The voltage E_L across the load at the various frequency ranges may be determined from the equivalent circuit shown in figure 7-5. For example, at the middle range of frequencies,

$$NE_L \; \frac{\mu E_g N^2 R_L}{(r_p + R_1) \; N^2(R_2+R_L)}$$

and

$$E_L \; \frac{\mu E_g N R_L}{(r_p+R_1) \; N^2(R_2+R_L)}$$

PUSH-PULL POWER AMPLIFIERS

A number of advantages are to be gained by the use of a push-pull amplifier as the output stage of an audio-frequency amplifier. Second harmonics and all even-numbered harmonics, as well as even-order combinations of frequencies, will be effectively eliminated if the tubes are properly balanced and if the frequencies are introduced within the output tubes themselves.

Hum from the plate power supply, which may be present in the single-tube amplifier, is substantially reduced in the push-pull amplifier because ripple components in the two halves of the primary of the output transformer are in phase and tend to counteract each other in the output.

Plate-current flow through the two halves of the primary winding is equal and in opposite directions. Therefore there is no d-c core saturation and the low-frequency response is improved.

Regeneration is also eliminated because signal currents do not flow through the plate-voltage supply when the circuit is operated as a class-A amplifier.

The last voltage amplifier preceding the push-pull power amplifier stage may be either resistance- or transformer-coupled to the power stage. If the power amplifier is operated class A or class AB, the driver commonly employs resistance coupling because it affords a better frequency response. A phase-inverter tube, or section of a tube, must be used in connection with the resistance-coupled driver to provide the correct phase relation at the input of the push-pull stage.

When the power tubes are operated class B, an input transformer employing a step-down turns ratio is commonly used. The transformer not only supplies the grid current necessary for class-B operation, but at the same time permits an instantaneous signal voltage of the correct polarity to be applied to the grids of the two power tubes.

Class-B power amplifiers draw practically no plate current when no signal is applied, and plate efficiency is much higher than that of class-A amplifiers. They are subject, however, to third-harmonic distortion, and the operating conditions are critical.

TRANSFORMER-COUPLED

A transformer-coupled push-pull amplifier is shown in figure 7-7. It is assumed that the following operating conditions are in effect: The grid bias is such that the plate current in each tube flows during the entire input-voltage cycle (class A), and during the positive half of the cycle the grid-voltage excursion extends over a portion of the lower bend of the i_p-e_g characteristic curve. The tubes are also properly matched and operate into the correct loads.

When no signal voltage is applied, equal plate currents flow through each tube. Equal currents also flow through each half of the primary of the output transformer toward the center tap. The magnetomotive forces resulting from the currents are equal and opposite and therefore cancel, leaving no magnetic field due to the d-c component of the plate current. This

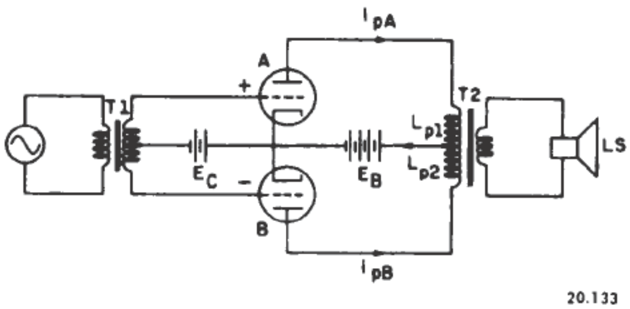


Figure 7-7.—Transformer-coupled push-pull amplifier.

cancellation effect is a big advantage over the single-tube output in which direct current flows continuously through the primary winding and establishes a d-c field component.

A signal voltage across the secondary of the input transformer, T1, will at a given instant have polarities as indicated. This voltage will be divided equally between tubes A and B. The push-pull arrangement thus requires, and will handle, twice the input voltage of a single tube under similar operating conditions. The grid of tube A is positive with respect to the center tap at the instant the grid of tube B is negative. Plate current increases in tube A and decreases a proportionate amount in tube B.

The increase in current flowing down through L_{p1} and the decrease in current flowing up through L_{p2} constitute two magnetomotive forces that combine additively to produce an output voltage in the secondary that is proportional to the sum of these two components.

Second harmonics are eliminated in the push-pull output, as shown in figure 7-8. The dynamic i_p-e_g curve for tube B is inverted with respect to that of tube A in order to show the phase relation between the signal components of the two tubes. Thus, when the input signal swings the grid voltage of tube A in a positive direction it is swinging the grid voltage of tube B the same amount in a negative direction. Plate current in tube A increases as plate current in tube B decreases. The plate current swing about the X-X axis for tube A is not symmetrical because the tube is operating on a nonlinear portion of the i_p-e_g characteristic curve. The same condition is true of the plate current swing about the X'-X' axis for tube B.

The plate-current curves of each tube may be resolved into a fundamental and second harmonic. Thus, the axis of the fundamental and its second harmonic is displaced from the axis of the original plate current curve by an amount equal to the peak value of the second harmonic. Combining the fundamental components of both tubes gives an output of twice the amplitude of one tube. However, when the second harmonics are combined, the resultant is zero because they are 180 degrees out of phase. The fundamental output current has the same waveform as the input voltage—an effect that would have been produced had both tubes been free of second harmonics.

POWER OUTPUT

The power output of a class-A push-pull amplifier is conveniently determined by using the equivalent circuit shown in figure 7-9, A. If e_g is the maximum input-signal voltage then the maximum current through the load is

$$I_{max} = \frac{2\mu e_g}{r_{p1}+r_{p2}+R_L}$$

and the peak voltage across the load is

$$E_{max} = R_L I_{max}$$

$$= \frac{2R_L \mu e_g}{2r_p + R_L}$$

The average power output consumed in the load is

$$P_o = \frac{E_{max} I_{max}}{2}$$

$$= 2R_L \left(\frac{\mu e_g}{2r_p + R_L}\right)^2$$

As an example of push-pull class-A triodes, assume that the mu of each tube is 4 and that the plate resistance is 800 ohms; also assume that the peak signal voltage is 43 volts and that the effective load impedance (plate-to-plate) is 5,000 ohms. The power output is

$$P_o = 2R_L \left(\frac{\mu e_g}{2r_p + R_L}\right)^2$$

$$= 2\text{x}5{,}000 \left(\frac{4\text{x}43}{2\text{x}800 + 5{,}000}\right)^2 = 6.8 \text{ watts}$$

For class-A power pentodes, assume that the mu of each tube is 120 and that the plate resistance is 22,000 ohms; also assume that the peak-signal voltage is 16 volts and that the total load impedance is 5,000 ohms. The average power output is

$$P_o = 2R_L\left(\frac{\mu e_g}{2r_p + R_L}\right)^2$$

$$= 2\text{x}5{,}000\left(\frac{120\text{x}16}{2\text{x}22{,}000+5{,}000}\right)^2 = 15.3 \text{ watts}$$

If one tube is removed from the circuit the voltage acting in the equivalent circuit is lowered and at the same time the impedance in the plate load is reduced. Since the output impedance varies as the square of the number of turns, the removal of one tube leaves only one-half as many turns in the plate circuit. The load impedance is therefore $(1/2)^2$, or 1/4 of its former value, as indicated in figure 7-9, B.

The maximum plate current now becomes

$$I_{max} = \frac{\mu e_g}{r_p + \frac{R_L}{4}}$$

and the maximum voltage across the load becomes

$$E_{max} = I_{max}\frac{R_L}{4}$$

$$= \frac{\mu e_g}{r_p + \frac{R_L}{4}} \text{ x } \frac{R_L}{4}$$

The average power output is then

$$P_o = \frac{I_{max}E_{max}}{2}$$

$$= \frac{\dfrac{\mu e_g}{r_p + \frac{R_L}{4}} \text{ x } \dfrac{\mu e_g}{r_p + \frac{R_L}{4}} \text{ x } \dfrac{R_L}{4}}{2}$$

$$= \frac{R_L}{8}\left(\frac{\mu e_g}{r_p + \frac{R_L}{4}}\right)^2$$

If one of the pentodes of the preceding example is removed from the circuit, the power output becomes

$$P_o = \frac{5{,}000}{8}\left(\frac{120\text{x}16}{22{,}000+\frac{5{,}000}{4}}\right)^2 = 4.25 \text{ watts}$$

This is a reduction in power of 72 percent.

The CLASS-B POWER AMPLIFIER is biased approximately to cutoff so that plate current flows in one tube during one half of the input cycle and in the other tube during the other half of the input cycle. Thus, one tube amplifies the positive half and the other the negative half of the input signal voltage. Both halves are combined in the secondary of the output transformer.

Since plate current flows in a given tube for only half of the input cycle and no appreciable current flows when no signal is applied, this type of amplifier has a higher efficiency than an amplifier operated class A or class AB. A class-B push-pull amplifier circuit is shown in figure 7-10, A.

The input-signal voltage curves shown in figure 7-10, B, make the grids alternately positive with respect to their associated cathodes on the positive peaks. When the grid voltage of triode A is positive maximum the plate voltage is minimum and the plate current is maximum. At this instant the reverse operation is taking place in triode B. The output signal voltages of the two tubes combine in series addition across the primary of the output transformer.

The plate voltage is alternately the sum and difference of the B-supply voltage and the induced voltage of one-half of the transformer primary.

The maximum plate current depends on the positive peak of the grid signal, and the minimum plate potential acting simultaneously on the tube electrodes. The minimum plate potential depends on the amount of voltage drop in the plate load. For best efficiency the minimum plate voltage should be as small as possible but not so small that the grid will draw excessive current.

The correct plate-to-plate load resistance in terms of I_{max}, E_{min}, and E_B of a single tube may be determined as follows: The current

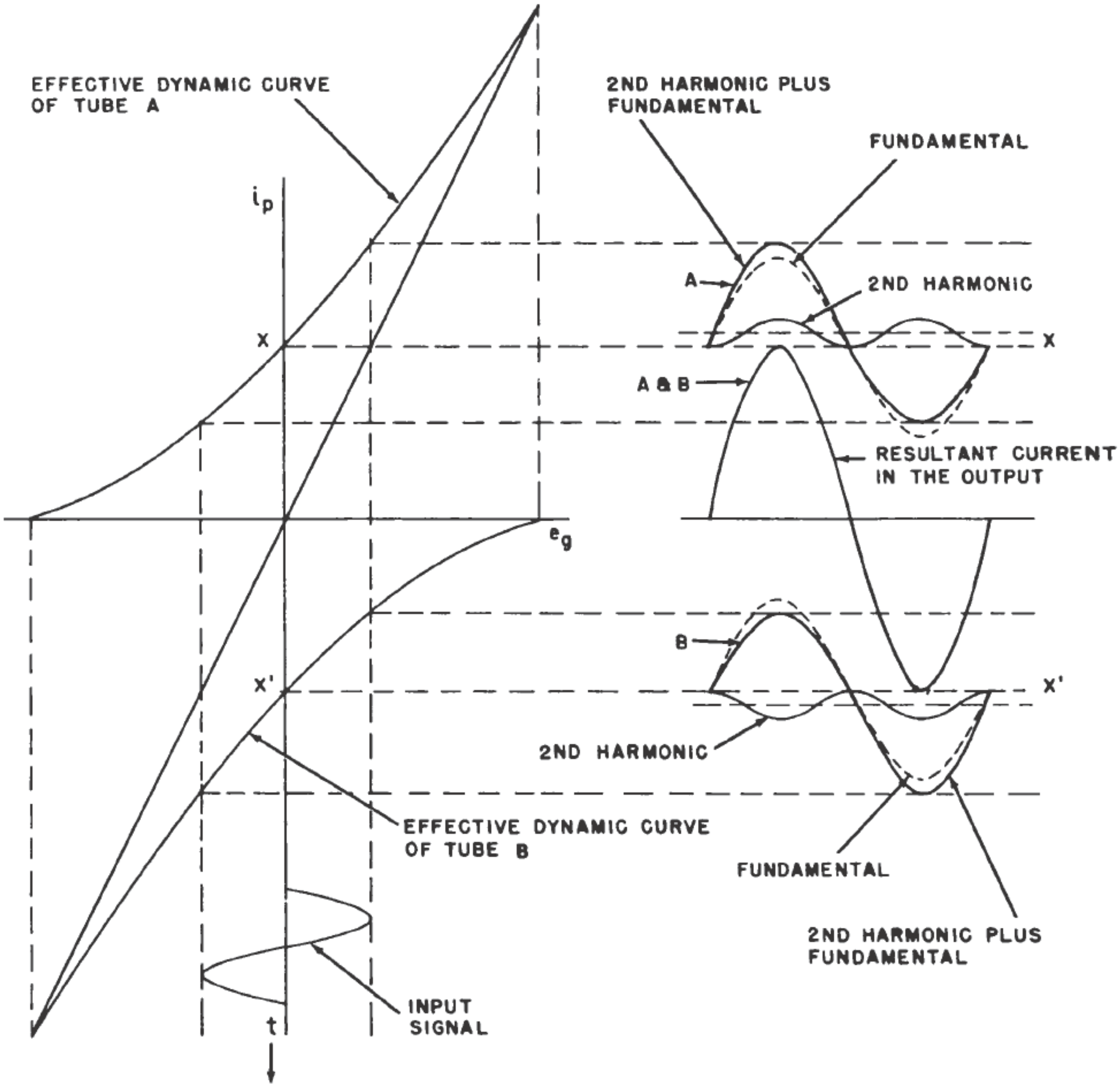
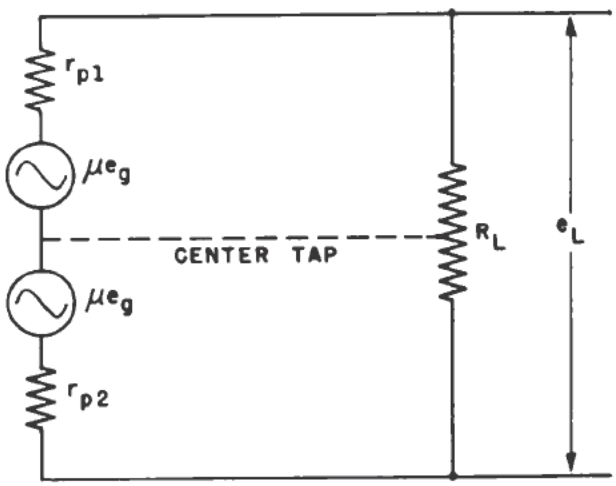


Figure 7-8.—Graph showing second-harmonic elimination.

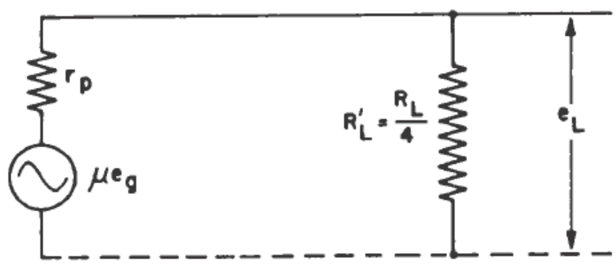
pulses from tubes A and B flow through only one-half of the primary of the output transformer; if they flowed through the entire primary, the amplitude would be only half as much, or $\frac{I_{max}}{2}$

The equivalent load resistance, R_L, acts in series with the triode plates and is equal to the impedance of the output transformer and its load as measured in terms of the primary. The voltage drop across R_L is $\frac{I_{max}}{2} R_L$. The alternating voltage drop between the plate and cathode of one tube is equal to one-half the voltage drop across R_L since the 2 tubes are in series. Thus,

$$e = \frac{1}{2} \frac{I_{max} R_L}{2} = \frac{I_{max} R_L}{4}$$



A
BOTH TUBES IN THE CIRCUIT



B
ONE TUBE REMOVED FROM THE CIRCUIT

20.135

Figure 7-9.—Equivalent circuit of a push-pull amplifier.

Since the voltage across a single tube is equal to $E_B - E_{min}$, the equation may be rewritten as

$$\frac{I_{max} R_L}{4} = E_B - E_{min}$$

from which

$$R_L = 4\,\frac{(E_B - E_{min})}{I_{max}}$$

The average power output is equal to one-half the peak instantaneous power, or

$$P = \frac{I_{max}(E_B - E_{min})}{2}$$

Using the values indicated on the curves of figure 7-10, B

$$R_L = \frac{4(400-100)}{0.120} = 10{,}000 \text{ ohms}$$

and the average power output is

$$P = \frac{0.120(400-100)}{2} = 18 \text{ watts}$$

The output power may also be determined by the use of the equivalent circuit shown in figure 7-10, C. If the maximum value of grid voltage, e_g, is assumed to be 87 volts, the maximum value of the signal current in the equivalent circuit will be

$$i = \frac{2\mu e_g}{2r_p + R_L} = \frac{2x4x87}{2x800+10{,}000} = 0.06 \text{ ampere}$$

The maximum value of the signal voltage appearing across the load is

$$e_L = iR_L$$

$$= 0.06x10{,}000 \quad 600 \text{ volts}$$

and the average power output is

$$P_o = \frac{e_L i}{2}$$

$$= \frac{600x0.06}{2} = 18 \text{ watts}$$

The output power of the class-B push-pull triode amplifier is about 2.6 times the output of the class-A triode amplifier given previously in this chapter.

Although approximately twice the signal voltage is applied to the grids of the class-B triodes as is applied to those of the class-A triodes, the output power is not 4 times as much because the load resistance of the class-B triode push-pull circuit is twice that of the class-A stage.

THE DECIBEL

The decibel is more practical than the bel for expressing power ratios. The bel is named in honor of Alexander Graham Bell, inventor of the telephone. Ten decibels equal one bel.

UNIT OF POWER GAIN OR LOSS

The international transmission unit, the BEL, is a unit of gain equivalent to 10 to 1 ratio of power gain. Thus the gain in bels is simply the number of times that 10 is taken as a factor to equal the ratio of the output power of an amplifier to the input power. If,

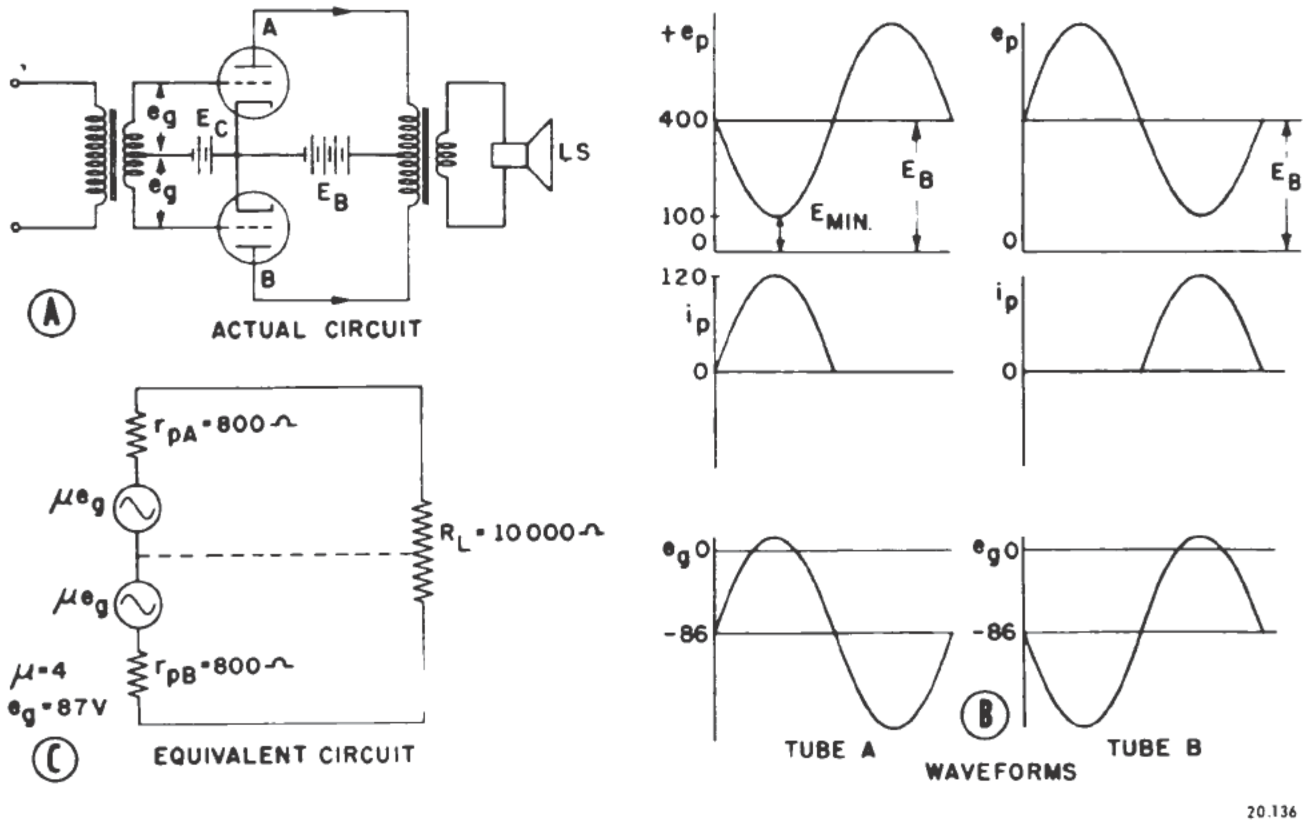


Figure 7-10.—Class-B operation.

for example, the output power is 100 times the input, the ratio is 100 to 1, or 10^2 to 1. The gain is therefore 2 bels: and the gain in decibels (db) is 10 times 2, or 20 decibels.

Where the power ratio is 1000 to 1, it may also be written as 10^3 to 1, or 3 bels; and the gain in db is 10 x 3, or 30 decibels.

Where the power ratio has been increased 10,000 times, that gain is 10^4 to 1, or 4 bels and the gain in decibels equals 40 db. This relationship between the numbers can be extended and readily associated in table 7-1.

The number of 10 factors contained in any ratio of the output power to the input power is the logarithm of the ratio to the base 10. The gain in decibels may therefore be expressed conveniently as

$$\text{db} = 10 \log_{10} \frac{P_2}{P_1}$$

where P_2 and P_1 are respectively the output and input power in watts.

The human ear responds to ratio changes in intensity rather than to changes in absolute value.

Table 7-1.—Relation between db and power ratio.

Power Ratio	No. of 10 factors	No. of BELS	No. of db
1,000,000 to 1	10^6	6	60
100,000 to 1	10^5	5	50
10,000 to 1	10^4	4	40
1,000 to 1	10^3	3	30
100 to 1	10^2	2	20
10 to 1	10^1	1	10
1 to 1	10^0	0	0

In other words, the ability of the human ear to detect changes in the intensity of sound is much greater at low levels of intensity than it is at high levels. A change in power level of 1 db is barely perceptible to the ear, and for this reason attenuators in audio systems are frequently calibrated in steps of 1 db.

Because the ear responds logarithmically to variations in sound-intensity levels, any practical system for measuring sound-intensity levels must necessarily vary logarithmically. The decibel system of measuring power levels is based on this concept.

Since the gains or losses in a system are expressed logarithmically, they are simply added or subtracted to determine the overall gain or loss. For example, transmission lines introduce a loss in power, amplifier stages produce a gain, and attenuators introduce a loss. The final result is the algebraic sum of the various gains and losses. A db gain or loss is readily determined by the use of equation (7-4).

CURRENT AND VOLTAGE RATIOS

Primarily, the decibel is a unit of measure of power ratios. It can be used readily to compute current ratios as well, provided the resistances through which the currents flow are taken into account. The db gain or loss expressed in terms of the currents and resistances is determined as follows:

$$\text{db} = 20 \log_{10} \frac{I_2 \sqrt{R_2}}{I_1 \sqrt{R_1}} \qquad (7\text{-}5)$$

where I_2 and I_1 are respectively the output and input currents in amperes, and R_2 and R_1 are respectively the output and input resistances in ohms. Thus, if the two currents and resistances are known, the db gain or loss can be determined by substitution in equation (7-5). If the resistances are equal they may be canceled out.

The same reasoning also applies to the voltage ratio provided the resistances across which the voltages are applied are properly considered. The equation for db gain or loss when voltages and resistances are employed directly is determined as follows:

$$\text{db} = 20 \log_{10} \frac{E_2 \sqrt{R_1}}{E_1 \sqrt{R_2}} \qquad (7\text{-}6)$$

where E_2 and E_1 are respectively the output and input voltages, and R_2 and R_1 are respectively the output and input resistances in ohms.

If the voltages and resistances are known, the db gain or loss may be determined by direct substitution in equation (7-6). If the resistances are equal they may, of course, be canceled out.

REFERENCE LEVELS

Considerable confusion has resulted from the use of various so-called zero-power reference levels. The term "zero reference level" is in itself somewhat confusing because it does not mean that no power is developed at that level. It means rather that the output level is referred to an arbitrary level designated as the reference, or zero, level; and as such it is perhaps one of the most convenient ways of expressing a power ratio. It is meaningless to say, for example, that a certain amplifier stage has an output of 30 db unless reference is made to some established power level.

It is common practice in naval and other telephone work to consider 6 milliwatts as the reference power level. Other values are also used in this and other fields; for example, 1, 10, and 12.5 milliwats, depending upon which unit is most convenient under the circumstances.

The VOLTAGE gain or loss of microphones, transmission lines, and voltage amplifiers is also generally expressed in decibels. In general, transmission lines introduce a loss and voltage amplifiers produce a gain. A reference voltage level and the resistance (if it differs from the one being compared) across which the signal appears must be given in order that the gain or loss may have meaning.

The voltage output of a microphone may be expressed in terms of decibels below 1 volt per dyne per cm^2. In other words, 1 dyne acting on 1 square centimeter and producing an output of 1 volt is taken as the zero-decibel output level.

When the voltage gain of an amplifier stage is given in decibels the input and output impedances must be given or they must be assumed to be equal. Thus if they are assumed to be equal, the gain in decibels may be expressed as

$$\text{db} = 20 \log_{10} \frac{E_2}{E_1}$$

where E_2 is the output voltage and E_1 the input voltage.

When certain arbitrary power reference levels are used, the db gain or loss is given a special designation. One of these designations is the dbm, or the power level in decibels referred to 1 milliwat, as

$$\text{dbm} = \log_{10} \frac{P}{0.001}$$

where P is the power in watts.

Typical power levels in dbm (0.001 watt in 600 ohms) are shown in figure 7-11 for various applications in audio systems.

The volume level of an electrical signal made up of speech, music, or other complex tones is measured by a specially calibrated

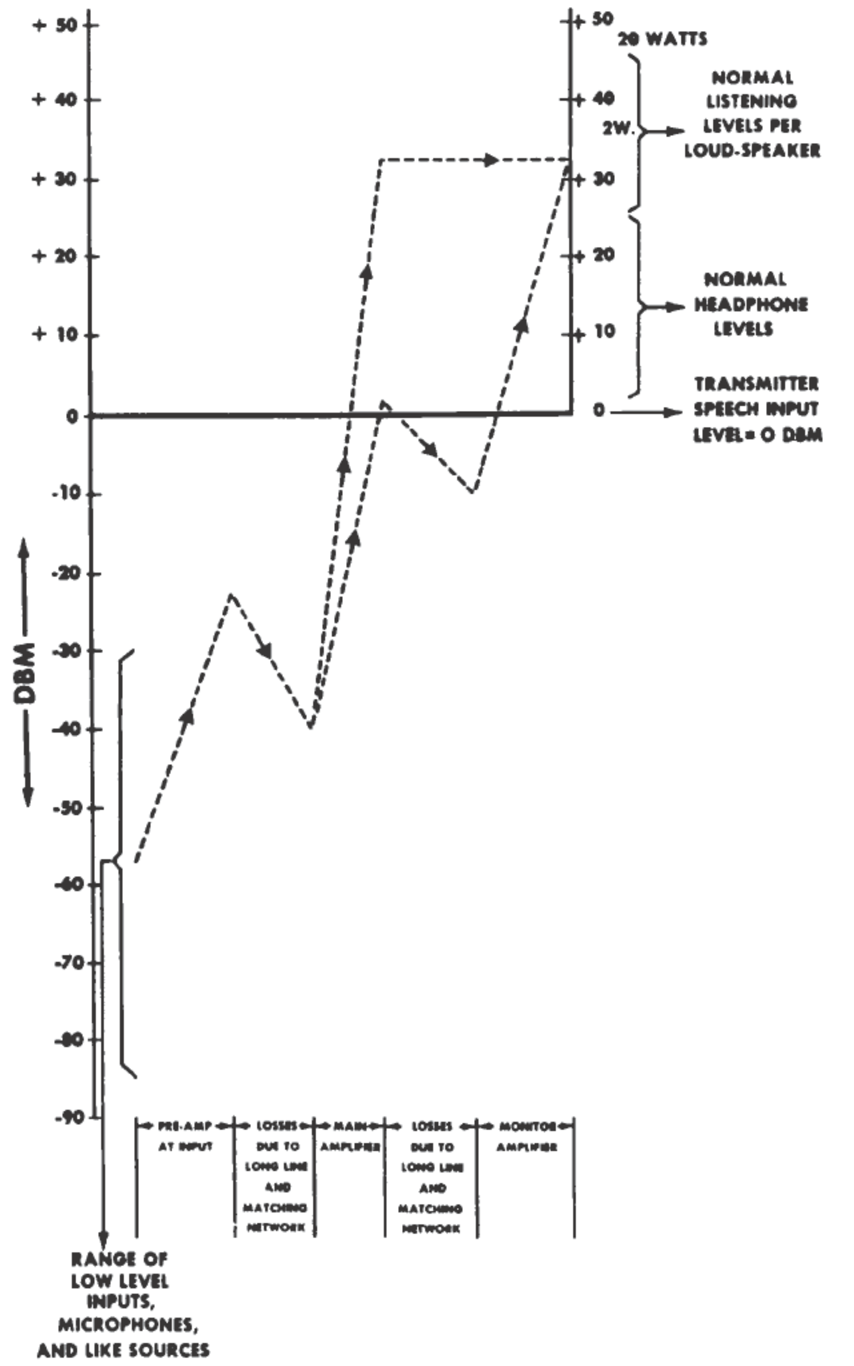


Figure 7-11.—Typical power levels in dbm (0.001 watt in 600 ohms) for various parts in audio systems.

voltmeter called a VOLUME INDICATOR. The volume levels registered on this indicator are expressed in volume units (VU). The number of units is numerically equal to the number of decibels above or below the reference volume level. Zero VU represents a power of 1 milliwatt dissipated in an arbitrary load resistance of 600 ohms (corresponding to a voltage of 0.7746 volts). Thus, when the VU meter is connected to a 600-ohm load, VU readings in decibels can be used as a direct measure of power above or below a 1-milliwatt reference level.

USE OF DECIBELS

Some practical examples illustrating the use of decibels will clarify the foregoing paragraphs. (Logarithms to the base 10 are called COMMON LOGARITHMS. When the subscript following the logarithm is omitted, the base is understood to be 10. The following examples involve common logarithms.)

1. How many decibels correspond to a power ratio of 100?

$$\text{db} = 10 \log \frac{P_2}{P_1}$$

$$= 10 \log 100$$

$$= 10x2 \quad 20$$

2. How many decibels correspond to a voltage ratio of 100 (assume equal resistances)?

$$\text{db} = 20 \log \frac{E_2}{E_1}$$

$$= 20 \log 100$$

$$= 20x2 = 40$$

3. If an amplifier has a 20-db gain, what power ratio does this gain represent?

$$\frac{P_2}{P_1} = x$$

$$\text{db} = 10 \log x$$

$$20 = 10 \log x$$

$$\log x = 2$$

$$x = 100$$

4. If an amplifier has a 30-db gain, what voltage ratio does this gain represent (assume equal resistances)?

$$\frac{E_2}{E_1} = x$$

$$\text{db} = 20 \log x$$

$$30 = 20 \log x$$

$$\log x = 1.5$$

$$x = 31.6$$

The voltage ratio is 31.6 to 1.

5. A certain microphone rated at -75 db is connected to a preamplifier through an attenuator rated at -10 db. The final audio amplifier is driven by the preamplifier and has a gain of 30 db. What must be the db gain of the preamplifier to balance out the losses in the microphone and attenuator? (All db gains or losses have the same reference level).

The total loss is 75+10 or 85 db and therefore the preamplifier must have a gain of 85 db to bring the losses to 0 db. From this point the main amplifier increases the gain 30 db above the common, or zero, reference level.

6. If the input to a certain loudspeaker is increased from 5,000 milliwats to 6,000 milliwatts, could the increase in volume level be readily detected by the human ear?

$$\text{db} = 10 \log \frac{P_2}{P_1}$$

$$= 10 \log \frac{6{,}000}{5{,}000}$$

$$= 10 \log 1.2$$

$$= 10x0.0792$$

$$= 0.792$$

Because a change of 1 db is barely discernible, a change of 0.792 db would probably not be detected.

7. If 1 volt is applied across the 600-ohm input of a certain amplifier and 500 volts is

developed across the 5,000-ohm output, what is the db power gain?

$$P_2 = \frac{E_{out}^2}{R_{out}} = \frac{500^2}{5{,}000} = 50 \text{ watts}$$

$$P_1 = \frac{E_{in}^2}{R_{in}} = \frac{1}{600} = 0.00166 \text{ watt}$$

$$\text{db} = 10 \log \frac{P_2}{P_1}$$

$$= \log \frac{50}{0.00166}$$

$$= 10 \log 30{,}000$$

$$= 10 \text{x} 4.4770$$

$$= 44.8$$

8. If the noise level in a certain transmission line is 50 db down from the desired signal level of 10 mw, how much power is contained in the noise level?

$$\text{db} = 10 \log \frac{P_2}{P_1}$$

where P_1 is the power in milliwatts contained in the noise level and P_2 is the power in milliwatts contained in the desired signal level.

Substituting,

$$50 = 10 \log \frac{10}{P_1}$$

$$5 = \log \frac{10}{P_1}$$

$$10^5 = \frac{10}{P_1}$$

$$P_1 = \frac{10}{10^5} = 10^{-4} = 0.0001 \text{ mw}$$

CHAPTER 8

OSCILLATORS

INDUCTANCE-CAPACITANCE OSCILLATORS

The primary function of an oscillator is to generate a given frequency and to maintain that frequency within certain limits. To that end inductance-capacitance oscillators depend for their operation on the resonant interchange of energy between a capacitor and an inductor, with an electron-tube amplifier supplying pulses of energy of the proper phase and magnitude to maintain the oscillations. Resonant circuits (tuned circuits) are treated in chapter 4, and electron-tube amplifiers are treated in chapters 6 and 7.

In addition to their use as amplifiers, electron tubes are used as oscillators for the generation of alternating voltages. When thus used as oscillators, electron tubes are essentially converters that change d-c electrical energy from the plate power supply into a-c electrical energy in the output circuit. To accomplish this energy conversion, the amplifying ability of the electron tube is used in such a manner as to generate sustained oscillations.

Two conditions are necessary if sustained oscillations are to be produced. FIRST, the feedback voltage from the plate circuit must be in phase with the original excitation voltage on the grid—that is, the feedback must be positive, or regenerative. SECOND, the amount of energy fed back to the grid circuits must be sufficient to compensate for the energy losses in the grid circuit.

Feedback may be accomplished by inductive, capacitive, or resistive coupling between the plate and the grid circuit. Various circuits have been developed to produce feedback of the proper phase and amount. Each of these circuits has certain characteristics that make its use advantageous under given circumstances.

If the proper values of inductance and capacitance are used, tuned-circuit oscillators may be designed to generate frequencies from the low frequencies in the audio range to the very high radio frequencies. The upper frequency limit is determined in general by the distributed inductance and capacitance of the circuit components and the interelectrode capacitance of the tubes.

The electron tube itself is not an oscillator. The oscillations actually take place in the tuned circuit, a part of which may be composed of the interelectrode capacitances of the electron tube and the distributed capacitances and inductances of the circuit. The electron tube functions primarily as an electrical valve that amplifies and automatically delivers to the grid circuit the proper amount of energy to maintain oscillation.

A basic oscillator showing the circuits necessary for its operation is shown in figure 8-1.

TICKLER-FEEDBACK OSCILLATOR

One of the simplest types of oscillator circuits is that employing tickler feedback, as shown in figure 8-2. Feedback voltage of the proper phase from the plate to the grid circuit is accomplished by mutual inductive coupling between the oscillator tank coil, L2, and the

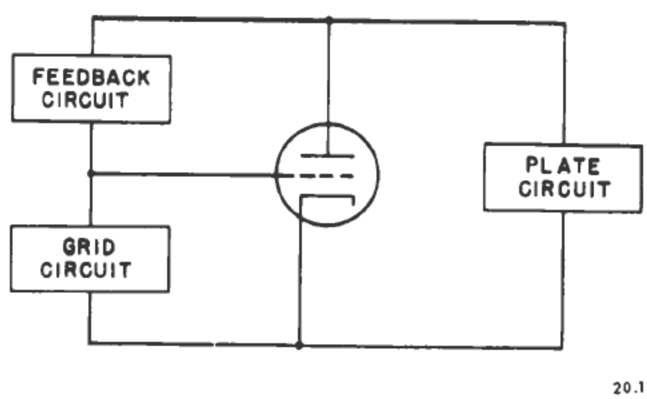


Figure 8-1.—Basic oscillator circuit.

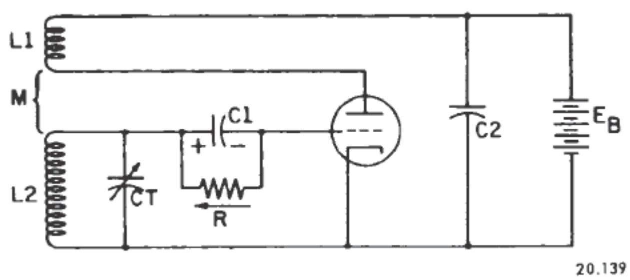


Figure 8-2.—Ticker feedback oscillator circuit.

tickler feedback coil, L1. The amount of feedback voltage is determined by the amount of flux from L1 that links L2. Thus the feedback voltage is varied by moving L1 with respect to L2 or by changing the setting of a variable resistor that is sometimes shunted across L1.

The frequency-determining part of the oscillator is the tank circuit, L2CT. The coil and tuning capacitor interchange energy at the resonant frequency rate and the excitation voltage developed across CT is applied to the grid in series with the grid-leak bias across RC1.

Grid current flowing through R establishes negative bias on the grid. Capacitor C1 charges up to the peak voltage across R and holds this voltage between r-f pulses because of the relatively long time constant of RC1 as compared with the time for each r-f cycle. A steady bias voltage is therefore developed when the oscillator is functioning properly. Thus one test for proper operation is a measurement of the d-c voltage across the grid resistor. To avoid blocking the oscillations, a high-resistance type of meter, such as the electron-tube voltmeter, should be used for this measurement.

SERIES-FED HARTLEY OSCILLATOR

Figure 8-3, A shows the circuit of a series-fed Hartley oscillator and the curves of grid voltage, plate current, and plate voltage for class-C operation. The plate circuit is shown in heavy lines. Inductor L1 is a part of the tuned circuit made up of L1, L2, and CT; it also serves to couple energy from the plate circuit back into the tuned grid circuit by means of the mutual inductance between L1 and L2. Capacitor C1 blocks the d-c component of grid current from L2, and, together with R_g, provides the necessary operating bias. Capacitor C2 and the radio-frequency choke (r-f-c) keep the alternating component in the plate circuit out of the B-supply.

The B-supply in figure 8-3, A is returned to the resonant tank coil, L1. The tuned circuit therefore contains a d-c component of plate current in addition to the a-c signal component. Thus the B-supply is connected in series with the plate and tank coil L1. This is a SERIES-FED connection.

The Hartley oscillator operates with both mutual inductive coupling between L1 and L2 and by the coupling provided by capacitor CT. The following explanation is based on mutual inductive coupling:

1. When the cathode is heated and plate voltage is applied, plate current rises.
2. The increase in plate current through L1 induces a voltage in L2 with the grid end positive, causing plate current to rise to point 1, figure 8-3, B.
3. Plate current through L1 stops rising at point 1 and the voltage induced in L2 is reduced to zero.
4. Plate current starts to fall.
5. The decrease in plate current through L1 induces a voltage in L2 that makes the grid swing in a negative direction and plate current further reduces.
6. When plate current in L1 stops falling (point 2) the mutually induced voltage in L2 again drops to zero.
7. In the absence of the negative going voltage in L2, plate current again starts to rise and the cycle repeats (point 3).
8. On each subsequent cycle the bias voltage builds up across C1 and R_g until it reaches a safety value as shown in figure 8-3, B.
9. Normal bias indicates class C operation (fig. 8-3, C). The interchange of energy between the tank coil and capacitor (flywheel effect) maintains the oscillations during the time that plate current is zero and no energy is being delivered to the oscillator from the B supply.

The Hartley oscillator will also operate if M, the mutual inductance between L1 and L2, is zero (fig. 8-3, A) because of the coupling effect of tank capacitor CT. For example, if plate current is increasing, CT will charge to a higher voltage and the accompanying charge current will flow down through L2. The decay in the charge current is accompanied by a self-induced voltage e2 in L2 having a positive polarity on

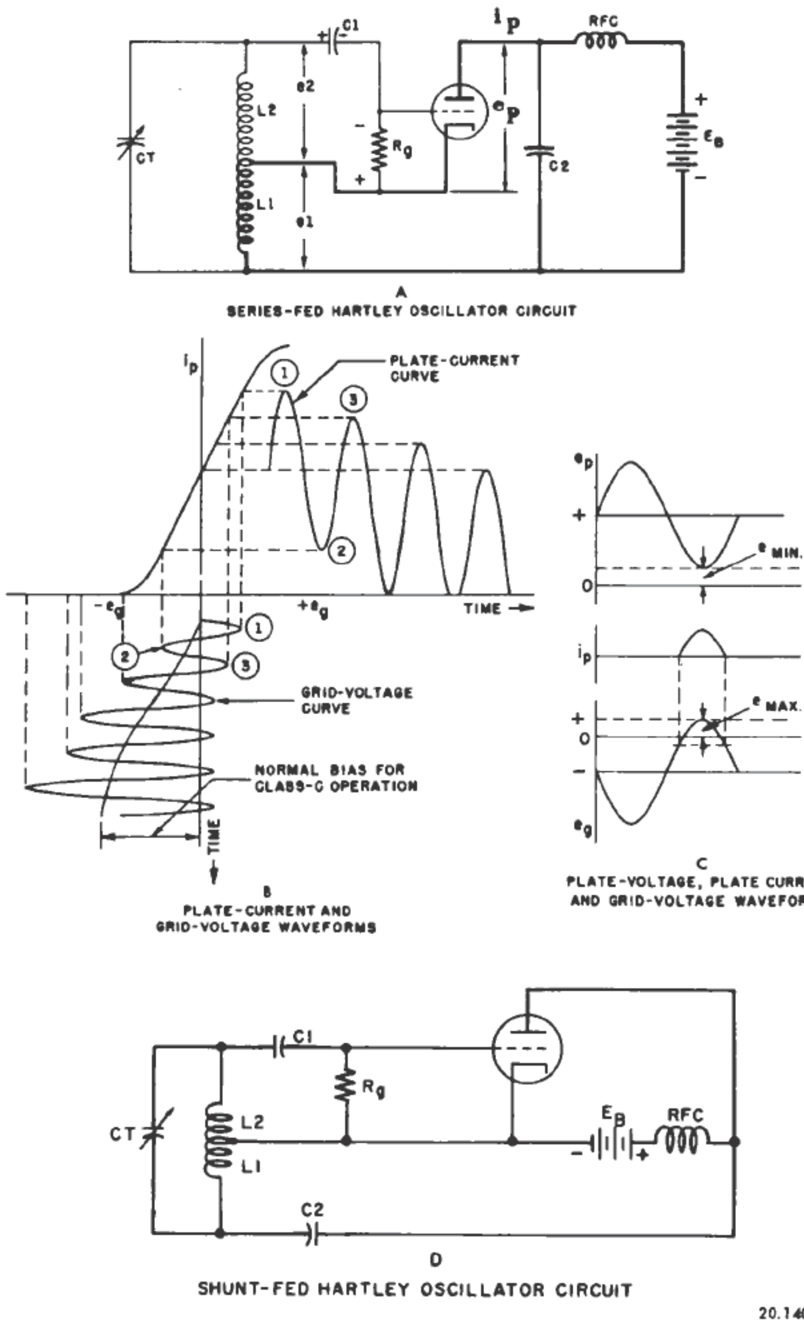


Figure 8-3.—Analysis of Hartley oscillator.

the grid end. (The self-induced voltage is in the same direction as the charge current and tries to sustain it.)

Conversely, if plate current is decreasing, tank capacitor CT will discharge. The path of the discharge current is up through L2. The self-induced voltage e2 in L2 now makes the grid negative with respect to the cathode causing a further decrease in plate current.

Thus, CT couples a self-induced voltage on both half cycles, back to the grid, in phase with the initial excitation voltage (positive feedback) and oscillations are sustained.

The time constant of R_gC1 should be relatively long compared with the time for one oscillator cycle. If it is too short, the bias will be lost, and if it is too long the oscillator will be blocked periodically because of excessive accumulated bias. Self-bias makes the oscillator self-starting when the filament is energized and the place voltage is applied.

Loading the oscillator may be accomplished by placing a relatively high resistance in shunt with the tank circuit or a low resistance in series with it. In either case the effective Q is reduced and the amplitude of the generated voltage is lowered. If the load is coupled inductively to the oscillator, the point of coupling should be near r-f ground. Loading the oscillator may detune it. Buffer stages are frequently inserted between the oscillator and its load to isolate the oscillator and prevent frequency changes.

The resonant frequency of the tuned circuit is essentially the frequency of the oscillator. Expressed in terms of the resonant tank inductance and capacitance, the resonant frequency is

$$f = \frac{1}{2\pi\sqrt{LC}}$$

where f is the resonant frequency in cycles per second, L, the inductance in henrys and C the capacitance in farads. This expression was derived in chapter 4 of this training course from the relation $X_L = X_C$ at resonance.

A more convenient form of the formula may be derived by making the following substitutions in the above formula:

$$L = L_{\mu h} \times 10^{-6}$$

$$C = C_{\mu\mu f} \times 10^{-12}$$

$$\frac{1}{2\pi} = 0.159$$

$$f = f_{mc} \times 10^{6}$$

where $L_{\mu h}$ is the inductance in mcirohenrys, C the capacitance in micromicrofarads, and f_{mc} is the frequency in megacycles.

$$f_{mc} = \frac{0.159 \times 10^{-6}}{\sqrt{L_{\mu h} \times 10^{-6} \times C_{\mu\mu f} \times 10^{-12}}} = \frac{159}{\sqrt{L_{\mu h} C_{\mu\mu f}}}$$

For example, if the inductance of the tank circuit is 100 microhenrys and the capacitance is 100 micromicrofarads, the resonant frequency of the oscillator is

$$f_o = \frac{159}{\sqrt{100 \times 100}} = 1.59 \text{ megacycles.}$$

SHUNT-FED HARTLEY OSCILLATOR

The shunt-fed parallel-fed Hartley oscillator (fig. 8-3, D) differs from the series-fed Hartley oscillator in that direct current does not flow through the tank circuit. Figure 8-3, D indicates that the B-supply is connected in shunt with the triode plate and the portion of the tank circuit that includes L1. The d-c component of plate current is kept out of the L1 tank circuit by the blocking capacitor, C2, and the a-c component is kept out of the plate power supply by the radio-frequency choke coil.

An advantage of shunt feed is that the high-voltage B-supply is isolated from the tuned circuit.

COLPITTS OSCILLATOR

The Colpitts oscillator shown in figure 8-4 is similar to the shunt-fed Hartley oscillator with the exception that the Colpitts oscillator

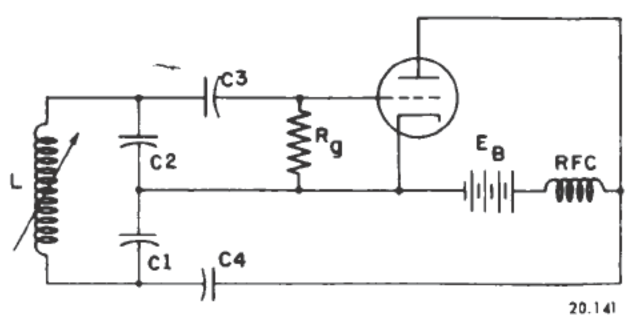


Figure 8-4.—Colpitts oscillator circuit.

uses a split tank capacitor as a part of the feedback circuit instead of a split tank inductor. The frequency of the oscillations is determined by the values of L, C1, and C2; and the grid excitation voltage appears across C2 instead of across L2 as in the shunt-fed Hartley oscillator. C3 and C4 perform the same function in this circuit as C1 and C2 in the shunt-fed Hartley circuit.

A simplified vector analysis of the Hartley and Colpitts oscillators is illustrated in figure 8-5. In the Hartley oscillator (fig. 8-5, A) C2 couples the triode plate to the lower end of L1, placing e_p across L1 and across the series combination of L2 and CT. The L2CT branch is predominately capacitive (X_c is greater than X_L). The current i_2 in this branch leads e_p by 90 degrees. The grid excitation voltage e_g is

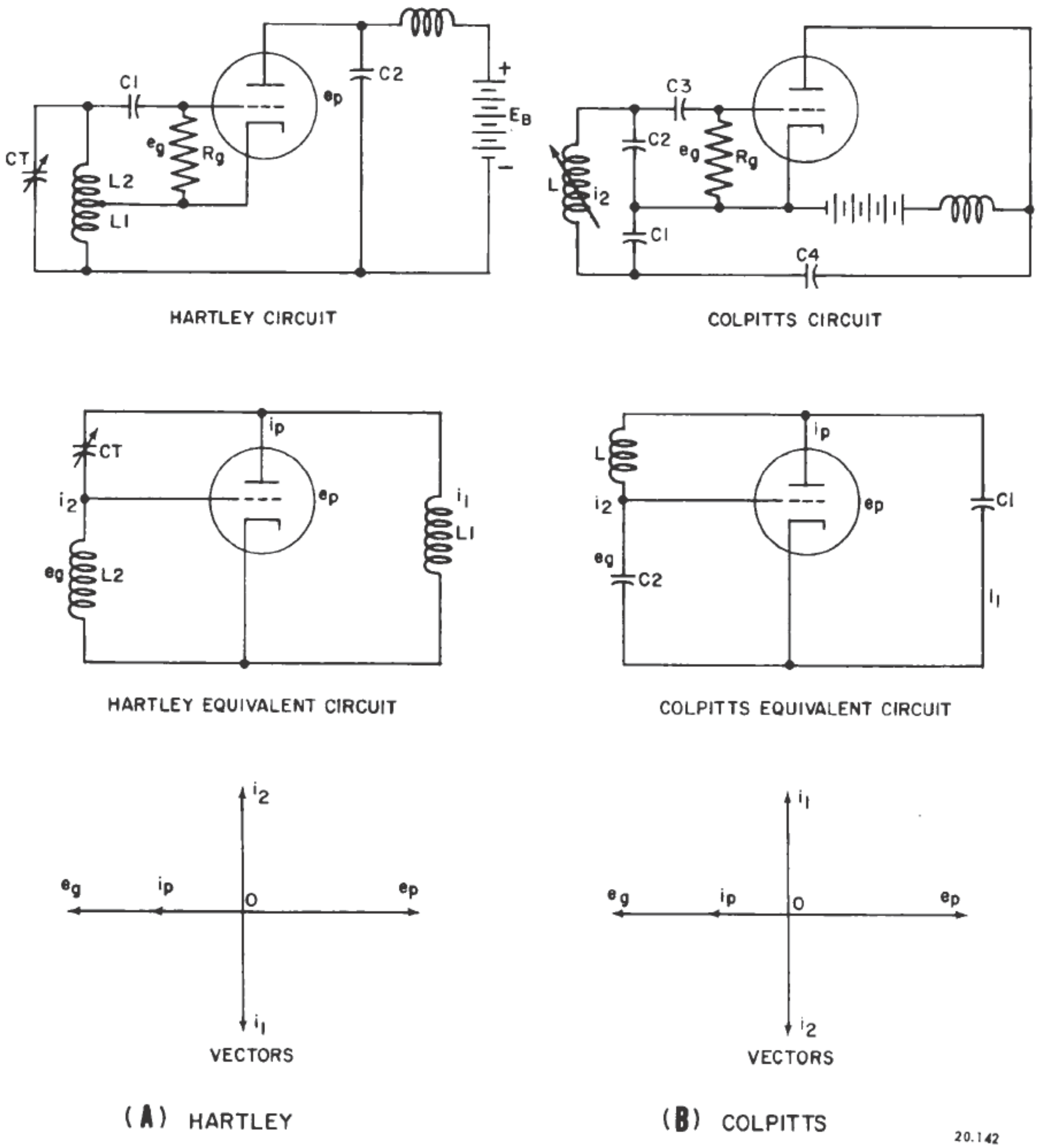


Figure 8-5.—Vector analysis of Hartley and Colpitts oscillators.

across L2 and this voltage leads the current i_2 by 90 degrees. Thus e_p across the plate load impedance is 180° out of phase with the grid excitation voltage e_g.

In the Colpitts oscillator (fig. 8-5, B) C4 couples the plate to the lower end of C1 placing e_p across C1 and the series combination of C2 and L. The a-c component of current i_1 in C1 leads e_p by 90 degrees. The C2L branch is predominately inductive because X_L is greater than X_C in this branch. The current i_2 in this branch lags the voltage e_p across the LC2 branch by 90 degrees. However the grid voltage e_g is across C2 and lags the current i_2 by 90 degrees. Again, the voltage e_p across the plate load impedance is 180° out of phase with e_g.

ULTRA-AUDION OSCILLATOR

The ultra-audion oscillator (fig. 8-6), frequently employed at ultrahigh frequencies, is similar to the Colpitts oscillator. The grid-to-cathode and plate-to-cathode interelectrode capacitances that make the operation of the ultra-audion oscillator similar to that of the Colpitts oscillator are indicated by dotted lines in the figure. Parallel feed is employed, and the radio-frequency choke prevents the a-c component of the plate voltage from entering the B-supply. Capacitor C2 provides a low-reactance path for r-f current and blocks direct current from the tank.

The voltage drop across C_{gk} is appreciable at the frequency employed and provides the grid excitation. Bias voltage is developed by the flow of grid current through R. The total tank capacitance is made up of CT in parallel with the series combination of C1, C_{gk}, C_{pk}, and C2. Capacitors C1 and C2 are relatively large so that they will offer negligible reactance to r-f currents.

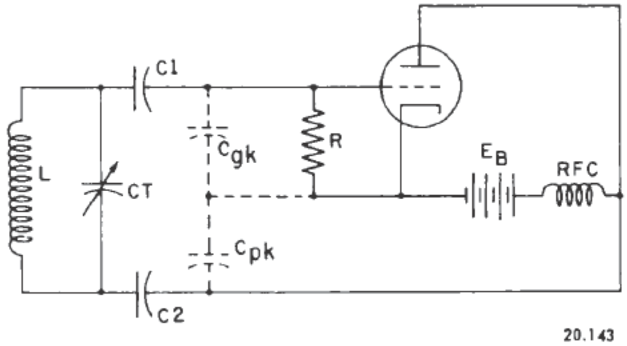


Figure 8-6.—Ultra-audion oscillator circuit.

TUNED-PLATE TUNED-GRID OSCILLATOR

The tuned-plate tuned-grid (t-p- t-g) oscillator utilizes a tuned circuit in both the plate and the grid circuits, as shown in figure 8-7, A. This type of oscillator may be employed in a wide range of frequencies from the low frequencies to ultrahigh frequencies. However, because of reduced feedback between plate and grid at low frequencies, the t-p- t-g oscillator is not so satisfactory at low frequencies as some of the circuits that have already been considered.

The feedback necessary to sustain oscillations is coupled from the plate circuit to the grid circuit by means of the interelectrode capacitance between the plate and the grid.

In the equivalent a-c circuit (fig. 8-7, B) both (parallel) tanks are below resonance a small amount so they appear as highly inductive coils L'_1 and L'_2.

The plate-grid capacitance C_{pg} is small and X_{cpg} correspondingly large; X_{cpg} is greater than $X_{L'1}$ and the left hand branch is capacitive. The right hand branch is inductive and at the operating frequency $X_C - X_{L'1} = X_{L'2}$.

Plate voltage, e_p (fig. 8-7, C) is the reference vector; e_p appears across both parallel

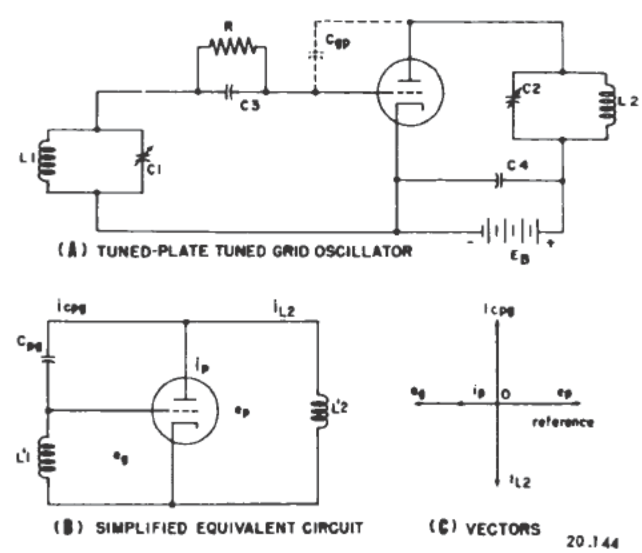


Figure 8-7.—Tuned-plate tuned-grid oscillator and equivalent circuits.

branches. The current i_{L2} lags e_p by 90° because it flows in the inductive circuit L'_2. The current i_{cpg} leads e_p by 90° because it flows in the left hand branch which contains more capacitive reactance than inductive reactance. The grid voltage e_g leads i_{cpg} by 90° because it appears across the inductive portion L'_1 of the left hand branch.

Notice i_p is in phase with e_g and 180° out of phase with e_p; the relation necessary to sustain oscillations. The tank having the highest Q (fig. 8-7,A) determines the oscillator frequency. If the plate tank is more heavily loaded than the grid tank, the grid tank will have the higher Q and determine the oscillator frequency.

PUSH-PULL OSCILLATOR

In order to obtain a power output larger than is possible with a single tube, an additional tube may be added in push-pull. As in push-pull amplifiers, the harmonic content of the output is reduced. The frequency stability of the push-pull oscillator is increased over that of single-ended types. The effect of interelectrode capacitance is reduced and the frequency range is extended. Push-pull oscillators are used generally at high and ultrahigh frequencies.

The push-pull oscillator (fig. 8-8) utilizes the interelectrode capacitance of each tube to feed back to the grid tank sufficient voltage to sustain oscillations.

When the oscillator is first energized it is improbable that the two tubes will be operating under exactly the same conditions. One tube therefore conducts more current than the other, and the voltages fed back to the two grids are unequal.

The push-pull oscillator is effectively two t-p t-g oscillators connected in push-pull. The action in V1 is similar to that previously described for the t-p t-g oscillator (fig. 8-7).

Like push-pull amplifiers the grid voltages are 180° out of phase. It is important that the center taps on L1 and L2 be at the exact center of the coils to avoid distortion in the output. Like push-pull amplifiers even order harmonics are cancelled in the output.

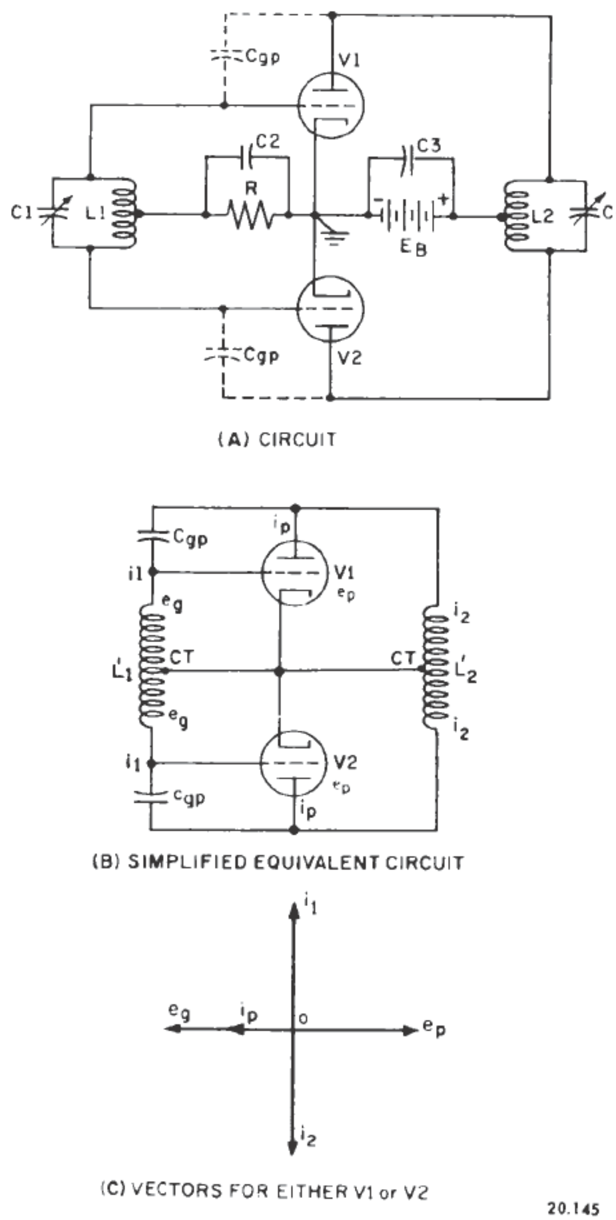


Figure 8-8.—Push-pull oscillator circuit.

ELECTRON-COUPLED OSCILLATOR

Three factors affect the stability of an oscillator: (1) changing values of C and L; (2) load changes; and (3) variations in the power supply voltage.

The electron-coupled oscillator (eco) is lightly loaded (especially when the plate tank is used as a frequency multiplier). Supply voltage variations cancel in the eco due to opposite effects on plate and screen. Thus the eco has greater frequency stability than other types of oscillators employing other types of coupling.

The electron-coupled oscillator shown in figure 8-9 combines the functions of an oscillator and a power amplifier. The control grid, tank circuit (CT, L1, and L2), cathode, and screen grid form a series-fed Hartley oscillator with the screen grid serving as the plate. Capacitor C2 places the screen at zero r-f potential, and like C3, bypasses the plate supply.

The output tuned circuit, C4L3, is in the plate circuit. The electron stream is the only coupling medium between the grid tank and the plate tank—hence the name ELECTRON-COUPLED OSCILLATOR. The two tank circuits are isolated by the screen grid, which is at r-f ground potential. This type of oscillator is relatively stable. Load variations have slight effect on the frequency of the oscillations.

An increase in screen voltage decreases the frequency of the oscillator, while an increase in plate voltage increases the frequency. If the screen and plate voltages are derived from the same power supply and the supply voltage increases, the tendency of the plate to increase the frequency and the screen to decrease the frequency cancel each other and the frequency remains unchanged. Conversely, a decrease in supply voltage has a similar effect. The tendency of the plate to decrease the frequency and the screen to increase the frequency again cancel each other. Potentiometer, R2 is adjusted for optimum screen voltage after which no further adjustment should be necessary. Frequency stability is best when the ratio of the plate-to-screen voltage is about 3 to 1.

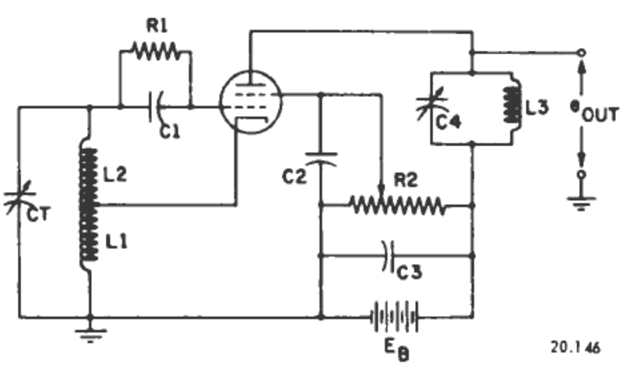


Figure 8-9.—Electron-coupled oscillator circuit.

NEGATIVE-RESISTANCE OSCILLATOR

A negative-resistance effect exists in a circuit if an increase in voltage across the circuit is accompanied by a decrease in current through it. This negative resistance effect was noted in the discussion of screen-grid tubes (ch. 1 of this training course). Thus, an increase in plate voltage of a tetrode is accompanied by a decrease in plate current, provided the plate voltage does not exceed that of the screen.

Another concept of negative resistance is applied to regenerative feedback in an oscillator. In supplying the input power to the oscillator, regenerative feedback introduces a negative resistance. In this sense negative resistance represents a source of power that supplies the losses in the oscillator circuit.

In the various feedback oscillator circuits that have been considered, enough power was available to overcome the circuit losses and to supply the necessary external load. If all of the circuit and load losses were represented by an equivalent resistance in the plate circuit, the amplifier action of the tube might be considered as presenting a negative resistance to the tuned circuit. The power supplied by the negative resistance must be equal to the power consumed by the positive resistance.

One of the oldest negative resistance oscillators is the dynatron oscillator. It depends on secondary emission from the plate for its operation. It uses a screen grid tube. Unlike other oscillators it does not depend on feedback to the grid to sustain oscillations.

The circuit of a dynatron oscillator using negative resistance is illustrated in figure 8-10.A. Plate current starts to flow when the tube warms up. At point b (fig. 8-10.B) plate current, i_p, starts to decrease because of secondary emission at the plate. The decrease in i_p from b to c is accompanied by a decrease in the voltage drop across the plate load impedance (parallel 1-c tank) and an increase in plate voltage e_p.

At point c (secondary emission ceases), i_p stops decreasing and starts to increase. The increasing voltage across the plate load impedance is accompanied by a decreasing e_p. At point b, i_p stops rising and starts to fall. The voltage across the plate load impedance again starts to fall as e_p rises and the cycle repeats.

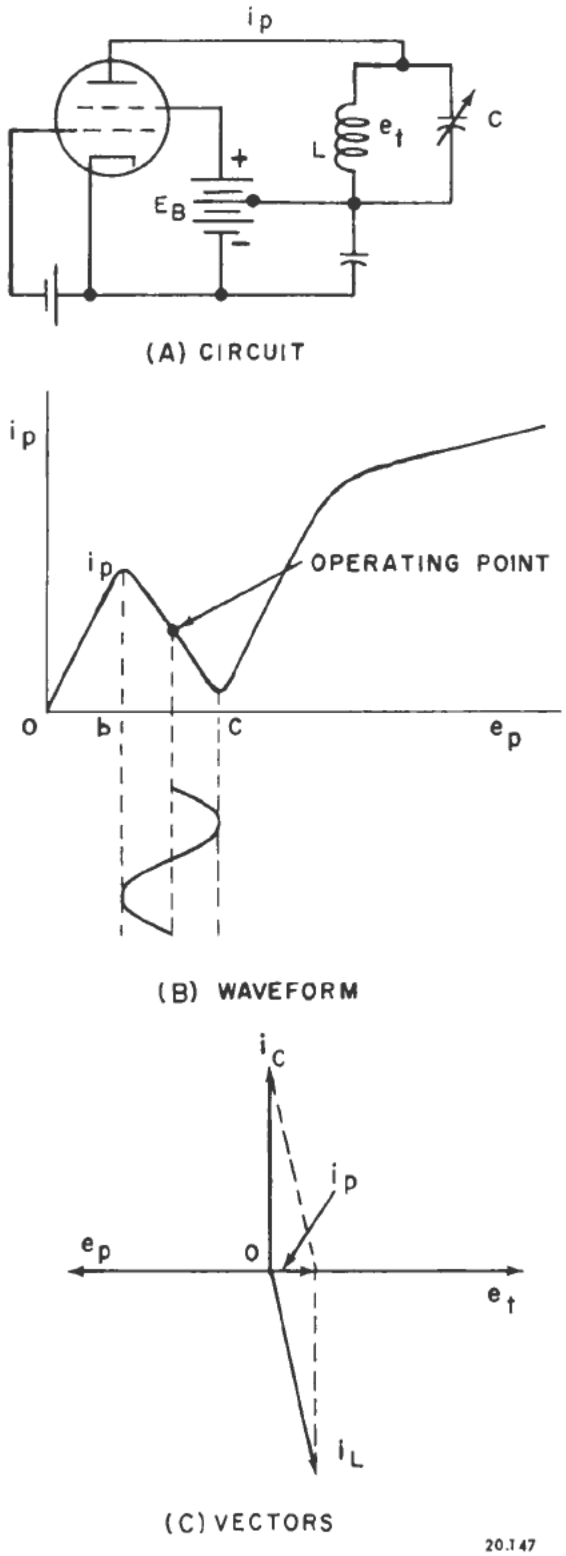


Figure 8-10.—Elementary dynatron oscillator circuit.

The vectors (fig. 8-10,C) indicate the phase relations between the a-c components of plate voltage e_p, plate current i_p, inductor L current i_L, and capacitor C current i_C. As in other oscillators previously described, the frequency of the oscillator depends on the values of L and C in the plate tank circuit.
The oscillator frequency is approximately equal to

$$f = \frac{1}{2\pi\sqrt{LC}}$$

Because the plate load impedance is in series with the triode plate and the B supply, the plate voltage will decrease as the voltage e_t (fig. 8-10), and current of the plate load impedance increase. This action causes e_p and i_p to be 180° out of phase, and also e_p and e_t to be 180° out of phase.

The current, i_L, in the inductive branch lags e_t by almost 90° (the losses in the tank are associated with the coil). Current i_c in the capacitive branch leads e_t by 90° (no losses assumed in the capacitive branch). The vector sum of i_L and i_c is the total input current flowing to and from the terminals of the plate load impedance. This current is the plate current (a-c component) i_p.

CRYSTAL-CONTROLLED OSCILLATOR

Some crystalline substances such as Rochelle salt, quartz, tourmaline, and even cane sugar, have the property of changing their shapes when an emf is impressed upon the crystal. The action is also reversible, that is, if the crystal is subjected to a mechanical strain, an emf will be produced across the surfaces of the crystal. A strain in one direction produces a certain polarity between the surfaces; if the direction of the strain is reversed, the polarity of the emf will also be reversed. This interrelation between mechanical and electrical stress in crystals is called the PIEZOELECTRIC EFFECT.

The magnitude of the response obtained from the crystal depends on the type of crystal used, the way it is cut, and the manner in which the emf is impressed. Rochelle salt is perhaps the most active, but it is also affected to a large extent by heating, aging, mechanical shock, and moisture. Quartz, while less active, is more rugged than Rochelle salt.

Another and more important advantage of the quartz crystal is its inherently higher Q. This results from the low damping, which in turn is due to the hardness of the crystal and its low internal friction when vibrating. Values of Q above 25,000 are possible; and if air damping is reduced sufficiently, Q's of the order of several hundred thousand are possible.

Since a high Q is necessary for frequency stability, quartz-crystal oscillators are widely used.

Quartz crystals used in oscillator circuits must be cut and ground to accurate dimensions. For example, the dimensions for a typical quartz crystal resonant at 1,000 kc would be approximately 1x1x0.1125 inch. All crystals, however, are not in the shape of a rectangular plate such as this. For use at the higher frequencies, some of the crystal elements are disk shaped, similar to a coin. For precise test work, crystals may be cut in the form of a flat ring.

Electrical contact with the crystal is made by a crystal holder consisting of two metal plates, between which the crystal is placed, and a spring device that places mechanical pressure on the plates. Electrical contact may also be made by soldering connecting wires to a metallic film deposited on the surface of the crystal.

Crystals are classified according to the way they are cut from the original quartz crystal. Figure 8-11 shows the approximate form of a raw crystal, a cross section of the crystal, and a section of the crystal in which three of the possible cuts are indicated.

The Z, or optical, axis is not important electrically because no piezoelectric effect is produced by the application of electrical stresses in this direction. The X axes drawn through the corners of the hexagon (fig. 8-11,B) are called the ELECTRICAL AXES, and the Y axes drawn perpendicular to the faces of the hexagon are called the MECHANICAL AXES. Designation of the axes as electrical or mechanial, aids in defining the type of cut. For example, in figure 8-11,C, the X-cut crystal is shown cut perpendicular to one of the X axes. During one-half of the r-f voltage cycle impressed across the flat surfaces of this X-cut crystal it will expand along the axis perpendicular to its flat surfaces, and during the other half cycle it will contract. Likewise, the Y-cut crystal is cut perpendicular to one of the Y axes. The AT-cut is made at approximately a 35° angle with the Z axis. Each of the other cuts such as the BT, CT, DT, and GT cuts, has distinctive characteristics.

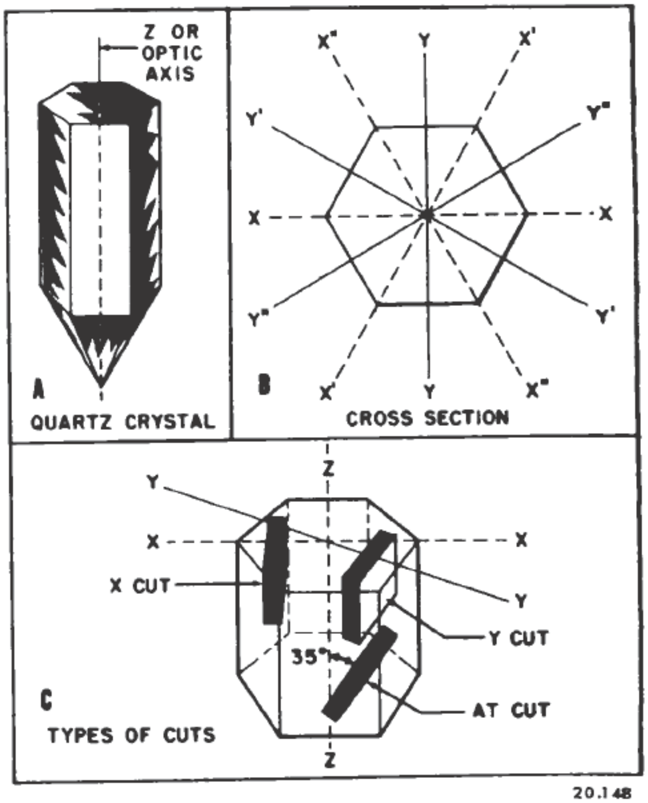


Figure 8-11.—Quartz crystal.

Temperature has different effects on the various types of cuts. The X-cut crystal has a negative temperature coefficient—that is, as the temperature increases, the frequency decreases. The Y-cut crystal has a positive temperature coefficient—as the temperature increases, the frequency increases. Both of these effects are undesirable, and therefore these two types of cuts have been replaced by others. The AT-cut shown in figure 8-11,C, has a temperature coefficient that is very nearly zero, that is, its frequency changes only slightly with changes in temperature. The GT-cut (not shown) has the lowest temperature coefficient of any of the cuts.

Where the frequency of the oscillator must be maintained within a few cycles of the assigned frequency, as in standard broadcast transmitters, the crystal is placed in a temperature-controlled chamber. Changes in frequency due to changes in temperature are thus maintained at a minimum value, consistent with the ability of the thermostats to keep the temperature constant.

When a crystal starts vibrating at its resonant frequency, it requires only a small force operating at the same frequency to obtain vibrations of a large amplitude. When an alternating voltage is applied to a crystal that has the same mechanical frequency as the applied voltage, it vibrates and only a small applied voltage is needed to keep it vibrating. In turn, the crystal generates a relatively large voltage at its resonant frequency.

If a crystal is placed between the grid and cathode of an electron tube, and a small amount of energy is fed back from the plate circuit to the crystal to keep it oscillating, the circuit will act as an oscillator. The natural frequency of a crystal is critical, and if the frequency fed back is slightly higher or lower than this value the crystal will stop vibrating. Thus, the frequency of the crystal-controlled oscillator must be the same as that of the natural frequency of the crystal.

A crystal-controlled oscillator employing a triode tube is shown in figure 8-12. The equivalent of crystal Y is shown in figure 8-12,A. In the equivalent crystal circuit, the inductance, L, represents the electrical equivalent of the crystal mass that is effective in causing mechanical vibration; R is the electrical equivalent of internal resistance due to friction; C2 is the capacity effect of the metal crystal holders; and C1 is the reciprocal of the crystal stiffness, that is, COMPLIANCE, which is the equivalent of capacitance in the electrical system.

Feedback takes place through the grid-to-plate capacitance, C_{gp}, within the electron tube, and the bias voltage is established across R1. Capacitor C3 has a small capacitance and is shunted across the crystal in order to obtain a fine adjustment of the operating frequency by changing the capacitance of the equivalent electrical circuit. The output voltage appears across R2.

Oscillations occur at the resonant frequency of the crystal, and the plate circuit is tuned to a slightly higher frequency by decreasing the capacitance of C5 so that the plate tank presents inductive reactance to the crystal grid circuit. Note the similarity to the t-p t-g oscillator in the equivalent oscillator circuit (fig. 8-12,C).

In adjusting a crystal oscillator, the factor of stable operation must be considered. In figure 8-12, a d-c milliameter is connected in series with the B+ lead to the plate tank circuit, and C5 is changed from a low to a high value of capacitance (tuned to a lower frequency). The plate current will slowly decrease to a MINIMUM at the exact resonant frequency of the crystal oscillator, as shown at point C in figure 8-13. At this point the output of the plate tank circuit is maximum (minimum direct current indicates maximum a-c output). Just to the right of point C the plate current suddenly increases to its maximum value, and oscillations cease.

The crystal frequency is slightly below the natural resonant frequency of the plate tank, and the tank, therefore, looks like an inductor. The feedback circuit is a series connection of the plate tank, grid crystal, and grid-to-plate capacitance of the triode. As long as the feedback is positive the oscillator operates at the crystal frequency. As the capacitance is increased beyond point C, the tank suddenly looks like a

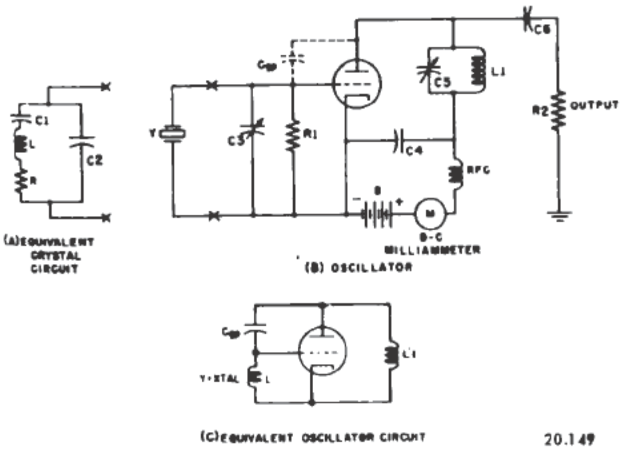


Figure 8-12.—Crystal controlled triode oscillators.

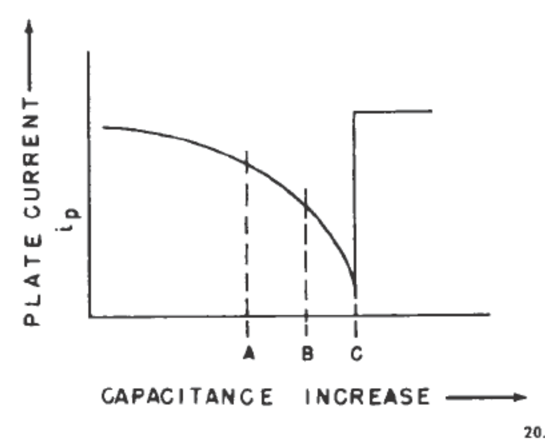


Figure 8-13.—Crystal-oscillator plate-current tuning curve.

capacitor instead of an inductor and the feedback becomes negative instead of positive. Oscillations therefore cease, and plate current rises rapidly.

In order to stabilize the operation, the capacitance is decreased (the resonant frequency is increased) to a value somewhere between A and C, for example, at B. The output is thereby reduced, but the operation is much more stable and slight changes in loading will not cause the oscillator to cease functioning.

The strength with which the crystal vibrates at its resonant frequency depends on the voltage fed back to it. The feedback is controlled by adjustment of the tuning of the plate tank circuit. If the feedback is too great, the vibrations may have sufficient magnitude to crack the crystal. The use of tetrodes and pentodes makes possible a reduction in feedback and thus overcomes this difficulty. Sufficient oscillations are still generated because these tubes are more sensitive than triodes and require less grid voltage for satisfactory operation. The circuit connections for tetrode or pentode crystal-controlled oscillators are modified to provide the correct amount of feedback and to supply the necessary screen-grid voltage.

RESISTANCE-CAPACITANCE OSCILLATORS

Resistance-capacitance oscillators depend for their operation on the charge and discharge of a capacitor in series with a resistor, whereas inductance-capacitance oscillators depend on the resonant interchange of energy between a capacitor and an inductor. A better understanding of the action of resistance-capacitance oscillators may be gained through a brief review of the growth and decay of current in an r-c series circuit, as treated in texts on basic electricity.

There are various types of resistance-capacitance oscillators such as sawtooth generators, multivibrators, blocking oscillators, and switching and counting circuits. The basic sawtooth generator and several of the typical multivibrator circuits are treated in this chapter. Other circuits will be treated as needed in advanced training courses.

SAWTOOTH GENERATOR

Voltages having sawtooth waveforms are widely used in television, radar, and many other electronic devices including test equipment. In each of these applications the sawtooth wave is used to sweep the electron beam across the fluorescent screen of a cathode-ray tube.

One of the simplest devices for developing this type of waveform is the gas-tube relaxation oscillator, shown in figure 8-14. Capacitor C is charged through resistor R until the potential across C reaches a value high enough to ionize the gas in the tube. Until this time the tube has a high impedance, but at the ionization potential its impedance drops to a low value and C discharges rapidly through it. When the voltage across C falls below the deionizing potential, the initial high impedance across the tube is reestablished and the capacitor stops discharging because the voltage across C is less than the value required to ionize the tube, the capacitor again charges.

For a given supply voltage, the frequency of the sawtooth voltage depends upon the R-C time constant and is varied by adjusting R.

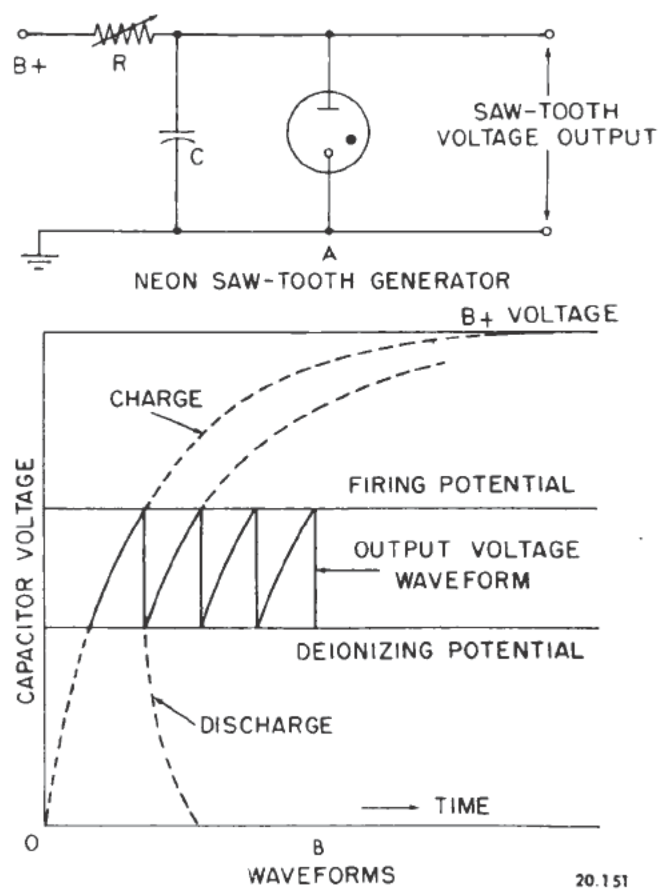


Figure 8-14.—Neon tube saw-tooth generator.

A consideration of figure 8-14,B, indicates that the output voltage varies between the de-ionizing potential and the firing potential of the gas tube. The full B-supply voltage is not applied across C because the firing potential is a much lower value and the difference appears across R. Likewise, C does not completely discharge because when the de-ionizing potential is reached, C stops discharging. The capacitor voltage follows a normal r-c charging curve between these two limits. The discharge follows a similar curve except that the discharge time is only a small fraction of the charge time because the resistance of the discharge path is only a small fraction of the resistance of the charge path and the curve is much steeper.

The output voltage curves for increased r-c time constant and for increased B potential are shown in figure 8-15. For example, in figure 8-15,A, increasing the resistance, R, increases the time necessary for C to charge to the ionizing potential, and the frequency is correspondingly decreased. Conversely, increasing the supply voltage (fig. 8-15,B) decreases the time necessary to charge C to the firing potential, and the frequency is correspondingly increased.

The thyratron, or gas-filled triode, is generally used to produce sawtooth waveforms and has certain advantages over the simple neon-tube sawtooth generator. For example, the thyratron is more stable, that is, changes in the applied voltage do not alter its characteristics so readily. The de-ionizing time also is reduced.

The thyratron operates much the same as the neon tube except that the IONIZING potential is controlled by the grid. The de-ioning potential is affected very little by the grid bias. The more negative the grid with respect to the cathode, the higher is the ionizing potential and the lower the frequency. A simple thyratron sawtooth generator circuit and the output waveforms with high and low values of grid potential are shown in figure 8-16. The approximate frequency of the sawtooth voltage is

$$f = \frac{1}{2.302RC\log_{10}\frac{E_b-E_2}{E_b-E_1}}$$

where f is the frequency in cycles per second, R the total charging resistance in megohms, C

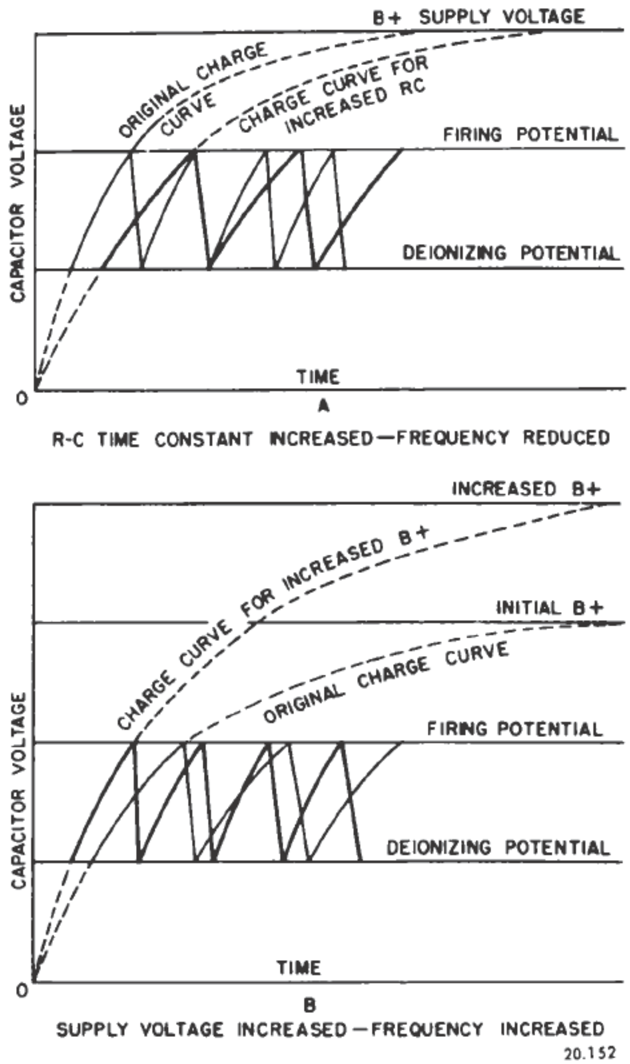


Figure 8-15.—Curves resulting from a variation in circuit constants.

the total capacitance in mocrofarads, E_b the plate supply, E_2 the de-ionization potential, and E_1 the ionization potential in volts.

In this circuit the B-supply voltage of course must be much larger than the ionizing potential of the tube. If alow potential is used, the output will have greater nonlinearity; in other words, a longer time will be required for capacitor C to become charged to the firing potential of the tube and therefore the full charge curve is used. On the other hand, if a high voltage is used, only the lower, more linear, portion of the curve is utilized before the ionization potential is reached.

A gas-tube relaxation oscillator does not produce oscillations that are stable in frequency. It can be synchronized with a constant frequency, however, by injecting a small voltage of the

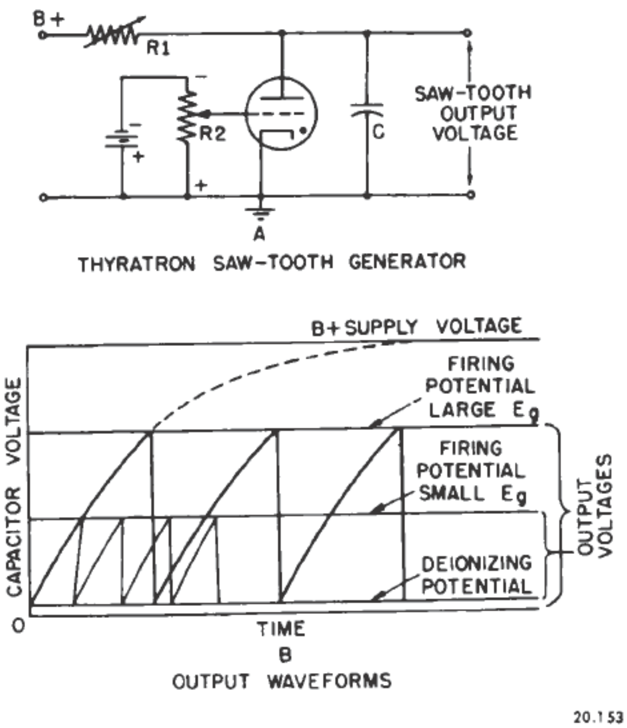


Figure 8-16.—Thyratron saw-tooth generator and output waveforms.

desired frequency into the grid circuit. A saw-tooth oscillator stabilized in this manner is shown in figure 8-17.

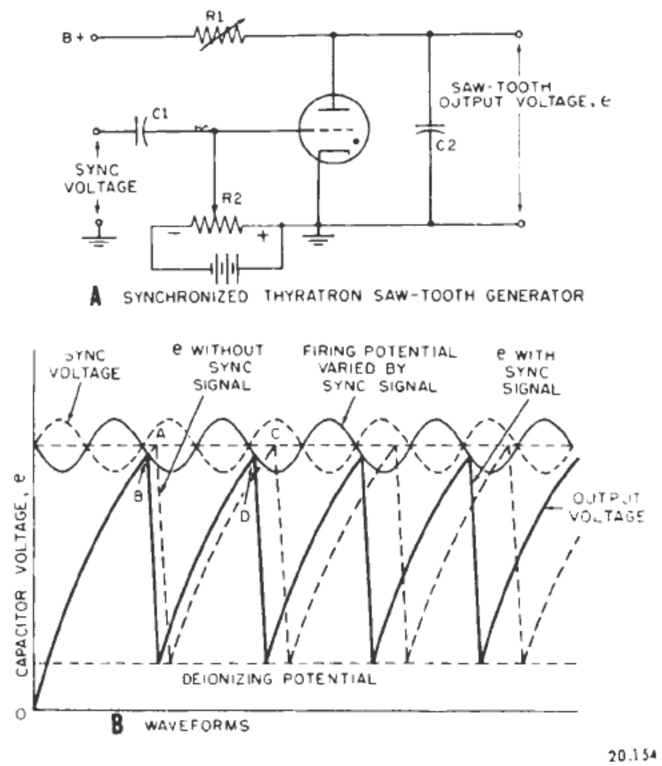


Figure 8-17.—Synchronized thyratron saw-tooth generator and waveforms.

Circuit operation may be explained as follows: the oscillator is adjusted until its natural frequency is somewhat lower than that of the synchronizing signal, as indicated by the dotted sawtooth curve (without sync signal) in figure 8-17,B. Without the sync signal the tube fires at points A, C, etc., but with the signal, the firing potential varies according to the instantaneous value of the grid potential. In other words, when the positive half of the sync signal is applied to the grid, the firing potential is reduced and the tube fires at points B, D, etc.; and, when the negative half is applied, the firing potential is increased. If the synchronizing voltage is applied, the time for each oscillation is reduced from AC to BD, and the oscillator is locked to the frequency of the sync voltage. The oscillator may also be locked to a multiple or submultiple of the sync voltage.

The vertical and horizontal sweep oscillators (not necessarily thyratrons) in television receivers are typical circuits controlled by sync pulses sent out by the transmitter.

MULTIVIBRATORS

A multivibrator is an electron-tube oscillator that utilizes two tubes or two sections of one tube to feed the output of one tube to the input of the other (and vice versa) by means of a resistance-capacitance coupling network. The output is essentially square wave and for a free-running multivibrator, the frequency is determined by the values of R and C. The frequency may be easily controlled by the application of an externally generated signal to the circuit. When this type of signal is applied, the circuit is referred to as a driven multivibrator.

Before the multivibrator cycle of operation is anaylzed, the following properties of electron-tube circuits are reviewed:

1. When the grid becomes less negative (more positive) with repsect to the cathode, the plate current increases, and vice versa.

2. An increase in current through the plate load resistor causes a greater IR drop across it, and therefore the plate voltage is reduced. Conversely, lower plate current results in higher plate voltage.

3. There is a 180° phase shift between the grid signal voltage and the a-c component of plate voltage.

4. The voltage cannot build up or decay instantaneously across a capacitor.

5. Current flow through a resistor is from the negative end to the positive end.

6. A capacitor requires a definite time to charge or discharge through a resistor. The time necessary for a capacitor to charge to 63 percent or to discharge to 37 percent of its final voltage is known as the TIME CONSTANT of the circuit. Its value in seconds is equal to the product of the resistance in ohms and the capacitance in farads.

Eccles-Jordan Oscillator

The basic multivibrator will be more easily understood if the action of the Eccles-Jordan trigger circuit is considered first. This type of circuit is used where an effect is desired that is electrically the same as opening or closing a switch. It is also used as a counting or scaling device.

In the strict sense of the word the Eccles-Jordan trigger circuit (fig. 8-18) is not an oscillator. It is instead a circuit that has two conditions of equilibrium. One condition is when V1 is conducting and V2 is cut off, and the other is when V2 is conducting and V1 is cut off. Both conditions are quiescent, that is, there is no change in current or in any of the potentials until the tube is triggered by an external signal. When the trigger pulse is applied, the nonconducting tube conducts and the conducting tube ceases to conduct. On the next pulse, the reverse operation occurs. This circuit is thus aptly called a FLIP-FLOP CIRCUIT.

The grids of V1 and V2 are connected to voltage divider networks. The voltage divider network for V1 includes R2, R4, R5, and E_C. The network for V2 includes R1, R3, R6, and E_C. The voltage drops across R5 and R6 are individually less than E_C and because these voltages subtract from E_C, the grids are always negative with respect to the cathodes.

The action of the Eccles-Jordan circuit is as follows:

1. Assume that the cathodes are heated and that B voltage is applied to both tubes. If both tubes and their corresponding elements were exactly alike, equal currents would flow through the plate circuits. It is not likely, however, that the two tubes and their circuit elements would be balanced so exactly as to permit this to occur. Actually one tube starts to conduct an instant before the other, or conducts more heavily than the other. Assume that V1 (fig. 8-18) conducts more current than V2.

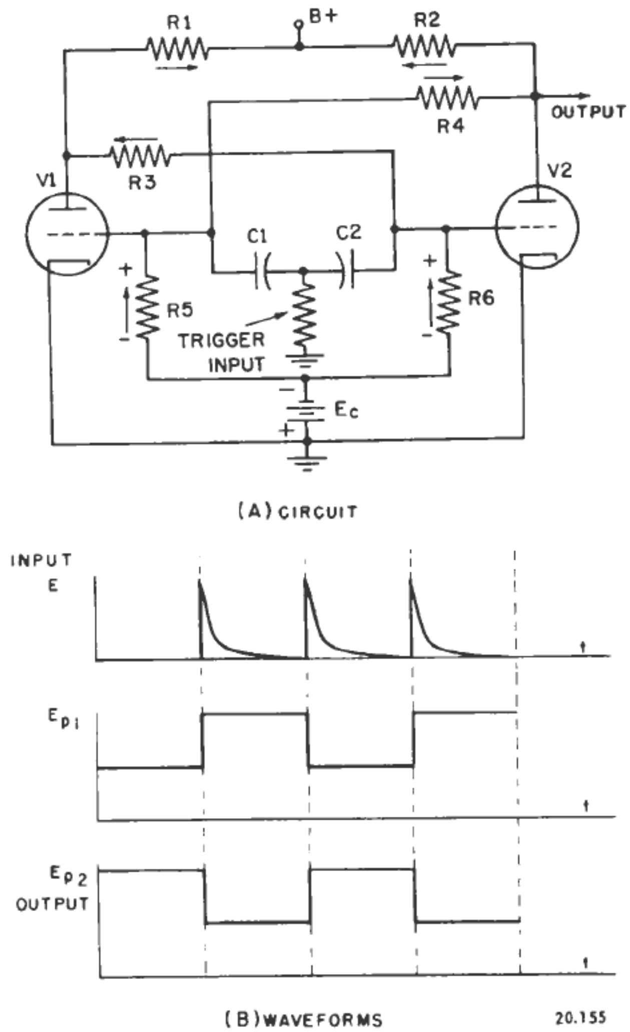


Figure 8-18.—Eccles-Jordan trigger (bistable or flip flop) circuit.

2. The voltage drop across R1 is greater than the drop across R2, and the voltage at the plate of V1 is lower than the voltage at the plate of V2.

3. The lower voltage on the plate of V1 reduces the voltage across R6 and increases the negative bias on V2. The current of V2 is further reduced.

4. Therefore the voltage at the plate of V2 is increased, and in turn the drop across R5 is

increased, reducing the negative bias on the grid of V1. Current in the plate circuit of V1 is further increased and the plate voltage is further decreased.

5. The action is cumulative, and very quickly a condition is reached when the plate current of V1 reaches a maximum and the plate current of V2 is cut off. This is one condition of stable equilibrium. During this quiescent period the drop across R5 is larger than the drop across R6.

6. Assume that a positive-going signal is applied simultaneously to the grids of both tubes at the trigger input. Since V1 is already passing a heavy current, the positive pulse on its grid has little effect on the flow of current through the tube. Tube V2, however, is cut off and the positive pulse on its grid, if of sufficient amplitude, removes the negative bias momentarily. Current then flows in the plate circuit of V2.

7. Plate voltage of V2 is reduced, and the reduced voltage across R5 makes the grid of V1 more negative.

8. Plate current in V1 is reduced and the plate voltage of V1 is increased.

9. The increased voltage of V1 increases the drop across R6, further reducing the bias on V2, and its plate current continues to increase, thus applying more negative bias to V1.

10. Current in V1 quickly ceases, and at the same time current in V2 reaches saturation.

If negative-going pulses had been used, the conducting tube would have been the first one affected. Its plate current would have been decreased, with the same end result (V1 cut off and V2 conducting). One alternation is thus completed for each pulse, and two pulses are necessary to complete a full cycle. It is possible however to bring about a complete cycle of operation with a single trigger pulse by making certain circuit changes.

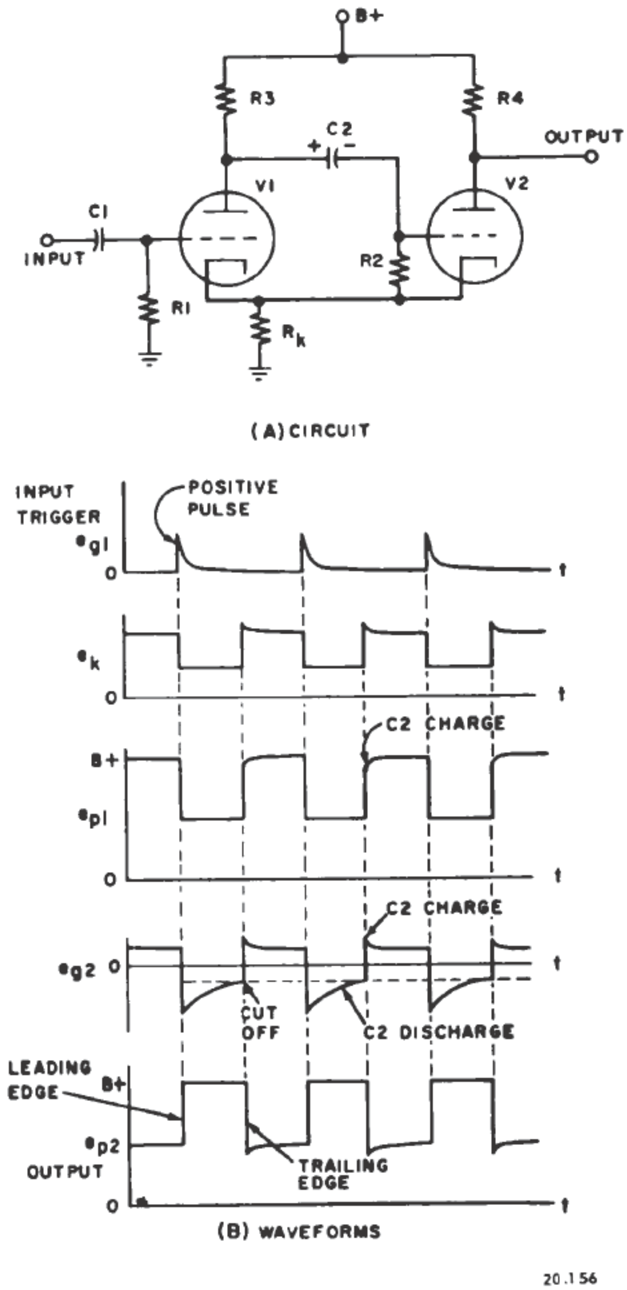


Figure 8-19.—One-shot (mono stable) multivibrator and waveforms.

One-Shot Multivibrator

The one-shot, or monostable, multivibrator (fig. 8-19) is a modification of the previously described bistable trigger circuit. It is essentially a two-stage r-c coupled amplifier (fig. 8-19,A). In its normal or balanced state one tube is cut off and the other is conducting. This condition is established by the biasing arrangement for the tubes. When a pulse is applied, the conducting tube is suddenly cut off and the other tube immediately conducts. After a certain time the tubes revert to their original condition of balance and the circuit is ready to respond to another input pulse.

The grid of V2 is returned to its cathode and no current normally flows through R2, therefore the V2 grid bias is normally zero. With zero bias, V2 conducts and plate current flows through cathode resistor R_k. The resulting voltage drop across R_k biases V1 to cut off. This condition is normal when there is no trigger input. It should be noted that when V2 is not conducting V1 cannot be cut off by the self bias developed across R_k.

A positive pulse, e_{g1} fig. 8-19, B, applied to the V1 grid via C1 sufficient to raise the V1 grid above cut off will cause V1 to conduct. The resulting drop in V1 plate voltage is applied through C2 as a negative going signal that will quickly cut off V2. The accompanying rise in plate voltage constitutes the leading edge of the output waveform.

Capacitor C2 discharges through R2 toward the lowered value of plate voltage of V1. The drop across R2 holds V2 cut off for a certain length of time depending on the time constant R2C2 and the cutoff bias of V2.

V1 continues to conduct after the input trigger is removed because as previously stated the drop across R_k is insufficient to cut off V1.

The circuit remains with V1 conducting and V2 cut off while C2 discharges toward the lowered value of V1 plate voltage until the V2 grid voltage rises above cutoff.

Then V2 conducts and the drop in V2 plate voltage is the trailing edge of the output waveform. The accompanying increase in voltage across R_k cuts off V1. The V1 plate voltage immediately rises and the positive-going signal applied to the V2 grid via C2 causes a further rise in V2 plate current as V2 returns to its original conducting state. The action is very rapid so that V2 conducts at practically the same time that V1 is cut off. The circuit is now in its original state of balance.

Free-Running Multivibrator

The basic free-running multivibrator circuit is shown in figure 8-20. Such a vibrator circuit is simply a two-stage r-c coupled amplifier with the output of the second state coupled through C1 to the input of the first stage and the output of the first stage coupled through C2 to the input of the second stage.

Because the voltage that is fed back in each case is of the proper polarity to reinforce the voltage on the grid of the tube receiving the feedback voltage, signals are reinforced and oscillation takes place.

The operation of the basic free-running multivibrator shown in figure 8-20, A, is described in the following paragraphs.

When the cathodes are heated and the plate potential is applied, both tubes begin to conduct. Initially the plate currents are nearly equal, but there is always a difference between them. The slight initial unbalance brings about a cumulative or regenerative switching action, which in this example is assumed to end with i_{p1} increased to a maximum value and i_{p2} reduced to zero. Although described as if it occurred slowly, this switching occurs with extreme rapidity—in a fraction of a microsecond in a well-designed multivibrator. This action is followed by a relatively long period in which the tubes are quiescent. During this interval one capacitor charges and the other discharges.

Assume that initially i_{p1} rises more rapidly than i_{p2}. Plate voltage e_{p1} falls (because of the increased drop in R3) and C2 discharges through R2, making the grid of V2 negative, thus reducing i_{p2}. Plate voltage e_{p2} rises (because of the decreased drop across R4) and C1 charges through R1, thus applying a positive bias to the grid of V1. The plate current of V1 rises to a maximum value, and V2 is cut off.

The charge path for C1 and the discharge path for C2 are shown in figure 8-20, B. The waveforms of plate current and plate and grid voltages are shown in figure 8-20, C.

The negative grid voltage applied to V2 results from the discharge of C2 through R2 and returns to zero as the capacitor discharge is completed. When the bias is reduced to the cutoff point, plate current i_{p2} begins to flow, and a second switching action takes place. This switching action is like the first except that i_{p2} is increasing and i_{p1} is decreasing. Plate voltage e_{p2} decreases (because of the increased drop across R4) and C1 discharges through R1, making the grid of V1 negative, thus reducing plate current i_{p1}. Plate voltage e_{p1} rises (because of the decreased drop across R3) and C2 charges through R2, making the grid of V2 positive. Thus the second switching action ends with V2 carrying maximum current and V1 cut off.

During the cycle of operation, current is maintained at a relatively steady value in one tube during the interval that the other tube is cut off. The action repeats continuously with first one tube and then the other conducting.

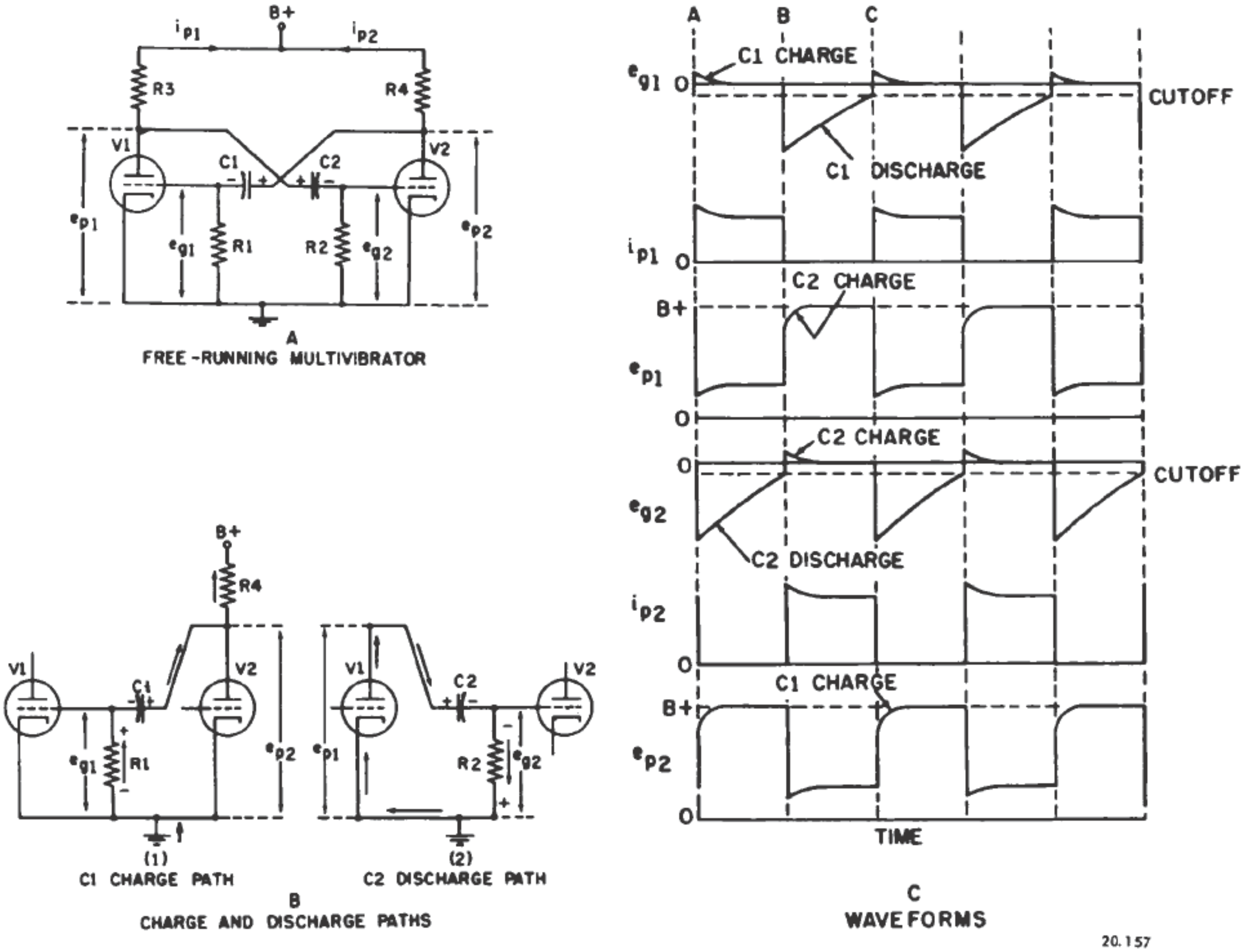


Figure 8-20.—Basic free-running multivibrator and waveforms.

The frequency of the oscillations depends on the time constants of the coupling networks, R_1C_1 and R_2C_2. For a uniform wave form, $R_1 = R_2$ and $C_1 = C_2$. The approximate frequency is expressed

$$f = \frac{1,000}{R_1C_1 + R_2C_2} = \frac{1,000}{2R_1C_1}$$

where f is in kilocycles, R_1 in ohms, and C1 in microfarads.

Synchronizing Multivibrators

Because free-running multivibrators have poor stability they are often synchronized with another frequency that forces the period of the multivibrator oscillation to be exactly the same as that of the synchronizing frequency. Such a multivibrator is said to be driven by the synchronizing voltage.

Sine waves or pulses are generally used for synchronizing purposes, although waveforms of almost any shape could be used. Synchronization by means of sine waves will be considered first. The synchronizing signal may be injected at the cathode or between the grid and cathode; however, only cathode injection is considered in this chapter.

Figure 8-21 shows a multivibrator with the synchronizing sine-wave voltage, e_k, applied to the cathode. A set of voltage curves is also shown in order to clarify the discussion.

The actual grid-to-cathode voltage of V1, which is the voltage controlling the flow of plate current, is the algebraic sum of e_{g1} and e_k. The source of the sinusoidal synchronizing voltage should have a low internal impedance

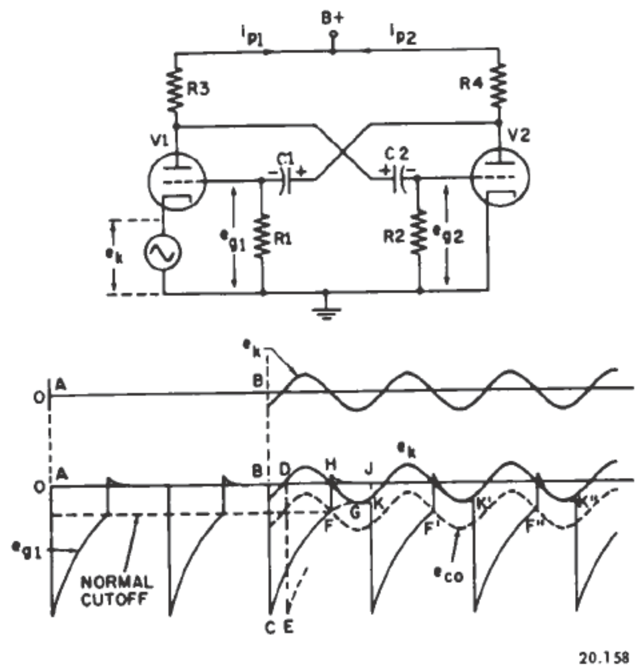


Figure 8-21.—Multivibrator with synchronizing voltage applied to the cathode.

so that the flow of i_{p1} through this source will not bring about an alteration in the shape of the wave.

The operation of the circuit may be explained thus:

If the multivibrator is properly balanced and running freely, the voltage curve of e_{g1} is that shown between time A and time B in figure 8-21. Because there is no synchronizing voltage on the cathode, its voltage remains constant at ground potential.

At time B the synchronizing voltage is applied and e_k begins to vary sinusoidally. Although e_{g1} is not affected by this variation, the grid-to-cathode potential now contains this sinusoidal voltage component, so that the effective cutoff voltage of the tube varies sinusoidally about its normal cutoff value in phase with the synchronizing voltage on the cathode. The cathode voltage curve, e_k, and the effective cutoff voltage curve, e_{co}, are shown with the e_{g1} curve in order to explain the synchronizing action.

The instant at which V1 conducts occurs when the e_{g1} curve crosses the e_{co} curve. At instant B, e_k starts to rise in a positive direction and i_{p1} is decreased. The positive-going voltage produced at the plate of V1 initiates the switching action via C2 to the grid of V2 and from the plate of V2 via C1 back to the grid of V1. Then e_{g1} drops along line BC instead of along DE, as it would in the free-running stage, and V1 is quickly cut off. C1 discharges along curve CFG; but since the e_{g1} curve intersects the e_{co} curve at F, the switching action by which V1 is made conducting and V2 is cut off takes place at F, instead of at G as it would in the free-running state.

The switching action drives the grid of V1 positive, but the resulting grid current quickly charges C1, and the grid returns to cathode potential. The grid voltage, e_{g1}, does not follow curve HJ, as it would in the free-running stage. Instead, the grid draws current because the synchronizing voltage causes the cathode to be negative with respect to ground at this time, and e_{g1} follows the cathode voltage along curve HK. When the cathode voltage begins to rise in a positive direction the plate current of V1 starts to decrease. At instant K the rise in voltage at the plate of V1, resulting from the decrease in i_{p1}, is large enough to drive V2 into conduction, and the tubes are rapidly switched.

The action of the sine-wave synchronizing voltage forces the time of one cycle to be shorter and the frequency to be higher than it would be without the synchronizing signal. Switching in one direction occurs at instants F, F', F'', etc., and switching in the other direction occurs at instants K, K', K'', etc. With the exception of the short transistion time, the period of the multivibrator is equal to the period of the synchronizing voltage. Thus, the multivibrator is said to be synchronized.

The synchronizing voltage can make the multivibrator operate above or below its natural frequency. However, if an attempt is made to pull the multivibrator to a frequency that is too high, it will synchronize at a frequency that is one-half or some other division of the synchronizing frequency, and FREQUENCY DIVISION may be obtained.

Multivibrators may be synchronized also by short positive or negative trigger pulses. Figure 8-22 shows the effect of positive pulses on the multivibrator grid-voltage waveform.

A positive pulse insufficient to drive the grid above cutoff applied to a nonconducting tube at instant A does not cause switching action. The only effect is to reduce the negative bias

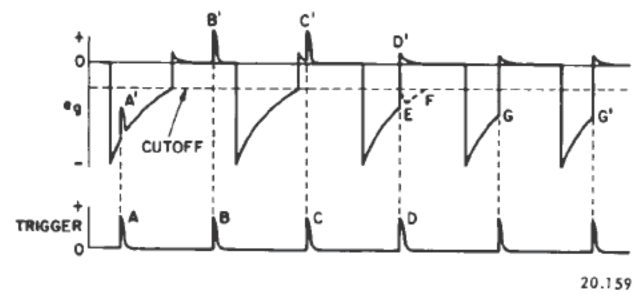


Figure 8-22.—Waveforms on the synchronized grid of a multivibrator driven by positive pulses.

slightly, as shown at A'. A positive pulse applied to a tube that is already conducting (points B or C) serves only to increase momentarily the grid voltage and thus to increase the plate current momentarily. It has no effect on the action of the multivibrator.

With the exception of the variations in the e_g waveform at times A', B', and C' the multivibrator is essentially free running. If applied at instant D, however, the positive trigger pulse is sufficient to overcome the negative voltage on the grid and drives the grid above cutoff. The cycle of the multivibrator is thereby shortened by an amount EF.

For proper synchronization, the natural period of the free-running multivibrator must be greater than the time interval between pulses. Under these circumstances the positive trigger pulses cause the switching action to occur earlier in the cycle than it would in the free-running state. Thus, the tube conducts at E, G, and G', whereas it would have conducted later in each instance had the pulses not been applied. Under these circumstances the frequency of the multivibrator is forced to become the same as the repetition frequency of the trigger pulses.

The multivibrator may be synchronized to a submultiple of the trigger frequency if both frequencies are such that every second, third, fourth, etc., synchronizing pulse occurs at the right time so that it will drive the grid voltage of the nonconducting tube above cutoff.

The multivibrator may be used as a sweep-frequency generator for cathode-ray tube applications—as in television, where magnetic deflection is commonly used. The horizontal and vertical oscillators in the receiver are triggered by pulses sent out from the transmitter so that they will be locked in step with similar oscillators at the transmitter. The multivibrator may be used as a source of square waves, as an electronic switch for various applications, or as a means of obtaining frequency division. It is often used to introduce a time delay between the operation of two circuits by using the leading edge of the square wave to trigger one circuit and the trailing edge to trigger another. The time delay can be controlled by varying the r-c time constants of the multivibrator circuit.

In television transmitters and in radar the action of the multivibrators is accurately timed by triggering them with pulses from a master oscillator circuit.

When used as an electronic switch, one multivibrator tube allows its associated amplifier to function normally while the second multivibrator tube holds its amplifier cut off, and vice versa. In radar, multivibrators are used principally as electronic switches to produce gate voltages that permit a part of a circuit to operate only during an accurately controlled time interval.

WIEN BRIDGE OSCILLATOR

An oscillator in which a frequency selective Wien-bridge circuit is used as the r-c feedback network is called a Wien-bridge oscillator. One extensively used circuit for this type of oscillator is shown in figure 8-23. In figure 8-23, A, the phase shifting element is shown as a bridge at the left. In figure 8-23, B, the circuit is redrawn to indicate more clearly the feedback paths.

Triode, V1, is the oscillator tube. Triode V2 acts as an amplifier and phase inverter. Thus, even without the bridge circuit the system oscillates because any signal that is fed to the V1 grid is amplified and inverted by both V1 and V2. The voltage feedback to the V1 grid is in phase with the initial input signal so that oscillations are set up and maintained. However, without the bridge the system amplifies voltages of a wide range of frequencies. The bridge circuit functions to eliminate feedback voltages of all frequencies except the single frequency desired at the output. The frequency of operation depends on the setting of variable capacitors C1 and C2, and the values of R1 and R2. Normally C1=C2 and R1=R2.

The bridge allows a voltage of only one frequency to be effective in the circuit because

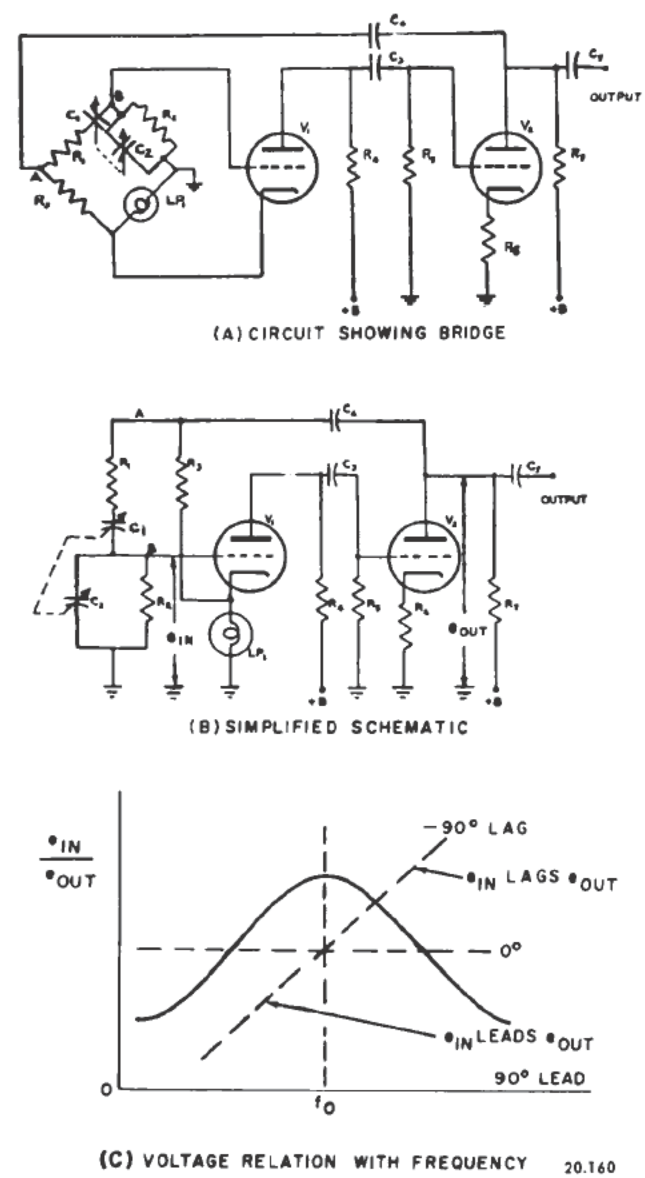


Figure 8-23.—Wien-bridge oscillator.

of the degeneration and phase shift provided by this circuit. Oscillations can take place only at the frequency, f_o, which permits the voltage across R2, the input signal to V1, to be in phase with the output voltage of V2 and for which the positive feedback voltage exceeds the negative feedback voltage.

Voltages of any other frequency cause a phase shift between the output of V2 and the input of V1 and are attenuated by the high degeneration of the bridge circuit so that the positive feedback voltage is not adequate to maintain oscillation at any frequency other than f_o.

A negative feedback voltage is provided via the voltage divider R3 and the lamp LP1. Because there is no phase shift across this divider and because the resistances are independent of frequency the amplitude of the negative feedback voltage is also independent of frequency.

The positive feedback voltage is provided by way of voltage divider R1C1-R2C2. If the frequency is very high the reactance of the capacitors is almost zero. In this case R2 is shunted by a very low reactance, making the V1 grid voltage almost zero.

On the other hand, if the frequency is reduced toward zero, the current that can flow through either C2 or R2 is reduced to almost zero by the very high reactance of C1. Therefore the V1 grid voltage falls to almost zero.

At some intermediate frequency the positive feedback voltage is a maximum as shown by figure 8-23,C. The curve is rather flat in the vicinity of f_o but the phase shift that occurs in the positive feedback circuit aids in permitting only a single frequency to be generated.

The voltage e_{IN}, across R2 is in phase with the output voltage, e_{OUT}, at frequency f_o. The intersection of the dashed slant and horizontal lines (fig. 8-23,C) indicates the in-phase condition. Below f_o, e_{IN} leads e_{OUT} and above f_o e_{IN} lags e_{OUT}. The frequency of the oscillator is

$$f_o = \frac{1}{2\pi\sqrt{(R1C1)(R2C2)}}$$

and when R1 = R2 = R and C1 = C2 = C the frequency is

$$f_o = \frac{1}{2\pi RC}$$

For example when R = 1000 ohms and C = 0.159 μf,

$$f_o = \frac{1}{2\pi \text{x} 1000 \text{x} 0.159 \text{x} 10^{-6}} = 1000 \text{ cycles}$$

At this frequency the positive feedback voltage on the V1 grid exceeds slightly the negative feedback voltage at the V1 cathode and e_{IN} is in phase with e_{OUT}.

When the oscillations are just starting to build up the resistance of the lamp, LP1 is smaller than the value required to balance the bridge, and e_{IN} is relatively large. As the oscillations build up, the lamp resistance increases. (The lamp filament is tungsten and its resistance increases with temperature.) This action brings the bridge nearer to a balanced condition and stabilizes the oscillations at f_o. The output waveform is sinusoidal and LP1 helps to prevent distortion due to overdriving V1.

The Wien bridge oscillator has an advantage of frequency stability and true sine wave output when used as an audio signal generator over a considerable range of frequencies (20-20,000 cycles).

Calculations are first given to show the magnitude and phase relation between e_{IN} and e_{OUT} for the conditions of f_o = 1000 cycles, R = 1000 ohms and C = 0.159 μf. The calculations are then repeated to show the lead of e_{IN} with respect to e_{OUT} when f = 500 cps, and the lag of e_{IN} with respect to e_{OUT} when f = 2000 cps, R and C remaining constant.

For the operating frequency of f_o = 1000 cps, the capacitive reactance of each capacitor in the bridge is

$$X_c = \frac{1}{2\pi f_o c} = \frac{1}{6.28 x 1000 x 0.159 x 10^{-6}}$$

$$= 1000 \angle -90^\circ \text{ ohms}$$

The total impedance of the bridge from terminal A to ground is

$$Z_t = R - jX_c \oplus \frac{R(-X_c)}{R - jX_c}$$

$$= 1000 - j1000 \oplus \left[\frac{(1000 \angle 0^\circ)\ (1000 \angle -90^\circ)}{1000 - j1000 = 1414 \angle -45^\circ} = 707 \angle -45^\circ = 500 - j500 \right]$$

$$= 1500 - j1500 = 2100 \angle 45^\circ \text{ ohms}$$

If the output voltage is assumed to be 2.1 volts $\angle 0^\circ$ the current at terminal A of this bridge will be

$$i_t = \frac{e}{Z_t} = \frac{2.1 \angle 0^\circ}{2100\ \ -45^\circ} = 0.001\ \ +45^\circ \text{ ampere}$$

The voltage e_{IN} across the parallel impedance, Z_b, of R2C2 is

$$e_{IN} = i_t Z_b$$

$$= (0.001 \angle +45^\circ)(707 \angle 45^\circ) = 0.707 \angle 0^\circ \text{ volt}$$

Thus e_{IN} is in phase with e_{OUT} at f_o = 1000 cps. The capacitive reactance of C at a frequency of f = 500 cps is

$$X_c = \frac{1}{2\pi fc}$$

$$= \frac{1}{6.28 x 500 x 0.159 x 10^{-6}}$$

$$= 2000 \angle -90^\circ$$

The total impedance of the bridge from terminal A to ground is

$$Z_t = R - jX_c \oplus \left[\frac{(R)(-jX_c)}{R - jX_c} = Z_b \right]$$

$$= 1000 - j2000 \oplus \left[\frac{(1000 \angle 0^\circ)(2000 \angle -90^\circ)}{1000 - j2000 = 2240 \angle -63.5^\circ} = 893 < -26.5^\circ = 800 - j398 \right]$$

$$= 1800 - j2398$$

$$= 3000 \angle -53^\circ \text{ ohms}$$

Assuming the same output voltage the total current at terminal A of the bridge is

$$i_t = \frac{e}{Z_t}$$

$$= \frac{2.1 \angle 0^\circ}{3000 \angle -53^\circ} = 0.0007 \angle +53^\circ \text{ ampere}$$

The input voltage, e_{IN}, across the parallel impedance Z_B of R2C2 is

$$e_{IN} = i_t Z_B$$

$$= (0.0007 \angle +53)(893 \angle -26.5^\circ)$$

$$= 0.615 \angle +26.5 \text{ volt}$$

Thus e_{IN} is reduced and leads e_{OUT} at f = 500 cps. The capacitive reactance of C at a frequency of f = 2000 cps is

$$X_c = \frac{1}{2\pi fc}$$

$$= \frac{1}{6.28 x 2000 x 0.159 x 10^{-6}}$$

$$= 500 \angle -90^\circ \text{ ohms.}$$

The total impedance of the bridge from terminal A to ground is

$$Z_t = R - jX_c \oplus \frac{(R)(-jX_c)}{R - jX_c} = Z_B$$

$$= 1000 - j500 \oplus \left[\frac{(1000\angle 0^\circ)(500\angle -90^\circ)}{1000 - j500 \quad 1130\angle -26.5^\circ} \right.$$

$$\left. = 443\angle -63.5^\circ = 197 - j395 \right]$$

$$= 1197 - j895$$

$$= 1495\angle -36.7^\circ$$

Again assuming the same output voltage of 2.1 volts $\angle 0^\circ$ the total current at terminal A of the bridge will be

$$i_t = \frac{e}{Z_t}$$

$$= \frac{2.1\angle 0^\circ}{1495\angle -36.7^\circ} = 0.0014\angle +36.7^\circ \text{ amp}$$

The input voltage, e_{IN}, across the parallel impedance Z_B of R2C2 is

$$e_{IN} = i_t Z_B$$

$$= (0.0014\angle +36.7^\circ)(443.\angle -63.5^\circ)$$

$$= 0.620\angle -26.8^\circ \text{ volt}$$

Thus e_{IN} is again reduced and now lags e_{OUT} at f = 2000 cps.

BLOCKING OSCILLATORS

A blocking oscillator periodically cuts itself off after one or more cycles of operation because of the gradual accumulation of charge on the grid capacitor with the grid side negative. When the grid swings positive with respect to the cathode (positive feedback) grid current flows and charges the grid capacitor with the grid side negative. On the following portion of the cycle the grid capacitor discharges through the grid lead resistor, and if the r-c time-constant is relatively long in relation to the period of the cycle, this action will bias the tube to cut off at or before the completion of one cycle (single swing type) or after several cycles (self-pulsing type). When the grid capacitor discharge has progressed sufficiently to raise the bias above cutoff the tube again conducts and the action repeats itself periodically. Thus the tube becomes an intermittent oscillator.

SINGLE SWING TYPE

A single swing blocking oscillator is shown in figure 8-24. The circuits (fig. 8-24,A) include a tuned grid having a high ratio of 1 to

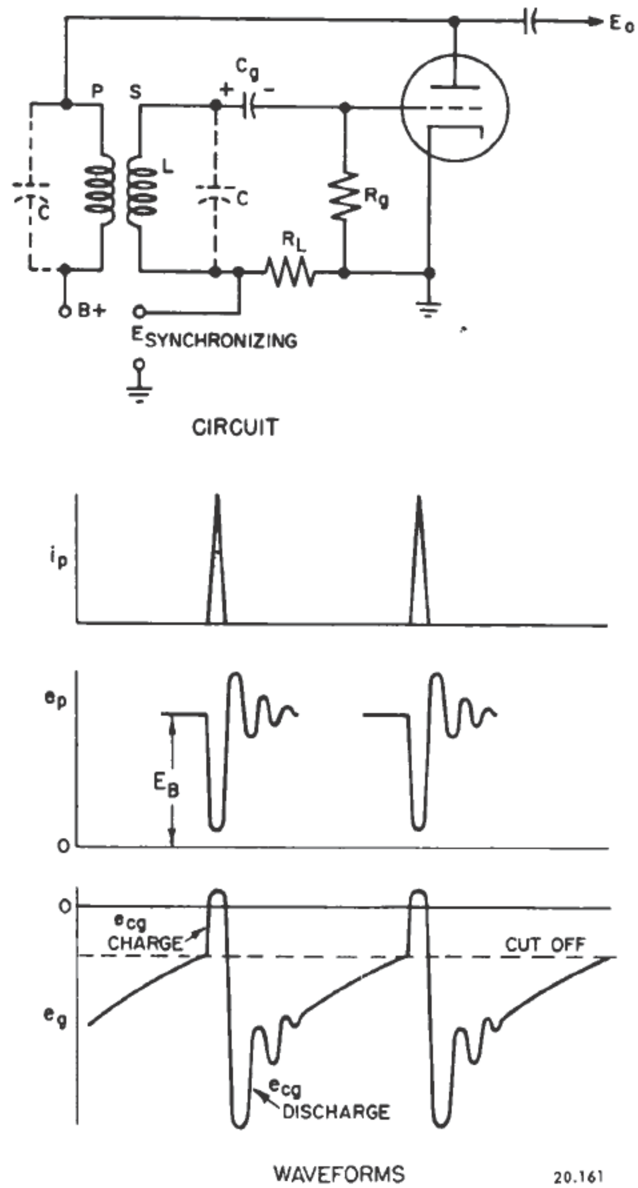


Figure 8-24.—Single swing blocking oscillator.

c. The capacitance is the distributed capacitance of the coil and the interelectrode capacitance of the triode. The coupling is high and there are many more turns on the secondary than the primary so that the grid drive is relatively large compared to that for normal oscillator operation. Grid leak resistor R_g is also relatively large and the R_gC_g time-constant is long compared with the period of the resonant tank circuit.

In the absence of bias, oscillation starts and quickly build up (fig. 8-24,B). The action is like that of a class C amplifier in which plate current flows during the time that plate voltage is decreasing and grid voltage is swinging positive. The oscillation starts to build up rapidly until the voltage drop across the plate coil is almost equal to the B supply voltage. The large grid driving voltage causes C_g to quickly charge up to this peak voltage during the portion of the first quarter cycle when the grid is positive. At the end of the first half cycle when the a-c voltage across the tank coils is zero the voltage across C_g biases the triode well beyond cutoff.

Because of energy losses in the coils, the positive peak of the next cycle of operation will not be sufficient to raise the grid above cutoff and succeeding oscillations will quickly die out as indicated by the damped oscillations in figure 8-24,B.

The period of the negative-going output pulse is approximately equal to half of the period of the resonant tank coils.

The cutoff period varies with the R_gC_g time constant. As C_g gradually discharges through R_g the grid voltage is positive-going and at cutoff, plate current starts to flow again and the cycle repeats.

A synchronizing voltage applied across R_L may be used to synchronize the repetition rate of the oscillator. The operation of the synchronizing voltage is similar to that previously described for the sawtooth oscillator and multivibrator.

SELF-PULSING TYPE

The Hartley oscillator (fig. 8-25,A) has a relatively large value of grid leak resistance, R_g and a long time constant, R_gC_g. The frequency of the r-f oscillations is determined by the l-c constants of the resonant tank. Capacitor C_g (fig. 8-25,B) charges quickly via the relatively low resistance of the conducting grid during the portion of the cycle when the grid is positive and grid current flows.

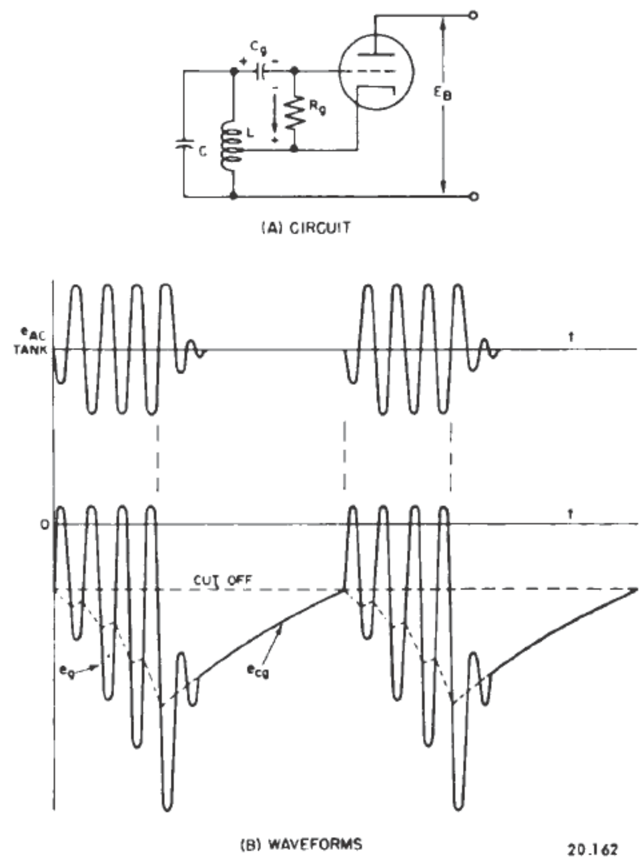


Figure 8-25.—Self pulsing Hartley oscillator.

Capacitor C_g discharges slowly through R_g during the portion of each cycle when the grid is negative. Thus during the time the tank circuit oscillations are being sustained by positive feedback from plate to grid, the grid capacitor C_g gradually accumulates a charge. Tank circuit oscillations cease when C_g lowers the grid bias below cut off and feedback is insufficient to sustain them.

The resting period is proportional to the discharge time of C_g. As C_g discharges, the grid bias is positive-going, with the decrease in voltage across C_g, and when the grid goes above cutoff, oscillations again build up in the tank.

The combined action results in periods of oscillation and periods of rest. Hence the operation is intermittent and is said to be self-pulsing.

Chapter 9

TRANSISTOR CIRCUITS

TRANSISTOR TRIODES

The majority of transistors in use today are triodes (three electrodes). They may be either NPN or PNP types as discussed in chapter 2 of this training course.

The schematic symbol for both types is similar but not identical. A heavy straight line represents the base and two slanting lines to the base represent the emitter and collector terminals. An arrowhead is placed on the emitter line and points toward the base for PNP transistors, and away from the base for NPN transistors.

CHARACTERISTICS

Transistor characteristics are analogous to those of a triode electron tube. Examples of the current curves are illustrated with corresponding test circuits in figure 9-1.

The PNP transistor in figure 9-1,A, has a grounded-base (common to the emitter input circuit and to the collector output circuit).

The PNP transistor in figure 9-1,B, has a grounded-emitter common to the input base circuit and the collector output circuit.

Collector current is measured along the X axis (fig. 9-1) and collector voltage is measured along the Y axis, in the characteristic curves. These curves represent the relation between collector current and collector voltage for various values of either emitter current i_e, or base current i_b.

In the grounded-base circuit (fig. 9-1,A) the associated curves indicate that the collector current i_c, and the emitter current i_e, are of the same order of magnitude. They also indicate that the collector current is practically independent of collector voltage over most of the range of the curves. Most important is the fact that i_c increases directly with i_e for a given collector voltage.

In the grounded-emitter circuit (fig. 9-1,B) the associated curves indicate that i_c is read on the milliamperes scale, while i_b is read on the microampere scale. Here too, the curves show that collector current is almost independent of collector voltage and primarily dependent upon i_b.

Current Gain

The current gain of a transistor is analogous to the amplification factor of a triode electron tube. The transistor, however, has either of two current gains depending upon the circuit connection.

Transistor current gain (fig. 9-1,A) is the ratio of the change in collector current to the corresponding change in emitter current for a constant collector voltage. In this circuit (grounded-base) current gain is designated α_{ce}. As mentioned in chapter 2 of this training course, the current gain for point contact transistors in the grounded-base circuit is approximately 2 or 3. In the PNP junction transistor grounded-base circuit of figure 9-1,A, the current gain is approximately 1.

Battery E_e, and potentiometer R_b, bias the input circuit in the forward (low resistance) direction. Battery E_c, and potentiometer R_c, bias the collector in the backward (high resistance) direction.

Moving the arm of R_b toward the positive terminal of E_e increases the forward bias current i_e, and this action will increase the collector current i_c. Moving the arm of R_c toward the negative terminal of E_c will increase the voltage across the collector-base circuit but this action will have little effect on collector current.

Transistor current gain (fig. 9-1,B) is the ratio of the change in collector current to the corresponding change in base current for a

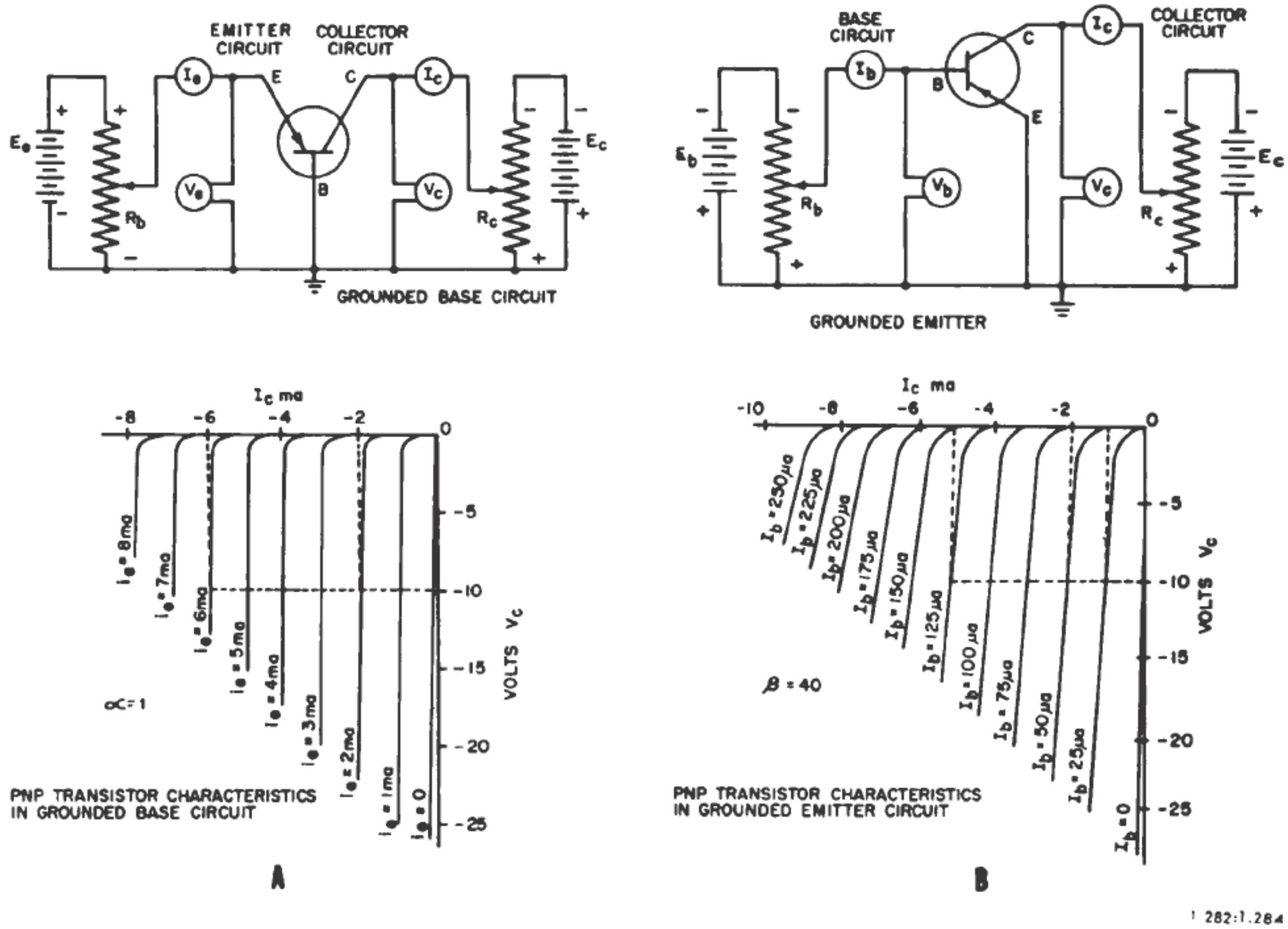


Figure 9-1.—Transistor characteristics.

constant collector voltage. In this circuit (grounded-emitter) transistor current gain is designated as β (pronounced beta).

With collector voltage adjusted to -10 volts and i_b = 25 μa (microamperes) collector current i_c = 1 ma or 1000 μa. Increasing i_b to 125 μa (with V_c constant) increases the collector current to approximately 5 ma or 5000 μa. Thus, an increase in input base current of 125-25 = 100 μa causes an increase in collector current of 5000-1000 = 4000 μa. The current gain is

$$\beta = \frac{\Delta i_c}{\Delta i_b} = \frac{4000}{100} = 40$$

Voltage and Power Gain

The voltage gain of a (grounded-base) transistor is the product of the current gain α_{ce}, and the ratio of collector resistance r_c, to emitter resistance r_e. Expressed as a formula

$$\text{V.G.} = \frac{\Delta i_c r_c}{\Delta i_e r_e} = \alpha_{ce} \frac{r_c}{r_e}$$

For example, a grounded-base point contact transistor having a current gain of $\frac{3 \text{ ma}}{1.5 \text{ ma}} = 2$ and a collector resistance of 15,000 ohms, and an emitter resistance of 300 ohms has a voltage gain of

$$\text{V.G.} = \frac{3}{1.5} \times \frac{15000}{300} = 2 \times 50 = 100.$$

The voltage gain of a (grounded-emitter) transistor is similar to that previously described except that current gain β is used in the formula instead of α_{ce}. For example, if the current gain of a grounded-emitter transistor is

$$\beta = \frac{\Delta i_c}{\Delta i_b} = \frac{4000}{100} = 40$$

(fig. 9-1,B), and the collector resistance and emitter resistance are approximately 1000 ohms each, the voltage gain will be

$$V.G. = \frac{4000x1000}{100x1000} = 40$$

The power gain of a transistor is the ratio of the change in output power to the corresponding change in input power. For a grounded-base transistor the power gain is

$$\frac{(\Delta i_c)^2 r_c}{(\Delta i_e)^2 r_e}$$

The power gain of the grounded-emitter transistor in the preceding example is

$$\frac{(4000)^2 x1000}{(100)^2 x1000} = 1600$$

TRANSISTOR AMPLIFIERS

Transistor triodes, like electron tube triodes, are amplifiers. The current, voltage, and power gain depend on the α or β values and the ratio of the output to input resistance. Although electron tube circuits are similar, they have widely differing characteristics. It is not possible to substitute a transistor for an electron tube without changing bias and plate supply voltages. For example, the transistor in some circuits is a current amplifier, and has a low input resistance; the corresponding electron tube amplifier has a high input impedance. The three elements in a triode electron tube and the corresponding elements in a PNP junction transistor are illustrated in figure 9-2.

The transistor has a collector terminal that is analogous to the triode plate, a base terminal that corresponds to the triode grid, and an emitter terminal that corresponds to the triode cathode.

BASIC CIRCUITS

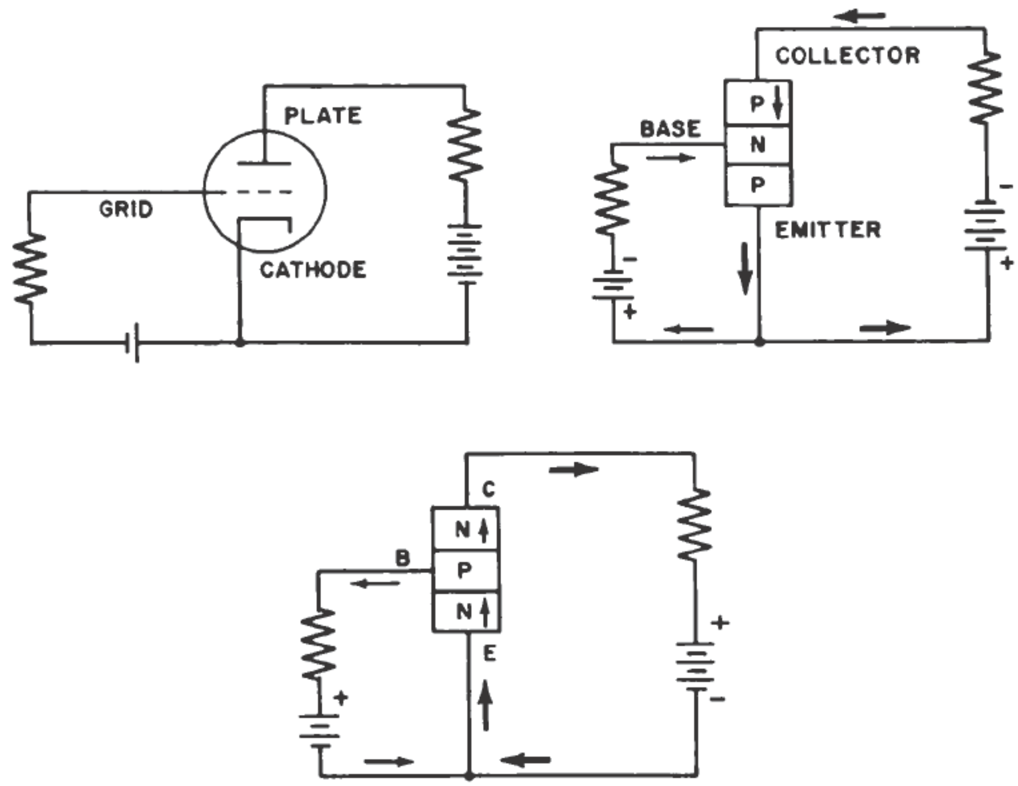
The transistor amplifier may be connected in any one of three basic circuits. These circuits are the (1) grounded-emitter, (2) grounded-base, and (3) grounded-collector. These three arrangements correspond respectively to three electron-tube basic circuits: (1) grounded cathode, (2) grounded-grid, and (3) grounded-plate amplifier. The significance of the expression "grounded" is that the element said to be grounded is, in reality, common to the input and output circuits, and does not necessarily have to be grounded to provide satisfactory operation.

Grounded-Emitter

The widely used grounded-cathode electron-tube amplifier and the corresponding grounded-emitter transistor amplifier are illustrated in figure 9-3. The transistor bias polarities are established for a PNP junction transistor in figure 9-3,B, and for an NPN junction transistor in figure 9-3,C.

The input signal to the triode (fig. 9-3,A) is developed across the grid resistor in series with the bias voltage between the grid and cathode. The output signal of the triode is developed between the plate and ground. The average grid voltage depends upon the magnitude of the grid bias supply voltage, and the average plate voltage depends upon the magnitude of the plate supply voltage.

The input signal to the transistor (fig. 9-3,B) is developed between the base and emitter in series with the bias voltage in this circuit. The bias polarity is in the forward, or low-resistance, direction of the base-emitter junction. The output signal of the transistor is developed between the collector terminal and ground. The average base voltage, as measured with respect to the emitter, depends upon the magnitude of the bias voltage in the base-emitter circuit. The magnitude of the bias current determines the mode (class) of operation of the transistor. The average collector voltage depends upon the magnitude of the collector voltage supply. In many instances, the magnitude of the collector voltage has only a small effect in determining the collector current. The collector current is primarily a function of the bias current in the input circuit.



1.280

Figure 9-2.—Corresponding elements in triode and transistor.

When a signal is applied to the input circuit of the transistor, the bias current varies about an average, or no-signal, value. This action causes the collector circuit current to vary through a much greater amplitude through the load impedance, which is connected in series with the collector voltage supply and the collector terminal of the transistor.

A phase shift of 180 degrees occurs between the input and output signals. Thus, a positive-going input signal (fig. 9-3,B) opposes the base-emitter bias and decreases the base-emitter current. This action decreases the collector current and voltage drop across the load impedance, resulting in an increase in collector-to-ground output voltage. Because the collector is negative with respect to ground, a positive-going input signal will result in a negative-going output signal.

A positive-going input signal (fig. 9-3,C) will cause an increase in base-emitter current with a corresponding increase in collector current and voltage drop across the load impedance. Because this drop subtracts from the collector voltage, the output signal is less positive and is therefore equivalent to a negative-going output signal. The latter action is similar to that occurring in the plate circuit of the triode (fig. 9-3,A) with a positive-going signal on the grid. The increase in plate current causes an increased voltage drop in the plate load impedance with a resulting decrease in the plate-to-ground voltage. A decrease in the positive plate voltage is equivalent to a negative-going output signal.

The grounded-emitter transistor has a medium input impedance. For junction transistors, this value may be of the order of 1 k-ohm and for point contact transistors, the value may be of the order of 35 k-ohms.

The output (load impedance) of the grounded-emitter transistor may be of the same order of magnitude as the input impedance or slightly higher.

Grounded-Base

The grounded-base transistor amplifier is analogous to the grounded-grid electron-tube amplifier. These circuits are illustrated in

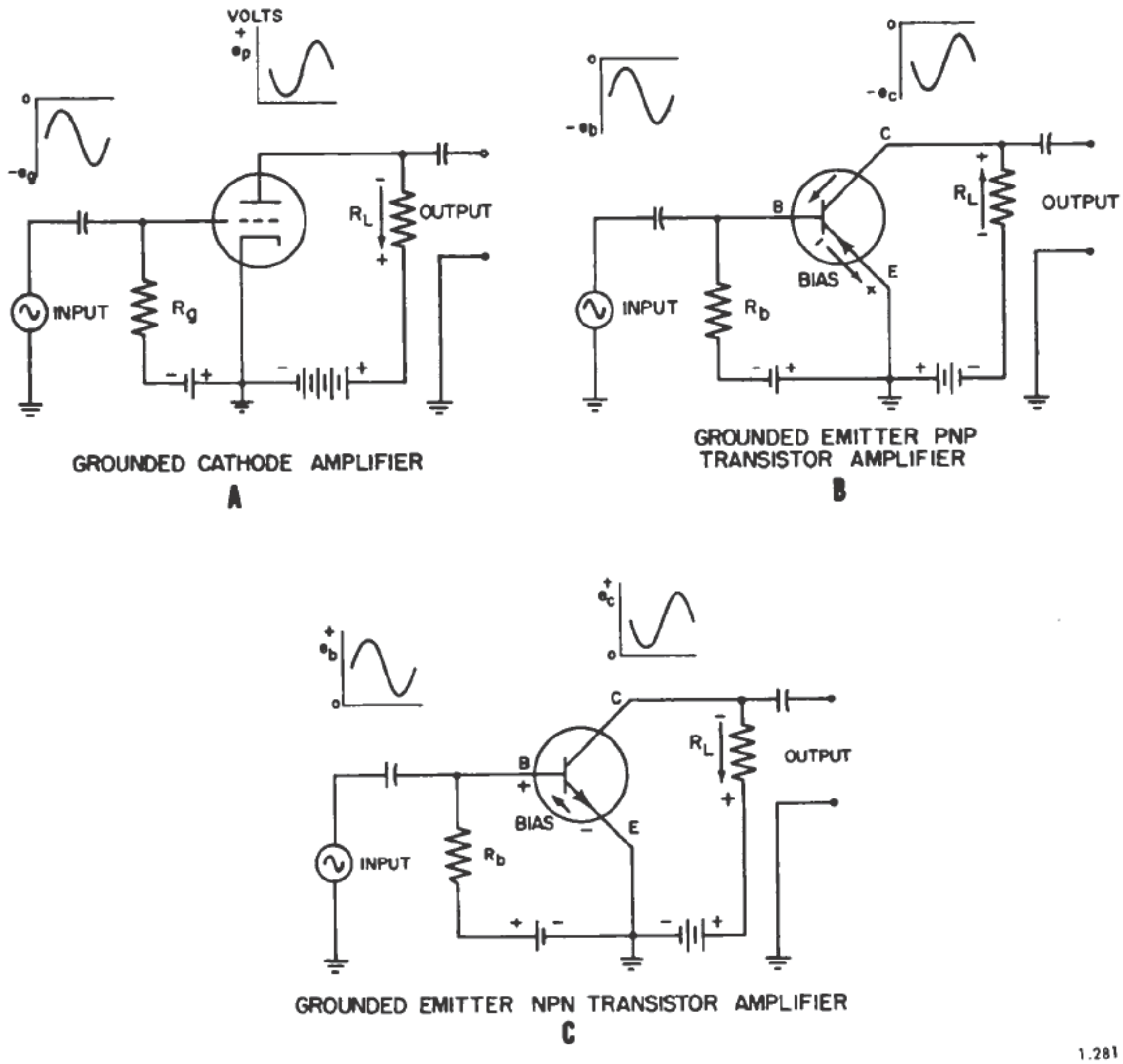


Figure 9-3.—Corresponding electron tube and transistor amplifiers.

simplified form in figure 9-4. The grounded-grid in the triode (fig. 9-4,A) is common to both the input and the output circuits. The grounded-base in the PNP junction transistor (fig. 9-4,B) is common to both input and output circuits.

The input signal to the triode is applied between the cathode and ground and varies the voltage across the cathode resistor in series with the bias voltage source in the grid cathode circuit. The output signal is developed between plate and ground as a result of the variations in plate current through the plate load impedance. Effectively, the input signal is developed between the cathode and the grid and the output signal is developed between the plate and the grid. Hence, the grounded-grid forms the common element between the input and output circuits. Plate current flows through the cathode resistor across which the input signal is developed. The action is degenerative and lowers the gain of the amplifier compared to that of a grounded-cathode amplifier.

The input signal to the transistor amplifier is applied between the emitter contact, E, and ground and appears across the resistor in series with the emitter base bias voltage. The average value of the emitter-to-base voltage depends upon the magnitude of the bias voltage in the input circuit. The output signal voltage is developed between the collector and grounded

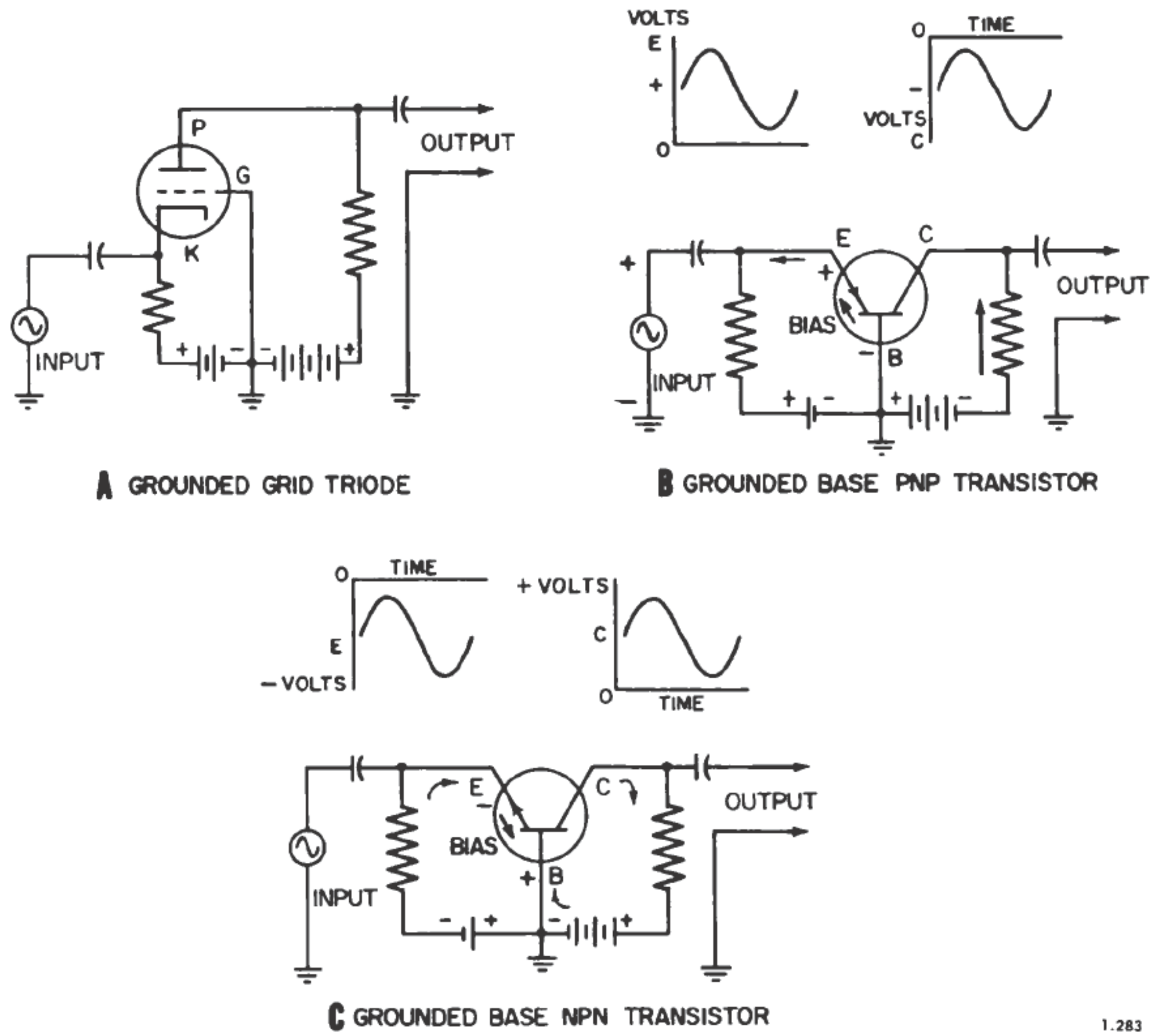


Figure 9-4.—Grounded-grid triode and corresponding grounded-base transistor amplifiers.

base. The average value of the collector-to-ground voltage depends upon the magnitude of the collector supply voltage. The base of the transistor is common to both input and output circuits and should not necessarily be grounded, provided the continuity of the common connection is maintained.

The input circuit bias polarity is in the forward, or low-resistance, direction of the emitter-base junction. As in the case of the grounded-emitter transistor, the magnitude of the bias current determines the mode of operation of the transistor amplifier, and the collector current is primarily a function of the bias current in the input circuit.

In contrast with the grounded-emitter amplifier, NO PHASE SHIFT OCCURS in the grounded-base amplifier between the input and the output signals. For example, a positive-going input signal (fig. 9-4,B) will aid the emitter-base-bias and increase the magnitude of the emitter current accordingly. This action will increase both the collector current and the voltage drop across the collector load impedance with a consequent decrease in collector-to-ground voltage. The collector is negative with respect to ground, hence the decrease in negative voltage to ground is equivalent to a positive-going output signal.

In the NPN transistor amplifier of figure 9-4,C, a positive-going input signal will oppose the input bias and reduce the emitter current accordingly. This action will reduce both the collector current and the voltage drop across the collector load impedance with a resulting increase in collector voltage. Because the collector is positive with respect to ground, the increase in positive voltage is equivalent to a positive-going signal. Thus a positive-going input signal will cause a positive-going output signal in both PNP and NPN grounded-base amplifiers.

The input impedance of a grounded-base transistor amplifier is of the order of 100 ohms or less. The output impedance is relatively high—that is, approximately 500 k-ohms for the junction transistor and 10 k-ohms for the point-contact transistor.

Grounded-Collector

The grounded-collector transistor amplifier corresponds to the grounded-plate electron-tube amplifier, or cathode follower. These circuits are illustrated in simplified form in figure 9-5. In the cathode follower the plate is at ground potential with respect to the signal component, and in the corresponding transistor circuits the collector is at ground potential with respects to the associated signal component. In the cathode follower the input signal is applied between the grid and the grounded side of the plate circuit, and the output signal is developed between the cathode and grounded side of the plate circuit. Thus the plate is common to the input and the output circuits.

In the grounded-collector transistor amplifier, the input signal is applied between the base and the grounded side of the collector circuit. The output signal is developed between the emitter and the grounded side of the collector circuit. Thus the collector is common to the input and the output circuits.

Bias current is supplied by the single-cell source in series with the base-emitter circuit, and collector voltage is obtained from a battery consisting of several cells in series between the collector and ground. The base-emitter circuits are biased in the forward, or low-resistance, direction, and the collector-base circuits are biased in the reverse, or high-resistance, direction. Thus the input impedance

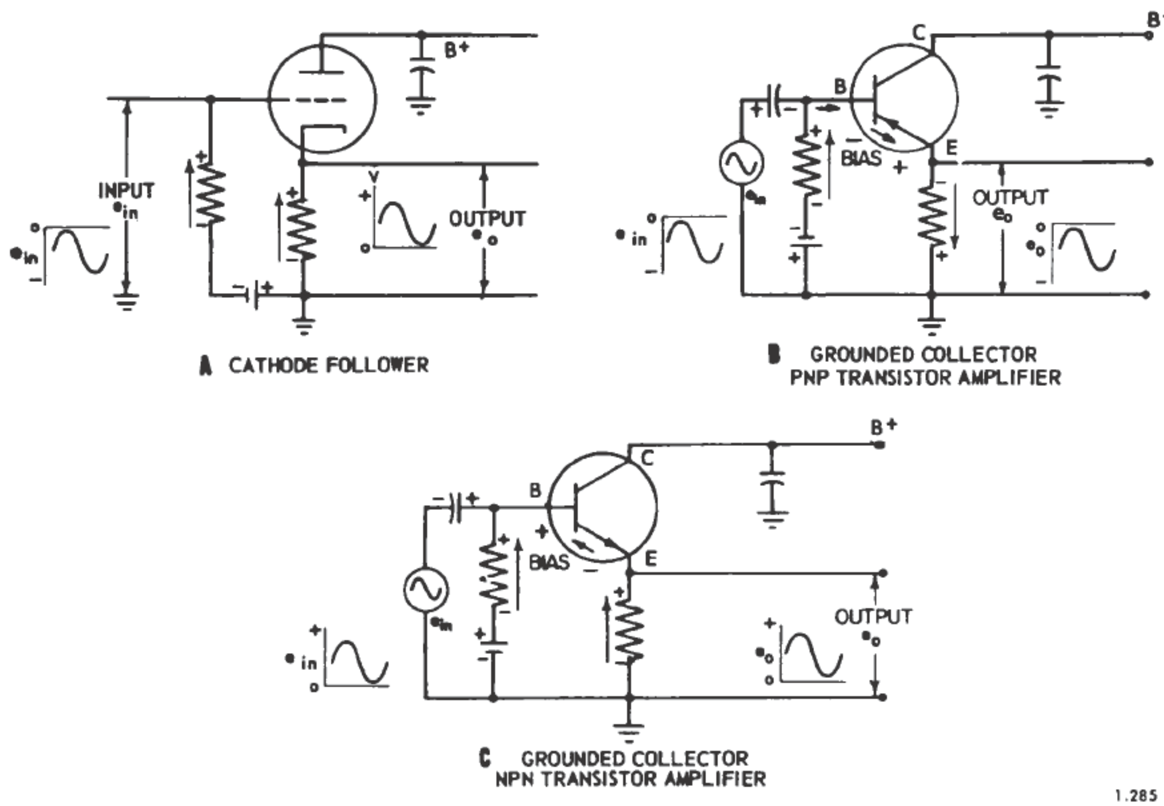


Figure 9-5.—Electron tube cathode follower and corresponding grounded collector transistor amplifiers.

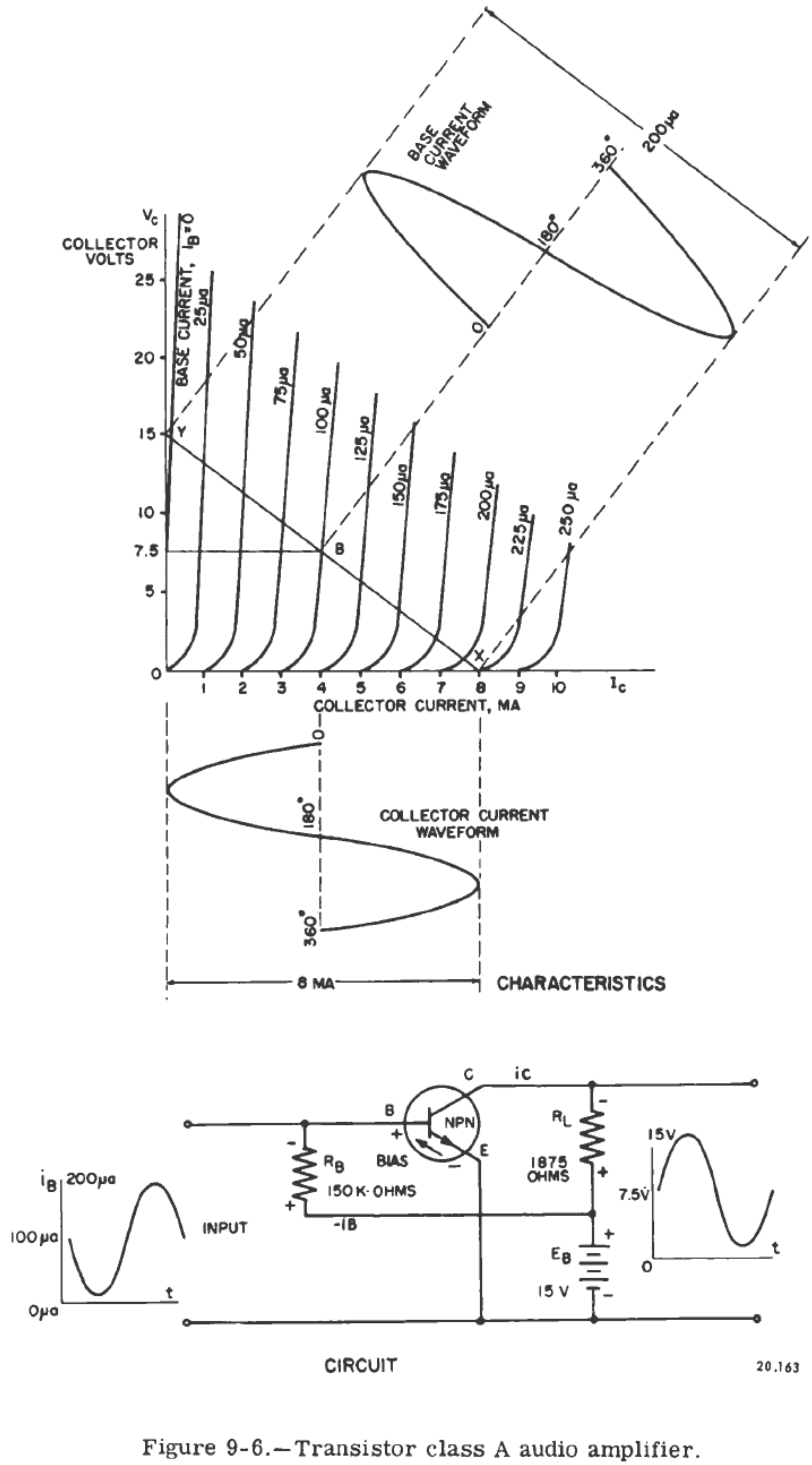


Figure 9-6.—Transistor class A audio amplifier.

of the grounded-collector transistor is relatively high, because the input circuit includes the collector-base circuit with the bias in the high-resistance direction. The output impedance is relatively low because the output circuit includes the base-emitter circuit with its bias in the forward, or low-resistance, direction.

The input impedance of the electron-tube cathode follower is high and the output impedance is low, hence the cathode follower may be used as an impedance changer or stepdown transformer. The output signal, e_o, is in the series circuit between the cathode and grid and opposes the input signal, e_{in}, in the same circuit. Hence, the net voltage acting between the grid and cathode is e_{in}-e_o, and e_o must always be less than e_{in} if the grid-to-cathode voltage is to be a finite value. Thus the voltage gain of the cathode follower is always less than 1.

Similarly, the output signal in the grounded-collector transistor opposes the input signal, and the net signal voltage acting between the base and emitter is e_{in}-e_o. For example, if the input signal causes the voltage occurring between the base and ground to swing 10 mv in a positive direction and the output voltage between emitter and ground to swing 4 mv in the same direction, the net voltage between base and emitter will be 10-4 or 6 mv. The output voltage, e_o, must always be less than e_{in} if the base-to-emitter voltage is to be a finite value. Thus the voltage gain of the grounded-collector transistor, like that of the cathode follower, is always less than 1.

The grounded-collector transistor provides a relatively large current amplification and also power amplification, depending upon the magnitude of the output impedance.

The input impedance of the grounded-collector transistor amplifier is a function of the load impedance. This action is in contrast to that in the electron-tube amplifier in which the input and output circuits are isolated and practically independent of each other. For example, the input impedance of a grounded-collector junction transistor amplifier may be of the order of 150 k-ohms when the load impedance is 10 k-ohms, but the input impedance may drop to 50 k-ohms or less when the load impedance is reduced to 1000 ohms.

Class A grounded-emitter transistor audio amplifiers are operated so that collector current flows during the entire input current cycle. If the correct operating point (load impedance) and input current are chosen, the output waveform will be essentially the same as the input in all respects, except for the amplitude.

CIRCUIT ANALYSIS

A simple method of determining the output voltage and current components of a transistor audio amplifier is by the use of the load line as illustrated in figure 9-6.

Load Line

The load line, XY, is a graph of the equation,

$$e_c = E_B - i_c R_L$$

where e_c is the instantaneous collector-to-emitter potential, E_B is the collector supply voltage and $i_c R_L$ is the voltage drop across the load resistor, R_L.

Point Y is established as $e_c = E_B$. This condition exists when the collector current is zero, and the full value of the collector supply voltage appears between the collector and emitter because there is no voltage drop across R_L. In this example, the load line intersects the Y axis at Y = 15 v. Thus, E_B = 15 v.

Point X is established as $i_c = \frac{E_B}{R_L}$. This condition exists when the base current is increased to the point where the effective internal resistance of the collector-emitter circuit is reduced to zero. For this condition the collector current would become a maximum value and the collector supply voltage would appear across the load resistance. In this example, the load line intersects the X axis at

$$i_c = \frac{E_B}{R_L} = \frac{15}{1875} = 0.008 \text{ amperes}$$

or 8 ma.

Load Resistance

The load resistance, R_L, is equal to the slope of the load line and may be determined as

$$\frac{E_B - E_C}{I_C}$$

where E_C is the collector-to-emitter voltage at the operating point B, and L_C is the collector current at this point.

In this example

$$R_L = \frac{15-7.5}{0.004} = 1875 \text{ ohms}$$

Bias

With the operating point at B (no-signal condition) the base-emitter bias voltage necessary to cause a base current of 100 μa may be obtained from the collector supply voltage by connecting a resistor of the proper magnitude between the base and the positive terminal of E_B. The magnitude of this bias resistor will be

$$R_B = \frac{E_B}{i_B} = \frac{15}{100\text{x}10^{-6}} = 150{,}000 \text{ ohms}$$

This value includes the emitter-to-base resistance, which is only a few hundred ohms and can be neglected in the calculation. Resistor R_B limits the base-emitter current in the forward direction to approximately 100 μa. However, the bias is not developed across R_B, but is developed across the junction as a result of the flow of no-signal current through the transistor between the emitter and the base. This action makes the emitter negative with respect to the base. This polarity indicates the forward direction for the base-emitter circuit of the NPN junction transistor, and is not to be confused with the polarity of the voltage drop across R_B, which is opposite to the proper bias polarity. Certain instabilities common to this bias method are described later.

The d-c operating point, B, indicates that the collector current will be 4 ma when the collector-to-emitter potential is 7.5 v, and the base current is 100 μa. The voltage drop across R_L for this condition is i_cR_L, or 0.004x1875=7.5 v, and the collector potential is E_B-i_cR_L, or 15-7.5 = 7.5 v. The no signal base current for this condition is

$$i_B = \frac{15}{0.150\text{x}10^6} = 100\text{x}10^{-6} \text{ amperes}$$

or 100 μa.

If the signal voltage applied across the input terminals has a peak magnitude slightly less than that of the voltage drop across the emitter-to-base circuit (E to B) and the polarity is opposite to that of the forward, or low-resistance, direction of electron flow, the base current will be reduced. In this example the input signal current is assumed to be of sine waveform and of a magnitude to vary the base current from 100 μa (no-signal value) to zero, to 100 μa for the first half cycle. For the second half cycle the base current is varied from 100 μa to 200 μa and back to 100 μa.

When the base current is reduced to almost zero, the collector current will be reduced to almost zero, and the collector voltage will increase to approximately 15 v. When the base current is increased to 200 μa, the collector current will increase to approximately 8 ma, and the collector voltage will be reduced to almost zero. Thus, the output voltage will vary sinusoidally about a no-signal value of 7.5 v and will have a peak-to-peak value of approximately 15 v.

Temperature Effects

The characteristics of transistors will change with temperature variations. Figure 9-7,A represents the constant base current curves of collector volts versus collector current for normal temperature operation. Figure 9-7,B, represents the curves for the same transistor with above normal temperature operation. The effects of increased temperature moves the curves to the right along the X axis, and may increase the spacing between them. This action shifts the operating point, B, on the load line toward point X, making the collector voltage too low. The normal collector voltage for class A operation, point B (fig. 9-7,A) is 7.5 v. The increase in temperature moves point B to a point indicated in figure 9-7,B, that corresponds to a collector voltage of 5.5 v.

Conversely, the effect of a decrease in temperature moves the curves to the left along the X axis and reduces the spacing between them (fig. 9-7,C). This action shifts the operating point along the load line toward point Y, making the collector voltage too high. The operating point now corresponds to a collector voltage of 9.5 v.

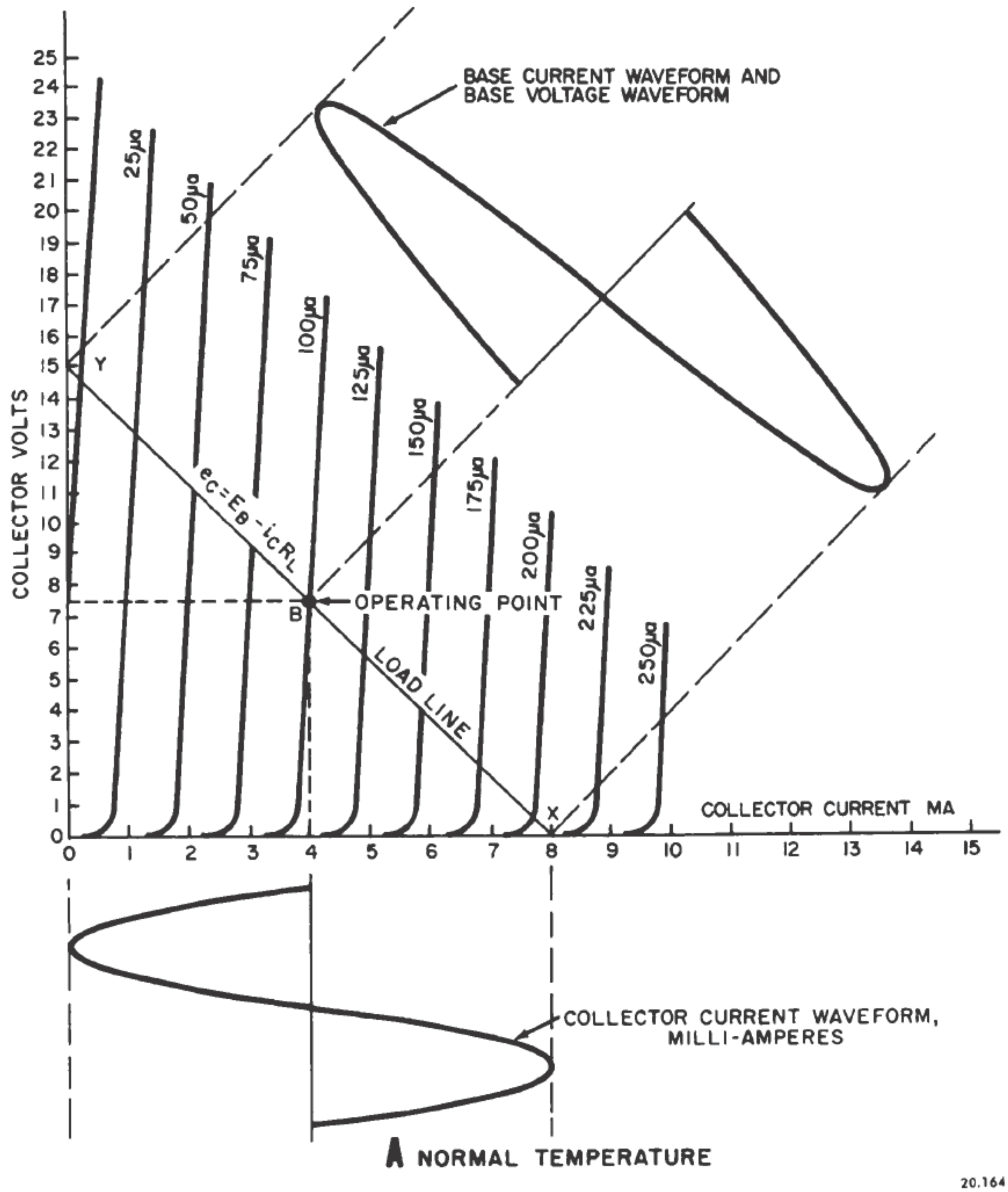


Figure 9-7.—Effect of temperature variations on NPN transistor characteristics.

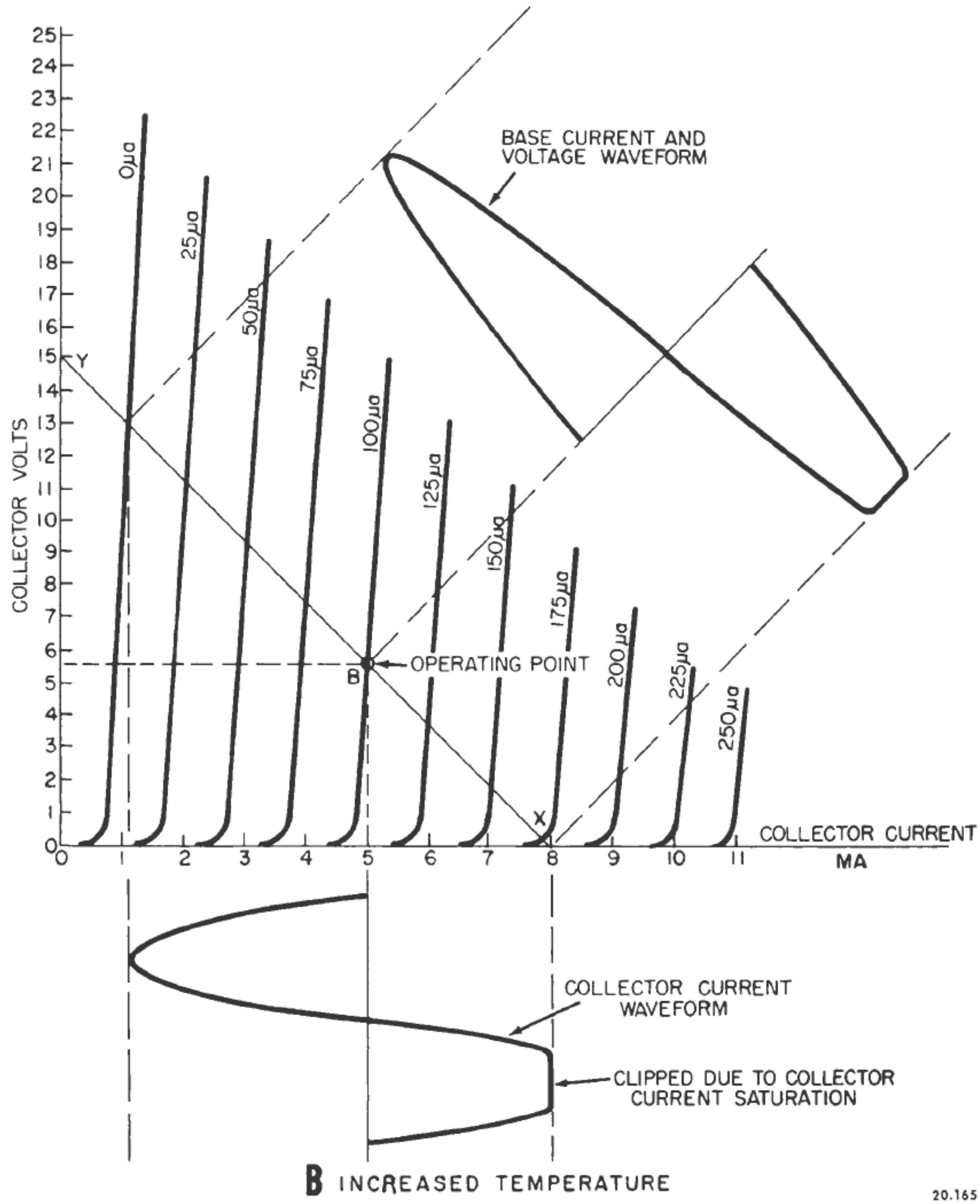


Figure 9-7.—Effect of temperature variations on NPN transistor characteristics—Continued.

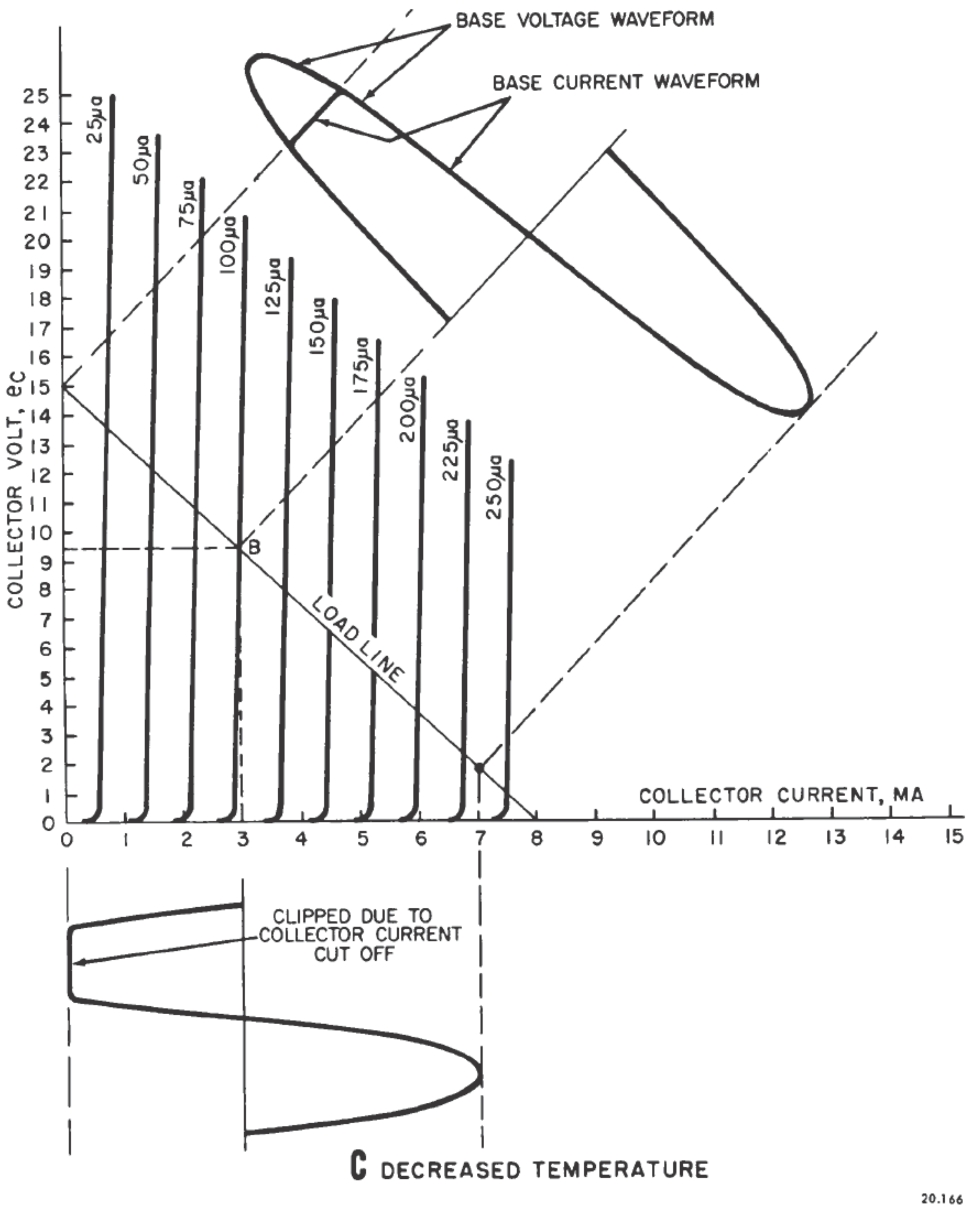


Figure 9-7.—Effect of temperature variations on NPN transistor characteristics—Continued.

The apparent shift of the operating point will not only cause distortion of the output signal, but in the case of increased temperature the increased collector current may cause the transistor to burn out. The condition may be partially remedied by connecting the bias resistor directly between the collector and base as illustrated in figure 9-8. In this circuit, R_B provides self bias for the transistor base-emitter circuit. As seen by tracing the circuit, R_B is in series with the base-emitter circuit. In the previous example the no-signal operating point corresponds to a collector voltage of 7.5 v and a base emitter current of 100 μa. For this condition,

$$R_B = \frac{E_c}{i_B} = \frac{7.5}{100 \times 10^{-6}} = 75{,}000 \text{ ohms}$$

As in the fixed bias arrangement the polarity of the voltage drop across the base-emitter junction makes the emitter negative with respect to the base. This polarity corresponds to the forward direction for the base-emitter NPN junction transistor.

With this type of self bias, a change in temperature will affect the magnitude of the bias current. For example, if an increase in temperature occurs, the tendency for the collector current to increase will be accompanied by a decrease in collector voltage across R_B with a corresponding decrease in base bias current through R_B. This action will tend to shift the operating point along the load line upward to the left of point B (fig. 9-7,B) and thus to reduce the amount of distortion caused by the temperature increase.

Conversely, a decrease in temperature will be accompanied by a tendency for the collector current to decrease and the collector voltage across R_B to increase. The increase in voltage across R_B will increase the base bias current through R_B and shift the operating point along the load line downward and to the right of point B (fig. 9-7,C). This action will again reduce distortion.

Thus, self bias provides an action that tends to partially compensate for temperature changes. In addition, self bias provides negative feedback similar to that provided by an unbypassed cathode resistor in an electron-tube amplifier. This action reduces the effective gain of the transistor amplifier.

In the transistor amplifier illustrated in figure 9-9, resistor R_B provides fixed bias for class A operation, and resistor R_E provides circuit stabilization to prevent temperature changes from altering the transistor characteristics. Resistor R_E also provides damage to the transistor by limiting the magnitude of the collector current to the maximum safe value for the highest temperature to be encountered during operation.

If the temperature increases, the increase in collector current through R_E will lower the base-emitter voltage. This action will tend to prevent further increase in collector current and therefore will prevent the shift of the characteristic curves to the right along the X axis (fig. 9-7,B). Conversely, a decrease in temperature will lower the collector current through R_E and the voltage drop across R_E. This action will increase the base-emitter voltage and will tend to prevent further reduction in collector current so that there will be less shift of the characteristic curves to the left along the X axis (fig. 9-7,C).

The value of R_E is equal to the value of R_L for maximum protection. In this case the efficiency is reduced from 50 percent in figure 9-8 for maximum power output to 25 percent in figure 9-9. As transistors become more uniform with improved methods of manufacture, the allowable tolerances will be reduced. It is believed that most transistor audio power amplifier circuits will be satisfactorily stabilized if the value of R_E is not more than 10 percent of the load resistance for maximum power output. In the preceding example of figure 9-8 according to this relation, $R_E = 0.1 \times 1875$, or approximately 188 ohms, and the power dissipated in R_E will be 10 percent of the power out of the stage.

If resistor R_1 is added to the circuit of figure 9-9, another return path for battery current will be provided. This path is around R_E. The effect of R_1 is to provide a relatively fixed base bias voltage that is independent of current change due to temperature effects. The total current in R_B is the sum of the current through R_1 and the base current I_B.

Increases in temperature tend to increase I_B and I_C and in the absence of R_1 might damage the transistor.

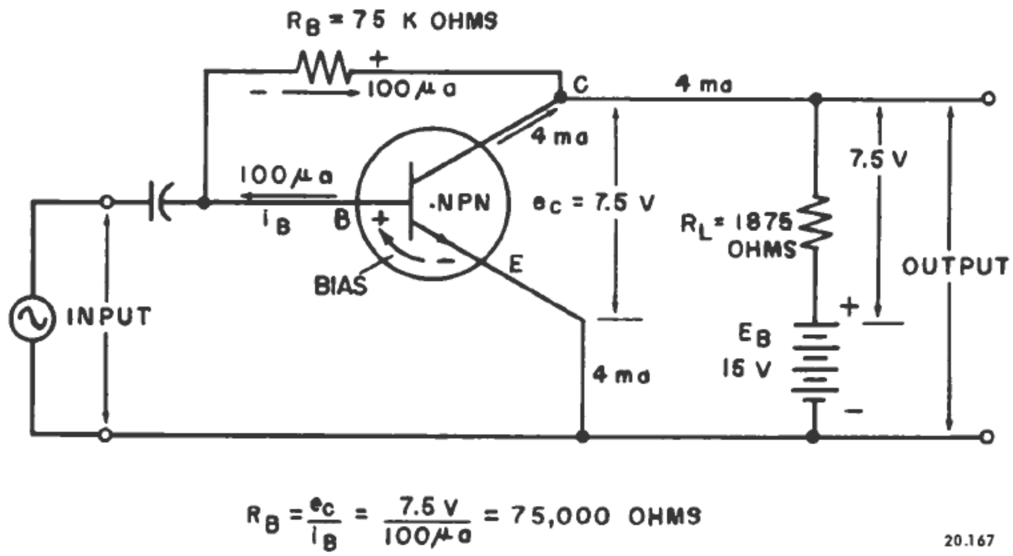


Figure 9-8.—Transistor self bias.

With R_1 in the circuit, temperature induced changes produce less voltage change across R_B due to the voltage divider action. At the same time the average d-c voltage across R_E will oppose the voltage across R_1 (like negative feedback) to compensate for the temperature effect. For example, an increase in temperature will cause an increase in collector current and voltage drop across R_E. The voltage across R_1 is equal to the sum of the voltage across R_E and the voltage across the base-emitter junction. The voltage across R_E increases directly with collector current. However, as mentioned before, the voltage across R_1 is approximately constant because of the voltage divider action. Thus an increase in voltage across R_E is accompanied by a decrease in forward bias voltage and current and the rise in I_C is limited. On the other hand more power will be wasted at the input, but there will be less change in input impedance and less change in input circuit loading.

Capacitors

In order to prevent degeneration from occurring across the stabilizing resistor, R_E (fig. 9-9), in the emitter circuit, a bypass capacitor, C_E, is connected in parallel with R_E. The action is similar to that occurring in the cathode bypass capacitor in parallel with the cathode resistor of a cathode biased electron-tube amplifier. In order to bypass the a-c component around the resistor without developing a voltage at the signal frequency across the resistor, the X_C ohms of the bypass capacitor should be low with respect to the resistance of the resistor. For most audio transistor amplifiers the bypass capacitor does not need to be larger than about 50 μf.

The coupling capacitor, C_C, must be large enough to pass the lowest frequency signal without appreciable phase shift or reduction in magnitude of the signal. The lowest frequency to be passed is the frequency at which there is a reduction in amplitude of 3 db. At this frequency the X_C ohms of the coupling capacitor are equal approximately to the input resistance of the transistor amplifier stage. In figure 9-9, the input resistance is approximately 1000 ohms and consists essentially of the resistance between the transistor base and emitter in the forward, or low-resistance, direction. If the lowest frequency to be passed is 100 cycles per second, the X_C ohms of C_C will equal the input resistance, or 1000 ohms. The capacitance of the coupling capacitor is found as follows:

$$C_C = \frac{10^6}{6.28 x f x X_C}$$

$$= \frac{10^6}{6.28x100x1000}$$

$$= 1.59 \mu f$$

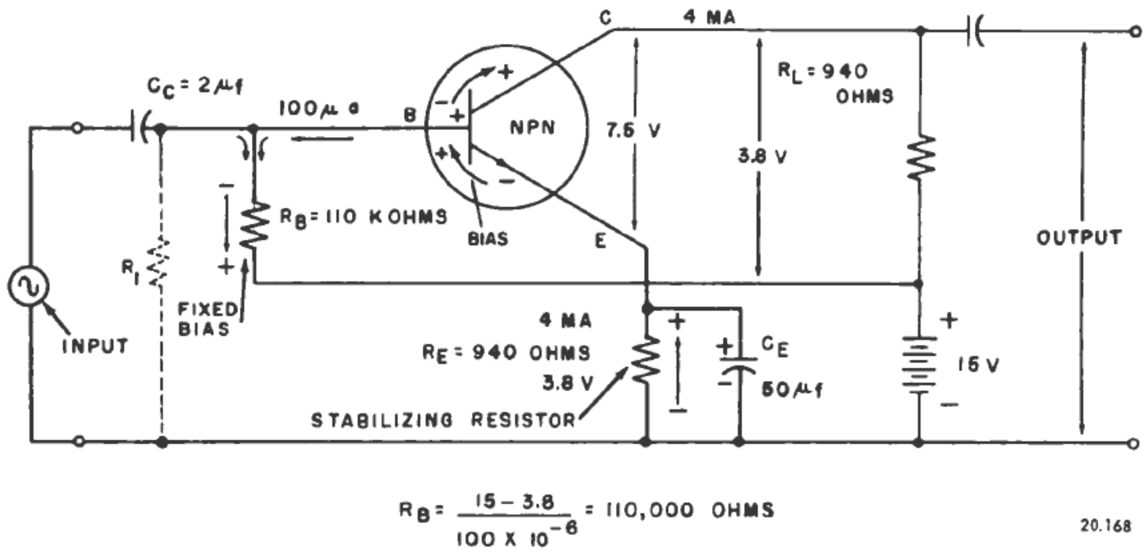


Figure 9-9.—Transistor amplifier with fixed bias and circuit stabilization.

A $2\mu f$ capacitor would extend the low frequency limit to about 80 cycles per second.

Voltage and Current Distribution

The action of a transistor amplifier will now be given with respect to the distribution of the voltages and currents through the input and output circuits before and after a signal is applied.

Figure 9-10,B, illustrates a transformer coupled PNP junction transistor audio amplifier stage employing the grounded-emitter as the common element between input and output circuits and a single battery for collector supply voltage and a base-emitter bias. The sine waveforms of signal current and voltage are illustrated in figure 9-10,A.

The input signal has an amplitude of 0.1 v. acts in series with the base-emitter bias of 0.1 v, and swings the base-emitter through a range of 0 to 0.2 v. This action swings the base-emitter current through a range of 0 to 200 μa, the collector current through a range of 0 to 8 ma, and the collector output voltage through a range of 15 to 0 v.

The circuit analysis is made for three different instants of time occurring within one complete cycle of applied signal. These instants are when (1) t = 0°, (2) t = 90°, and (3) t = 270°.

At the instant when t = 0°, the input signal voltage is zero, and the 7.5-v battery supplies the no-signal base-emitter bias current of 100 μa and the no signal collector current of 4 ma. The distribution of voltage and currents is illustrated in figure 9-10,B. Only the input circuits are analyzed in detail. Kirchhoff's law of voltages is applied to three closed circuits, all on the input side of the transistor. The algebraic sum of the instantaneous voltages around these circuits is equated to zero. If the algebraic sum of the numerical values is zero. the voltages are assumed to be correct.

The first circuit to be traced starts at point A and includes the 7.5-v battery, the bias resistor, R_B, the base, B, of the transistor. the emitter, E, and returns to the battery at the starting point, A. The circuit trace (red) is designated as ADMNBEGA. The designation for the algebraic sum of the voltages,

Σ (sum) volts = 0, identifies the individual circuit by means of the subscripts following the symbol, Σ. These letters appear on the amplifier circuit (fig. 9-10,B,C, and D). Thus the equation that applies to

$$\sum_{\text{ADMNBEGA}}^{\text{VOLTS}} = 0$$

is derived by tracing the red circuit ADMNBEGA and equating the algebraic sum of the voltages

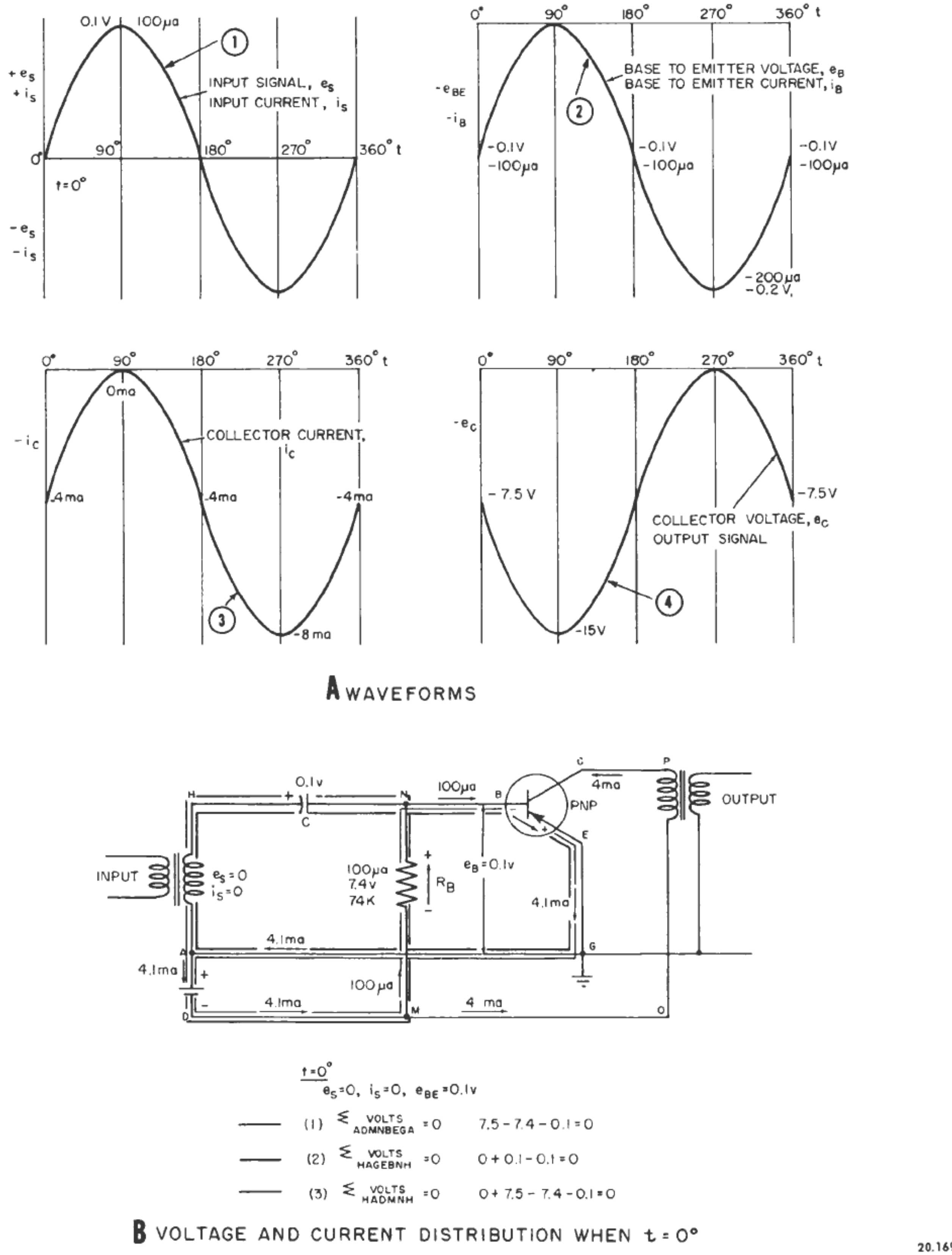
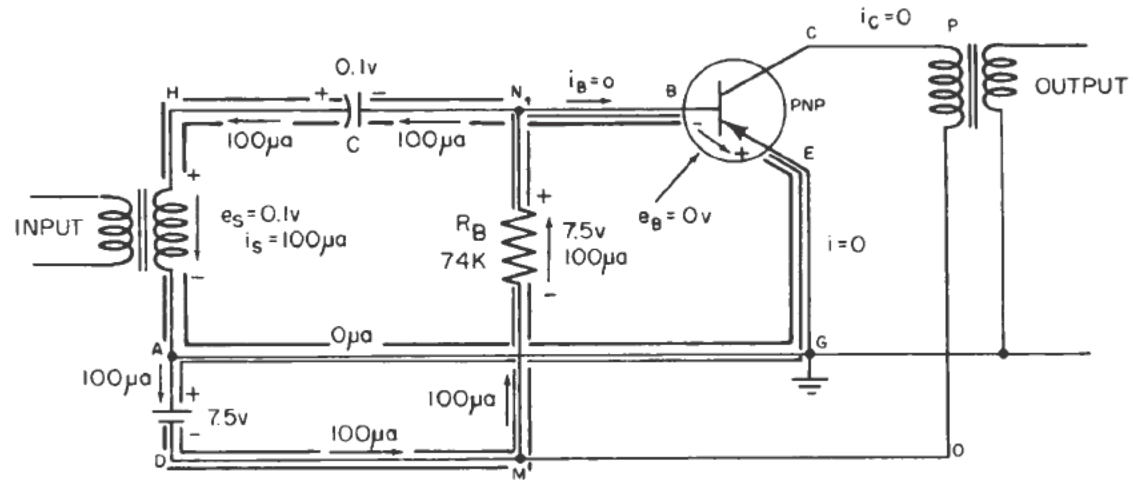


Figure 9-10.—Circuit analysis for a junction transistor amplifier having a common battery for input and output circuits.



$\underline{t = 90°}$

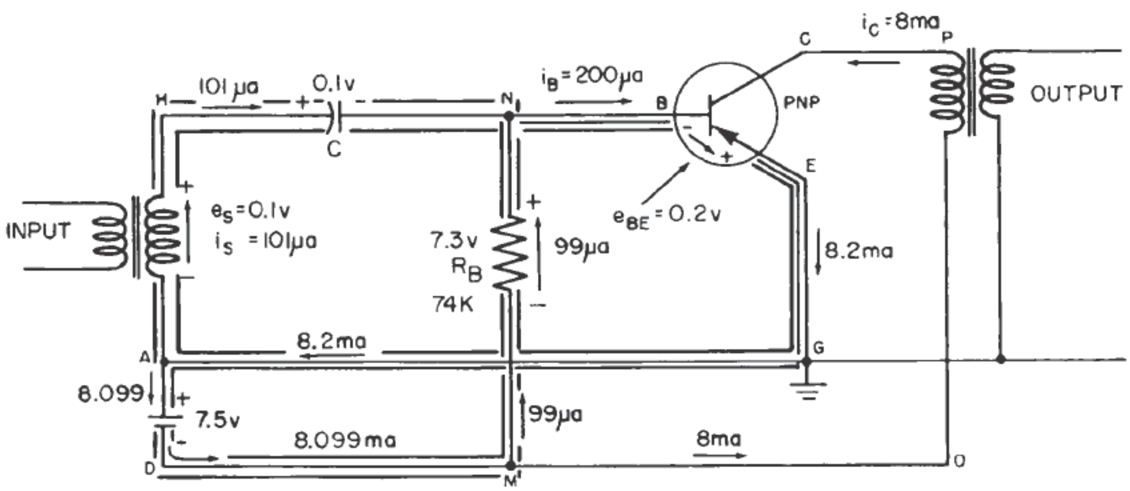
$e_S = +0.1v \quad i_S = +100μa \quad e_{BE} = 0$

(1) $\sum_{ADMNBEGA} \text{VOLTS} = 0 \qquad 7.5 - 7.5 + 0 = 0$

(2) $\sum_{HAGEBNH} \text{VOLTS} = 0 \qquad 0.1 + 0 - 0.1 = 0$

(3) $\sum_{HADMNH} \text{VOLTS} = 0 \qquad 0.1 + 7.5 - 7.5 - 0.1 = 0$

C VOLTAGE AND CURRENT DISTRIBUTION WHEN t=90°



$\underline{t = 270°}$

$e_S = -0.1v \quad i_S = -101μa \quad e_{BE} = -0.2v \quad i_B = 200μa$

(1) $\sum_{ADMNBEGA} \text{VOLTS} = 0 \qquad 7.5 - 7.3 - 0.2 = 0$

(2) $\sum_{HAGEBNH} \text{VOLTS} = 0 \qquad 0.1 + 0.2 - 0.1 = 0$

(3) $\sum_{HADMNH} \text{VOLTS} = 0 \qquad 0.1 + 7.5 - 7.3 - 0.1 = 0$

D VOLTAGE AND CURRENT DISTRIBUTION WHEN t=270°

20.1

Figure 9-10.—Circuit analysis for a junction transistor amplifier having a common battery for input and output circuits—Continued.

equal to zero in this circuit. The voltage equation for this circuit is

$$E_{Bat} - E_{RB} - e_{BE} = 0 \qquad (1)$$
$$7.5 - 7.4 - 0.1 = 0$$

The base-emitter no-signal bias current is limited principally by resistor R_B because most of the battery voltage appears across R_B. The voltage drop across the base-emitter terminals of 0.1 v in the forward direction with a base-emitter current of 100 μa or 0.1 ma indicates a base-emitter resistance of $\frac{0.1 \text{ v}}{0.1 \text{ ma}}$, or 1 k-ohm.

Capacitor C blocks the d-c component of voltage and current across the base-emitter junction from the secondary of the input transformer and charges up to the peak voltage drop across the junction. Capacitor C is relatively large (2 mf), and the voltage across C is assumed to remain at 0.1 v through the entire cycle of applied signal voltage.

The second voltage equation for the instant when $t = 0°$ is derived by tracing around the blue circuit HAGEBNH and equating the algebraic sum of the instantaneous voltages to zero. Thus the equation that applies to

$$\sum_{\text{HAGEBNH}} \text{VOLTS} = 0 \text{ is}$$

$$e_s + e_{BE} - E_c = 0 \qquad (2)$$
$$0 + 0.1 - 0.1 = 0$$

As mentioned previously, C will block the d-c component of base-emitter current and d-c voltage from the secondary of the input transformer.

The third voltage equation for the instant when $t = 0°$ is derived by tracing around the brown circuit HADMNH and equating the algebraic sum of the instantaneous voltages to zero. Thus the equation that applies to

$$\sum_{\text{HADMNH}} \text{VOLTS} = 0 \text{ is}$$

$$e_s + E_B + E_{RB} - E_c = 0 \qquad (3)$$
$$0 + 7.5 - 7.4 - 0.1 = 0$$

The no-signal condition upon which the preceding three equations depend may be regarded as the operating point, B (fig. 9-6,A) on the load line, XY. The load impedance (noninductive) is assumed to be the slope of the load line, or 1875 ohms. The principal difference between the example of figure 9-10 and that of figure 9-6 is that the load resistor of figure 9-6 has been replaced by a transformer having negligible d-c resistance in its windings so that the collector voltage supply may be reduced from 15 v to 7.5 v, without changing the slope of the load line. Thus, the no-signal collector current of 4 ma that flows through the primary of the output transformer produces a negligible d-c voltage drop across the primary, and the collector voltage for this condition is 7.5 v.

During the interval between 0° and 90° the input signal voltage increases from 0 to a positive maximum value of 0.1 v, and the input signal current increases from 0 to 101.5 μa as represented by curve (1) of figure 9-10,A. During this interval, the base-emitter current decreases from 100 μa to 0 curve (2), and the collector current decreases from 4 ma to 0 curve (3). At the same time, the induced voltage acting in the primary of the output transformer causes the collector voltage to increase from -7.5 v to -15 v, curve (4).

At the instant $t = 90°$ (fig. 9-10,C) the input signal voltage is a positive maximum value of 0.1 v. The polarity of the signal voltage indicates that this voltage acts across the base-emitter junction in the backward, or high-resistance, direction, hence no base current will flow. Instead, signal current will flow through the secondary of the input transformer (brown circuit) in series with the battery, the bias resistor, R_B, and the capacitor, C. The signal voltage acts in series additively with the battery voltage, thereby increasing the voltage drop across R_B from 7.4 v to 7.5 v. During the interval from 0° to 90° the base current decreases from 100 μa to 0 as the signal current increases from 0 to 101.5 μa. This action indicates that an approximately constant current flows through R_B into junction N. As the current flow into the base, B, decreases, the current flow into the capacitor circuit, NH, increases. It is assumed that the flow of signal current through the circuit from N to H will not increase the voltage drop across the capacitor to any appreciable extent. Now apply Kirchhoff's law again.

The voltage equation for the instant when $t = 90°$ that corresponds to the red circuit designation

$$\sum_{ADMNBEGA} VOLTS = 0 \text{ is} \tag{4}$$

$$E_{Bat} - E_{RB} - e_{BE} = 0$$

$$7.5 - 7.5 - 0 = 0$$

The slight increase in voltage across the 74 k-ohm resistor, R_B, from 7.4 v to 7.5 v produces a negligible increase in current (1.5μa) through R_B and a decrease in voltage across the base-emitter junction from -01 v to 0, curve (2) (fig. 9-10,A).

The voltage equation for the instant when $t = 90°$ that corresponds to the blue circuit designation

$$\sum_{HAGEBNH} VOLTS = 0 \text{ is} \tag{5}$$

$$e_S + e_{BE} - E_C = 0$$

$$0.1 + 0 - 0.1 = 0$$

The voltage equation for the instant when $t = 90°$ that corresponds to the brown circuit designation

$$\sum_{HADMNH} VOLTS = 0 \text{ is}$$

$$e_S + E_{Bat} + E_{RB} - E_c = 0 \tag{6}$$

$$0.1 + 7.5 - 7.5 - 0.1 = 0$$

The peak positive input signal current flows in a counterclockwise direction around this circuit.

During the interval from 90° to 180° the input signal voltage decreases from positive maximum to 0, curve (1) (fig. 9-10,A). At the same time the voltage drop across R_B decreases from 7.5 v to 7.4 v, and the base-emitter voltage increases from 0 to -0.1, curve (2). Thus during the first half cycle (0° to 180°) of applied signal, the base-emitter current varies sinusoidally through a maximum change of 100 μa.

In the collector circuit the collector current also varies sinusoidally through a maximum change of 4 ma as the collector voltage varies sinusoidally through a maximum change of 7.5 v, curves (3) and (4).

The input signal voltage for the second half cycle is of opposite polarity to that for the first half cycle, curve (1). Thus the signal voltage opposes the battery voltage and lowers the voltage drop across R_B to 7.3 v at the instant when $t = 270°$. At this instant the signal voltage acting in series addition with the capacitor voltage develops a voltage of -0.2 v across the base-emitter junction in the forward or easy direction of current flow. Thus the base current increases to a peak of -200 μa. This action causes the collector current to increase to a maximum value of -8 ma, as the self-induced voltage in the output transformer primary causes the collector voltage to swing to zero, curves (3) and (4).

The voltage equation for the instant when $t = 270°$ that corresponds to the (red circuit) designation

$$\sum_{ADMNBEGA} VOLTS = 0 \text{ is}$$

$$E_{Bat} - E_{RB} - e_{BE} = 0 \tag{7}$$

$$7.5 - 7.3 - 0.2 = 0$$

The current flow through R_B (fig. 9-10,D) decreases slightly from 100 μa to 99 μa. The signal current flowing into junction N from terminal H increases from 0 to 101 μa and combines with the bias current flowing through R_B into junction N to increase the base-emitter current from -100 μa to -200 μa, curve (2). This current returns to junction A where it divides almost equally with 101 μa, returning to the secondary of the input transformer and 99 μa returning to the battery.

The voltage equation for the instant when $t = 270°$ that corresponds to the blue designation

$$\sum_{HAGEBNH} VOLTS = 0 \text{ is}$$

$$-e_S + e_{BE} - E_c = 0 \tag{8}$$

$$-0.1 + 0.2 - 0.1 = 0$$

The voltage equation for the instant when t = 270° that corresponds to the brown designation

$$\sum_{HADMNH} VOLTS = 0 \text{ is}$$

$$-e_s + E_{Bat} - E_{RB} - E_c = 0 \qquad (9)$$

$$-0.1 + 7.5 - 7.3 - 0.1 = 0$$

As the voltage across R_B decreases from 7.4 v to 7.3 v, the voltage across the base-emitter junction increases from -0.1 v to -0.2 v in the forward direction of the base-emitter circuit. The second half cycle is completed as the signal voltage decreases from negative maximum to 0, the base-emitter current decreases from -200 to -100 μa, and the collector current decreases from -8 ma to -4 ma and the collector voltage swings from 0 v back to -7.5 v.

The peak-to-peak signal voltage appearing across the primary of the output transformer is 15 v. The peak-to-peak signal component of current through the primary of the output transformer is 8 ma. The power output of the amplifier is $e_{rms} i_{rms}$, or $\frac{15}{2} \times 0.707 \times \frac{8}{2} \times 0.707 = 15$ mw.

CASCADE AMPLIFIERS

Transistor amplifier stages may be connected in cascade as illustrated in figure 9-11. Transistors Q1 and Q2 are PNP junction transistor audio voltage amplifiers, and Q3 is a PNP junction transistor power amplifier.

Single-Ended

Each stage is single-ended. Transformer T1 is a step-down matching transformer that couples a high-impedance microphone to a low-impedance input circuit. Interstage coupling transformers T2 and T3 match the output impedance of one stage to the input impedance of the next. These impedances are not widely different. For example, the output impedance of Q1 may be 2000 ohms and the input impedance of Q2 may be 1000 ohms. Transformer T4 is a stepdown transformer that couples the output signal in the collector circuit of the third stage to the low-impedance voice coil of the reproducer. Capacitors C1, C2, and C3, block the d-c bias from the secondaries of T1, T2, and T3, and couple the signal to the input circuits. Potentiometer P, serves as a volume control for the amplifier. Volume increases as the arm is moved toward the upper end of T2. Resistors R1- R2, and R3 limit the no-signal base-emitter bias current to the proper value for each stage. The grounded-emitter is common to all input and output circuits, and the polarities of the input and output d-c voltages for each stage correspond to those required by PNP junction transistors.

Push-Pull

Transistor power amplifiers are usually connected in push-pull because of the advantages over single-ended operation. The most important advantages are the reduction in distortion and the removal of d-c core saturation from the output transformer. Second harmonic distortion and other distortion caused by even-order harmonics are cancelled in push-pull class A

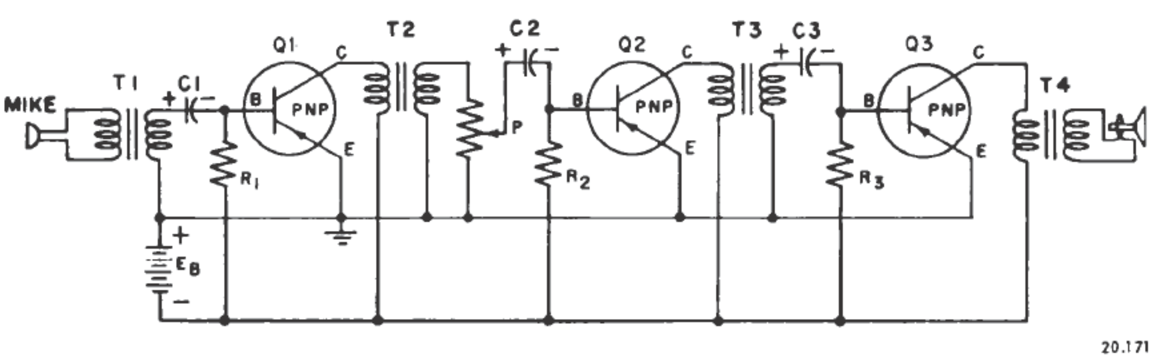


Figure 9-11.—Cascade connection of transistor amplifiers.

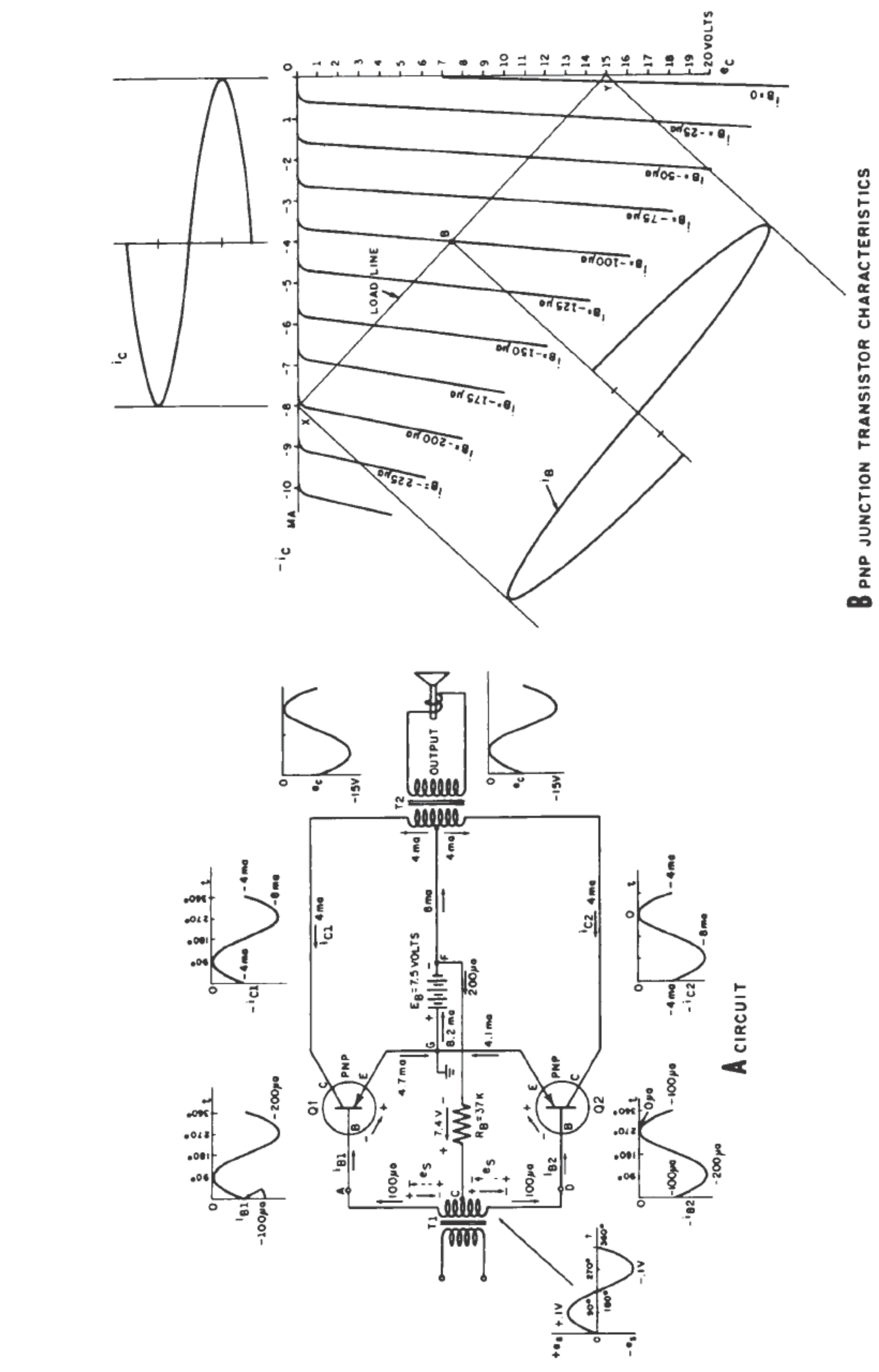


Figure 9-12.—Transistor push-pull amplifier.

amplifiers. The load impedance and the output power are twice the values for single-ended operation. The incremental inductance of the output transformer is higher as a result of the elimination of the d-c component of primary current.

A class A push-pull amplifier using PNP transistors is illustrated in figure 9-12, A. The characteristic curves for these transistors are illustrated in figure 9-12, B. When biased for class A operation the no-signal collector current is 4 ma (point B on the load line). The corresponding base current is 100 μa. The voltage across the bias resistor, R_B, is the difference between the battery voltage and the drop across the base-emitter circuit, or 7.5 - 0.1 = 7.4 v. The no-signal base-emitter current through R_B is the sum of the base-emitter current supplied to each transistor, or 200 μ a. The resistance of R_B is

$$\frac{7.4}{200 \times 10^{-6}} = 37{,}000 \text{ ohms}$$

Consider the action of the input signal on the base-emitter current of each transistor. When t=0° the input signal is 0, and the base-emitter current through R_B divides at C, flowing in opposite directions through the two halves of the secondary of T1.

When t=90° the signal voltage has a peak value of 0.2 v, and the direction is represented by the solid arrows. The signal voltage is distributed equally between both halves of the secondary of T1 with 0.1 v acting in each half. By tracing around the input circuit of Q1, GFCABEG, the voltage equation is developed. By Kirchhoff's law of voltages, $\sum_{\text{GFCABEG}} \text{VOLTS} = 0$, and the corresponding equation is 7.5-7.4-0.1=0. The first term of this equation represents the battery voltage, the second term represents the drop across R_B, the third term represents the signal voltage across the upper half of the secondary of T1, and the fourth term represents the drop across the base-emitter circuit of Q1. At this instant (t=90°) the signal voltage induced in the upper half of the secondary of T1 opposes the flow of base-emitter current in Q1, and the base current of Q1 is 0.

At the same instant the signal voltage induced in the lower half of the secondary of T1 aids the flow of base-emitter current in Q2. The voltage equation corresponding to the designation $\sum_{\text{GFCDBEG}} \text{VOLTS} = 0$ is 7.5-7.4+0.1-0.2=0. The first term represents the battery voltage, the second term represents the drop across R_B, the third term represents the voltage induced in the lower half of the secondary of T1, and the fourth term represents the drop across the base-emitter circuit of Q2. At this instant (t=90°) the base-emitter current of Q2 is 200 μ a.

One half cycle later when t=270° the polarities of the signal voltage are reversed as indicated by the dotted arrows in the secondary of T1. At this instant the voltage in the upper half of the secondary of T1 aids the base-emitter bias voltage of Q1, and the base current of Q1 increases to 200 μ a. At the same instant the voltage in the lower half of the secondary of T1 opposes the base-emitter bias of Q2, and the base current of Q2 decreases to 0. For the upper circuit, GFCABEG $\sum_{\text{GFCABEG}} \text{VOLTS} = 0$, and the voltage equation is 7.5-7.4+0.1-0.2=0. For the lower circuit, GFCDBEG $\sum_{\text{GFCDBEG}} \text{VOLTS} = 0$, and the voltage equation is 7.5-7.4-0.1-0=0. From the fourth term in these equations the voltage drop across the base-emitter terminals of the transistors is 0.2 v for Q1 and 0.0 v for Q2. Thus the base-emitter current of Q1 has increased to 200 μ a while that of Q2 had decreased to 0.

The current through R_B remains constant over the input cycle, hence the voltage drop across R_B is constant, and R_B does not require a bypass capacitor when the amplifier is operated class A.

The waveform of current in the collector circuits of Q1 and Q2 is like that of the input circuits. On no-signal the collector currents flow in opposite directions through the primary of T2 from the center tap. Because these currents are equal in magnitude, the ampere turns are equal. Because they are opposite in direction, they produce no effect on the magnetization of the iron, and there is no magnetization of the core when the input signal is 0. When the input signal current increases the collector current of Q2 from 4 ma to 8 ma, it decreases the collector current of Q1 from 4 ma to 0. The increasing current in the lower half of the primary of T2 and the decreasing current in the upper half combine additively in the secondary to produce the output voltage of T2. Similarly on the next half cycle the increasing current in the upper half of the primary of T2 and the decreasing current in the lower half combine additively in the secondary. The effect is the same as that of combining the output signal voltages of Q1 and Q2 in series addition across the two halves of the primary of T2. Thus, the Q1 output signal voltage of 15 v (peak-to-peak) combines effectively in series addition with the Q2 output signal voltage of 15 v (peak-to-peak) to produce a peak-to-peak output voltage of 30 v. The peak-to-peak signal current through the primary of T2 is 8 ma. Thus, the effective impedance looking into the primary of T2 is $\frac{30}{0.008} = 3750$ ohms. The power output is $e_{rms}i_{rms} = \frac{30}{2} \times 0.707 \times \frac{8}{2} \times 0.707 = 30$ mw. This value of power output is twice that of the example of figure 9-10 in which a single-ended amplifier employs a transistor of the same characteristics as those of the push-pull stage.

TRANSISTOR OSCILLATORS

Because transistors have the ability to amplify, they can be used in oscillator circuits which are similar to those using electron tubes as described in the preceding chapter.

TICKLER FEEDBACK

A PNP transistor oscillator using a tickler coil for inductive feedback is illustrated in figure 9-13, A. The base is common to the input and output circuits. The battery at the left biases the emitter-base input circuit in the forward direction. The battery at the right biases the collector-base output circuit in the reverse or high resistance direction. Waveforms of emitter current (i_e) and collector current (i_c) are indicated in figure 9-13, B.

Oscillations begin when power is applied (S is closed). Emitter current, i_e, and collector current, i_c, increase from O to Y (fig. 9-13, B) because of regenerative feedback in T1 from the 3-4 winding to the 1-2 winding.

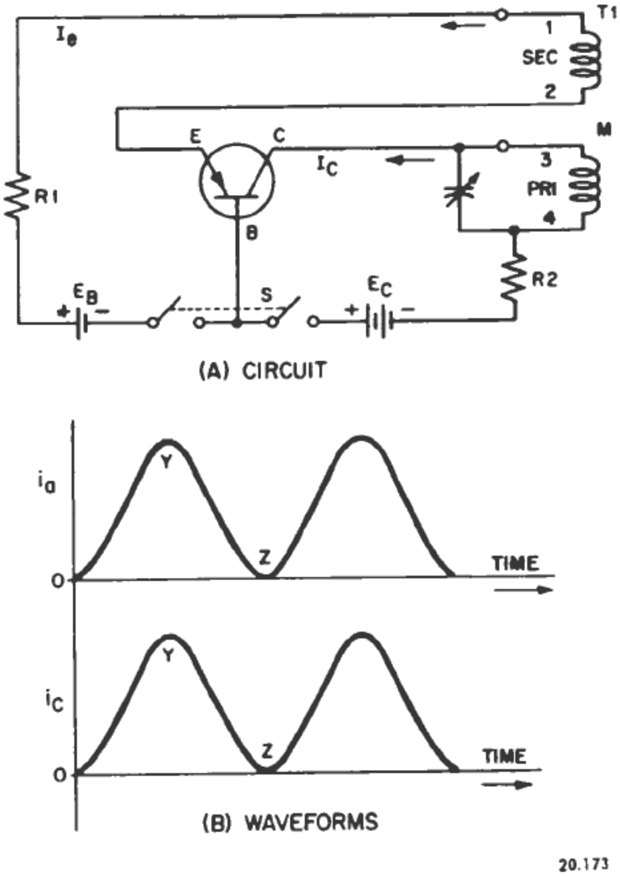


Figure 9-13.—Tickler feedback oscillator.

At point Y the transistor is saturated and no further increase in current will occur. Feedback ceases and i_e starts to decrease. Collector current i_c, starts to decrease. The feedback in T1 reverses polarity and causes i_e to decrease to zero. At zero (point Z, fig. 9-13, B) feedback again ceases.

Once the feedback voltage drives the transistor to cutoff, the bias voltages revert to their original condition and the process is repeated. The transistor is driven to saturation, then to cutoff, then back to saturation and so forth. The time for change from saturation to cutoff is primarily determined by the tank circuit which in turn determines the frequency of oscillation.

HARTLEY TRANSISTOR OSCILLATOR

An NPN junction transistor employed in a Hartley oscillator is shown in figure 9-14. The tapped coil returned to the common emitter identifies the circuit as the Hartley. The collector circuit may be either shunt fed (fig. 9-14,

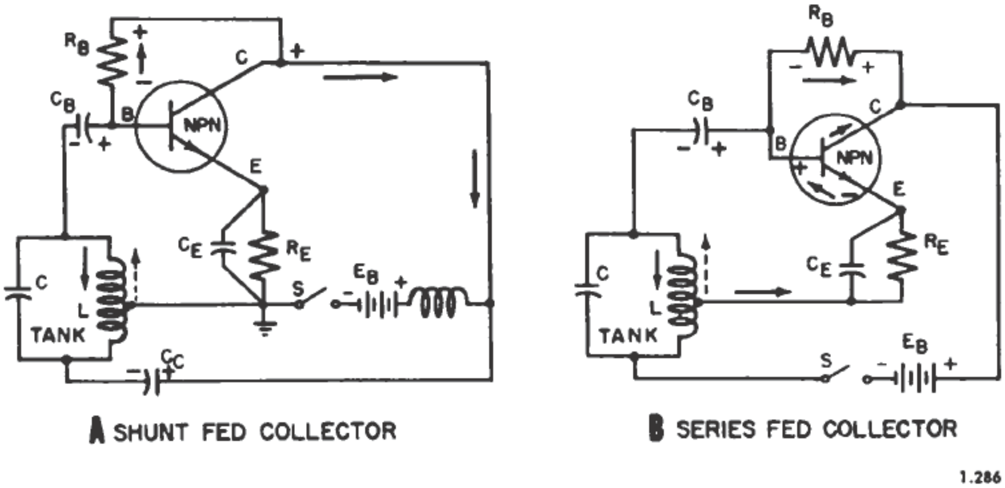


Figure 9-14.—NPN junction transistor Hartley oscillator.

A) or series fed (fig. 9-14, B). Resistor R_B limits the base-emitter bias current to the proper value. The L and C values of the tank circuit control the frequency of the oscillator. Capacitor C_B prevents the tank coil from shorting the base-emitter junction of the transistor. Resistor R_E limits the collector current to a safe value and C_E prevents negative feedback. Battery E_B provides bias current for the input and collector voltage for the output circuit.

When S is closed (fig. 9-14, B), collector current flows from the negative terminal of the battery through the lower portion of the tank coil to terminal E of the transistor by way of the center tap of L. The rising value of the collector current induces a voltage in the upper half of L that aids the flow of base-emitter current. This induced voltage is represented by the solid arrow. The accompanying increase in base-emitter current causes a further increase in collector current. The action continues until saturation is reached. Resistor R_E determines the operating point on the a-c load line (not shown).

At saturation the collector current stops rising and the induced voltage in the upper half of L reduces to zero. This action reduces the magnitude of the base-emitter bias current, which reduces the magnitude of the collector current. The collapse of collector current induces a voltage in the upper half of L of opposite polarity to that originally induced by the rise of collector current. The direction is indicated by the dotted arrow. This voltage opposes the base-emitter bias current and reduces the collector current to zero. When the collector current is zero, the induced voltage in the upper half of L becomes zero and the base-emitter bias current again increases. The increase in collector current initiates the second cycle of operation. During each half cycle the tank capacitor charges and discharges with an interchange of energy occurring between the coil and the capacitor. This action controls the oscillator frequency. Increasing L or C decreases the frequency. The resonant frequency of the tank is

$$f = \frac{159}{\sqrt{LC}}$$

where f is in mc, L is in μh, and C is in $\mu\mu$f. The oscillator frequency is less than the value indicated in the formula because of the loading effect that the transistor imposes on the tank.

COLPITTS TRANSISTOR OSCILLATOR

A common emitter PNP transistor connected as a Colpitts oscillator is illustrated in figure 9-15, A. The capacitor voltage divider C1C2 replaces the tapped coil of the previously described Hartley oscillator.

When S is closed, oscillations start. C1 and C2 acquire the polarities indicated. The base-emitter is forward biased and the collector emitter reverse biased as in amplifier circuits. Base current, i_b, flows on the initial charge

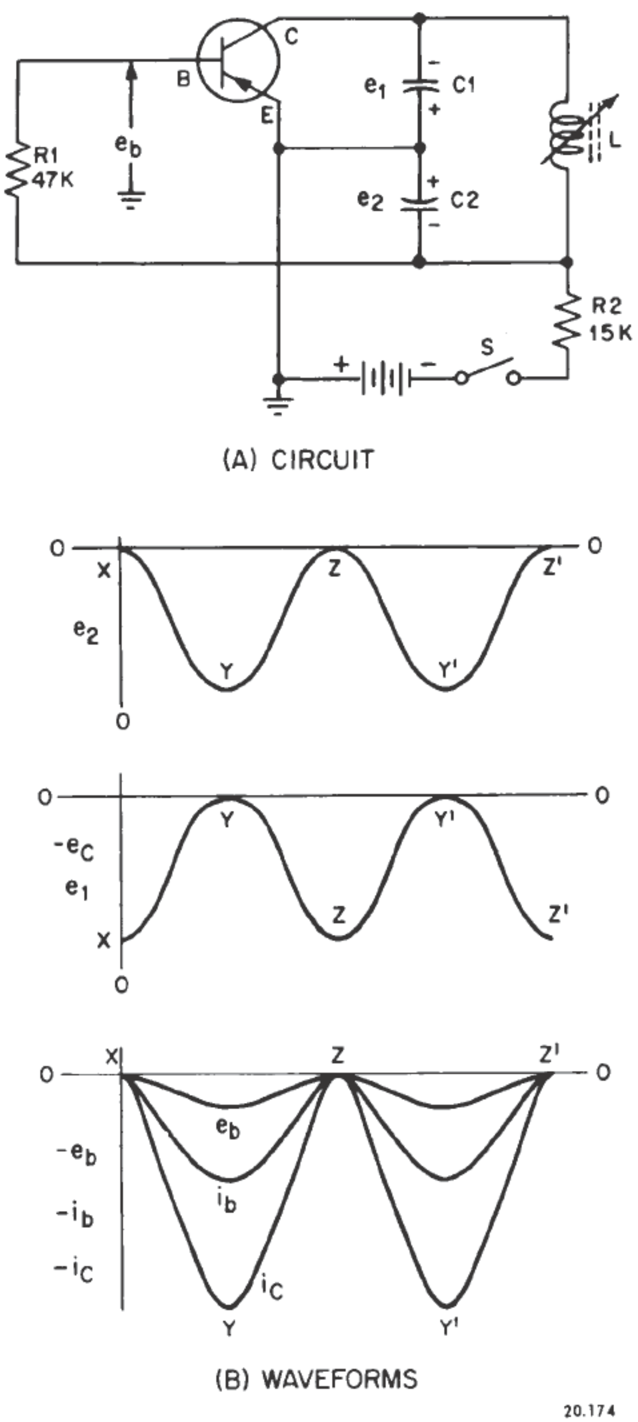


Figure 9-15.--Colpitts transistor oscillator.

of C2. Collector current flows through R2 and collector voltage, e_c, decreases (X to Y, fig. 9-15, B) due to drop through R2.

Capacitor C2 charges to approximately three-quarters of the battery voltage via R2; i_b and i_c increase during this time. Then C2 starts to discharge via tank coil L into C1.

The voltage across C1 is initially low (point Y, fig. 9-15, B). As the C2 voltage falls, i_b decreases, and i_c decreases (both to cutoff); e_c increases with the charge in C1 (Y to Z, fig. 9-15, B).

Capacitor C1 discharges into C2 via L. During the next portion of the cycle, Z to Y: C2 assumes a charge, i_b increases, and i_c starts to flow again. The collector voltage e_c drops because the voltage increases across R2. The battery supplies additional charge to C2 during this part of the cycle to overcome tank circuit losses.

Capacitor C2 stops charging; e_b and i_b stop increasing (point Y'). The base-emitter bias reverts to the normal value; i_c decreases as C2 begins to discharge into C1 via tank coil L (points Y' to Z'). Capacitor C2 voltage, e_2, decreases; e_b, i_b, and i_c decrease (Y' to Z', fig. 9-15, B); e_1 and e_c increase in magnitude (Y' to Z'). As i_b decreases to cutoff, i_c decreases to zero, and C1 starts to discharge again into C2, to repeat the cycle (point Z').

FREE-RUNNING MULTIVIBRATOR

Two PNP transistors are connected in a common emitter circuit to form a free-running multivibrator (fig. 9-16, A). The waveforms are illustrated in figure 9-16, B. This oscillator is basically a two stage r-c coupled amplifier with the output of the second stage coupled through C1 to the input of the first stage, and the output of the first stage coupled through C2 to the input of the second stage. The feedback action is similar to that of the free-running multivibrator using triode electron tubes.

Assume that when S is closed, Q1 conducts more than Q2. The Q1 collector voltage, e_{c1}, falls, that is, goes less negative (due to the drop through R2). Capacitor C2 discharges through the collector-emitter circuit of Q1 via R3 and the battery. This action places reverse bias on the base-emitter circuit of Q2 and holds Q2 cut off.

Capacitor C1 charges via R4 and the base-emitter circuit of Q1 to add forward bias current to the Q1 input. Collector current of Q1 goes to saturation.

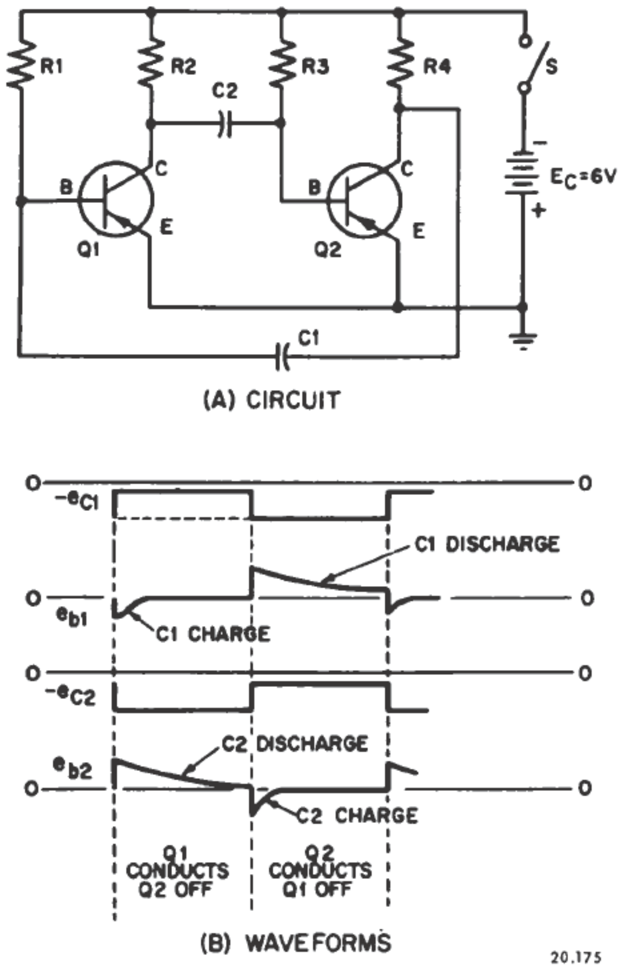


Figure 9-16.—Free-running transistor multivibrator.

As C2 continues to discharge,. its voltage decreases and the base of Q2 becomes less positive until at zero volts, the Q2 forward bias current starts to flow from the supply battery via R3, and Q2 conducts.

As Q2 conducts, e_{c2} falls (due to the drop through R4) and C1 discharges via the collector-emitter circuit of Q2, E_c, and R1. The direction of electron flow to and from the Q1 input applies reverse bias to the base-emitter circuit of Q1 and therefore cuts off the Q1 collector current.

Capacitor C2 recharges via R2 and the base-emitter circuit of Q2 to increase the forward bias of the Q2 input; e_{c2} falls as the Q2 collector current goes to saturation.

The discharge time of C1 and C2 is relatively long compared to the charge time. The capacitor that is charging has reached its final value before the other capacitor has completely discharged.

The frequency of the transistor multivibrator depends on the discharge time-constant, the type of transistor, and the magnitude of the supply voltage.

ONE-SHOT MULTIVIBRATOR

The transistor one-shot multivibrator (fig. 9-17, A) has a single (mono) stable condition of equilibrium. Like its electron tube counterpart, it requires a trigger pulse to cycle and return to its initial state.

For example, in the absence of a trigger, the base-emitter circuit of Q2 is forward biased and Q2 conducts. The base-emitter circuit of Q1 is reverse biased (the drop across R6 exceeds that across R2) and Q1 is cut off.

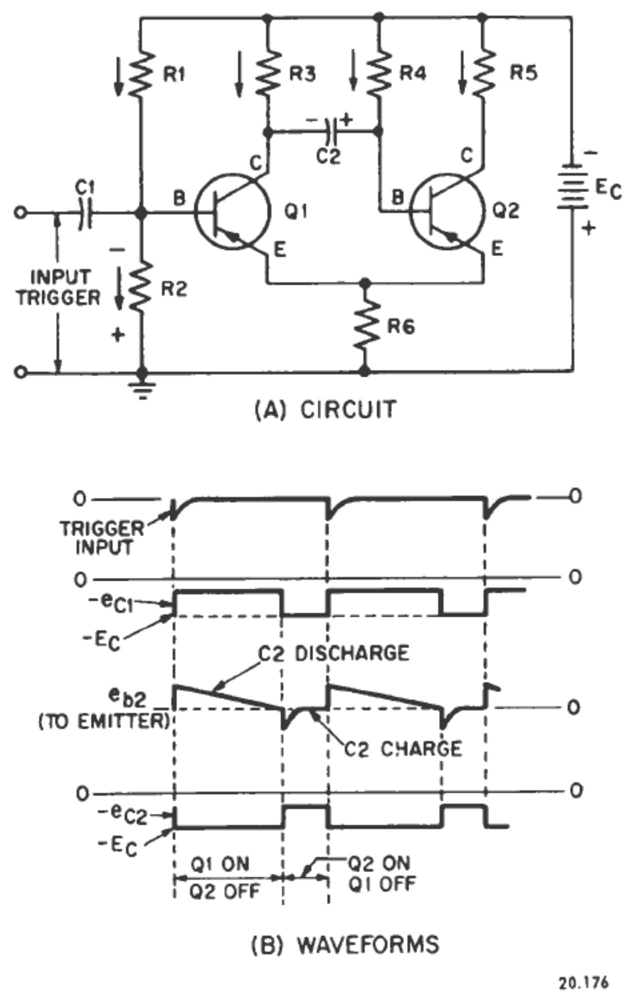


Figure 9-17.—One-shot multivibrator.

A negative trigger (fig. 9-17, B) is applied to the base-emitter circuit of Q1 (fig. 9-17, A) to start the cycle. The drop across R2 exceeds that across R6 to momentarily forward bias the Q1 input; Q1 collector current flows and the Q1 collector voltage, e_{c1}, falls (due to the drop through R3); C2 discharges via R4, the collector-emitter circuit of Q1 and the battery, E_c.

This action biases the base-emitter circuit of Q2 in the reverse direction to cut off Q2. Q2 is held cut off until C2 discharges to zero; now Q2 is no longer reverse biased; Q2 collector current starts to flow; the drop across R6 exceeds that across R2 to place reverse bias on the Q1 input.

Transistor Q1 is cut off; e_{c1} rises and C2 charges via R3 and the Q2 input to place a forward bias current through the Q2 input. Collector current of Q2 goes to saturation and the drop across R6 holds Q1 cut off until the next input trigger via C1.

The input trigger is of short duration and lasts only long enough to initiate the cycle.

The relatively long interval that Q2 is off and Q1 is conducting is caused by the relatively large value of R4 (in the C2 discharge path) compared with that of R3 (in the C2 charge path).

The repetition rate of the oscillator is that of the external trigger.

The one-shot action is thus seen to be that of temporarily cutting Q1 on and Q2 off; after C2 discharges, Q2 reverts to its initial conducting condition and Q1 returns cut off.

FLIP-FLOP (ECCLES-JORDAN) MULTIVIBRATOR

The flip-flop (Eccles-Jordan) multivibrator (fig. 9-18, A) is a bi-stable oscillator (either Q1 conducts when Q2 is off, or Q1 is off when Q2 conducts.

In this example the circuit is assumed to be symmetrical, that is, R3=R4, R5=R6, R1=R2, C1=C2, C3=C4, and transistor Q1 is similar to transistor Q2. The waveforms are shown in figure 9-18, B.

Assume that when battery power is applied, Q1 conducts more rapidly than Q2. Collector voltage e_{c1} falls because of the increased drop through R3. This action reduces the voltage applied to the divider R6, R2 and E_b cuts off Q2 (E_b is greater than the drop across R2). Collector voltage e_{c2} rises to almost the collector supply voltage E_c (because there is only a small drop through R4). This action supplies increased voltage across divider R5R11. The drop across R1 exceeds E_b to place forward bias on the Q1 input circuit. Transistor Q1 conducts heavily. Thus prior to time, t_o (fig. 9-18, B), Q1 conducts and Q2 is cut off.

At time, t_o, a positive pulse is applied between the base and ground of Q1 which reduces the current through R1 and the collector current in Q1; e_{c1} rises (there is less drop in R3) and the voltage supplied to divider R6R2 increases; the drop across R2 exceeds E_b and the resulting forward bias on Q2 causes Q2 to conduct heavily; e_{c2} falls (due to increased drop across R4) and less voltage is supplied to divider R5R1. The drop across R1 falls below E_b and Q1 cuts off.

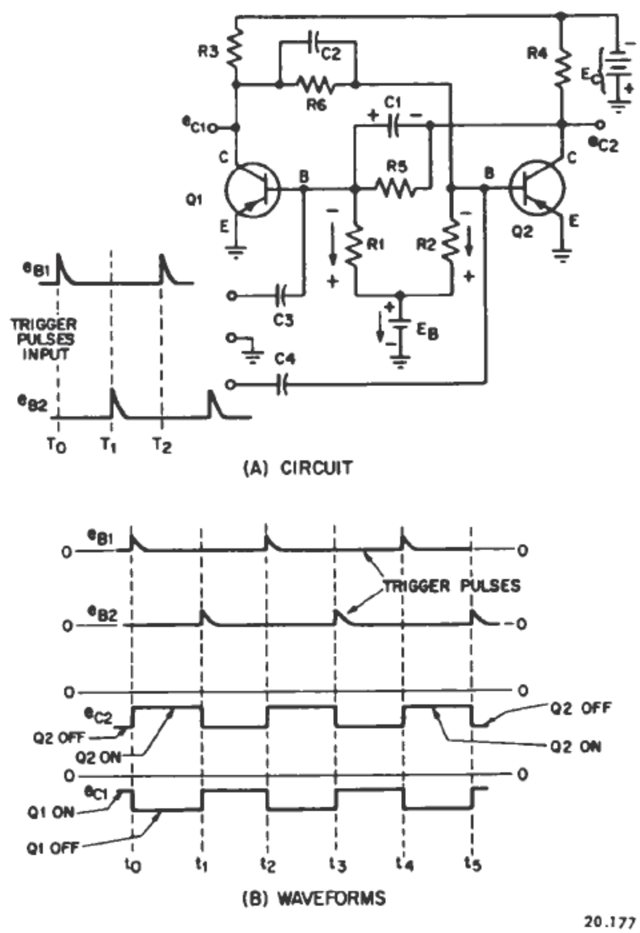


Figure 9-18.—Flip-flop (Eccles-Jordon) bistable multivibrator.

Thus a positive pulse applied to the Q1 base via C3 at time, t_o, cuts off Q1 and cuts on Q2. This condition is stable.

(The same positive pulse applied to the Q2 base at time, t_o, would have no effect on Q2 because Q2 was already in cutoff condition due to the faster current rise in Q1 at the time battery power was applied.)

At time, t_1, a positive pulse applied to the Q2 base (Q2 conducting) will cut off Q2. The rise in collector voltage, e_{c2}, will apply forward bias to Q1 (the drop across R1 will exceed E_b) and Q1 will conduct with an accompanying decrease in e_{c1} (caused by the increase in drop across R3).

Thus a second stable condition exists at t_1 when Q1 conducts and Q2 is cut off.

Note that the repetition rate of the output waveform between both collectors and ground is one-half that of the input pulses (2 input pulses for 1 output pulse).

BLOCKING OSCILLATOR

A transistor free running oscillator (fig. 9-19, A) may be made to block periodically by inserting a capacitor, C1, across the base-emitter circuit and a resistor, R1, in a self-biasing circuit provided the time constant, R1C1, is relatively long compared to the period of one cycle. Positive feedback is accomplished from collector to base by mutual induction from primary to secondary of the transformer. The waveforms are illustrated in figure 9-19, B.

When S is closed, i_c increases (A to B, fig. 9-19, B) because of positive feedback and the increase of i_b (forward bias from the battery via R1); e_c falls as i_c rises (A to B).

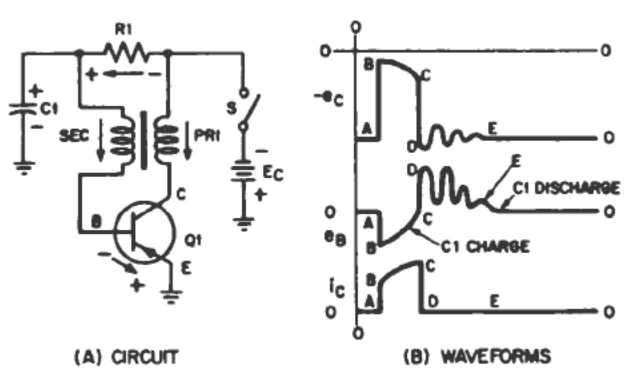


Figure 9-19.—Transistor blocking oscillator.

Positive feedback through the transformer resulting from the increase in collector current through the primary charges C1 (B to C). At point C, i_c goes to zero as e_c increases from point C to point D; e_b goes positive as C1 discharges in series addition to the battery voltage. (The voltage across R1 is greater than E_c and this action places reverse bias on the base-emitter circuit (points C to D).)

As C1 discharges (D to E) its voltage falls and eventually will unblock Q1 to permit the cycle to repeat.

The period from D to E indicates damped oscillations in the transformer primary when an interchange of energy takes place between the distributed capacitance and the inductance of the winding.

The repetition rate of the blocking oscillator depends on the time constant R1C1.

SERVICING TECHNIQUES

Transistors may be damaged beyond repair by applying the incorrect polarity to the collector circuit or excessive voltage to the input circuit or by careless soldering techniques that overheat the transistor. Also the low-voltage electrolytic capacitors used in transistor circuits are easily damaged by reversed polarity connections or excessive voltage.

Because of the small physical size of transistors and their associated circuit components, small-sized tools are used. Small cutting pliers and needle-nose pliers are more useful than the conventional sizes. Narrow-blade screw drivers are more useful than larger, conventional types. A sharp-pointed, thin metal probe is also useful for cleaning solder from small openings or areas. Soldering is performed more satisfactorily with a small low-wattage soldering iron or pencil (35 to 40 watts) having a narrow point or wedge.

Always ground the frame of the soldering iron or gun to the chassis when soldering around transistors of transistor circuitry. Enough leakage voltage is present in most irons to cause damage to the transistor. Soldering with an ungrounded iron to a transistor and its associated circuitry held in one's hand may damage the transistor because the body has considerable capacitance to ground; the charging of this capacitance through the transistor can cause permanent damage. The junction is made very thin in order to operate at very high frequencies and is easily damaged by excess current. The same precaution (grounding the frame or chassis) is followed in applying any a-c operated test equipment.

When soldering transistor leads, the terminal lead is held with needle-nose pliers (or a heat shunt) positioned between the transistor body and the lead end. This arrangement allows the heat to travel into the pliers (or shunt), thereby diverting it away from the transistor. To make sure that all the heat is drawn away from the transistor, the pliers should remain securely on the lead for a short interval after removing the iron. Transistor leads should be relatively long and the soldering operation should be as short as possible. Low-temperature rosin core solder is the proper type to use.

Where transistor leads are stiff they may be plugged into appropriate type sockets. In this case the socket terminals should be soldered only when the transistor is out of the socket.

Transistor components are small and the connecting wires of resistors, capacitors, and coils are easily broken. These wires should be handled carefully both in installations and in replacements.

Transistors are biased in the forward, or low-resistance direction, of the base-emitter input circuits. These circuits are particularly vulnerable to any excess voltage over their rated value. This excess voltage can cause a current to flow that will overheat the transistor junction and permanently damage it. Proper operation depends on the crystal lattice structure and the impurity atoms that are present. Heat will distort the lattice and affect the behavior of the impurity atoms so that normal transistor action is seriously impaired. Excessive collector current will generate more heat in the relatively high impedance of the collector circuit and damage the transistor. Maximum allowable collector current depends on the ambient temperature. If the ambient temperature goes up, the maximum allowable collector current must be decreased (derated).

Both the value and the polarity of voltages to be applied to transistor circuits should be carefully checked before they are applied. It is important to know the type of transistor being used. The NPN types require positive collector voltages and negative emitter voltages (fig. 9-9); whereas, the PNP types require negative collector voltages and positive emitter voltages (fig. 9-10). Transistors should be inserted properly in the sockets before voltages are applied in order to avoid the possibility of transient current which might damage a transistor if allowed to flow. Similarly, voltages should be disconnected before removing transistors from sockets. If the magnitude of the collector current is in doubt a milliammeter of suitable range should be connected in series with the collector terminal, and battery potential should be applied gradually by means of a potentiometer.

When signal generators are used in testing transistor circuits the magnitude of the signal should be limited to a low value, especially in low-level input stages. Indirect coupling, for example (either capacitive or inductive) is preferred to direct coupling. The ungrounded lead from the output terminal of the signal generator can be terminated on the insulated portion of a capacitor or resistor in the circuit under test. Another method is to connect a coil to the output terminals of the signal generator and bring the coil into proximity with inductive elements of the circuit under test.

If an ohmmeter is used to check components in a transistor circuit, the range in use should not employ a battery of more than about 3 v. Higher voltages can damage the transistor. Electrolytic capacitors may give incorrect readings if the rule for polarities of the ohmmeter leads is not observed when connecting the leads to the capacitor terminals. The positive lead of the ohmmeter should connect to the positive terminal of the electrolytic capacitor, and the negative lead should connect to the negative terminal for correct indications to be possible.

The markings of the test leads do not necessarily indicate the polarity of the voltage on the leads when using a multimeter as an ohmmeter. The lead marked "common" may be either positive or negative with respect to the other leads depending on the design of that particular multimeter. Make sure by reference to the technical manual or by checking with another meter of known polarity. For example, when using the AN/PSM-4A as an ohmmeter, the lead marked "common" is positive with respect to the lead marked "volt-ohm-amps".

A transistor may be tested for its amplifying property by inserting it in a circuit like that of figure 9-9 and applying a suitable input signal, preferably from an audio signal generator. The input and output voltages are then measured with a vacuum-tube voltmeter and the gain is calculated. This value is then compared with that of a transistor known to be in good condition.

When the output of a transistor radio receiver is distorted or weak, the first thing to check is the battery. Its voltage should be checked with a vacuum-tube voltmeter or high resistance volt-ohmmeter while the set is turned on. If the battery is weak, its voltage will be down 20 percent or more. Before replacing the battery the resistance between the battery clips should be checked. For example, if the resistance measured with an ohmmeter (requiring a battery of not over 3 v) is 8000 ohms and the manufacturer's limits are 6000 to 15,000 ohms, it is safe to insert a new battery. An alternative method of some manufacturers is to indicate the allowable current drain of the battery instead of the resistance between the battery clip with the battery removed. To obtain an indication of the current that the battery supplies to the set, a milliammeter can be connected in series with the battery and either clip.

TRANSISTOR POWER SUPPLIES

Transistor circuits require relatively low amounts of power compared to electron-tube circuits, and for this reason, small batteries like the carbon-zinc types and the newer mercury types are used.

The mercury cell employs a negative electrode of zinc and a positive electrode of mercuric oxide. The electrolyte is a solution of potassium hydroxide. Local action is very small and the cell has a long shelf life.

The range of transistor power supply ratings is from the low voltage and current ratings of small flashlight cells to the medium ratings of 15-v to 30-v "B" batteries. The smaller voltage and current requirements pertain to subminiaturized equipments like hearing aids. The medium requirements pertain to transistor power amplifiers like those terminating in a loudspeaker.

Where transistor circuits are to be operated from an a-c source, the transistor power supply utilizes components that are smaller than those used in electron-tube power supplies. The components may include a 2-winding power supply transformer, a crystal rectifier, a voltage divider, and a low pass filter. The filter components may include resistors and electrolytic capacitors. To prevent damage to transistors due to excessive current, series limiting resistors are sometimes provided.

CHAPTER 10

MODULATION AND DEMODULATION

INTRODUCTION

MODULATION is the process by which the amplitude or frequency of a sine wave voltage (the CARRIER) is made to vary with time according to the voltage or current variations of another (MODULATING) signal. The carrier is usually of a higher frequency than the modulating signal.

DEMODULATION, or DETECTION, is the process by which the modulating signal is recovered from the carrier at the receiver. The modulating signals to be considered in this chapter will be in the audio frequency range.

Audio frequencies extend from 15 cycles per second to 20,000 cycles per second, as shown by the audio frequency spectrum in figure 10-1. The human voice extends from about 87 cps to 1175 cps. The violin has a range of from about 200 cps to 3000 cps, and the bass viol extends from about 40 cps to 250 cps. The pure tones of the piccolo extend to about 5000 cps. However, combinations of sound frequencies produce harmonics that extend up to 20,000 cps. These combinations of frequencies give to the speech or music the identifying characteristics that distinguish one person from another and one type of musical instrument from another.

For the following reasons it is not practical to transmit electromagnetic waves at audio frequencies—that is, without the use of a modulated r-f carrier: (1) The range of such transmission would be very limited because of the poor radiation efficiency of antennas at these low audio frequencies; (2) all transmitters would operate in the same frequency range, and therefore the signals could not be separated in the receivers; (3) the antennas would have to be excessively long to be in resonance at the middle frequency in the audio range (it follows that the antenna would be considerably out of tune at the end frequencies); and (4) the inductors and capacitors would have to be very large in order to produce resonance at these low frequencies.

These problems may be overcome by the use of a modulated r-f carrier at the transmitter end and a method of removing the modulation at the receiver end. The efficiency of radiation is thereby improved; each carrier with its associated modulation component is confined to a relatively narrow band in the r-f spectrum; and interference (while still a problem in certain instances) is not intolerable.

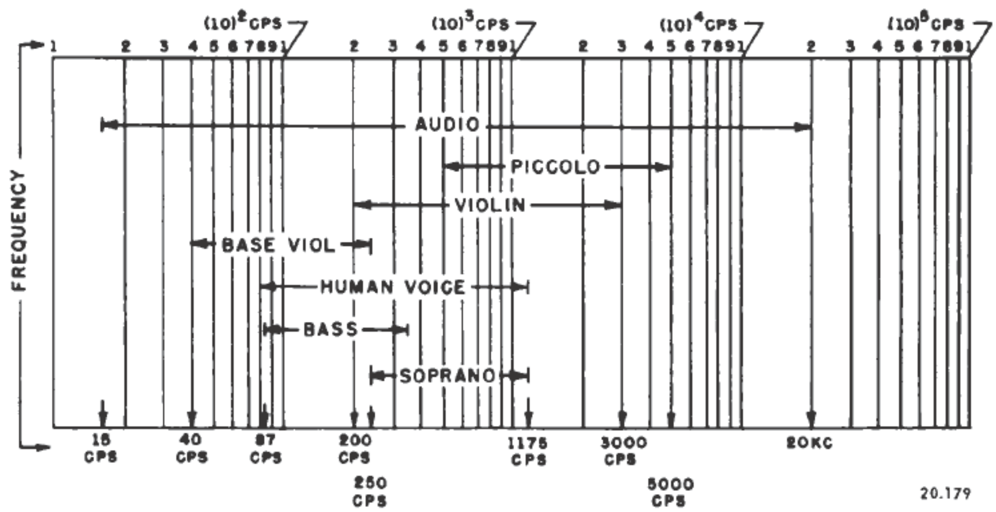


Figure 10-1.—Audio frequency spectrum.

AMPLITUDE MODULATION

Amplitude modulation (a-m) may be defined as the variation of the strength of the r-f output of a transmitter at an audio rate. In other words, the r-f energy is made to increase and decrease in power according to the audio (sound) frequencies. If the audio frequency is high, the radio frequency must vary in amplitude more rapidly than if the audio frequency were low. If the a-f tone is loud in volume, the r-f energy must increase and decrease by a larger percentage than if the a-f tone were soft. Thus, the r-f variation must correspond in every respect with the a-f variations.

Figure 10-2 indicates that the resultant wave with single-tone amplitude modulation consists of three separate waves. The lower sideband has a frequency equal to the difference between the modulation and carrier frequencies and is shown directly below the carrier in figure 10-2. The upper sideband has a frequency equal to the sum of the carrier and modulation frequencies and is shown above the carrier in figure 10-2. The carrier and the sidebands are not merely a mathematical abstraction; they may be separated from one another by filters and used individually.

In an a-m wave only the sidebands contain the intelligence to be transmitted; the audio frequency as such is not transmitted. Because the modulating frequencies are the information to be transmitted, as much power as possible should be put into the sidebands. In other words, the amplitude of the modulated carrier wave should be varied as much as possible. When the amplitude is carried completely to zero during the modulation cycle, the modulation is 100 percent; and the sidebands contain the maximum permissible amount of power, or one-half the carrier power, because modulation greater than 100 percent causes distortion.

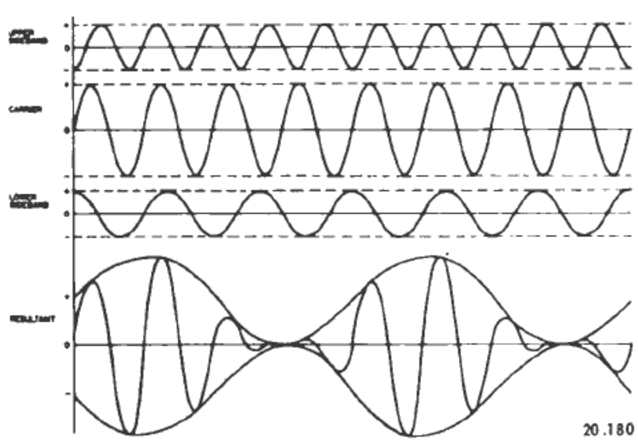


Figure 10-2.—Curves showing how the carrier and sidebands combine to produce an a-m wave with 100-percent modulation.

BANDWIDTH REQUIREMENTS

Because an a-m wave has sidebands on each side of the carrier, the transmission of information by amplitude modulation requires the use of a band of frequencies rather than a single frequency. Music may contain frequency components as high as 15,000 cps, so that music modulated upon a carrier would produce sideband components extending to 15,000 cycles on each side of the carrier frequency.

Local a-m broadcast stations are allocated a total bandwidth of only 10 kc (5 kc on each side of the carrier frequency) because of the large number of stations on the air. Since the total bandwidth is only 10 kc, audio frequencies above 5 kc cannot be transmitted without causing interference between stations.

Naval a-m communication equipments operate within the relatively narrow bandwidth required for voice or code transmission. Navy receivers such as the AN/SRR 12 and 13 have narrow bandwidth capabilities on the sharp selectivity settings as indicated in figure 10-3.

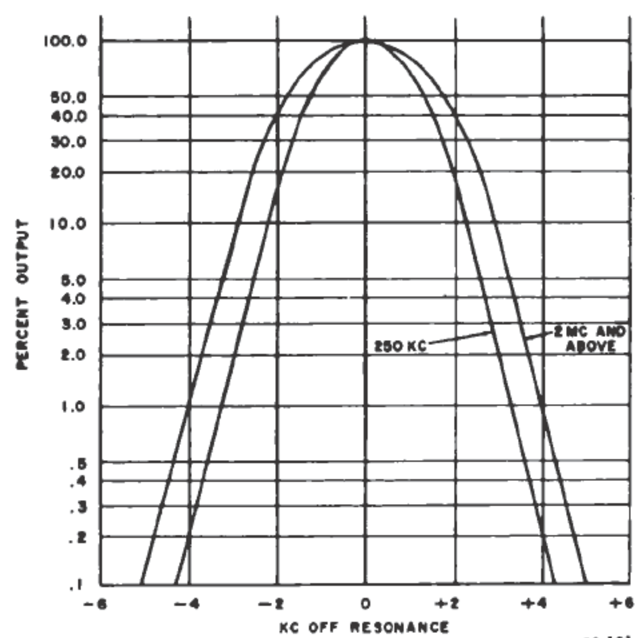


Figure 10-3.—Sharp selectivity of AN/SRR 12 and 13 receivers.

A narrow bandwidth permits the use of more selective circuits. These selective circuits reduce noise and the number of interfering signals.

PERCENTAGE OF MODULATION

The degree of modulation in an a-m wave is expressed by the percentage of maximum deviation from the normal value of the r-f carrier. The effect of such a modulated wave, as measured by receiver response, is proportional to the degree, or percentage, of modulation.

An a-m wave is shown in figure 10-4. The percentage of modulation may be determined by the equation

$$\text{percentage of modulation} = \frac{e_{max} - e_{min}}{2e_0} \times 100,$$

where e_{max} is the maximum instantaneous value of the r-f voltage across the transmitter tank circuit, e_{min} the minimum instantaneous peak value of the r-f voltage, and e_0 the maximum instantaneous value of the r-f voltage in the absence of modulation.

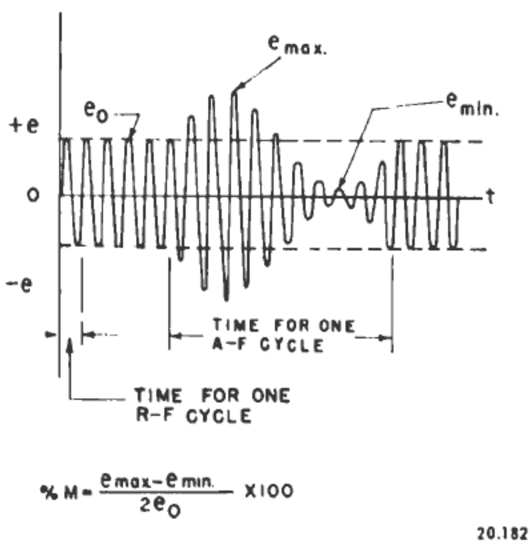


Figure 10-4.—Percentage of modulation.

It is important that the amplitude be varied as much as possible, because the output of a detector in a radio receiver varies with the amplitude variations of the received signal. Thus, a comparatively low-powered, but well modulated, transmitter often produces a stronger signal at a given point than a much higher powered, but poorly modulated, transmitter located the same distance from the receiver.

If modulation exceeds 100 percent there is an interval during the audio cycle when the transmitter is removed completely from the air. For example, in figure 10-5, the modulation is shown in excess of 100 percent. The audio voltage is assumed to have a peak value of 1500 volts, and the unmodulated carrier voltage has a peak value of 1000 volts. A more detailed analysis of overmodulation is included later in this chapter under "Plate Modulation."

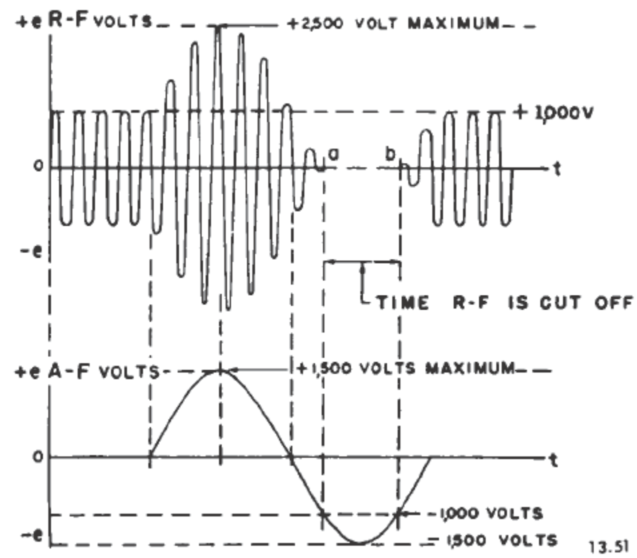


Figure 10-5.—Amplitude modulation in excess of 100 percent.

SYSTEMS OF AMPLITUDE MODULATION

A block diagram of an a-m radiotelephone transmitter is shown in figure 10-6. The top row of blocks indicates the r-f section. The next row of blocks indicates the a-f section; and the lower block indicates the power supply, which provides all d-c potentials.

The r-f section generates the high-frequency carrier radiated by the antenna. The methods by which the r-f signal is generated are treated in chapter 8.

The a-f section includes a speech amplifier that receives a few millivolts of a-f signal from the microphone (chapter 11) and builds it up to several volts at the input to the driver stage. This stage is made up of power amplifiers (chapter 7) that convert the signal into a relatively large voltage and appreciable current at the input to the modulator. The modulation

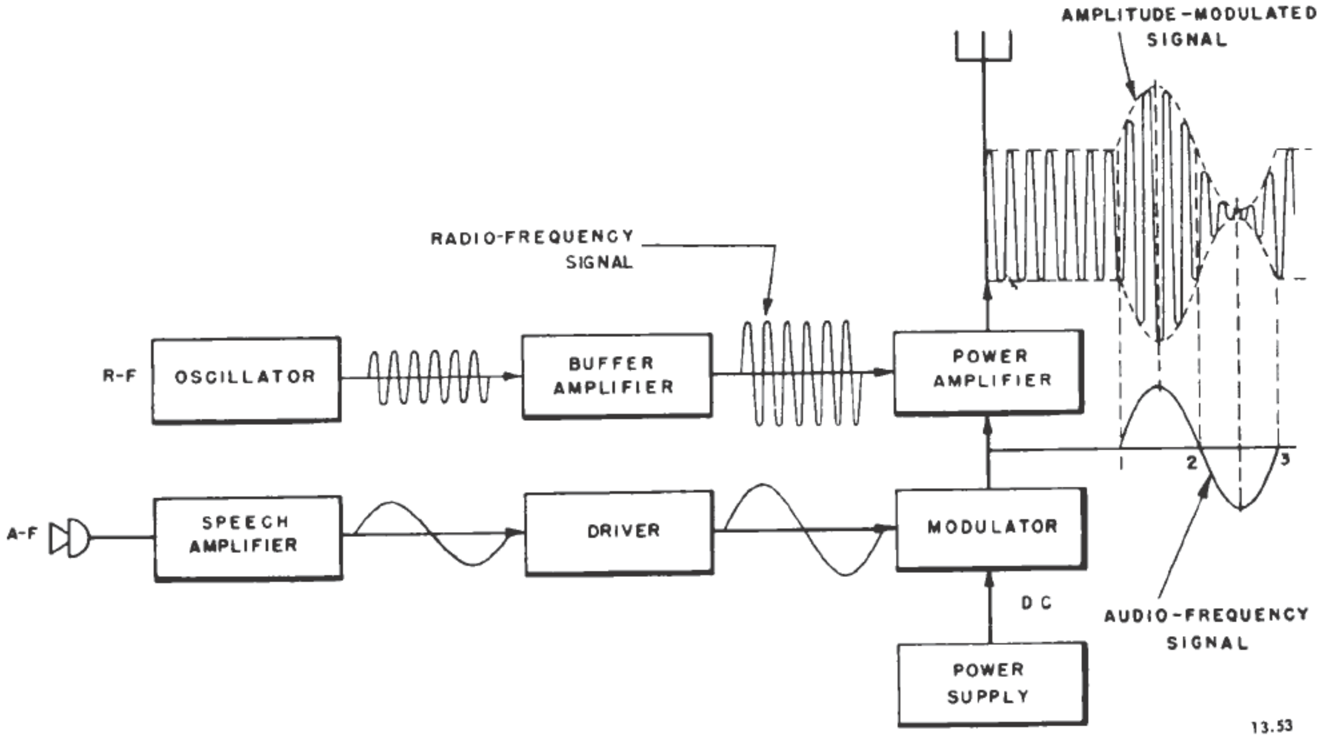


Figure 10-6.—An a-m radio telephone transmitter.

transformer is capable of handling considerable audio power. Its output is fed to the final r-f power amplifier in such a way as to alternately add to and subtract from the plate voltage of the r-f amplifier.

The result is that the amplitude of the r-f field at the antenna is gradually increased in strength during the time the a-f output is increasing the r-f power and gradually decreased in strength during the time the a-f output is decreasing the r-f power.

In other words, during the positive alternation of the audio signal (between point 1 and point 2 in figure 10-6), the amplitude of the r-f output wave is increased, and during the negative alternation (between point 2 and point 3) it is decreased. Amplitude modulation consists of varying the amplitude of the r-f antenna current (and r-f output wave) gradually over the relatively long a-f cycle. Thus, the r-f field strength is alternately increased and decreased in accordance with the a-f signal and at the a-f rate.

There are a number of methods of producing amplitude modulation, such as plate modulation, grid modulation, screen-grid modulation, and so forth, but the two most important are plate modulation and grid modulation.

Plate Modulation

One method of accomplishing amplitude modulation, called HIGH-LEVEL PLATE MODULATION, is shown in figure 10-7. It is called high-level modulation because the audio signal is injected in the plate circuit at a high level of plate voltage. The triode class-C power amplifier, V1, is tuned to a resonant carrier frequency of 1 megacycle. The plate current is 100 milliamperes, and the plate voltage supply is 1000 volts. The grid input driving voltage is produced by the r-f oscillator and buffer amplifier (used to isolate the oscillator from the power stage), shown by block diagram at the left of figure 10-7.

An audio signal of approximately 1000 volts (peak) having a sine waveform and a frequency of 1000 cps is produced by the a-f section shown by block diagram at the right of figure 10-7 and appears as an output of the modulation transformer, M. This audio output voltage is in series with the r-f voltage across the tank circuit, LC, and the plate power supply, B+. In this example, conditions are established for the modulation of the 1-mc r-f output voltage with a single a-f signal having a sine waveform.

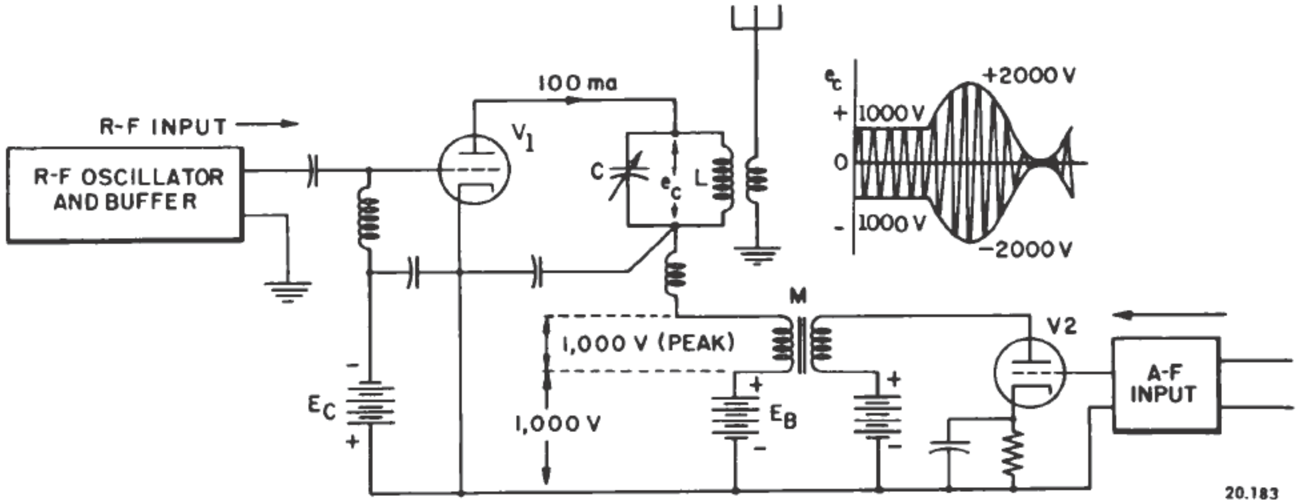


Figure 10-7.—High level plate modulation.

The 1-mc r-f carrier wave combines with the a-f signal to form sidebands. The frequency of the r-f voltage developed across the plate tank is 1000 times the audiofrequency. Thus, the time for one complete audio cycle is long enough to include 1000 cycles of r-f energy in the tank circuit, LC.

Before an audio-modulation signal is introduced, the r-f signal applied to the grid of V1 causes the triode to conduct periodically. During each conducting period, capacitor C charges. When the grid swings below cutoff, the triode stops conducting and the capacitor discharges through the coil.

The exchange of energy between the coil and capacitor accounts for the a-c voltage developed across the tank. In this example, the peak a-c voltage across the tank capacitor, C, which received its charge from the 1000-volt B supply, is approximately 1000 volts, neglecting losses.

Plate voltage varies between 1000 + 1000, or 2000 volts and 0 volts, as the capacitor voltage varies between + 1000 and -1000 volts. Plate voltage is above the B-supply voltage during that part of the cycle when the grid voltage is below cutoff and the triode is not conducting. Plate voltage is below the B-supply voltage during the time the grid is above cutoff and the triode is conducting. Energy to supply the tank circuit losses comes from the B supply, and the flywheel effect in the tank circuit accounts for the sine waveform of r-f voltage and current within the tank.

Now assume that the a-f voltage of sine waveform is introduced at M in series with the r-f tank and the B supply. Consider that portion of the a-f cycle in which the voltage is gradually rising according to the sine wave variation so that the polarity aids the voltage of the B supply. For 250 cycles of radio frequency (one quarter of one a-f cycle), the total voltage available for charging capacitor C in the plate-tank circuit is increasing from 1000 volts to a maximum of 2000 volts. Assume that the capacitor charges to 2000 volts. The plate voltage at the end of 250 r-f cycles is a maximum of 2000 + 2000, or 4000 volts. Thus, with 100-percent modulation, and on the positive audio peaks, the value of the peak-to-peak voltage across capacitor C is approximately double the peak-to-peak value it would have without modulation, or 4 times the B-supply voltage.

The tank voltage starts to decrease between the 250th r-f cycle and the 500th r-f cycle as the a-f voltage falls from a maximum of 1000 volts to zero volts during this period. From the 500th to the 750th r-f cycle the polarity of the a-f voltage is reversed, and reaches a maximum negative value at the 750th r-f cycle, and the tank capacitor charges up to the difference voltage between the 1000-volt B supply and the instantaneous a-f value. At the instant the a-f voltage is maximum negative (-1000 volts, at the 750th r-f cycle), the tank capacitor is not charging at all. The tank voltage is zero, and the r-f output energy momentarily becomes zero. Then the a-f voltage starts to fall from maximum negative to zero and the opposition that it offers to the power-supply voltage (B+) is reduced. The plate-tank capacitor starts charging again, and the r-f tank current again gradually increases during the subsequent 500 r-f cycles of operation.

The transmitting antenna is coupled to the tank, and the instantaneous antenna current increases and decreases in accordance with the tank-current variations. An r-f ammeter in the antenna circuit indicates the effective value of the current and not instantaneous values.

To understand the effective current indication corresponding to an a-m signal, study the distribution of r-f and a-f power in the tank circuit of figure 10-7. Before any a-f signal is injected, the tank circuit is assumed to be resonant. Plate current is 100 milliamperes and is composed of a rapid succession of nonsinusoidal pulses due to class-C operation. The power supplied to the r-f tank comes chiefly from the plate power-supply source. This source supplies 1000 volts d-c and 100 milliamperes, or a power supply of 100 watts. The peak value of the a-f voltage introduced at M is assumed to be 1000 volts, which produces in this instance 100-percent modulation. The a-f voltage is assumed to be of a sine waveform; and the peak value of the current is assumed to be 100 milliamperes. The current is the same in all parts of a series circuit, and the output winding of the modulation transformer, M, is connected in series between the tank and B+. The a-f power supplied to this circuit is equal to one-half the product of the maximum voltage and the maximum current. Thus, the power supplied by the a-f modulator is

$$\frac{1000 \times 0.1}{2} = 50 \text{ watts.}$$

The transmitter output power (neglecting losses) before modulation is 100 watts, and after modulation is 150 watts. If the equivalent antenna load resistance is assumed to be 100 ohms, the antenna current before modulation is

$$I = \sqrt{\frac{P}{R}} = \sqrt{\frac{100}{100}} = 1 \text{ ampere.}$$

When modulation occurs, the power in the preceding example increases to 150 watts. The antenna current is then

$$I = \sqrt{\frac{P}{R}} = \sqrt{\frac{150}{100}} = 1.224 \text{ amperes,}$$

which is an increase of 22.4 percent.

In the preceding example the modulation is 100 percent, which is the condition that exists when the audio power input is equal to one-half the r-f power input. When 100-percent modulation occurs, antenna current increases 22.4 percent above the unmodulated value.

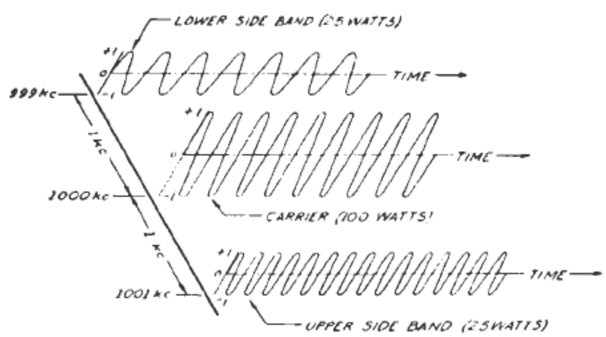
As previously stated, if the modulation exceeds 100 percent, there is an interval during the audio cycle when the transmitter output is zero. For example, in figure 10-5 the modulation is shown in excess of 100 percent. Assume that the audio input voltage of figure 10-7 has a maximum value of 1500 volts instead of 1000 volts. At the instant the audio voltage is maximum and adds to the power-supply voltage, the r-f voltage across the plate-tank curcuit rises to 1000 + 1500, or 2500 volts.

On the negative alternation of the audio cycle, when the audio voltage subtracts from the power-supply voltage, the plate voltage reverses and becomes negative with respect to ground during the interval a to b (fig. 10-5) when the audio voltage exceeds the voltage of the power supply. Tank-circuit oscillations cease during this interval because power cannot be supplied to the tank circuit while the plate of the triode is negative. This condition is called OVERMODULATION and results whenever the audio-modulation voltage exceeds the d-c voltage of the power-supply circuit. Overmodulation not only produces a distorted envelope, but also produces excessive interference to stations operating on adjacent channels because of the production of broad sidebands.

The circuit of figure 10-7 may be regarded as being composed of two sources of signal voltage and a common load in series. One source is the r-f generator, the other is the a-f generator, and the load is the resonant tank circuit coupled to the antenna load.

Whenever two or more voltages of different frequencies are introduced into a circuit having a common nonlinear load impedance (the class-C amplifier), these two voltages combine to produce two additional frequencies. These additional frequencies are called SIDEBANDS or SIDEBAND FREQUENCIES, as explained in the beginning of this chapter. They are the sum-and-difference frequencies of the r-f and a-f generators.

Thus, in the example under consideration, the r-f generator supplies 100 watts of power at a frequency of 1 mc; and the a-f generator (in the case of 100-percent modulation) supplies 50 watts of a-f power at a frequency of 1 kc. The tank circuit and the antenna coupled to it may be regarded as containing not only the r-f carrier current and its associated field (fig. 10-8) but also two closely associated currents and their fields—one current having a frequency of 1000 + 1, or 1001 kc, and the other having a frequency of 1000 - 1, or 999 kc.



20.184

Figure 10-8.—Carrier wave and its sideband frequencies.

The field radiated by the antenna into space at the speed of light may be regarded as being composed of the carrier and these two sidebands. The carrier has a power of 100 watts, neglecting circuit losses. The 50 watts of audio power are equally distributed in each sideband. Thus, the 999-kc sideband has a power of 25 watts, and the 1001-kc sideband also has a power of 25 watts. When the r-f carrier is 100-percent modulated, one-sixth of the TOTAL power is contained in each sideband.

During modulation, the r-f amplifier must handle peak currents that are twice the normal (unmodulated) value. Thus, during modulation an amplifier must be capable of handling up to four times the power it dissipates during steady intervals of unmodulated carrier output. For this reason, in a transmitter that is designed for both continuous-wave (c-w) and radiotelephone service, the modulated amplifier stages are ordinarily reduced in carrier power output for phone operation. Even if the power output is not reduced, the phone signal is weaker because it depends on the sideband power, which cannot exceed one-half of the carrier power.

The example under discussion has involved the simplest form of amplitude modulation. A single tone having sine waveform—like that produced by a tuning fork—constitutes the a-f input. Two sideband components accompany the r-f carrier (fig. 10-8); and when the modulation is 100 percent the power in each sideband is one-sixth of the total input power. When the audio signal becomes more complex, the number of different frequency components increases.

For each a-f component a sum-and-difference frequency is generated in the output of the transmitter. The power output of the audio modulator is divided equally among the sideband frequencies. Thus, as the number of sideband frequencies increases, the amount of power contained in any one sideband frequency is reduced. For example, for a given audio power input a signal containing speech does not have as much strength in any one frequency component as a code signal modulated by a single 1000-cps tone.

In most Navy transmitters the output power is reduced when going from MCW to voice, to reduce the possibility of exceeding the power capabilities of the transmitter on modulation peaks. These peaks do not occur when using MCW.

The usable range of a signal depends upon the ease with which it can be interpreted at the receiving station. A signal-to-noise level which renders a voice signal completely unintelligible may still permit the copying of an MCW signal by a skilled operator. For this reason a tone-modulated code signal has a slightly greater distance range than voice modulation, for the same transmitter.

There are various methods of producing an amplitude-modulated wave, but the most important one is high-level plate modulation. It has been shown that the audio signal is introduced in series with the plate circuit of a class-C amplifier where the plate voltage is at a relatively high level—the highest power level of the entire system. A class-C amplifier using high-level plate modulation is more efficient than a class-B or class-A amplifier that must be used with low-level modulation. Furthermore, class-C amplifiers are more easily adjusted and have proportionately less plate-power loss. For these reasons, high-level plate modulation is widely used.

Grid Modulation

Low-level grid modulation requires less bulky equipment than high-level plate modulation, with consequent savings in space, weight, and input power. This type of modulation has many Navy shipboard and aircraft applications. The a-f signal is applied in series with the grid circuit of the r-f power amplifier tube. The a-f signal varies the grid bias, which in turn varies the power output of the r-f amplifier. This variation in power output causes a modulated wave to be radiated. This method is known as GRID-BIAS MODULATION.

A circuit using grid-bias modulation is shown in figure 10-9. A modulation transformer is placed in series with the grid return lead of the r-f power amplifier. The a-f voltage from a modulating amplifier adds to or subtracts from the fixed grid-bias voltage and thus controls the output power from the r-f amplifier. The audio modulator tube supplying the modulation transformer, M, must be operated as a class-A amplifier. Varying the grid bias of the r-f stage does not require a great amount of power. It is difficult to achieve any large degree of modulation by this method, and the r-f carrier output power is about one-quarter of that of the plate-modulated transmitter; thus, the intelligibility of the signal is decreased.

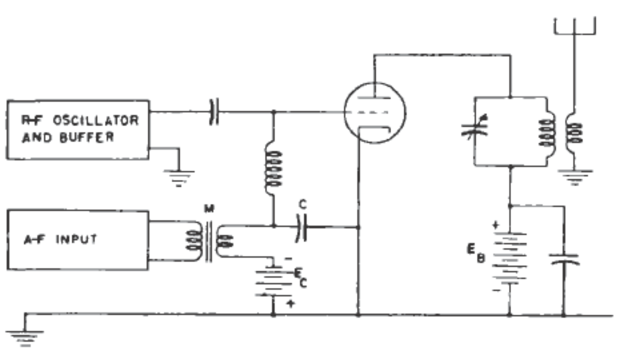


Figure 10-9.—Grid-bias modulation.

Tone Transmission

In tone transmission the r-f carrier is modulated at a fixed audio frequency of about 1000 cps. In some Navy transmitters the output of the transmitter is keyed in the same manner as for c-w transmission. An additional relay reduces the output of the tone generator to zero when the key is up and the r-f output is zero. Other Navy transmitters employ a continuous carrier and key only the modulating signal when operating MCW. Because a buzzer or audio oscillator is generally used as the tone source, the amplitude of the modulated r-f output wave is practically constant and the modulation can be 100 percent. Tone modulation has a slightly greater distance range than voice modulation for the same transmitter. However, the range of tone modulation is always less than that of c-w transmission on the same transmitter. Receiver tuning on tone signals is broader than in straight c-w reception, and code signals are less apt to be lost as a result of transmitter-frequency drift.

FREQUENCY MODULATION

The intelligence to be transmitted may be superimposed on the carrier in the form of changes in the frequency of the carrier. This type of modulation is called FREQUENCY MODULATION and has certain inherent advantages over conventional a-m transmission, particularly when static-free transmission is desired. However, in extensive tests conducted at the Naval Research Lab (NRL), it was found that for general Navy use amplitude modulation was in many ways more desirable than narrow-band frequency modulation. Nevertheless, the Naval Communication System uses a limited number of f-m transmitters and receivers. Aircraft altimeters use frequency modulation, as do some other radar and sonar equipments.

F-M SIDEBANDS

An a-m wave contains one upper and one lower sideband for each modulating frequency. An f-m wave may contain more than one pair of sideband frequencies for each modulating frequency.

As shown in figure 10-10, the f-m wave consists of a carrier wave of frequency f_o and associated sideband frequencies of $f_o \pm f_m$, $f_o \pm 2f_m$, $f_o \pm 3f_m$, and so forth, where f_m is the modulating frequency and f_o is the carrier frequency. Each line on each side of the center line represents a particular component of the f-m wave. The center line represents the carrier.

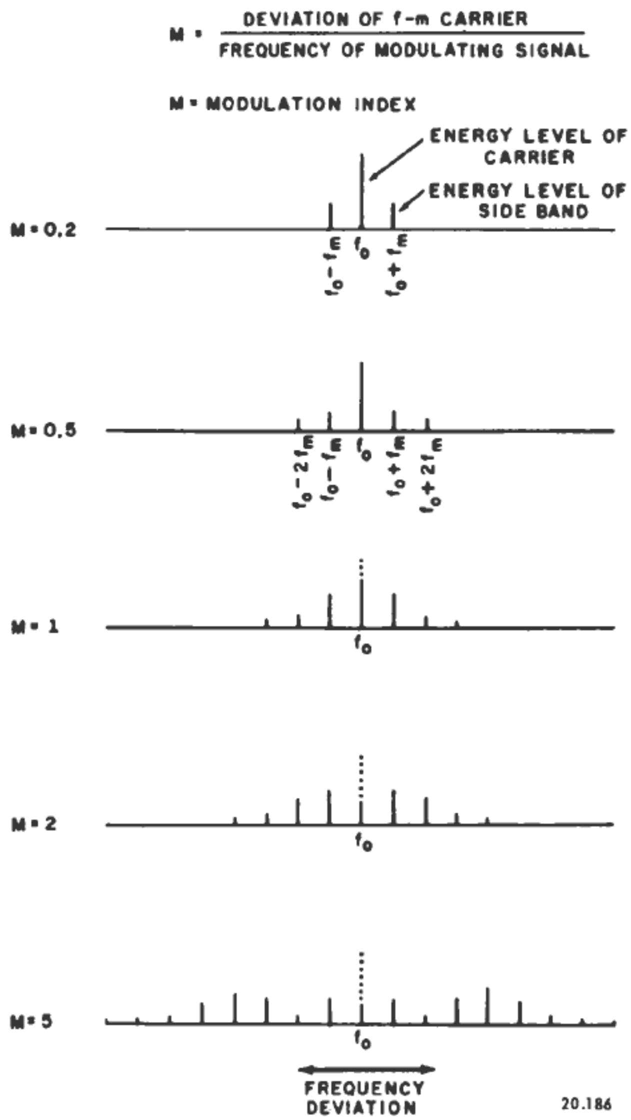


Figure 10-10.—Frequency and energy distribution for 5 values of modulation index of an f-m wave.

The lines to the right of center represent the upper sideband frequency components, and those to the left of center represent the lower sideband components. The lengths of the lines represent the energy levels of the various components. The horizontal distance between the center line and the last significant sideband (farthest removed from the center line) is proportional to the deviation frequency of the carrier, which depends in turn on the amplitude of the modulating frequency, f_m. The sideband frequency components are spaced an amount equal to the modulating frequency.

The number of sideband frequencies that contain sufficient energy to be important depends on the frequency deviation imposed on the carrier by the modulating signal. For example, if the modulating frequency, f_m, causes the carrier to deviate an amount equal to f_m and no more, the first pair of sidebands, $f_o \pm f_m$, is the only pair of importance. No additional energy is supplied an f-m wave during modulation. In contrast, additional energy is supplied an a-m wave during modulation. The energy of the f-m wave is redistributed during modulation, as shown in figure 10-10.

The carrier energy is reduced when the modulation index (M) exceeds 0.5. The modulation index is $\frac{f_d}{f_m}$, where f_d is the frequency deviation of the carrier and f_m is the modulating frequency. During modulation, energy is taken from the carrier and redistributed in the sideband components. For the cases M = 2 and M = 5, some of the sideband frequency components contain more energy than the carrier. During strong modulating signals the energy level of the carrier approaches zero.

Starting at the carrier and counting the sideband components consecutively in each direction, the upper odd-numbered sideband frequency component is 180° out of phase (the phase relations are not shown in the figure) with its associated lower sideband component. Thus the upper odd-numbered sideband component, $f_o + f_m$, is 180° out of phase with the odd-numbered lower sideband components, $f_o - f_m$. All even-numbered sideband frequency components are in phase with each other. Thus $f_o + 2f_m$ is in phase with $f_o - 2f_m$. The energy levels of a given pair equally spaced from the carrier are always equal.

As with amplitude modulation, the bandwidth for frequency modulation or phase modulation is determined by the number of sidebands associated with the carrier. An f-m wave with single-tone modulation theoretically has an infinite number of sideband pairs instead of just one pair, as in amplitude modulation. Fortunately, however, only a limited number of sidebands contain sufficient energy to be significant. An approximation of the number of significant sideband frequencies may be made by assuming that the important sideband components extend over a frequency range on each side of the carrier by an amount equal to the sum of the modulation frequency and the carrier frequency deviation.

For example, in figure 10-10 the frequency deviation of the carrier, f_o, is assumed to be 50 kc when the modulating frequency, f_m, is equal to 10 kc (M = 5) and the bandwidth on each side of the carrier is approximately 50 + 10, or 60 kc, making a total bandwidth of 60 x 2, or 120 kc. This bandwidth is an approximation derived from the "rule of thumb" given in the preceding paragraph. Because the sideband components are spaced an amount equal to the modulation frequency, the product of the number of significant sideband components and the modulation frequency is equal to the bandwidth. In this example there are 8 significant sideband frequencies on each side of the carrier, as shown in the illustration at the bottom of figure 10-10. Thus the bandwidth is 8 x 10, or 80 kc, on each side of the carrier, or a total of 2 x 80, or 160 kc. The bandwidth thus depends upon (1) the modulation frequency and (2) the total frequency deviation of the carrier.

Thus the bandwidth requirement of an f-m system is greater than twice the frequency deviation of the carrier by an amount equal to at least twice the modulating frequency. For most f-m systems the bandwidth is greater than that required for a-m systems. F-m transmission is made on higher carrier frequencies (88 to 108 mc for commercial channels) to obtain the necessary number of wideband channels.

For commercial high-fidelity broadcast transmission, 15 kc is the highest modulation frequency. The maximum frequency deviation of the carrier is limited by the Federal Communications Commission (FCC) to 75 kc on each side of the carrier frequency. The ratio of frequency deviation to modulation-frequency, or modulation index M, is therefore $\frac{75}{15} = 5$. As previously noted for a modulation index of 5, there are 8 important sidebands. Because the sidebands are spaced 15 kc apart, the bandwidth requirement is 8 x 15, or 120 kc, on each side of the carrier.

Although the FCC regulation limits the carrier shift to ±75 kc, some significant sidebands may extend beyond this frequency. A guard band of 25 kc on each side of the allowable frequency swing of ±75 kc is established to take care of most of the significant sidebands beyond the established limits.

Naval communications sets do not need high-fidelity response and therefore use modulating frequencies up to only a few thousand cycles. As mentioned earlier the important sideband components are approximately the sum of the modulation frequency and the carrier frequency deviation. With a carrier frequency deviation of 12 kc and a modulating frequency of 3 kc the approximate bandwidth is 2(3+12) = 30 kc for narrow band frequency modulation.

DEGREE OF MODULATION

To explain 100-percent modulation in an f-m system, it is desirable to first review the same condition for an a-m wave. As has been stated, 100-percent modulation exists when the amplitude of the carrier varies between zero and twice its normal unmodulated value. There is a corresponding increase in power of 50 percent. The amount of power increase depends upon the degree of modulation; and because the degree of modulation varies, the tubes cannot be operated at maximum efficiency continuously.

In frequency modulation, 100-percent modulation has a different meaning. The a-f signal varies only the frequency of the oscillator. Therefore, the tubes operate at maximum efficiency continuously and the f-m signal has a constant power input at the transmitting antenna regardless of the degree of modulation. A modulation of 100 percent simply means that the carrier is deviated in frequency by the full permissible amount. For example, an 88-mc f-m station has 100-percent modulation when its audio signal deviates the carrier 75 kc above and 75 kc below the 88-mc value, when this value is assumed to be the maximum permissible frequency swing. For 50-percent modulation, the frequency would be deviated 37.5 kc above and below the resting frequency.

SYSTEMS OF FREQUENCY MODULATION

A successful f-m transmitter must fulfill two important requirements—(1) the frequency deviation must be symmetrical about a fixed frequency, and (2) the deviation must be directly proportional to the amplitude of the modulation and independent of the modulation frequency.

There are several systems of frequency modulation that fulfill these requirements. A MECHANICAL MODULATOR employing a capacitor microphone is the simplest system of frequency modulation, but is seldom used. The two most important systems of frequency modulation are REACTANCE-TUBE and ANGLE, or PHASE-ANGLE, MODULATION. The main difference between these two systems is that in reactance-tube modulation the r-f wave is modulated at its source (the oscillator), while in phase modulation the r-f wave is modulated in some stage following the oscillator. The results of each of these systems are the same—that is, the f-m wave created by either system can be received by the same receiver.

Capacitor-Microphone System

The simplest form of frequency modulation is that of a capacitor microphone, which shunts the oscillator-tank circuit, LC, as shown in figure 10-11. The capacitor microphone is equivalent to an air-dielectric capacitor, one plate of which forms the diaphragm of the microphone. Sound waves striking the diaphragm compress and release it, thus causing the capacitance to vary in accordance with the spacing between the plates. This type of transmitter is not practicable (among other reasons, the frequency deviation is very limited), but it is useful in explaining the principles of frequency modulation. The oscillator frequency depends on the inductance and capacitance of the tank circuit, LC, and therefore varies in accordance with the changing capacitance of the capacitor microphone.

If the sound waves vibrate the microphone diaphragm at a low frequency, the oscillator frequency is changed only a few times per

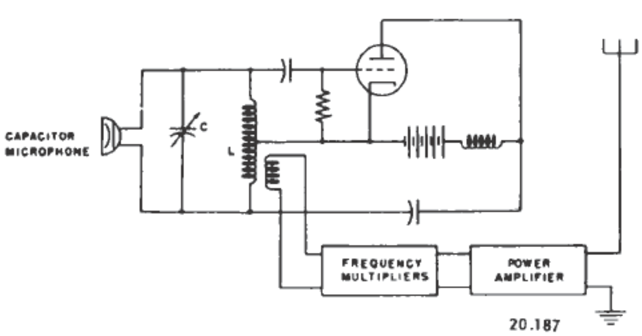


Figure 10-11.—F-m transmitter modulated by a capacitor microphone.

second. If the sound frequency is higher, the oscillator frequency is changed more times per second. When the sound waves have low amplitude, the extent of the oscillator frequency change from the no-signal, or resting, frequency is small. A loud a-f signal changes the capacitance a greater amount and therefore deviates the oscillator frequency to a greater degree. Thus, the deviation frequency of the oscillator tank depends upon the amplitude of the modulating signal.

In some military systems, in order to prevent interference between adjacent channel f-m transmitters, a bandwidth of 80 kc plus a guard channel of 20 kc is allowed for each transmitter. Thus, the strongest audio signal that can be used for modulating an f-m transmitter is limited to the value that causes a maximum deviation of 40 kc on each side of the average, or resting, carrier frequency. This allowance makes available a total of 80 kc, known as the CARRIER SWING, over which the frequency of any one transmitter may vary.

To increase the initial deviation frequency of the oscillator (which is greatly restricted in the case of the capacitor-microphone modulator) to a suitable value in the output, a system of frequency multiplication is used. The circuits used to accomplish this frequency multiplication are contained in the block diagram labeled "frequency multipliers" in figure 10-11. One method used a broadly tuned plate-tank circuit in a class-C amplifier. The tank is tuned to the second harmonic of the grid-input signal and thus builds up a tank current and output signal at double frequency. The output of the first doubler is fed into another similar doubler. Actually several stages may be used. For example, a 5-mc signal fed into a 3-stage amplifier using frequency doubling becomes a 40-mc signal at the output. An initial deviation of 1 kc produced by an audio-modulation signal becomes a frequency deviation of 8 kc at the output of the third doubler stage. Actually, frequency triplers or quadruplers may be used in some systems.

Reactance-Tube System

The reactance-tube system of frequency modulation is shown in figure 10-12. The reactance tube is an electron tube operated so that its reactance varies with the modulation

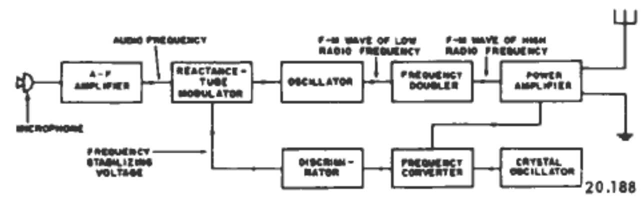


Figure 10-12.—Block diagram of a reactance-tube f-m transmitter.

signal and thereby varies the frequency of the oscillator stage.

In this circuit the reactance tube is connected in parallel with the oscillator tank and functions like a capacitor whose capacitance is varied in accordance with the audio signal, as in the capacitor-microphone system of frequency modulation. The frequency of the oscillator is thus changed, and the resulting f-m signal is passed through a frequency doubler to increase the carrier frequency and the deviation frequency. A power amplifier feeds the final signal to the antenna. The transmitter is kept with its assigned frequency limits by comparing the output of the transmitter with that of a standard crystal-controlled oscillator, and feeding back a suitable correcting voltage from a frequency-converter and discriminator (frequency-detector) stage.

The theory of operation of a reactance-tube circuit may be explained with the aid of a figure 10-13. The reactance tube, V1 (fig. 10-13,A), is effectively in shunt with the oscillator tank, LC, and the phase-shift circuit, R_gC1. The capacitive reactance of the capacitor is large compared with the resistance of the resistor; and the current, i, in this circuit leads the voltage, e_p, across the circuit by approximately 90 degrees. The voltage, e_p, is the alternating component of the plate-to-ground voltage appearing simultaneously across the reactance tube, the phase-shift circuit, and the oscillator tank.

The coupling capacitor, C2, has relatively low capacitive reactance to the a-c component of current through it, and at the same time it blocks the d-c plate voltage from the phase-shift circuit and the tank. The reactance tube receives its a-c grid-input voltage, e_g, across R_g. This voltage is the IR drop across R_g and is in phase with plate current i_p. This relation is characterisitic of amplifier tubes.

Because e_g is in phase with both i and i_p and e_g leads e_p by approximately 90°, both

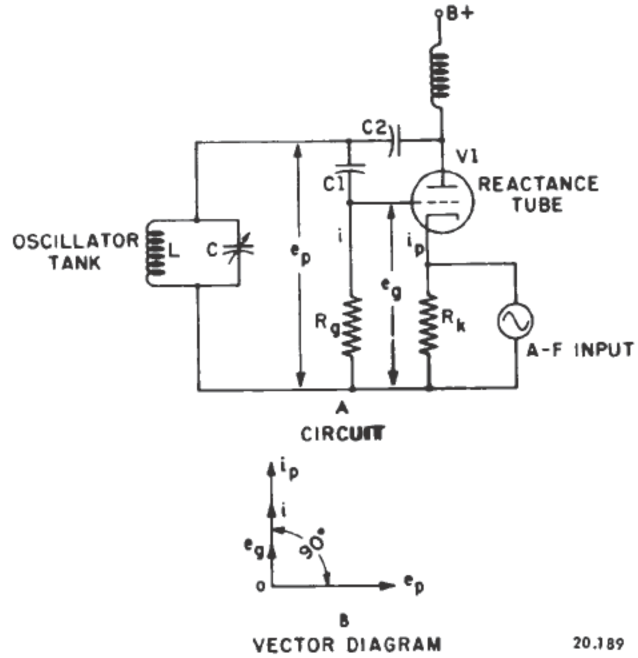


Figure 10-13.—Frequency modulation with a reactance-tube modulator.

i and i_p lead e_p by approximately 90 degrees. These relations are shown in the vector diagram in figure 10-13,B. Both i and i_p are supplied by the oscillator tank circuit, and because both are leading currents with respect to the tank voltage, e_p, they act like the current in tank capacitor C. Therefore, the EFFECT of these currents on the frequency of the tank is the same as though additional capacitance were connected in parallel with it.

Consider now the effect of introducing an audio signal across R_k. With zero audio voltage, r-f plate current i_p is a succession of rapid pulses of constant amplitude, and the oscillator tank operates at a constant frequency, called the NO-SIGNAL, or RESTING FREQUENCY. When the audio voltage rises with a polarity that swings the cathode negative with respect to the grid, the pulses of plate current gradually increase in amplitude. This leading r-f plate current is drawn through the oscillator tank and is equivalent to an increasing value of tank capacitance. Thus, the oscillator frequency is lowered.

Conversely, when the audio signal swings the grid of the reactance tube negative with respect to the cathode, the r-f plate-current pulses gradually decrease in amplitude and the oscillator frequency increases.

The frequency of the a-f signal determines the number of times per second that the oscillator-tank frequency changes. On the other hand, the amplitude of the a-f signal determines the extent of the oscillatory-frequency change—that is, the deviation frequency. Thus, the reactance tube with its audio-signal input produces an f-m output having the same characteristics as that of the capacitor microphone modulator.

Phase Modulation

Any process that changes the instantaneous frequency of the r-f energy already generated at a constant frequency is referred to as angle, or phase-angle, modulation. All radio modulating processes are based on changing the r-f carrier wave in some respect. The variation normally is directly proportional to the instantaneous value of the modulating voltage. When the instantaneous frequency of the carrier is varied in a direct relation to the modulating wave, the result is frequency modulation. If the instantaneous phase of the carrier is varied by an electrical angle directly proportional to the instantaneous modulating voltage, phase modulation is obtained. Varying the carrier frequency also changes the instantaneous phase relation of the carrier frequency to its own fixed unmodulated state. Likewise, varying the carrier phase changes the carrier frequency.

Thus, frequency modulation and phase modulation are basically the same. In fact, frequency modulation is equivalent to phase modulation in which the phase-angle variation is inversely proportional to the modulation frequency. An f-m signal is one in which the carrier deviation from the resting frequency is proportional to the amplitude of the modulating signal and is independent of the modulating frequency. A pure phase-modulated signal is one in which the carrier deviation is proportional to both the amplitude and frequency of the modulating signal.

In all cases the carrier variation occurs at a rate of change equal to the frequency of the modulating wave. For example, a 1000-cps tone changes the carrier frequency plus and minus 75 kc 1000 times per second in a broadcast f-m system at maximum modulation.

A phase-modulation system is shown in the block diagram in figure 10-14. The transmitter oscillator is maintained at a constant frequency

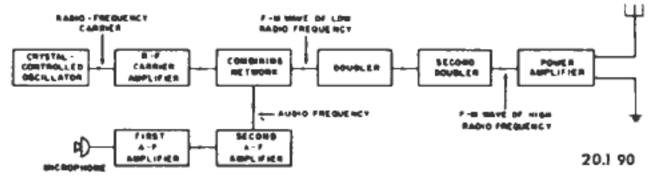


Figure 10-14.—Block diagram of a phase-modulated f-m transmitter.

by means of a quartz crystal. This constant-frequency signal passes through an amplifier that increases the amplitude, or energy level, of the wave. The audio signal is applied to the r-f carrier by means of a combining network. The output of the combining network is fed into a series of class-C amplifiers, the plate circuits of which are tuned to a multiple (doubling is indicated in this figure) of the frequency of the grid input. The output of these frequency multipliers is fed to a power amplifier that couples the f-m signal to the antenna.

A diagram of the combining network in which the phase shift is accomplished is shown in figure 10-15, A. The r-f and a-f voltages are applied across the grid-input voltage divider, which consists of R2, R1, and R4. The triode plate load is a broadly tuned L-C tank. The r-f signal of constant frequency and amplitude appears across R2 as e_g. As the instantaneous value of the audio signal varies through each audio cycle, the triode bias is increased and decreased at the a-f rate because of the a-f voltage that appears across resistors R4 and R1. Consequently, the triode gain is varied in accordance with the a-f signal. (This circuit is similar to the one used in grid-bias amplitude modulation discussed earlier in this chapter (fig. 10-9) except that changes in amplitude of the carrier are not utilized; instead, changes in phase of the carrier are developed and utilized.)

Now consider how this varying gain is translated into phase shift. The instantaneous plate-load voltage, e_o (fig. 10-15, B), is the resultant of two r-f voltages in the triode plate circuit. These two r-f voltages are (1) that portion of the grid-input signal that is coupled to the plate circuit by means of the grid-plate capacitance of the triode (this voltage is designated e_1) and (2) the grid-input signal amplified in the triode plate by normal amplifier action (this voltage is designated e_2).

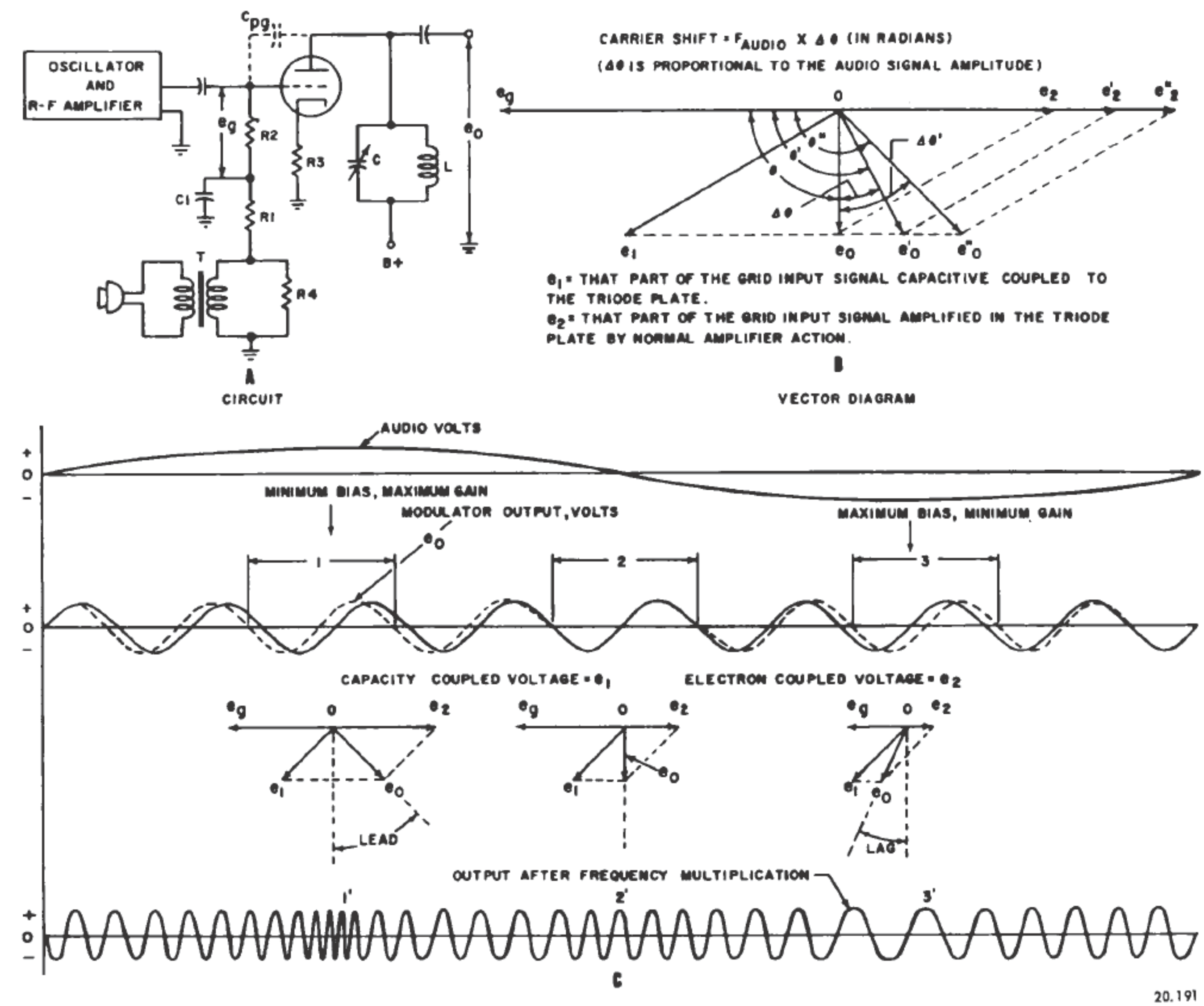


Figure 10-15.—Phase-modulated network.

The resultant of e_1 and e_2 is indicated in the vector diagram as e_o. In the same diagram, e_g is the grid-input r-f signal. The triode-amplifier voltage, e_2, is relatively low because of inverse feedback obtained by omitting the usual cathode-bypass capacitor across R3. Voltage e_1 leads e_g and is of less amplitude than e_g because the interelectrode capacitive reactance acts in series with the plate load and causes a leading current to flow through it. The plate load is resonant at the oscillator frequency and hence acts as a pure resistance. Thus e_1 across the load is in phase with the leading current through the load and leads e_g by some angle less than (90°) depending on the magnitude of the interelectrode capacitive reactance and the plate-load resistance.

At the time the audiofrequency swings the triode bias to the MAXIMUM the triode gain is a MINIMUM, and e_2 is relatively small. For this condition e_o leads e_g by angle θ. The amplified voltage, e'_2, represents the condition existing when the a-f signal swings the triode bias to a minimum, and the tube gain is higher than before. Therefore, voltage e'_2 is larger and combines with e_1 to produce the resultant plate-load voltage, e'_o, which leads e_g by angle $\theta + \Delta\theta$. The resultant voltage, e'_o, undergoes a change in phase angle, $\Delta\theta$, with respect to e_g in accordance with the change in triode bias, gain and a-f instantaneous values.

The difference in the angles θ and θ'—that is, $\Delta\theta$—is the change in the PHASE-SHIFT ANGLE of e_o, and is a factor of carrier swing.

When voltage variations, e_o, e'_o, and so forth, are applied to a tuned circuit, a smooth wave having both positive and negative alternations is formed as the result of the flywheel effect in the tank circuit. This wave has the same varying time interval between positive peaks as the applied voltage variations, and therefore its frequency is shifted in accordance with the audio-modulation signal during the time the phase angle is changing.

The variation in phase of the carrier output voltage, e_o, with audio modulation is illustrated in figure 10-15, C. As the audio signal voltage varies the gain of the modulating stage from maximum to minimum, the carrier output voltage is shifted in phase from leading to lagging the zero-signal position.

On the positive half of the audio cycle, e_o is advanced in phase ahead of the zero-signal position. The angle of lead is equivalent to a slight increase in frequency of e_o and a corresponding decrease in the time for each carrier cycle as indicated at 1 compared with the time at 2.

On the negative half cycle of the audio signal the carrier voltage lags in phase behind the zero-signal position. The lag is equivalent to a slight decrease in frequency of e_o and a corresponding increase in the time for each carrier cycle as indicated at 3 compared with the time at 2.

The output voltage of the phase modulated carrier is fed through several frequency multiplier stages and the output waveform then contains many more cycles as illustrated in the lower curve of the figure. Frequency multiplication increases the carrier frequency and the changes in the carrier as a result of modulation. Thus the frequency is increased to a higher value in the vicinity of 1' and decreased to a correspondingly lower value at 3' compared with the no-signal value at 2'.

In a phase-modulation system such as the one being discussed the carrier shift is proportional to the product of the audio frequency and the phase-shift angle. It is therefore necessary to introduce an action that will prevent a signal that changes in audio frequency, but remains at constant amplitude, from influencing carrier swing. Otherwise low-frequency audio components would be underemphasized at the receiver. Only the AMPLITUDE of the modulating signal and NOT its instantaneous frequency should influence the extent of the frequency swing of the carrier. This action may be accomplished by introducing a preemphasis circuit such as the one composed of R1C1 in figure 10-15, A.

When the audio signal of constant amplitude is decreased from a high frequency to a lower frequency, the audio voltage across C1 is increased in amplitude. Therefore, the tube-bias swing and the change in tube gain cause the normal amplifier component of output voltage to be increased from e'_2 to e''_2. Carrier shift is proportional to the product of the audio frequency, f_{audio}, and the phase-shift angle, $\Delta\theta$, as indicated in figure 10-15, B. Then an increase in the phase-shift angle from $\Delta\theta$ to $\Delta\theta'$ compensates for the decrease in audiofrequency, f_{audio}. Thus, the product of f_{audio} and $\Delta\theta$ remains constant, and the carrier swing is now independent of the audiofrequency. The output f-m signal from the phase-modulated transmitter is therefore similar to that of a frequency-modulated transmitter. Thus, an f-m receiver performs equally well on either output.

DEMODULATION OF WAVES

DEMODULATION, or DETECTION, is the process of recovering the intelligence from a modulated wave. When a radio carrier wave is amplitude-modulated, the intelligence is imposed on the carrier in the form of amplitude variations of the carrier. The demodulator of an amplitude-modulated (a-m) wave produces currents or voltages that vary with the amplitude of the wave. Likewise, the frequency-modulation (f-m) detector and the phase-modulation (p-m) detector change the frequency variations of an f-m wave, and the equivalent phase variations of a p-m wave, into currents or voltages that vary in amplitude with the frequency or phase changes of the carrier.

The instantaneous amplitude, e_o, of the carrier may be represented as

$$e_o = E_o \sin (2\pi f_o t + \theta),$$

where E_o is the maximum amplitude of the original carrier, f_o the frequency of the carrier, and θ the phase angle (for a-m signals, θ may be considered as zero). One or more of the independent variables (those on the right-hand side of the equation) may be made to vary in accordance with the modulating signal to produce a variation in E_0. However, the general practice is to vary only one of the values—E_0 (for a-m), f_0 (for f-m), or θ (for p-m)—and to prevent any variation in the others.

The detector in the receiver must therefore be designed so that it will be sensitive to the type of modulation used at the transmitter, and insensitive to any other.

Most Navy equipment is designed for amplitude modulation. A clear understanding of the mechanism of a-m detection is therefore very important.

A-m modulators and demodulators are nonlinear devices. A NONLINEAR DEVICE is one whose current-voltage relation is not a straight line. Because the ratio of current to voltage is not constant, the device has a nonlinear impedance—for example, one of the electron-tube detectors to be considered later—the average output current is the difference between each successive positive and negative swing of the output signal current, as shown in figure 10-16. The average output (signal component) follows the envelope of the incoming modulated wave more or less closely, depending on the shape of the nonlinear curve. Because the envelope of the incoming a-m wave contains the desired audio frequency, a nonlinear device demodulates the a-m wave.

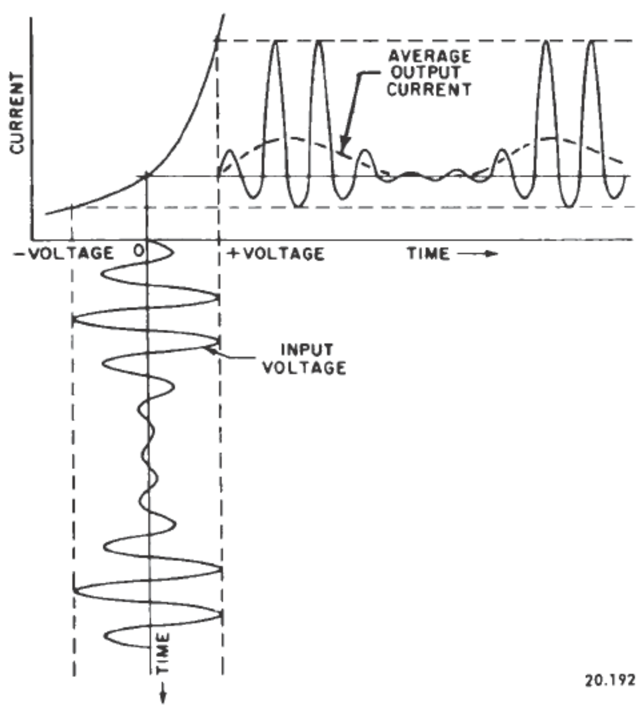


Figure 10-16.—Results of impressing an a-m wave on a nonlinear device.

For an understanding of the differences in the output frequencies of the various detectors it is necessary to examine the frequencies involved in both modulation and demodulation.

COMPARISON OF AMPLITUDE MODULATION AND DEMODULATION

If at the transmitter an r-f carrier and a single-frequency audio-modulating signal of sine waveform are impressed on a LINEAR device, the output waveform from the linear device will contain the same r-f and a-f signal frequencies. The tuned r-f amplifiers in the transmitter will amplify the r-f carrier, but will eliminate for all practical purposes the a-f component. Under these circumstances only the carrier will be radiated; and it will be ineffective in "carrying" the intelligence component.

A very different result is obtained if an r-f carrier and a single-frequency audio-modulating signal of sine waveform are impressed on a NONLINEAR device. In this instance distortion is introduced and, as a result, additional frequencies are produced. In addition to the original frequencies, sum-and-difference frequencies are generated, and a zero-frequency, or d-c component, is added. The tuned circuits at the transmitter now respond to the carrier and the upper and lower sidebands; but, as before, the a-f modulating signal is discriminated against. However, this a-f component is replaced, or generated, by the demodulator in the receiver.

In the receiver the carrier and the two sidebands are impressed on a second nonlinear device called the DEMODULATOR. If the demodulator has an IDEAL nonlinear curve it will distort the incoming waveform (the positive halves of the cycle will be different from the negative halves). Therefore, in addition to the r-f carrier and the two sidebands, the SIGNAL FREQUENCY (which is the difference between the upper sideband and the carrier or the difference between the carrier and the lower sideband) and a zero-frequency (or d-c component) will be produced. As shown in chapter 14, the d-c component may be used for automatic volume control.

If the demodulator used in the receiver does not have an ideal nonlineear curve, but has a PRACTICAL realizable curve such as the square-law curve, additional frequencies will be produced. These frequencies will be harmonics of all frequencies present in the input. They are produced because input voltages having larger amplitudes are distorted differently from input voltages having smaller amplitudes. The r-f harmonics may be filtered in the output of the demodulator, but the a-f harmonics are not easily eliminated.

Thus, modulation and demodulation are essentially the same in that the waveform is distorted in each case and new frequencies are produced.

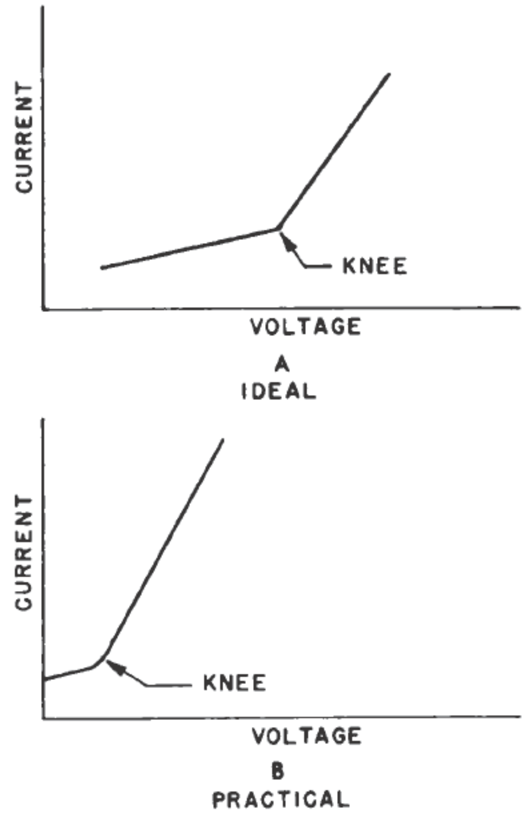
TYPES OF A-M DETECTORS

Detectors are classified according to the shape of their current-voltage (characteristic) curve. If the curve is smooth, as in figure 10-16, the detector is called a SQUARE-LAW DETECTOR. It is called a square-law detector because, for a first approximation, the output voltage is proportional to the square of the effective input voltage.

If the current-voltage curve of the detector is shaped like an obtuse angle, as in figure 10-17,A the curve is still nonlinear because of the abrupt change in shape at the knee. Because the detector action takes place on the linear portions of the curve on both sides of the knee, this type of detector is called a LINEAR DETECTOR. It is called a linear detector merely to distinguish it from a square-law detector. Both the square-law detector and the linear detector are actually NONLINEAR devices.

The rectified output voltage of a linear detector is proportional to the input voltage. The output of a square-law detector is proportional to the square of the input voltage.

Detectors are also described as power detectors or as weak-signal detectors. If a detector is to handle r-f carrier voltages having amplitudes greater than approximately 1 volt it is called a POWER DETECTOR. If the input signal strength is less than this amount the detector is called a WEAK-SIGNAL DETECTOR. Thus, the approximate value of 1 volt is the dividing line between the two detectors. Because a linear detector cannot be obtained with a sharp discontinuity (angle) at the knee, weak signals are always detected on a curved portion of the characteristic, as shown in figure 10-17,B. Thus, weak-signal detectors are always of the square-law type. Power detectors may be either linear or square-law, depending on the application.



20.193
Figure 10-17.—Linear detectors.

DIODE DETECTOR

The diode detector (fig. 10-18,A) is one of the simplest and most widely used detectors, and has nearly an ideal resistance characteristic. Diodes have a point of sharp transition between the conducting (forward) and nonconducting (reverse) directions and therefore make good detectors.

Early radio receivers used a crystal detector that was made of galena, a mineral compound of sulfur and lead. The end of a short length of fine wire touched the surface of the crystal and was held against it by the pressure of a spring. The wire could be adjusted manually. When the operator found a sensitive spot the crystal rectified the signal and the operator heard the signal in his earphones.

Later applications of the crystal detector are found in radar and television. In these

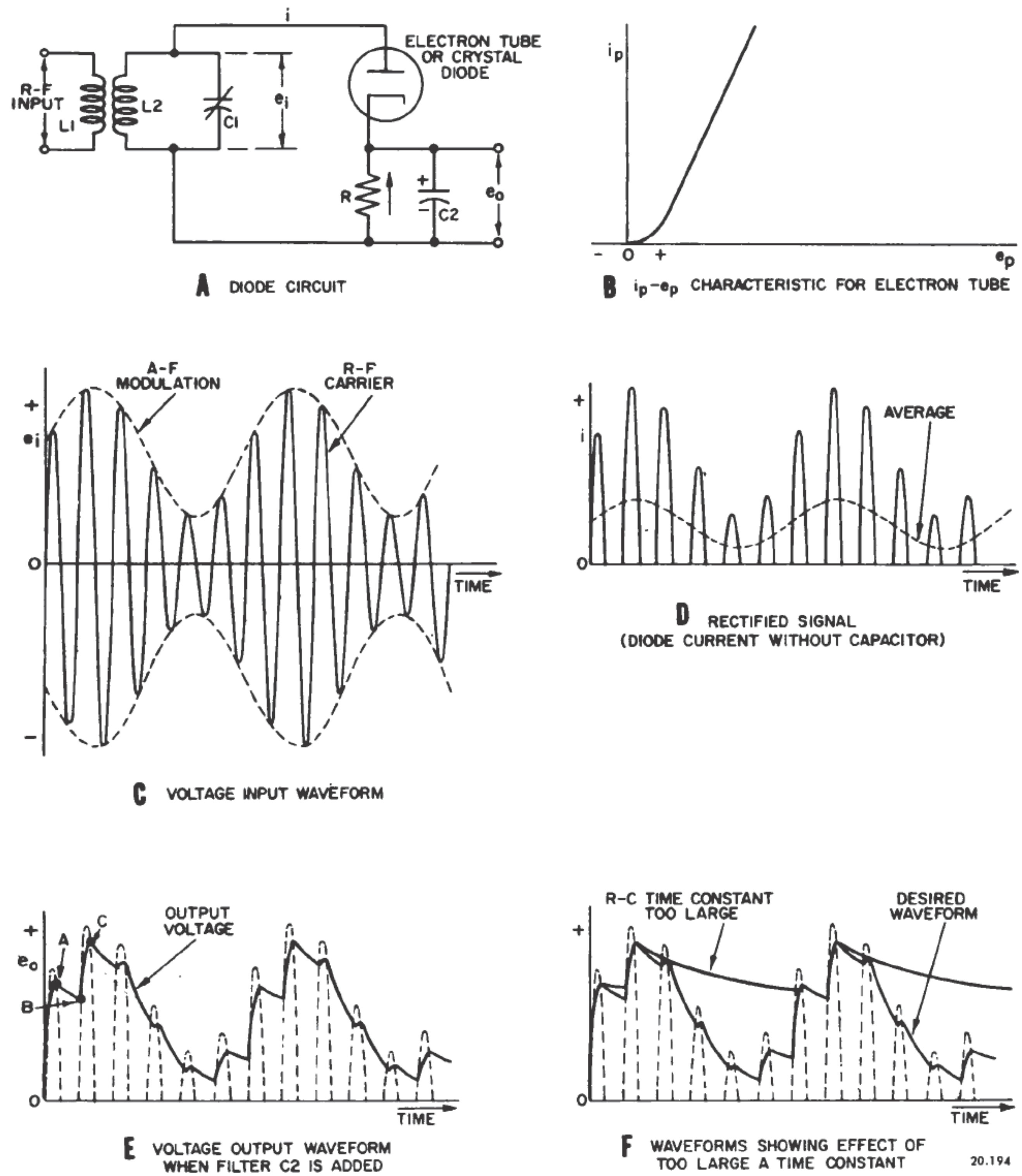


Figure 10-18.—Diode detector and waveforms.

applications the crystal diode is of a different form than the older type. It comprises a small crystal with a fine wire point contact bearing on the crystal surface. The components are contained in a sealed ceramic cartridge. A germanium crystal diode is described in chapter 2 (fig. 2-12) of this training course. The advantage of the crystal diode as a rectifier of high frequency signals is the relatively low shunting capacitance between the crystal and the point contact, compared with the interelectrode capacitance between the plate and cathode of the electron tube diode.

The slight bend in the lower portion of the i_p-e_p characteristic curve for the electron tube diode (fig. 10-18,B) results from contact potential.

Contact potential is the potential difference existing between the surfaces of metals of different electron affinities that are in direct contact or are connected by means of an external circuit. This is true of the metal elements of electron tubes; and the voltages developed may be of the order of 1 volt or more. In the case of the diode, contact potential keeps the plate current from decreasing to zero when the plate voltage approaches zero. The current, however, is very slight, and the characteristic curve is generally shown as zero when the applied voltage is zero. Nevertheless, on low signal voltage, plate current does not increase as rapidly as it does after the contact potential has been exceeded.

Because the diode characteristic is nearly straight on both sides of the knee, the diode detector is a linear detector. However, with weak signals, the output of the detector follows the square law because weak signals force the operation to take place on the lower, curved portion of the characteristic curve (fig. 10-18,B). Because the diode detector normally handles large input signals with minimum distortion, it is classified as a power detector.

The modulated signal voltage (fig. 10-18,C) is developed across the tuned circuit, L2C1, of the detector stage. Signal current flows through the diode only when the plate is positive with respect to the cathode—that is, only on the positive half cycles of the r-f voltage wave.

The rectified signal flowing through the diode (fig. 10-18,D) actually consists of a series of r-f pulses and not a smooth outline or envelope. The average of these pulses, with little or no filtering, does increase and decrease at the a-f rate, as shown by the dotted line. Therefore, there is an audio voltage output even if no filtering is employed. However, some stray capacitance exists, and consequently some r-f filtering takes place.

If a capacitor (C2 in figure 10-18,A) of the proper size is used as a filter, the output voltage of the detector is increased and more nearly follows the envelope. On the first quarter cycle of applied r-f voltage, C2 charges up to nearly the peak value of the r-f voltage (point A in figure 10-18,E). The small voltage drop in the tube prevents C2 from charging up completely. Then as the applied r-f voltage falls below its peak value, some of the charge on C2 leaks through R, and the voltage across R drops only a slight amount to point B. When the r-f voltage applied to the plate on the next cycle exceeds the potential at which the capacitor holds the cathode (point B), diode current again flows and the capacitor charges up to almost the peak value of the second positive half cycle at point C.

Thus the voltage across the capacitor follows very nearly the peak value of the applied r-f voltage and reproduces the a-f modulation. The detector output, after rectification and filtering, is a d-c voltage that varies at an audio rate, as shown by the solid line in figure 10-18,E. The curve of the output voltage across the capacitor is shown somewhat jagged. Actually, the r-f component of this voltage is negligible and, after amplification, the speech or music originating at the transmitter is faithfully reproduced.

The correct choice of R and C2 (fig. 10-18,A) in the diode-detector circuit is very important if maximum sensitivity and fidelity are to be obtained. The load resistor, R, and the plate resistance of the diode act as a voltage divider to the received signal. Therefore, the load resistance should be high compared with the plate resistance of the diode so that maximum output voltage will be obtained. The value of C2 should be such that the RC time constant is long compared with the time of one r-f cycle. This is necessary because the capacitor must maintain the voltage across the load resistor during the time when there is no plate current. Also, the RC time constant must be short compared with the time of one a-f cycle in order that the capacitor voltage can follow the modulation envelope.

The values of R and C2 therefore place a limit on the highest modulation (audio) frequency that can be detected. Figure 10-18,F, shows the type of distortion that occurs when the RC time constant is too large. At the higher modulation frequencies the capacitor does not discharge as rapidly as required, and negative peak clipping of the audio signal results.

The efficiency of rectification in a diode is the ratio of the peak voltage appearing across the load to the peak input signal voltage. The efficiency increases with the size of R compared with the diode plate resistance, because R and the diode are in series across the input circuit and their voltages divide in proportion to their resistance. With audio frequencies, a large value of R may be used (of the order of 100,000 ohms), and consequently the efficiency is relatively high (95%). When high modulation frequencies, such as those used in television, are necessary the value of R must be reduced to keep the RC time constant low enough to follow the envelope. Consequently the efficiency is reduced.

The diode detector can handle large signals without overloading, and it can provide automatic-volume-control voltage without extra tubes or special circuits. However, it has the disadvantage of drawing power from the input tuned circuit because the diode and its load form a low-impedance shunt across the circuit.

Consequently, the circuit Q, the sensitivity, and the selectivity are reduced. The interelectrode capacitance of the diode detector limits its usefulness at high carrier frequencies, and the bend in the lower portion of current-voltage characteristic indicates that it distorts on weak signals. Therefore considerable amplification is needed before detection.

GRID-LEAK DETECTOR

The grid-leak detector functions like a diode detector combined with a triode amplifier. It is convenient to consider detection and amplification as two separate functions. In figure 10-19,A, the grid functions as the diode plate.

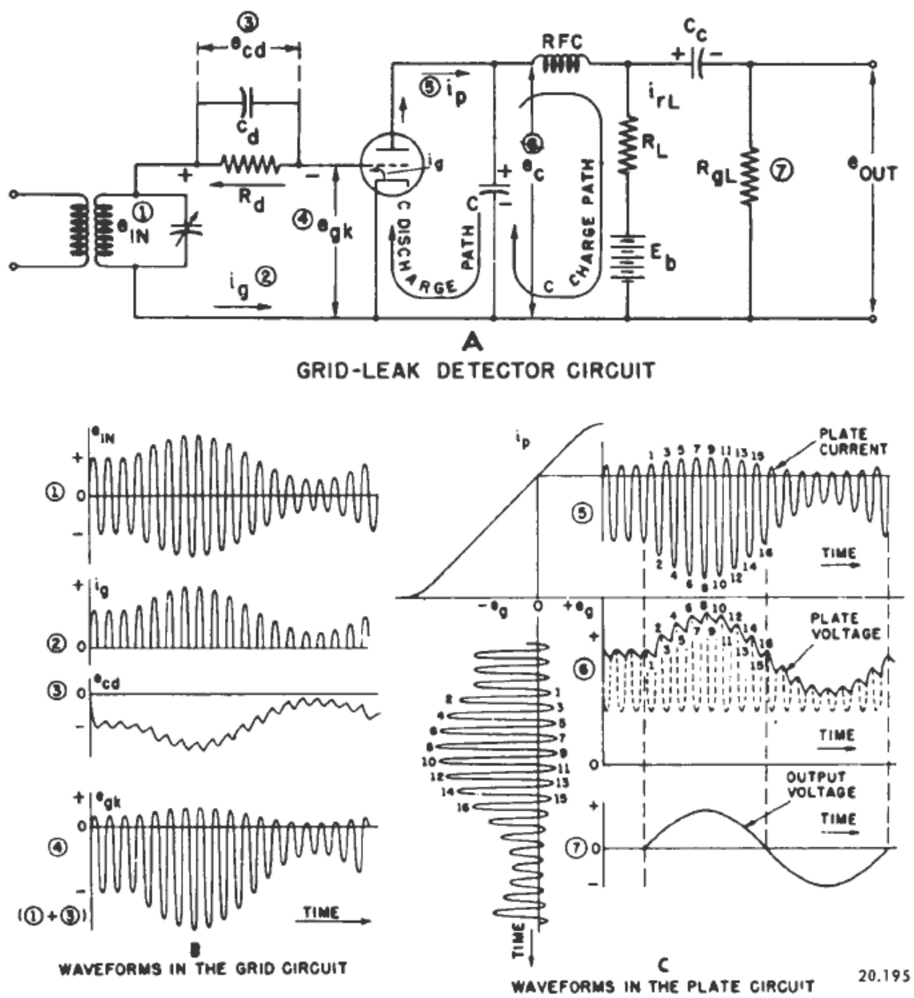


Figure 10-19.—Grid-leak detector and waveforms.

The values of C_d and R_d must be so chosen that C_d charges during the positive peaks of the incoming signal and discharges during the negative peaks. The time constant of R_dC_d should be long with respect to the r-f cycle and short with respect to the a-f cycle.

An approximate analysis of the waveforms existing in the diode (grid) circuit is shown in figure 10-19,B. Part ① shows the input waveform which is also the waveform in the input tuned circuit. Because r-f current i_g flows in only one direction in the grid circuit, part ② shows a rectified current waveform in this circuit. Part ③ shows the waveform developed across C_d. This audio waveform is produced in the same way as the audio waveform in the diode detector. However, the waveform shown in part ③ is not the output voltage. In the grid-leak detector the waveform produced across C_d is combined in series with the r-f waveform in the tuned circuit to produce the grid-to-cathode waveform shown in part ④.

An approximate analysis of the waveforms existing in the triode plate circuit is shown in figure 10-19,C. Part ⑤ is the plate-current waveform, and part ⑥ is the plate-voltage waveform.

Capacitor C discharges on the positive half cycles of grid input voltage (points 1, 3, 5, 7, 9, 11, 13, and so forth). The discharge path is clockwise through the circuit including the tube and capacitor. The time constant of the discharge path is the product of the effective tube resistance and the capacitance of capacitor C, and this time constant is short because the effective tube resistance is low. The increase in plate current is supplied by the capacitor rather than the B supply, thus preventing any further increase in current through the r-f choke and plate load resistor R_L. Therefore, any further change in plate and capacitor voltage is limited.

Capacitor C charges up as plate voltage rises on the negative half cycles of r-f grid input voltage (fig. 10-19,C, points 2, 4, 6, 8, 10, 12, 14, and so forth). The charging path is clockwise through the circuit containing the capacitor, r-f choke, load resistor R_L, and the B supply. The rise in plate voltage is limited by the capacitor charging current which flows through the r-f choke and through R_L. The plate current decrease is approximately equal to the capacitor charging current; thus the total current through the r-f choke and R_L remains nearly constant, and the plate and capacitor voltage rise is checked.

Positive grid swings cause sufficient grid current flow to produce grid-leak bias. Low plate voltage limits the plate current on no signal in the absence of grid bias. Thus, the amplitude of the input signal is limited, since with low plate voltage the cutoff bias is low, and that portion of the input signal that drives the grid voltage below cutoff is lost. The waveform of the voltage across capacitor C is shown by the solid line in part ⑥ of figure 10-19,C. The plate voltage ripple is removed by the r-f choke (RFC). Part ⑦ shows the output-voltage waveform. This waveform is the difference between the voltage at the junction of R_L and RFC with respect to the negative terminal of E_b and the voltage across coupling capacitor C_c, which for most practical purposes is a pure d-c voltage.

Because the operation of the grid-leak detector depends on a certain amount of grid-current flow, a loading effect is produced which lowers the selectivity of the input circuit. However, the sensitivity of the grid-leak detector is moderately high on low-amplitude signals.

PLATE DETECTOR

In a GRID-LEAK DETECTOR the incoming r-f signal is detected in the grid circuit and the resultant a-f signal is amplified in the plate circuit. In a PLATE DETECTOR, the r-f signal is first amplified in the plate circuit, and then it is detected in the same circuit.

A plate detector circuit is shown in figure 10-20,A. The cathode bias resistor, R1, is chosen so that the grid bias is approximately at cutoff during the time that an input signal of proper strength is applied. Plate current then flows only on the positive swings of grid voltage, during which time average plate current increases. The peak value of the a-c input signal is limited to slightly less than the cutoff bias to prevent driving the grid voltage positive on the positive half cycles of the input signal. Thus, no grid current flows at any time in the input cycle, and the detector does not load the input tuned circuit, LC1.

Cathode bypass capacitor C2 is large enough to hold the voltage across R1 steady at the lowest

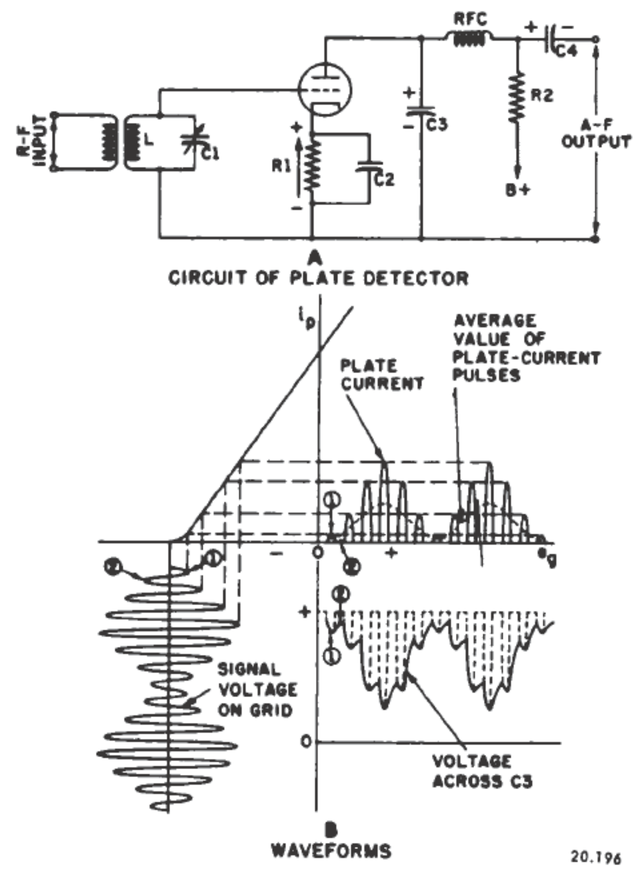


Figure 10-20.—Plate detector and waveforms.

audiofrequency to be detected in the plate circuit. C3 is the demodulation capacitor across which the a-f component is developed. R2 is the plate load resistor. Thr r-f choke blocks the r-f component from the output. R2C3 has a long time constant with respect to the time for one r-f cycle so that C3 resists any voltate change which occurs at the r-f rate. R2C3 has a short time constant with respect to the time for one a-f cycle so that the capacitor is capable of charging and discharging at the audio rate.

The action of the plate detector may be demonstrated by the use of the i_p-eg curve in figure 10-20,B. On the positive half cycle of r-f input signal (point 1) the plate voltage falls below the B supply because of the increased drop across R2 and the r-f choke. Capacitor C3 discharges. The discharge current flows clockwise through the circuit including the tube and C3. Plate current is supplied by C3 rather than the B supply. The drop across R2 and the r-f choke is limited, and the decrease in plate voltage is slight.

On the negative half cycle of r-f input signal (point 2) plate current is cut off and plate voltage rises. Capacitor C3 charges. The charging current flows clockwise around the circuit including the r-f choke, R2, and the B supply. The drop across R2 and the r-f choke contributed by the charging current of C3 checks the rise in plate voltage.

Thus, C3 resists voltage change at the r-f rate. Because C3R2 has a short time constant with respect to the lowest a-f signal, the voltage across C3 varies at the a-f rate.

The plate detector has excellent selectivity. Its sensitivity (ratio of a-f output to r-f input) is also greater than that of the diode detector. However, it is inferior to the diode detector in that it is unable to handle strong signals without overloading. Another disadvantage is that the operating bias will vary with the strength of the incoming signal and thus cause distortion unless a means is provided to maintain the signal input at a constant level. Thus, a-v-c or manual r-f gain control circuits usually precede the detector.

REGENERATIVE DETECTOR

When high sensitivity and selectivity are the most important factors to be considered, a regenerative detector may be used. However, the linearity as well as the ability to handle strong signals without overloading is very poor.

The process of feeding some of the output voltage of an electron-tube circuit back into the input circuit so that it adds to or reinforces (is in phase with) the input voltage is known as REGENERATION or POSITIVE FEEDBACK. The use of regeneration in a circuit greatly increases the amplification of the circuit because the output voltage fed back into the input circuit adds to the original input voltage, thus increasing the total voltage to be amplified by the tube.

A grid-leak detector may be modified to operate as a regenerative detector, as indicated in figure 10-21. Because an amplified r-f component is present in the plate circuit of the grid-leak detector, regeneration can be obtained by connecting a coil, L2, known as a TICKLER COIL, in series with the plate circuit so that it is inductively coupled to the grid coil, L3.

With an r-f signal across L3 an r-f component of plate current flows through L2. L2 is

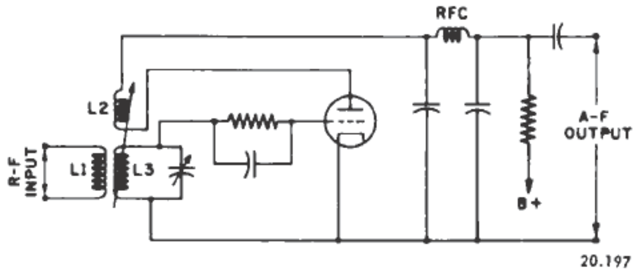


Figure 10-21.—Regenerative detector.

connected so that the voltage it induces in L3 is in phase with the incoming signal voltage applied to the grid. Thus, the voltage gain of the stage is increased.

It is important that the voltage fed back by the ticker coil be in phase with the incoming signal voltage. Otherwise the feedback will be degenerative and the amplification will be reduced. Furthermore, if the coupling between L2 and L3 is too great, oscillation will take place. For receiving code signals, oscillation is desirable in order to produce an audible beat tone. However, it is not desirable for voice or music reception because of the objectionable squeal from the beat tone. The regenerative detector is the most sensitive triode detector circuit possible when it is operated just below the point of oscillation.

A comparison of the sensitivity (ratio of a-f output to r-f input), linearity (accuracy of reproduction of the waveform of the modulation component), selectivity (ability to separate signals), and relative ability of the various detectors to handle large signals without overloading is given in table 10-1.

HETERODYNE DETECTOR

The process of combining two frequencies to obtain the difference frequency is called HETERODYNING. The heterodyne principle has a number of important Navy applications. It is used in heterodyne code reception to change continuous-wave (c-w) telegraph signals to an audio-frequency. It is widely used in superheterodyne receivers to change the carrier frequency to the fixed intermediate frequency. Heterodyne action is also employed to separate frequencies that differ from one another by small amounts.

If two a-c signals of different frequencies are combined, or mixed, in a suitable circuit, a third signal, called a BEAT FREQUENCY, will be produced. The frequency of the beat is equal to the difference between the frequencies that are mixed to produce it.

Thus, if two a-f voltages having frequencies of 500 and 600 cps are properly mixed, a beat frequency of 100 cps will be produced by a suitable reproducer.

If two r-f signals differing in frequency by an audible frequency are mixed, an a-f voltage will be produced. Thus, if two r-f voltages having frequencies of 500 and 501 kc are properly mixed, a beat frequency of 1 kc will be produced.

An important application of this principle is in the detection of c-w signals. In c-w transmission the unmodulated carrier is keyed, or interrupted, according to the code that is being transmitted. Because no modulation component is present in the transmitted energy, the detectors previously considered are not effective in detecting the signal. At best, there may be some noise or sound variation in the reproducer at the beginning and end of each interruption, but reception will be unsatisfactory. In order to receive c-w signals from a radiotelegraph transmitter it is therefore necessary to mix with the incoming signal another locally generated signal that differs in frequency from the incoming signal by some frequency in the audible range.

Table 10-1.—Comparison of A-m Detectors

Detector type	Sensitivity	Linearity	Selectivity	Ability to handle strong signals without overloading
Diode	Low	Good	Poor	High
Grid-leak.	High	Poor	. . . .do. . . .	Limited
Plate.	Medium	Fair.	Excellent . . .	Medium
Regenerative . . .	Very high. . .	Poor.	. . . .do. . . .	Poor

One method of producing an audible beat tone is by the use of an oscillating regenerative-detector circuit. If the regeneration, or positive feedback, in a regenerative detector is increased beyond a certain critical point, the circuit will oscillate at a frequency approximately equal to the frequency of the tuned circuit. Thus, if the regenerative detector is made an oscillating detector, and is tuned so that the frequency it generates differs from the incoming r-f signal frequency by an audible amount, it is possible to detect unmodulated r-f signals. This process is known as heterodyning, and an oscillating detector is called a HETERODYNE DETECTOR. A brief analysis of the heterodyne principle is included under superheterodyne receivers in chapter 14.

Many communication receivers are equipped to receive both a-m signals and c-w signals. These receivers are commonly designed with an oscillator, called the BEAT-FREQUENCY OSCILLATOR (BFO), coupled to the plate circuit of the diode detector. The beat-frequency oscillator is then tuned to a frequency that differs from the intermediate-frequency by the desired audiofrequency. In t-r-f receivers, the local oscillator must differ from the incoming signal by the desired audiofrequency.

The first detector, or frequency converter, of superheterodyne receivers is a heterodyne detector. In this instance there is a modulation component, but it is desirable to reduce the carrier frequency to a new, lower frequency, called the INTERMEDIATE FREQUENCY. Therefore, the incoming signal is mixed with a locally generated signal to produce a frequency that is the difference between the two signals. The modulation envelope is not appreciably affected by the heterodyne action.

DEMODULATION OF F-M WAVES

In f-m transmission the intelligence to be transmitted causes a variation in the instantaneous frequency of the carrier either above or below the center, or resting, frequency. The detecting device must therefore be so constructed that its output will vary linearly according to the instantaneous frequency of the incoming signal. Also, the detecting device must be insensitive to amplitude variation produced by interference or by receiver nonlinearities; thus a special limiting device, called a LIMITER, must precede the f-m detector.

TYPES OF F-M DETECTORS

A number of f-m detectors might be used. Each has certain inherent advantages and disadvantages. Two of the most common are the DISCRIMINATOR and the RATIO DETECTOR. There are also various types of LOCKED-OSCILLATOR f-m detectors. The simplest type of detector is the SLOPE detector. Although it is rarely used, this type of f-m detector will be considered first, because of its simplicity.

Slope Detector

Even an a-m receiver may give a distorted reproduction of an f-m signal under certain conditions of operation. When the carrier frequency of the f-m signal falls on the sloping side of the r-f response curve in an a-m receiver, the frequency variations of the carrier signal are converted into equivalent amplitude variations. This conversion results from the unequal response above and below the carrier center frequency (point B), as shown in figure 10-22.

Thus, when the incoming f-m signal is less than the center frequency—for example, at point A, which is the minimum value—the output voltage is at a minimum in the negative direction. When the incoming signal swings to point C (the maximum value), the output voltage is maximum in the positive direction. The resultant a-m signal may be coupled to the regular a-m detector where the original audio voltage is recovered.

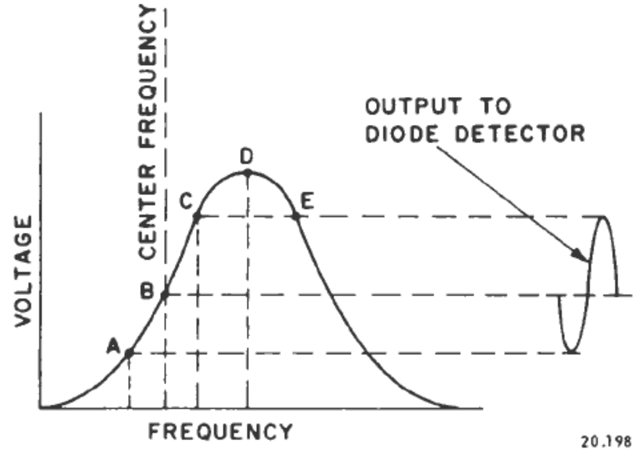


Figure 10-22.—Slope detection.

The obvious disadvantage of this type of detection is the nonlinearity of the response curve. At best, the most linear portion of the curve has a limited frequency range. Consequently the undistorted output voltage is low.

If in figure 10-22 the center frequency falls at point D and the maximum frequency swings are between points C and E, there is no effective output signal voltage because the curve is relatively flat.

Discriminator

One form of discriminator is shown in figure 10-23. It must be preceded by one or more limiter stages (discussed in chapter 14) because this type of f-m detector is sensitive to both amplitude changes and frequency changes. Basically, what is wanted is an output voltage, e_{10}, that varies in amplitude according to the instantaneous frequency of the incoming signal. Amplitude variations therefore must be removed ahead of the discriminator to keep their instantaneous values from adding to those values produced by the instantaneous frequency changes.

The input voltage, e_1, is applied across the input tuned circuit. The current, i_1, lags e_1 by 90 degrees. The mutually induced voltage, e_2, lags i_1 by 90 degrees. Thus, e_2 is 180° out of phase with e_1, as shown in part ① of figure 10-23,B.

Inductor L4 is shunted across the input tuned circuit via C2 and C5, which have negligible reactance at the resonant frequency. Thus e_1 is also applied across L4.

Assume first that the incoming signal is at the resting frequency. The induced current, i_2, is in phase with e_2, as shown in parts ② and ③ of figure 10-23,B. The voltages e_3 and e_4 are the iX_L drops across L2 and L3 respectively. From figure 10-23,A, and part ③ of 10-23,B, it may be seen that e_6, the voltage applied to V1 is the vectorial sum of e_1 and e_3; and e_7, the voltage applied to V2, is the vectorial sum of e_1 and e_4. The rectified output voltage of V1 is e_8 and that of V2 is e_9. The output voltage, e_{10}, is the algebraic sum of e_8 and e_9. In part ③ of figure 10-23,B, e_6 and e_7 are equal because the incoming signal is at the resting frequency. Therefore e_8 is equal to e_9; and since they are in opposite directions, the output voltage is zero.

Below the resting frequency i_2 leads e_2 because X_C is greater than X_L. Voltages e_3 and e_4 are still in phase opposition, but each is 90° out of phase with i_2, as shown in figure 10-22,C. Therefore, e_7 is greater than e_6, and e_9 is greater than e_8. Point A becomes negative with respect to ground, thus producing an output signal voltage.

Above the resting frequency, i_2 lags e_2 because X_L is greater than X_C. Voltages e_3 and e_4 bear the same phase relation with each other and with i_2 as they did for each of the above conditions. However, from figure 10-23, D, e_6 is now greater than e_7. Therefore, e_8 is greater than e_9, and point A becomes positive with respect to ground, thus producing the other half of the audio signal-voltage waveform.

Ratio Detector

One form of the ratio detector is shown in figure 10-24,A. It differs from the discriminator in figure 10-23, A in that the diodes are connected in series across the transformer secondary; whereas each diode plate in the discriminator is connected to opposite ends of the transformer. There are differences also in the method of obtaining the output voltage and in the amount of limiting preceding the detector. However, the vector analysis is essentially the same in both circuits.

The induced voltage, e_2, is 180° out of phase with e_1, as indicated by the vector diagrams in figure 10-24. At resonance (fig. 10-24,B), i_2 is in phase with e_2, as in all tuned circuits. Voltages e_3 and e_4, developed by the i_2X_{L2} and i_2X_{L3} voltage drops respectively, are 180° out of phase with each other and 90° out of phase with i_2. This out-of-phase relation holds true at resonance as well as below or above resonance.

From figure 10-24,B, e_5, the a-c voltage applied across V1, is the vector sum of e_1 (coupled through C3 between the center tap and ground) and e_3. Also, e_6, the voltage applied across V2, is the vector sum of e_1 and e_4. At resonance, e_5 and e_6 are equal.

The angle of current flow through V1 and V2 may be of the order of 40° or less. In figure 10-24,B, vector e5 leads e6 by approximately 50°. Thus V1 will conduct for 40° and be cut off for 10° of the input cycle

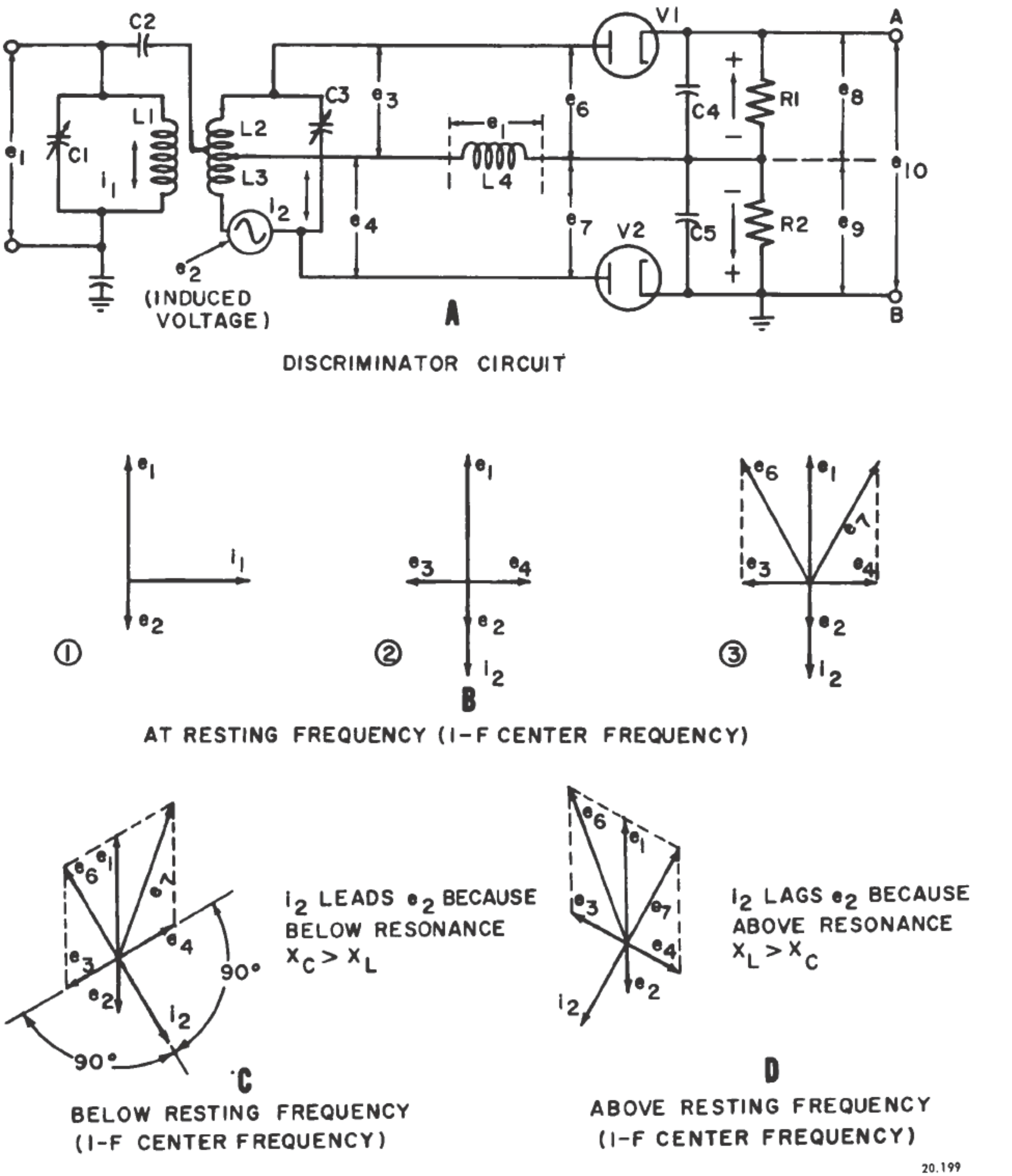


Figure 10-23.—Discriminator circuit and vector diagram.

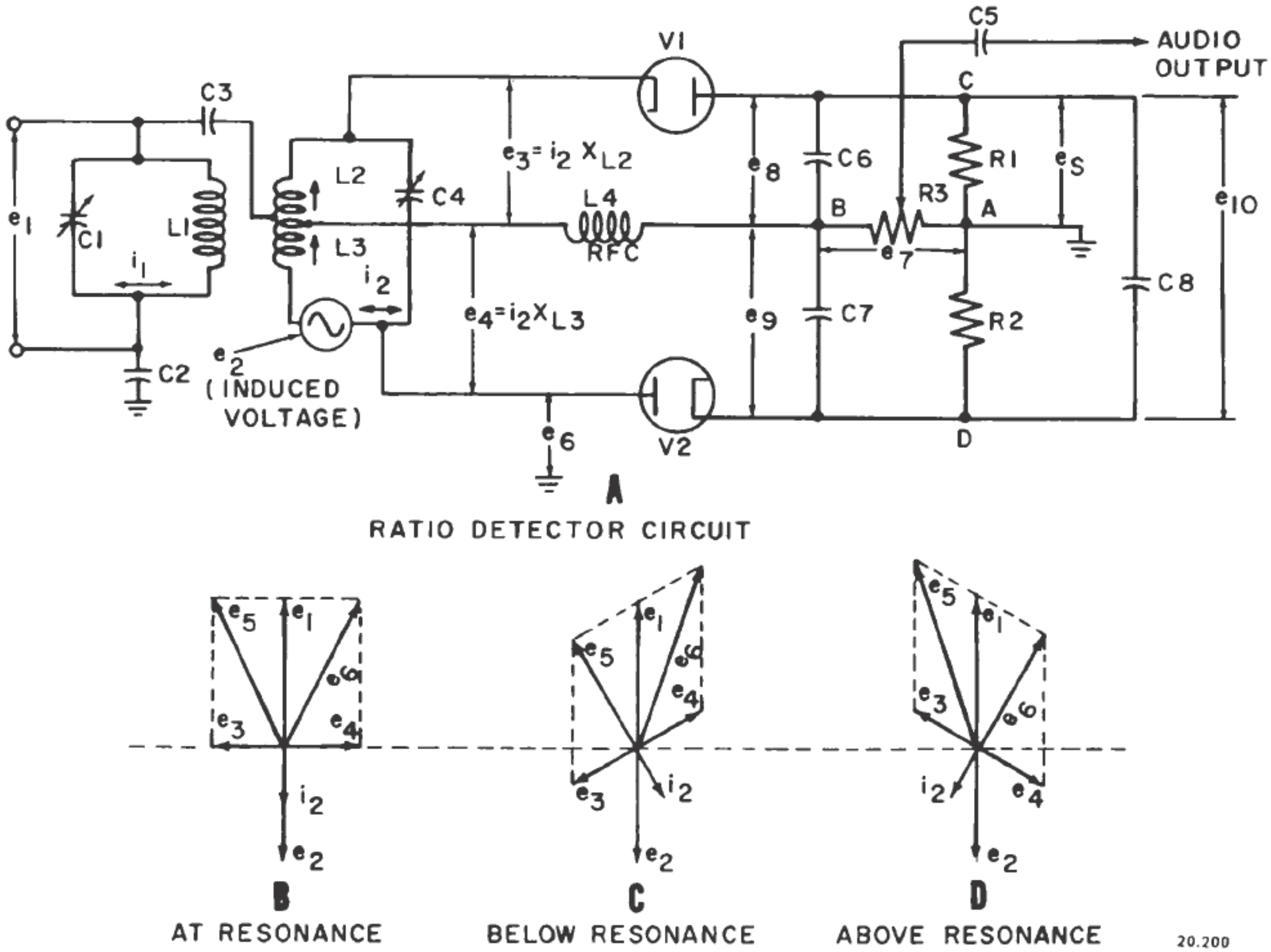


Figure 10-24.—Ratio detector and vector diagrams.

before V2 begins conduction. The current path of V1 is from C to A, B, L4, L2, V1 and back to C. The current path of V2 is from D to V2, L3, L4, B, A, and back to D. Hence r-f rectified current flow via V1 from A to B and via V2 from B to A. Without capacitors C6 and C7 the r-f voltage developed across R3 as a result of these current is small because of RF choke L4 in series with R3.

However, C6 and C7 will charge to d-c voltages e8 and e9 respectively, and will keep the voltage across R3 at the average of these two currents which is zero at the rest (resonant) frequency. In other words when V1 conducts, the path for current is via C6 instead of R1 and R3. When V2 conducts, the path for current is via C7 instead of R2 and R3. During the interval when neither diode is conducting, C6 and C7 will discharge only slightly. Thus the output voltage e7 across R3 is zero at the rest (resonant) frequency because the V1 and V2 currents are equal.

Below resonance (fig. 10-24,C) i2 leads e2 because X_C is greater than X_L. Therefore, e6, the vector sum of e1 and e4, is greater than e5, the vector sum of e1 and e3. The current paths are the same as previously stated but because e_6 is greater than e_5 the current through V2 is greater than through V1. Thus the current through R3 from B to A (V2 conducting) will be greater than from A to B (V1 conducting). Point B therefore swings negative with respect to point A to develop the audio output voltage e7.

Above resonance (fig. 10-24,D), i2 lags e2 because X_L is greater than X_C. Therefore, e5 is greater than e6. The current through V1 exceeds the current through V2. The current through R3 from A to B is greater than from B to A and the potential at point B swings positive with respect to point A. Thus the other half of the audio output cycle appears across R3.

If there is a sudden undesirable increase in amplitude of the input to the ratio detector due to noise, static, and so forth, capacitor C8 having a large capacitance will resist any change in amplitude of voltage across terminals CD. Because of the low impedance of the capacitor, C8 will effectively "absorb" all but very low frequency amplitude variations.

Although the ratio detector requires fewer i-f amplifier stages and minimum limiting, it presents alignment difficulties, and the signal may be distorted at high input voltages if some form of limiting is not applied.

CHAPTER 11

TRANSMITTERS

INTRODUCTION

A transmitter is a device for converting intelligence, such as voice or code, into electrical impulses for transmission either on closed lines, or through space from a radiating antenna. Transmitters take many forms, have varying degrees of complexity, develop various levels of power, and employ numerous methods of sending the desired information or energy component from one point to another.

A telephone handset has both a transmitter and a receiver section. The transmitter section converts the human voice into electrical impulses that may be amplified and conveyed along the closed telephone line to the receiving station.

Radar transmitters develop bursts of energy of the required frequency and duration and radiate this energy in the direction in which the antenna is pointing. The echo may be utilized to give such information as range and bearing. The tremendous energy radiated by some radar sets is made possible because of the relatively long resting time between pulses of energy. Radar will be treated briefly in chapter 15.

Loran transmitters are especially constructed for use in navigation. The principle of loran is based on the difference in time required for pulse radio signals to arrive at a point from a pair of synchronized transmitters. These transmitters operate on low frequencies, are accurately synchronized, and develop considerable power because they are transmitting only a small fraction of the time. Loran transmitters, as well as sonar and other transmitters, are treated in the rating texts.

In this chapter the discussion is confined to radio transmitters; however, many of the principles involved in this discussion apply in general to other basic electronic transmitters.

The function of a radio transmitter is to supply power to an antenna at a definite radio frequency and to convey intelligence by means of the radiated signal. Radio transmitters radiate waves of two general types.

One type of radiation is the CONTINUOUS wave (c-w), or UNMODULATED WAVE, which has a waveform like that of the r-f current in the tuned tank circuit of a power output stage. In this type of wave the peaks of all the waves are equal, and they are evenly spaced along the time axis. The waveform is sinusoidal.

The other type of radio wave is the MODULATED WAVE. The amplitude may be modulated by means of a signal of constant frequency, as in MODULATED-CONTINUOUS-WAVE (m-c-w) telegraphy. Likewise, the amplitude may be modulated by means of speech, music, and so forth; and in this case is called AMPLITUDE MODULATION (a-m). If the frequency of the wave is varied with time it is called FREQUENCY MODULATION (f-m).

A given transmitter operated on c-w has a greater range than the same transmitter (for the same power output) operated on m-c-w or voice modulation. This condition results from the fact that all the intelligence is contained in the side bands (treated in ch. 10), and the fewer the number of side-band frequencies the greater will be the signal strength in the remaining side-band frequencies. In c-w operation the side bands do not extend very far on each side of the carrier, and all of the energy is therefore contained in a narrow band and not wasted in nonessential bands.

On m-c-w the side bands are necessarily wider, and more energy is needed to supply the side bands; that is, each side band contains proportionately less energy. In order to get the same signal level (at the required bandwidth) to a receiver, the transmitter must increase its output power over what it would be if c-w were used.

When voice modulation is used, the necessary side bands are increased over those needed for m-c-w. Each side band requires a certain amount of energy, and therefore, in order to keep the energy level of all the essential side bands up to the required level at the receiver the transmitter must deliver more energy than for m-c-w.

Navy transmitters operate on very-low-frequency (v-l-f), low-frequency (l-f), medium-frequency (m-f), and high-frequency (h-f) bands as well as on very-high-frequency (v-h-f) and ultrahigh frequency (u-h-f) bands.

The VERY-LOW-FREQUENCY BAND, from 10 to 30 kilocycles, is not covered by shipboard transmitters. The antennas needed for such low frequencies are too long to be erected aboard a ship. However, there are some v-l-f stations on shore. One of the frequencies of the Primary Fleet Broadcast, NSS, is in the very-low-frequency band. Powerful v-l-f stations with their huge antennas are capable of transmitting signals through magnetic storms that blank out the higher radio-frequency channels.

The LOW-FREQUENCY BAND, from 30 to 300 kc, is used mostly for long-range direction finding. This band provides, however, a means of reliable medium- and long-range communication. Useful amounts of radiation can be produced in this band with shipboard antennas. The frequencies in the low-frequency band do not depend on sky waves and provide stable communication with little variation from season to season.

The area that may be covered by the MEDIUM-FREQUENCY BAND, extending from 300 to 3000 kc, depends on the ground wave. Sky-wave reception of medium-frequency waves is also possible. (Propagation is treated in chapter 13.) At the upper end of this band the ionosphere has a great effect on the sky waves. Relatively long distances can be covered by using this band if the correct frequency is used at the correct time. The international distress frequency, 500 kc, is in this band. Commercial broadcast stations as well as Navy stations operate in this band.

The HIGH-FREQUENCY BAND is also used by the Navy. This band includes the frequencies between 3 and 30 megacycles. The sky wave is increasingly important in this band, and long-range communication is possible. Propagation characteristics of waves in the h-f band change with the time of day and the season. The choice of frequency depends on the variables in the ionosphere. For long-range communications either the h-f band or the upper part of the m-f band may be employed, depending mostly on the time of day.

Single side band (ssb) communications are becoming increasingly important in Navy communications. If one of the two sidebands in double sideband (dsb) transmission is filtered out before it reaches the transmitter power amplifier, the intelligence can be transmitted on the remaining single sideband. All of the power is then transmitted on one sideband rather than being divided between the carrier and the two sidebands as in dsb transmissions.

Thus, theoretically, a radiated power of 50 watts at the ssb transmitter is equivalent to a radiated power of 150 watts at the dsb transmitter (100 watts carrier power and 50 watts combined upper and lower sideband power).

The AN/URC-32 transceiver is designed primarily for single sideband transmission or reception on either the upper or lower sideband or on both sidebands simultaneously with separate channels for each sideband. In addition to ssb operation it can be used for cw and a-m radio telephone transmission in the 2-30 mc range.

The very-high-frequency (v-h-f) band extends from 30 to 300 mc. The portion from 225 to 300 mc (called UHF by the Navy) is used extensively by both surface and air units for communications purposes. The older TDQ transmitter has been replaced largely by the AN/URT-7 (115-156 mc range). The latter is used for short range communications by surface ships not yet converted to u-h-f equipment. It is similar in size and appearance to the Navy model TED u-h-f transmitter. Portions of the v-h-f band, however, are used for airborne communications.

The v-h-f band is used with early warning radar, television, and f-m broadcasting stations.

The ultra high frequency (u-h-f) band includes the frequencies between 300 and 3000 mc. The low end of this band is used for communications and portions of the high end are used for radar.

Navy model TED transmitter is a short range (line-of-sight) u-h-f transmitter used widely for ship-to-ship, ship-to-aircraft, and harbor communications.

The AN/GRC-27 and AN/GRC-27A equipments are u-h-f transmitter receiver sets covering the range from 225 to 400 mc. These sets have replaced the older Navy type TDZ transmitter. They are used extensively for u-h-f radio telephone and m-c-w communications.

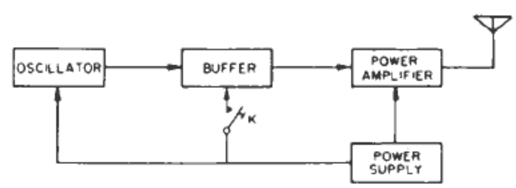
Navy transmitters must be flexible as far as type of modulation (c-w, m-c-w, or voice) and frequency range (frequency multiplication is often used to extend the range) are concerned. The maximum in utility must be obtained from the equipment for a minimum in size and weight. Therefore, care in design and modification where needed characterize these transmitters. A few of the more important features that must be incorporated in every Navy transmitter are excellent frequency stability, ruggedness, long life, flexibility of operation, remote-control operation, ease of tuning, and high efficiency. A given series of transmitters may include several models usually differing only in the power supplies or in minor mechanical or electrical changes. Navy equipment is somewhat different from corresponding commercial equipment. Some of the differences may seem of little significance, but over a period of years the equipment designed specifically for Navy use has proved superior.

CONTINUOUS-WAVE TRANSMITTERS

The continuous wave is used principally for radio telegraphy—that is, for the transmission of short or long pulses of r-f energy to form the dots and dashes of the Morse code characters. C-w transmission was the first type of radio communication used, and it is still used extensively for long-range communications. Some of the advantages of c-w transmission are narrow bandwidth; high degree of intelligibility, even under severe noise condition and long range operation.

A modern development for the transmission of messages is the teletype, which replaces the c-w operator for changing the plain language into code and code into plain language. To transmit a message with teletype the operator presses keys on a keyboard similar to a standard typewriter. As each key is pressed, mechanical cams and linkages cause a sequence of mark and space signals to be sent out by the transmitter. At the receiving end, the received teletype signal actuates selector magnets in a similar machine that causes the character transmitted to be printed on a paper and the carriage to be advanced one space. The characters for either system may be cut in their respective tapes by an operator at relatively slow speeds and later transmitted at high speeds. Automatic transmission of c-w messages has the advantages of speed and rythmic character transmission which is recorded on tape but reproduced later at a reduced speed that suits the transcribing operator. However, any rapid transmission has the great disadvantage of being very easily interfered with. When interference is too great, an operator must take over at slow speeds and transcribe on a typewriter a straight c-w transmission being substituted for the high-speed signals.

The four essential components of a c-w transmitter are: (1) a generator of r-f oscillations, (2) a means of amplifying these oscillations, (3) a method of turning the r-f output on and off (keying) in accordance with the code to be transmitted, and (4) an antenna to radiate the keyed output of the transmitter. A block diagram of a master-oscillator power-amplifier transmitter together with the power supply is shown if figure 11-1.



20.201

Figure 11-1.—Master-oscillator power-amplifier transmitter.

OSCILLATOR

One of the most important sections of a transmitter is the one containing the oscillator. Here the frequency on which the transmitter operates (or a subharmonic of the transmitter frequency, if frequency doublers or multipliers are used) is generated and maintained within the required limits. Two of the problems encountered in this section of the transmitter are (1) maintaining a stable frequency, and (2) switching from one frequency to another with a minimum of adjustments. Electron tube oscillators are treated in chapter 8.

Only some of the special problems as applied to transmitters will be considered in this chapter.

One of the major problems encountered in the operation of transmitters is that of frequency drift. The frequency of a transmitter must be stable enough to permit a receiver to stay on the transmitter frequency. However, the master oscillator in a transmitter tends to change frequency when is is being warmed up and when the load on the oscillator varies. The law requires that the carrier frequency be held very close to the specific frequency assigned by the Federal Communications Commission. For example, the frequency tolerance allowed an international broadcast station is 0.005 percent of the assigned frequency.

The frequency of a transmitter can be stabilized by the use of a crystal oscillator. However, this arrangement would require a large number of crystals to cover the many frequency channels used by the Navy. A more flexible means of obtaining stability (especially at the lower frequencies) is to control the frequency of a transmitter with a variable master oscillator. An electron-coupled oscillator (ECO) is commonly used.

Most of the frequency drift in master oscillators is due to changes in the physical size of the components with variations in temperature; therefore, changes in the electrical characteristics of the oscillator circuit, as well as changes in the oscillator tube characteristics, are introduced. Placing the frequency-determining components of the oscillator in a temperature-controlled oven minimizes this drift. To ensure further stability the oscillator is loaded very lightly and isolated by a buffer stage.

The frequency of the master oscillator can be affected also by vibration and sudden shocks. In some transmitters all of the oscillator elements are mounted in a single oscillator unit. The oscillator unit is then suspended on springs and snubbed by sponge rubber cushions to keep the shock and vibrations reaching the oscillator unit to a minimum.

Frequency stability becomes even more important when a transmitter uses frequency multiplier stages because any drift in the oscillator frequency will be multiplied in these stages. For example, if the output frequency is eight times the oscillator frequency, any drift of the oscillator frequency will be multiplied by eight.

Not all Navy transmitters use all of these refinements, but every transmitter has some means of ensuring frequency stability.

The majority of master oscillator circuits in the older type low- and medium-frequency Navy transmitters are electron-coupled oscillators because of their stability. The frequency of the oscillator is varied by either a variable capacitor or a variable inductor. Different frequency ranges are obtained by using a tapped oscillator coil or by switching in various values of capacitance. Sometimes both methods are used together. Usually the frequency of the oscillator is doubled in the plate circuit. With this arrangement, any energy fed back to the grid circuit is twice the frequency of the energy in the grid circuit and does not affect the stability of the oscillator.

In most oscillators any change in oscillator frequency will affect the circuit stability of the oscillator. Modern Navy transmitters like the AN/SRT 14, 15, and 16 use a crystal controlled master oscillator with a fixed frequency. The crystal is often mounted in a crystal oven that maintains a constant temperature, thus increasing the frequency stability of the oscillator. By means of heterodyning and frequency multiplication processes taking place in several subunits associated with the master oscillator, a wide range of accurately calibrated output frequencies is obtained.

Most of the transmitters operating in the u-h-f band are crystal controlled. However, crystals having a fundamental frequency in the u-h-f band are not practical because of the physical restrictions, such as the difficulty in grinding the crystals, and their extreme fragility. Therefore, the transmitter usually employs a low-frequency oscillator followed by a number of frequency multipliers, the number used depending on the desired output frequency. Of course, the multiplier stages must be accurately tuned to the correct harmonic frequency. In this arrangement the crystal is larger and more substantial than it would be if it were operated at the higher frequencies.

BUFFER AMPLIFIER

As mentioned previously, a buffer amplifier is placed between the oscillator and the power amplifier to isolate the oscillator from the load and thus improve the frequency stability of the transmitter. If the frequency of the plate tank circuit of the buffer amplifier is the same as that of the oscillator driving it, the stage is a conventional type of amplifier, usually class C.

If the plate tank circuit of the buffer amplifier is tuned to the second harmonic (in order to increase the frequency of the radiated signal) of the driving signal applied to the grid, the stage becomes a frequency doubler and the output voltage has a frequency equal to twice that of the input. Likewise, the buffer amplifier may become a tripler or a quadrupler.

A frequency-doubler stage is shown in figure 11-2,A. The plate tank is tuned to twice the frequency of the grid tank. If L1 is equal to 10 μh and C1 is equal to 25.3 $\mu\mu$f, the resonant frequency of the grid tank is

$$f_1 = \frac{159}{\sqrt{LC}} = \frac{159}{\sqrt{10 \times 25.3}} = 10 \text{ mc.}$$

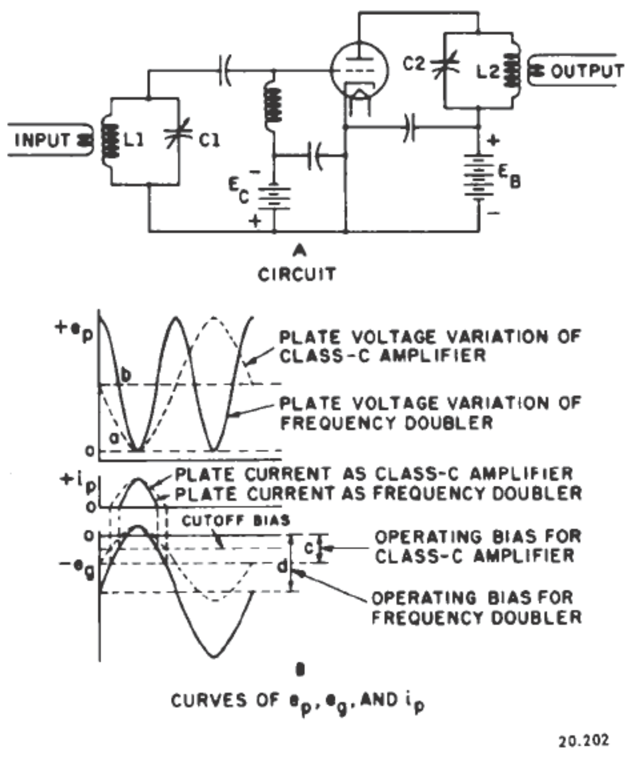


Figure 11-2.—Frequency doubler.

If the plate tank coil has an inductance of 10 μh and the resonant frequency of the plate tank is 20 mc, the plate tank capacitor, C2, will have

$$\frac{C1}{4} = \frac{25.3}{4} = 6.3 \ \mu\mu\text{f.}$$

The capacitance of C2 may be verified by substituting the values of f and L in the following formulas:

$$LC = \left(\frac{159}{f}\right)^2;$$

thus,

$$10C = \left(\frac{159}{20}\right)^2,$$

from which

$$C = \frac{159^2}{20^2 \times 10} = 6.35 \ \mu\mu\text{f.}$$

The curves of plate voltage, grid voltage, and plate current are shown if figure 11-2,B. The dotted curves indicate operation as a class-C amplifier without frequency multiplication, and the solid curves indicate operation as a frequency doubler. Unless the operating bias is increased, the triode plate will overheat when the plate tank is tuned to the second harmonic of the grid-tank circuit. Plate voltage is higher (a to b) during the interval the grid voltage is above cutoff, and the duration of plate current flow is reduced by increasing the operating bias from c to d.

Although the angle of plate-current flow is reduced, the efficiency is maintained. The output of the frequency multiplier varies inversely with the extent of frequency multiplication. If the plate tank is tuned to the second harmonic of the grid tank, the duration of flow of plate current ranges from 90° to 120° and the power output is about 65 percent of the output of a class-C amplifier. If the plate tank is tuned to the third harmonic of the grid tank, the angle of flow of plate current ranges from 80° to 120° and the output is reduced to 40 percent of that of a class-C amplifier. If the plate tank is tuned to the fourth harmonic of the grid input signal, the angle of plate current flow is reduced to between 70° and 90°, and the output is 30 percent of that of a class-C amplifier. If the frequency is multiplied by 5, the angle of plate current flow ranges from 60° to 72° and the output power is 25 percent of that of a class-C amplifier.

In every case it is necessary to increase the operating bias and the grid driving signal as the frequency multiplication increases in order not to overheat the triode plate. The flywheel effect in the plate tank supplies the missing cycles of grid drive and the output is approximately an undamped wave having sine waveform.

Three important conditions must prevail in order to obtain frequency multiplication—(1) high grid-driving voltage, (2) high grid bias, and (3) plate tank tuned to the desired harmonic. If the second harmonic is selected, the stage is called a FREQUENCY DOUBLER; if the third is used, the circuit is called a FREQUENCY TRIPLER, and so forth.

Certain amplifier circuits are suited to the generation of even harmonics and others to the generation of odd harmonics. Push-pull amplifiers produce only odd harmonic frequency multiplication—third, fifth, seventh, and so forth. If the grids of two triodes are connected in push-pull and the plates in parallel (fig. 11-3), even-order harmonics can be produced.

The grid signals are 180° out of phase. When one grid voltage is positive maximum, the other is negative maximum, and the second alteration of the cycle reverses the respective potentials. Thus pulsating plate current flows first in one tube and then in the other. Because the plates are connected in parallel, the output pulses are in the same direction and the plate tank circuit receives two pulses for each input cycle at the grids. This type of doubler is capable of greater output and higher plate efficiency than the single-tube type.

POWER AMPLIFIER

Whereas the buffer stage isolates the oscillator from the varying load caused by the keying, the power amplifier increases the magnitude of the r-f current and voltage by the resonant action of the plate tank circuit. The power amplifier shown in figure 11-4,A is a class-C amplifier. The triode amplification factor, μ, is 20 and the plate supply voltage is 1,000 volts. The cutoff bias, e_{co}, is therefore

$$e_{co} = \frac{-e_p}{\mu} = \frac{-1,000}{20} = -50 \text{ volts.}$$

The operating bias, e_o, is three times the cutoff, and therefore

$$e_o = 3(-50) = -150 \text{ volts.}$$

The maximum value of the r-f input signal is 180 volts. Thus, when the grid end of the r-f input is positive, the peak positive grid-to-cathode voltage is 180-150 = +30 volts. When the grid end of the r-f input is negative, the peak negative grid-to-cathode voltage is (-180) + (-150) = -330 volts.

When the grid voltage is above the cutoff value, plate current flows, and at the instant the grid voltage is +30 volts, the plate current is 150 ma (fig. 11-4,B). The tuning capacitor C4, charges up to nearly the full value of the B-supply voltage, or 950 volts. During this charging process, the lower capacitor plate is positive and the upper plate is negative. Thus, the instantaneous triode plate-to-ground voltage, when the capacitor voltage is 950 volts, is 1,000-950 = 50 volts. This value is called e_{min} and represents the lowest value of plate-to-cathode voltage in the entire cycle.

The relations between plate voltage, plate current, grid excitation voltage, and resonant plate tank circuit voltage and current are shown in figure 11-5. The flywheel effect in the plate tank circuit causes the capacitor to periodically reverse its polarity and continue the a-c cycle within the tank when the grid voltage is below cutoff and no energy is being supplied to the tank circuit from the power supply. The plate voltage swings from 1,000 volts to a minimum of 50 volts and then to a maximum of 1,950 volts before it completes the cycle.

Thus, the plate tank circuit converts pulses of unidirectional current in the triode plate circuit into sine-wave variations of current in the resonant tank. These surging currents give the tank circuit the so-called FLYWHEEL EFFECT, in which the tuned circuit makes up the portion of the sine wave missing in the plate

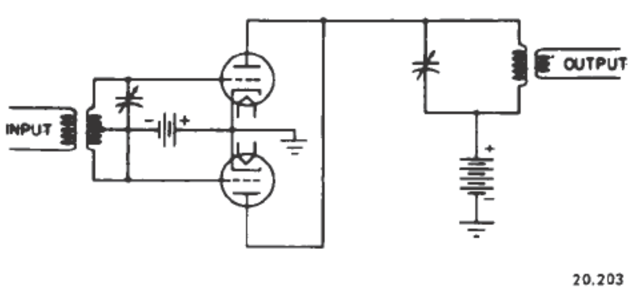


Figure 11-3.—Even-order harmonic frequency multiplier.

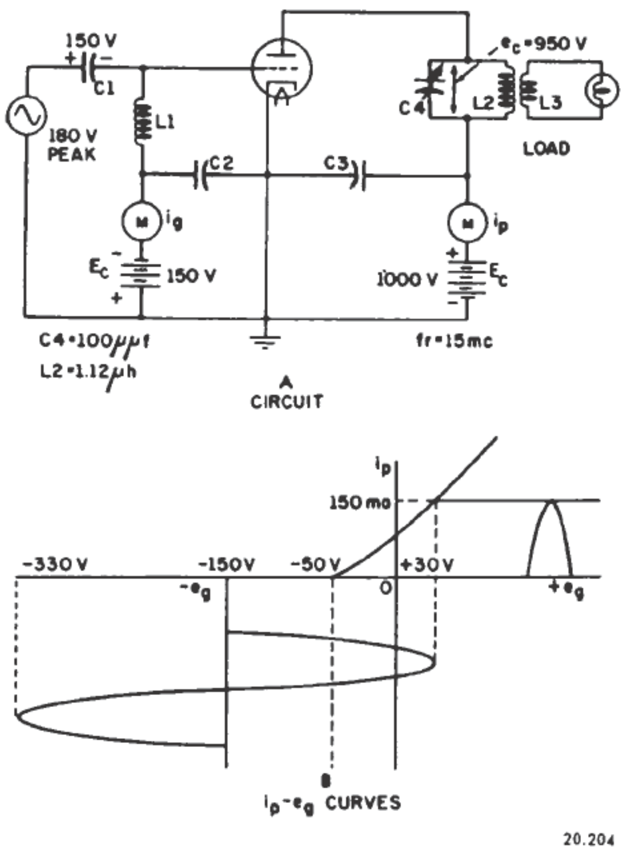


Figure 11-4.—Power amplifier.

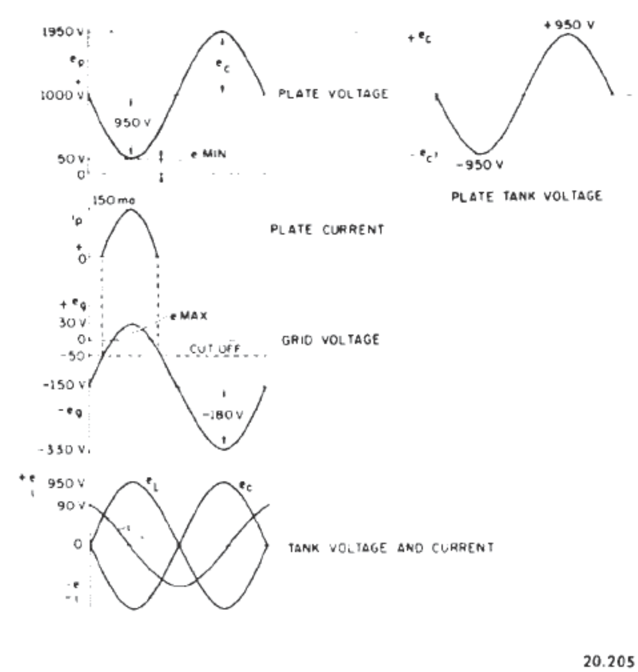


Figure 11-5.—Power amplifier current and voltage relation.

current pulses, and supplies a voltage of sine waveform to the load. The tube acts as a valve merely to supply the necessary power at just the right time.

Plate current flows for about one-third of each cycle. Energy is supplied to the tank circuit during the time plate current flows. Plate voltage, e_p, is below the power-supply value during this portion of the cycle, because the tank capacitor charges up with a polarity that opposes the polarity of the power supply, and $e_p = E_b - e_c$, where e_c is the voltage across the tank capacitor. Thus, energy is supplied to the tank with minimum plate losses because it is supplied at a time when the plate voltage (hence plate losses) is at a minimum. The efficiency of the class-C amplifier may be as high as 70 percent.

Grid voltage is positive with respect to the cathode for a short time in each cycle, and for optimum conditions the minimum value of plate voltage, e_{min}, should be equal to the maximum positive value of grid during voltage, e_{max}. In other words, $e_{max} = e_{min}$. The maximum positive grid voltage, e_{max}, should never be allowed to exceed the minimum plate voltage, e_{min}. Otherwise, plate current would decrease and grid current would become excessive resulting in a reduction in output power and excessive grid losses. Thus in figure 11-5, e_{max} is made 20 volts less than e_{min} in order to ensure that e_{max} will never exceed e_{min}.

The plate tank circuit in figure 11-4 may have an artificial load applied to it for the purpose of tuning the amplifier prior to coupling the antenna to it. This load may take the form of an incandescent lamp of approximately the same power rating as the amplifier supplying it. The load is coupled to the tank by means of the link coil, L3.

The plate supply voltage is reduced and tuning capacitor C4 is adjusted for resonance, as indicated by the dip in the plate milliammeter (line current is a minimum at resonance). The sharp decrease in plate current is accompanied by a corresponding increase in tank current. As resonance is approached, grid current increases as plate current decreases. The load on the tank circuit may be increased by moving link coupling inductor L3 closer to tank coil L2 and increasing the plate supply voltage to the normal value.

The increased load on the amplifier increases the current through the lamp and decreases the current in the plate tank circuit. The decrease in current in the resonant tank is accompanied by a decrease in voltage, e_c, across the tank. Thus in figure 11-5, e_{min} becomes larger ($e_{min} = E_b - e_c$) and plate current increases with the load. The space current increases because plate voltage is increased during that portion of the cycle when the triode is conducting. For a given filament emission, plate current increases and grid current decreases as the load on the tank increases.

The transmitter output may be coupled to the antenna by tuning the antenna to resonance and coupling it to the final amplifier tank by means of a link coil similar to L3. The artificial load is removed by moving L3 away from L2, as the antenna load is applied to the tank.

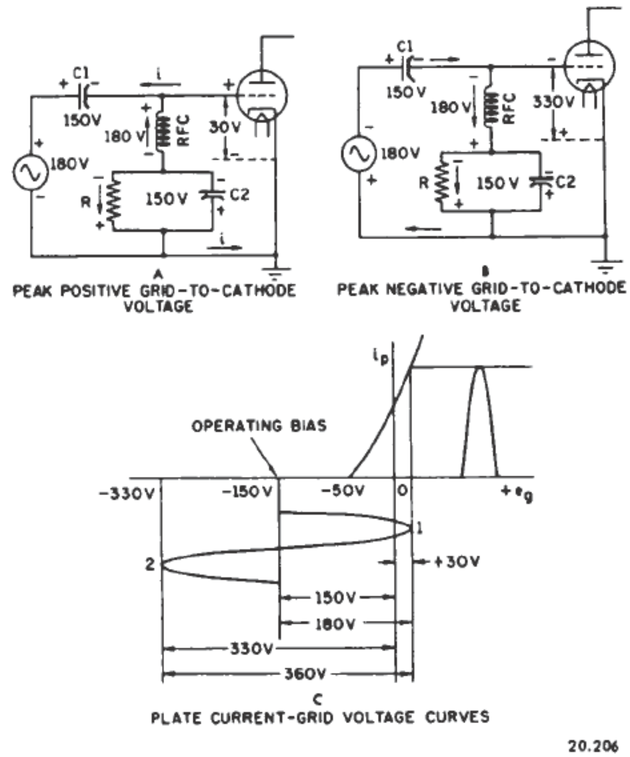


Figure 11-6.—Analysis of grid leak bias.

BIAS METHODS

Most of the bias methods used in receivers can also be employed in transmitters. However, because of the power output requirements of transmitters, class-B or class-C r-f amplifiers are used most often and these normally employ grid-leak bias. Grid-leak bias depends on grid current flow for a portion of the input cycle. This type of bias is not generally used in receivers because most r-f amplifiers in receivers operate class A with no grid current.

A grid-leak bias circuit is shown in figure 11-6. The triode is assumed to operate as a class-C amplifier with a peak driving voltage of 180 volts, a cutoff bias of -50 volts, and an operating bias of -150 volts. The polarities and magnitudes of the voltages for the condition of maximum positive grid-to-cathode voltage are shown in figure 11-6,A. The polarities and magnitudes of the voltages for the condition of maximum negative grid-to-cathode voltage are shown in figure 11-6,B. The wave forms of the grid-driving voltage and plate current are shown in figure 11-6,C.

Capacitor C1 blocks grid current from the signal source. Capacitor C2 is an r-f bypass capacitor that holds the lower end of the r-f choke at r-f ground potential. Grid-leak bias voltage is developed across grid resistor R.

In this example, R has a resistance of 15 k-ohms and the grid current is 10 milliamperes. The voltage across R is 10 X 15, or 150 volts. The voltage across R without C2 is a series of half-wave pulses. The fact that C2 is shunted across R smoothes these pulses into a steady d-c bias voltage. The electron flow through R makes the grid end negative and the cathode end positive.

The a-c driving voltage is developed across the r-f choke which presents high inductive reactance to the r-f input and low effective resistance to the d-c grid current. The capacative reactances of capacitors C1 and C2 are low so that the a-c voltage across them is negligible.

In figure 11-6,A, the grid conducts and the positive peak grid-to-cathode voltage is 180-150, or + 30 volts. During the time grid current flows, C1 charges up to 180-30, or 150 volts, and the low impedance of the conducting cathode-to-grid circuit bypasses grid circuit around the r-f choke and resistor R.

Also during the time the grid is conducting, C2 discharges through R, thus maintaining the operating bias of -150 volts. C2 is sufficiently large that its voltage does not change appreciably during discharge.

On that portion of the input cycle when the grid is negative with respect to the cathode, no grid current flows. At the instant pictured in figure 11-6,B, the grid is maximum negative with respect to the cathode. The path for the a-c input voltage is to the right through C1 and down through the r-f choke and grid resistor R. Capacitor C2 charges up as C1 discharges. C1 is sufficiently large that its d-c potential does not fall appreciably during discharge.

Grid-leak bias has the desirable characteristic of adjusting its value automatically when the amplitude of the grid driving voltage varies in magnitude. For example, an increase in driving voltage increases the operating bias, which checks the increase in grid current; or if the grid driving voltage decreases, the decrease in grid current is checked by a shift of the operating bias in a positive direction. The correction is automatic in either case because it is the flow of grid current through the grid resistor that produces the operating bias. Thus the grid current is maintained at the proper value automatically over an appreciable range of input voltage.

In order to maintain grid-leak bias, grid current must flow a part of each cycle. Removing the driving voltage or lowering its amplitude below the value that drives the grid positive causes a loss in grid current and operating bias. Plate current then becomes dangerously high and the tube may be damaged.

Separate bias may be employed to prevent the condition of excessive plate current when grid bias is removed. Figure 11-7 shows a circuit that uses protective bias in series with grid-leak bias. In this example, the grid bias is developed as a result of the flow of 10 ma of grid current through a 10 k-ohm resistor. The voltage across R is 10 X 10, or 100 volts. The protective bias is provided by a 50-volt battery. Cutoff bias for the triode is assumed to be -50 volts. The total operating bias is -100 -50, or -150 volts.

Grid current flows down through R and into the negative terminal of the battery. The direction is such as to charge the battery and raise its terminal voltage when grid current flows. The energy to charge the battery comes from the a-c driving voltage source. Thus, the voltage regulation of a separate bias supply, such as the bias battery, is negative (the greater

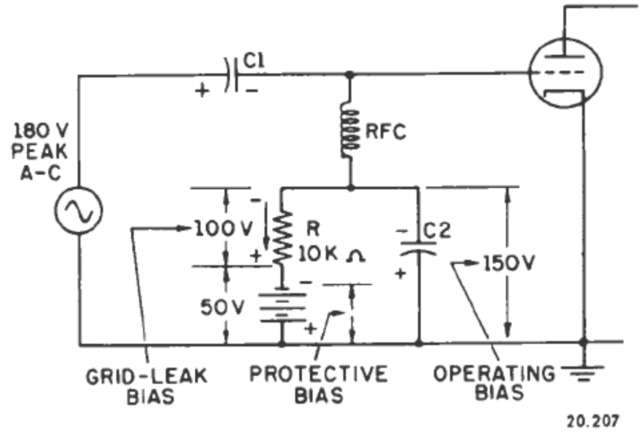


Figure 11-7.—Combination grid-leak and battery bias.

the grid current, the higher is the bias supply voltage).

Other forms of protective bias, including cathode bias, and separate bias supplies are discussed in chapter 3.

TRANSMITTER ELECTRON TUBES

Electron tubes used in transmitters differ very little from those used in receivers except in size. Because most transmitter tubes are power tubes designed to amplify high voltage and heavy current, they must be of much larger and heavier construction. The plate dissipation of a tube is the difference between the plate power input and the power output. If this dissipation is greater than normal, the plate will become very hot, sometimes glowing with a red color from this heat. If the heat becomes intense, gases may develop within the tube, making it unsatisfactory for further use. A transmitter should not be operated for any appreciable time if the plates become red unless the service manual for the set states that this condition is normal for operation. Loss of bias, insufficient grid excitation, or improper tuning may cause overheating of a transmitter tube.

A common screen-grid tube similar to the type 860 used as a class-C r-f power amplifier in Navy transmitters is shown in figure 11-8. It requires no neutralization and is used in medium-power transmitters.

The thoriated tungsten filament has a rating of 10 volts and 3.25 amperes. The two filament leads and the screen-grid are brought out through

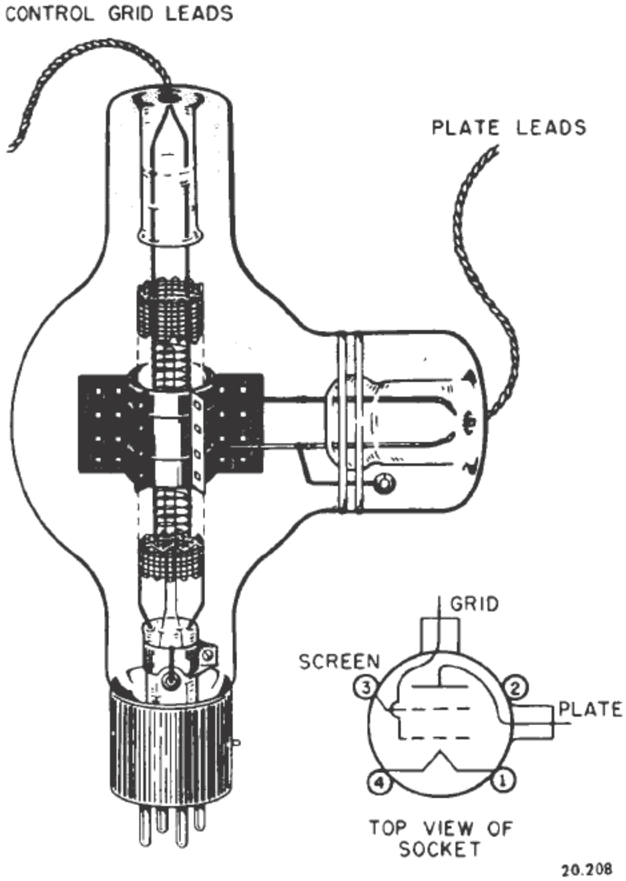


Figure 11-8.—Medium-power transmitting tetrode.

the base pin connections. The screen-grid voltage is 300 volts. The control grid is connected to stranded leads that are brought out through a separate seal in the upper arm of the bulb. Control-grid voltage for class-C operation is -150 volts and grid current is 15 milliamperes. The driving power required is about 7 watts. The plate is connected to stranded leads that are brought out through a separate seal in the side arm of the bulb. Plate voltage for class-C operation is 3,000 volts.

This tube is mounted in a vertical position. The plate shows a dull red color when it is operated at the maximum plate dissipation rating of 100 watts. Normal plate current is 85 ma when the power output is 165 watts. Plate efficiency of this amplifier is about 65 percent when the plate dissipation is 90 watts. The amplification factor of this tetrode is 200, its transconductance is 1,100 micromhos, and its a-c plate resistance is approximately 182,000 ohms.

Because plate current flows for only a portion of each cycle, tubes are better able to dissipate the heat developed and thus have a longer life if they are operated class C. Additional details on the construction and operation of electron tubes are given in chapter 1.

NEUTRALIZATION

A transmitter r-f amplifier having a plate tank and grid tank circuit both tuned to the same frequency resembles a tuned-plate tuned-grid oscillator. Unless some precaution is taken to prevent it, the amplifier may break into oscillation, causing a very unstable operating condition. If a voltage is fed back from plate to grid in phase with the grid signal, oscillation will occur. If the voltage fed back is 180° out of phase with the grid signal, the action is degenerative and oscillations will be stopped.

Neutralization is a process of balancing the voltage fed back by the interelectrode capacitance of the tube with an equal voltage of opposite polarity. Dividing the plate circuit so that the neutralization voltage is developed across a part of it is called PLATE NEUTRALIZATION. If the voltage of neutralization is developed in the grid circuit the arrangement is called GRID NEUTRALIZATION.

In figure 11-9, the plate tank coil has a center-tap connection. The voltage between point A and ground is 180° out of phase with the voltage from point B to ground. Feedback through the plate-to-grid capacitance of the triode produces a voltage across the grid input circuit that is in phase with the grid excitation voltage and therefore tends to cause the amplifier to break into oscillations.

The neutralizing capacitor, C_N, couples a portion of the voltage between point B and ground to the grid input circuit. This action is degenerative and tends to block oscillations. A simplified equivalent circuit is shown in figure 11-10.

By the adjustment of C_N, the voltage fed back to the grid through C_N is made equal to the voltage fed back through the tube.

One method of determining the correct adjustment for C_N is to apply the input r-f voltage with normal filament but no plate voltage. A pick-up coil near the plate tank is fed to

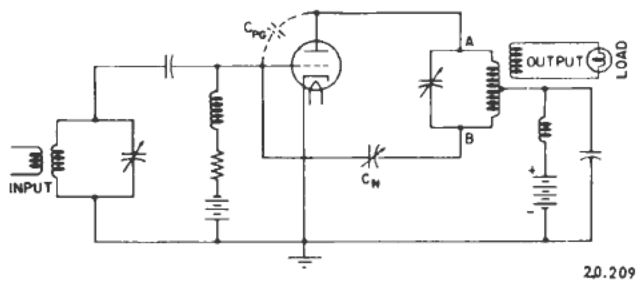


Figure 11-9.—Plate neutralization.

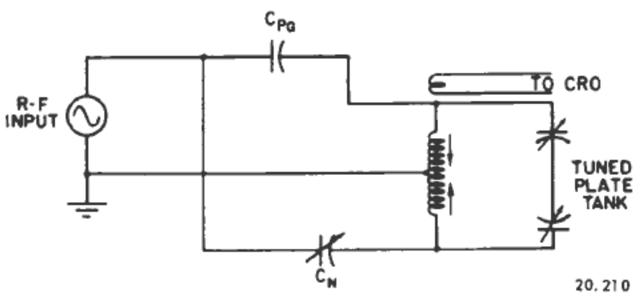


Figure 11-10.—Equivalent circuit for plate neutralization.

the vertical input of a cathode-ray oscilloscope. C_N is adjusted so that no r-f voltage appears on the scope when the plate tank is tuned to resonance. Under these circumstances, the r-f current divides equally through C_{PG} and C_N. The resulting r-f currents in the plate tank flow in opposite directions and cancel the tank inductive effect so that no resonant build-up occurs between the coil and capacitors. A neon glow bulb, a loop of wire attached to a small flashlight bulb, or a sensitive r-f galvanometer may be used if an oscilloscope is not available.

If there is a milliammeter in the amplifier grid circuit, the adjustment of C_N may be made by observing the grid meter as the plate tank is tuned through resonance, with no plate voltage applied. When there is an unbalance between C_{pg} and C_N, the plate becomes alternately positive and negative as the plate tank strikes resonance. On positive swings, plate current flows. As the plate tank circuit is tuned to the resonant frequency, some of the electrons that were going to the grid now go to the plate, thereby causing a dip in grid current.

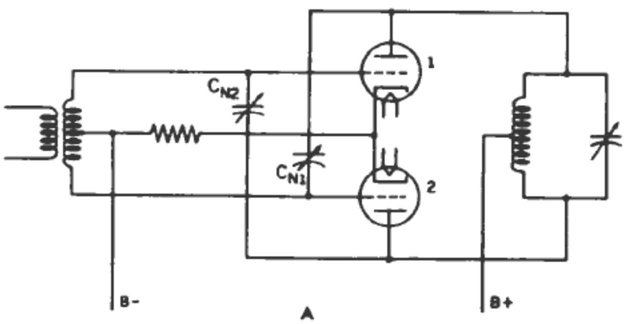
However, as C_N is adjusted to neutralize the amplifier stage, the r-f current from the input stage divides equally and flows in opposite directions in the two halves of the plate-tank coil, thus canceling the inductive effect of the coil and preventing the buildup of resonance in the tank. There is no rise in tank current and voltage, and the triode plate remains at zero potential. Therefore, with C_N properly adjusted, no dip in grid current occurs as the plate tank is tuned through the resonant frequency.

Another indication of the neutralized condition of the amplifier stage is obtained by observing the reaction on the plate and grid currents of the INPUT stage as the amplifier plate tank (with no voltage applied) is tuned through resonance. If the amplifier is not properly neutralized, the resonant build-up in the plate tank varies the load on the exciter stage as the plate tank is tuned through resonance. Thus, in the exciter stage, the grid current decreases and the plate current increases as the amplifier plate tank strikes resonance. However, when C_N is properly adjusted to neutralize the amplifier stage, no increase in loading occurs on the exciter stage as the plate tank is tuned to the resonant frequency, and the grid and plate currents in the exciter stage remain the same.

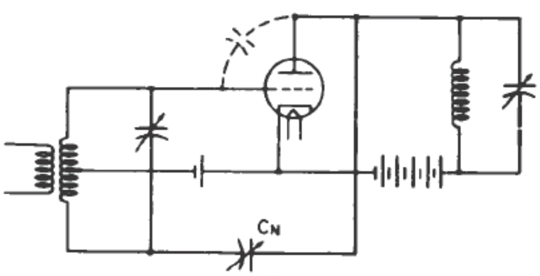
In some transmitter circuits it is more convenient to turn off the filament voltage on the amplifier stage instead of removing plate voltage. If this is done, the process of neutralizing the amplifier is carried out the same way, except that no current flows in the amplifier grid circuit. The absence of radio frequency in the amplifier plate tank, as evidence of the correct adjustment of C_N, may be determined by the effect on the exciter stage or on an r-f pickup coil and associated indicator, as previously mentioned.

To obtain complete neutralization the transmitter must be designed so that there is no coupling between the input (grid) and output (plate) circuits of the amplifier stages other than through the interelectrode capacitance of the tubes. The input and output inductors must be shielded from each other or mounted at right angles to reduce any coupling between them to a negligible amount. The wiring and arrangement of component parts must reduce stray capacitive or inductive coupling to a minimum.

Cross neutralization of a push-pull amplifier is accomplished as shown in figure 11-11, A. The plate of tube 1 is connected to the grid of tube 2 through neutralizing capacitor C_{N1}, and the plate of tube 2 is connected to the grid of tube 1 through neutralizing capacitor C_{N2}. The voltage fed back through the interelectrode capacitance (plate-to-grid) of tube 1 to the input circuit for that tube is counteracted by the voltage fed back through C_{N2}. The voltage fed back through the interelectrode capacitance (plate-to-grid) of tube 2 to the input circuit for tube 2 is counteracted by the voltage fed back through neutralizing capacitor C_{N1}.



A
PUSH-PULL NEUTRALIZATION



B
RICE SYSTEM OF NEUTRALIZATION

20.211

Figure 11-11.—Neutralization circuits.

A special method of amplifier neutralization, known as the Rice system, is shown in figure 11-11, B. This arrangement is similar to that of figure 11-11, A, except that the Rice system utilizes a split input circuit in place of a split output circuit. The voltage fed back from the plate to the grid through C_{pg} is counteracted by the voltage fed back through C_N. This circuit is a form of grid neutralization.

The use of a well-shielded tetrode or pentode makes neutralization unnecessary, because the plate and grid are shielded from each other by the screen grid and its associated r-f bypass capacitor which holds the screen at r-f ground potential. However, the overall efficiency of these tubes is not as great as that of triodes, since there is a screen-grid power loss. The high impedance of such tubes makes them more suitable for voltage amplifiers than for final output stages where power output is the principal factor. Low excitation requirements make tetrodes and pentodes especially suitable for use in the intermediate stages of a transmitter.

PARASITIC OSCILLATIONS

Circuit conditions in an oscillator or amplifier may be such that secondary oscillations occur at frequencies other than that desired. The frequency of these oscillations is neither that of the fundamental nor its harmonics. Such oscillations are appropriately termed PARASITIC OSCILLATIONS and are to be avoided. The energy required to maintain parasitic oscillations is wasted so far as useful output is concerned. A circuit afflicted with parasitics has low efficiency and frequently operates erratically.

Figure 11-12 shows some of the incidental circuits that may give rise to parasitics in a transmitter amplifier. The dotted lines in figure 11-12, A, outline a possible h-f circuit, and those in figure 11-12, B, outline a possible u-h-f circuit. The part of a transmitter that constitutes a possible l-f parasitic circuit is shown in figure 11-12, C.

Parasitic oscillations may be suppressed by placing resistors or chokes at appropriate positions in the circuits, or by slightly modifying the existing values of certain circuit elements. Also, care should be used in the physical arrangement and wiring of parts. Parasitic suppressors consisting of an inductor and resistor in parallel are sometimes inserted in the grid and plate leads of an r-f amplifier to suppress high-frequency parasitic oscillations. The resistor has a resistance of from 50 to 100 ohms. The path through the r-f choke has a low impedance to normal frequencies and a high impedance to high frequencies (parasitics). Thus, normal frequency currents flow through the r-f choke without attenuation. The path of the h-f parasitic currents is through the resistor which dissipates the feedback energy in the form of heat and reduces the magnitude of the parasitic oscillations to a negligible amount.

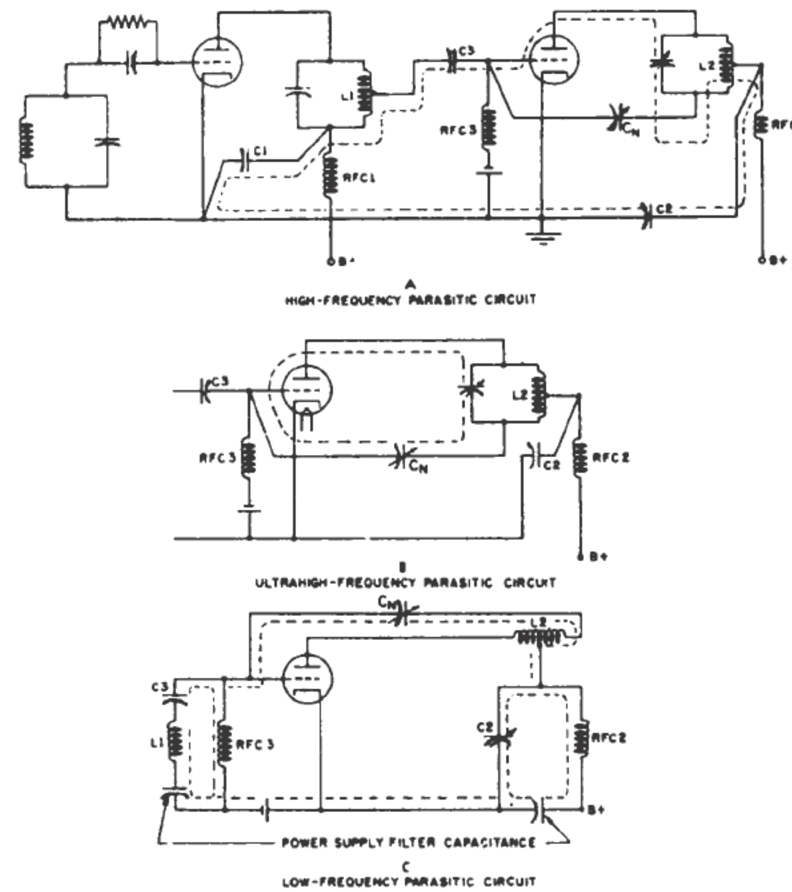


Figure 11-12.—Parasitic oscillatory circuits in a transmitter.

KEYING SYSTEMS

Keying a c-w transmitter causes an r-f signal to be radiated ONLY when the key contacts are closed. When the key is open the transmitter does not radiate energy. Keying is accomplished in either the oscillator or amplifier stages of a transmitter. A number of different keying systems are used in Navy transmitters.

In most Navy transmitters the hand telegraph key is at low potential with respect to ground. The keying bar is usually grounded to protect the operator. Generally a keying relay with its contacts in the center-tap lead of the filament transformer is used to key the equipment. Because one or more stages use the same filament transformer, these stages are also keyed. The class C final amplifier, when operated with fixed bias is usually not keyed because with no excitation applied no current flows. Hence, keying the final amplifier along with the other stages is not necessary.

Two methods of oscillator keying are shown in figure 11-13. In figure 11-13, A, the grid circuit is closed at all times, and the key opens and closes the negative side of the plate circuit. This system is called PLATE KEYING. When the key is open, no plate current can flow and the circuit does not oscillate. In figure 11-13, B, the cathode circuit is open when the key is open and neither grid current nor plate current can flow. Both circuits are closed when the key is closed. This system is called CATHODE KEYING. Although the circuits of figure 11-13 may be used to key amplifiers, other keying methods are generally employed because of the larger values of plate current and voltage encountered.

Two methods of blocked-grid keying are shown in figure 11-14. The key in figure 11-14, A, shorts cathode resistor R1, allowing normal plate current to flow. With the key open, reduced plate current flows up through resistor R1, making the end connected to grid resistor R_g

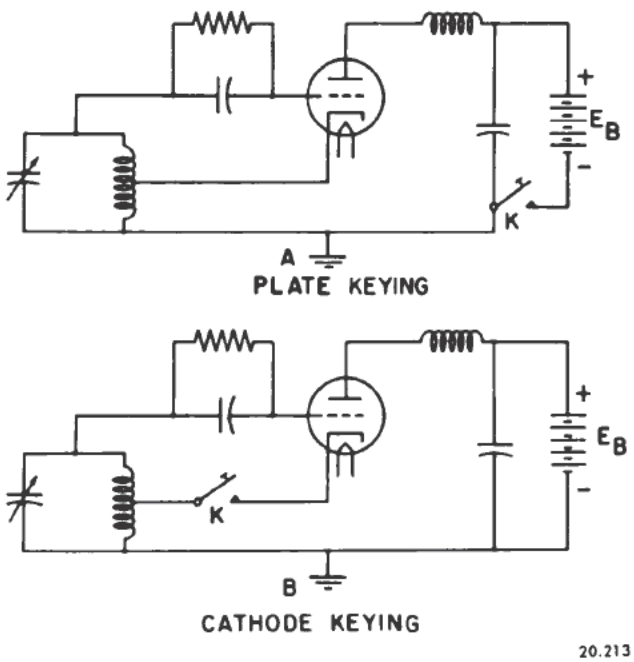


Figure 11-13.—Oscillator keying.

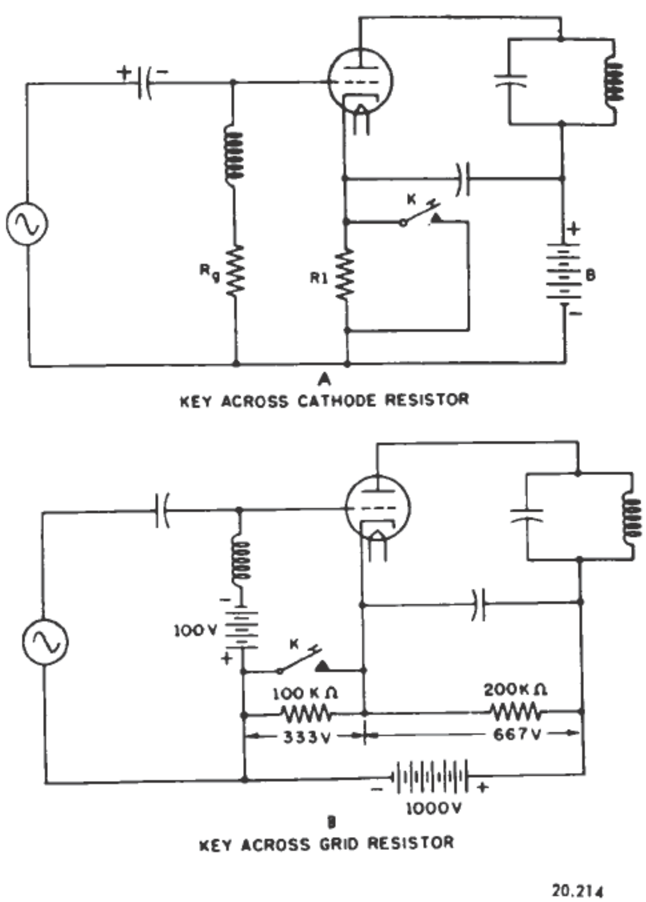


Figure 11-14.—Blocked grid keying.

negative. If R1 has a high enough value, the bias developed is sufficient to cause practical cutoff of plate current. Complete cutoff is not possible because the bias voltage developed across R1 depends on the flow of some plate current through it. However, the blocking is sufficient for practical keying. Depressing the key short-circuits R1, thus increasing the bias above cutoff and allowing the normal flow of plate current. Grid resistor R_g is the usual grid-leak resistor for normal bias. This method of keying is applied to the buffer stage in a c-w transmitter.

The blocked-grid keying method shown in figure 11-14, B, affords complete cutoff of plate current and is one of the best methods for keying amplifier stages in c-w transmitters. In the voltage divider, with the key open, two-thirds of 1,000 volts, or 667 volts, are developed across the 200 k-ohm resistor and one-third of 1,000 volts, or 333 volts, are developed across the 100 k-ohm resistor. The grid bias is -100 -333, or -433 volts. Because this is below cutoff, no plate current flows. The plate voltage is 667 volts. With the key closed, the 100 k-ohm resistor is shorted out and the voltage across the 200 k-ohm resistor is increased to 1,000 volts. Thus, the plate voltage becomes 1,000 volts at the same time the grid bias becomes -100 volts. Grid bias is now above cutoff and the amplifier triode conducts. Normal amplifier action follows.

Where greater frequency stability is required, the oscillator should remain in operation continuously while the transmitter is in use. This procedure keeps the oscillator tube at normal operating temperature and offers less chance for frequency variation to occur each time the key is closed. If the oscillator is to operate continuously and the keying is to be accomplished in an amplifier stage following the oscillator, the oscillator circuit must be carefully shielded to prevent radiation and interference to the operator while he is receiving.

In transmitters using a crystal-controlled oscillator the keying is almost always in a stage following the oscillator. In the large transmitters (75 watts or higher) the ordinary hand key cannot accommodate the plate current without excessive arcing. Moreover, because of the high plate potentials used it is dangerous to operate a hand key in the plate circuit. A slight slip of the hand below the key knob might result in a bad shock; or, in the case of defective r-f plate chokes, a severe r-f burn might be incurred. In these larger transmitters, some local low-voltage supply, such as a battery or the filament supply to the transmitter, is used with the hand key to open and close a circuit through the coils of a keying relay. The relay contacts in turn open and close the keying circuits of the amplifier tubes. A schematic diagram of a typical relay-operated keying system is shown in figure 11-15. The hand key closes the circuit from the low-voltage supply through coil L of the keying relay. The relay armature closes the relay contacts as a result of the magnetic pull exerted on the armature. The armature moves against the tension of a spring. When the hand key is opened, the relay coil is deenergized and the spring opens the relay contacts.

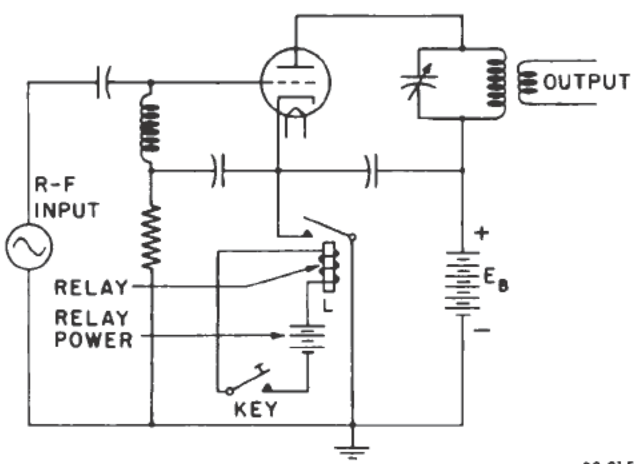


Figure 11-15.—Circuit for a relay-operated keying system.

Theoretically, keying a transmitter should instantly start and stop radiation of the carrier completely. However the sudden application and removal of power creates rapid surges of current which cause interference in nearby receivers. Even though such receivers are tuned to frequencies far removed from that of the transmitter, interference is present in the form of clicks or thumps. To prevent such interference, key-click filters are used in the keying systems of radio transmitters. Two types of key-click filters are shown in figure 11-16.

The capacitors and r-f chokes in both circuits of figure 11-16 prevent surges of current. The choke coil, L, causes a lag in the current when

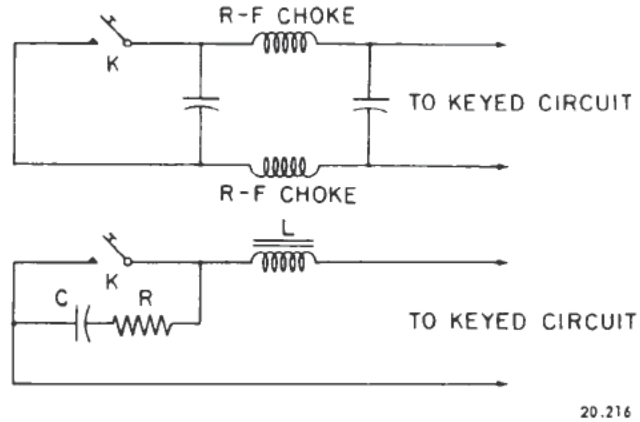


Figure 11-16.—Key-click filters.

the key is closed, and the current builds up gradually instead of instantly. Capacitor C charges up as the key is opened and slowly releases the energy stored in the inductor magnetic field. Resistor R controls the rate of charge and discharge of capacitor C and also prevents sparking at the key contacts by the sudden discharge of C when the key is closed.

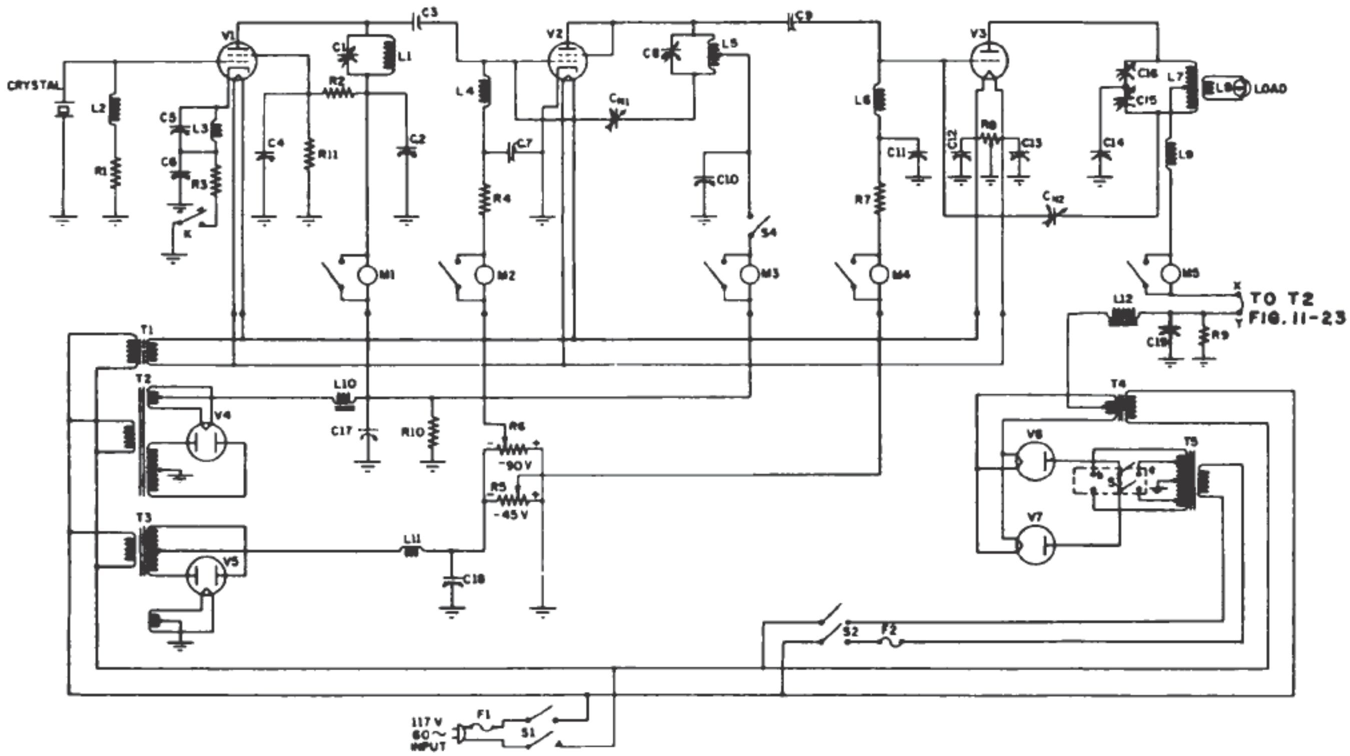
Another difficulty that may be encountered in keying a transmitter is the presence of a back wave. A back wave results when some r-f energy leaks through to the antenna even though the key is open. The effect is as though the dots and dashes were simply louder portions of a continuous carrier. It may be difficult to distinguish the dots and dashes under such conditions. Back-wave radiation is usually the result of incomplete neutralization.

CIRCUIT OF A C-W TRANSMITTER

The circuits of a small, shore-based, c-w transmitter and its power supply are shown in figure 11-17. The transmitter includes 3 tubes, oscillator, buffer, and power amplifier. The power supply is designed to plug into a 117-volt 60-cycle source.

To simplify the wiring diagram, individual meters are shown in the plate and grid leads for each stage. In practice, one meter and a multi-terminal switch for the grid circuits and a similar combination for the plate circuits would suffice.

The crystal oscillator stage is keyed in the cathode circuit of beam-power tetrode V1, which is similar to a tuned-grid tuned-plate oscillator



20.217

Figure 11-17.—C-w transmitter and power supply.

except that the crystal replaces the tuned-grid circuit. The stage oscillates with the key closed when the plate tank, L1C1, is tuned to a frequency close to that of the crystal. Grid-leak bias is developed across R1. The key-click filter circuit consists of R3, L3, C5, and C6.

The r-f excitation voltage appears across the r-f choke, L2.

Screen voltage is applied through resistor R2. Capacitor C4 is the screen bypass capacitor, which holds the screen at r-f ground potential. Capacitor C2 is the plate bypass capacitor, which places the lower end of the plate tank circuit at r-f ground. The output voltage across the tank is coupled to the buffer stage through capacitor C3. Meter M1 indicates the sum of the plate and screen currents in V1 and the current in R11. The output power of the oscillator is about 5 watts at the crystal frequency.

The buffer stage, V2, employs a triode-connected beam-power tetrode biased for class-C operation and using plate neutralization. Neutralizing capacitor C_{N1} couples the correct amount of feedback voltage to the grid to neutralize the stage.

The oscillator output voltage is developed across the r-f choke, L4. Capacitor C7 is an r-f bypass capacitor that places the lower end of the choke at r-f ground potential. In addition to separate bias, the grid circuit uses the voltage developed across resistor R4 as automatic bias. Meter M2 indicates the buffer grid current. The stage may be used as a frequency doubler, in which case the plate tank tuning capacitor, C8, is adjusted so that the tank strikes resonance at the second harmonic of the crystal fundamental frequency. Capacitor C10 is a plate bypass capacitor, which holds the center tap of tank coil L5 at r-f ground potential. Plate current is indicated by meter M3. The stage develops an output power of about 15 watts. The output voltage is coupled through capacitor C9 to the grid circuit of power amplifier triode V3. The excitation voltage appears across r-f choke L6.

The power amplifier triode is biased for class-C operation and the plate tank (L7, C15, and C16) is tuned to the frequency of the grid-excitation voltage. The antenna is coupled to the final tank through link coil L8. Resistor R8 is center-tapped to provide a common return for the plate and grid circuits and to prevent the 60-cycle filament voltage from modulating the r-f grid voltage. Capacitor C11 is an r-f bypass capacitor that holds the lower end of r-f choke L6 at r-f ground potential. Capacitors C12 and C13 are filament bypass capacitors that keep r-f current out of the filament leads. Capacitor C14 effectively places the center tap of plate tank coil L7 at r-f ground potential. The stage employs plate neutralization. The neutralizing capacitor, C_{N2}, couples a portion of the voltage between the lower end of the plate tank and ground back to the grid, to neutralize the voltage fed back through the plate-to-grid capacitance of V3. The r-f choke, L9, keeps r-f currents out of the plate supply lead. Meter M4 indicates grid current and M5 plate current. Power output is approximately 100 watts when the plate supply voltage is 1,000 volts.

The power supply includes three full-wave rectifiers to provide plate, screen, and grid voltages for the transmitter tubes. It also includes filament supply transformers for both transmitter and power supply tubes. Transformer T1 supplies filament power to V1, V2, and V3.

Switches S1 and S2 are arranged so that S1 must be closed before S2 can be energized. Thus, filament and bias voltages are provided for V2 and V3 before plate voltage can be applied to V3. The primaries of all the power supply transformers are connected in parallel and are supplied by a 117-volt 60-cycle source. Power supplies are treated in chapter 3.

TUNING A C-W TRANSMITTER

All radio transmitters must be properly tuned to ensure efficient operation on the assigned frequency. Transmitters are always tuned on c-w even if m-c-w or voice modulation is to be used. Plate-current meters are generally used to indicate proper adjustment of the r-f stages. All stages, with the exception of the oscillator, are always adjusted or tuned for minimum plate current. If a stage is not tuned to resonance, the plate current will be high and high plate dissipation, power loss, and low output will result. When a stage is loaded by another stage or an antenna, the plate current of the stage in question must be rechecked for circuit resonance (minimum plate current) after loading.

If a gassy tube is present in the equipment, plate current in that stage cannot be brought to the proper minimum, and grid current will remain too low or may even reverse. The tube will act as though there were a short between the grid and cathode and much of the energy supplied to the stage will be grounded and lost. This condition can be recognized by any of the indications just mentioned and by a violet-colored glow between the tube elements. The only remedy for this condition is a new tube.

The coupling of the tuned antenna to the transmitter is accompanied by three principal effects—

1. The antenna current (r-f energy) increases.
2. The plate current of V3, as indicated on meter M5, increases.
3. The grid current of V3, as indicated on meter M4, decreases.

In the final tune-up process the act of moving the antenna link coil closer to the final power-amplifier tank coil usually detunes the final stage slightly. This detuning results in an increase in the indication on plate-current meter M5. To correct this condition, plate tank capacitors C15 and C16 should be readjusted until minimum current is indicated on M5. This adjustment results in a further increase in output power and antenna current. The transmitter is then ready for keying.

CAPABILITIES OF A C-W TRANSMITTER

In view of the comparative slowness and inconvenience of keying the dots and dashes of Morse code, it might seem that radiotelegraphy would be superseded by radiotelephony, which uses modulated waves. C-w transmission, however, has four distinct advantages over radiotelephony.

1. Radiotelegraph transmitters have a greater transmission range than radiotelephone transmitters of the same power output because, in the latter, speech from a distant point may be audible, but not intelligible.
2. C-w signals may be picked up by code receivers that are capable of rejecting most of the interference characteristic of all r-f waves.
3. The comparable radiotelegraph transmitter is smaller and much simpler to operate.

4. Within a given frequency band, many more radiotelegraph transmitters than radiotelephone transmitters may be operated without interference.

AMPLITUDE-MODULATED RADIOTELEPHONE TRANSMITTER

Amplitude modulation has been defined as the variations of the magnitude of the r-f output of a transmitter at an audio rate. In other words, the r-f energy increases and decreases in accordance with the energy delivered by the audio modulator. If the audio frequency is high, the radio frequency varies in amplitude more rapidly than if the audio frequency were low. If the audio note is loud in volume, the r-f energy is increased and decreased by a larger percentage than if the audio note were soft. Thus, the r-f variations correspond with the a-f variations.

A microphone or a similar device is used to produce the electrical equivalent of the audio signal. The signal is then amplified by means of an a-f amplifier before it is fed to the modulator. Because of the importance of microphones in the communications chain a brief description of some of the more common microphones, together with their characteristics, follows.

MICROPHONES

A microphone is essentially an energy converter that changes acoustical (sound) energy into corresponding electrical energy. When one speaks into a microphone, the audio pressure waves strike the diaphragm of the microphone and cause it to move in and out in accordance with the instantaneous pressure delivered to it. The diaphragm is attached to a device that causes current to flow in proportion to the instantaneous pressure applied to the diaphragm. Many devices can perform this function, each having characteristics that make its use advantageous under a given set of circumstances.

Most microphones, with the exception of the carbon microphone, are relatively inefficient—that is, the output in electrical energy is considerably less than the input in acoustical energy. Some, however, are more efficient than others; and some have a better frequency response than others. The characteristics of microphones, therefore, will be discussed before the various types of microphones are discussed. Microphones are rated according to their (1) frequency response, (2) impedance, and (3) sensitivity.

Frequency Response

For good quality, the electrical waves from a microphone must correspond closely in magnitude and frequency to the sound waves that cause them, so that no new frequencies are introduced. The frequency range of the microphone (that range of frequencies over which the microphone is capable of responding) must be no wider than the desired over-all response limits of the system with which it is to be used. The microphone response should be uniform, or flat, within its frequency range and free from sharp peaks or dips such as those caused by mechanical resonances. To aid in attaining this condition, some form of damping may be employed.

Impedance

Crystal microphones have impedances of several hundred thousand ohms; whereas magnetic and dynamic microphones have impedances that range from 20 to 600 ohms. The impedance of a microphone is usually measured between its terminals when some arbitrary frequency in the useful audio range—for example, 1,000 cycles—is used.

The impedance of magnetic and dynamic microphones varies with frequency in much the same manner as that of any coil or inductance—that is, the impedance rises with increasing frequency. The actual impedance of a microphone is of importance chiefly as it is related to the load impedance into which the microphone is designed to operate. If the load has a high impedance, the microphone should have a high impedance, and vice versa. Of course, impedance-matching devices may be used between the microphone and its load.

A long transmission line between the microphone and the amplifier input tends to seriously attenuate the high frequencies, especially if the impedance of the microphone is high. This action results from the increased capacitive effect of the line at the higher frequencies. If the microphone has a high impedance the high-frequency currents drawn through the inherent capacity of the line cause an increased voltage loss within the microphone, and therefore less voltage is available at the load. Because the voltage generated by the microphone is very minute, all losses in the microphone and the line must be kept to a minimum. At the lower frequencies the capacitive effect is less and the losses are correspondingly less. If the microphone has a low impedance a correspondingly lower voltage drop will occur in the microphone and more voltage will be available at the load.

Because many microphone lines aboard ship are long, it is necessary to use low-impedance microphones in order to preserve a satisfactory signal voltage level over the required audio band at the input grid of the amplifier. Because the crystal microphone is essentially a capacitor, the length of the transmission line used with this microphone will affect all audio frequencies equally.

Sensitivity

The sensitivity or efficiency of a microphone is usually expressed in terms of the electrical power level which the microphone delivers to a terminating load (the impedance of which is equal to the rated impedance of the microphone) compared to the acoustical intensity level or pressure of the sound energy that is being picked up. Because sound energy at the input is being compared with electrical energy at the output, some basis of comparison must be established.

One method is to assume that a microphone has a sensitivity of 0 db (the level of comparison) if a force of 1 dyne per square centimeter on the diaphragm produces an output of 1 volt on open circuit. The pressure of 1 dyne per square centimeter was chosen because that is the approximate pressure produced by normal speech on the diaphragm of a microphone held a few inches from the mouth. The usual method, however, is to assume that the 0 db level represents an input of 1 dyne per square centimeter (as in the first method) and an output of 1 milliwatt. If it is further assumed that the milliwatt is developed in 600 ohms, then dbm or volume units (vu) may be used. Decibels and the various power-level units are given in chapter 7. a power-level chart is shown in figure 7-11.

Suppose a microphone is rated at -80 db. This rating means that the energy output is much less than the energy input. Actually, the output is 10^{-8} milliwatt for an equivalent input of 1 milliwatt, and this is equivalent to -80 db. This rating may be demonstrated by the use of the equation.

$$\text{db} = 10 \log_{10} \frac{P2}{P_1}$$

$$\text{db} = 10 \log_{10} \frac{(10)^{-8}}{1} = -80 \text{ db.}$$

It is important to have the sensitivity of the microphone as high as possible. High sensitivity means a high electrical power output level for a given input sound level. High microphone output levels require less gain in the amplifiers used with them and thus provide a greater margin over thermal noise, amplifier hum, and noise pick-up in the line between the microphone and the amplifier.

When a microphone must be used in a noisy location, an additional desirable characteristic is the ability of the microphone to favor sounds coming from a nearby source over random sounds coming from a relatively greater distance. Microphones of this type tend to cancel out random sounds and to pick up only those sounds originating a short distance away. When talking into this type of microphone the lips must be held as close as possible to the diaphragm. Directional characteristics also aid in discriminating against background noise.

Carbon Microphone

The carbon microphone is the most common type of microphone. It operates on the principle that a change in sound pressure on a diaphragm that is coupled to a small volume of carbon granules will cause a corresponding change in the electrical resistance of the granules.

The single-button carbon microphone (fig. 11-18, A) consists of a diaphragm mounted against carbon granules that are contained in a small cup. In order to produce an output voltage, this microphone is connected in a series circuit containing a battery and the primary of a microphone transformer. The pressure of the sound waves on the diaphragm, which is coupled to the carbon granules, causes the resistance of the granules to vary. Thus a

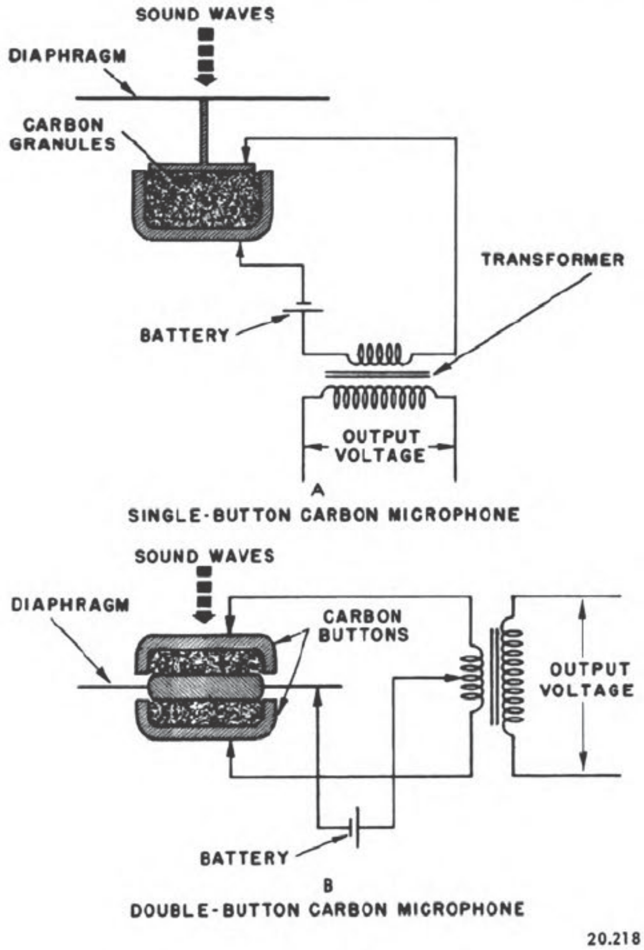


Figure 11-18.—Schematic diagram of carbon microphones.

varying direct current in the primary produces an alternating voltage in the secondary of the transformer. This voltage has essentially the same waveform as that of the sound waves striking the diaphragm. The current through a carbon microphone may be as great as 0.1 ampere, and the resistance may vary from about 50 to 90 ohms. The voltage developed across the secondary depends upon the turns ratio and also upon the rate of change in primary current. Normal output voltage of a typical circuit is from 3 to 10 volts peak across the secondary terminals.

The double-button carbon microphone is shown schematically in figure 11-18, B. Here one button is positioned on each side of the diaphragm so that an increase in pressure and resistance on one side is accompanied simultaneously by a decrease in pressure and resistance on the other. Each button is in series with the battery and one-half the transformer primary. The decreasing current in one half of the primary and the increasing current in the other half produce an output voltage in the secondary that is proportional to the sum of the primary signal components. This action is similar to that of push-pull amplifiers (chapter 7).

Commercial types of carbon microphones give essentially faithful reproduction from 60 to 6,000 cycles, and their output is of the order of -50 db.

The carbon microphone has the disadvantage of requiring an external voltage source; it may be noisy; and unless the necessary precautions are taken in the design the microphone tends to peak up (have mechanical resonance) at certain frequencies.

Dynamic Microphone

The dynamic, or moving-coil, microphone (fig. 11-19) consists of a coil of wire attached to a diaphragm and is so constructed that the coil is suspended and free to move in a radial magnetic field. Sound waves impinging on the diaphragm cause the diaphragm to vibrate. This vibration moves the voice coil through the magnetic field so that the turns cut the lines of force in the field. This action generates a voltage in the coil that has the same waveform as the sound waves striking the diaphragm.

The dynamic microphone requires no external voltage source, has good fidelity (approx. 20 to 9,000 cycles with proper damping), is

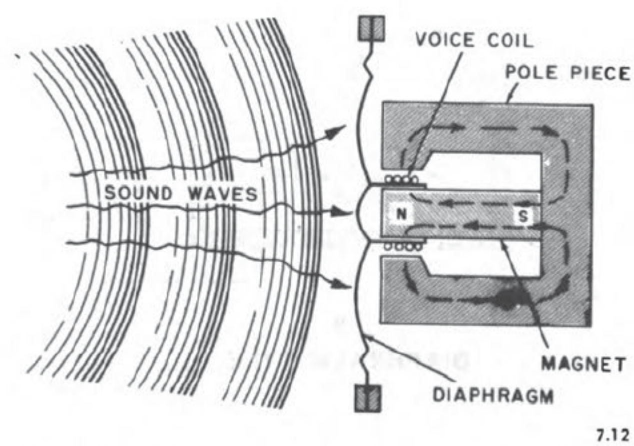


Figure 11-19.—Action of a dynamic microphone.

directional for high-frequency sounds, and has an output of the order of -85 db. The impedance of the dynamic microphone is low (50 ohms or less). Therefore, it may be connected to relatively long transmission lines without excessive attenuation of the high frequencies.

Crystal Microphone

The crystal microphone utilizes a property of certain crystals—such as quartz and Rochelle salt—known as the PIEZOELECTRIC EFFECT, treated in chapter 8. The bending of the crystal, resulting from the pressure of the sound wave, produces an emf across the faces of the crystal. This emf is applied to the input of an amplifier.

The crystal microphone (fig. 11-20) consists of a diaphragm that may be cemented directly on one surface of the crystal (fig. 11-20, A), or in some cases it may be connected to the crystal element through a coupling member (fig. 11-20, B). A metal plate, or electrode, is attached to the other surface of the crystal. When sound waves strike the diaphragm, the vibrations of the diaphragm produce a varying pressure on the surface of the crystal, and therefore an emf is induced across the electrodes. This emf has essentially the same waveform as that of the sound waves striking the diaphragm.

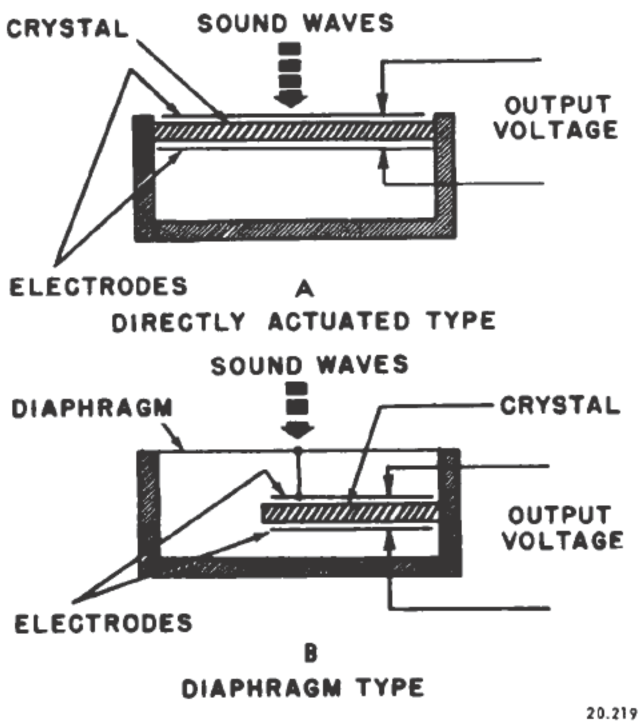


Figure 11-20.—Schematic diagrams of crystal microphones.

Rochelle salt is most commonly used in crystal microphones because of its relatively high voltage output. Schematic diagrams showing how crystal microphones function are given in figure 11-20.

Actually, a large percentage of crystal microphones employ some form of the bimorph cell. In this type of cell two crystals, so cut and oriented that their voltages will be additive in the output, are cemented together and used in place of the single crystal.

This type of microphone has high impedance (several hundred thousand ohms), is light in weight, requires no battery, is nondirectional, has good frequency response (up to 17,000 cps for the directly actuated type and between 80 to 6,000 cps for the diaphragm type), and has an output of the order of -70 db. However, the crystal microphone is sensitive to high temperature, humidity, and rough handling and therefore its use is restricted where these conditions prevail. Nevertheless, it is used extensively in broadcast work where its relatively high output is an advantage.

Magnetic Microphone

The magnetic, or moving-armature, microphone (fig. 11-21) consists of a permanent magnet and a coil of wire enclosing a small armature. Sound waves impinging on the diaphragm cause the diaphragm to vibrate. This vibration is transmitted through the drive rod to the armature, which vibrates in a magnetic field, thus changing the magnetic flux through the armature and consequently through the coil.

When the armature is in its normal position midway between the two poles, the magnetic flux is established across the air gap, and there is no resultant flux in the armature.

When a compression wave strikes the diaphragm, the armature is deflected to the right. Although a considerable amount of the flux continues to move in the direction of the arrows, some of it now flows from the north pole of the magnet across the reduced gap at the upper right, down through the armature, and around to the south pole of the magnet. The amount of flux flowing down the left-hand pole piece is reduced by this amount.

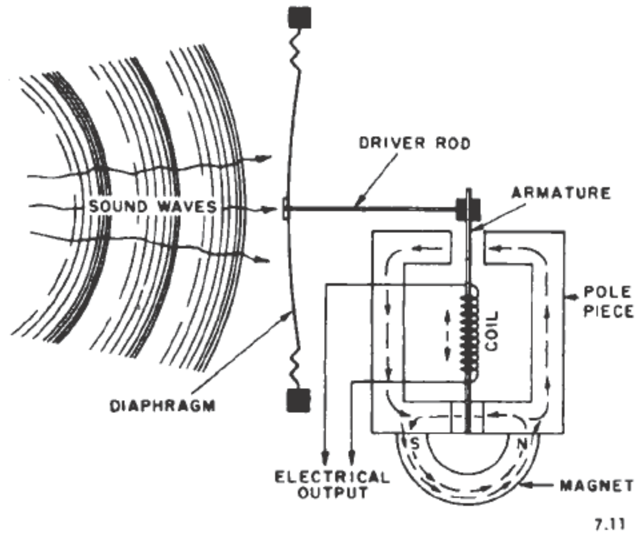


Figure 11-21.—Action of a magnetic microphone.

When a rarefaction wave strikes the diaphragm, the armature is deflected to the left. Some of the flux is now directed from the north pole of the magnet, up through the armature, through the reduced gap at the upper left, and back to the south pole. The amount of flux now moving up through the right-hand pole piece is reduced by this amount.

Thus, the vibrations of the diaphragm cause an alternating flux in the armature. The alternating flux cuts the stationary coil wound around the armature and induces an alternating voltage in the coil. This voltage has essentially the same waveform as that of the sound waves striking the diaphragm.

The magnetic microphone is the type most widely used in shipboard announcing and communicating systems because it is more resistant to vibration, shock, and rough handling than other types of microphones.

There are other types of microphones, such as the ribbon velocity microphone and the capacitor microphone, that are treated in the rating texts. All of the foregoing microphones may be used for radiotelephone broadcast, but the circumstances usually limit the choice to one or two types.

CIRCUITS OF AN A-M RADIO-TELEPHONE TRANSMITTER

The block diagram of a simple a-m radiotelephone transmitter is shown in figure 11-22. The oscillator, buffer stage, and power amplifier closely resemble the c-w transmitter.

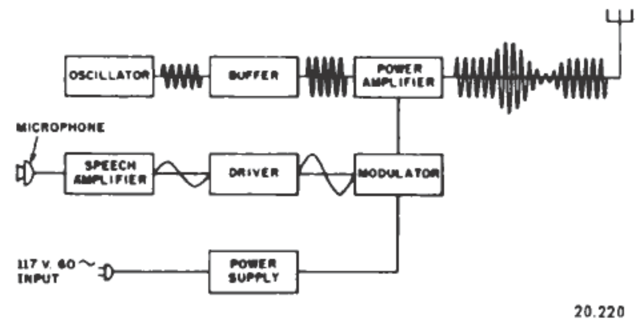


Figure 11-22.—Block diagram of an a-m radio telephone transmitter.

In figure 11-17 the c-w transmitter is keyed by opening and closing the cathode circuit of oscillator V1, which turns the r-f output of V2 and V3 off and on. If it is desired to vary the output of the transmitter instead of merely turning it off and on, the voltage on one of the elements of the final r-f power-amplifier tube, V3, may be varied. For example, if the plate voltage of V3 were varied at the audio-frequency rate, the output of the amplifier, and hence of the transmitter, would be varied at the same rate. This method, known as PLATE MODULATION, is used in the following example and is the most popular type of amplitude modulation. Plate modulation is discussed in greater detail in chapter 10.

In order to vary the plate voltage of the final r-f amplifier it is necessary first to produce an audio voltage. An audio voltage may be produced with a microphone. The output of a microphone is, however, very low (usually less than 1 volt), while the plate voltage of the r-f amplifier is quite high. The insertion of a small audio voltage in series with a high plate voltage would result in only a small variation of the plate voltage.

It is necessary, therefore, to amplify the electrical output of the microphone before it is applied in series with the plate of the final r-f power amplifier. This amplification is accomplished in three units of the block diagram of figure 11-22. These units are the speech amplifier, the driver, and the modulator. The output of the microphone is fed to the grid of V1, a class-A voltage amplifier pentode (fig. 11-23), which is the input tube of a two-tube speech amplifier. Voltage amplifier V2A and phase inverter V2B provide push pull output to the driver unit (V3), which further amplifies the audio signal to drive the audio modulator tubes (V4 and V5). The output voltage of the modulator tubes is fed in series with the plate supply voltage of the final r-f power amplifier of the transmitter. The modulator can be any type of audio power amplifier capable of providing sufficient undistorted power. Thus, it may be a class-A, class-AB, or class-B amplifier. In this example it is a class-AB push-pull stage.

The power supply unit includes the necessary transformers, rectifiers, and filters, to supply the filaments, plates, screens, and grid-bias voltages from a single-phase 117-volt 60-cycle source.

The output of the modulator unit may be applied in series with the plate of the final r-f power amplifier (shown in fig. 11-17 below M5).

In the absence of a modulating signal, a continuous r-f wave is radiated by the antenna. Assume that an audio voltage of sine waveform is applied across R1 to the grid of V1. The amplified signal appears across plate resistor R4 and is coupled to the grid of V2A through C4. The amplified signal is applied to the grid of driver V3A through C6, and a part of it is applied to the phase-inverter stage, V2B, by means of the tap on R10. The amplified signal from V2B is simultaneously applied via C8 to the grid of V3B in opposite phase to that applied to the grid of V3A. The driver stage (V3A and V3B) provides excitation for the grids of modulator tubes V4 and V5 through impedance-matching transformer T1.

The push-pull amplifier (V4 and V5) develops a relatively large audio output voltage across the secondary of modulation transformer T2 to amplitude-modulate the carrier output 100 percent. The following three relationships exist in the case of 100-percent modulation.

1. The modulation is 100 percent when the peak value of the audio voltage across the secondary of transformer T2 is approximately equal to the plate supply voltage of V3 (fig. 11-17) that appears across resistor R9. The degree of modulation depends on the volume of sound striking the microphone, which in turn determines the magnitude of the audio signal voltage developed across R1- the grid input to V1 (fig. 11-23).

2. The modulation is 100 percent when the energy supplied to the final r-f amplifier by

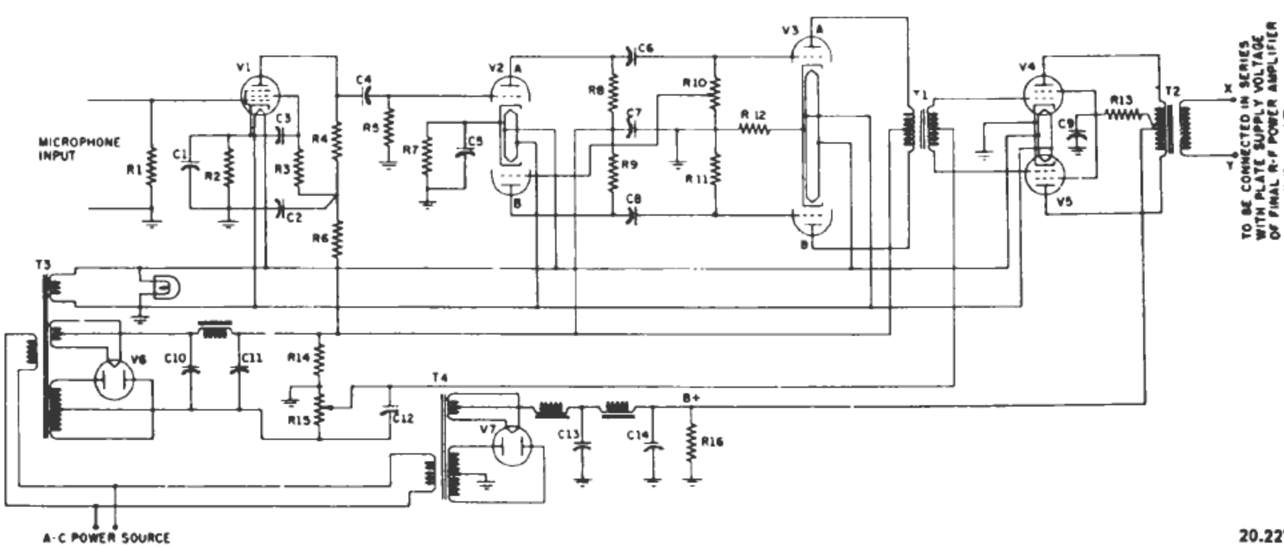


Figure 11-23.—Schematic diagram of a speech amplifier, driver, and modulator.

modulation transformer T2 is equal to one-half the r-f energy delivered to the final tank by the high-voltage power supply. (T5, V6, and V7 in fig. 11-17). The mixing of the a-f voltage from T2 (fig. 11-23) with the r-f voltage developed across final tank coil L7 (fig. 11-17) produces side-band frequencies (sum and difference frequencies) that are coupled into the antenna circuit by the mutual inductance existing between L7 and the coupling coil L8 of the antenna.

3. The modulation is 100 percent when the r-f energy delivered to the antenna (as a result of the injection of the a-f voltage from T2) is increased 50 percent above the amount delivered to the antenna when no audio signal is present. This condition represents an increase of 22.5 percent in the antenna current. Thus, the antenna r-f ammeter (not shown) may be used as a modulation indicator.

In this example there are four frequencies present in the final tank (C15, C16, and L7 of fig. 11-17). These are (1) the crystal frequency, (2) the audio frequency, (3) the sum of these two frequencies, and (4) the difference between these two frequencies. All except the audio frequency are coupled into the antenna circuit. The audio frequency is so far removed from the carrier and its associated side bands that the mutual inductive coupling between L7 and L8 for this frequency is effectively zero.

The details of adjustment and operation of both a-m and f-m transmitters are included in instruction books on these equipments.

FREQUENCY-MODULATED RADIOTELEPHONE TRANSMITTER

Intelligence may be also be conveyed by varying the frequency of a continuous radio wave of constant amplitude. The carrier frequency can be varied a small amount on either side of its average, or assigned, value by means of the a-f modulating signal. The amount the carrier is varied depends on the magnitude of the modulating signal, and the frequency with which the carrier is varied depends on the frequency of the modulating signal. The amplitude of the r-f carrier remains constant with or without modulation.

A radio receiver that is sensitive only to variations in the frequency of the incoming carrier and that discriminates to a large extent against variations in amplitude is used to receive the f-m signals.

CIRCUIT OF AN F-M RADIO-TELEPHONE TRANSMITTER

The block diagram of a narrow-band f-m transmitter is shown in figure 11-24. All power from the unit is obtained from a 115-volt 60-cycle source. The total drain is low enough so that the power may be taken from any convenient branch circuit outlet.

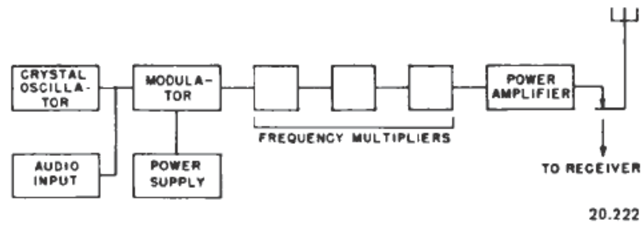


Figure 11-24.—Block diagram of a narrow-band f-m transmitter.

Oscillations are produced in the crystal-oscillator stage, the output of which is fed to a phase-shift network that supplied the grid voltage of the modulator tubes. The phase of the output voltage of the modulator varies in accordance with the input signal from the microphone. The phase shift is equivalent to to a relatively low deviation of the output signal frequency of the modulator stage.

The frequency of the output of the modulator stage is quadrupled in the first multiplier stage, again quadrupled in the second multiplier stage, and doubled in the last multiplier the output of which drives the power-amplifier stage, which consists of two beam-power tubes in parallel. To obtain the final operating frequency, the crystal frequency is multiplied by 32.

A schematic drawing of the narrow-band f-m transmitter, shown in the block diagram in figure 11-24 is shown in figure 11-25. For convenience in making adjustments, a meter, M1, and meter switch, S1, are provided on the chassis to indicate the grid-circuit of each stage.

With the exception of the first, the position numbers on the meter switch correspond with the numbers on the tops of the r-f transformers. For example, r-f transformer T2 is tuned for

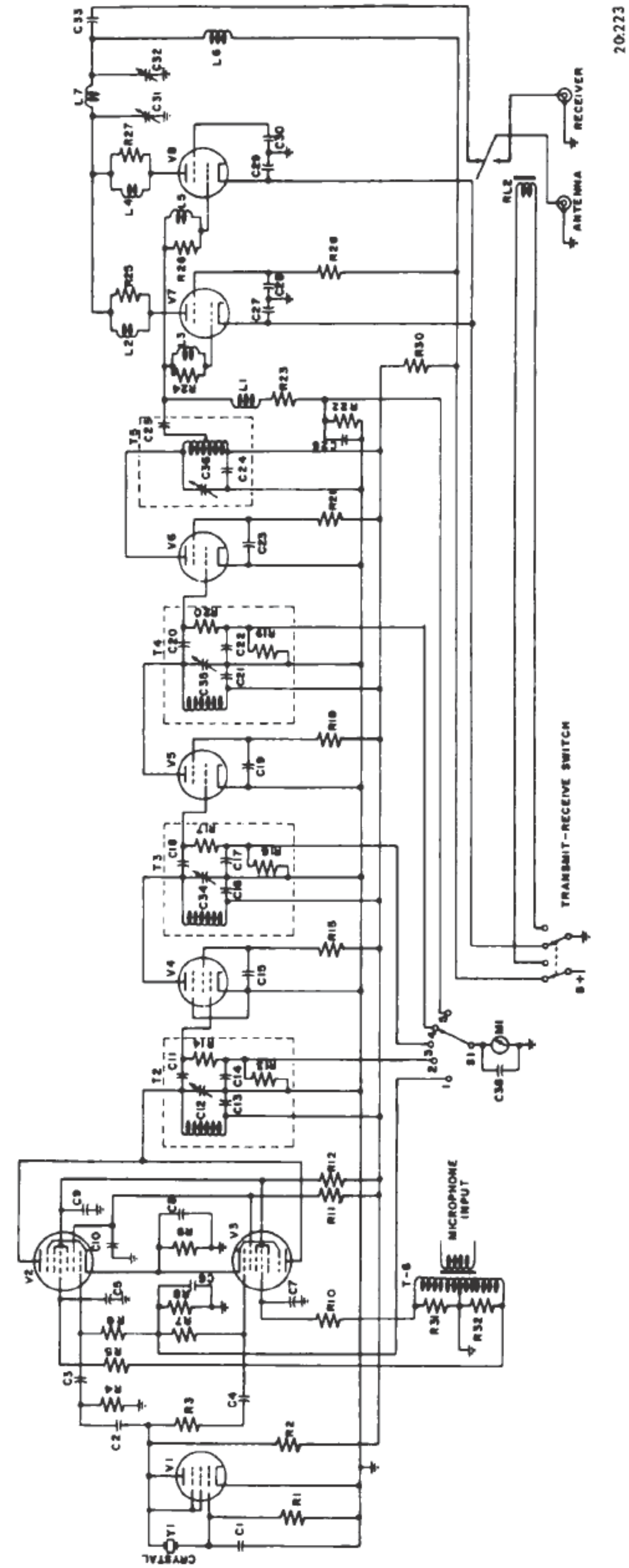


Figure 11-25.—Schematic diagram of a narrow-band f-m transmitter.

maximum current through the meter when the meter switch is in position 2. The transmitter employs the phase-shift method of obtaining frequency deviations. This method is discussed in chapter 10.

This transmitter exhibits characteristics that differ from those of the usual a-m type of transmitter. Intelligence is conveyed in the f-m transmitter by varying the frequency of the constant-amplitude carrier wave about an average assigned value. This process is in marked contrast with the amplitude-modulated transmitter previously described in which intelligence is conveyed by varying the amplitude of the constant-frequency carrier wave.

The phase-shift method of obtaining frequency deviations in this f-m transmitter permits direct crystal control of the average carrier frequency. Frequency multiplication after modulation is necessary in order to generate the required frequency deviation of ±15kc on either side of the carrier.

OPERATION

Crystal oscillator V1 is a pentode connected as a triode and operated as a conventional triode crystal oscillator. The crystal is of the low-drift AT-cut type and operates at the 32nd subharmonic of the output frequency. The output-frequency range is from 30 to 40 megacycles, and thus the crystals that are used range in frequency from 937.5 kc to 1,250 kc. The crystal oscillator utilizes a resistance-coupled circuit so that no oscillator tuning is necessary when changing crystals. The crystal is connected between the grid and plate of V1, and R2 acts as the plate-circuit load.

The control grids of the two balanced modulators, V2 and V3 are driven from the plate of the r-f oscillator through a phase-shifting network that displaces their associated r-f driving voltages by 90°.

The plate currents of V2 and V3 are about 90° out of phase and equal in magnitude because the driving voltages from the oscillator are equal when there is no modulating signal on the number 4 grids of the tubes. The two plate currents add vectorially to produce a resultant current in T2 that is approximately 45° out of phase with each component. The output voltage of T2 varies in phase and magnitude with this resultant current.

When a modulating signal is applied to grid number 4 of V2 and V3, the plate currents are varied about the average values they would have if no modulation were present. For example, as grid number 4 of V2 swings in a positive direction the plate current of V2 increases. Simultaneously the voltage on grid number 4 of V3 swings in a negative direction, and the plate current of V3 decreases.

Because these two currents are 90° out of phase, the resultant current (their vector sum) changes its phase with respect to the components as the components change in magnitude. Thus the current in T2, and the output voltage of T2, change in phase with the modulating signal. This change in phase is equivalent to a limited change in frequency occurring during the time that the output voltage phase shift is occurring.

The modulator grids of V2 and V3 are connected to the secondary of push-pull audio transformer T6. This transformer is driven directly from the microphone.

The modulator grids are fed through the frequency correcting networks R10C7 and R5C5. These R-C combinations attenuate the a-f range (above 2,000 cycles) so that excessive frequency deviation is not obtained. Resistors R31 and R32 are terminating resistors for the secondary of microphone transformer T6. Cathode bias is obtained across R9 and C8. R11 and R12 are screen voltage dropping resistors and C10 and C9 their respective screen bypass capacitors.

The phase shift depends on the ratio of the signal strength of the carrier to the modulating signal strength. A ratio of 2 to 1 is equivalent to about 0.5 radian, or 30°, phase shift. The frequency shift is equal to the product of the modulating frequency and the phase-shift angle in radians.

Frequency Multipliers

The frequency deviation that may be produced by the balanced modulator stage, V2 and V3, is small—usually not more than half the modulating frequency, since the phase-shift angle is of the order of 0.5 radian. To get sufficient deviation (±15 kc) the frequency of the modulated wave is multiplied by 32. As mentioned previously, this multiplication is accomplished by two quadruplers, V4 and V5 and a doubler, V6. All of these tubes act as class-C amplifiers with the plate tanks tuned to either the second or fourth harmonic of their respe tive grid signals. The grid drive in each case is such that plate current is well above saturation so that slight changes in tuning or reduction in tube emission have little effect on succeeding stages. All stages up to this point use receiving-type tubes working at relatively low plate and filament currents.

Power Amplifier

The power amplifier utilizes two beam-transmitting tubes, V7 and V8, in parallel as a class-C amplifier. Grid-leak bias is used; and, as in the previous stages, grid current is metered for alignment and testing The plate tank and antenna circuit is of the pi-type for harmonic suppression and ease of adjustment. This circuit consists of plate tuning capacitor C31, tank coil L7, and antenna loading capacitor C32. The output is fed through blocking capacitor C33 to antenna relay RL2.

CHAPTER 12

TRANSMISSION LINES

The transmission line (or antenna feedline, as it is assumed to be in this chapter) conducts or guides electrical energy from the input, or transmitter, end of the line to output, or antenna, end of the line. If this function is to be performed with a minimum loss, such elements as impedance matching and line losses must be considered.

Transmission lines may be classified as resonant or nonresonant lines, each of which may have advantages over the other under a given set of circumstances. There are various types of transmission lines such as the parallel two-wire line, the twisted pair, the coaxial line, and waveguides. The use of a particular type is dependent on the frequency, the voltage, the amount of power, the efficiency required, or the kind of installation to be used.

Resonant lines may have important uses other than the transmission of power. Among other uses, they may be employed as impedance-matching devices, phase shifters, and inverters, wave filters and chokes, and oscillator frequency controls.

Of primary importance in the study and application of transmission lines is the characteristic impedance of the line.

CHARACTERISTIC IMPEDANCE OF A TRANSMISSION LINE

In conventional circuits that contain inductors and capacitors, the inductance and capacitance are present in definite "lumps." In an r-f transmission line, however, these quantities are distributed throughout the entire line and cannot be separated from each other.

The characteristic impedance (or surge impedance) of a transmission line having infinite length is the impedance in ohms at the operating frequency, presented by the line to the source feeding the line. This impedance across the input of a theoretically infinite line has a very valuable use. If a load equal to this impedance is connected to the output end of the line, regardless of the length of the line, the impedance presented to the source by the input terminals of the line is still equal to the characteristic impedance of the transmission line. Only one value of impedance for any particular type and size of line acts in this way.

A section of two-wire transmission line of unit length has a certain amount of resistance (no material is a perfect conductor) that varies directly with the length and inversely with the cross sectional area of the conductor.

The same section of line has the property of distributed inductance. This property exists because of magnetic flux linkages which are established within the section when current flows. For example, an open line composed of two No. 12 conductors spaced 6 inches apart has an inductance of approximately 0.6 microhenry per foot.

This section of line also has the property of capacitance because the two wires, separated by a dielectric, act as the two plates of a capacitor. The capacitance of the two-wire line in the previous example is approximately 1.7 micromicrofarads per foot.

Finally, the transmission line of unit length has leakage resistance in the path through the insulating material that separates the two conductors (no substance is a perfect insulator). For convenience in working out problems dealing with longer lines, this property usually is expressed as the reciprocal of the leakage resistance, which is conductance. The conductance is of the order of a few micromicromhos per foot.

Resistance R1 and R2, inductances L1 and L2, capacitance C, and conductance G of a unit length of two-wire transmission line are shown in figure 12-1,A. In many cases the effect of conductance G is very small compared with that produced by the inductance and capacitance and may therefore be neglected. Conductance G and resistances R1 and R2 in figure 12-1, B are omitted and L1 and L2 are treated as if they were in one side of the transmission line.

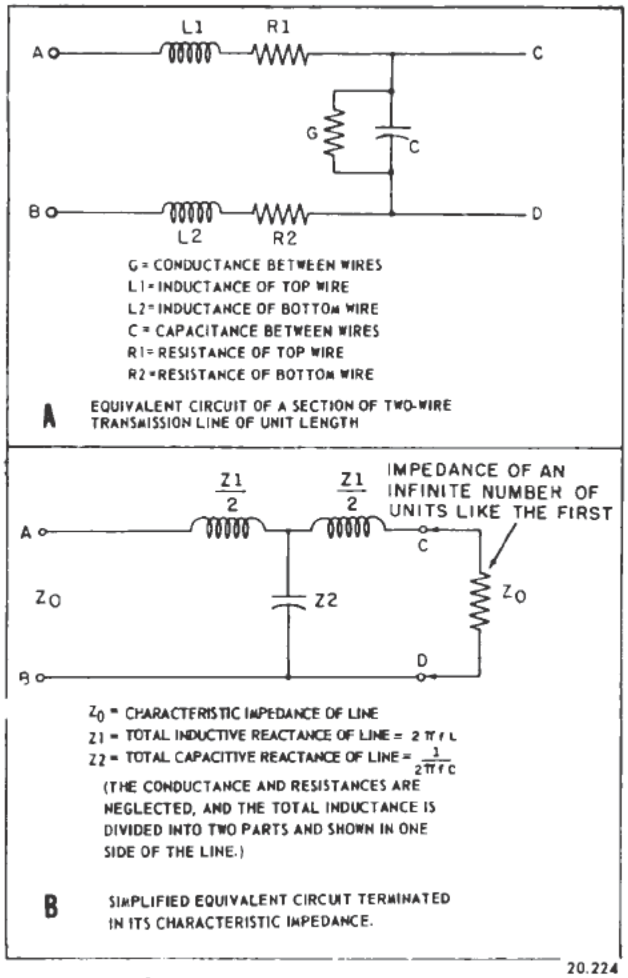


Figure 12-1.—Equivalent circuits of a two-wire transmission line of unit length.

In any circuit such as the one shown in figure 12-1, some current will flow if a voltage is applied across terminals A and B. The ratio of the voltage to the current is the impedance, Z—that is, $Z = \frac{E}{I}$. The impedance presented to the input terminals of a transmission line is much more than the simple resistance of the wires in series with the impedance of the load. The effects of series inductance and shunt capacitance distributed along the line are appreciable at the relatively high operating frequency and constitute the principal components of the equivalent network.

The formula for the characteristic impedance as a function of the L and C of a unit length of transmission line may be determined from the simplified equivalent T-network circuit shown in figure 12-1.B. The conductor resistance and the insulation leakage conductance are low and considered negligible, hence are not shown in the figure. The distributed inductance of the line is divided equally in two parts in horizontal arms of the T. The distributed capacitance of the line is lumped in one value in the central leg of the T. The line is terminated in a resistive load having a value equal to that of the characteristic impedance of the line as seen looking into the T-network terminals, A and B.

The impedance, Z_O, looking into the T-network terminals, AB, is

$$Z_O = \frac{Z_1}{2} + \frac{Z_2\left(\frac{Z_1}{2} + Z_O\right)}{Z_2 + \frac{Z_1}{2} + Z_O}$$

$$= \frac{Z_1}{2} + \frac{\frac{Z_1 Z_2}{2} + Z_O Z_2}{Z_2 + \frac{Z_1}{2} + Z_O}$$

$$= \frac{Z_1 Z_2 + \frac{Z_1^2}{2} + Z_1 Z_O + \frac{2Z_1 Z_2}{2} + 2Z_O Z_2}{2\left(Z_2 + \frac{Z_1}{2} + Z_O\right)}$$

If both sides of this equation are multiplied by the denominator of the right-hand side, the result is

$$2Z_2 Z_O + \frac{2Z_1 Z_O}{2} + 2Z_O^2 = Z_1 Z_2 + \frac{Z_1^2}{2} + \frac{2Z_1 Z_2}{2} + 2Z_O Z_2$$

and this equation simplified becomes

$$2Z_O^2 = 2Z_1 Z_2 + \frac{Z_1^2}{2}$$

or

$$Z_O^2 = Z_1 Z_2 + \left(\frac{Z}{2}\right)^2$$

If the transmission line is to be accurately represented by an equivalent network, the T-network section of figure 12-1, B must be replaced with an infinite number of similar sections. Thus, the distributed inductance of the line will be divided into n sections, instead of 2 as indicated in the last term of the preceding formula. As the number of sections approaches infinity, the last term, $\left(\frac{Z_1}{n}\right)$, will approach zero as a limit—that is, as $n \to \infty$, $\left(\frac{Z_1}{n}\right)^2 \to 0$. Therefore,

$$Z_O = \sqrt{Z_1 Z_2}$$

$$= \sqrt{\frac{2\pi f L}{2\pi f C}}$$

and

$$Z_O = \sqrt{\frac{L}{C}}$$

The last formula indicates that the characteristic impedance depends on the distributed inductance and capacitance of the line. An increase in the separation of the wires increases the inductance and decreases the capacitance. This effect takes place because the effective inductance is proportional to the flux which may be established between the wires. If the two wires carrying current in opposite directions are placed farther apart, more magnetic flux is included between them (they cannot cancel their magnetic effects as completely as they could if the wires were closer together) and the distributed inductance is increased. The capacitance is of course lowered if the plates of the capacitor (in this case the plates are the two wires) are more widely separated.

Thus, the effect of increasing the spacing of the wires is to increase the characteristic impedance, because the $\frac{L}{C}$ ratio is increased. Similarly, a reduction in the diameter of the wires also increases the characteristic impedance. The reduction in the size of the wire affects the capacitance more than the inductance, for the effect is equivalent to decreasing the size of the plates in a capacitor in order to decrease the capacitance. Any change in the dielectric material between the two wires also changes the characteristic impedance. Thus, if a change in dielectric material increases the capacitance between the wires, the characteristic impedance is reduced.

The characteristic impedance of a two-wire line with air as the dielectric may be obtained from the formula

$$Z_O = 276 \log_{10} \frac{b}{a}$$

where b is the spacing between the centers of the conductors and a is the radius of one of the conductors.

The characteristic impedance of a concentric, or coaxial, line also varies with L and C. However, because the difference in construction of the two lines causes L and C to vary in a slightly different manner, the following formula must be used in determining the characteristic impedance of the concentric line:

$$Z_O = 138 \log_{10} \frac{b}{a}$$

where b is the inner diameter of the outer conductor and a is the diameter (or the outer diameter if a hollow tube is used) of the inner conductor.

WAVE MOTION ON AN INFINITE LINE

Figure 12-2 shows sine waves of voltage and current that travel at high speed along a two-wire transmission line of infinte length. Because this is a line of infinite length, no reflections occur; therefore, the voltage and the current are in phase with each other everywhere along the line. Because of line losses, the curves diminish in amplitude as the waves progress along the line. If a voltage is impressed on a line such as this, an electric field will be established between the wires. Likewise, current will flow in the wires, and a magnetic field will be established around each wire. These two fields constitute an electromagnetic wave that travels down the wire at a velocity somewhat less than that of light.

Figure 12-2 illustrates what would happen if the voltage and current could be stopped for an instant in time. An instant later all waves would have moved to the right a slight

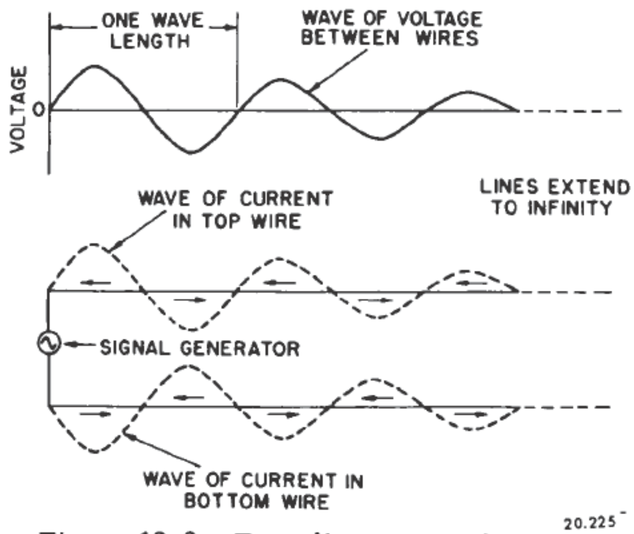


Figure 12-2.—Traveling waves of current and voltage on a line of infinite length.

amount. In this figure the waves are stopped at the instant when the alternating source voltage has just reached zero.

Traveling waves exist on the line because it takes a certain amount of time to propagate them down the line. The production of these waves may be better understood from the following considerations. First of all, it must be understood that the waveforms shown in figure 12-2 are "stopped" for an instant in time and the observer examines the entire train of waves along the line. Assume that at a given instant the voltage at the generator terminals is zero. An instant later one terminal becomes more positive and the other becomes more negative. The electric field between the wires increases in strength; the current and also the magnetic field increase proportionately. The perpendicular distance from any point along the wire to the current curve (fig. 12-2) indicates the relative magnitude and direction of the current at that point. The perpendicular distance from any point along the voltage axis to the voltage curve represents the relative magnitude and polarity of the voltage across the line at the corresponding location.

At 90° in the electrical cycle the electric and magnetic field are at their maximum, and from 90° to 180° they decrease in amplitude to zero. At 180° the voltage at the generator terminals reverses polarity, and the electric field between the wires reverses direction. Similarly, the current reverses direction, and this causes the magnetic field to reverse direction. The fields increase in strength from 180° to 270°, and then decrease in strength to zero at 360°. These electric and magnetic impulses do not return to the generator once they start down a line of infinite length.

The characteristics of a theoretically infinite line may be summarized as follows:

1. The voltage and current are in phase throughout the line.
2. The ratio of the voltage to the current is constant over the entire line and is known as the CHARACTERISTIC IMPEDANCE.
3. The input impedance is equal to the characteristic impedance.
4. Since the voltage and current are in phase, the line operates at maximum efficiency.
5. Any length of line can be made to appear like an infinite line if it is terminated in its characteristic impedance.

LINE REFLECTIONS

If a transmission line is infinitely long, or if it is terminated in its characteristic impedance, reflections do not occur. However, if there is an abrupt discontinuity in the line (such as an open circuit or a short circuit) complete reflection will occur. A discontinuity of less importance (such as a poorly made splice) will cause some reflection, the amount of reflection depending on the value of the resistance at the splice. In this section, the discussion of reflection will be limited to the two extreme conditions—that is, to open-and closed-end lines.

OPEN-END LINES

One type of r-f transmission line is the open-end line, in which the impedance at the output end can be considered as practically infinite because no load is attached. When energy is applied to the generator end, the first surge consists of a wave of current and a wave of voltage that sweep down the line in phase with each other—that is, their positive maxima are together. The initial current and voltage waves must travel down the line in phase because the characteristics of the line are the same as those of a line that is truly infinite WHILE THE INITIAL WAVE IS TRAVELING TOWARD THE OUTPUT END. The in-phase condition of these waves can be changed only when they encounter a difference in the impedance between the two wires of the line, and reflections occur. Thus, when the wave of current reaches the open-circuited output end (terminal point) of the line the current must collapse to zero. When the current wave collapses, the magnetic field that was set up by it also collapses. The collapsing magnetic field cuts the conductors near the output end and induces additional voltage across the line. This voltage acts, in a way, like a reverse generator and sets up new current and voltage waves that travel back along the line toward the input end.

An open-end transmission line one wavelength long and having no attenuation is shown in figure 12-3,A. It should be pointed out that all of the waveforms shown are instantaneous values that exist along a transmission line. These curves are unlike conventional sine curves in which distance along the X axis represents lapse in time from the origin, proceeding to the right. Instead, these curves represent instantaneous values of current or voltage as they exist all along the line for the same instant of time.

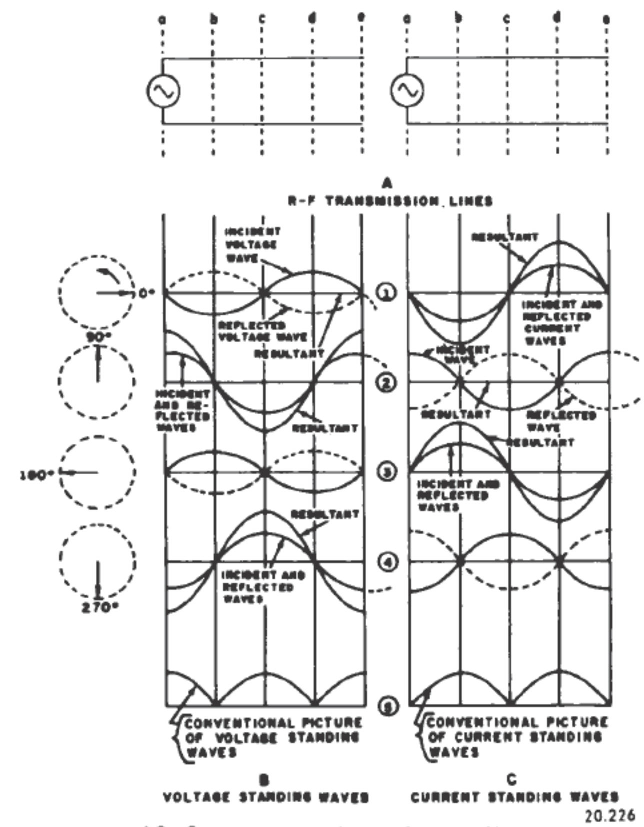


Figure 12-3.—Formation of standing waves on an open-end transmission line one wavelength long.

Although only four positions of the generator voltage vector are shown, the picture could be made more complete by showing the waveforms at 45° intervals. At intervals of 45° between the generator positions shown—for example, at 45° 135°, 225°, and 315 —the instantaneous values of the resultants will be 0.707 of their maximum values. The line is represented twice (fig. 12-3,A) in order to orient it properly with respect to the waves of voltage and current. These waves are shown separately (fig 12-3,B and C) in order to simplify the analysis, although in reality they both appear simultaneously along the same transmission line.

Voltage Standing Waves

It is assumed that in part (1) of figure 12-3, B, the generator voltage vector has gone through at least two complete revolutions so that the voltage wave as had time to travel down the line and return to the generator end. The waveforms are stopped in time in this figure at the instant that the generator voltage vector is at the zero position.

It may be observed that the initial voltage wave is reflected at the output end of the line in phase with the voltage wave that would have continued along the line in the original direction of travel if the line had been longer. For example, the dotted waveform extending slightly beyond the end of the line in part (1) indicates that the incident wave would have started going negative. Therefore, the reflected wave will start back in a negative direction. At the instant in time being considered here the incident and reflected wave add vectorially to give a zero resultant wave.

Ninety degrees later (part (2) of figure 12-3, B) the incident and reflected waves are in phase and add vectorially to give the resultant voltage wave, as shown. At 180° (part (3)), the resultant voltage is again zero; and at 270° (part (4)), the incident and reflected waves are once again in phase, and the resultant voltage wave is shown 180° out of phase with the resultant wave in part (2).

Next, consider the voltage variations across the line that occur with respect to time at certain locations along the line. At point a, the voltage is zero (part (1)), then maximum in one direction (part (2)), then zero again (part (3)), and finally maximum in the other direction (part ④). This is true also at points c and e; at points b and d the voltage is always zero. A suitable voltage-indicating device (to be discussed later in the chapter) located at a, c, or e will indicate voltage loops (points of maximum voltage); and at b and d the device will indicate voltage nodes (points of minimum voltage).

Standing waves of voltage are shown in part ⑤ of figure 12-3,B. This curve represents effective values of voltage at the various points along the line. These values are actually the effective values of the sinusoidal voltage variations occurring across the line at points where the measurements are being made. Thus, at point c the voltage will be zero at one instant of time (part ①), then it will build up to a maximum with one polarity (part ②), then it will become zero (part ③), and finally it will build up to a maximum with the opposite polarity (part ④).

Current Standing Waves

Standing waves of current are shown in figure 12-3,C. They are occurring simultaneously with the voltage waves on the transmission line, but they are shown here separately in order to simplify the figure. The initial current wave is reflected at the output end of the line 180° out of phase with the current wave that would have continued along the line in the original direction of travel if the line had been extended. (See dotted-line extensions.) In other words, the reflected current reverses direction at the open end of the line and is shown 180° out of phase with the incident wave except when the incident and reflected waves have zero values at the end of the line (0°, 180°, and so forth). This reversal is opposite to the condition for voltage, because the reflected voltage wave has the same polarity that the incident wave would have if it continued down the extended line in the direction of travel. (See dotted-line extensions.)

Because the incident and reflected current waves are 180° out of phase at the open end of the line, they cancel at this point and the resultant current at the open end is always zero. The rotating vectors at the left of the figures indicate the generation of sine waves of both voltage and current. At b the incident and reflected current waveforms combine to produce a current loop (maximum current); the same is also true at point d. At a, c, and e the incident and reflected waveforms combine to produce current nodes (points of zero current).

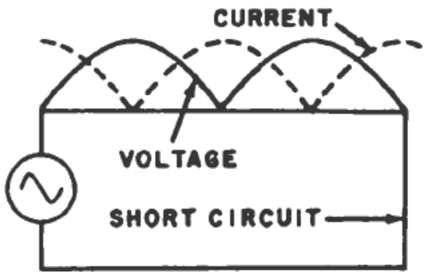
In part ② of figure 12-3,C, the current vector has rotated 90 degrees. Combining the incident and reflected waves at this instant gives a resultant current waveform that has an amplitude of zero throughout the entire length of the line. In part ③ of the figure the current vector has completed 180° of its cycle. Again combining the incident and reflected waves at this instant gives a resultant current that has a maximum amplitude at points b and d although the direction of the current is opposite to that in part 1 at these points. In part ④ the resultant current is again zero all along the line.

Part ⑤ of figure 12-3,C, is a plot of the effective values of current at the various points along the line. These current values are actually the effective values of the resultant current variations through the cycle at the respective points where the measurements are made. For example, at point b the current is maximum in one direction at one instant of time (part ① then it becomes zero (part ②), then it builds up to a maximum in the opposite direction (part ③), and finally returns to zero (part ④). The effective value of this current variation is plotted at b in part ⑤ .

CLOSED-END LINES

In a closed-end line the voltage and current waveforms exchange places with respect to their locations in an open-end line. Thus, in figure 12-3,A, if the line is closed at the end, part B becomes current standing waves and part C becomes voltage standing waves. At the short-circuited end of the line (point e), the current varies from zero to a maximum in one direction and back to zero and a maximum in the other direction. The effective value of the current will therefore be a maximum at point e (a current loop). The voltage ($E = IR$) across the short circuit approaches zero because the resistance of the short circuit is negligible.

At $\frac{\lambda}{4}$ and $\frac{3\lambda}{4}$ wavelengths from the shorted end, the effective current is a minimum at all times and the effective voltage is a maximum. At $\frac{\lambda}{2}$ and λ wavelengths from the shorted end, the effective value of the current is a maximum and the effective voltage is zero. The curves representing the variation of effective current and voltage along a shorted line are shown in figure 12-4.



20.227

Figure 12-4.—Conventional picture of current and voltage standing waves on a closed-end line.

NONRESONANT LINES

A nonresonant line can be defined as a line that has no standing waves of current and voltage. Such a line is either infinitely long or is terminated in its characteristic impedance. Because there are no reflections, all of the energy passed along the line is absorbed by the load (except for the small amount of energy dissipated by the line). The voltage and current waves are traveling waves that move in phase with each other from the source to the load.

On lines carrying radio frequencies, the characteristic impedance is almost always pure resistance. Therefore, it is customary to say that a nonresonant line is terminated in a resistive load equal to the characteristic impedance of the nonresonant line.

RESONANT LINES

A resonant transmission line is one that has standing waves of current and voltage. The line is of finite length and is not terminated in its characteristic impedance, and therefore reflections are present.

A resonant line, like a tuned circuit, is resonant at some particular frequency. The resonant line will present to its source of energy a high or a low resistive impedance at multiples of a quarter-wavelength. Whether the impedance is high or low at these points depends on whether the line is short- or open-circuited at the output end. At points that are not exact multiples of a quarter-wavelength, the line acts as a capacitor or an inductor.

A resonant transmission line thus may assume many of the characteristics of a resonant circuit that is composed of lumped inductance and capacitance. The more important circuit effects that resonant transmission lines have in common with resonant circuits having lumped inductance and capacitance are as follows:

Series Resonance—Resonant rise of voltage across the reactive circuit elements, and low impedance across the resonant circuit.

Parallel Resonance—Resonant rise of current in the reactive circuit elements, and high impedance across the resonant circuit.

RESONANCE IN OPEN-END LINES

The open-end resonant line may be better understood by an analysis of figure 12-5. The transmission lines considered in this figure and in the following figure have no losses. If losses are present (and they are in a practical line), the voltage at voltage nodes is not zero; neither is the current zero at current nodes. However, in these figures losses are neglected in order to simplify the analysis. Figure 12-5 illustrates the relation of voltage, current, and impedance for various lengths of open-end transmission lines. The impedance that the generator "sees" at various distances from the output end is shown directly above in the impedance curves. The curves above the letters (R, X_L, X_C) of various heights indicate the relative magnitudes of the impedances presented to the generator for the various lengths of lines indicated. The letters themselves indicate the type of impedance offered at the corresponding inputs. The circuit symbols above the various transmission lines indicate the equivalent electrical circuits for the transmission line at that particular length (measured from the output end). The curves of effective E and I whose ratio, $\frac{E}{I}$ is the impedance, Z, are shown above each line.

At all ODD quarter-wavelength points $\left(\frac{\lambda}{4}, \frac{3\lambda}{4}, \frac{5\lambda}{4}, \text{etc.}\right)$ measured from the open end of the line, the current is a maximum and the impedance is a minimum. In addition, there is a resonant rise of voltage from the odd quarter-wavelength points toward the open end. Thus at all odd quarter-wavelengths points the open-end transmission line acts like a series resonant circuit. The impedance is therefore very low and is

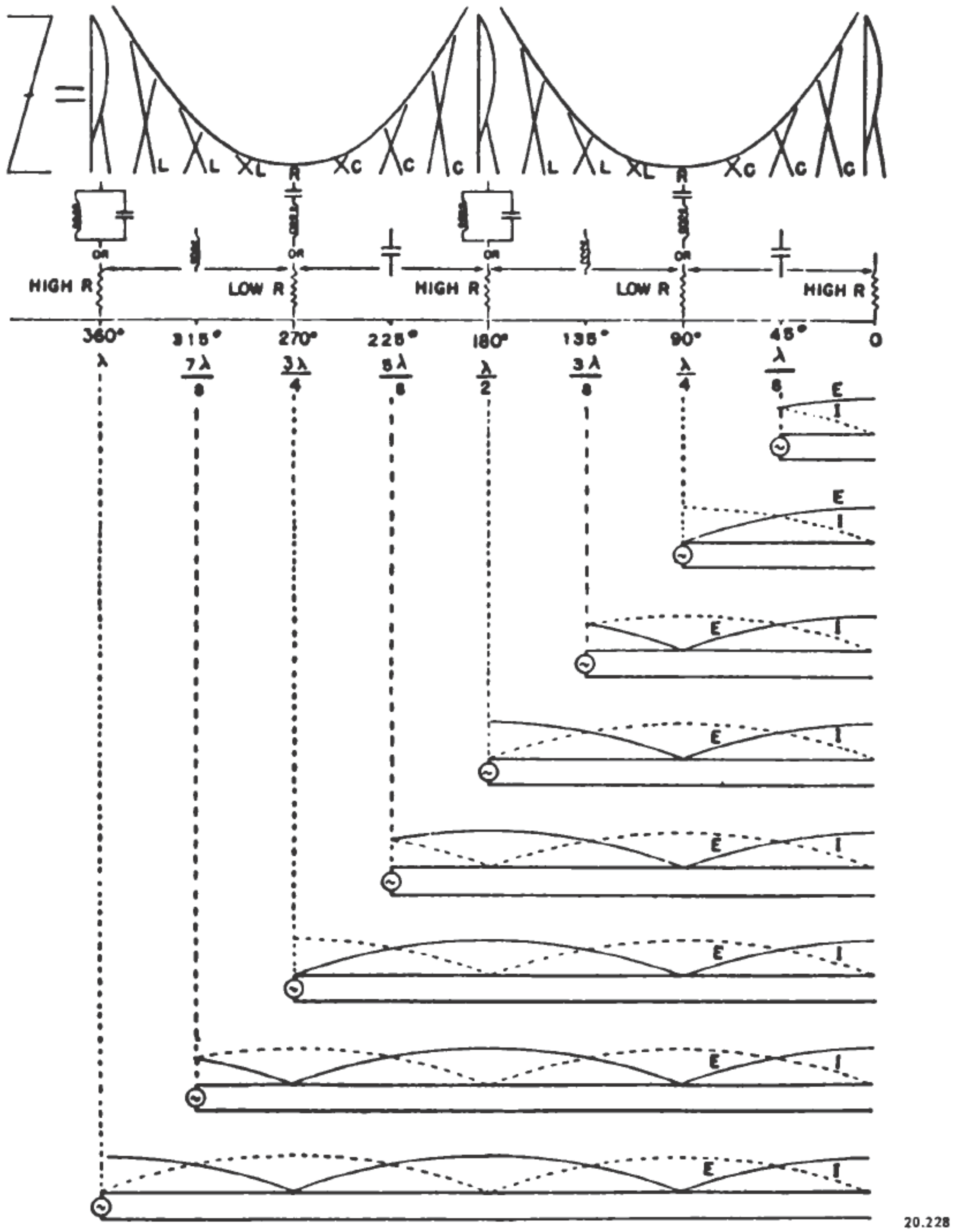


Figure 12-5.—Impedance characteristics of open-end resonant lines.

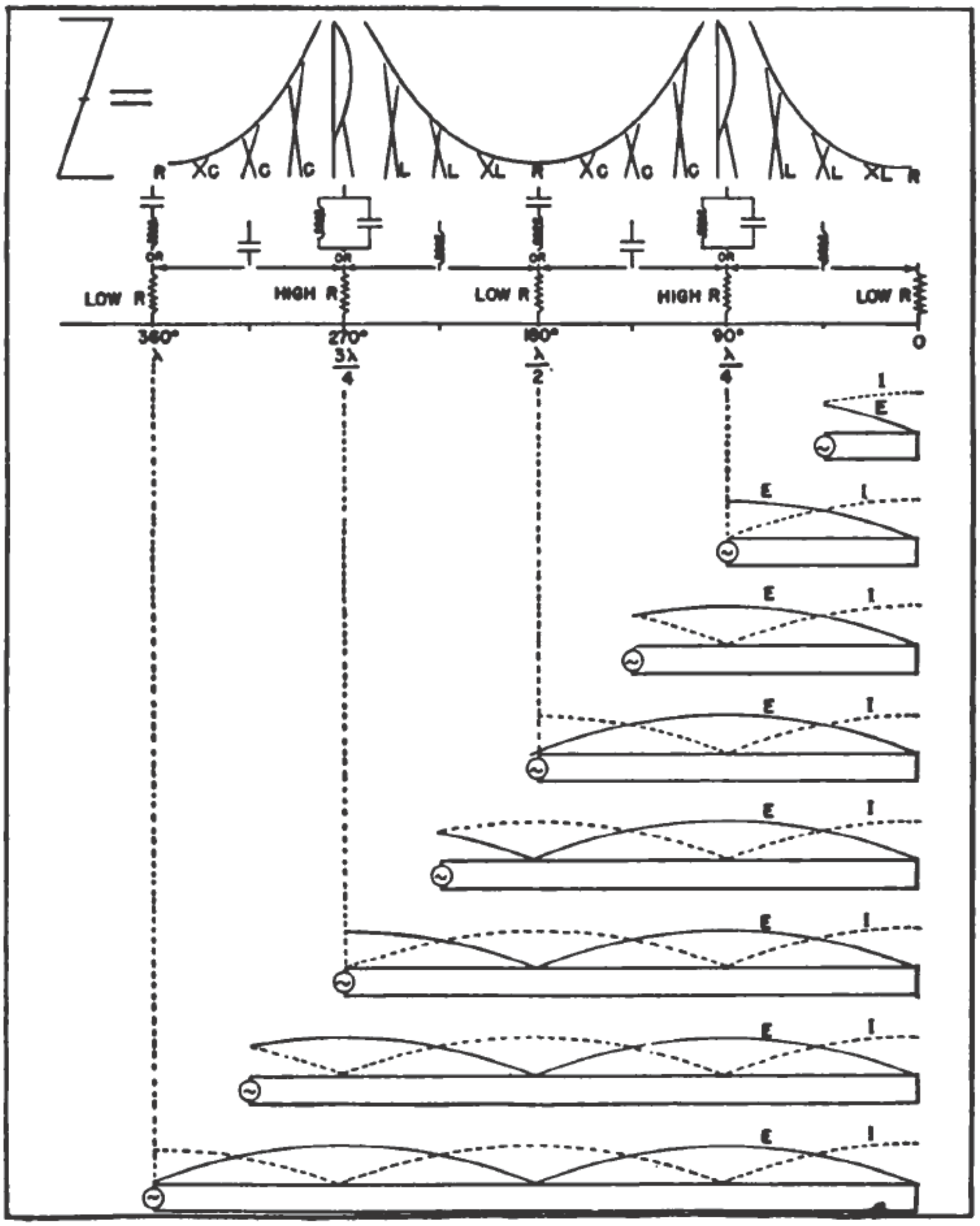
prevented from being zero only by the small circuit losses.

At all EVEN quarter-wavelength points $\frac{\lambda}{2}$, λ, $\frac{3\lambda}{2}$, etc. the voltage is a maximum, and therefore the impedance is a maximum. A comparison of this type of transmission line with an L-C resonant circuit shows that at even quarter-wavelengths (from the output end) the line acts like a parallel resonant circuit.

In addition to acting as series or parallel L-C resonant circuits, resonant open-end lines also may act as nearly pure capacitances or inductances when the lengths of the lines are not an exact multiple of the fundamental quarter-wavelength corresponding to the frequency of the applied voltage at the input terminals. Figure 12-5 shows that an open end line less than a quarter-wavelength long acts like a capacitance; between $\frac{\lambda}{4}$ and $\frac{\lambda}{2}$ wavelength, as an inductance; between $\frac{\lambda}{2}$ and $\frac{3\lambda}{4}$ wavelength, as a capacitance; between $\frac{3\lambda}{4}$ and λ wavelength, as an inductance; and so forth.

RESONANCE IN CLOSED-END LINES

The closed-end line may likewise be studied with the aid of figure 12-6. At ODD quarter-wavelengths from the closed end of the line the voltage is high, the current low, and the impedance high. Because conditions are similar to those in a parallel resonant circuit, the shorted transmission line of odd quarter-wavelengths acts like a parallel resonant circuit. The voltage across a circuit of this type cannot exceed the applied voltage.



20.229

Figure 12-6.—Impedance characteristics of closed-end resonant lines.

At EVEN quarter-wavelength points (measured from the shorted end) the voltage is a minimum, the current is a maximum, and the impedance is a minimum. Because this action is similar to series resonance in an L-C circuit, a shorted transmission line of even quarter-wavelengths acts like a series resonant circuit.

Resonant closed-end lines, like open-end lines, may also act as nearly pure capacitances or inductances when the length of the lines are not exact multiples of the fundamental quarter-wavelength corresponding to the frequency of the applied voltage at the input terminals.

LINE TERMINATED IN A REACTANCE

A line terminated in a resistance equal to its characteristic impedance normally has no reflections present. However, if a transmission line is terminated in a reactance equal to its characteristic impedance or to any other impedance, standing waves are not eliminated. Figure 12-7,A, shows the standing waves that exist on a line terminated in a capacitive reactance equal to its characteristic impedance. Note that the last current loop is less than a quarter-wavelength from the capacitive termination of the line. With capacitive termination the voltage and current distribution has essentially the same character as with the open-end line, except that the curves are shifted toward the output end of the line by an amount that increases as the capacitive reactance is reduced—that is, as the line approaches the closed-end condition of zero impedance.

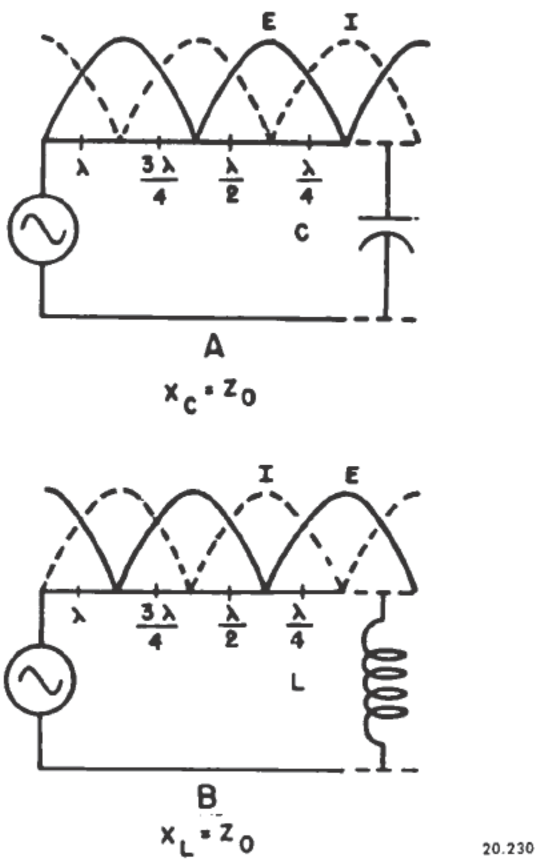


Figure 12-7.—Transmission lines terminated in reactance.

Figure 12-7,B, shows the standing waves that occur on a line terminated in an inductive reactance equal to the characteristic impedance. Note that the last voltage loop is less than a quarter-wavelength from the inductive termination of the line. With inductive termination the voltage and current distribution has essentially the same character as with a short-circuited output, except that the curves are shifted toward the inductive termination by an amount that increases as the terminating inductive reactance approaches infinity—that is, as the line approaches the open-end condition.

STANDING-WAVE RATIO

The ratio of the effective voltage at a loop to the effective voltage at a node, or the effective current at a loop to the effective current at a node is called the STANDING-WAVE RATIO (SWR) of a transmission line. It is also equal to the ratio of the characteristic impedance of the line to the impedance of the load, or vice versa. When the line is terminated in a perfect match, all of the energy sent down the line is absorbed by the load and none is reflected. Under these conditions no standing waves are present. The maximum and minimum values are the same, and therefore the standing-wave ratio is equal to 1.0.

Two mismatched lines are shown in figure 12-8. In each of these lines the characteristic impedance, Z_0, of the line is 3000 ohms. In figure 12-8,A, the load impedance is 60 ohms. The ratio of the effective current at a to the effective current at b is equal to $\frac{y}{x}$, or $\frac{5}{1}$, which is also equal to $\frac{300}{60}$.

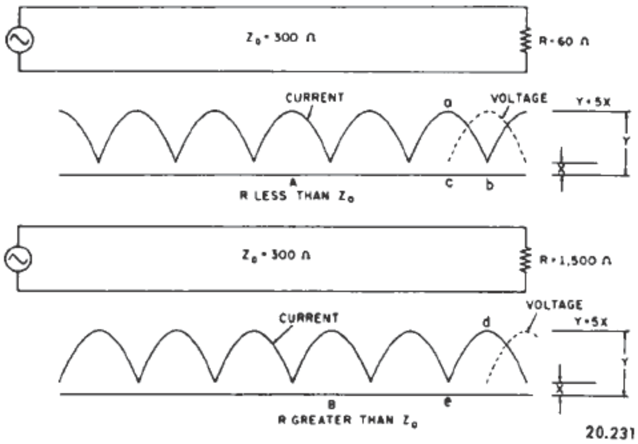


Figure 12-8.—Mismatched lines showing standing-wave ratio.

In figure 10-8,B, the line impedance is less than ($\frac{1}{5}$ of) the impedance of the load. The ratio of the effective current at d to the effective current at e is equal to $\frac{1,500}{300}$, or $\frac{5}{1}$. In the first example, the SWR is equal to the ratio of the Z_0 of the line to the Z of the load. In the second example it is equal to the ratio of the Z of the load to the Z_0 of the line. In both examples, the SWR is equal to the ratio of the effective current at a loop to the effective current at a node.

In general, the higher the SWR, the greater is the mismatch between the line and the load. A knowledge of the position of the current and voltage loops and nodes along the line will indicate whether the load resistance is less than or greater than the characteristic impedance. For example, in figure 12-8,A, there are a voltage node and a current loop at the load. This occurs because the load resistance is less (approaching a shorted condition) than the characteristic impedance of the line. Thus, it is a simple matter (by the use of one of the r-f measuring devices to be discussed later) to determine whether the load resistance is greater or smaller than Z_0. If the load resistance is greater than Z_0 (fig. 12-8,B), the output end of the line will appear more like an open circuit, and r-f measuring devices will indicate maximum effective voltage and minimum effective current at that point.

TYPES OF TRANSMISSION LINES

There are five general types of transmission lines—the parallel two-wire line, the twisted pair, the shielded pair, the concentric (coaxial) line, and waveguides. As mentioned in the introduction, the use of a particular type of line depends among other things on the frequency and the power to be transmitted and on the type of installation.

PARALLEL TWO-WIRE LINE

One of the most common types of transmission lines consists of two parallel conductors that are maintained at a fixed distance by means of insulating spacers or spreaders that are placed at suitable intervals. This type of line is shown in figure 12-9,A. The line is used frequently because of ease of construction, economy, and efficiency. In practical applications two-wire transmission lines (with individual insulators rather than spacers) are used for power lines, rural telephone lines, and telegraph lines. This type of transmission line is also used as the connecting link between an antenna and transmitter or an antenna and receiver.

In practice, such lines used in radio work are generally spaced from 2 to 6 inches apart on 14-mc and lower frequencies. The maximum spacing for 38-mc or higher frequencies is 4 inches. In any case, in order to effect the best

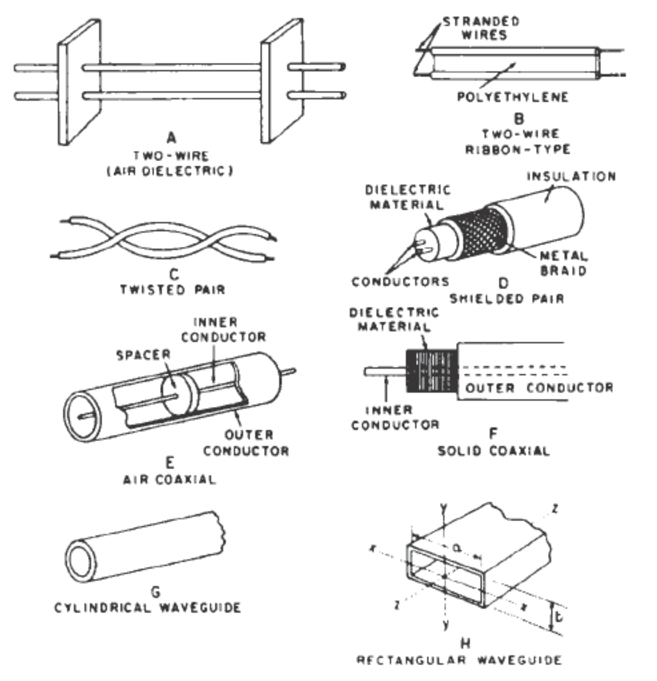


Figure 12-9.—Types of transmission lines.

cancellation or radiation, it is necessary that the wires be separated by only a small fraction of a wavelength. For best results, the separation should be less than 0.01 wavelength.

The principal disadvantage of the parallel-wire transmission line is that it has relatively high radiation loss and therefore cannot be used in the vicinity of metallic objects, especially when high frequencies are used, because of the greatly increased loss which results.

Uniform spacing of a two-wire transmission line may be assured if the wires are imbedded in a solid low-loss dielectric throughout the length of the line, as indicated in figure 12-9,B. This type of line is often called a two-wire ribbon type. The ribbon type is commonly made with two characteristic impedance values, 300 ohms and 75 ohms. The 300-ohm line is about one-half inch wide and is made of stranded wire. Because the wires are imbedded in only a thin ribbon of polyethylene, the dielectric is partly air and partly polyethylene. Moisture or dirt will change the characteristic impedance of the line. This effect becomes more serious if the line is not terminated in its characteristic impedance.

The wires of the 75-ohm line are closer together, and the field between the wires is confined largely to the dielectric. Weather and dirt affect this line less than they affect the 300-ohm line. The ribbon type of line is widely used to connect television receivers to antennas.

TWISTED PAIR

The twisted pair is shown in figure 12-9,C. As the name implies, it consists of two insulated wires twisted to form a flexible line without the use of spacers. It is used as an untuned line (on a tuned line the insulation might be broken down by arc-over at voltage loops) for low frequency transmission. It is not used for the higher frequencies because of the high losses occurring in the rubber insulation. When the line is wet, the losses increase greatly. The characteristic impedance of such lines is about 100 ohms, depending on the type of cord used.

SHIELDED PAIR

The shielded pair (shown in figure 12-9,D) consists of two parallel conductors separated from each other and surrounded by a solid dielectric. The conductors are contained within a copper-braid tubing that acts as a shield. This assembly is covered with a rubber or flexible composition coating to protect the line against moisture and friction. Outwardly, it looks much like an ordinary power cord for an electric motor.

The principal advantage of the shielded pair is that the two conductors are balanced to ground—that is, the capacitance between each conductor and ground is uniform along the entire length of the line and the wires are shielded against pickup of stray fields. This balance is effected by the grounded shield that surrounds the conductors at a uniform spacing throughout their length.

If radiation from an unshielded line is to be prevented, the current flow in each conductor must be equal in amplitude in order to set up equal and opposite magnetic fields that are thereby canceled out. This condition may be obtained only if the line is clear of all obstructions, and the distance between the wires is small. If, however, the line runs near some grounded or conducting surface, one of the two conductors will be nearer that obstruction than the other. A certain amount of capacitance exists between each of the two conductors and the conducting surface over the length of the line, depending upon the size of the obstruction. This capacitance acts as a parallel conducting path for each half of the line, causing a division of current flow between each conductor. Since one conductor may be nearer the obstruction than the other, the current flow will accordingly be increased, resulting in an inequality of current flow in the two conductors and therefore incomplete cancellation of radiation. The shielded line, therefore, eliminates such losses to a considerable degree by maintaining balanced capacitances to ground.

AIR COAXIAL

The air coaxial line has advantages that make it practical for operation at the ultrahigh frequencies. It consists of a wire mounted inside of, and coaxially with, a tubular outer conductor (fig. 12-9,E). In some cases the inner conductor also is tubular. The inner conductor is insulated from the outer conductor by insulating spacers or beads at regular intervals. The spacers are made of pyrex, polystrene, or some other material possessing good insulating qualities and having low loss at high frequencies.

The chief advantage of the coaxial line is its ability to keep down radiation losses. In the two-wire parallel line the electric and magnetic fields extend into space for relatively great distances and tend to cause radiation losses and noise pickup from other lines. In a coaxial line, however, no electric or magnetic fields extend outside the outer conductor. They are confined to the space between the two conductors. Thus, the coaxial line is a perfectly shielded line.

The disadvantages of such a line are: (1) it is expensive; (2) at extremely high frequencies its practical length is limited because of the considerable loss that occurs; and (3) it must be kept dry in order to prevent excessive leakage between the conductors. To prevent condensation of moisture, the line may be filled in certain applications with dry nitrogen at pressures ranging from 3 to 35 pounds per square inch. The nitrogen is used to dry the line when it is first installed, and a pressure is maintained to ensure that the leakage will be outward.

SOLID COAXIAL

Concentric cables are also made with the inner conductor consisting of flexible wire insulated from the outer conductor by a solid and continuous insulating material, as shown in figure 12-9,F. Flexibility may be gained if the outer conductor is made of a metal braid, but the losses in this type of line are relatively high.

WAVEGUIDES

Two common types of waveguides are the cylindrical type (fig. 12-9,G) and the more often used rectangular type (fig. 12-9,H). The term "waveguide" is applicable to all types of transmission lines in the sense that they are used to direct or guide the energy from one point to another. In this sense it does not matter whether the line is composed of a single conductor, two or more conductors, a coaxial line, a hollow metal tube, or a dielectric rod. Usage, however, has limited the meaning of the word to the hollow metal tube and the dielectric transmission line. The term "waveguide," as used in this text, means a hollow metal tube.

The transmission of an electromagnetic wave along a waveguide is closely related to its transmission through space. At powerline frequencies the current flow through the conductors was long considered to be the means by which energy is transmitted over a line, and the external electric and magnetic fields were regarded as coincidental to that transmission. That this may not necessarily be the case may be deduced from the fact that today energy may be transmitted along a waveguide with no longitudinal current flow along the guide. Thus it is believed that the energy transmitted is contained in the electromagnetic fields that travel down the waveguide and current flow in the guide walls only provides a boundary for these electric and magnetic fields.

Types of Waveguides

Waveguides may be classed according to cross section (rectangular, elliptical, or circular) or according to material (metallic or dielectric). Dielectric waveguides are seldom used because the losses for all known solid dielectric materials are too great for efficient transmission.

Of the three types of hollow-tube waveguides the rectangular cross section type is most commonly used. Circular waveguides are seldom used because it is difficult to control the plane of polarization and the mode of operation. (Modes are described later.) Circular waveguides involve the further difficulty of joining curved surfaces when a junction is required. They do find use in rotating joints, however, because of their circular symmetry, both physical and electrical. Elliptical waveguides are not used because of fabrication, joining, and bending difficulties.

Advantages of Hollow Waveguides

A hollow waveguide has lower loss than either an open-wire line or a coaxial line in the frequency range for which it is practical. An open-wire line has three kinds of loss—(1) radiation, (2) dielectric, and (3) copper. In the coaxial line there is no radiation loss because the outer conductor acts as a shield which confines the magnetic and electric fields to the space between the inner and outer conductors. Both the coaxial line and the hollow pipe are perfectly shielded lines and have no radiation loss.

Dielectric loss in the insulating beads of a coaxial line is considerable at very high frequencies, but air has negligible dielectric loss at any frequency. Because hollow metal waveguides are usually filled with air, they have negligible dielectric loss.

The third kind of loss is the copper loss. At high frequencies the current flows in a thin layer near the surface of the conductor. As the frequency increases, the thickness of this layer decreases, thus reducing the effective cross section of the conductor and causing the copper loss to increase as the effective resistance of the conductor becomes greater. In a coaxial line most of the resistance and most of the copper loss are in the inner conductor because the circumference of this conductor is less, and for a given penetration of current the effective cross section is less than that of the outer conductor.

For example, if the current flows in a very thin layer at the surface of both conductors, and if the inner circumference of the outer conductor is five times that of the outer surface of the inner conductor, the area through which current flows in the outer conductor is five times that of the inner conductor. The resistance, R, of a conductor is

$$R = \frac{\rho L}{A}$$

where ρ is the resistivity of the metal, L the length of the conductor, and A the area of the cross section through which the current flows. Therefore, the resistance of the inner conductor is five times that of the outer conductor. If the inner conductor were eliminated, the copper losses would be greatly reduced. A coaxial line without the inner conductor is equivalent to a round hollow waveguide.

Because the waveguide has less copper loss than a coaxial line, and because it has negligible dielectric loss and no radiation loss, the total losses of a waveguide above the cutoff frequency are less than those of a coaxial line of the same size operating at the same frequency.

The waveguide is simpler in construction than the coaxial line because the inner conductor and its supports are eliminated. Because there is no inner conductor which may be displaced or broken by vibration or shock, the waveguide is more rugged than the coaxial line.

Disadvantages of Hollow Waveguides

The minimum size of the waveguide that can be used to transmit a certain frequency is proportional to the wavelength at that frequency. This proportionality depends upon the shape of the waveguide and the manner in which the electromagnetic fields are set up within the pipe. In all cases there is a minimum frequency that can be transmitted. The lowest cutoff frequency is determined by the inside dimensions shown in figure 12-9,H. The cutoff wavelength, λ_{co} associated with the cutoff frequency, is equal to twice the inside width of the guide, or

$$\lambda_{co} = 2a$$

Higher frequencies, however, can be transmitted. The width of the guide for these frequencies is greater than their corresponding free-space half-wavelengths.

The distance, b, is not critical with regard to frequency. However, this distance determines the voltage level at which the waveguide arcs over. Therefore, for high power and voltage, the distance, b, should be large. In practice, b may be from 0.2 to 0.5 of the wavelength in air, and a may be about 0.7 times the wavelength in air.

Because the cutoff frequency corresponds to a wavelength that is equal to twice the inside width of the guide, waveguides are not used extensively at frequencies below approximately 1000 mc (30 centimeter). At lower frequencies the guide would be too large. For example, to transmit 30-centimeter waves, a rectangular waveguide would have to be wider than 15 centimeters. For 1-meter waves the waveguide would have to be about 2 1/3 feet wide, and for 10-meter waves, 23 feet wide.

The installation of a waveguide transmission system is somewhat more difficult than the installation of other types of lines. The radius of bends in the guide must be greater than two wavelengths to avoid excessive attenuation and the cross section of the guide must be maintained uniform around the bend. These difficulties hamper installations in restricted spaces. If the guide is dented, or if solder is permitted to run inside the joints, the attenuation of the line is greatly increased. In addition to the increased attenuation that they cause, dents and beads of solder also reduce the breakdown voltage of the waveguide and cause standing waves in the guide.

Although such faults may not cause arc-over in the guide, they limit the power handling capacity of the system and make the possibility of arc-over more likely. Thus, unless great care is exercised in the installation, one or two carelessly made joints could nullify completely the initial advantage obtained from the use of the waveguide.

Modes of Transmission

For convenience of reference, a system employing letters and subscript numbers was devised for describing waveguide modes. Figure 12-10,A, shows the TE mode of operation of a rectangular waveguide. The letters TE indicate a mode of operation in which the electric field composed of parallel E lines) lies in transverse planes that contain the X and Y axes and in which the E lines are parallel to the Y axis and are perpendicular to the longitudinal (Z) axis of the guide. Similarly the letters TM (fig. 2-10,B) indicate that the magnetic field (composed of closed loops) lies in transverse planes that contain the X and Y axes and are wholly transverse to the guide axis.

For rectangular waveguides the accepted system of subscripts is that the first number subscript indicates the number of half-wave variations of the transverse field in the wide dimension of the guide, and the second number indicates the number of half-wave variations of the same field in the narrow dimension. The TE_{10} mode, for example, means that the electric field has one half-wave variation in the wide dimension, and none in the narrow dimension. The TM_{11} mode means that the magnetic field has one half-wave variation in both the wide and the narrow dimensions. The mode having the lowest cutoff frequency for a given size of guide is called the DOMINANT MODE for that guide. The dominant mode, TE_{10}, for rectangular waveguides is the mode commonly used. There are many reasons for this. This mode is easily excited, it is planepolarized, it is easily matched to a radiator, and the plane of polarization is easily controlled. Other reasons are that its cutoff frequency is dependent upon only one of the guide dimensions while many of the other modes depend on two dimensions, with the result that it is easy to design the waveguide so that only this one mode may exist in it.

Coupling

There are three principal ways in which energy can be put into and removed from waveguides. The first is by placing a small loop of wire so that it "cuts" or couples the H lines of the magnetic field, as in a simple transformer. The second is by providing an "antenna" or probe which can be placed parallel to the E lines of the electric field. The third method is to link or conduct the fields inside the guide by external fields through the use of slots or holes in the walls.

The foregoing is a general description of waveguides. A fuller discussion of the theory and operation of waveguides and cavity resonators will be included in an advance course.

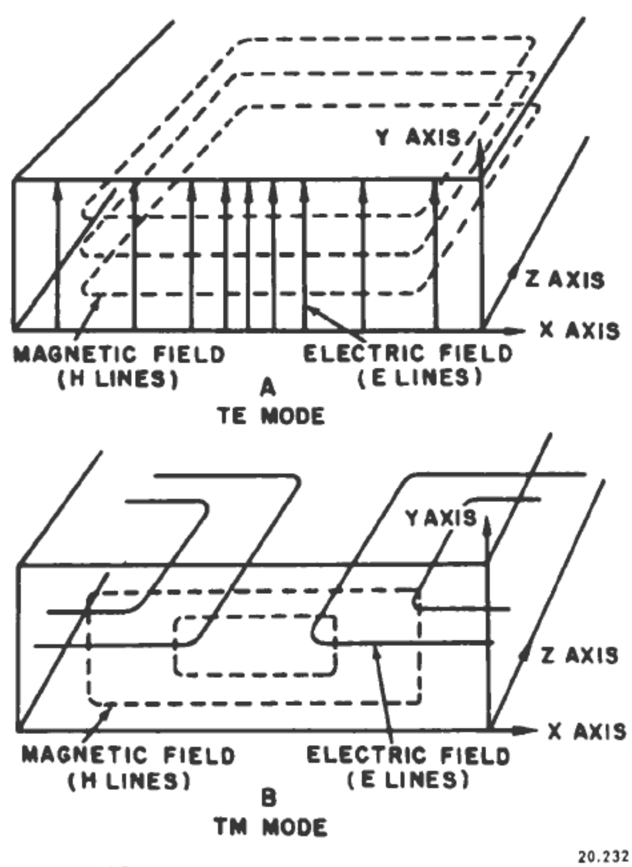


Figure 12-10.—TE and TM modes of operating a waveguide.

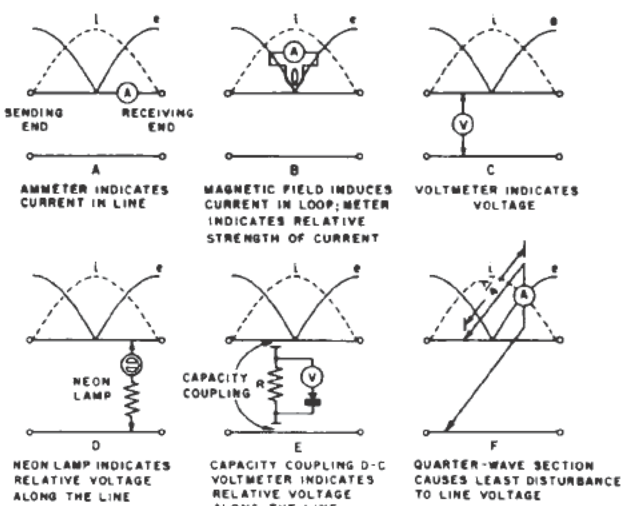
MEASUREMENTS ON R-F LINES

It is often necessary to determine if standing waves are present on a transmission line and, if present, where the loops and nodes of voltage and current occur. It may also be necessary to determine the SWR for voltage and current. Therefore, it is necessary to make electrical measurements on the line.

METHODS OF MAKING MEASUREMENTS

There are several methods of determining the magnitude of voltage or current at any point on an r-f line. In making these measurements it must be remembered that the magnetic field around a line varies directly with the current, and that the electrostatic field about the line varies directly with the voltage.

Current may be observed at any point on a line either by cutting the line and inserting a suitable ammeter (fig. 12-11,A), or by placing a loop (fig. 12-11,B) connected to an ammeter in the magnetic field. In figure 12-11,A, the a-c ammeter is connected in series with the line. The ammeter indicates the standing wave of current at the point where the measurement is made. If this method is not practical, the method shown in figure 12-11,B, may be used instead. The coil is moved along the line, and the meter will indicate maximum current at the point where maximum current is induced in the coil—that is, at a current loop—and will indicate minimum current at a current node.



20.233

Figure 12-11.—Methods of measuring standing waves of voltage and current.

In order to determine the voltage between lines at any point along the line, an a-c voltmeter is connected across the line, as indicated in figure 12-11,C. The voltage between the lines may also be measured by the method shown in figure 12-11,D. When the neon lamp is connected across the line the lamp will glow with a degree of brightness that is proportional to the voltage across the line at different locations along the line. If the field is strong enough, the lamp will glow when it is in close proximity to the line even if there is no physical contact. This method is convenient, although it lacks the precision of other methods. Greater precision may be obtained by the use of a sensitive type d-c voltmeter and rectifier as shown in figure 12-11,E. The resistor, R, is capacity-coupled to the two sides of the line, and the voltage drop across R is measured by the voltmeter.

Each of the methods described thus far has the disadvantage that a certain amount of energy is absorbed from the line. This absorption of energy at the point of measurement represents a change in impedance at this point and causes line reflections. Thus, the foregoing methods of making r-f measurements temporarily alter the normal characteristics of the line during the time the measurements are being taken.

A method of r-f measurement that causes the least disturbance to the line is shown in figure 12-11,F. It is composed of a $\frac{\lambda}{4}$ section shorted by means of an r-f ammeter. This device presents an extremely high impedance to the line, and therefore little current is needed to energize it.

When coaxial lines are used, the magnetic and electric fields are contained within the space between the outer and inner conductors and are not accessible for measurement as in open lines. In making measurements on this type of line, the arrangements shown in figure 12-12 are used. A probe is inserted in the slot, but not far enough to touch the inner conductor. The probe is a slender rod that acts as an antenna. It is excited by the electric field which is parallel to the probe. Since the line current flows parallel to the slot, the effective resistance of the coaxial line is not appreciably reduced by the presence of the slot.

Because the coupling is slight, very little energy is extracted by the probe. The r-f energy is detected by a crystal rectifier in figure 12-12,A, and the resulting d-c, which varies

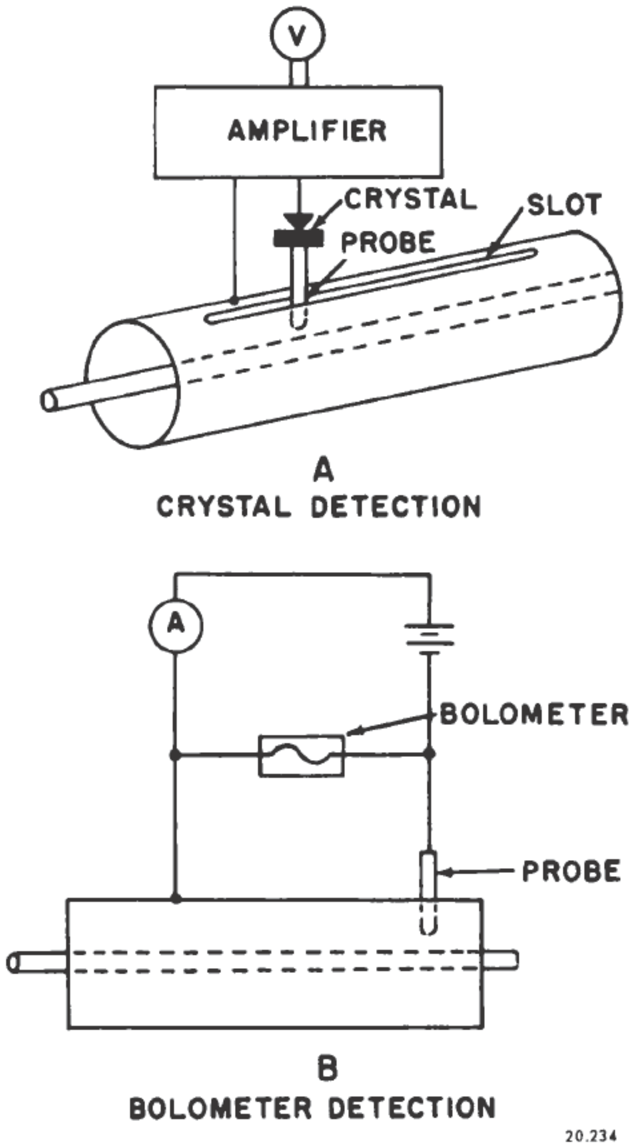


Figure 12-12.—Making measurements on coaxial lines.

in magnitude with the a-c signal voltage, is amplified and fed to the voltmeter. The r-f energy is detected in figure 12-12,B, by a bolometer the resistance of which varies with temperature. This action varies the d-c current in ammeter A. The temperature of the bolometer varies with the amount of r-f current from the probe. The bolometer itself is generally a 0.01 ampere fuse having a positive temperature coefficient.

WAVELENGTH MEASUREMENTS

Because the distance between a voltage loop and the next adjacent voltage node or a current loop and the next adjacent current node is equal to a quarter-wavelength, one wavelength is equal to four times the quarter-wave, as shown in figure 12-13.

Because energy travels more slowly on a wire than in free space, the wavelength is a little shorter on the wire than in space. The electrical length of a wire therefore differs slightly from the length in terms of the free-space wavelength. This results from the capacitive effects between the wires and ground that decrease the velocity of propagation on the line. The spacers and insulating material used have a dielectric constant greater than air, and this also increases the effective capacitance.

The electrical quarter-wavelength for various types of lines may be calculated from the formula

$$L = \frac{246 \times k}{f}$$

where L (in feet) is the quarter-wavelength, k is a constant that depends on the type of line, and f is the frequency in megacycles. The constant, k, for a parallel line is 0.975, and for an air-insulated concentric (coaxial) line is 0.85.

LECHER LINES

Lecher lines are two-wire transmission lines that are used as tuned-circuit elements or as resonant lines for the purpose of obtaining wavelength. Such lines are, in general, between 1/4 and 5 wavelengths long and usually have a shorting bar that is adjustable over a considerable range of length.

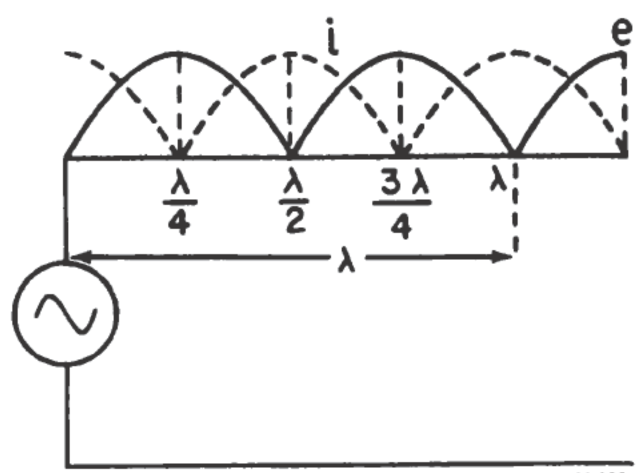


Figure 12-13.—Determination of wavelength by means of standing waves.

In the Lecher lines shown in figure 12-14,A and B, two parallel wires are extended a distance equal to slightly more than 5 quarter-wavelengths. By the use of these lines the wavelength and the frequency of the r-f signal may be determined. The wavelength may be determined by measuring the distance between successive maxima and minima of the current or voltage waveforms. The frequency is determined by making proper substitutions in the preceding formula, transposing and solving for f.

The use of the shorting bar and the pickup coil to determine current maximums and minimums is illustrated by the two positions of the shorting bar shown in the figure. The current wave shown in figure 12-14,A, indicates that the standing wave of current is minimum at the location of the coupling coil. Very little current flows through the coil, and the weak magnetic field that results induces only a slight current in the pickup coil and the indication of the ammeter is a minimum. In figure 12-14,B, the shorting bar has been moved a quarter-wavelength toward the r-f generator, and the current at the coupling coil is now maximum. Maximum current is induced in the meter pickup coil, and the ammeter indication is a maximum.

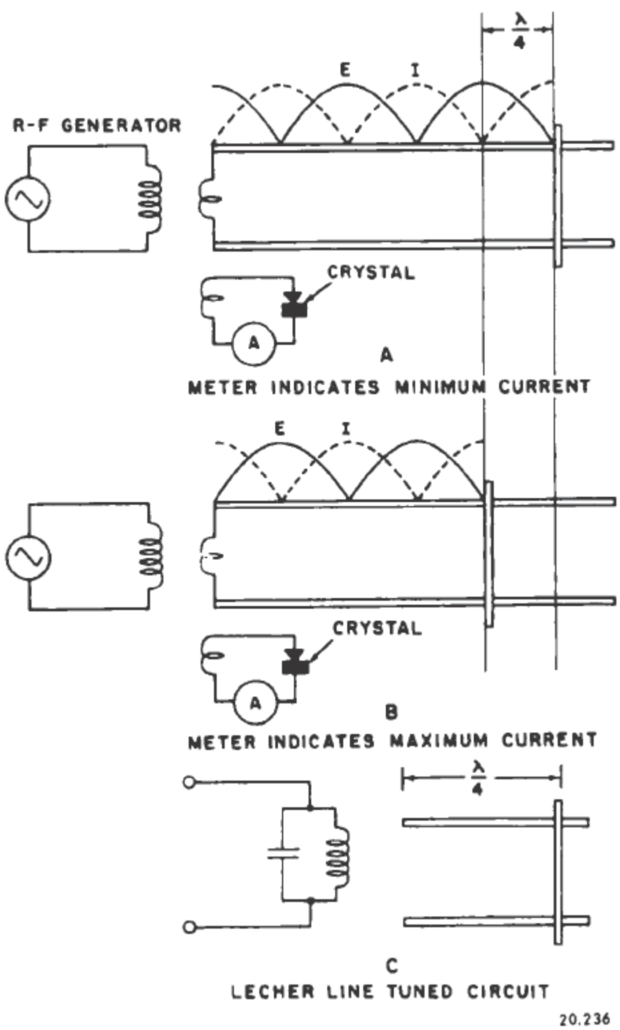


Figure 12-14.—Lecher lines.

A Lecher line one-quarter wavelength long has the characteristics of a parallel-resonant circuit (fig. 12-14,C) and therefore may be used as a tuned circuit in an ultrahigh-frequency oscillator. At 400 mc a quarter-wavelength line is only a little more than 7 inches long and therefore is of a practical length for oscillators having frequencies that lie in the upper end of the v-h-f band and the lower end of the u-h-f band.

APPLICATIONS OF RESONANT LINES

In many applications, standing waves on transmission lines must be eliminated or reduced to the lowest possible level. In radar, for example, the SWR must be very near unity if satisfactory operation is to be obtained. Standing waves have the following bad effects:

1. The power-handling capacity of the line is reduced because at some points the voltage is greater than at others; and at other points the current is excessively high. At high-voltage points the insulation may break down, and at high-current points the temperature rise may be excessive.

2. The efficiency of the line is lowered because of the excessive current and accompanying I^2R loss. The line current and voltage are not in phase, hence the line power factor is low. The efficiency of transmission becomes a maximum for a given amount of power being transmitted only when the line power factor becomes unity and the effective current becomes a minimum. These conditions can exist only on a nonresonant line (no standing waves).

3. The effective resistance of the line is increased by the introduction of standing waves. There is also increased radiation loss and reduced efficiency.

For these reasons, resonant lines are seldom used to transmit large amounts of power over any considerable distance.

Resonant lines, however, have many important uses besides that of transmitting power from one point to another. For example, they may be used as metallic insulators, as wave filters and chokes, and as impedance-matching devices.

METALLIC INSULATORS

When a quarter-wave line is shorted at the output end and is excited to resonance at the other end by the correct frequency, there are standing waves of current and voltage on the line. At the short circuit, the voltage is zero while the current is at a maximum. At the input end (the end supporting the transmission line) the current is nearly zero and the voltage is a maximum. Therefore, at the input end the $\frac{E}{I}$ ratio, and thus the impedance, is very large. Because an exceedingly high impedance across the input end looks like an insulator to the transmission line, the quarter-wave line shorted at the output end may be used as an insulator at its two open terminals (those to which the transmission line is attached).

Figure 12-15,A, shows a quarter-wave section of line acting as a stand-off insulator for a two-wire transmission line. Naturally, for direct current this section acts as a direct short on the line, but for the particular frequency that makes the section a quarter-wavelength, it acts as a highly efficient insulator. At terminals A and B there is a high voltage and a low current. Because $Z=\frac{E}{I}$, the impedance between A and B must be very high. The insulator obtains a negligible amount of energy from the line to make up any losses caused by the circulating current. If the frequency varies too widely from the value for which the section is designed, the section rapidly becomes a poor insulator and begins to act as a capacitor or inductor across the line.

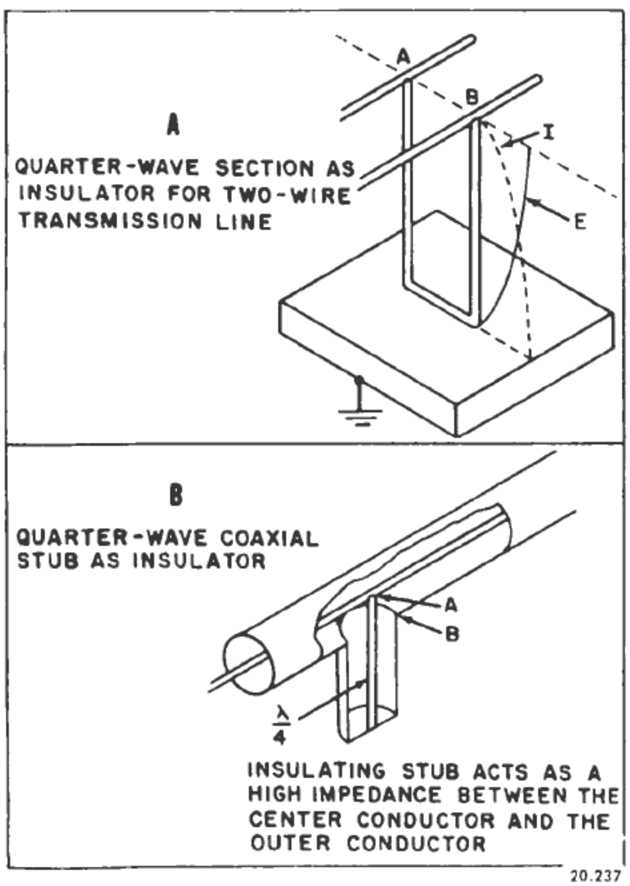
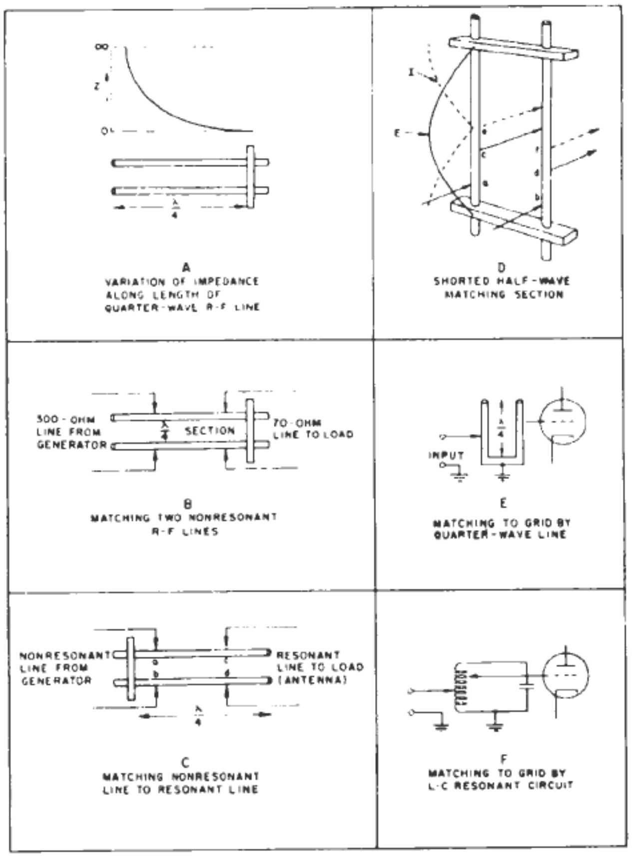


Figure 12-15.—Quarter-wave insulators.

Figure 12-15,B, shows a quarter-wavelength of coaxial line that is "teed" into a coaxial transmission line to support the center conductor. If the quarter-wave stubs are placed close enough together to provide adequate mechanical support, they are usually more efficient than beads of dielectric material—that is, if the coaxial line is operated at one frequency only. Metallic insulators are practical only at the higher frequencies where the quarter-wave stub is of a practical length

IMPEDANCE-MATCHING DEVICES

The impedance of a quarter-wave section of transmission line shorted at one end varies widely over its length, as is indicated in figure 12-16,A. At the shorted end the current is high and the voltage is low. Because $Z=\frac{E}{I}$, the impedance at the shorted end is low. At the open end, the conditions are reversed, and the impedance is high. When this section of r-f transmission line is excited, it is possible to match almost any impedance somewhere along the line. For example, a 300-ohm line may be matched to a 70-ohm line without the production of standing waves on either of the two lines that are being matched. Figure 12-16,B, shows how this connection may be made. Energy from the 300-ohm line sets up standing waves on the quarter-wave section. The connection between the 300-ohm line and the quarter-wave matching section is made at a point where the impedance of the quarter-wave line is 300-ohms. When this adjustment is made, the SWR on the 300-ohm line should be at a minimum, essentially unity.



20.238

Figure 12-16.—Transmission line as an impedance-matching device.

The 70-ohm line is similarly adjusted to bring about an impedance match near the shorted end.

The quarter-wave line may also be used to match a nonresonant line to a resonant line, as shown in figure 12-16,C. In order to be nonresonant, a line must be terminated in its characteristic impedance, and the terminating impedance should be approximately a pure resistance. The impedance of a shorted quarter-wave resonant section is zero at the shorting bar and increases along the line toward the open end. The shorting bar is adjusted to make a voltage maximum appear at cd; and the contacts, ab, between the nonresonant line and the quarter-wave section, are adjusted for the best match.

A half-wave section of line shorted at both ends is also used as an impedance-matching device, particularly in antenna coupling problems. Figure 12-16,D, shows a half-wave section excited at ab and having resonant current and voltage values as shown by the curves labeled E and I. The input (from the generator) to ab "sees" an impedance, Z_{ab}, equal to the $\frac{E}{I}$ ratio at that point, and the output (load) looking into cd "sees" a larger $\frac{E}{I}$ ratio, hence a larger impedance, Z_{cd}. The greatest impedance will be obtained at ef where the voltage is highest and the current lowest. Conversely, the lowest impedance points will be at the shorting bars where the current is high and the voltage low. Because the upper half of the half-wave section, or half-wave frame, repeats the impedance of the lower half, there will always be two points on the frame that have the same impedance. There will be a difference, however, in the phase of the currents involved, the current in one half being 180° out of phase with that of the other half.

Another example of the use of shorted quarter-wave section as an impedance-matching device is shown in figure 12-16,E. In this figure a relatively low impedance input is transformed to a high impedance to match the high input impedance to a grid. The equivalent lumped circuit is shown in figure 12-16.F.

A nonshorted resonant transmission line may also be used as an impedance-matching device, as shown in figure 12-17,A. A nonshorted quarter-wave transmission line having the correct characteristic impedance may be used to match two dissimilar impedances. The necessary characteristic impedance, Z_O, of the quarter-wave matching section is

$$Z_O = \sqrt{Z_s \times Z_r}$$

where Z_s is the impedance of line 1 and Z_r is the impedance of line 2. In this figure,

$$Z_O = \sqrt{300 \times 70} = 145 \text{ (approx.)}$$

The transformer analogy is shown in figure 12-17,B. A method of connecting a 300-ohm line by means of a quarter-wave matching section is shown in figure 12-17,C.

QUARTER-WAVE LINES AS FILTERS

The characteristic of a quarter-wave line allows it to be used as an efficient filter or suppressor of EVEN harmonics. Other types

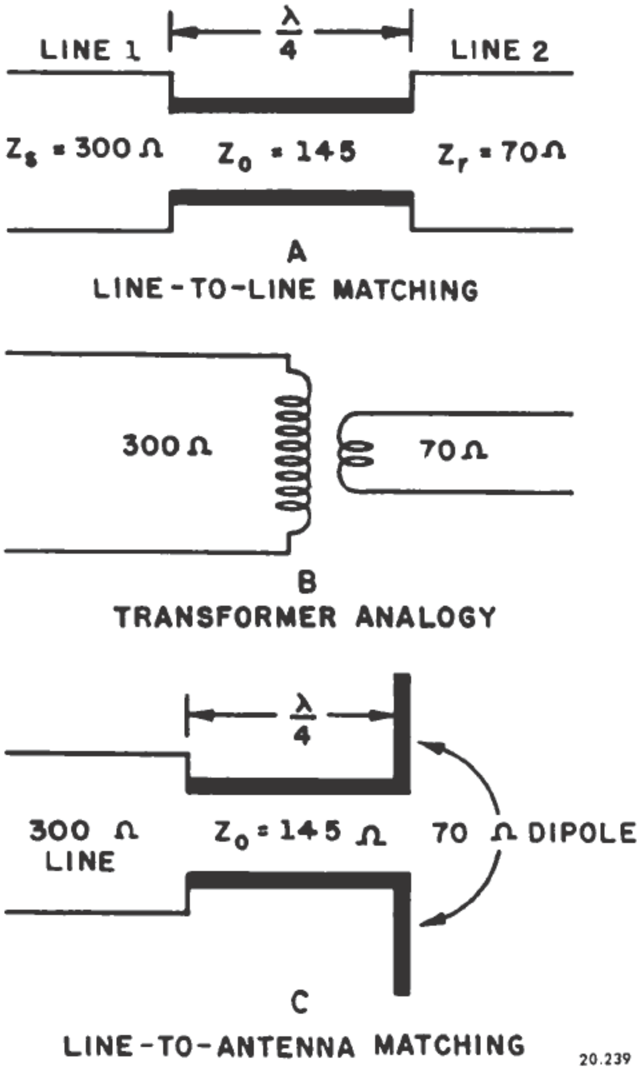


Figure 12-17.—Impedance matching with unshorted quarter-wave line.

of filters may be used for the elimination of ODD harmonics. In fact, filters may be designed to eliminate efficiently the radiation of an entire single side band of the modulated carrier.

Suppose that a transmitter is operating on a frequency of 5 mc and it is found that the transmitter is causing excessive interference on 10 and 20 mc. In addition to the other means of eliminating radiation at these even harmonic frequencies, a resonant transmission line may be used as a harmonic suppressor.

A quarter-wave line shorted at one end offers a high impedance at the unshorted end to the fundamental frequency. At a frequency twice the fundamental, such a line is a half-wave line, and at a frequency four times the fundamental, the line becomes a full-wave line. A half-wave or a full-wave line that is shorted at the output end offers zero impedance at its input end. Therefore, the radiation of even harmonics from the transmitting antenna can be eliminated almost completely by the circuit shown in figure 12-18,A.

The resonant filter line, ab, as shown, is a quarter-wave in length at 5 mc and offers almost infinite impedance at this frequency. In other words, the quarter-wave section looks like an insulator (to the transmission line) connected between the lines at the point where the antenna is connected. At the second harmonic, 10 mc, the line, ab, is a half-wave line and offers zero impedance at the antenna, thus shorting this frequency to ground. Again at 20 mc, the filter is a full-wave line and offers zero impedance. Thus, energy at this frequency is also grounded. The quarter-wave filter may be inserted anywhere along the nonresonant transmission line with similar effect—for example, at a in figure 12-18,B.

Both open and closed quarter-wave resonant lines may be used as wave filters. Figure 12-18,C, shows how more than one line filter may be connected between a transmitter and an antenna to eliminate the radiation of undesired frequencies. In this instance, a quarter-wave filter, b, that is open at the output end is inserted

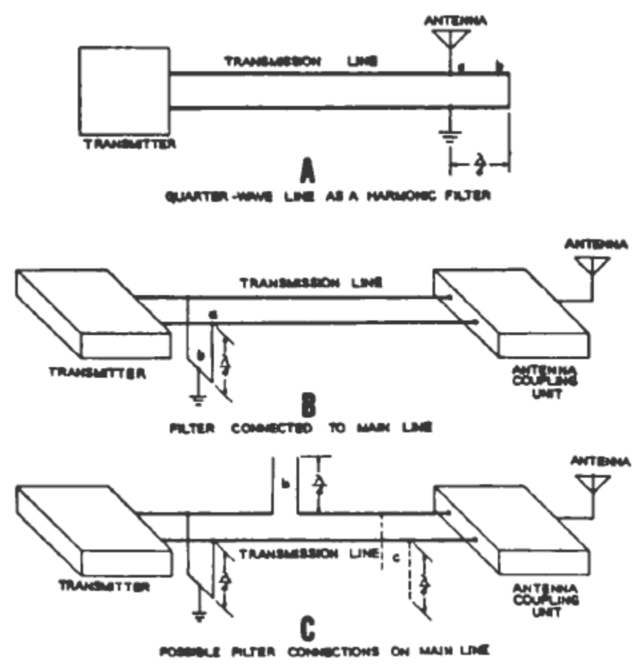


Figure 12-18.—Quarter-wave filters.

in series with the transmission line. A quarter-wave line that is open at the output end offers low impedance at the input end to the fundamental frequency. At each odd harmonic such a line is an odd multiple of a quarter-wave and therefore offers little impedance to odd harmonics. Actually, at the fundamental and odd harmonics the impedance at b is so low that it may be considered a continuous line, as if a short were placed across the base of the quarter-wave line. Thus, the quarter-wave open-filter line, b, in figure 12-18,C, passes the fundamental and odd harmonics along the line to the antenna-coupling unit. At even harmonics, however, the length of the open line (at b) becomes a half wave, or some multiple of a half wave, so that line b offers high impedance to the even harmonics and blocks their passage to the antenna-coupling unit.

Unfortunately, the foregoing methods of inserting wave filters in shunt with a line cannot be used to eliminate odd harmonics, because any attempt to eliminate the odd harmonics also results in serious loss to the fundamental frequency. For example, assume that line c of figure 12-18,C, is a quarter-wave at the third harmonic (15 mc). This frequency would be eliminated effectively before it could reach the antenna-coupling unit. However, the fundamental that is to be transmitted would also be greatly attenuated. If a line is a quarter-wave in length at 15 mc it is a twelfth-wave in length at 5 mc (wavelength varies inversely with frequency). A line a twelfth-wave in length would act as a capacitor and offer a low impedance to 5 mc. Therefore, although 15-mc radiation would be suppressed, the desired carrier would also be suppressed considerably.

CHAPTER 13

ANTENNAS AND PROPAGATION

PRINCIPLES OF RADIATION

A radio-frequency current flowing in a wire of finite length can produce electromagnetic fields that may be disengaged from the wire and set free in space. The principles of the radiation of electromagnetic energy are based on the laws that a moving electric field creates a magnetic field, and conversely, a moving magnetic field creates an electric field. The created field (either electric or magnetic) at any instant is in phase in time with its parent field, but is perpendicular to it in space. These laws hold true whether or not a conductor is present.

The electric (E) and magnetic (H) fields are perpendicular to the direction of motion through space. A right-hand rule may be applied that relates the directions of the E field, the H field and the propagation. This rule states that if the thumb, forefinger, and middle finger of the right hand are extended so that these three digits are mutually perpendicular, the thumb will point in the direction of the electric field, the forefinger in the direction of the magnetic field, and the middle finger in the direction of propagation.

In the instantaneous cross section of a radio wave shown in figure 13-1, the E lines represent the electric field and the H lines represent the magnetic field. If the right-hand rule is applied, the thumb points downward, representing the direction of the E lines; the forefinger, to the left, representing the direction of the H lines; and the middle finger away from the observer, representing the direction of propagation.

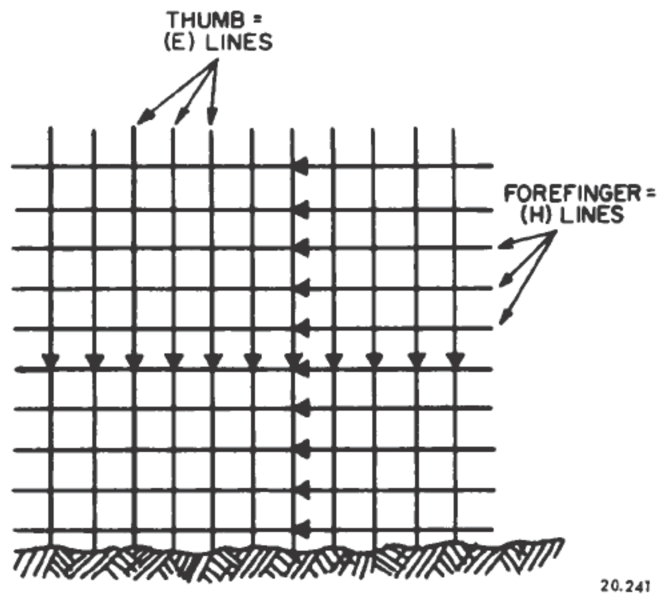


Figure 13-1.—Instantaneous cross section of a radio wave.

FUNDAMENTAL CONCEPTS

When r-f current flows through a transmitting antenna, radio waves are radiated from the antenna in all directions in much the same way that waves travel on the surface of a pond into which a rock has been thrown. It has been found that these radio waves travel at a speed of approximately 186,000 miles per second (300 million meters per second). The frequency of the radio wave radiated by the antenna will be equal to the frequency of the r-f current.

Because the velocity of the radio wave is constant regardless of its frequency, to find the wavelength (which is the distance traveled by the radio wave in the time required for one cycle) it is necessary only to divide the velocity by the frequency of the wave—

$$\lambda = \frac{300,000,000}{f}$$

where λ (the Greek letter lambda, used to symbolize wavelength) is the distance in meters from the crest of one wave to the crest of the next, f the frequency in cycles per second, and 300,000,000 the velocity of the radio wave in meters per second. This relationship is important in radio communications. It can also be expressed as

$$f = \frac{300}{\lambda}$$

where f is in megacycles, λ is in meters, and 300 is the velocity of propagation of the radio wave in millions of meters per second.

For example, the frequency of the current in a transmitting antenna that is radiating an electromagnetic wave having a wavelength of 2 meters is $\frac{300}{2}$, or 150 megacycles. Radio waves are usually referred to in terms of their frequency and are discussed in considerable detail later in this chapter.

The transmission line guides electrical energy from the generator to the antenna. The following explanations omit this device. The diagrams show the generator connected to the antenna.

In figure 13-2, two wires are attached to the terminals of a high-frequency a-c generator. The frequency of the generator output is chosen so that each wire is one-quarter of the wavelength, $\frac{\lambda}{4}$, corresponding to the generator frequency. The result is a common type of antenna known as a DIPOLE and is shown in figure 13-2,A.

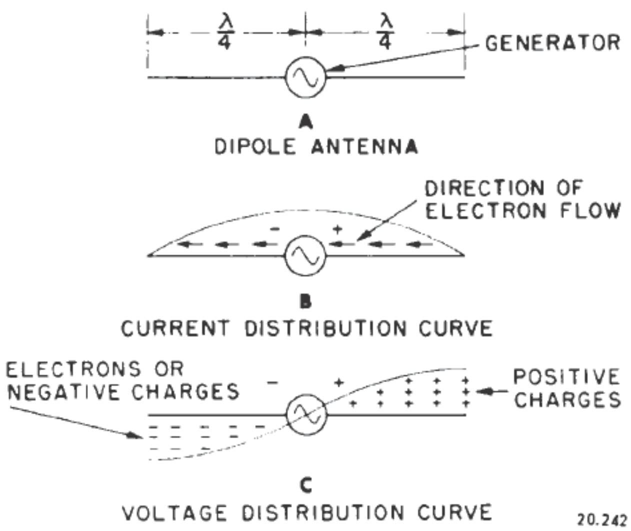


Figure 13-2.—Dipole antenna showing current and voltage distribution.

At a given instant, the right-hand terminal of the generator is positive and the left-hand terminal is negative. As like charges repel, electrons will flow away from the negative terminal as far as possible, while the positive terminal will attract electrons to it. Figure 13-2,B, shows the direction and distribution of electron flow at this instant. The current distribution curve indicates that the current flow is greatest at the center of the dipole and zero at the ends. At any given point along the antenna, except at the ends, the current variation is assumed to be sinusoidal with respect to time (the generator voltage has sine waveform). The relative current distribution is also sinusoidal with respect to the antenna length. Thus, an r-f ammeter inserted near the center of the antenna will indicate a relatively large effective current, and one inserted near the end will indicate a small effective current. The relative current distribution over the antenna will always be the same no matter how much or how little current is flowing, but the current amplitude at any given point on the antenna will vary directly with the amount of voltage developed at the generator terminals.

The generator voltage initiates the flow of antenna current. The action of the antenna is partly like that of a capacitor. When a capacitor becomes fully charged its voltage is maximum and the charging current ceases. In figure 13-2,C, the antenna voltage near the ends is maximum at the instant that the charging current is zero. Although no current flows at this instant, there is a maximum accumulation of electrons at the left end of the antenna and a deficit at the right end. Most of the charges are at the ends trying to get as far from the generator terminals as possible (like charges repel). The antenna voltage, like the antenna current, varies sinusoidally with respect to time. Also the antenna voltage varies sinusoidally with respect to the antenna length. Thus an r-f voltmeter connected between ground and one end of the antenna indicates a relatively large effective (rms) voltage. As the end probe is moved toward the antenna center the effective voltage is decreased to a low value. The antenna has both distributed inductance and capacitance and acts like a resonant circuit. At the center the current and voltage are in phase with each other; in the antenna wire between the center and the ends they are out of phase.

Summarizing:

1. A current having sine waveform flows in the antenna. Its distribution is sinusoidal, as shown in figure 13-2,B.

2. A sinusoidal distribution of charge, as shown in figure 13-2,C, exists on the antenna. Every half cycle the charges reverse position.

3. The sinusoidal variation in charge (voltage) is out of phase with the sinusoidal variation in current by one-quarter of a cycle, or 90 degrees, except at the center and the ends where the current and voltage are inphase.

INDUCTION FIELD

An alternating current flows in the antenna; therefore an alternating magnetic field, H, is set up around the antenna as shown at one instant in figure 13-3,A. Alternate positive and negative charges also appear on the antenna, causing an electric field (E in fig. 13-3,B) to be set up. This field is represented by lines of force drawn between the positive and negative charges (fig. 13-3,B). The arrow heads indicate the direction a unit positive charge would move at those points.

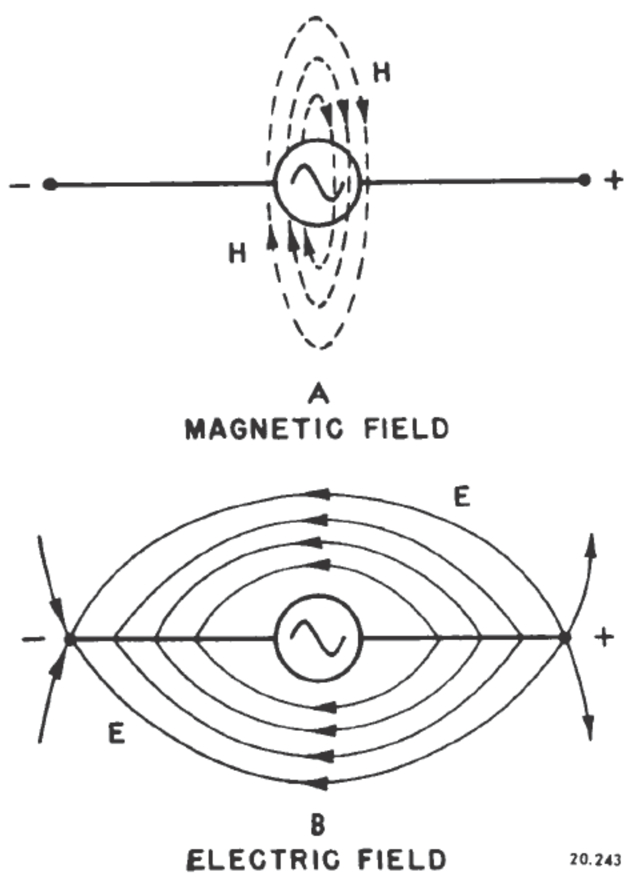


Figure 13-3.—Instantaneous field around an antenna.

Because the current and voltage that produce these fields are 90 degrees out-of-phase the two fields must also be out-of-phase by 90 degrees. Thus in spite of the fact that they are mutually perpendicular, these fields do not constitute the radiated electromagnetic field that passes through space from the transmitting antenna to the receiving antenna.

On the other hand, the magnetic and electric components of the radiated field are in phase with each other. The energy contained in the induction field cannot be detached from the antenna. The amplitude of the induction field energy varies inversely as the square of the distance from the antenna, and consequently its effect is entirely local. However, its effect must be considered in making field strength measurements of the radiation field in the vicinity of the antenna—that is, if only the field strength of the radiation field is to be measured. At $\frac{\lambda}{2\pi}$ wavelengths away from the antenna the field strengths of the two fields are equal. This distance is approximately one-sixth wavelength. At distances of a few wavelengths away from the antenna the induction field becomes negligible.

RADIATION FIELD

Although both a magnetic field and an electric field are radiated into space simultaneously, only the electric field is considered at present. The charges producing the electric field are constantly moving from one end of the antenna to the other as the polarity of the voltage at the generator changes. At one instant, one end of the antenna is positive; an instant later the antenna is uncharged. A negative charge next appears where the positive charge was. Then the antenna is again uncharged, and the cycle repeats.

In figure 13-4,A, electric flux lines are drawn between positive and negative charges. An instant later (fig. 13-4,B) the antenna is nearly discharged as the charges approach each other, thus bringing together the two ends of the flux lines associated with them. When the charges do touch, they seem to disappear, and their flux lines should also disappear. Most of the flux lines that represent the INDUCTION FIELD do disappear, but some flux is repelled by other lines nearer the antenna, and as in figure 13-4,C, the repelled flux lines are left with their

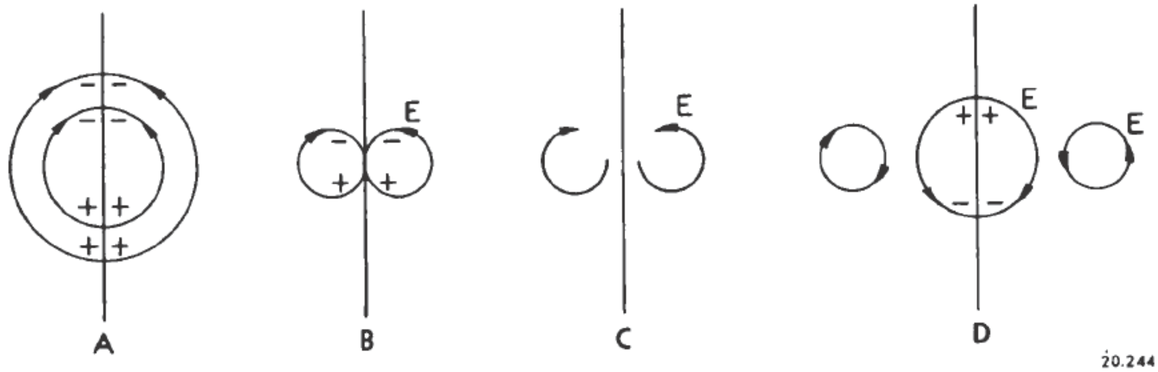


Figure 13-4.—Creation of closed electric flux lines on a half-wave antenna.

heads touching their tails. A closed electric field is thus created without an associated electric charge.

An instant after the independent field has been formed, the antenna is charged again in the opposite direction and produces lines of force that repel the recently formed independent electric field. Figure 13-4,D, shows that the repelling field is of the proper polarity to do this. The radiated field is forced away from the antenna at the speed of light.

As previously stated, a moving electric field generates a perpendicular magnetic field in phase with it. Therefore, because the radiated electric field is moving, it generates a magnetic field in accordance with this principle. The result is a radiated electromagnetic field that can travel great distances and deliver a usable part of its energy to a receiving antenna.

In the preceding discussions, the magnetic field generated by the antenna current has been ignored as a factor in generating the radiated field, but, by similar reasoning, magnetic lines of force may become detached from the antenna. Because the detached lines move away from the antenna, they generate a perpendicular in-phase electric field. The result is also a radiated electromagnetic field.

The electromagnetic radiation from the antenna is apparently made up of two components—the electric generated field and the magnetic generated field. These two fields can be shown to add and give a single sinusoidally varying radiated field.

The strength of the radiated field varies inversely with the distance.

RECEPTION

If a radiated electromagnetic field passes through a conductor, some of the energy in the field will set electrons in motion in the conductor. This electron flow constitutes a current that varies in accordance with the variations of the field. Thus, a variation of the current in a radiating antenna causes a similar varying current (of much smaller amplitude) in a conductor at a distant location. Any intelligence being produced as current in a transmitting antenna will be reproduced as current in a receiving antenna. The characteristics of receiving and transmitting antennas are similar, so that a good transmitting antenna is also a good receiving antenna.

BASIC ANTENNA PRINCIPLES

An antenna is a conductor or system of conductors that serves to radiate or intercept energy in the form of electromagnetic waves. In its elementary form an antenna, or aerial, may be simply a length of elevated wire like the common receiving antenna for an ordinary broadcast receiver. However, for communication and radar work, other factors make the design of an antenna system a more complex problem. For instance, the height of the radiator above ground, the conductivity of the earth below it, and the shape and dimensions of an antenna all affect the radiated-field pattern in space. Also, the antenna radiation often must be directed between certain angles in either the horizontal or the vertical plane, or both.

An antenna may be constructed to resemble a resonant two-wire line with the wires so arranged that the fields produced by the currents in the wires add in some directions instead of canceling completely. Figure 13-5,A, shows one way to prevent cancellation of the fields by making the earth one conductor. This permits considerable separation of the conductors. In this manner the fields resulting from the current expand considerably farther into space than if the other conductor were nearby, and therefore can be detached from the radiating conductor by rapid reversals more easily. Another way to accomplish the radiation is to spread the ends of the two-wire line as shown in figure 13-5,B, so these ends are 180 degrees apart (as shown in fig. 13-5,C). The currents which cancel the fields of each other, in figure 13-5,B, now aid in producing a field in space (fig. 13-5,C) similar to that produced in figure 13-5,A.

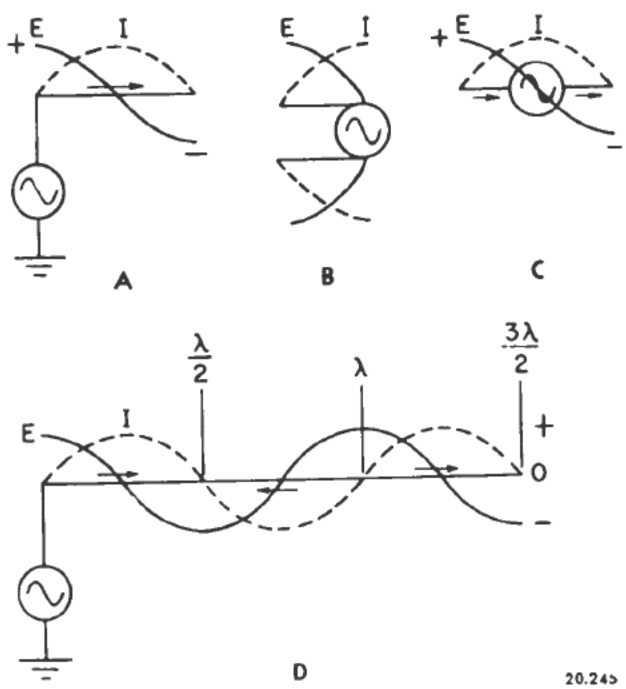


Figure 13-5.—Half-wave and multiple half-wave antenna.

The antenna shown in figure 13-5,A, can be extended, as shown in figure 13-5,D. Current flowing to the right is represented by the positive portion of the current curve, and current flowing to the left is represented by the negative portion of the curve. Similarly, voltages at any point on the antenna are positive or negative with respect to ground according to the position of the voltage curve above or below the axis represented by the antenna. The effectiveness of this device (fig. 13-5,D) as an omnidirectional antenna is not greatly increased by extending it horizontally close to the earth because currents flowing in opposite directions side by side produce canceling fields in some directions. However, if the antenna extends vertically above the earth, it is possible to elevate the effective radiation field a greater distance in all directions by operating the antenna at some harmonic, such as the third, fifth, or seventh harmonic of the fundamental frequency. The result is that the intensity of the radiated field at various points in space is considerably changed when compared with the field of the simple dipole.

Nonresonant lines also can be expanded to antennas, but they are not efficient radiators (except in a given direction between fixed stations). Resonant conductors are more efficient omni-directional radiators because they have large standing waves of voltage and current, and hence they produce intense fields with a minimum of generator current and voltage. The antenna shown in figure 13-5,A, which is cut to an electrical half wavelength, also radiates other frequencies, but its effectiveness as a radiator diminishes as the standing waves of current and voltage decrease.

ELECTRICAL LENGTH

If an antenna is made of very small wire and is isolated perfectly in space, its electrical length corresponds closely to its physical length. Thus, in free space, a 1-wavelength antenna for .10 meters would be 10 meters in length, and a half-wave length antenna for the same signal would be 5 meters in length. In actual practice, however, the antenna is never isolated completely from surrounding objects. For example, the antenna will be supported by insulators with a dielectric constant greater than 1. Therefore the velocity of the wave along the conductor is always slightly less than the velocity in space, and the physical length of the antenna is correspondingly less (by about 5 percent) than the corresponding wavelength in space. The physical

length, L, in feet, of a half-wave antenna for a given frequency is derived as follows:
Since

$$\lambda = \frac{300}{f} \text{ and } \frac{\lambda}{2} = \frac{300}{2f}$$

$$L = \frac{300 \times 3.28 \times 0.95}{2f} = \frac{468}{f}$$

where f is the frequency in magacycles, 3.28 feet equal 1 meter, and 0.95 represents the velocity of the wave in the antenna compared to that in free space. This formula does not apply to antennas longer than one-half wavelength.

ANTENNA INPUT IMPEDANCE

The antenna input impedance determines the antenna current at the feed point for a given value of r-f voltage at that point. The input impedance may be expressed mathematically by Ohm's law for alternating current—

$$Z = \frac{E}{I}$$

where Z is the antenna impedance and E and I are the r-f voltage and current respectively. Impedance is also expressed as

$$Z = R \pm jX$$

where R and X are the input resistance and reactance respectively.

In a half-wave antenna, the current is a maximum at the center and zero at the ends; whereas the voltage is a maximum at the ends and minimum at the center. The impedance, therefore, varies along the antenna and is minimum at the center and a maximum at the ends. Thus, if energy is fed to a half-wave antenna at its center, it is said to be CENTER FED (current fed); if energy is fed at the ends it is said to be END FED (voltage fed). In the case of a half-wave antenna isolated in free space, the impedance is approximately 73 ohms at the center and 2,500 ohms (allowing for losses) at the ends. The intermediate points have intermediate values of impedance.

Figure 13-6,A, is a plot of the input resistance of center-fed antennas for various wavelengths. Resistance values for both a thin and a thick antenna are plotted so that the effect of the diameter of the wire is apparent. In figure 13-6,B, the reactance is plotted as a function of wavelength. The curves show that an antenna may be either inductive or capacitive,

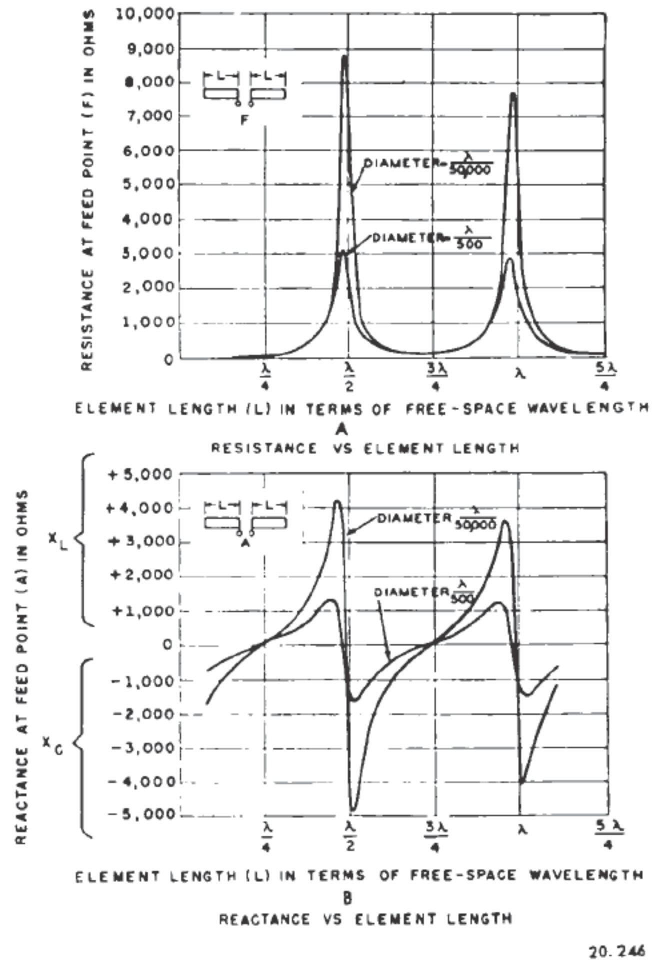


Figure 13-6.—Impedance curves for a center-fed antenna.

depending on its length, and that abrupt changes of impedance occur at multiples of a half-wavelength. The points in figure 13-6,B, where the reactance curves cross zero indicate the resonant lengths of the antenna. Because the curves are plotted in terms of the free-space wavelength, the effect of the reduced velocity of the wave motion along the antenna is shown by the curves. For example, a half-wave antenna element is resonant only when it is less than the free-space half-wavelength. This foreshortening is caused by the increased capacitance associated with the elements. If the diameter of the radiator is large, for example $\frac{\lambda}{500}$, the increased capacitance is greater than a thin element. As a result, the large-diameter radiator is foreshortened more than the thin radiator.

Figure 13-6 may be used to calculate the input impedance of center-fed antennas. For example, let it be required to find the impedance of a thin $\left(\text{diameter} = \frac{\lambda}{50,000}\right)$ antenna with a half length five-eighths of the wavelength being fed to it. In this case the antenna is not fully resonant. The impedance includes both resistance and reactance. The resistance is located on the proper curve halfway between $\frac{\lambda}{2}$ and $\frac{3\lambda}{4}$ and is approximately 150 ohms. Similarly, the reactance is found to be capacitive and approximately 1,100 ohms. The impedance in ohms is

$$Z = 150\text{-}j1,100 = 1,110\angle\text{-}82.2°$$

Thus, for maximum transfer of energy to the antenna a feedline to a $\frac{5\lambda}{8}$ center-fed antenna in free space must be designed to present an impedance of $150+j1,100 = 1,110\angle+82.2°$. In this case the feedline has a resistance of 150 ohms and an inductive reactance of 1,100 ohms.

The input impedance of an antenna is affected by the presence of nearby conductors (for example, the rigging on ships). Any object that can be affected by the induction field will distort the field and also the antenna voltage and current distribution. Therefore, the input impedance will be changed, and necessary corrections must be made to obtain the best match to each antenna. Because this effect is almost always difficult if not impossible to calculate, corrections are usually determined by trial-and-error methods.

RADIATION RESISTANCE

The antenna at the end of the transmission line is equivalent to a resistance that absorbs a certain amount of energy from the generator. Neglecting the losses that occur in the antenna, this is the energy that is radiated into space. The value of resistance that would dissipate the same power that the antenna dissipates is called the RADIATION RESISTANCE of the particular antenna. The power dissipated in a resistor is equal to I^2R. Likewise, the power dissipated in (radiated from) an antenna is equal to the current (at the feed point) squared times the radiation resistance of the antenna.

Figure 13-7 shows how the radiation resistance varies with antenna length, for an antenna

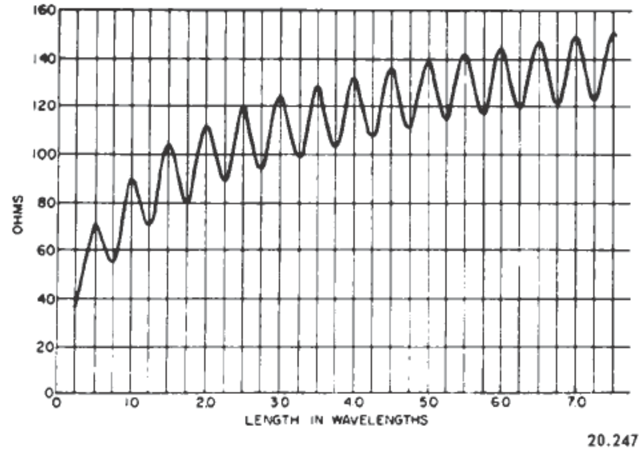


Figure 13-7.—Radiation resistance of antennas in free space plotted against length.

in free space. For a half-wave antenna the radiation resistance is approximately 73.2 ohms, measured at the current maximum, which is at the center of the antenna. For a quarter-wave antenna the radiation resistance measured at the current maximum is approximately 36.6 ohms. The radiation resistance is also affected somewhat by the height of the antenna above the ground and by its proximity to nearby objects. Other small antenna losses are caused by the ohmic resistance of the conductor, corona discharge, and insulator losses.

WAVE POLARIZATION

The position of a simple antenna in space determines the polarization of the emitted wave; that is, the direction of the electric lines of force determines the polarization of the wave. An antenna that is vertical with respect to the earth radiates a vertically polarized wave, while a horizontal antenna radiates a horozontally polarized wave. Figure 13-8,A, shows the vertical electric field component of a vertical antenna as a sine wave in the plane of the paper. Figure 13-8,B, shows the horizontal electric field component of a horizontal antenna as a sine wave lying in a horizontal plane. The first wave is vertically polarized; the second, horizontally polarized. For low frequencies the polarization is not distrubed and the radiation field has the same polarization at the distant receiving station that it had at the transmitting antenna. At high frequencies, however, the

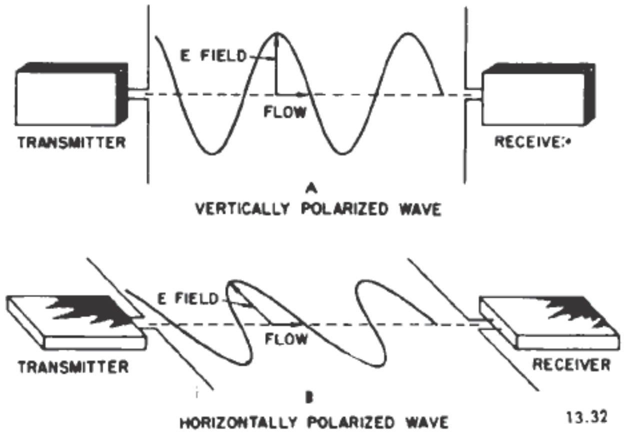


Figure 13-8.—Vertical and horizontal polarization.

polarization usually varies, sometimes quite rapidly, because the wave splits into several components which follow different paths. These paths will not be the same length; therefore, the recombined electric vectors representing the several components generally will not be parallel. If this is the case, the path traced by the point of the resultant vector may be circular or elliptical, and such a radiated field is known as either a CIRCULARLY or an ELLIPTICALLY POLARIZED FIELD.

When the antennas are close to the ground, vertically polarized waves yield a stronger signal close to the earth than do horizontally polarized waves. However, when the transmitting and receiving antennas are at least 1 wavelength above ground, the two types of polarization give approximately the same field intensities near the surface of the earth. When the transmitting antenna is several wavelengths above ground, horizontally polarized waves result in a stronger signal close to the earth than is possible with vertical polarization.

POLAR DIAGRAMS

The variation of signal strength around an antenna can be shown graphically by polar diagrams as in figure 13-9. Zero distance is assumed to be at the center of the chart (indicating the center of the antenna) and the circumference of the tangent circles is laid off in angular degrees. Computed or measured values of field strength then may be plotted radially in a manner that shows both magnitude and direction for a given distance from the antenna. Field strengths in the vertical plane are plotted on a semicircular polar chart (not shown in the figure) and are referred to as vertical polar diagrams.

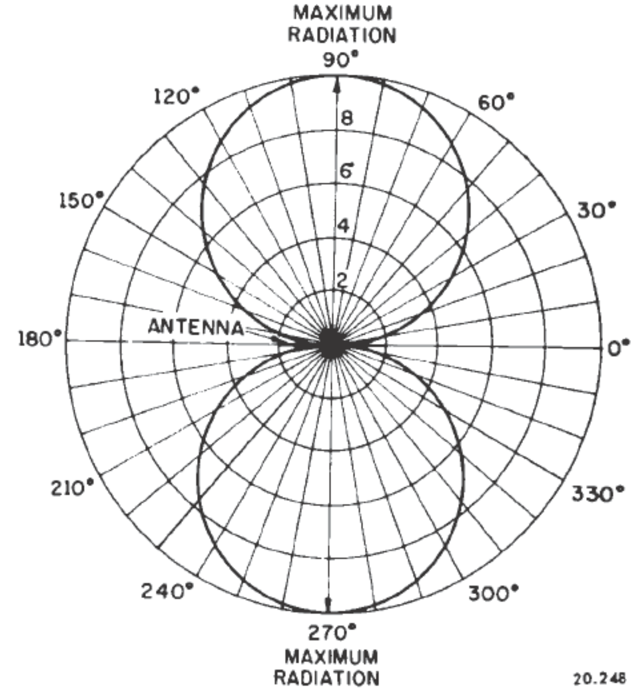


Figure 13-9.—Polar diagram of an antenna showing relative field strength.

BASIC TYPES OF ANTENNAS

An invention often borrows the name of its inventor. This is true about two basic antennas, the Hertz and Marconi.

HERTZ ANTENNA

Any antenna that is one-half wavelength long, or any even or odd multiple thereof, is a Hertz antenna and may be mounted either vertically or horizontally. A distinguishing feature of all Hertz antennas is that they need not be connected conductively to the ground, as are other antennas to be described. At the low and medium frequencies, these antennas are rather long and have little use in the Navy, except at shore stations where there is room for them. Vertical half-wave and five-eighths wave antennas are widely used with a-m broadcasting stations and have been built to heights of 1,000 feet or more for the lower broadcast frequencies. At the medium and high frequencies they are used extensively in fixed service when operation is not required at a large number of frequencies. This type of antenna is not particularly suited to services where a large number of different and unrelated frequencies must be transmitted using the same antenna, such as aboard ship.

Half-wave antennas showing two different methods of connecting the feedline together with the equivalent resonant circuits are shown in figure 13-10. For a half-wave dipole, the effective current is maximum at the center and minimum at the ends, while the effective voltage is minimum at the center and maximum at the ends. The voltage and current relationships are similar to those of the simple dipoles shown in figure 13-10.

MARCONI ANTENNA

A grounded antenna whose length is one-fourth wave or any odd multiple thereof is known as a Marconi antenna. Figure 13-11 illustrates the principle of a Marconi antenna mounted on the surface of the earth. The transmitter may be connected between the bottom of the antenna and the earth. Although the antenna is only one-quarter wavelength long, the reflection in the earth is equivalent to another quarter-wave antenna. By this arrangement, half-wave operation can be obtained from an antenna only one-quarter wavelength long. The impedance, voltage, and current relationships are similar to those in a half-wave antenna except that the input impedance at the base of a Marconi antenna is 36.6 ohms; whereas the input impedance of a Hertz antenna at the center is 73.2 ohms.

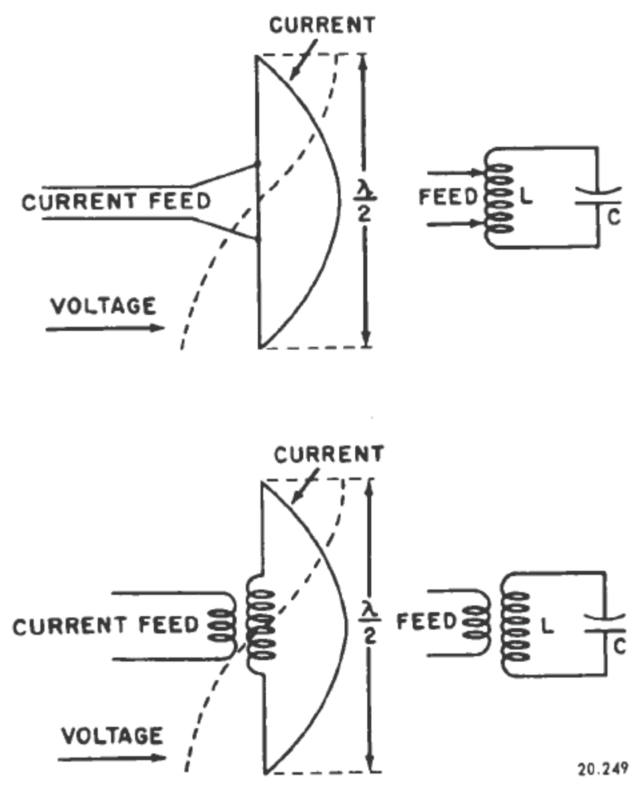


Figure 13-10.—Hertz antennas and equivalent circuits.

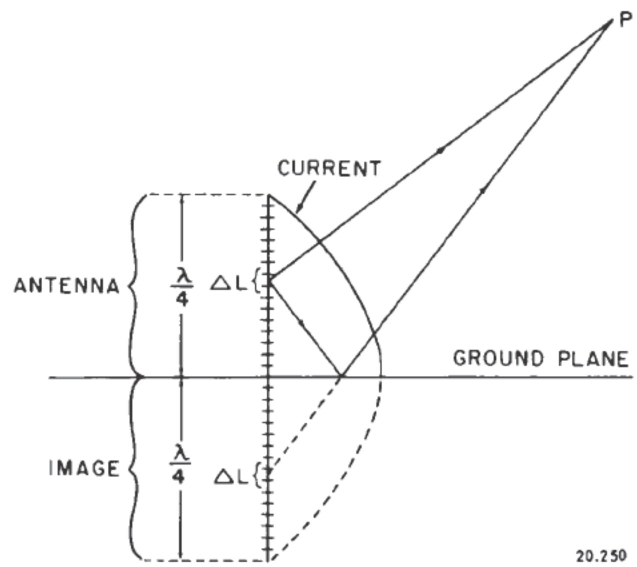


Figure 13-11.—Marconi antenna and image.

The quarter-wave antenna is used extensively with portable transmitters. On an airplane a quarter-wave mast or trailing wire is the antenna, and the fuselage produces the image. Similar installations are made on ships. A quarter-wave mast or horizontal wire is the antenna, and the hull and superstructure provide the image.

The effective current in the Marconi quarter-wave grounded antenna is maximum at the base and minimum at the top. The voltage is maximum at the top and minimum at the base.

ANTENNA TUNING

Aboard ship, antennas used for communications at the medium frequencies are not usually of the proper length to give optimum performance at the operating frequency. This condition exists because these antennas are all of standard size and shape or are installed in whatever space may be available for them and because they are each operated at more than one frequency. All equipment must be able to operate at any frequency within its tuning range. In this case, then, it is necessary to employ some means at the transmitter to adjust the antenna for reasonable efficiency at any frequency regardless of the physical dimensions or arrangement of any antennas that might be available.

As each transmitter is usually associated with only one antenna, which is of fixed length, the adjustment of the effective length of the antenna must be made by electrical means. This process is called ANTENNA TUNING and is accomplished by adding either inductance or capacitance to the antenna at the point where it is fed from the transmitter or transmission line, as shown in figure 13-12. Added inductance has the property of increasing the effective electrical length, while capacitance decreases it. In this manner the antenna can be made to respond as if the physical length was changed.

By tuning the antenna properly, the standing waves are increased and the radiated energy is increased.

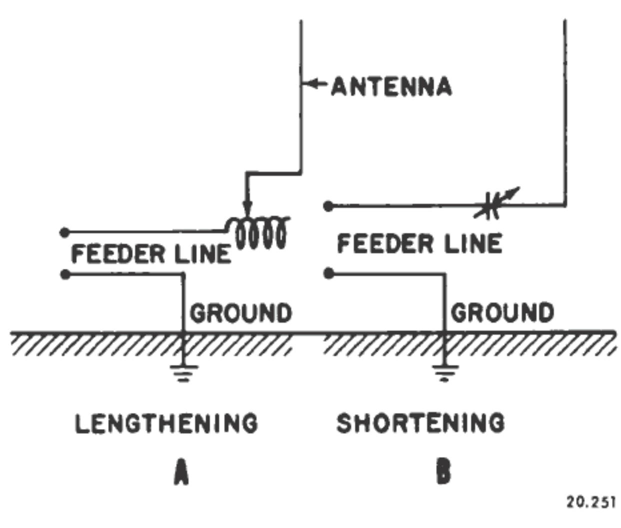


Figure 13-12.—Methods of correcting the electrical length of a grounded antenna.

Sometimes, particularly at low frequencies, it is not practical to make the quarter-wave grounded antenna the full physical quarter-wave in height shown in figure 13-13,A. Instead, it may be made shorter physically and then made the correct length electrically by top-loading it with a series inductor (fig. 13-13,B) or a parallel capacitor (fig. 13-13,C).

If the antenna is slightly more than a quarter wavelength high, the input at the base will be inductive, requiring the addition of a capacitor in series with the feed to bring the antenna into resonance.

RADIATION PATTERN FOR HALF-WAVE ANTENNAS

Because the current is greatest at the center of a dipole, maximum radiation takes place at this point and practically no radiation takes place from the ends. If this antenna could be isolated completely in free space, the points of maximum radiation would be in a plane perpendicular to the plane of the antenna at its center. The doughnut-shaped surface pattern is shown in figure 13-14,A, and the horizontal cross section pattern is shown in figure 13-14,B. Because a circular field pattern is created, the field strength is the same in any compass direction.

Theoretically, a vertical dipole in free space has no vertical radiation along the direct line of its axis. However, it may produce a considerable amount of radiation at other angles measured to the line of the antenna axis. Figure 13-14,C, shows a vertical cross section of the radiation pattern of figure 13-14,A. The radiation along OA is zero; but at another angle, represented by angle AOB, there is appreciable radiation. At a greater angle, AOC, the radiation is still greater. Because of this variation in field strength pattern at different vertical angles, a field-strength pattern of a vertical half-wave antenna taken in a horizontal plane must specify the vertical angle of radiation for which the pattern applies.

Figure 13-14,D, shows half of the doughnut pattern for a horizontal half-wave diople. The maximum radiation takes place in the plane perpendicular to the axis of the antenna and crosses through its center. A polar diagram representing the radiation pattern of a horizontal dipole is shown in figure 13-9.

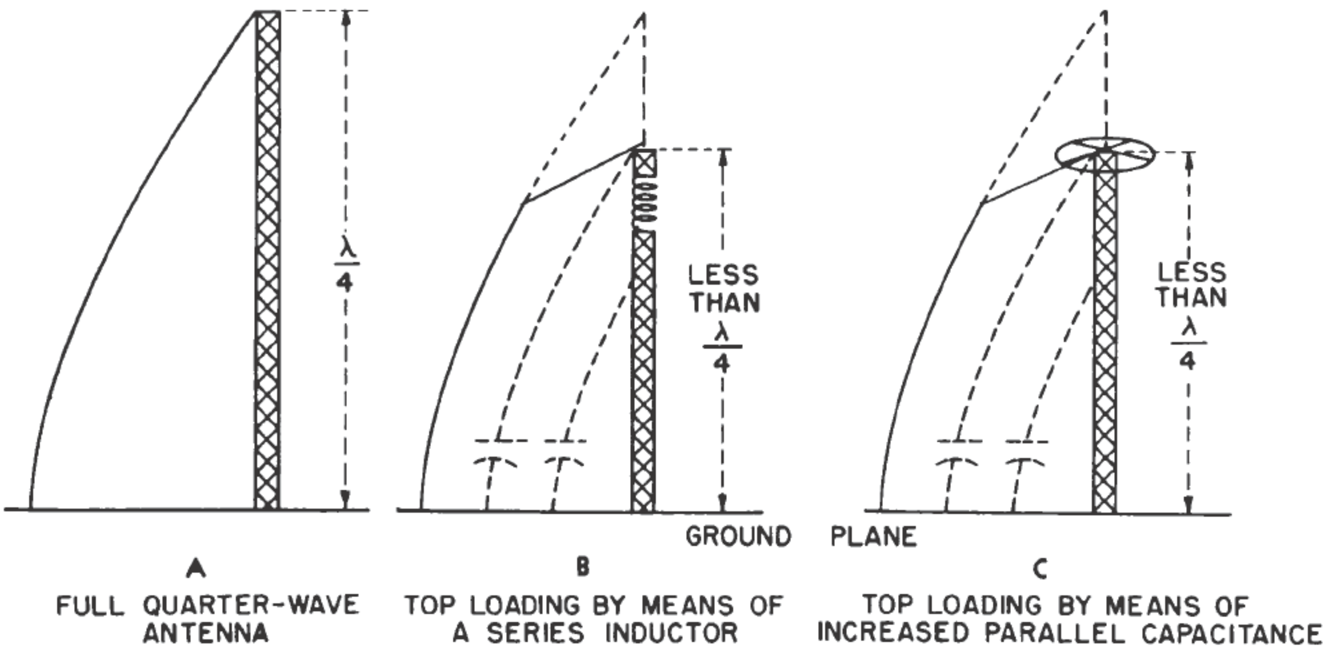


Figure 13-13.—Vertical quarter-wave antennas and methods of loading.

ANTENNA COUPLING

A common method of coupling the shipboard communications antenna to its associated transmitter is shown in figure 13-15. The antenna tuning system is made up of the antenna coupling capacitor, C1; the antenna tuning inductor, L1; the antenna tuning capacitor, C2; and the antenna feed switch, S1. The d-c blocking capacitor, C3, is connected in series with the antenna coupling capacitor, C1, to protect the antenna system from d-c potentials that might be occasioned by damage or voltage breakdown of the variable antenna coupling capacitor C1. The antenna tuning capacitor, C2, is variable, and is operated in one of two circuit arrangements. With S1 in position P, capacitor C2 is connected in parallel with L1, and the antenna is voltage (or shunt) fed. With S1 in position S, capacitor C2 is connected in series with L1 and the antenna is current (or series) fed.

The antenna system is tuned by first adjusting C1 to minimum coupling and tuning the final power amplifier stage to resonance. Then capacitor C2 and inductor L1 are tuned for antenna resonance.

The antenna now appears as a pure resistance to the final amplifier. The capacitance of C1 is then increased in small steps until the required loading of the final amplifier is obtained. Each time the coupling capacitor is changed, however, the final amplifier and antenna tuning circuit L1C2 must be retuned to resonance. Take care not to overcouple with C1. After the final amplifier and the antenna circuit are resonanted, the load on the final amplifier is purely resistive, and the maximum transfer of energy from the final amplifier to the antenna is obtained.

PROPAGATION OF RADIO WAVES

Any consideration of antennas must include a clear idea as to how the several paths operate by which a wave may travel between the sender and receiver equipments.

RADIO WAVE

When a radio wave leaves a vertical antenna the field pattern of the wave resembles a huge doughnut lying on the ground with the antenna in the hole at the center. Part of the wave moves outward in contact with the ground to form the GROUND WAVE, and the rest of the wave moves upward and outward to form the SKY WAVE, as shown in figure 13-16. The ground and sky portions of the radio wave are responsible for

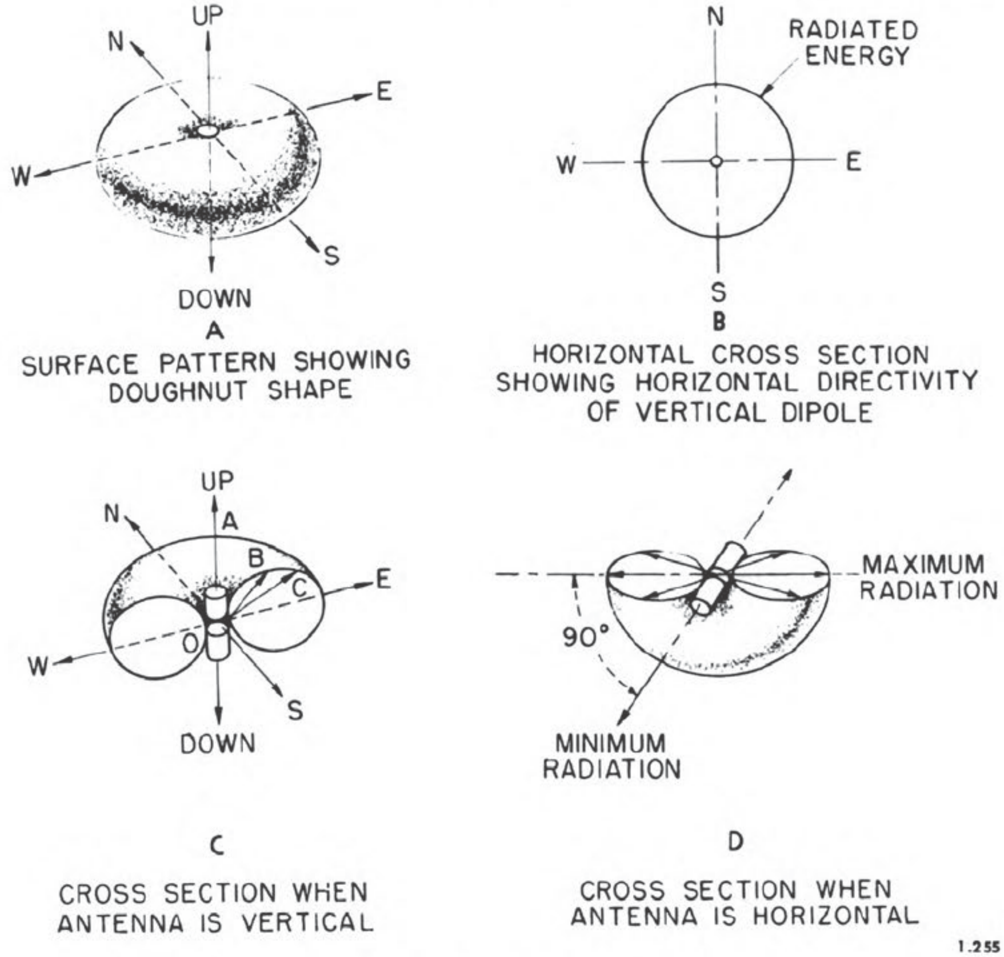


Figure 13-14.—Radiation pattern of a dipole.

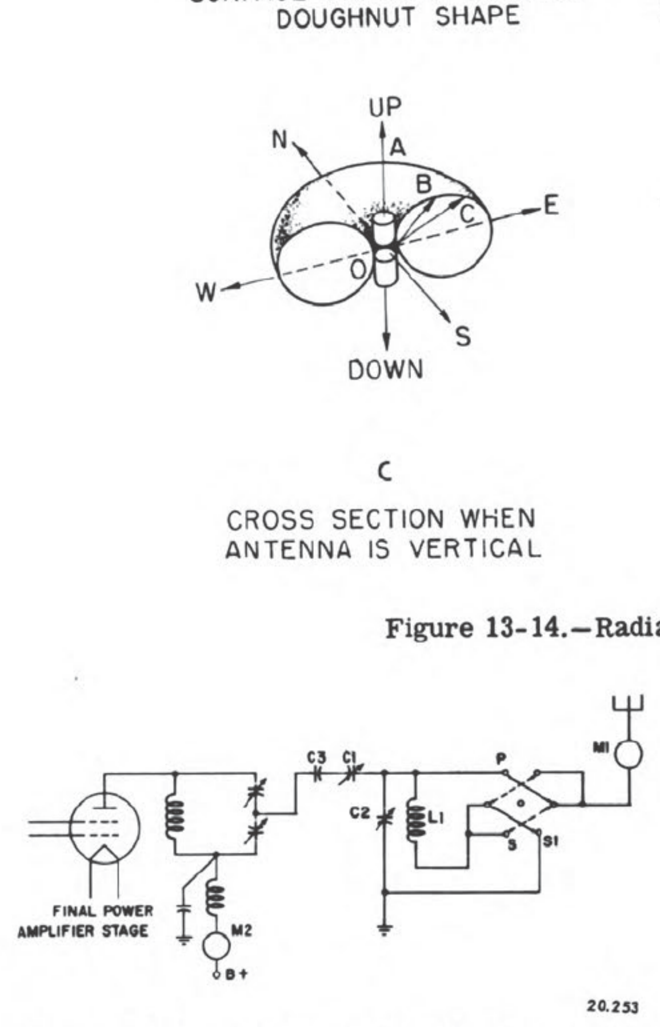


Figure 13-15.—Antenna coupling to transmitter.

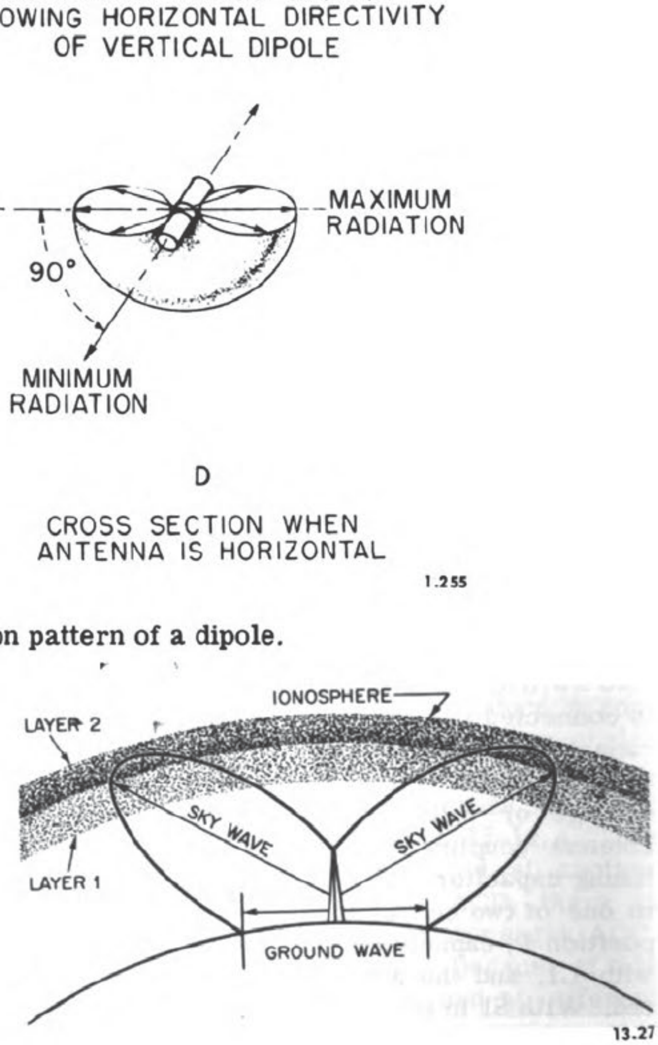


Figure 13-16.—Formation of the ground wave and sky wave.

two different methods of carrying the messages from transmitter to receiver. The ground wave is used both for short-range communication at high frequencies with low power, and for long-range communications at low frequencies with very high power. Daytime reception from most nearby commercial stations is carried by the ground wave.

The sky wave is used for long-range high-frequency daylight communication. At night, the sky wave provides a means for long-range contacts at somewhat lower frequencies.

GROUND WAVE

The ground wave is commonly considered to be made up of two parts, a surface wave and a space wave. The surface wave travels along the surface of the earth.

The space wave travels in the space immediately above the surface of the earth in two paths—one directly from transmitter to receiver, and the other a path in which the space wave is reflected from the ground before it reaches the receiver. The space wave follows two paths of different lengths, therefore, the two components may arrive in or out of phase with each other. As the distance from the transmitter is changed, these two components may add or they may cancel. Neither of these component waves is affected by the reflecting layer of atmosphere high above the surface of the earth, called the IONOSPHERE.

The space-wave part of the ground wave becomes more important as the frequency is increased or as the transmitter and receiver antenna height is increased. When the transmitting and receiving antennas are both close to the ground, the space wave components cancel. This is true because the ground-reflected component is shifted 180 degrees in phase upon reflection, has the same magnitude as the direct component, and travels a path of approximately the same length as that of the direct component. The surface wave part of the ground wave, therefore, is responsible for most of the daytime broadcast reception.

As it passes over the ground, the surface wave induces a voltage in the earth, setting up eddy currents. The energy to establish these currents is absorbed from the surface wave, thereby weakening it as it moves away from the transmitting antenna. Increasing the frequency, rapidly increases the attenuation so that surface wave communication is limited to relatively low frequencies.

Shore-based transmitters are able to furnish long-range ground wave communications by using frequencies between 18 and 300 kc with extremely high power.

Since the electrical properties of the earth along which the surface wave travels are relatively constant, the signal strength from a given station at a given point is nearly constant. This holds true in nearly all localities except those that have distinct rainy and dry seasons. There the difference in the amount of moisture causes the conductivity of the soil to change.

The conductivity of salt water is 5,000 times as great as that of dry soil. High-power low-frequency transmitters are placed as close to the edge of the ocean as practical because of the superiority of surface wave conduction by salt water.

SKY WAVE

That part of the radio wave that moves upward and outward and that is not in contact with the ground is called the SKY WAVE. It behaves differently from the ground wave. Some of the energy of the sky wave is refracted (bent) by the ionosphere so that it comes back toward the earth. A receiver located in the vicinity of the returning sky wave will receive strong signals even though several hundred miles beyond the range of the ground wave.

IONOSPHERE

The ionosphere is found in the rarefied atmosphere approximately 40 to 350 miles above the earth. It differs from the other atmosphere in that it contains a much higher number of positive and negative ions. The negative ions are believed to be free electrons. The ions are produced by the ultraviolet and particle radiations from the sun. The rotation of the earth on its axis, the annual course of the earth around the sun, and the development of sun spots all affect the number of ions present in the ionosphere, and these in turn affect the quality and distance of radio transmission.

The ionosphere is constantly changing. Some of the ions are recombining to form neutral atoms, while other atoms are being ionized by the removal of electrons from their outer orbits. The rate of formation and recombination of ions depends upon the amount of air present, and the strength of radiation from the sun.

At altitudes above 350 miles, the particles of air are too sparse to permit large-scale ion formation. Below about 40 miles altitude, only a few ions are present because the rate of recombination is so high. Ultraviolet radiations from the sun are absorbed in passage through the upper layers of the ionosphere so that below an elevation of 40 miles too few ions exist to affect materially skywave communication.

Densities of ionization at different heights make the ionosphere appear to have layers. Actually there is thought to be no sharp dividing line between layers, but for the purpose of discussion, a sharp demarcation is indicated.

The ionized atmosphere at an altitude of between 40 and 50 miles is called the D layer. Its ionization is low and it has little effect on the propagation of radio waves except for the absorption of energy from the radio waves as they pass through it. The D layer is present only during the day. Its presence greatly reduces the field intensity of transmissions that must pass through daylight zones.

The band of atmosphere at altitudes between 50 and 90 miles contains the so-called E layer. It is a well-defined band with greatest density at an altitude of about 70 miles. This layer is strongest during the daylight hours, and is also present but much weaker at night. The maximum density of the E layer appears at about noon local time.

The ionization of the E layer at the middle of the day is sometimes sufficiently intense to refract frequencies up to 20 mc back to the earth. This action is of great importance to daylight transmissions for distances up to 1,500 miles.

The F layer extends approximately from the 90-mile level to the upper limits of the ionosphere. At night only one F layer is present; but during the day, especially when the sun is high, this layer often separates into two parts, F_1 and F_2, as shown in figure 13-17. As a rule the F_2 layer is at its greatest density during early afternoon hours, but there are many notable exceptions of maximum F_2 density existing several hours later. Shortly after sunset the F_1 and F_2 layers recombine into a single F layer.

In addition to the layers of ionized atmosphere that appear regularly, erratic patches of ionized atmosphere occur at E-layer heights in the manner that clouds appear in the sky. These patches are referred to as sporadic-E ionizations. They are often present in sufficient number and intensity to enable good v-h-f radio transmission over distances not normally possible.

Sometimes sporadic ionizations appear in considerable strength at varying altitudes and actually prove harmful to radio transmissions.

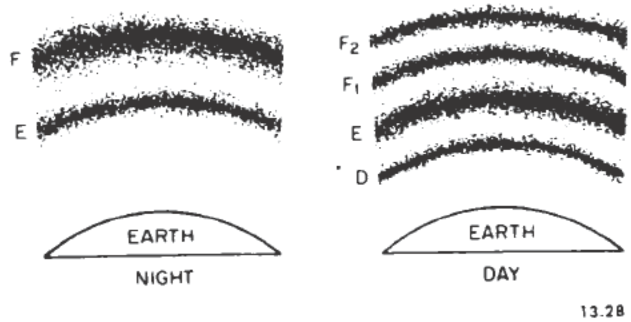


Figure 13-17.—E layer and F layer of the ionosphere.

EFFECT OF IONOSPHERE ON THE SKY WAVE

The ionosphere has many characteristics. Some waves penetrate and pass entirely through it into space, never to return. Other waves penetrate but bend. Generally, the ionosphere acts as conductor, and absorbs energy in varying amounts from the radio wave. The ionosphere also acts as a radio mirror and refracts (bends) the sky wave back to the earth, as illustrated in figure 13-18. Here, the ionosphere does by refraction what water does to a beam of light.

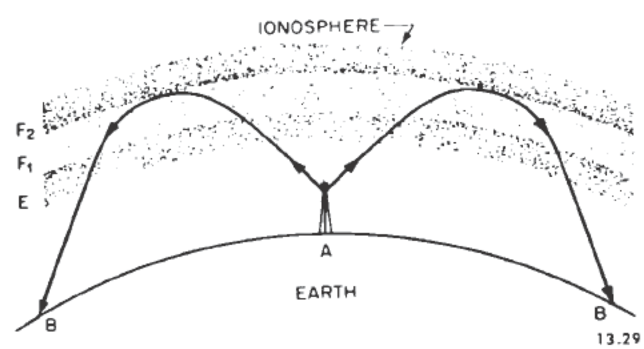
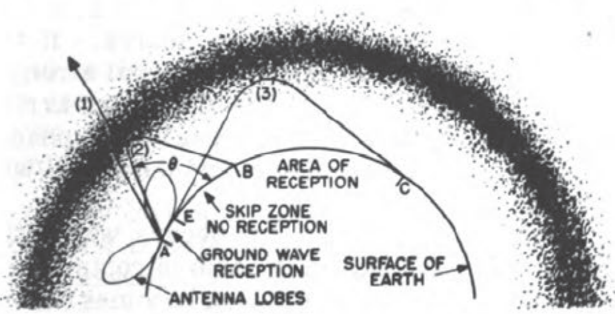


Figure 13-18.—Refraction of the sky waves by the ionosphere.

The ability of the ionosophere to return a radio wave to the earth depends upon the angle at which the sky wave strikes the ionosphere, the frequency of the transmission, and ion density.

When the wave from an antenna strikes the ionosphere at an angle the wave begins to bend. If the frequency is correct, (the ionosphere is sufficiently dense and the angle is proper), the wave will eventually emerge from the ionosphere and return to the earth. If a receiver is located at either of the points B in figure 13-18, the transmission from point A will be received. The antenna height in the figure is not drawn to scale. The tallest antennas are not over 1,000 feet in height. The sky wave in figure 13-19 is assumed to be composed of rays that emanate from the antenna in three distinct groups that are identified according to the angle of elevation. The angle at which the group 1 rays strike the ionosophere is too nearly vertical for the rays to be returned to the earth. The rays are bent out of line, but pass completely through the ionosophere and are lost.



13.30

Figure 13-19.—Effect of the angle of departure on the area of reception.

The angle made by the group 2 rays is called the CRITICAL ANGLE for that frequency. Any ray that leaves the antenna at an angle greater than this angle (θ) will penetrate the ionosphere.

Group 3 rays strike the ionosphere at the smallest angle that will be refracted and still return to the earth. At any smaller angle the rays will be refracted but will not return to the earth.

As the frequency increases, the critical angle decreases. Low-frequency fields can be projected straight upward and will be returned to the earth. The highest frequency that can be sent directly upward and still be returned to the earth is called the CRITICAL FREQUENCY. At sufficiently high frequencies, regardless of the angle at which the rays strike the ionosphere, the wave will not be returned to the earth. The critical frequency is not constant but varies from one locality to another, with the time of day, with the season of the year, and with the sunspot cycle.

Because of this variation in the critical frequency, nomograms and frequency tables are issued that predict the maximum usable frequency (muf) for every hour of the day for every locality in which transmissions are made.

Nomograms and frequency tables are prepared from data obtained experimentally from stations scattered all over the world. All this information is pooled, and the results are tabulated in the form of long-range predictions that remove most of the guess work from radio communications.

In the example in figure 13-19, the area between points B and C will receive the transmission by way of the refracted sky wave. The area between points A and E will receive the transmission by ground wave. All receivers located in the SKIP ZONE between points E and B will receive no transmissions from point A, because neither the sky wave nor the ground wave reaches this area.

EFFECT OF DAYLIGHT ON WAVE PROPAGATION

The increased ionization during the day is responsible for several important changes in sky wave transmission. It causes the sky wave to be returned to the earth nearer to the point of transmission. The extra ionization increases the absorption of energy from the sky wave; if the wave travels a sufficient distance into the ionosphere, it will lose all of its energy. The presence of the F_1 and E layers with the F_2 layer make long-range high-frequency communications possible, provided the correct frequencies are used.

Absorption usually reduces the effective daylight communication range of low-frequency and medium-frequency transmitters to surface wave ranges.

H-F LONG-RANGE COMMUNICATIONS

The high degree of ionization of the F_2 layer during the day, enabling refraction of high frequencies which are not greatly absorbed, has an important effect on transmissions of the high frequency band. Figure 13-20 shows how the F_2 layer completes the refraction and returns the transmissions of these frequencies to the earth, thereby making possible long-range high-frequency communication during the daylight hours.

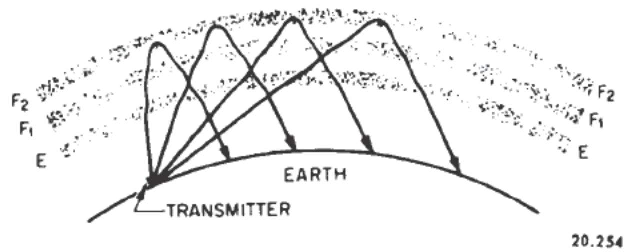


Figure 13-20.—Effect of the F2 layer on transmission of high frequency signals.

The waves are partially bent in going through the E layer and the F_1 layer, but are not returned to the earth until the F_2 layer completes the refraction process. V-h-f waves pass directly through the ionosphere.

The exact frequency to be used to communicate with another station depends upon the condition of the ionosphere and the distance between stations. As the ionosphere is constantly changing, the nomograms and frequency tables are used to select the correct frequency for the desired distance and for the time of day the transmission is to be made.

MULTIPLE REFRACTION

The radio wave may be refracted many times between the transmitter and receiver locations, as shown in figure 13-21. In this example the radio wave strikes the earth at location A, with sufficient intensity to be reflected back to the ionosphere and there to be refracted and returned to the earth a second time. Frequently a sky wave has sufficient energy to be refracted and reflected several times, greatly increasing the range of transmission. Because of this so-called multiple-hop transmission, transoceanic and around-the-world transmission is possible with moderate power.

FADING

Fading is a term used to describe the variations in signal strength that occur at the receiver during the time a signal is being received. There are several reasons for fading; some are easily understood while others are more complicated.

One cause is probably the direct result of interference between single-hop (fig. 13-18) and double-hop transmission (fig. 13-21) occurring simultaneously from the same source. If the two waves arrive in phase, the signal strength will be increased, but if the two waves arrive in phase opposition (180 degrees out-of-phase) they will cancel each other, and the signal will be weakened.

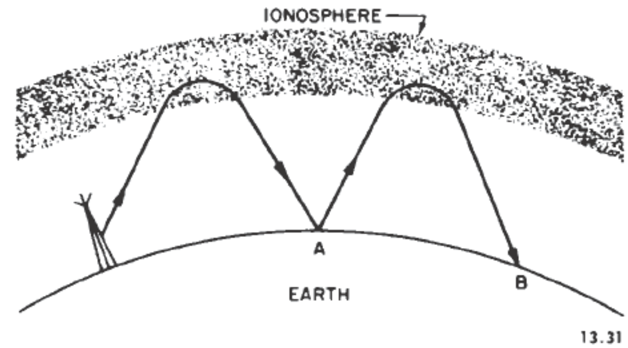


Figure 13-21.—Multiple refraction and reflection of a sky wave.

Interference fading also occurs where the ground wave and sky wave come in contact with each other. This type of fading becomes severe if the two waves are approximately equal in strength. Fluctuations in the sky wave with a steady ground wave can cause worse fading than sky wave transmission alone.

Variations in absorption and in the length of the path in the ionosphere are also responsible for fading. Occasionally, sudden disturbances in the ionosphere cause complete absorption of all sky wave radiation.

Receivers located near the outer edge of the ship zone are subjected to fading as the sky wave alternately strikes and skips over the area. This type of fading sometimes causes the received signal strength to fall to nearly the zero level.

FREQUENCY BLACKOUTS

Frequency blackouts are closely related to certain types of fading, some of which are severe enough to completely blank out the transmission.

Changing conditions in the ionosphere shortly before sunrise and shortly after sunset may cause complete blackouts at certain frequencies The higher frequency signals pass through the ionosphere while the lower frequency signals are absorbed by it.

Ionospheric storms (turbulent conditions in the ionosphere) often cause radio communication to become erratic. Some frequencies will be completely blacked out, while others may be reinforced. Sometimes these storms develop in a few minutes, and at other times they require as much as several hours to develop. A storm may last several days.

When frequency blackouts occur, the operator must be alert if he is not to lose contact with other ship or shore stations. In severe storms the critical frequencies are much lower, and the absorption in the lower layers of the ionosphere is much greater.

V-H-F AND U-H-F COMMUNICATION

In recent years there has been a trend toward the use of frequencies above 30 mc for short-range ship-to-ship and ship-to-airplane communications.

Early concepts suggested that these transmissions traveled in straight lines. This leads to the assumption that the u-h-f transmitter and receiver must be within sight of each other in order to supply radio contact.

Extensive use and additional research show the early line-of-sight theory to be frequently in error because radio waves of these frequencies may be refracted. The receiver does not always have to be in sight of the transmitter. Although this type of transmission still is called line-of-sight transmission, it is better to call it v-h-f and u-h-f transmission.

In general the v-h-f and u-h-f waves follow approximately straight lines, and large hills or mountains cast a radio shadow over these areas in the same way that they cast a shadow in the presence of light rays. A receiver located in a radio shadow will receive a weakened signal and in some cases, no signal at all. Theoretically, the range of contact is the distance to the horizon, and this distance is determined by the heights of the two antennas. However, as stated previously, communication is sometimes possible many hundreds of miles beyond the assumed horizon range. This fact must be observed when transmission is to occur under conditions of radio security.

EFFECT OF ATMOSPHERE ON H-F TRANSMISSIONS

Unusual ranges of v-h-f and u-h-f contacts are caused by abnormal atmospheric conditions a few miles above the earth. Normally, the warmest air is found near the surface of the water. The air gradually becomes cooler as the altitude increases. Sometimes unusual situations develop where warm layers of air are found above cooler layers. This condition is known as TEMPERATURE INVERSION.

When a temperature inversion exists, the amount of refraction (index of refraction) is different for the particles trapped within the boundaries from those outside them. These differences form channels or ducts that will conduct the radio waves many miles beyond the assumed normal range.

Sometimes these ducts are in contact with the water and may extend a few hundred feet into the air. At other times the duct will start at an elevation of between 500 and 1,000 feet and extend an additional 500 to 1,000 feet in the air.

If an antenna extends into the duct or if the wave enters a duct after leaving an antenna, the transmission may be conducted a long distance. An example of this type of transmission of radio waves in ducts formed by temperature inversions is shown in figure 13-22.

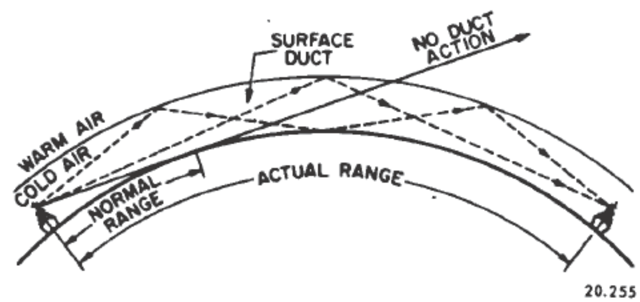


Figure 13-22.—Duct effect in high frequency transmission.

With certain exceptions, ducts are formed over water where the following conditions are observed aboard ship:

1. A wind is blowing from land.
2. There is a stratum of quiet air.
3. There are clear skies, little wind, and high barometric conditions.

4. A cool breeze is blowing over warm open ocean, especially in the tropic areas and in the trade-wind belt.

5. Smoke, haze, or dust fails to rise, but spreads out horizontally.

6. The moisture content of the air at the bridge is less than at the surface of the sea.

7. The temperature at the bridge is higher than at the surface of the water.

8. The received signal is fading rapidly.

GENERAL USE OF FREQUENCIES

Each frequency band has its own special uses. These uses depend upon the nature of the waves—surface, sky, or space, and the effect that the sun, the earth, the ionosphere, and the atmosphere of the earth have on them.

It is difficult to establish fixed rules for the choice of a frequency for a particular purpose. Some general statements can be made however as to which frequency bands are best suited to the type of transmission to be made.

For example, if a long-range communication is to get through to a distant receiver, high power and low frequency should be used. The large international communication systems and the fleet broadcast stations use this combination of high power and low frequency. (Fleet broadcast stations broadcast their messages without requiring a reply by the stations receiving the messages.) However, this combination requires an antenna array that may be too large for use with shipboard transmission. An alternative is to transmit the message to the nearest shore station for relay to its destination.

During daylight the sky wave builds up to a peak of usefulness in the h-f band. At night the peak of sky wave usefulness is in the top third of the m-f band. The usefulness of the ground, or surface wave declines steadily as the higher frequencies are reached until it is of no value in the h-f band. The only means of radio communication in the v-1-f band and for a limited range above v-1-f band is the space wave component of the ground wave.

Sky wave transmission (1,600 to 30,000 kc) is almost always associated with skip distances. Great range can be obtained, but in the process many receiving stations may be skipped in between the source and the most remote point of expected receptions. Thus, one of the stations to which it is desired to get the signal may be skipped if the receiving station is located within any one of the skip zones associated with the frequency being transmitted.

The most important frequencies for long-range transmission are those from 2,000 to 18,100 kc (2 to 18.1 mc). This band is standard for long-distance naval communications from ship to ship and ship to shore. It is the band that is used most frequently and the one that is covered by the standard Navy transmitters such as the AN/SRT-14, 15, and 16, and AN/URC-32. The band is in the short-wave region, and transmissions are accomplished by means of the sky wave and are affected by skip distances. When long range is desired in daytime, the frequencies that should be used are approximately from 7 mc to 18 mc. For night communications, frequencies below 10 or 15 mc should be used.

Nonregistered Publications Memoranda (DNC-14A), which are supplied to the various ships of the Navy, contain tables that show the best frequencies within the band for communication with various shore stations. These tables give the recommended frequency for every hour of the day for distances varying from 250 to 5,000 miles for some stations. The direction of the receiving station from the ship transmitting the signals is also taken into account. DNC-14A covers a 3-month period with a separate table for each month and for each major shore station. For specific and current information, consult these tables.

CHAPTER 14

ELEMENTARY COMMUNICATIONS RECEIVERS

INTRODUCTION

Many of the principles, circuits, and components discussed in previous chapters are directly applicable to radio receivers. A basic knowledge of them is therefore assumed in treating the material included in this chapter.

At the radio transmitter the carrier frequency is modulated by the desired signal, which may consist of coded characters, voice, music, or other types of signals. AMPLITUDE MODULATION (a-m) occurs if the signals cause the amplitude of the carrier to vary. FREQUENCY MODULATION (f-m) occurs if the signals cause the frequency of the carrier, or center frequency, to vary. Although there are other types of modulation, only a-m and f-m receivers will be treated in this chapter.

The r-f carrier wave with the modulating signal impressed upon it is transmitted through space as an electromagnetic wave to the antenna of the receiver. As the wave passes across the receiving antenna, small a-c voltages are induced in the antenna. These voltages are coupled into the receiver via the antenna coupling coil. The function of the receiver is to select the desired carrier frequency from those present in the antenna circuit and to amplify the small a-c signal voltage. The receiver then removes the carrier by the process of detection (rectification and the removal of the r-f component) and amplifies the resultant audio signal to the proper magnitude to operate the loudspeaker or earphones.

Three major types of radio receivers are reviewed in this chapter—the TUNED-RADIO-FREQUENCY (t-r-f) receiver, the SUPERHETERODYNE receiver, and the FREQUENCY-MODULATION (f-m) receiver. The a-m and f-m superheterodyne receivers are often combined in one set, because many of the circuit components are common to both types of receivers.

T-R-F RECEIVERS

OPERATING PRINCIPLE

The tuned-radio-frequency receiver, generally known as the t-r-f receiver, consists of one or more r-f stages, a detector stage, one or more a-f stages, a reproducer, and the necessary power supply. A block diagram of a t-r-f receiver is shown in figure 14-1. The waveforms that appear in the respective sections of the receiver are shown below the block diagram.

The amplitude of the a-m signal at the input of the receiver is relatively small because it has been attenuated in the space between the transmitter and the receiver. It is composed of the carrier frequency and the modulation envelope. The r-f amplifier stages amplify the waveform, but they do not change its basic shape if the circuits are operating properly. The detector rectifies and removes the r-f component of the signal. The output of the detector is a weak signal made up only of the modulation component, or envelope, of the incoming signal. The a-f amplifier stages following the detector increase the amplitude of the a-f signal to a value sufficient to operate the loudspeaker or earphones.

COMPONENTS

R-F Section

The ANTENNA-GROUND SYSTEM serves to introduce the desired signal into the first r-f amplifier stage via the antenna coupling transformer. For best reception the resistance of the antenna-ground system should be low. The antenna should also be of the proper length for the band of frequencies to be received; and the antenna impedance should match the input impedance of the receiver. The gain of most commercial receivers, however, is

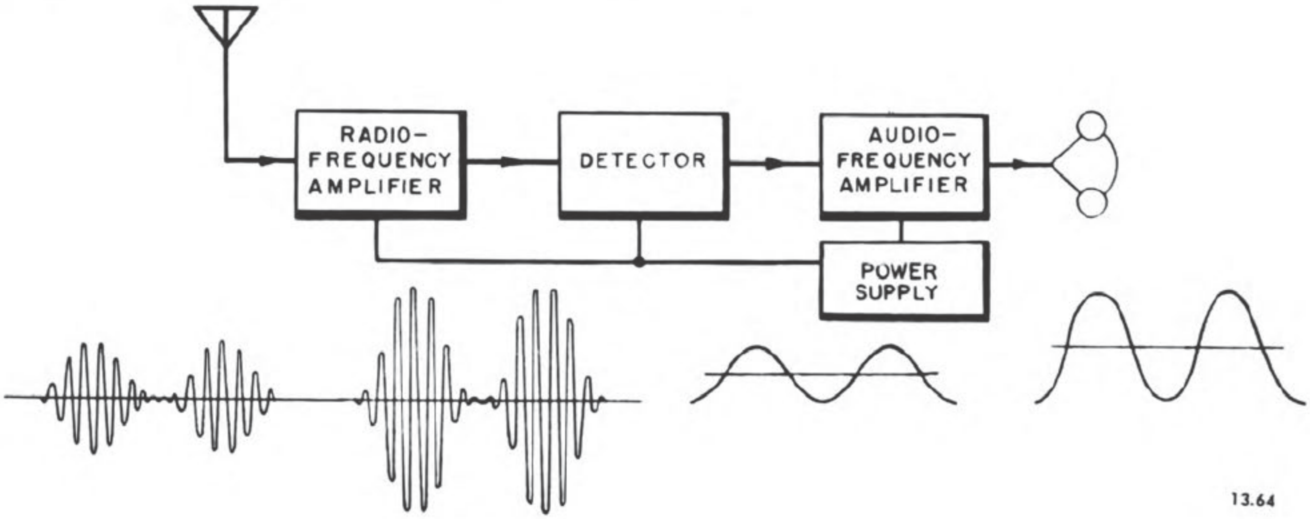


Figure 14-1.—Block diagram of a t-r-f receiver and waveforms.

generally sufficient to make these values non-critical.

The R-F AMPLIFIERS in the t-r-f receiver have tunable tanks in the grid circuits. Thus, the receiver may be tuned so that only one r-f signal within its tuning range is selected for amplification. When the tank is tuned to the desired frequency, it resonates and produces a relatively large circulating current. The grid of the r-f amplifier then receives a relatively large signal voltage at the resonant frequency, and minimum signals at other frequencies.

The relative ability of a receiver to select one particular frequency and to reject all others is called the SELECTIVITY of the receiver. The relative ability of the receiver to amplify small signal voltages is called the SENSITIVITY of the receiver. Both of these values may be improved by increasing the number of r-f stages. When this is done, the tuning capacitors in the grid tank circuits are usually ganged on the same shaft and trimmers are added in parallel with each capacitor to make the stages track at the same frequency. In addition, the outer plates of the rotor sections of the capacitors are sometimes slotted to enable more precise alignment throughout the tuning range.

Tetrodes or pentodes are generally used in r-f amplifiers because, unlike triodes, they do not usually require neutralization. They also have higher gain than triodes.

A typical r-f amplifier stage employing a pentode is shown in figure 14-2. Tuned circuit L2C1 is inductively coupled to L1, the antenna coil. R1 and C3 provide operating bias for the tube. C4 and R2 are the screen bypass capacitor and dropping resistor, respectively. The tuned circuit, L4C6, couples the following stage inductively to L3. Both transformers are of the air-core type. The dotted lines indicate mechanical ganging of C1 and C6 on the same shaft. The tuning capacitor in the next stage is also ganged on the same shaft.

If it is desired to cover more than one frequency range, additional coils having the proper inductance are used. They are sometimes of the plug-in variety, but more generally they are mounted on the receiver and their leads are connected to a multicontact rotary switch. The latter method is preferable for BAND SWITCHING because the desired band

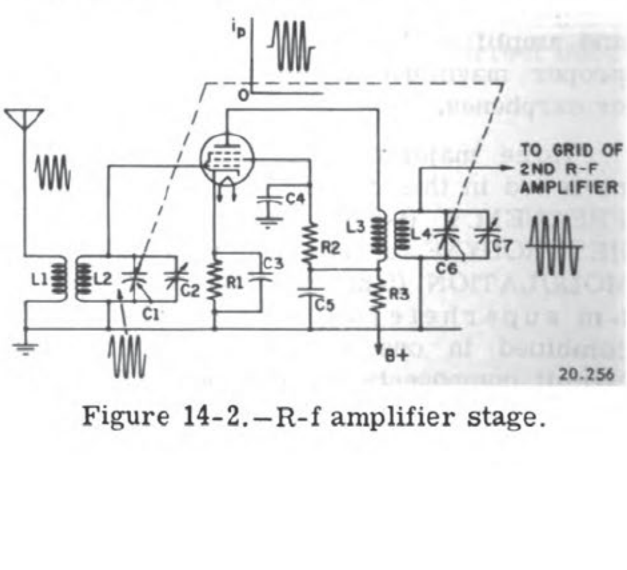


Figure 14-2.—R-f amplifier stage.

can be selected simply by turning the switch. The same tuning capacitor is used for each band. However, when band switching is employed, the trimmers are connected across the individual tuning inductors and not across the main tuning capacitors.

Decoupling circuits are designed for both r-f and a-f amplifiers to counteract feedback. Thus, in the r-f amplifier in figure 14-2, C5 and R3 make up the decoupling circuit. R3 offers a high impedance to the signal current, but C5 offers a low impedance. Consequently the signal current is shunted to ground around the B supply. Since R3 also offers a high resistance to d-c current, it may be replaced by a choke coil having a high impedance only to the signal current. Each stage is similarly equipped with a decoupling circuit.

A mechanical or an electrical bandspread may be used as an aid in separating stations that are crowded together on the tuning dial. MECHANICAL BANDSPREAD is simply a micrometer arrangement to reduce the motion of the capacitor rotor as the tuning knob is turned. When ELECTRICAL BANDSPREAD is used a small variable capacitor is connected in parallel with the tuning capacitor. Because of its small size, this variable capacitor may be moved a considerable amount before it causes an appreciable change in the frequency of the tuned circuit. If the tuning capacitors are ganged, the bandspread capacitors are also ganged.

Detector

The process of removing the intelligence component of the modulated wave form from the r-f carrier is called DETECTION or DEMODULATION. In the a-m system the audio or intelligence component causes both the positive and the negative half cycles of the r-f wave to vary in amplitude. The function of the detector is to rectify the modulated signal. A suitable filter eliminates the remaining r-f pulses and passes the audio component on to the a-f amplifiers.

Details of the various methods of detection are treated in chapter 10. Each of the several methods that might be used in the t-r-f receiver have certain inherent weaknesses. For example, the diode detector requires several stages of amplification ahead of the detector. It loads its tuned input circuit, and therefore the sensitivity and selectivity of the circuit are reduced. However, it can handle strong signals without overloading, and its linearity is good.

The grid-leak detector is sensitive (and therefore requires fewer stages of amplification), but it has poor linearity and selectivity and it may be overloaded on strong signals.

The regenerative detector is very sensitive and has excellent selectivity, but it has poor linearity and easily overloads on strong signals.

The circuit shown in figure 14-3 employs plate detection. It has medium sensitivity and the ability to handle strong signals without overloading. The selectivity of this circuit is excellent, but because the i_p-e_g graph is curved near the cutoff point (where the plate detector operates) some distortion in the output cannot be avoided.

In figure 14-3, the tube is biased nearly to cutoff by the average plate current that flows through R1. This average value increases as the signal strength increases. On positive half cycles of the incoming signals the plate current varies with the amplitude of the modulating wave and produces the desired a-f output voltage. On negative half cycles no appreciable plate current flows. Between positive half cycles the bias voltage is held constant across R1 by the action of C3, because the time constant of R1C3 is long compared with the time for the lowest a-f cycle.

The r-f pulses are filtered out by means of a low-pass filter (consisting of C4, L2, and C5), which rejects the r-f component and passes the a-f component. C6 couples the a-f component to the first audio amplifier. R2 is the plate load resistor, and the combination R3C7 makes up the decoupling circuit.

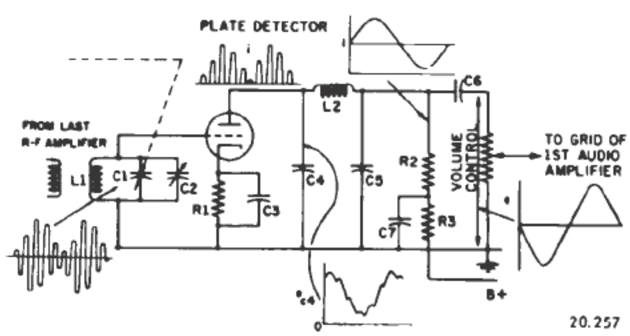


Figure 14-3.—Plate detector circuit.

A-F Section

The function of the a-f section of a receiver is to further amplify the audio signal, which is commonly fed via the volume control to the grid of the first audio amplifier tube. In most cases the amount of amplification that is necessary depends on the type of reproducer used. If the reproducer consists of earphones, only one stage of amplification may be necessary. If the reproducer is a large speaker or other mechanical device requiring a large amount of power, several stages may be necessary. In most receivers the last a-f stage is operated as a power amplifier.

A necessary part of the a-f section is some means of manual control of the output signal level of the receiver.

A MANUAL VOLUME CONTROL may be employed in a number of receiver circuits. Normally this control varies the amplitude of the signal applied to the grid of an amplifier tube, as shown in figure 14-3. Increasing the resistance between ground and the sliding contact increases the amplitude of the signal applied to the grid of the driven stage.

An A-F OUTPUT STAGE is shown in figure 14-4. C1 couples the first a-f amplifier to the output stage, and R1 is the grid coupling resistor. R2 and C2 provide a steady bias. Because of the low frequencies involved, C2 should have a larger value of capacitance than similar bypass capacitors in the r-f section.

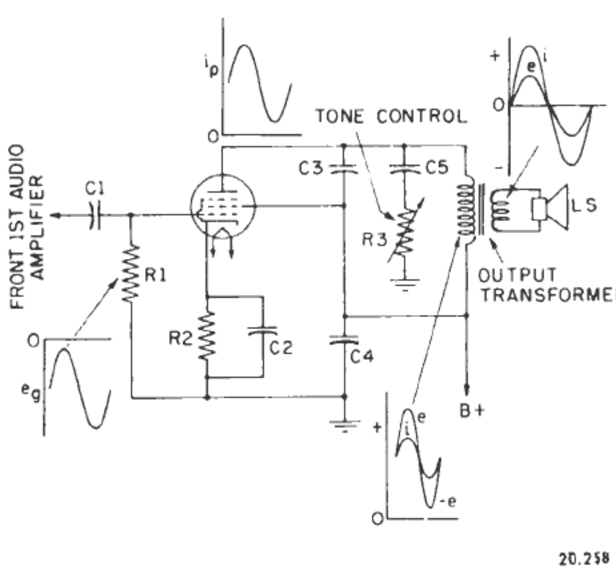


Figure 14-4.—Audio amplifier output stage.

C4 is the plate-bypass, or decoupling capacitor. C3 has a small value of capacitance and bypasses some of the higher frequencies around the output transformer, thus emphasizing the bass. The impedance of the output-transformer primary commonly represents a compromise between maximum power transfer and minimum distortion. The impedance of the secondary is chosen to match the impedance of the voice coil. Some secondaries have taps on the windings to permit an impedance match to a variety of voice-coil impedances.

TONE CONTROL may be used in communications receivers. The purpose of tone control is to emphasize either the low or the high frequencies, by shunting the undesired frequencies around the remainder of the circuit components in the audio section. A simple tone-control circuit, such as the series capacitor C5 and variable resistor R3 combination shown in figure 14-4, may be connected between plate and ground or between grid and ground in any of the audio stages of a receiver. In this figure it is connected between plate and ground. The value of the series capacitor is such that it will bypass to ground the high-frequency components. The amount of high-frequency energy removed by the tone-control circuit is determined by the setting of the variable-resistor control arm. When the resistance is low, the high frequencies are attenuated; when it is high they appear in the output.

Feedback voltage from output to input is sometimes developed across the impedance of the common power supply. For frequencies within the usable audio range, this impedance is sufficiently low so that insufficient feedback is obtained to cause oscillation. However, for extremely low frequencies, the capacitors in the power supply will sometimes have enough impedance to cause oscillation.

MOTORBOATING IN AUDIO STAGES.—When two or more audio amplifier stages are supplied from a common B supply, feedback occurs as a result of common coupling between the plate circuits, and some method of decoupling must be employed. The coupling consists of the internal impedance of the source of plate voltage. The feedback may either increase or decrease the amplification depending on the phase relation between the input voltage and the feedback voltage. In a multistage amplifier the greatest transfer of feedback energy occurs between the final and first stages because of the high amplification through the multistage amplifier.

The effects of feedback are important if the feedback voltage coupled into the plate circuit of the first stage is appreciable compared to the signal voltage that would be developed if feedback did not exist. For example, a three-stage resistance-coupled amplifier may develop a feedback voltage (coupled via the B supply into the plate circuit of stage 1) which is in phase with the signal voltage of stage 1 and hence may cause oscillations to be set up. In audio amplifiers having high gain and a good low-frequency response this regeneration causes a low-frequency oscillation known as "motorboating" because of the "putt-putt" sound in the speaker.

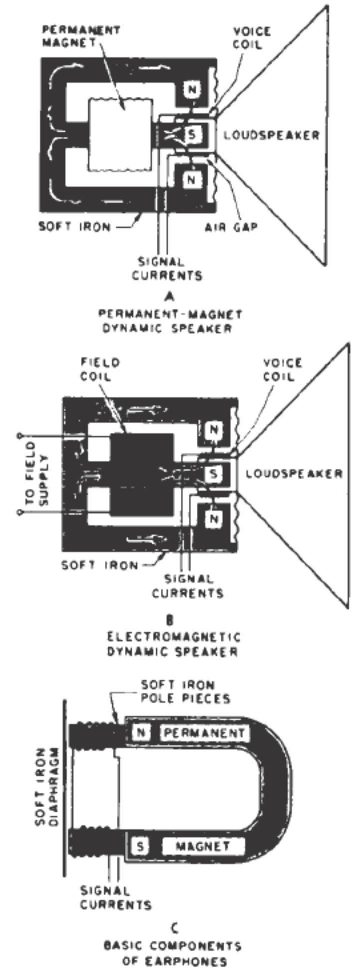
Design engineers usually decouple plate circuits by adding a series resistor to the input stage, between its plate load and B+, and bypassing that resistor to ground. The appearance of motorboating reveals the need of replacing either the decoupling resistor or its bypass capacitor.

Loudspeakers and Earphones

Figure 14-5 shows three types of audio reproducers.

In the permanent-magnet dynamic type of reproducer (fig. 14-5,A) a strong field is established between the pole pieces by means of a powerful permanent magnet. The flux is concentrated in the air gap between a permeable soft-iron core and an external yoke. The voice coil is mounted in the air gap. When a-c signal currents flow in the coil a force proportional to the strength of the current is applied to the coil, and the coil is moved axially in accordance with the a-c signal. The loudspeaker diaphragm is attached to the voice coil and moves in accordance with the signal currents, thus setting up sound waves in the air. The corrugated diaphragm to which the speaker cone is attached keeps the cone in place and properly centered.

As in figure 14-5,B, an electromagnet may be used in place of the permanent magnet to form an electromagnetic dynamic speaker. However, in this instance sufficient d-c power



20.259

Figure 14-5.—Types of audio reproducers.

must be available to energize the field. The operation is otherwise much the same as that of the permanent-magnet type.

The basic components of earphones are shown in figure 14-5,C. When no signal currents are present, the permanent magnet exerts a steady pull on the soft-iron diaphragm. Signal current flowing through the coils mounted on the soft-iron pole pieces develops a magnetomotive force that either adds to or subtracts from the field of the permanent magnet. The diaphragm thus moves in or out according to the resultant field. Sound waves will then be reproduced that have amplitude and frequency (within the capability of the reproducer) similar to the amplitude and frequency of the signal currents.

CIRCUIT OF THE T-R-F RECEIVER

The complete circuit of a t-r-f radio receiver operated from an a-c power supply is shown in figure 14-6. The receiver uses two pentodes in the r-f section, one triode operated as a plate detector, and two pentode a-f amplifier stages that feed the loudspeaker.

From previous discussions, the various circuits may be identified and the signal may be traced from the antenna-ground system to the loudspeaker. The dotted lines indicate that the three main tuning capacitors are ganged on a single shaft. Across each of the main tuning capacitors is connected a trimmer capacitor to enable circuit alignment. The ground circuit and the various decoupling circuits may be readily identified. The power supply voltage is obtained from a conventional full-wave rectifier. Rectifier and tube filament currents are obtained from two low-voltage windings on the power transformer.

CHARACTERISTICS OF THE T-R-F RECEIVER

The principal disadvantage of the t-r-f receiver is that its selectivity, or its ability

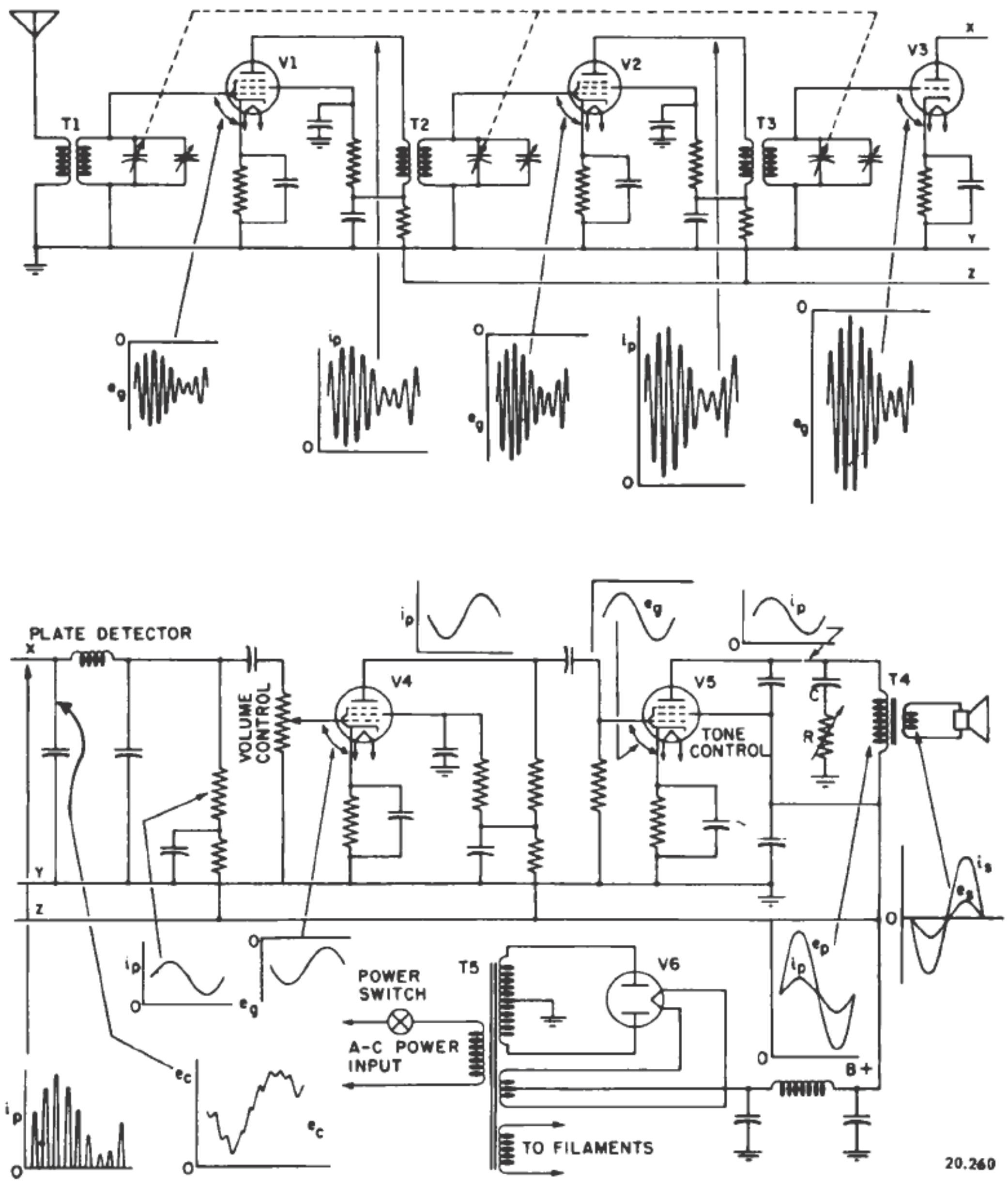


Figure 14-6.—Circuit of a t-r-f receiver.

to separate signals, does not remain constant over its tuning range. As the set is tuned from the low-frequency end of its tuning range to the high frequency end, its selectivity decreases.

Also, the amplification, or gain, of a t-r-f receiver is not constant over the tuning range of the receiver. The gain depends on r-f transformer gain, which increases with frequency. In order to improve the gain at the low-frequency end of the band, r-f transformers employing high-impedance (untuned) primaries are designed so that the primary inductance will resonate with the primary distributed capacitance at some frequency slightly below the low end of the tunable band. Thus, the gain is good at the low end of the band because of the resonant buildup of primary current. The near-resonant condition of the primary at the low end more than offsets the effect of reduced transformer action. However, the shunting action of the primary distributed capacitance lowers the gain at the high-frequency end of the band. To make up for the resultant poor gain at the high end of the band, a small capacitor is connected between the plate and grid leads of adjacent r-f stages to supplement the transformer coupling. At the low end of the band the capacitive coupling is negligible.

The superheterodyne receiver has been developed to overcome many of the disadvantages of the t-r-f receiver.

SUPERHETERODYNE RECEIVERS

OPERATING PRINCIPLE

The essential difference between the t-r-f receiver and the superheterodyne receiver is that in the former the r-f amplifiers preceding the detector are tunable over a band of frequencies; whereas in the latter the corresponding amplifiers are tuned to one fixed frequency called the INTERMEDIATE FREQUENCY (i-f). The principle of frequency conversion by heterodyne action is here employed to convert any desired station frequency within the receiver range to this intermediate frequency. Thus an incoming signal is converted to the fixed intermediate frequency before detecting the audio signal component, and the i-f amplifier operates under uniformly optimum conditions throughout the receiver range. The i-f circuits thus may be made uniformly selective, uniformly high in voltage gain, and uniformly of satisfactory bandwidth to contain all of the desired sideband components associated with the amplitude-modulated carrier.

The block diagram of a typical superheterodyne receiver is shown in figure 14-7. Below corresponding sections of the receiver are shown the waveforms of the signal at that point. The r-f signal from the antenna passes first through an r-f amplifier (preselector) where the amplitude of the signal is increased. A locally generated unmodulated r-f signal of

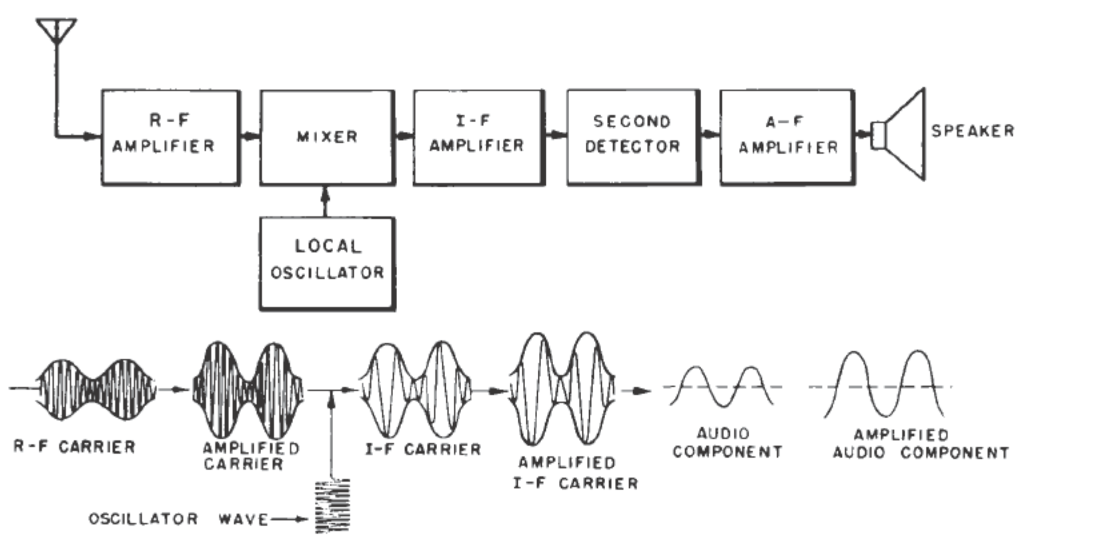


Figure 14-7.—Block diagram of a superheterodyne receiver and waveforms.

constant amplitude is then mixed with the carrier frequency in the mixer stage. The mixing or heterodyning of these two frequencies produces an intermediate-frequency signal which contains all of the modulation characteristics of the original signal. The intermediate frequency is equal to the difference between the station frequency and the oscillator frequency associated with the heterodyne mixer. The intermediate frequency is then amplified in one or more stages called INTERMEDIATE-FREQUENCY (i-f) AMPLIFIERS and fed to a conventional detector for recovery of the audio signal.

The detected signal is amplified in the a-f section and then fed to a headset or loudspeaker. The detector, the a-f section, and the reproducer of a superheterodyne receiver are basically the same as those in a t-r-f set, except that diode detection is generally used in the superheterodyne receiver. Automatic volume control or automatic gain control also is commonly employed in the superheterodyne receiver.

R-F AMPLIFIER

If an r-f amplifier is used ahead of the mixer stage of a superheterodyne receiver it is generally of conventional design. Besides amplifying the r-f signal, the r-f amplifier has other important functions. For example, it isolates the local oscillator from the antenna-ground system. If the antenna were connected directly to the mixer stage, a part of the local oscillator signal might be radiated into space. This signal could be picked up by a sensitive direction finder on any enemy ship. For this reason and others, Navy superheterodyne receivers are provided with at least one r-f amplifier stage.

Also, if the mixer stage were connected directly to the antenna, unwanted signals, called IMAGES, might be received, because the mixer stage produces the intermediate frequency by heterodyning two signals whose frequency difference equals the intermediate frequency. (The heterodyne principle is treated later in this chapter.)

The image frequency always differs from the desired station frequency by twice the intermediate frequency—Image frequency = station frequency ± (2 X intermediate frequency). The image frequency is higher than the station frequency if the local oscillator frequency tracks (operates) above the station frequency (fig. 14-8,A). The image frequency is lower than the station frequency if the local oscillator tracks below the station frequency (fig. 14-8,B). The latter arrangement is generally used for the higher frequency bands, and the former, for the lower frequency bands.

For example, if such a receiver having an intermediate frequency of 455 kc is tuned to receive a station frequency of 1500 kc (fig. 14-8,A), and the local oscillator has a frequency of 1955 kc, the output of the i-f amplifier may contain two interfering signals—one from the 1500-kc station and the other from an image station of 2410 kc (1500 + 2 X 455 = 2410 kc). The same receiver tuned near the low end of the band to a 590-kc station has a local oscillator frequency of 1045 kc.

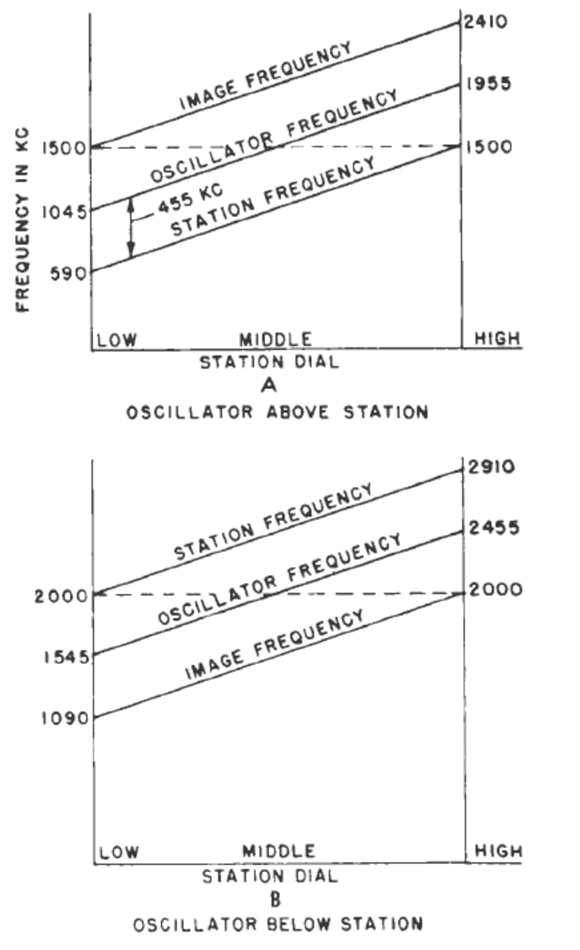


Figure 14-8.—Relation of image frequency to station frequency in a superheterodyne receiver.

The output of the i-f amplifier contains the station signal (1045-590 = 455 kc) and an image signal (1500-1045 455 kc). Thus the 1500-kc signal is an image heard simultaneously with the 590-kc station signal.

It may also be possible for ANY two signals having sufficient strength, and separated by the intermediate frequency to produce unwanted signals in the reproducer. The selectivity of the preselector tends to reduce the strength of these images and unwanted signals.

The ratio of the amplitude of the desired station signal to that of the image is called the IMAGE REJECTION RATIO and is an important characteristic of a superheterodyne receiver. Better superheterodyne receivers are therefore equipped with one or more preselector stages, a typical example of which is shown in figure 14-9.

The preselector stage employs a variable-mu tube and cathode bias. L1 is the antenna coil, L2 and C1 make up the tuned input circuit, and C2 is the trimmer used for alignment purposes. The dotted line indicates ganged tuning capacitors. Usually these are the tuning capacitor of the mixer input tank circuit and the local oscillator tuning capacitor. C3 provides low impedance coupling between the lower end of L2 and the grounded end of C2, thus bypassing the decoupling filters in the automatic-volume-control (a-v-c) circuit. (Automatic-volume, or automatic-gain, control is treated later in this chapter.) The r-f transformer in the output circuit consists of an untuned high-impedance primary, L3, and a tuned secondary, L4, which resonates with tuning capacitor C5 at the station frequency. R-f bypass capacitor C6 serves a function similar to that of C3.

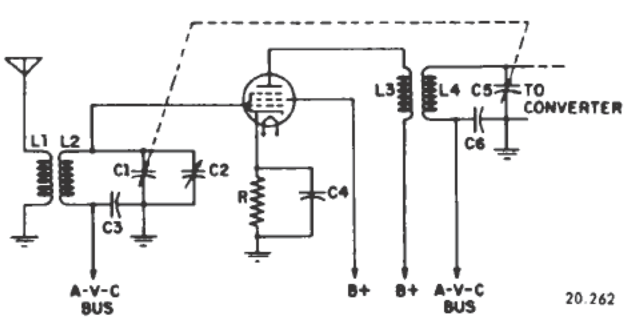


Figure 14-9.—Typical superheterodyne preselector stage.

First Detector

The first detector, or frequency-converter, section of a superheterodyne receiver is composed of two parts—the oscillator and the mixer. In many receivers, particularly at broadcast frequencies, the same vacuum tube serves both functions, as in the pentagrid converter shown in figure 14-10. The operation of the tube may be simplified somewhat if both stages (oscillator and mixer) are considered as exerting two different influences on the stream of electrons from cathode to plate. These electrons are influenced by the oscillator stage (grids, 1, 2, and 4) and also by the station input signal on grid number 3. Thus, coupling between the input signal and the oscillator takes place within the electron stream itself.

There is a tendency for the local oscillator to synchronize with the station frequency signal applied to grid 3. At high frequencies where the two signals have nearly the same frequency, the pentagrid converter is replaced with a mixer tube and a separate oscillator tube. This type of circuit provides frequency stability for the local oscillator.

The oscillator stage employs a typical Hartley circuit in which C5 and the oscillator coil make up the tuned circuit. C4 is the trimmer capacitor which is used for alignment (tracking) purposes. C3 and R2 provide grid-leak bias for the oscillator section of the tube. Grid 1 is the oscillator grid, and grids 2 and 4 serve as the oscillator plate. Grids 2 and 4 are connected together and also serve as a shield for the signal input grid, 3.

Grid 3 has a variable-mu characteristic, and serves as both an amplifier and a mixer grid. The tuned input is made up of L1 and C1, with the parallel trimmer C2. The dotted lines

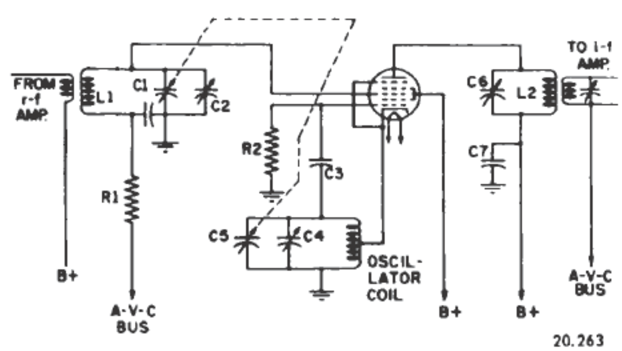


Figure 14-10.—First detector employing a pentagrid converter.

drawn through C1 and C5 indicate that both of these capacitors are ganged on the same shaft (in this example with the preselector tuning capacitor). The plate circuit contains the station frequency and the oscillator frequency signals both of which are bypassed to ground through the low reactance of C6 and C7. The heterodyne action within the pentagrid converter produces additional frequency components in the plate circuit, one of which is the difference frequency between the oscillator and the station frequency. The difference frequency is the intermediate frequency and is developed across C6 and L2. This signal is coupled to the first i-f amplifier grid through the desired band-pass coupling which is wide enough to include the sideband components associated with the amplitude-modulated signal applied to grid 3 of the pentagrid converter.

The conversion gain in a pentagrid converter is,

$$\mu = V_d S_c,$$

where V_d is the a-c plate resistance with the station r-f carrier applied, and S_c is the conversion transconductance (30% to 40% of the g_m of the pentode amplifier). Conversion gain is the change in plate voltage at the intermediate frequency divided by the change in grid voltage at the r-f station frequency for equal changes in plate current at the intermediate frequency. Expressed as a formula,

$$\text{conversion gain} = \frac{\text{i-f output volts}}{\text{r-f input volts}}.$$

The conversion gain of a typical pentagrid converter used in broadcast receivers ranges between 30 and 80.

Heterodyne Principle

The production of audible beat notes is a phenomenon that is easily demonstrated. For example, if two adjoining piano keys are struck simultaneously, a tone will be produced that rises and decreases in intensity at regular intervals. This action results from the fact that the rarefactions and compressions produced by the vibrating strings will gradually approach a condition in which they reinforce each other at regular intervals of time with an accompanying increase in the intensity of the sound. Likewise at equal intervals of time, the compressions and rarefactions gradually approach a condition in which they counteract each other, and the intensity is periodically reduced.

This addition and subtraction of the intensities at regular intervals produces BEAT FREQUENCIES. The number of beats produced per second is equal to the difference between the two frequencies.

The production of beats in a superheterodyne receiver is somewhat analogous to the action of the piano, except that with the receiver the process is electrical and the frequencies are much higher. Figure 14-11 indicates graphically how the beat frequency (intermediate frequency) is produced when signals of two different frequencies are combined in the mixer tube. The resultant envelope varies in amplitude at the difference frequency, as indicated by the dotted lines.

In this example, one voltage, e_s, has a frequency of 8 cycles per second and the other voltage, e_o, has a frequency of 10 cycles per second. Initially, the amplitudes of the two voltages add at instant A, but at instant B the relative phase of e_o has advanced enough to

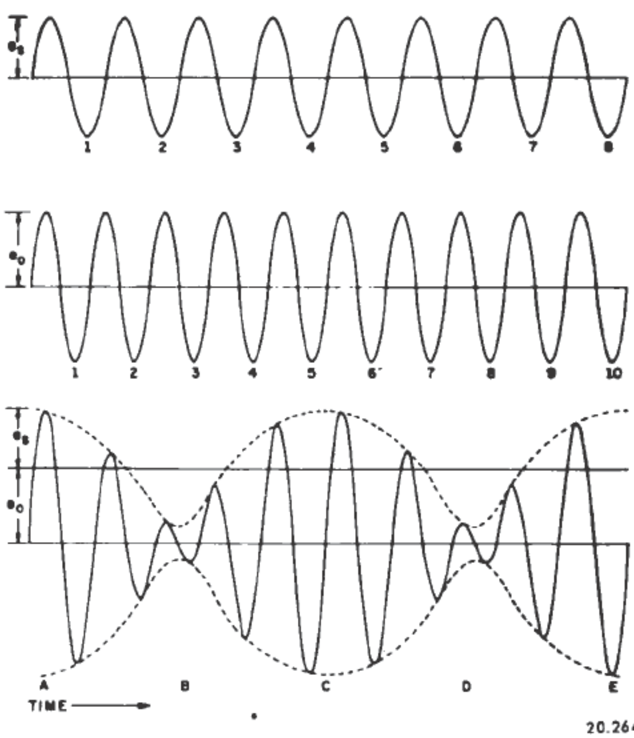


Figure 14-11.—Simplified graphical analysis of the formation of beats.

oppose e_s, and the amplitude of the resultant envelope is reduced to a value dependent upon e_s. At instant C the relative phase of e_o has advanced enough to permit the amplitudes to add again. Thus, 1 cycle of amplitude variation of the envelope takes place in the time interval that e_o needs to gain 1 cycle over e_s. From figure 14-11 it may be seen that e_o gains 2 cycles in the interval A to E. Therefore, the beat or difference frequency is 2 cycles per second. In the superheterodyne receiver the amplitude of the oscillator signal is designed to be greater than that of any received signal.

I-f Amplifier

The i-f amplifier is a high-gain circuit commonly employing pentode tubes. This amplifier is permanently tuned to the frequency difference between the local oscillator and the incoming r-f signal. Pentode tubes are generally employed, with one, two, or three stages, depending on the amount of gain needed. As previously stated, all incoming signals are converted to the same frequency by the frequency converter, and the i-f amplifier operates at only one frequency. The tuned circuits, therefore, are permanently adjusted for maximum gain consistent with the desired band pass and frequency response. These stages operate as class-A voltage amplifiers and practically all of the selectivity of the superheterodyne receiver is developed by them.

Figure 14-12 shows the first i-f amplifier stage. The minimum bias is established by means of R1C1, and automatic volume control is applied to the grid through the secondary of the preceding coupling transformer.

The output i-f transformer, which couples the plate circuit of this stage to the grid circuit of the second i-f stage, is tuned by means of capacitors

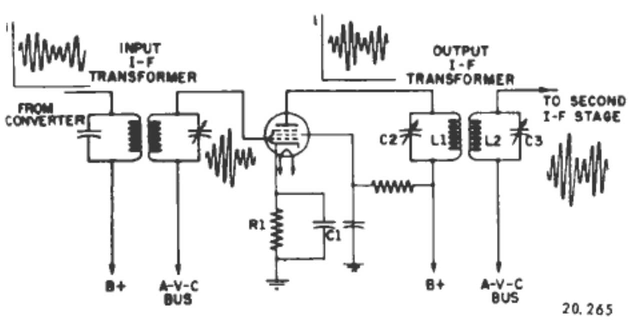


Figure 14-12.—First i-f amplifier stage.

C2 and C3. Mica or air-trimmer capacitors may be used. In some instances the capacitors are fixed, and the tuning is accomplished by means of a movable powdered-iron core. This method is called PERMEABILITY tuning. In special cases the secondary only is tuned. The coils and capacitors are mounted in small metal cans which serve as shields, and provision is made for adjusting the tuning without removing the shield.

The input i-f transformer has a lower coefficient of coupling than the output transformer in some receivers in order to suppress noise from the pentagrid converter. The output i-f transformer is slightly overcoupled with double humps appearing at the upper and lower sideband frequencies. The overall response of the stage is essentially flat, and in typical broadcast receivers has a voltage gain of about 200 with a bandpass of 7 to 10 kc and an i-f of about 456 kc.

The chief characteristic of the double-tuned bandpass coupling is that at frequencies slightly above and slightly below the intermediate frequency the impedance coupled into the primary by the presence of the secondary is reactive. This cancels some of the reactance existing in the primary, and the primary current increases. Thus the output voltage of the secondary does not fall off and the response is uniform within the pass band. The double-tuned i-f amplifier is discussed in detail in chapter 6.

Crystal Filter

A quartz crystal, used as a selective filter in the i-f section of a communications receiver, is one of the most effective methods of achieving maximum selectivity. It is especially useful when the channel is crowded and considerable noise (both external and internal) is present. The crystal acts as a high-Q tuned circuit, which is many times more selective than tuned circuits consisting of inductors and capacitors. The crystal dimensions are so chosen that the crystal will be in resonance at the desired intermediate frequency.

One of the simplest of a number of possible circuit arrangements is shown in figure 14-13. The crystal is in one arm of a bridge circuit. The secondary of the input transformer is balanced to ground through the center-tap connection. The phasing capacitor, C4, is in another arm of the bridge circuit. The crystal acts as

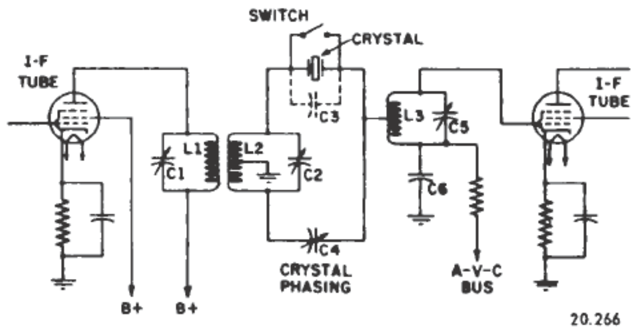


Figure 14-13.—Crystal filter used in the i-f section of a superheterodyne receiver.

a high-Q series-resonant circuit and allows signals within the immediate vicinity of resonance to pass through the crystal to the output coil, L3. The desired signal appears between the center tap of L3 and ground.

The capacity between the crystal holder plates may bypass unwanted signals around the crystal. Therefore, some method must be provided to balance out this capacitance.

In this circuit, balancing is accomplished by taking a voltage 180° out of phase with the instantaneous voltage across the crystal and applying it via C4 in such a way as to neutralize the undesired signal voltage. The balanced input circuit in this case is obtained by the use of a center-tapped inductor. The tap on L3 permits the proper impedance match.

Second Detector

Most superheterodyne receivers employ a diode as the second detector. This type of detector is practical because of the high gain as well as the high selectivity of the i-f stages. The diode detector has good linearity and can handle large signals without overloading. For reasons of space and economy, the diode detector and first audio amplifier are often included in the same envelope in modern superheterodyne receivers.

A simple diode detector is shown in figure 14-14. The rectified voltage appears across R1, which also serves as the volume-control potentiometer. Capacitor C2 bypasses the r-f component to ground, and C3 couples the output of the detector to the first audio amplifier stage. The tuned circuit L2C1 is the secondary of the last i-f transformer.

The time constant of R1C2 is long compared to the time for one i-f cycle but short compared

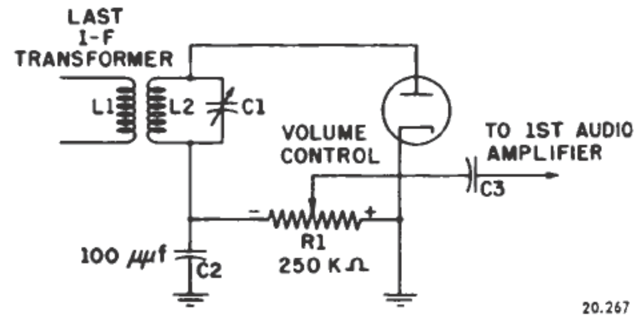


Figure 14-14.—Diode detector.

to the time for one a-f cycle. If the intermediate frequency is 456 kc the time for one i-f cycle in microseconds is

$$\frac{1}{0.456} = 2.19\ \mu s.$$

If R1 is 250 k-ohms and C2 is 100$\mu\mu$f the time constant in microseconds is

$$0.25 \times 100 = 25\ \mu s.$$

The demodulation capacitor, C2, discharges through R1 in one-half the time for one a-f cycle $\left(\frac{1}{2f}\right)$. The time required to discharge C2 is 5R1C2 seconds. Thus,

$$\frac{1}{2f} = 5R1C2$$

$$f = \frac{1}{10R1C2}$$

$$= \frac{1}{10 \times 0.250 \times 10^6 \times 100 \times 10^{-12}}$$

$$= \frac{10^6}{10_3 \times 0.250} = 4{,}000 \text{ cps.}$$

Thus, the highest audio frequency which C2 is capable of following without distortion is, in this example, 4,000 cps. In order to increase the response of the diode detector the time constant of R1C2 is reduced, for example, by decreasing R1 to 100 k-ohms. The highest audio frequency now becomes

$$f = \frac{1}{10R1C2}$$

$$= \frac{1}{10 \times 0.100 \times 10^6 \times 100 \times 10^{-12}}$$

$$= \frac{10^6}{10^2} \quad 10{,}000 \text{ cps.}$$

Demodulation capacitor C2 cannot discharge rapidly enough to follow modulation frequencies higher than 10,000 cps (in this case), and clipping results with all higher audio frequencies. The diode detector is discussed in detail in chapter 10.

Automatic Gain Control

Under ideal conditions, once the manual volume or gain control has been set, the output signal should remain at the same level even if the input signals vary in intensity. The development of variable-mu tubes makes it possible to devise a practical a-v-c or a-g-c circuit, since the amplification of the tube may be controlled by varying the grid-bias voltage. All that is needed is a source of bias voltage that varies with the signal strength. If this voltage is applied as bias to the grids of the variable-mu r-f amplifier stages, the grids will become more negative as the signal becomes stronger. The amplification will thus be reduced, and the output of the receiver will tend to remain at a constant level. Unless the selectivity of the i-f stages is good, strong adjacent-channel signals will reduce receiver gain when a weak signal is tuned in. When no interference is present, a-v-c holds the audio output constant as the input signal amplitude varies over a wide range.

The LOAD RESISTOR of a diode detector is an excellent source of this voltage, since the rectified signal voltage will increase and decrease with the signal strength. A filter is used to remove the a-f component of the signal and at the same time to prevent the a-v-c circuit from shorting the audio output. Only the slower variations due to fading or change of position of the receiving antenna, and so forth, will then affect the gain of the r-f amplifier stages because the a-v-c circuit cannot compensate for very fast or extreme variations.

Figure 14-15 shows how the a-v-c voltage is obtained. The a-v-c voltage is tapped off at the negative end of the diode load resistor, R2 (fig. 14-15,A), which is also the manual volume control. The a-f component is removed by the filter circuit that is composed of C2 and R1. One or more of the r-f amplifiers may be controlled by the voltage thus obtained. A customary value for R1 is 2 megohms and for C2 is 0.05 μf.

Figure 14-15,B, shows an a-v-c circuit used with a duodiode triode in a conventional diode detector circuit. The two plates of the diode are connected together to form a half-wave rectifier in the r-f portion of the circuit. The output of the diode detector is fed to the grid of the triode section which acts as a class-A voltage amplifier.

Low voltage bias is obtained by utilizing contact potential developed across R3 resulting from the dissimilar elements in the grid and cathode.

The variable-mu tube is designed to operate with a minimum bias of about -3 volts. The minimum bias is usually provided by a cathode resistor, and the a-v-c bias is applied in series with it. A disadvantage of ordinary automatic volume control is that even the weakest signals produce some a-v-c bias, which reduces the amplification slightly.

Delayed Automatic Gain Control

The disadvantage of automatic gain control, that of attenuating even the very weak signals, is overcome by the use of delayed automatic gain control, as shown in figure 14-16. In this circuit the a-v-c diode, plate 2, is separated

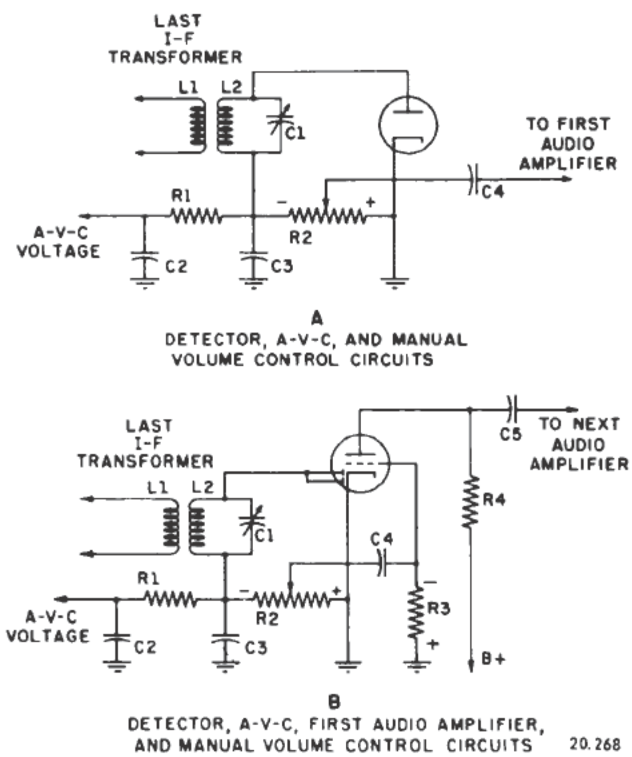


Figure 14-15.—Manual and a-v-c circuits.

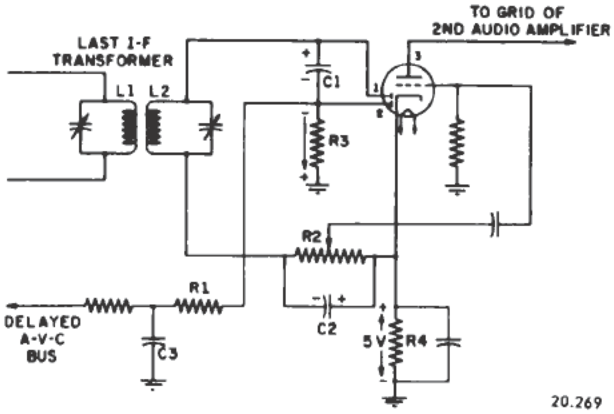


Figure 14-16.—Delayed a-v-c circuits.

from the detector diode, plate 1, and both are housed in the same envelope with a triode amplifier.

In this example a bias of 5 volts on the delayed a-v-c diode, plate 2, prevents it from conducting until the signal exceeds 5 volts. The signal across the secondary of the i-f transformer is coupled to diode, plate 2, by capacitor C1. Until the signal exceeds 5 volts no charge is acquired by the a-v-c capacitor, C3; no additional bias is applied to the grids of the i-f amplifier, preselector, or converter tubes; and their gain is maximum on weak signals. The 5-volt bias applied to the delayed a-v-c diode, plate 2, is developed across cathode resistor R4 by the current flowing through the triode section of the tube. The triode section serves as a class-A voltage amplifier driven by the audio voltage developed across diode load resistor R2.

When the signal across the secondary of the i-f transformer exceeds the 5-volt bias value across R4, the a-v-c diode (plate 2) conducts on alternate half cylces and C3 acquires a charge. The voltage developed across C3 constitutes the delayed a-v-c voltage. It is supplied to the grids of the various stages ahead of the second detector in series with the cathode bias developed by the individual tubes.

Noise Limiter

Sudden bursts of noise in a receiver may be attenuated by the use of a series noise limiter, such as the one shown in figure 14-17. The diode detector circuit includes V1, R1, R2, and C2. The cathode of V2 is connected through R3 to the a-v-c line, which is negative with respect to ground. The plate of V2 is connected to the common connection, B, between the diode load resistors, R1 and R2.

When a normal signal is detected, the plate of V2 is negative with respect to ground by an amount equal to the voltage drop across R1. Normally the plate is less negative than the cathode, and V2 conducts, thus providing a continuous circuit through V2, for the audio voltage tapped off at point D.

If a sudden burst of noise comes through the receiver the voltage across R1 suddenly increases. A large negative potential with respect to ground is thus applied to the plate of V2. The cathode cannot follow this sudden change because of the long time constant of C4 + R4. The plate is now negative with respect to the cathode, and V2 ceases to conduct. Thus the output circuit to the audio amplifiers is opened and the receiver becomes momentarily quiet. The point at which V2 begins limiting depends on the average strength of the received signal in relation to the amplitude of the noise. On weak signals the a-v-c voltage and cathode bias are both small; thus a low-amplitude noise pulse swings the plate voltage negative with respect to the cathode bias, and V2 limits the audio output. Conversely, a strong signal is limited by a large amplitude noise pulse but is not limited by the same amplitude of noise pulse that cut off the weak signal.

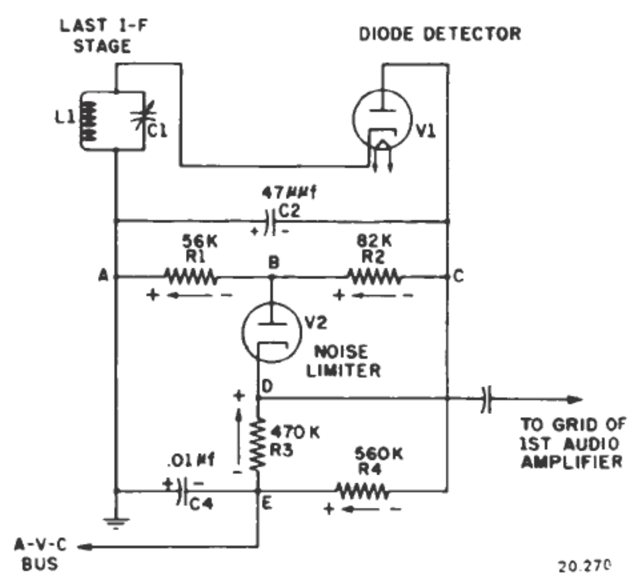


Figure 14-17.—Simplified circuit of a diode detector and series noise limiter.

Beat-Frequency Oscillator

The beat-frequency oscillator (bfo) is necessary when c-w signals are to be received because these signals are not modulated with an audio component. In superheterodyne receivers the incoming c-w signal is converted to the intermediate frequency at the first detector as a single frequency signal with no side-band components. The i-f signal is heterodyned (with a separate tunable oscillator known as the beat-frequency oscillator) at the second detector to produce an a-f output. In the circuit shown in figure 14-18, the Hartley oscillator (bfo) is coupled to the plate of the second detector by capacitor C3.

If the intermediate frequency is 455 kc and the bfo is tuned to 456 kc or 454 kc, the difference frequency of 1 kc is heard in the output. Generally the switch and capacitor tuning control are located on the front panel of the receiver.

The bfo should be shielded to prevent its own output from being radiated and combined with desired signals ahead of the second detector. If a-v-c voltage is to be used it should be obtained from a separate diode isolated from the second detector. One way is to couple the output of an i-f amplifier stage ahead of the second detector to the a-v-c diode. Otherwise, the output of the bfo would be rectified by the second detector and would develop an a-v-c voltage even on no signal.

Silencer

A silencer is sometimes employed in the a-f section of a receiver to disable the receiver when no signals are being received. One type of silencer circuit is shown in figure 14-19.

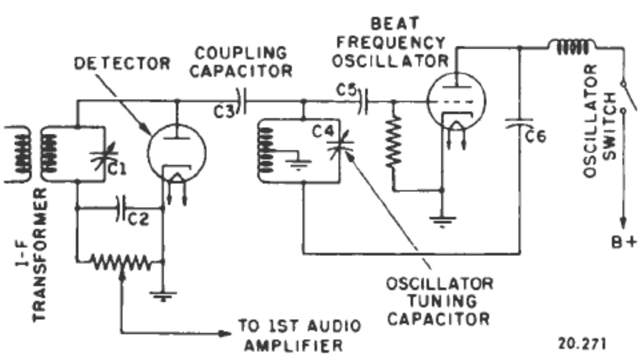


Figure 14-18.—Beat-frequency oscillator.

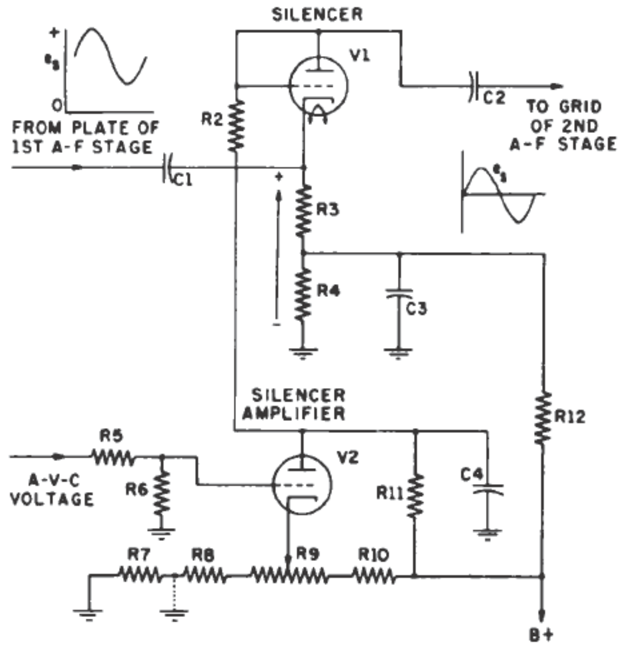


Figure 14-19.—Silencer circuit. 20.272

The silencer, V1, a diode-connected triode, connects the output of the first a-f stage to the input of the second audio amplifier. Silencer amplifier V2 serves as the control tube for the silencer. The plate voltage of V1 is supplied via R2 from the plate of V2 (which is in turn supplied from the B supply via R11) and is positive with respect to ground. The cathode voltage of V1 is also positive with respect to ground, since it is connected to the B supply through a voltage divider made up of R12 and R4. With no input signal, R9 is adjusted until V2 draws enough plate current to reduce its plate voltage and that of V1 to a value below the voltage on the cathode of V1. Thus the silencer plate voltage is negative with respect to the cathode. Conduction ceases, and the silencer cuts off. The output is reduced to zero, and the receiver is mute.

The grid of V2 is connected to the a-v-c line. When a signal enters the receiver, the negative a-v-c voltage is applied to the grid of V2, thereby reducing the plate current and increasing the plate voltage of both V2 and V1. When the plate of V1 becomes positive with respect to its cathode, the tube conducts and the signal is passed to the second a-f amplifier.

CIRCUIT OF A SUPERHETERODYNE RECEIVER

The complete circuit of a superheterodyne receiver is shown in figure 14-20. In this circuit

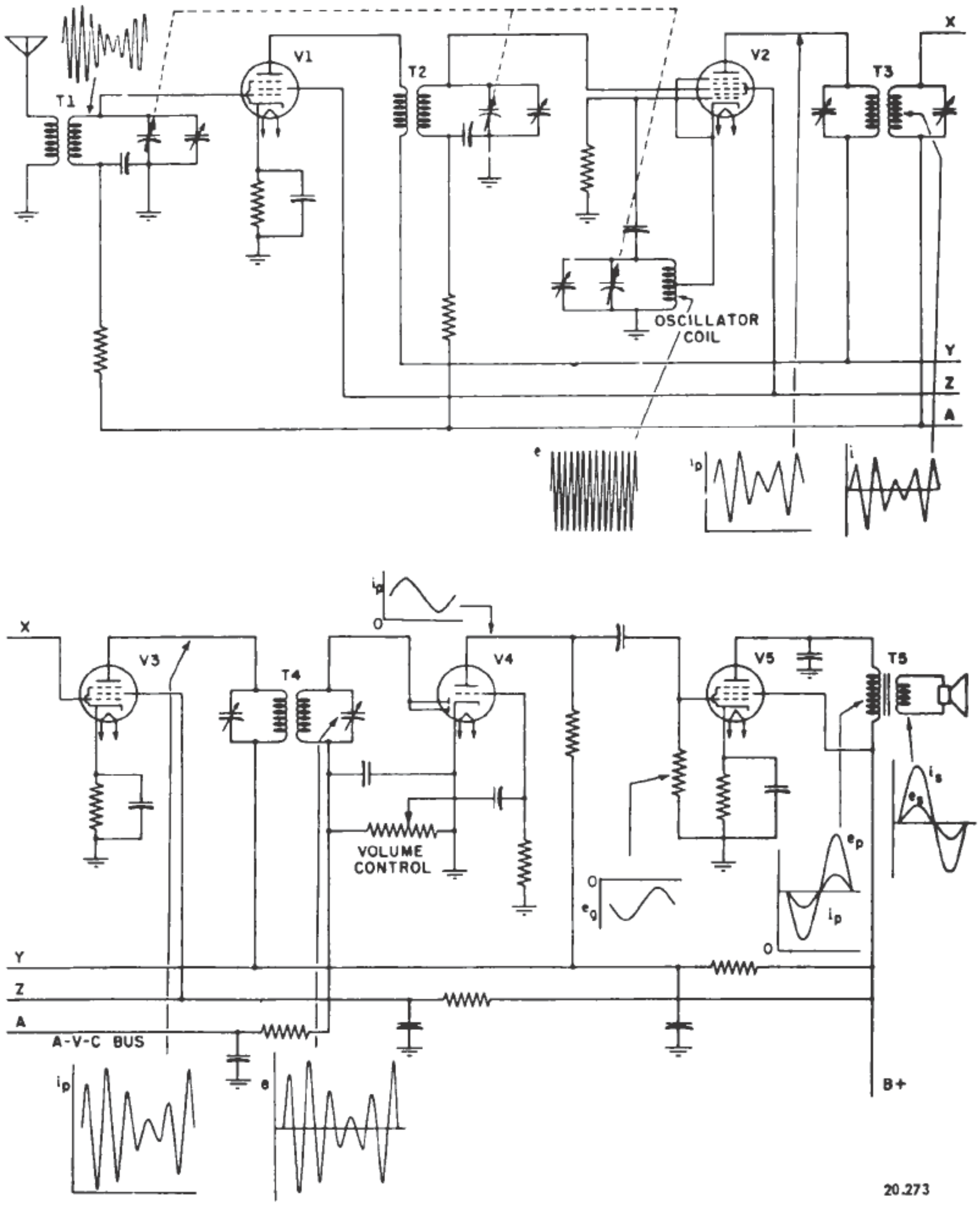


Figure 14-20.—Circuit diagram of a superheterodyne receiver.

one r-f amplifier (preselector) stage is used. Tube V2, a pentagrid converter, serves both as the mixer tube and oscillator tube. Three tuning capacitors (one each in the preselector, mixer, and oscillator stages) are ganged on a common shaft to assure proper tracking. Trimmers are connected in parallel with each tuning capacitor to permit alinement. The oscillator tuning capacitor is smaller than the tuning capacitor in the preselector or the converter stages. The oscillator operates above the station frequency and tracks closely at three points on the dial—(1) low end, (2) middle, and (3) high end. The oscillator tuning capacitor split-rotor plates allow closer adjustment for tracking at the low end and at the middle of the band. Shunt trimmer capacity adjustments on the oscillator tuning capacitor provide close tracking of the oscillator at the high end of the band.

Tube V3 is the i-f amplifier with input and output i-f transformers tuned to the receiver intermediate frequency.

Tube V4 serves as the second detector and first audio amplifier. Conventional automatic volume control is tapped off at the end of the volume control potentiometer farthest from ground. Plate and screen potentials are obtained from the B supply through the corresponding voltage dropping resistors. The power supply is a conventional full-wave rectifier.

F-M RECEIVERS

The t-r-f and superheterodyne receivers that have been described in the preceding paragraphs of this chapter are designed to receive r-f signals that vary in amplitude according to the audio modulation at the transmitter. The amplitude of the r-f signal is increased by one or more r-f amplifier stages, and the modulation component is removed by the detector. Each of the tuned circuits preceding the detector is designed to pass only a relatively narrow band of frequencies containing the necessary upper and lower sideband frequencies associated with the amplitude-modulated carrier.

F-m receivers are supplied r-f signals that vary in frequency according to the information being transmitted. The amount of the variation or deviation from the CENTER, or RESTING, FREQUENCY at a given instant depends on the amplitude of the impressed audio signal. The frequency with which the variations from the center frequency occur depends on the frequency of the impressed audio signal. The function of the f-m receiver is basically the same as that of the a-m superheterodyne receiver—that is, the amplitude of the incoming r-f signals is increased in the r-f stages; then the frequency is reduced in the mixer stage to the intermediate frequency and amplified in the i-f amplifier section. Finally, the amplitude is clipped in the limiter stage and the modulation component is removed by the second detector, or DISCRIMINATOR as it is called in the f-m receiver.

There are a few major differences between the f-m and the a-m receiver. The greatest difference is in the method of detection. Also the tuned circuits of the f-m receiver have wider bandpass and the last i-f stage is especially adapted for limiting the amplitude of the incoming signal. However, in both systems the audio amplifiers and reproducers are similar.

A comparison between a superheterodyne receiver designed for a-m reception and one designed for f-m reception is shown in figure 14-21.

COMPONENTS

The function of the f-m antenna is to provide maximum signal voltage to the receiver input. Unlike most broadcast a-m receiver antennas, f-m receiver antennas act as resonant lines having standing waves on them. Therefore f-m antennas are cut to the required length in order to receive a signal of sufficient amplitude to drive the first r-f amplifier.

R-F Section

If a single frequency is to be received, the antenna may be designed for maximum response at that frequency. If, however, a band of frequencies is to be received, the antenna length will represent a compromise. Usually the length is so chosen that it will be in resonance at the geometric center of the band. The GEOMETRIC CENTER or MEAN is equal to $\sqrt{\lambda_1 \lambda_2}$, where λ_1 and λ_2 are the wavelengths at the two ends of the band.

There are many types of f-m antennas, but probably the simplest is the half-wave dipole. The length of the half-wave dipole, in feet, is

$$\frac{468}{\sqrt{f_1 f_2}},$$

where f_1 and f_2 are the frequencies in megacycles at the two ends of the band. Because the resistance at the center of the half-wave dipole is about 72 ohms, the transmission line connecting the antenna with the receiver should have a characteristic impedance of 72 ohms in order to operate as a nonresonant transmission line with no standing waves. The transmission line feeds the signal to the receiver via a matching transformer at the input to the preselector stage.

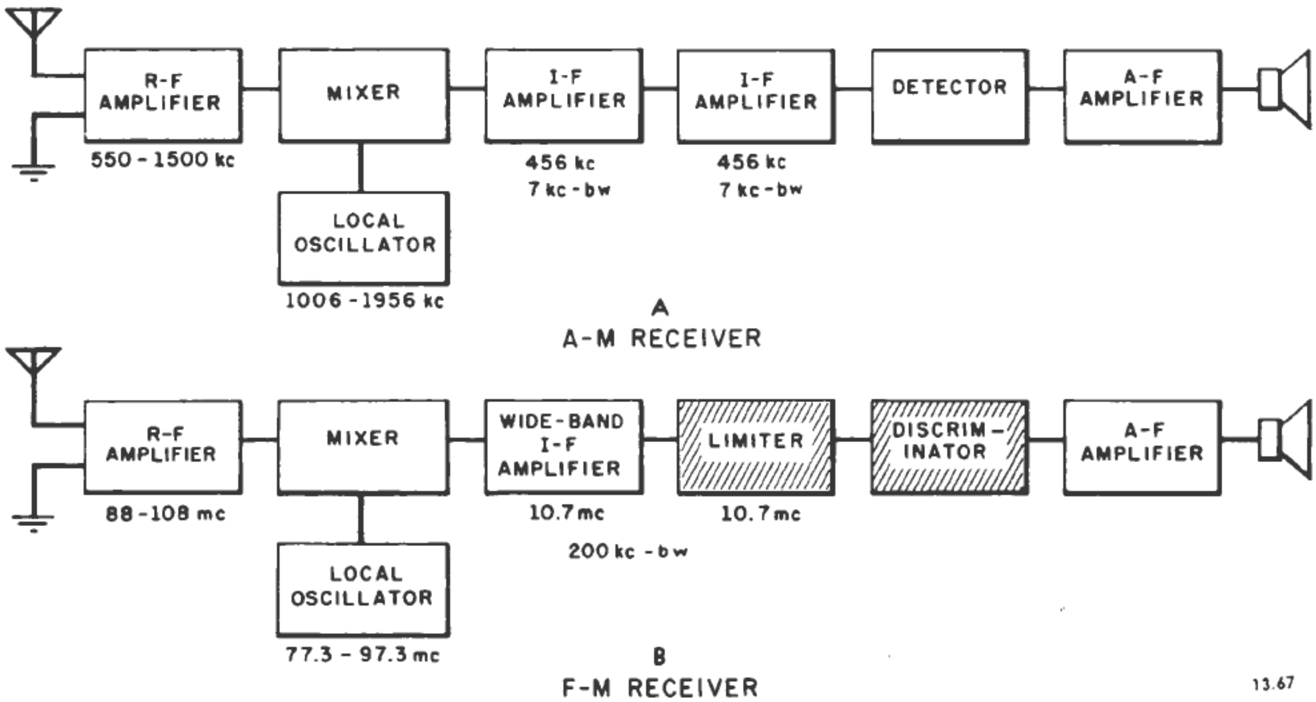


Figure 14-21.—Receiver block diagram.

The r-f amplifier, or preselector, performs essentially the same function in the f-m receiver as it does in the a-m receiver—that is, it increases the sensitivity of the receiver. Such an increase in sensitivity is often a practical necessity in fringe areas. However, the gain of the i-f stages is relatively much greater, perhaps 10 times that of the preselector, since the chief advantage of the superheterodyne lies in the uniformity of response and gain of the i-f stages within the receiver band. The principal functions of the r-f stage are to discriminate against undesired signals (images) and to increase the amplitude of weak signals so that the signal-to-noise ratio will be improved.

If the receiver is designed to receive both amplitude modulation and frequency modulation, a suitable band-switching arrangement is necessary. Many combination receivers are designed to receive more than one a-m band. Under such circumstances additional tuned circuits are needed. Thus, if two a-m bands and one f-m band are used and one r-f stage is used ahead of the mixer, three tuned circuits are needed for each band to be covered. This circuit arrangement includes one for the r-f stage, one for the mixer stage, and one for the oscillator stage for each of the three bands, or a total of nine tuned circuits. The f-m tuned circuits have wider bandpass characteristics than do the a-m tuned circuits, as shown in figure 14-21.

Frequency Converter

The frequency converter employed in the f-m receiver functions in much the same manner as the one employed in the a-m superheterodyne receiver. However, additional problems are involved.

For example, at the frequencies employed in the commercial f-m band the stability of the local oscillator becomes a major problem. As mentioned before, there is a tendency for the local oscillator to become synchronized with the incoming signal and thus to lose the intermediate frequency output entirely. The tendency is more pronounced at f-m frequencies because the station and oscillator are relatively closer together. Therefore, for maximum frequency stability, separate oscillator tubes are used. This results in increased space requirements and expense. Especially designed pentagrid converters that have reasonably good frequency stability, high conversion transconductance, and oscillator transconductance are employed in some less expensive commercial sets.

Even in a normal well-designed f-m receiver such factors as the change in internal capacitance of the oscillator tube (or oscillator section of a tube) and the expansion of coil windings and capacitor plates during warmup may cause the local oscillator frequency, and consequently the intermediate frequency, to drift an appreciable amount. A relatively small shift in oscillator frequency (always downward with respect to the center frequency in an uncompensated oscillator circuit) may shift the i-f signal beyond the range of the i-f stages with a consequent loss in output signal.

Various methods are used to combat oscillator drift. For example, the second harmonic of the local oscillator frequency is sometimes used for mixing. In this instance the local oscillator may be operated at a lower fundamental frequency, where the stability is improved. Another method is to use capacitors having a negative temperature coefficient. These are connected in shunt with capacitors having a positive temperature coefficient to counteract the change in capacitance when the temperature of the oscillator stage varies. Proper voltage regulation as well as the choice of oscillator tubes having low internal capacitances, will also increase the stability of the local oscillator.

Frequency stability of the local oscillator, in the standard f-m band, makes it advantageous to operate the local oscillator at a frequency below that of the incoming signal. (See fig. 14-21.)

However, if the local oscillator is operated above the frequency of the incoming signal it is not so likely to interfere with television receivers in the same vicinity that are operating on the lower video channels. Therefore, some commercial f-m receivers have local oscillators operating above the incoming signal.

I-F Amplifier

The i-f amplifier in an f-m receiver is usually tuned to a center frequency of from 8 to 10 megacycles. It generally employs double-tuned transformers having equal primary and secondary inductances. The bandpass is from 150 to 200 kc. The last one or two i-f stages function as a limiter.

The gain of each wide-band i-f stage is considerably less than that of the narrow-band a-m type of i-f amplifier. Therefore, an f-m receiver employs more i-f stages than a corresponding a-m receiver. The gain of an i-f amplifier employing a double-tuned transformer has been given in chapter 5 as

$$V.G. = \frac{g_m \omega K \sqrt{L_p L_s}}{K^2 + \frac{1}{Q_p Q_s}}.$$

In the case of a wide-band i-f amplifier having a double-tuned transformer, critical coupling $K = \left(\frac{1}{Q}\right)$, and primary and secondary inductances and Q's that are equal, the gain becomes

$$V.G. = \frac{g_m \omega L Q}{2}.$$

In this formula g_m is the transconductance of the tube, and ωL is the inductive reactance of the circuits at the intermediate frequency.

A low value of intermediate frequency is undesirable because local oscillator drift might force the set to operate outside the i-f range. Also, it would be pointless to have the intermediate frequency lower than the total frequency deviation (bandwidth) of any one f-m station.

In the choice of the optimum i-f value such factors as image response, response to signals at the same frequency as the intermediate frequency, response to beat signals produced by two stations separated in frequency by the i-f value, and response to harmonic frequencies must be considered.

Two stations separated in frequency by the i-f value will, if sufficiently powerful, produce a beat frequency that will pass through the receiver. This type of interference may be eliminated if the intermediate frequency chosen is greater than the entire f-m bandwidth. It may be minimized by adequate discrimination in the preselector stage.

Harmonics of the local oscillator may combine with harmonics produced when a strong incoming signal overloads the input stage to produce the intermediate frequency.

Interfering signals may develop as a result of the interaction of these harmonic frequencies. For example, consider an f-m receiver having an intermediate frequency of 9.1 mc. and tuned to an 86-mc station. The oscillator frequency is 86 + 9.1, or 95.1 mc. It is possible that a strong 90.5-mc signal picked up at the f-m receiver input would develop at that point its second harmonic of 181.0 mc. The oscillator second harmonic frequency is 95.1 x 2, or 190.2 mc. The difference frequency is 190.2 - 181.0, or 9.20 mc. This difference frequency would appear in the output of the mixer stage and be accepted by the i-f amplifiers tuned to 9.1 mc. Thus the receiver output would contain the 86-mc station and simultaneously the 90.5-mc interfering signal.

Harmonics produced at the input may be reduced by increasing the selectivity of the tuned circuits and using variable-mu tubes that do no overload easily. The production of harmonics by the local oscillator may be reduced by maintaining a satisfactorily high circuit Q and by reducing its loading.

The uniform gain within the band pass of an a-m i-f amplifier, and the selectivity of i-f amplifiers, are treated in chapter 6. In commercial f-m i-f amplifiers the bandpass is considerably greater than it is in a-m i-f amplifiers because of the greater frequency swing used in frequency modulation. An ideal frequency response curve is difficult to obtain economically. Therefore, a practical compromise that gives the necessary uniform gain and discrimination against adjacent channel frequencies is chosen.

The i-f stage may be designed for f-m only or for both a-m and f-m. An i-f transformer designed for both a-m and f-m is shown in figure 14-22. In order to have the desired high L/C ratio for increased gain and increased bandwidth permeability tuning is employed. Circuits C1L1 and L2C2 are tuned to the higher f-m intermediate frequency, about 10 mc, and have greater bandpass, about 200 kc. Circuits C3L3 and L4C4 are tuned to the lower a-m intermediate frequency, perhaps 455 kc, and the bandpass is lower, about 7 kc.

When the receiver is adjusted for f-m reception, only the f-m section of the i-f transformer is effective in coupling signal voltage to the next tube. Capacitor C3, having a low reactance to the higher f-m signals, shunts the a-m section of the transformer. Likewise, when the receiver is adjusted for a-m reception, only the a-m section of the i-f transformer is effective in coupling signal voltage to the next tube. In this case L1 becomes an effective short circuit for the lower frequency a-m

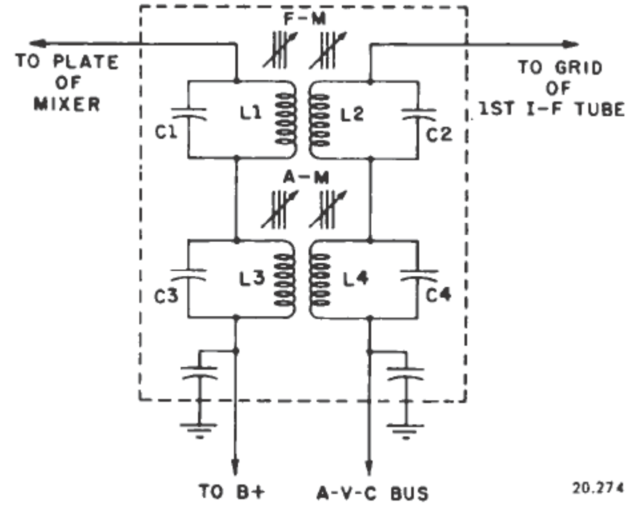


Figure 14-22.—I-f transformers for a-m and f-m.

signals. Usually the last i-f stage is modified to operate as a limiter.

Limiter

The limiter in an f-m receiver removes amplitude modulation and passes on to the discriminator an f-m signal of constant amplitude.

As the f-m signal leaves the transmitting antenna it is varying in frequency according to an audio-modulating signal, but it has essentially a constant amplitude. As the signal travels between the transmitting and receiving antenna, however, natural and man-made noises, or static disturbances, are combined with it to produce variations in the amplitude of the modulated signal. Other variations are caused by fading of the signal. Fading might be caused, for example, by movement of the ship carrying the transmitter or the receiver. Still other amplitude variations are introduced within the receiver itself because of a lack of uniform response of the tuned circuits.

All of these undesirable variations in the amplitude of the f-m signal are amplified as the signal passes through the successive stages of the receiver up to the input of the limiter. This condition in which both frequency modulation (desired) and amplitude modulation (undesired) are present at the same time is shown in figure 14-23,A.

The character of the signal after leaving the limiter should be as indicated in figure

14-23,B, in which all amplitude variations have been removed, leaving a signal that varies only in frequency.

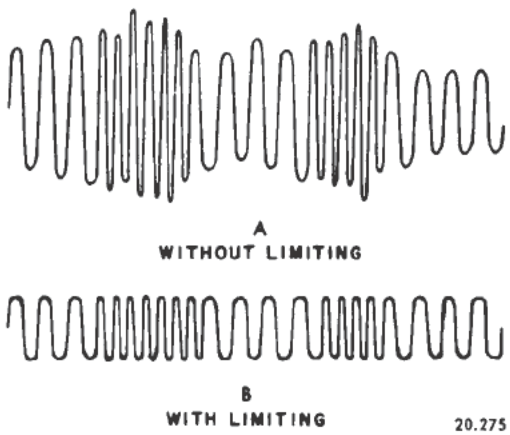


Figure 14-23.—F-m signals.

A grid-leak bias limiter is shown in figure 14-24. The tube is a sharp-cutoff pentode operated with grid-leak bias. Because the plate and screen voltages are purposely made low, plate-current saturation as well as plate-current cutoff, is produced readily by input signals having a magnitude of only a few volts.

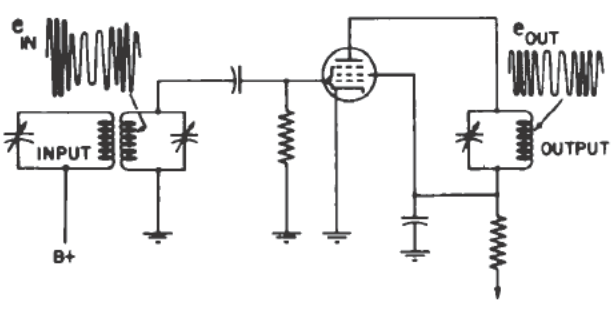


Figure 14-24.—Grid-leak bias limiter.

The manner in which the limiter functions is illustrated by the i_p-e_g curve shown in figure 14-25. Grid-leak bias is used so that with varying signal amplitudes, the bias can adjust itself automatically to a value that allows just the positive peaks of the signal to drive the grid positive and cause grid current to flow.

Suppose that a signal having a peak amplitude greater than the cutoff bias is impressed on the grid of the tube. A bias voltage having

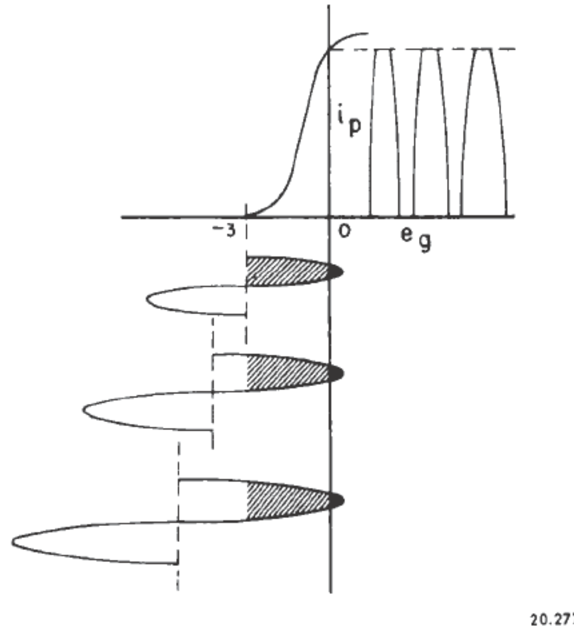


Figure 14-25.—Limiter i_p-e_g curve.

a magnitude approximately equal to the peak value of the signal will be developed. Accordingly, grid current will flow for a very small part of the positive half cycle at the peak of signal swing, as shown by the shaded area. Plate current flows for almost the entire positive half cycle. When the signal amplitude increases, a greater bias is developed, but the grid cutoff voltage remains the same and the average plate current changes very little. Thus, the amount of plate-current flow in the limiter stage is approximately constant for all signals having an amplitude great enough to develop a grid-leak bias voltage that is greater than the cutoff voltage. The frequency variations in the f-m signal are maintained in the output because the plate-current pulses are produced at the signal frequency and excite the plate-tuned tank circuit which has a relatively low Q and a wide bandpass. Thus, because of the "flywheel" effect, a complete a-c waveform is passed to the secondary of the discriminator transformer for each cycle of input signal.

When the peak amplitude of the grid signal is less than the cutoff voltage, the limiting action fails because the stage is practically a class-A amplifier for such signals, and the average plate current varies as the grid-leak bias changes with varying signal amplitudes. For this reason, the stages preceding the limiter must have sufficient gain to provide satisfactory limiting action on the weakest signal to be received.

Discriminator

Another major difference between the a-m receiver and the f-m receiver is in the method used to detect the signal. The DETECTOR in an a-m receiver interprets the AMPLITUDE VARIATIONS of the amplitude-modulated r-f energy in terms of the audio signal. In the f-m receiver, the discriminator interprets the FREQUENCY VARIATIONS of the frequency-modulated r-f energy in terms of the audio signal.

Several types of f-m detectors have been developed and are in use, but perhaps two of the most common types are the discriminator and the ratio detector. The operation of these detectors is treated in chapter 10.

The discriminator requires a limiter, which in turn requires considerable amplification ahead of its input.

An f-m detector that would be insensitive to amplitude variations would eliminate the need for a limiter, and in addition one or more i-f amplifier stages might be eliminated. Such an improved discriminator circuit that meets these requirements to a larger degree than the discriminator, is the RATIO DETECTOR.

CIRCUIT OF AN FM TUNER

A schematic diagram of an f-m tuner is shown in figure 14-26. Tubes V1 and V3 are remote-cutoff tubes using cathode bias without automatic volume control. Automatic volume control is not so important in f-m as it is in a-m, since in f-m, particularly if the second detector is a discriminator, the i-f stages are operated at maximum gain. The ratio detector shown in the figure provides a convenient source of a-v-c voltage, which is supplied to the grids of V4 and V5. The tuning range of the input tank, and also that of the tuned circuit in the mixer input, is 88 to 108 megacycles.

If the intermediate-frequency stages are tuned to 10 mc, and the local oscillator, V2, is operated above the station frequency, then the local oscillator is tunable from 88 + 10, or 98 mc, to 108 + 10, or 118 mc. Oscillator tube V2 is especially designed for high-frequency operation. Tube V3 is a pentagrid mixer used for mixing the incoming signal with the locally generated signal.

The i-f amplifiers are remote-cutoff pentodes. As mentioned previously, the i-f transformers must have the desired wide bandpass characteristic (200 kc).

Tube V6, a twin diode, is operated as a ratio detector. The audio output from this detector is fed to a conventional audio amplifier, not shown in the circuit diagram.

The B supply is obtained from a full-wave rectifier, V7, as shown in figure 14-26.

AUTOMATIC FREQUENCY CONTROL

Automatic frequency control (afc) is sometimes used with f-m receivers to hold the receiver automatically on the station frequency because the local oscillator has a tendency to drift.

The system keeps the local oscillator frequency separated from the station frequency by exactly the value of the intermediate frequency at all times.

The representative afc system includes a discriminator or ratio detector and a reactance tube modulator (described in chapter 10 of this training course). The output of the discriminator contains a d-c component whose polarity is dependent upon the direction of the deviation of the i-f from the mean or center value (that is, above or below the i-f center value). The magnitude of the d-c component depends upon the amount of deviation of the i-f from the center or mean value. A filter separates the d-c from any a-c components.

The filtered d-c component is applied to the grid of the reactance tube in shunt with the local oscillator so as to change the frequency of the local oscillator to the proper value to produce the correct i-f. When the i-f output is correct the d-c component in the discriminator output becomes zero. In the f-m tuner (fig. 14-26) reactance tube V controls the oscillator, V2, frequency when a d-c component appears between terminals a and b of the ratio detector, V6.

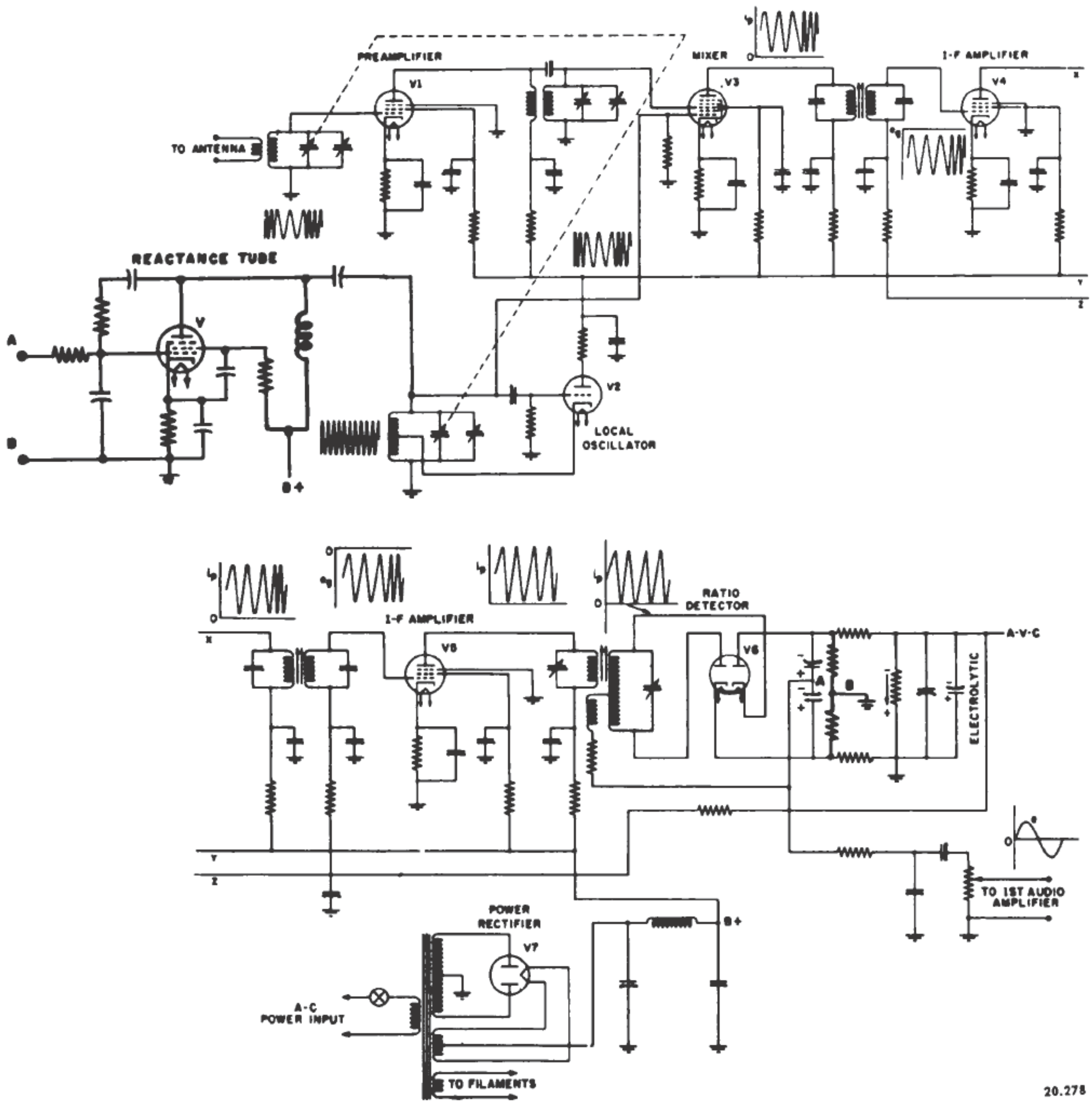


Figure 14-26.—F-m tuner.

CHAPTER 15

INTRODUCTION TO RADAR

ELEMENTS OF RADAR

The word RADAR is formed as an abbreviation for RAdio Detection And Ranging. RADAR is an electronic device that may be used to detect the presence of objects such as airplanes or ships even in darkness, fog, or storm. In addition to indicating their presence, radar may be used to determine their bearing and distance. In special types, elevation and speed may also be indicated. It is one of the greatest scientific developments that has emerged from World War II. Its development, like that of most great inventions, was mothered by necessity—that of detecting the enemy before he detected us. The basic principles on which its functioning depends are relatively simple, and the seemingly complicated series of electrical events encountered in radar can be resolved into logical series of functions, which, taken individually, may be identified and understood.

PRINCIPLES OF OPERATION

The principle upon which radar operates is very similar to the principles of sound echoes or wave reflection.

Sound Wave Reflection

If a person shouts in the direction of a cliff, or some other sound-reflecting surface, he hears his shout "return" from the direction of the cliff. What actually takes place is that the sound waves, generated by the shout, travel through the air until they strike the cliff. There they are reflected or "bounced off," and some are returned to the originating spot, where the person is then able to hear the echo. Some time elapses between the instant the sound originates and the time when the echo is heard because sound waves travel through air at approximately 1100 feet per second. The farther the person is from the cliff, the longer this time interval will be. If a person is 2200 feet from the cliff when he shouts, about 4 seconds elapses before he hears the echo—that is, 2 seconds for the sound waves to reach the cliff and 2 seconds for them to return.

If a directional device is built to transmit and receive sound, the principles of echo together with a knowledge of the velocity of sound can be used to determine the direction, distance, and height of the cliff shown in figure 15-1. A sound transmitter, which can generate pulses of sound energy, can be so placed at the focus of the reflector that it radiates a beam of sound. The sound receiver can be a highly directional microphone located inside a reflector (at its focal point, and facing the reflector) to increase the directional effect. The microphone is connected through an amplifier to a loudspeaker.

Then, to determine the distance and direction of the cliff the transmitting and receiving apparatus are placed so that the line of travel of the transmitted sound beam and the received echo will very nearly coincide. They would coincide exactly if the same reflector could be used for both transmitting and receiving, as is done in radar systems. The apparatus (both the transmitter and receiver) is rotated until the maximum volume of echo is obtained. The horizontal distance to the cliff can then be computed by multiplying one-half of the elapsed time in seconds by the velocity of sound. This will be essentially the distance along the line RA (fig. 15-1, A). If the receiver has a circular scale that is marked off in degrees, and if it has been properly orientated with a compass, the direction or azimuth of the cliff can be found. Thus, if the angle indicated on the scale is 45°, the cliff is northeast from the receiver position.

To determine height (fig. 15-1, B), the transmitter and receiver antennas are tilted from the horizontal position (shown by dotted lines) while still pointing in the same direction. At first the echo is still heard, but the elapsed time is

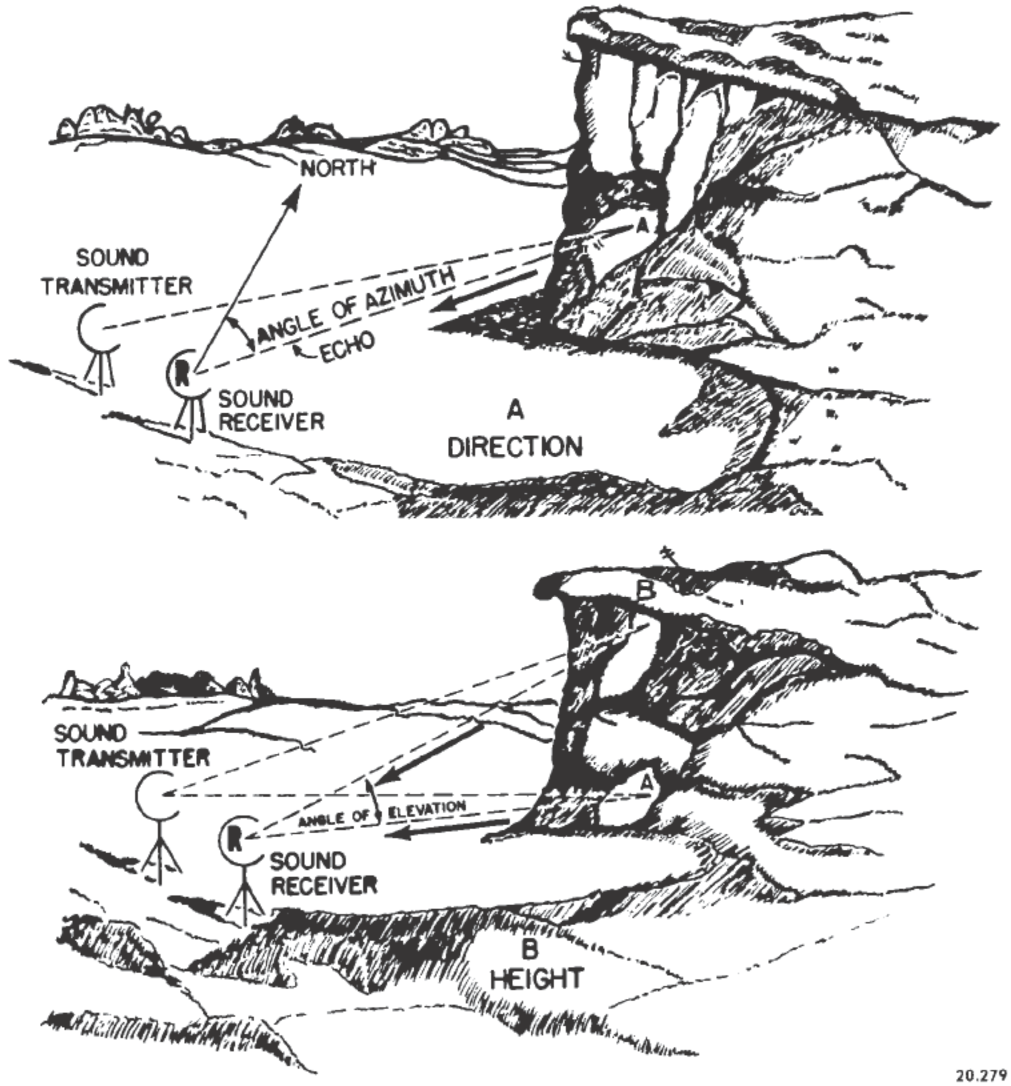


Figure 15-1.—Determination of direction and height.

increased slightly. As the angle of elevation is increased, an angle is found where the echo disappears. This is the angle at which the sound is passing over the top of the cliff and is therefore not reflected back to the receiver. The angle at which the echo just disappears is such that the apparatus is pointing along line RB. If the receiver is equipped with a scale that permits a determination of the angle of elevation, the height of the cliff, AB, can be calculated from this angle and either the distance RA or RB, by the use of one of the basic trigonometric ratios.

Radio-Wave Reflection

All radar sets work on a principle very much like that described for sound waves. In radar sets, however, a radio wave of extremely high frequency is used instead of a sound wave. The energy sent out by a radar station (fig. 15-2) is similar to that sent out by an ordinary radio transmitter.

The radar station has one outstanding characteristic different from a radio in that it picks up its own signals. It transmits a short pulse and receives its echoes, then transmits another pulse and receives those echoes. This out-and-back cycle is repeated 60 to 4000 times per second, depending on the design of the set. If the outgoing wave is sent into clear space, no energy is reflected back to the receiver. The wave and the energy that it carries simply travel out into space and are lost for all practical purposes.

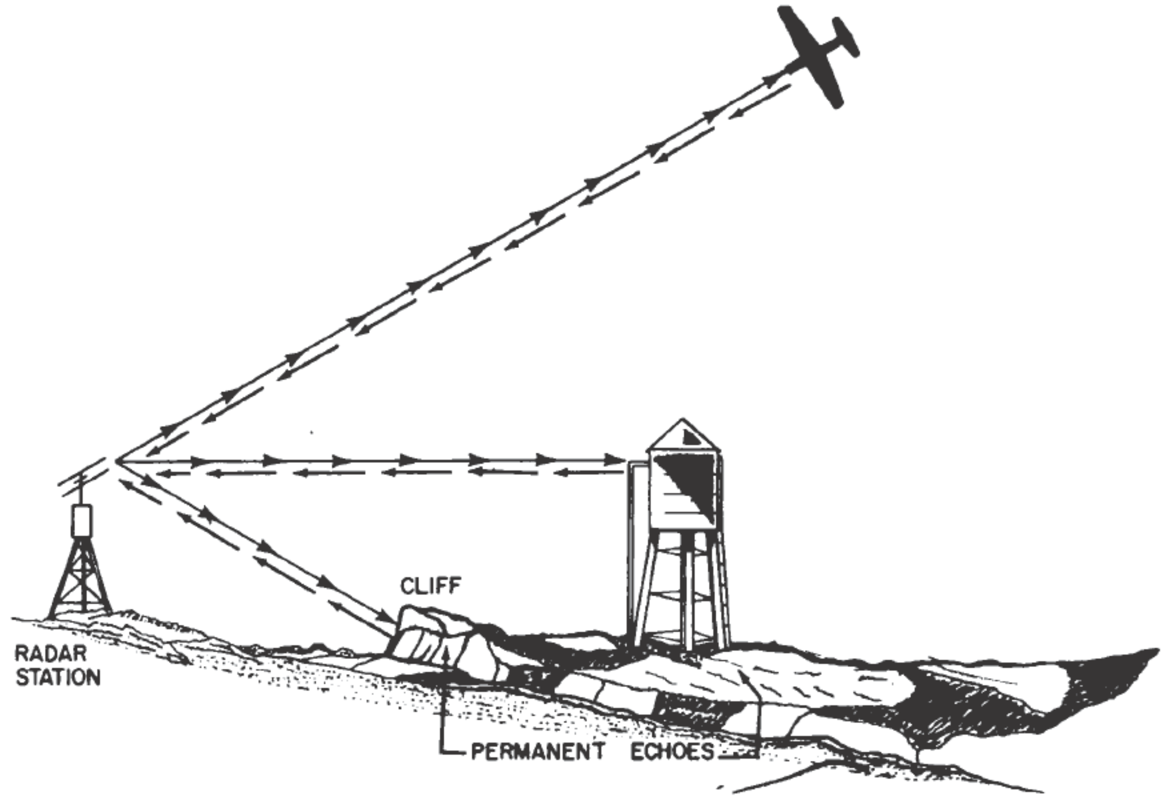


Figure 15-2.—Transmission and reflection of radar pulses.

If, however, the wave strikes an object such as an airplane (fig. 15-2), a ship, a building, or a hill, some of the energy is sent back as a reflected wave. If the object is large compared to a quarter-wavelength of the transmitted energy, a strong echo (but only a fraction of the transmitted energy) is returned to the antenna. If the object is small, the reflected energy is small and the echo is weak.

Radio waves travel at the speed of light, approximately 186,000 land miles per second or 162,000 nautical miles per second. A physical concept of this speed may be gained by considering the circumference of the earth as approximately 21,770 nautical miles. A radio wave could encircle the earth in approximately 134,400 μs (microseconds), or in slightly less than one-seventh of a second. Review the article in chapter 13 on "multiple refraction" to see how a radio wave is guided around the earth one or more times. Fifteen trips around the earth would be accomplished in slightly more than 2 seconds. This amounts to 2,000,000 μs. Consider that the moon is about 206,000 nautical miles (240,000 land miles) from the surface of the earth. Radar signals have been directed toward the moon and their echoes were returned. The elapsed time was approximately 2 1/2 seconds.

Most of the radio waves of the u-h-f and s-h-f bands are only slightly affected by the earth's atmosphere and travel in straight lines. Accordingly, there will be an extremely short time interval between the sending of the pulse and the reception of its echo. It is possible, however, to measure the interval of elapsed time between the transmitted and received pulse with great accuracy—even to one ten-millionth of a second (1×10^{-7} seconds). The forming, timing, and presentation of these pulses are accomplished by a number of special circuits and devices.

The directional antennas employed by radar equipment transmit and receive the energy in a fairly sharply defined beam. Therefore, when a signal is picked up, the antenna can be rotated until the received signal is maximum. The direction of the target is then determined by the position of the antenna.

The echoes received by the radar receiver appear as marks of light on an oscilloscope (called "scope" for short). This scope may be marked with a scale of miles (or yards), or degrees, or both. Hence, from the position of a signal echo on the scope, an observer can tell the range and bearing of the corresponding target.

RADAR DETECTING METHODS

Continuous-Wave Method

The continuous-wave (c-w) method of detecting a target makes use of the Doppler effect. The frequency of a radar echo is changed when the object which reflects the echo is moving toward or away from the radar transmitter. This change in frequency is known as the DOPPLER EFFECT. A similar effect at audible frequencies is recognized readily when the sound from the whistle of an approaching train appears (to the ear) to increase in pitch. The opposite effect (a decrease in pitch) occurs when the train is moving away from the listener. The radar application of this effect permits a measurement of the difference in frequency between the transmitted and reflected energy and thus a determination of both the presence and speed of the moving target. This method works well with fast-moving targets, but not well with those that are slow or stationary. C-w systems are therefore limited in present usage.

Frequency-Modulation Method

In the frequency-modulation (f-m) method the transmitted energy is varied continuously and periodically over a specified band of frequencies. The instantaneous frequency of the energy being radiated by the antenna therefore differs from the instantaneous frequency being received by the antenna.

The frequency difference depends on the distance traveled and can be used as a measure of range. Moving targets produce a frequency shift in the returned signal because of the Doppler effect, however, and this affects the accuracy of range measurements. This method, therefore, works better with stationary or slow-moving targets than with fast-moving ones.

Pulse-Modulation Method

In the pulse modulation method the r-f energy is transmitted in short pulses in which the time duration may vary from 0.1 to 50 μs. If the transmitter is turned off before the reflected energy returns from the target the receiver can distinguish between the transmitted pulse and the reflected pulse. After all reflections have returned, the transmitter can be turned on again and the process repeated. The receiver output is applied to an indicator that measures the time interval between the transmission of the energy and its return as a reflection. Because the energy travels at a constant velocity, one-half the time interval between the outgoing pulse and its echo becomes a measure of the distance traveled by the pulse to the target, or the range. Because this method does not depend on the relative frequencies of the emitted and returned signals or on the motion of the target, difficulties experienced in the c-w and f-m methods are not present. The pulse-modulation method is used almost completely in military applications. Therefore, it will be the only method discussed in this text.

HISTORICAL DEVELOPMENT

One of the first observations of "radio echoes" was made in the United States in 1922 by Dr. A.H. Taylor at the Naval Research Laboratory. Dr. Taylor observed that a ship passing between a radio transmitter and receiver reflected some of the waves back toward the transmitter. Between 1922 and 1930 further tests proved the military value of this principle for the detection of objects that would normally be hidden by smoke, fog, or darkness. During this same period Dr. Breit and Dr. Tuve of the Carnegie Institute published reports on the reflection of pulse transmissions from electrified layers in the upper atmosphere. This led to the application of the principle to the detection of aircraft. Other countries carried on further experiments independently and with utmost secrecy. By 1936, the United States Army was engaged in the development of a radar warning system for coastal frontiers. By the end of 1940 the British had developed radar to such a point that they were able to bring down great numbers of enemy airplanes with guns being accurately controlled by radar systems. Beginning in 1941, British-American cooperation in the development of radar gave the Allies the best radar equipment in the world.

Along with the development of radar came the development of effective countermeasures. Since 1941 great advances have been made in radar and in countermeasures in the various research and development centers throughout the country.

USES OF RADAR

Naval scientists pioneered in finding practical uses for radar. Those uses were made chiefly for detecting and destroying an enemy and his armaments; this is still the most important naval use today. Civilian uses followed those made for military purposes. For example, some familiar civilian uses are (1) radar speed determination on highways for controlling traffic, (2) radar weather prediction, (3) commercial radar air navigation, and (4) safeguarding air vehicles and merchant ships from collision hazards.

Since no single radar appliance can yield all of the information required of naval shipboard systems, several classes, or types, have been designed. Each of these types, or categories, has its limitations and capabilities within the purpose for which built. Naval shipboard radar equipments are grouped in three general categories: search, fire-control, and special.

Search radars are of two categories: air search and surface search. These equipments are used for early warning networks and for general navigational purposes. Search radars produce detection at maximum ranges while sacrificing some degree of accuracy and resolution (detail).

Fire-control radars, integral parts of certain gunfire control systems, are used after targets have been located by search radars.

Special radars are used for specific purposes, which include recognition, or identification of friend or foe (IFF), ground-controlled and carrier-controlled approach (GCA and CCA, respectively), range rate or speed, and height finding.

Types of Presentation

To furnish usable information, a radar set must produce some type of visual presentation of the target echo so as to suggest a mental image of the target to the observer. Cathode-ray tubes are used for this purpose. The scope images contain data of measurable quantities, including range, time, height, speed, and azimuth. Several types of data presentation have been developed to give the required information.

A radar beam systematically reveals what it SCANS (scrutinizes or examines in great detail). The results of each scan are revealed (presented) as a picture or presentation on the scope. About 15 types of scans have been designed, but naval requirements can be met by use of only a few of them as can be seen in figure 15-3. Each type of scan is identified by a letter of the alphabet.

TYPE-A SCAN.—Type-A presentation is used to determine range. The screen of this scope has a short presistence. The echo causes a vertical displacement of the spot, the amplitude of which depends on the strength of the returned signal pulse. The point on the horizontal base line at which the vertical displacement occurs indicates the range. Type-A presentation is shown in figure 15-3,A.

TYPE-B SCAN.—The type-B presentation (fig. 15-3,B) indicates both range and azimuth angle (bearing). The vertical displacement of the echo signal indicates range, and the horizontal displacement of the echo signal indicates azimuth angle. This scope has long persistence.

TYPE-PPI SCAN.—The PPI (Plan Position Indicator) presentation is another type of scan for presenting range and bearing (azimuth) information. See figure 15-3,C. You can think of the PPI scan as a modified type-B scan, in which rectangular coordinates are replaced by polar coordinates.

The antenna is rotated uniformly about the vertical axis so that searching is accomplished in a horizontal plane. The radar beam is usually narrow in azimuth and broad in elevation. Large numbers of pulses are transmitted for each rotation of the antenna. As each pulse is transmitted, the spot starts from the center of the indicator and moves toward the edge along a radial line. Upon reaching the edge of the scope,

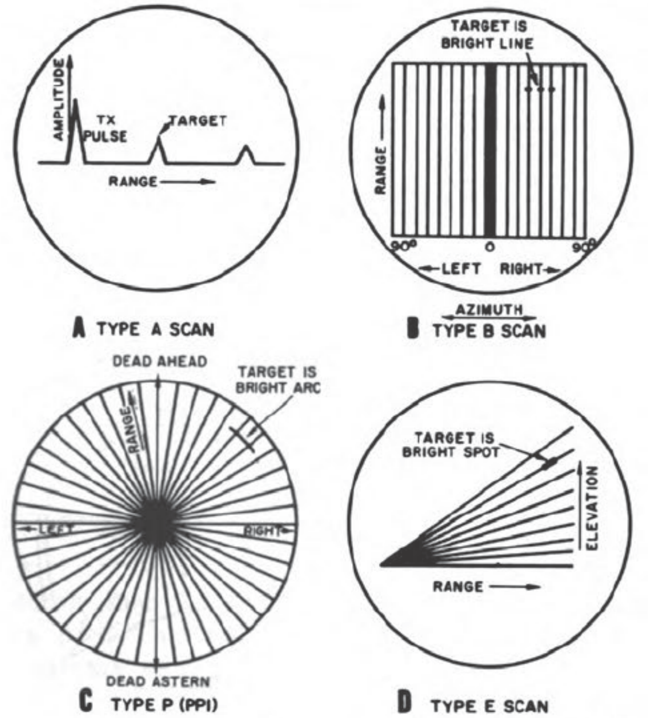


Figure 15-3.—Types of scans.

the spot quickly jumps back to the center and begins another trace as soon as the next pulse is transmitted. As the antenna rotates, the path of the spot also rotates around the center of the scope so that the angle of the radial line on which the spot appears indicates the azimuth of the antenna beam, and distance (out from the center of the scope) indicates the range.

When an echo is received, the intensity of the spot is increased considerably, and a brighter spot remains at that point on the screen, even after the scanning spot has passed it. Thus, it is possible with this scan to produce a map of the territory surrounding the observing station on the scope. This type of scan is useful when the radar set is used as an aid to navigation.

In type-A presentation the echo signal causes vertical deflection of the electron beam—in other words, it is a DEFLECTION MODULATED. In type-B and type-PPI presentation the echo signal makes the electron beam brighter. This action is called INTENSITY MODULATION.

TYPE-E SCAN (RHI).—The RHI (Range Height Indicator) presentation is another type of scan for presenting range and height information. The RHI scan is also known as the type-E scan shown in figure 15-3,D. The type-E scan is a modification of the type-B scan on which an echo appears as a bright spot with the range indicated by the horizontal coordinate and the elevation (height) as the vertical coordinate. This type is used (1) in directing planes in blind landing, (2) for ground-controlled approach, (3) for carrier-controlled approach, and (4) in determining altitude. This scan is also used in the Mk 56 gunfire control system (GFCS), which employs the Mk 35 radar. There the presentation is specified as the ΔE (delta E) scan, which means that the elevation changes are presented.

Range Determination

The successful employment of pulse-modulated radar systems depends primarily on the ability to measure distance in terms of time and a knowledge of the velocity of light. Radio-frequency energy, once it has been radiated into space, continues to travel with a constant velocity. When it strikes a reflecting object there is no loss in time, but merely a redirecting of the energy. Its velocity is that of light, or, in terms of distance traveled per unit of time, 186,000 land miles per second, 162,000 nautical miles per second, or 328 yards per microsecond. This means that it takes approximately 6.18 microseconds for radio energy to travel 1 nautical mile, or approximately 6080 feet (2027 yards). All radar ranging is based on a flat figure of 6080 feet per mile and, because the speed of light (and radio waves) is so great, microseconds (μs) are used for all time determinations.

This constant velocity of radio-frequency energy is applied in radar to determine range by measuring the time required for a pulse to travel to a target and return. The time lapse between the transmitted pulse and the echo return may be readily determined with the aid of the oscilloscope. For the purpose of illustrating how this may be done, assume that a target ship is 20 nautical miles away from the radar transmitter-receiver combination. Because radio energy travels 1 nautical mile in 6.18 microseconds, 123.6 microseconds will be required for the transmitted pulse to reach the target, or a total of 247.2 microseconds before the echo will return to the radar receiver.

The horizontal sweep frequency of the scope is adjusted so that it makes one complete sweep (from left to right) during the time the transmitted pulse is going to the target (maximum range) AND THE ECHO IS RETURNING TO THE RECEIVER. In other words, the time of one sweep is 247.2 microseconds, and the frequency is therefore approximately 4045 cps. Assume that a translucent scale with uniform divisions in miles from 0 to 20 is placed over the face of the scope; and assume further that the extent of the sweep extends from the 0 mark to the 20-mile mark. In this case the maximum range is 20 miles.

Figure 15-4 shows how the range to the target is determined. In part ① the transmitted pulse is just leaving the antenna. A part of the generated energy is fed to the vertical deflection plates at the instant the pulse is transmitted and causes a vertical line (pip) to appear at the zero-mile mark on the scope.

In part ②, 61.8 microseconds later, the transmitted pulse has traveled 10 miles toward the target. The horizontal trace on the scope, however, has reached only the 5-mile mark—that is, one-half the distance the transmitted pulse has traveled (the sweep frequency is timed to indicate one-half the distance).

In part ③, 123.6 microseconds after the initial pulse left the transmitter, the transmitted pulse has reached the target, 20 miles away and the echo has started back. The scope reading is 10 miles.

In part ④, 185.4 microseconds after the initial pulse, the echo has returned half the distance from the target, and the scope reading is 15 miles.

In part ⑤, 247.2 microseconds after the initial pulse, the echo has returned to the receiving antenna. This relatively small amount of energy is amplified and applied to the vertical deflection plates, and an echo pip of smaller amplitude than the initial pip is displayed on the scope at the 20-mile mark.

If two or more targets are in the path of the transmitted pulse each will return a portion of the incident energy as echoes. The targets farthest away (assuming they are similar in size and type of material) will return the weakest echo.

In conjunction with the scope there is a handcrank and mechanical counter assembly that enables the operator to determine the range to a greater degree of accuracy. When a target is indicated on the base line the operator turns the handcrank to move the range indicator, or gate (fig. 15-5), to the target and then reads the range, in yards, directly from the counter assembly. This process is known as "gating the target."

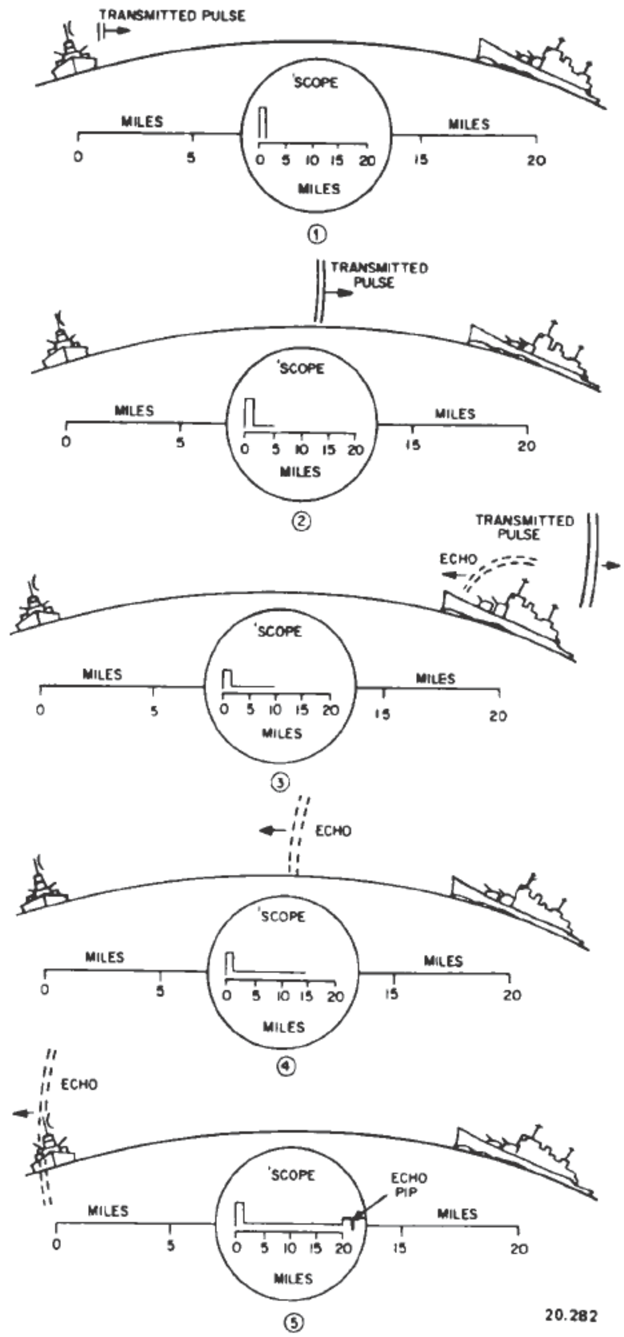


Figure 15-4.—Radar range determination.

Bearing Determination

The bearing (true or relative) of the target may be determined if the direction in which the

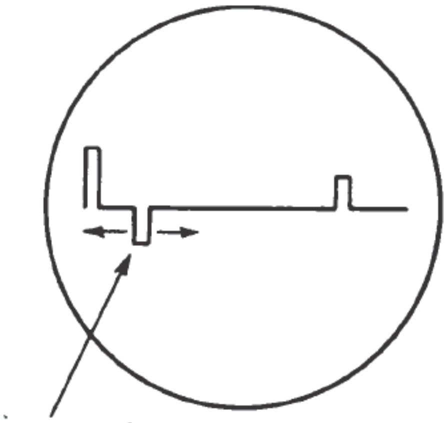


Figure 15-5.—Target gating.

directional antenna is pointing when the target is picked up is known. Control and indicator systems have been devised that make this possible.

The measurement of the bearing of a target as "seen" by the radar is usually given as an angular position. The angle may be measured either from true north (true bearing), or with respect to the heading of a vessel or aircraft containing the radar set (relative bearing). The angle at which the echo signal returns is measured by utilizing the directional characteristics of the radar antenna system. Radar antennas are constructed of radiating elements, reflectors, and directors to produce a single narrow beam of energy in one direction. The pattern produced in this manner permits the beaming of maximum energy in a desired direction. The transmitting pattern of an antenna system is also its receiving pattern. An antenna can therefore be used to transmit energy, to receive reflected energy, or to do both.

The simplest form of antenna for measuring azimuth or bearing is one that produces a single-lobe pattern. The system is mounted so that it can be rotated. Energy is directed across the region to be searched, by moving the beam back and forth in azimuth until a return signal is picked up. The position of the antenna is then adjusted to give maximum return signal.

Figure 15-6 shows the receiving pattern for a typical radar antenna. In this figure, relative signal strength is plotted against the angular position of the antenna with respect to the target. A maximum signal is received only when the axis of the lobe passes through the target. The sensitivity of this system depends on the angular width of the lobe pattern. The operator adjusts the position of the antenna system for maximum received signal. If the signal strength changes appreciably when the antenna is rotated through a small angle, the accuracy with which the on-target position can be selected is great. Thus, in figure 15-6, the relative signal strengths A and B have very little difference. If the energy is concentrated in a narrower beam, the difference is greater and the accuracy better.

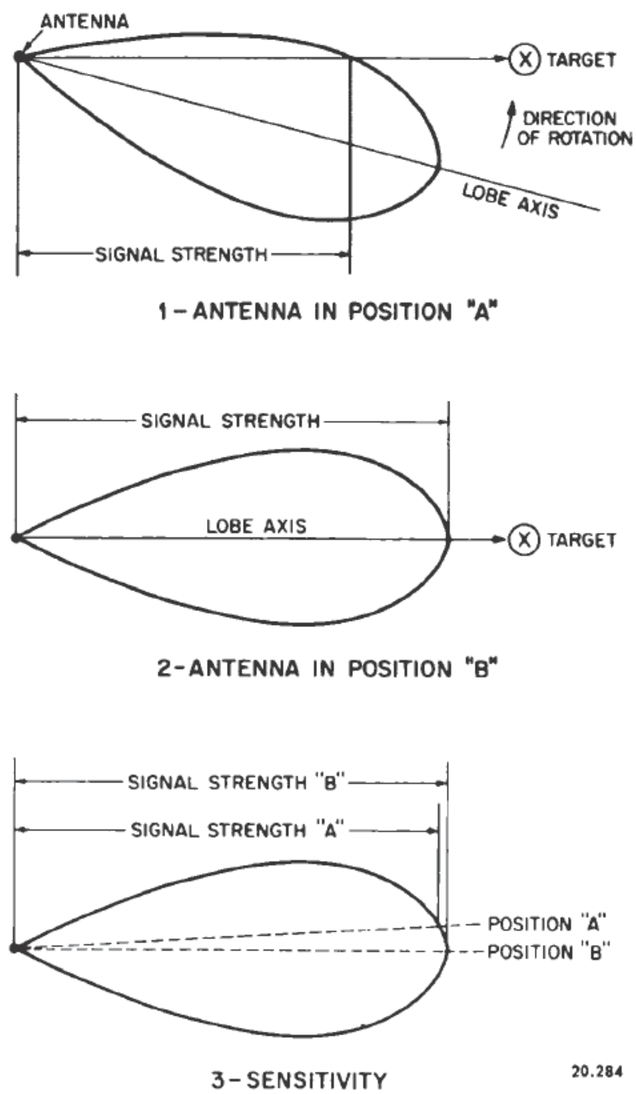


Figure 15-6.—Radar determination of azimuth or bearing.

Altitude Determination

The remaining dimension necessary to locate completely an object in space can be expressed either as an angle of elevation or as an altitude. If one is known, the other can be calculated from one of the basic trigonometric ratios. A method of determining the angle of elevation or the altitude is shown in figure 15-7. The slant range (fig. 15-7,A) is obtained from the radar scope indication as the range to the target. The angle of elevation is that of the radar antenna (fig. 15-7,B). The altitude is equal to the slant range multiplied by the sine of the angle of elevation.

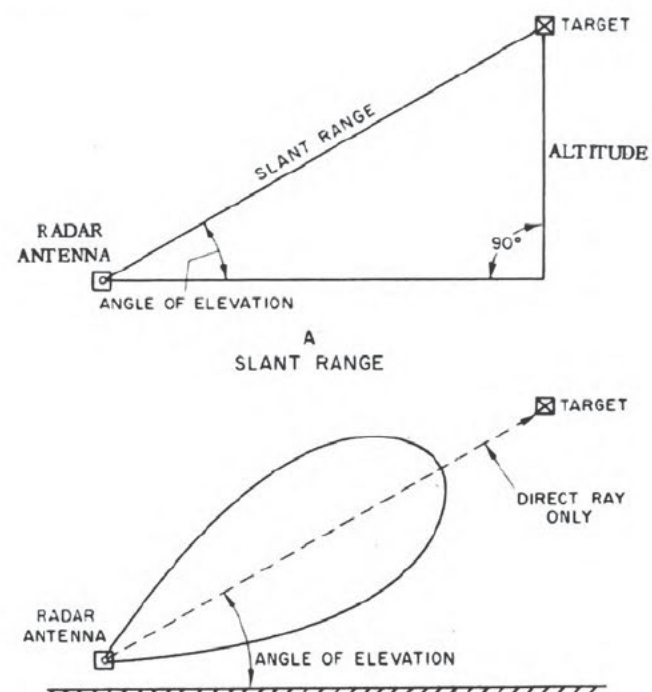


Figure 15-7.—Radar determination of altitude.

In radar equipments with antennas that may be elevated, altitude determination by slant range is automatically computed electronically. In equipments (air search) where the antennas do not elevate, the altitude may be calculated by means of "fade charts." A method for producing the chart and its application are described in the Navy training course for *Radarman 3&2*, NavPers 10146.

PLAN POSITION INDICATOR.—The range scope has certain limitations when it is desired to know what is happening instantaneously in all directions because it indicates only the targets in the direction in which the antenna is instantaneously pointing.

A master PPI allows the radar operator to see the screen images of all objects surrounding his craft (within the range limitations of the equipments) because it displays a graphic plot of 360° of antenna rotation and has a screen of the necessary persistence to retain the targets visible after the antenna has rotated past the target bearing.

The range scope presents the target information on a horizontal base line as shown in figure 15-3,A. The PPI has a radial base line originating at the center of the screen (fig. 15-3,C) which indicates the physical antenna location, and this line follows the antenna rotation.

A view of the PPI scope is shown in figure 15-8. The bright spots on the screen are images of objects (ships, planes, land masses, etc.) in the vicinity of the craft carrying the PPI equipment. Around the outer edge of the scope are relative and true bearing circles. Spaced evenly across the face of the tube are range circles, calibrated in miles. Thus, from the position of the images, their approximate range and bearing may be determined from the scope. A particular object of interest may be singled out for more accurate ranging by referring to the range scope.

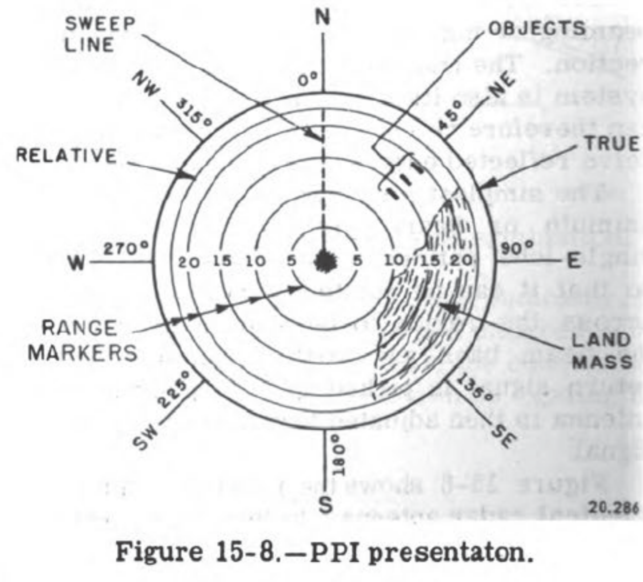


Figure 15-8.—PPI presentaton.

Another principal difference betwen the two systems (range and PPI) is the method of applying the signal to the scope. In the RANGE scope the echo signal is amplified and applied to the vertical deflection plates in such a way as to produce a pip on the horizontal time-base line, on the screen. In the PPI SCOPE, the echo signal is amplified and applied to the control grid of the scope in such a way that the trace is brightened momentarily on the radial time-base line. If the intensity of the trace is kept sufficiently low, the scope will be essentially dark until an echo is received, and then the contrast will be very pronounced.

The PPI uses electromagnetic deflection instead of electrostatic deflection. Current flowing from the sweep generator through a single pair of electromagnets mounted across the neck of the tube at right angles to the axis of the tube causes the electron beam to be swept from the center of the tube to one edge and back again to the center.

In some types of PPIs the deflection electromagnets are mounted so that they can be rotated around the neck of the tube. The rotating assembly is synchronized with the antenna rotation so that when the antenna turns, the sweep trace is rotated about the screen at the same rate.

Thus, for example, in figure 15-8 when the antenna is pointing in the NE direction the deflection magnets will force the beam across the screen from the center to the outer edge in the NE direction. The beam will be deflected across the screen many times during the course of a small angular rotation of the magnets. In this area on the screen the echoes from the three targets will cause three areas of intensification on the screen.

SEARCH RADAR.—Search radars used for early warning nets do not require great precision in ranging or bearing, but do require the ability to locate targets at fairly long ranges. Therefore, they are normally designed with high power, wide beam angle, and fairly long pulse widths. Their target resolution (ability to accurately determine bearing and range) is not as good as that of radars used for another purpose such as fire control.

Each type of radar equipment has been designed for definite purposes. An air search radar performs inadequately for surface search and vice versa. Each of these types may be used in an emergency as fire control radar, but cannot be expected to furnish bearings, ranges and position angles with the same degree of accuracy as radar designed for that purpose.

AIR SEARCH RADAR.—Primary function of an air search radar is the detection and determination of ranges and bearings of aircraft targets at long ranges maintaining complete 360° surveillance from the surface to high altitudes. System constants must be chosen with this function in mind. Relatively low radar frequencies are chosen (P or L band 30-1000 mc) to permit long-range transmissions with minimum attenuation. Wide pulse widths (2 to 4 μs) and high peak power are used to aid in detecting small targets at great distances. Low pulse repetition rates are selected to permit greater maximum measurable range. Wide vertical beam width is used to ensure detection of targets from the surface to relatively high altitudes, and to compensate for the pitch and roll of the ship. Medium horizontal beam width is employed to permit fairly accurate bearing resolution while maintaining 360° search coverage.

SURFACE SEARCH RADAR.—The primary function of surface search radar is the detection and determination of accurate range and bearing of surface targets while maintaining 360° surveillance for all surface targets within line-of-sight distance of the radar antenna.

Since the maximum range requirement of a surface search radar is primarily limited by the radar horizon, very high frequencies (X band) are employed to permit maximum reflection from small target-reflecting areas, such as ship mast-head structures and submarine periscopes. Narrow pulse widths (0.37 to 2 μs) are used to permit a high degree of range resolution at short ranges, and to achieve greater range accuracy. High-pulse repetition rates (600 to 1000) are used to permit maximum illumination of targets. Medium peak powers can be used to permit detection of small targets at line-of-sight distances. Wide vertical beam widths (10° to 30°) permit compensation for pitch and roll of own ship and to detect low-flying aircraft. Narrow horizontal beam widths (1° to 3°) permit accurate bearing determination and good bearing resolution.

FIRE CONTROL RADAR.—The primary function of fire control radar is the acquisition of targets originally detected and designated from search radars, and the determination of extremely accurate ranges, bearings, and position angles of targets within firing range. The antennas can be stabilized to compensate for pitch and roll of own ship. Very high frequencies are chosen to permit the formation of narrow beam widths with comparatively small antenna arrays, detection of targets with small reflecting areas, and high detail of all targets. Very narrow pulse widths provide a high degree of range accuracy, at short range, and excellent range detail. Very high repetition rates afford maximum target illumination while using very narrow pulse widths. Since long ranges are not required, low peak power permits the use of smaller components by keeping the average power low. Narrow, vertical, and horizontal beam widths provide accurate bearing and position angles and a high degree of bearing and elevation resolution.

AIRBORNE RADAR.—Radar equipments for aircraft are of the same general types as for land or shipboard except that they are physically much smaller. Both search and fire control radars are successfully used in aircraft. While radar is a powerful aid to aircraft, the aircraft in turn increases the range of radar by supplying it with an elevated platform from which its effective range in detecting objects is greatly extended because the line-of-sight distance is increased toward a farther horizon.

Radar information picked up in a plane may be relayed by radio transmitter to another distant location on board ship or elsewhere, thereby effectively increasing the range.

AUXILIARY EQUIPMENT.—Some means of distinguishing a friendly target from an enemy target is necessary An electronic system for "identification, friend of foe" (IFF system) is used for this purpose. This auxiliary equipment provides an accurate and rapid means of determining the friendly or enemy character of objects detected by radar.

The IFF system equipment consists of two groups containing two units in each group. Two units (transmitter-receiver) are located with the radar. The other two (receiver-transmitter) are located in friendly craft. The transmitter-receiver group is referred to as the RECOGNITION SET (also INTERROGATOR) and the receiver-transmitter group is called the IDENTIFICATION SET (also TRANSPONDOR or transponder).

When a radar operator observes an unidentified target on his radar, he sets the first group (the INTERROGATOR) in operation. The interrogator transmitter is a pulse-type transmitter, which emits coded challenging pulses. The transpondor's receiver receives these pulses, which trigger (automatically) a coded reply known only by the friendly operators. The transpondor's transmitter releases this coded reply, which is received by the interrogator's receiver and placed on an indicator for evaluation. The indicator may be either an integral section within the radar scope or a separate scope.

INFORMATION PRODUCED BY RADAR.—Radar increases the effectiveness of naval craft by adding new powers and capabilities to the human senses. It is unhampered by the ordinary obstacles to unaided vision such as darkness, fog, haze, and smoke. Radar reveals the presence and location of certain kinds of objects situated far beyond the range of normal vision, indicating their distance and bearing directly and with a high degree of accuracy. Radar pierces the surrounding darkness or overcast and reveals aircraft, ships, land areas, cities, clouds, and hazards to navigation.

FUNCTIONAL CONCEPTS

Radar systems now in existence vary greatly in detail. They may be very simple; or, if more accurate data are required, they may be highly refined. The principles of operation, however, are essentially the same for all systems. Thus a single basic radar system can be visualized in which the functional requirements hold equally well for all specific equipments.

In general, the degree of refinement of radar circuits increases with the frequency. The microwave region lends itself to a higher degree of precision in angular measurement, and for this reason modern radars operate at superhigh frequencies.

The functional breakdown of a pulse-modulated radar system generally includes six major components, as shown in the block diagram of figure 15-9. The components may be summarized as follows:

1. The modulator produces the synchronizing signals that trigger the transmitter the required number of times per second. It also triggers the indicator sweep and coordinates the other associated circuits. In some sets an external trigger generator is used to synchronize all triggered units.

2. The transmitter generates the r-f energy in the form of short, powerful pulses.

3. The antenna system takes the r-f energy from the transmitter, radiates it in a highly directional beam, receives any returning echoes, and passes these echoes to the receiver.

4. The receiver amplifies the weak r-f pulses returned by the target and reproduces them as video pulses to be applied to the indicator.

5. The indicator produces a visual indication of the echo pulses in a manner that furnishes the required information.

6. The power supply furnishes all a-c and d-c voltages necessary for the operation of the system components.

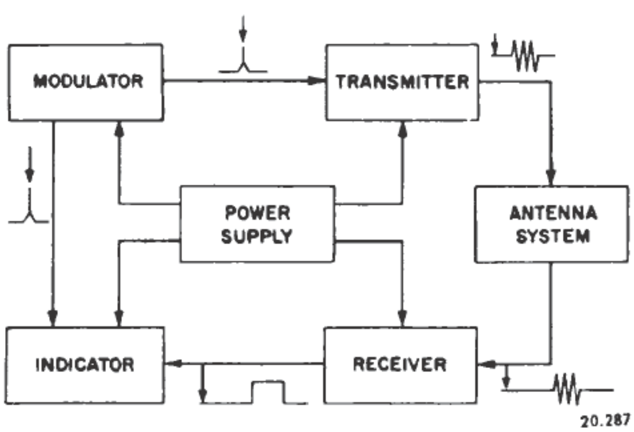


Figure 15-9.—Functional diagram of a fundamental pulse-modulated radar system.

PERFORMANCE CHARACTERISTICS OF A RADAR SYSTEM

Any radar system has associated with it certain specifications such as CARRIER FREQUENCY, PULSE-REPETITION FREQUENCY (the number of pulses sent out per second), PULSE WIDTH (in microseconds), and POWER RELATION (relationship of peak and average power). The choice of these arbitrary constants for a particular system is determined by its tactical use, the accuracy required, the range to be covered, the practical physical size, and the problem of generating and receiving the signal.

CARRIER FREQUENCY

The carrier frequency is the frequency at which the r-f energy is generated. The principal factors influencing the selection of the carrier frequency are the desired directivity and the generation and reception of the necessary microwave r-f energy.

For the determination of direction and for the concentration of the transmitted energy so that a greater portion of it is useful, the antenna should be highly directive. The higher the carrier frequency, the shorter the wavelength will be. Hence the antenna array is smaller for a given sharpness of pattern, because the individual radiating element is normally a half-wave long. For an antenna array of a given physical size the pattern is sharper for a higher frequency.

The problem of generating and amplifying reasonable amounts of r-f energy at extremely high frequencies is complicated by the physical construction of the tubes to be used. The common triode becomes impractical and must be replaced by tubes of special design.

In general, the modifications for extremely high-frequency operation are designed to reduce interelectrode capacitances, transit time, and stray inductance and capacitance in the tube leads.

At the receiver end, it is very difficult to amplify microwave signals; as a result, r-f amplifiers are not employed. Instead, the frequency of the incoming signal is mixed with that of a local oscillator in a crystal mixer to produce a difference frequency called the INTERMEDIATE FREQUENCY (i-f). The intermediate frequency is low enough to be amplified in suitable i-f amplifier stages employing electron tubes.

PULSE-REPETITION FREQUENCY

Sufficient time must be allowed between each transmitted pulse for an echo to return from any target located within the maximum workable range of the system. Otherwise, the reception of the echoes from the more distant targets would be obscured by succeeding transmitted pulses.

The range of a radar set depends upon the pulse repetition rate provided the power is sufficient. For example, when the peak power is sufficient, and the repetition rate is 250 PPS, the period will be $\frac{10^6}{250} = 4000\ \mu s$. At 12.2 μs per mile, the range will be $\frac{4000}{12.2} = 328$ miles. This necessary time interval fixes the highest pulse-repetition frequency that can be used to avoid interference with the returning echo by the next output pulse.

When the antenna system is rotated at a constant speed, the beam of energy strikes a target for a relatively short time. During this time, a sufficient number of pulses of energy must be transmitted in order to return a signal that will produce the necessary indication on the oscilloscope screen. For example, an antenna rotated at 6 rpm having a pulse repetition frequency of 800 cps will produce approximately 22 pulses for each degree of antenna rotation. The persistence of the screen and the rotational speed of the antenna therefore determine the lowest pulse repetition frequency that can be used.

PULSE WIDTH

The minimum range at which a target can ideally be detected is determined largely by the width of the transmitted pulse. If a target is so close to the transmitter that the echo is returned to the receiver before the transmitter is turned off, the reception of the echo obviously will be masked by the transmitted pulse. For example, a pulse width of 1 μs will have a minimum range of 164 yards, meaning that a target within this range will not show, or will be "blocked out" on the screen. In this respect, equipments for "close in" ranging or navigation work use pulses of the order of 0.1 μs. For long-range equipment the pulse width is normally from 1 μs to 5 μs.

POWER RELATION

A radar transmitter generates r-f energy in the form of extremely short pulses and is turned off between pulses for comparative long intervals. The useful power of the transmitter is that contained in the radiated pulses and is termed the PEAK POWER of the system. Power is normally measured as an average value over a relatively long period of time. Because the radar transmitter is resting for a time that is long with respect to the operating time, the average power delivered during one cycle of operation is relatively low compared with the peak power available during the pulse time.

A definite relationship exists between the average power dissipated over an extended period of time and the peak power developed during the pulse time. The overall time of one cycle of operation is the reciprocal of the pulse repetition frequency (PRF). Other factors remaining constant, the greater the pulse width the higher will be the average power; and the longer the pulse-repetition time, the lower will be the average power. Thus,

$$\frac{\text{Average power}}{\text{peak power}} = \frac{\text{pulse width}}{\text{pulse-repetition time}}$$

These general relationships are shown in figure 15-10.

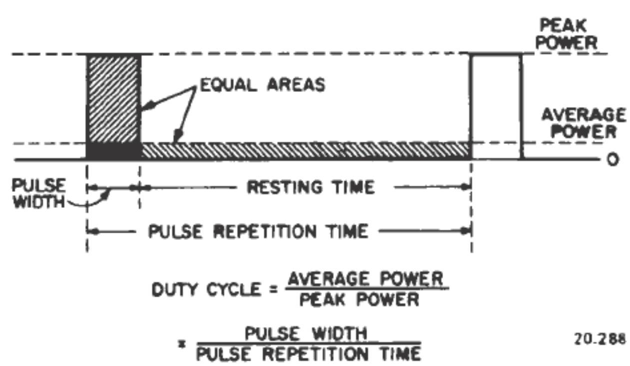


Figure 15-10.—Relationship of peak and average power.

The operating cycle of the radar transmitter can be described in terms of the fraction of the total time that r-f energy is radiated. This time relationship is called the DUTY CYCLE and may be represented as

$$\text{duty cycle} = \frac{\text{pulse width}}{\text{pulse repetition time}}$$

For example, the duty cycle of a radar having a pulse width of 2 microseconds and a pulse-repetition frequency of 500 cycles per second (pulse repetition time $= \frac{10^6}{500}$, or 2,000 microsecond) is

$$\text{duty cycle} = \frac{2}{2{,}000} = 0.001$$

Likewise, the ratio between the average power and peak power may be expressed in terms of the duty cycle. Thus,

$$\text{duty cycle} = \frac{\text{average power}}{\text{peak power}}$$

In the foregoing example it may be assumed that the peak power is 200 kilowatts. Therefore, for a period of 2 microseconds a peak power of 200 kilowatts is supplied to the antenna, while for the remaining 1998 microseconds the transmitter output is zero. Because

$$\text{average power} = \text{peak power} \times \text{duty cycle},$$
$$\text{average power} = 200 \times 0.001 = 0.2 \text{ kilowatts}$$

High peak power is desirable in order to produce a strong echo over the maximum range of the equipment. Low average power enables the transmitter tubes and circuit components to be made smaller and more compact. Thus, it is advantageous to have a low duty cycle. The peak power that can be developed is dependent upon the interrelation between peak and average power, pulse width and pulse-repetition time, or duty cycle.

ELEMENTARY RADAR TRANSMITTER AND RECEIVER

The function of the modulator is to ensure that all circuits connected with the radar system operate in a definite time relationship with each other and that the interval between pulses is of the proper length. In general, there are two practical methods of supplying the timing requirements—timing by means of a separate unit and timing within the transmitter.

A separate timing source may be used to give rigid control of the pulse-repetition frequency. In this case the source consists of any stable type of audio oscillator such as the Wien-bridge oscillator. The output is then applied to the necessary pulse-shaping circuits to produce the required timing pulse. Figure 15-11 shows in block form the functional components associated with the timer. These include the oscillator and other stages and components that are necessary to generate, shape, and amplify the waveform so that it may properly trigger the magnetron in the transmitter.

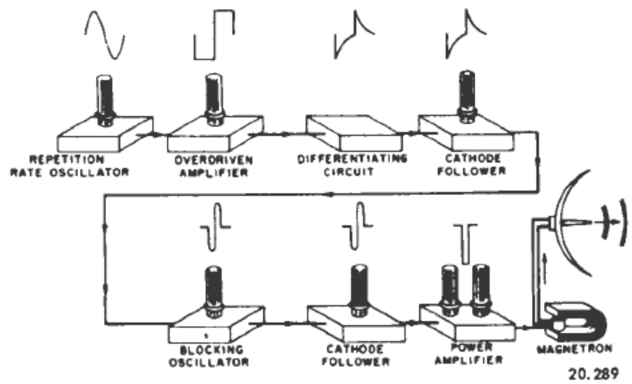


Figure 15-11.—Simplified block diagram of a modulator and transmitter.

The oscillator generates a steady output at a given frequency (usually any frequency between 625 and 650 cps and generally less than 1000 cps), and this output establishes the PRF of the set.

The sine wave output of the oscillator is of the correct frequency but it does not have the correct shape and its amplitude is insufficient to fire the magnetron. Therefore, the signal is changed first into a square wave in the overdriven amplifier stage. The square wave is sharpened into a peaked wave in a differentiating circuit (a resistor and capacitor in series with the input, and the output taken across the resistor) and fed via a cathode follower to a blocking oscillator.

The blocking oscillator is triggered at the correct frequency by this peaked wave. The blocking oscillator generates the type of square wave needed by the magnetron, except that it is of insufficient amplitude.

The square-wave signal generated by the blocking oscillator is fed via a cathode follower to the power amplifier (preceded in actual circuits by driver amplifiers) where the square-wave pulse is amplified sufficiently to drive the magnetron. Only the negative portion of the pulse is used to drive the magnetron oscillator, and therefore the positive portion of the pulse is removed.

The magnetron goes into oscillation the moment it is triggered by the negative-going square wave from the power amplifier. The frequency of the magnetron oscillation may be of the order of 6500 megacycles. The width of the pulse is determined by the width of the negative-going pulse from the power amplfier and may be of the order of 1 microsecond. During the pulse, the power output may be of the order of 125 kw.

TRANSMITTER

The transmitter is basically an r-f oscillator. It may be turned on and off by the negative-going pulse from the modulator. The radar oscillator (in this instance a magnetron) differs from other oscillators treated in chapter 8 in that it produces a much higher frequency and has a much higher power output. The higher frequency permits smaller waveguides and antennas to be used; and the higher power permits stronger echoes and a greater useful range.

Because of the superhigh frequencies in a radar set, buffers, frequency multipliers, and power output tubes following the magnetron would have little value in increasing the output power, and hence are not used in a radar set.

The more powerful sets are capable of putting out 1 megawatt (1000 kw) of peak power. A simplified diagram of a magnetron is shown in figure 15-12,A. The magnetron is essentially a diode that has its plate at ground potential and its cathode at a high negative potential during the time it is oscillating. The diode is placed in a powerful magnetic field produced by a permanent magnet.

When a negative pulse is applied to the cathode and there is no magnetic field present, electrons move in straight lines from the cathode to the plate, as shown in part 1 of figure 15-12,B. When a weak magnetic field (part 2) is applied, the electron paths become curved; and as the magnetic field becomes stronger (parts 2, 3, 4, and 5), the electron paths become progressively more curved. Finally, the paths become so curved that the electrons are moving in closed circular orbits that miss the plate entirely, and no plate current flows. The plate in figure 15-12,C, is a copper cylinder the internal surface of which is separated into a number of segments by holes in the cylinder that serve as

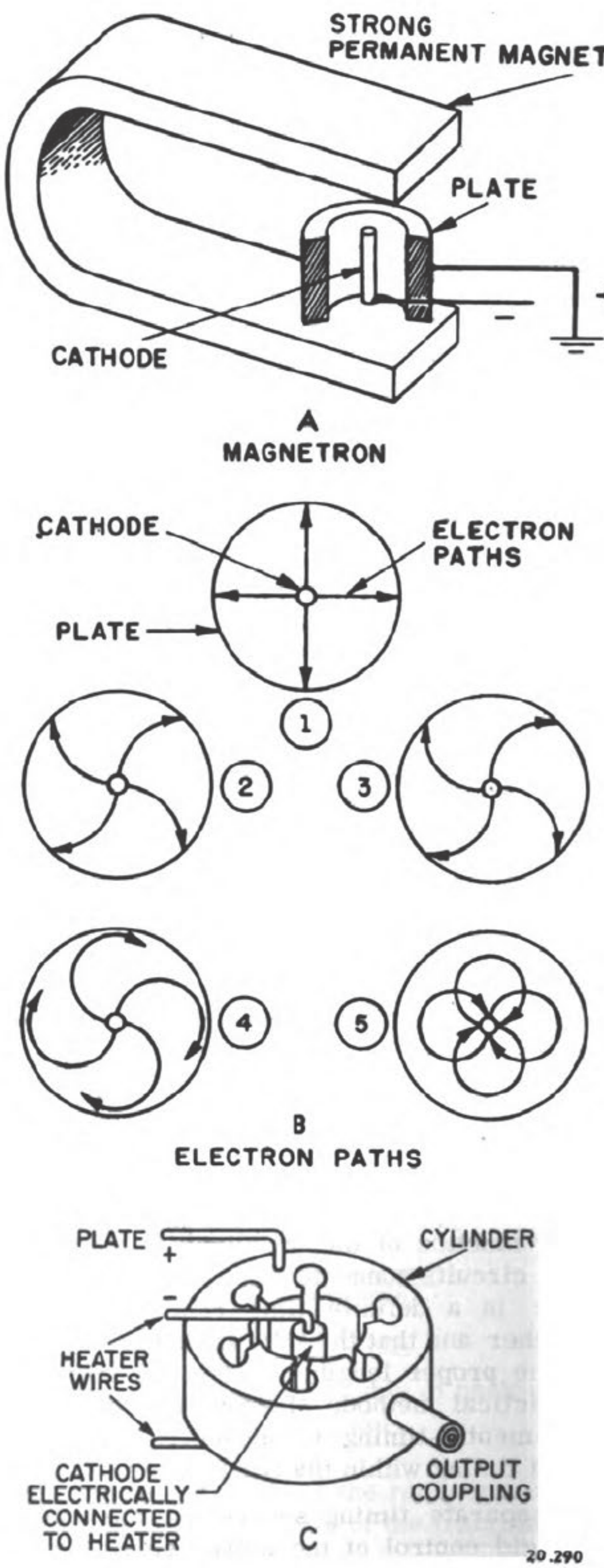


Figure 15-12.—Simplified diagram of a magnetron.

tuned circuits. As the electrons move in circles past the plate segments they induce currents electrostatically in the walls of the holes. The energy of the magnetron output pulse is contained in the field associated with these currents. The smaller the circles the electrons make, the higher is the frequency of the oscillations. The frequency depends on the size of the cylinder, the strength of the magnetic field, and the difference in potential between the cathode and plate.

Energy is coupled out of the magnetron by means of a loop or probe; it is then transmitted to the antenna via a waveguide.

The tremendous peak power produced in short pulses by the magnetron requires high plate-to-cathode potential and high cathode emission. Because of the relatively long resting time between pulses, the problem of cooling is reduced and the physical size of the magnetron is not as large as would be expected from the peak power rating.

TRANSMITTING AND RECEIVING ANTENNA SYSTEM

One function of an antenna system is to take the energy from the transmitter, radiate that energy in some chosen manner (by using a directional system when bearings are desired but by using a nondirectional antenna system where a bearing indication is not necessary). Another function of an antenna system is to pick up the returning echo, pass it to the receiver with a minimum of losses.

Some original radar installations contained two separate antenna systems: one for transmitting and one for receiving. The more practical radar system uses a single antenna system and an electronic switch capable of rapidly shifting the antenna performance from transmit to receive functions and vice versa. The switch is needed to protect the receiver from damage by the potent transmitter energy during the pulse time and, also, to keep the transmitter from absorbing some (or all) of the very weak echo during the receiving time.

Nondirectional Antenna

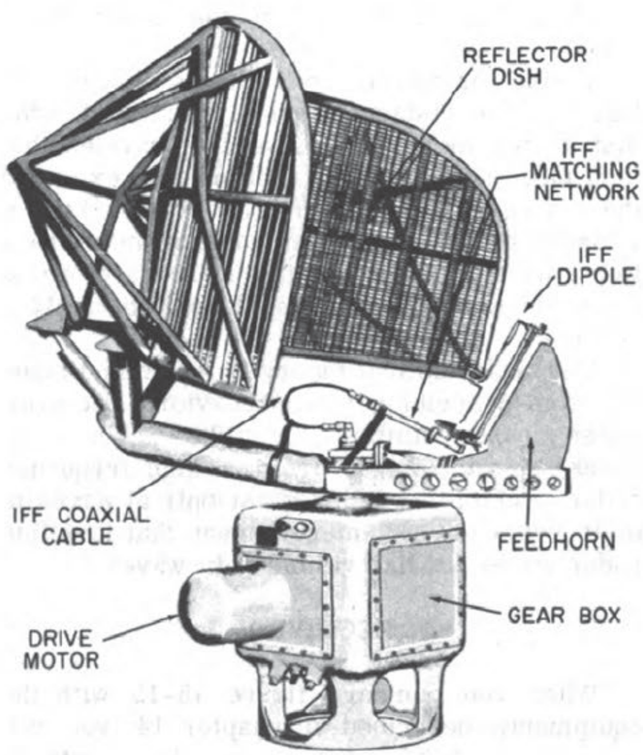
Some applications of radar can use a simple nondirectional antenna, for example, the vertical dipole. Nondirectional antennas are used in navigation aids, as radar beacons (called racon), and some forms of IFF equipment.

Directional Antenna

That radar system which indicates the bearings of targets must have some means of pointing its radiated energy in known directions. Practically all such radar systems accomplish this by constant 360° rotation of a motor-driven energy-transfer device such as an antenna, waveguide, reflector or director, or energy feedhorn.

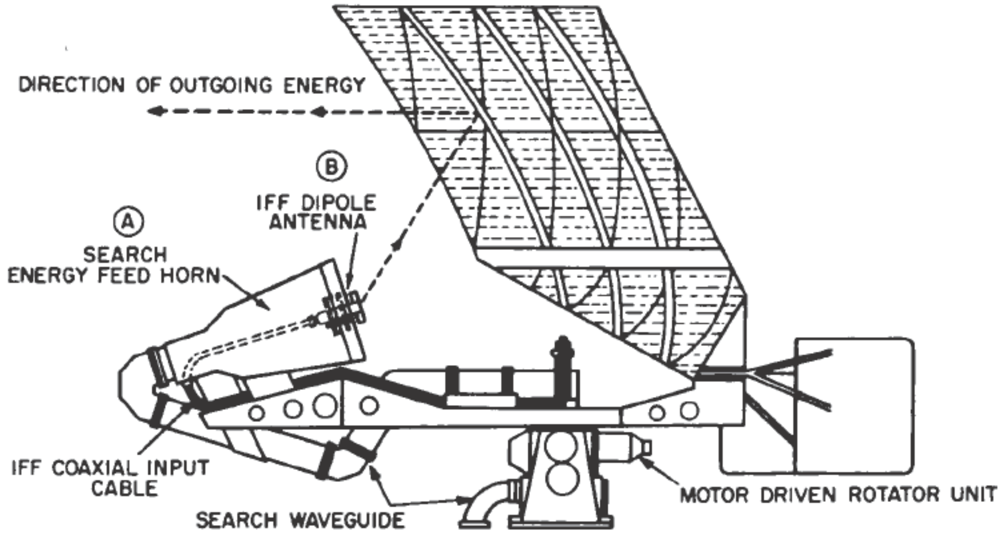
Each antenna type has abilities to couple and project electromagnetic energy into space; also each has an ability to convert received energy into the forms that activate receiver equipments.

Combining the characteristics from several antenna types results, obviously, in an improved system. Thus, a reflector may be added as shown in figure 15-13, and driven by motor for continual 360° rotation, to direct concentrated energy toward the horizon. This provides a highly-directive antenna system for use at radar frequencies in VHF region and above. Figure 15-14 shows two antenna feed units at the focus; a



20.291

Figure 15-13.—Radar antenna with reflector.



20.292

Figure 15-14.—This section of reflector dish is fed from two kinds of energy-transfer devices, (A) Energy feedhorn and (B) dipole antenna.

dipole antenna for IFF purposes and flared feedhorn for search purposes. This dish is a section of a parabola.

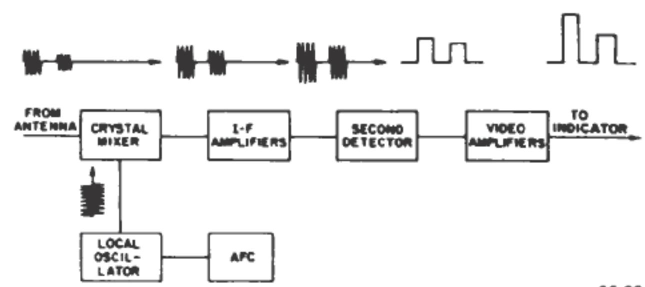
Experience shows that a parabolic dish, when properly focused for projecting energy, will also serve at its best for accepting echo energy from space and returning it into the transmission system.

If the parabolic reflector is sufficiently large so the distance from any point within the dish to the focal point is several wavelengths, then QUASI-OPTICAL conditions will exist and the emerging wave is a narrow beam. Sizes of reflectors, which are practicable for microwave work, have a diameter of 10 to 20 wavelengths to produce a beam width of approximately 5 degrees.

The quasi-optical theory is mentioned many times in describing radar behavior. The word quasi means "similar" or "like." When you speak of microwaves from a high-frequency radar transmitter being quasi-optical waves in their behavior, you merely mean that invisible radar waves act like visible light waves.

RECEIVER

When you compare figure 15-15 with the equipments described in chapter 14, you will recognize that the radar receiver is essentially a special type of superheterodyne receiver. Its function is to receive the weak echoes from the antenna system, combine them in a crystal mixer (half-wave crystal rectifier) with the r-f signals from a local oscillator, amplify the resultant i-f signal, detect the pulse envelope, amplify the resulting d-c pulses, and feed them to the indicator. At the higher frequencies used in radar, it is not possible to use a stage of r-f amplification ahead of the mixer, and therefore, the r-f signals are fed directly to the mixer. This converter is shown in figure 15-16.



20.293

Figure 15-15.—Block diagram of a radar receiver.

In order to keep radar receivers in tune with their companion transmitters, a system of automatic frequency control is used in the receivers. Briefly, the system functions as follows: A small fraction of the r-f energy from the transmitter line is applied to a special automatic-frequency-control (a-f-c) mixer along with a small fraction of the r-f energy

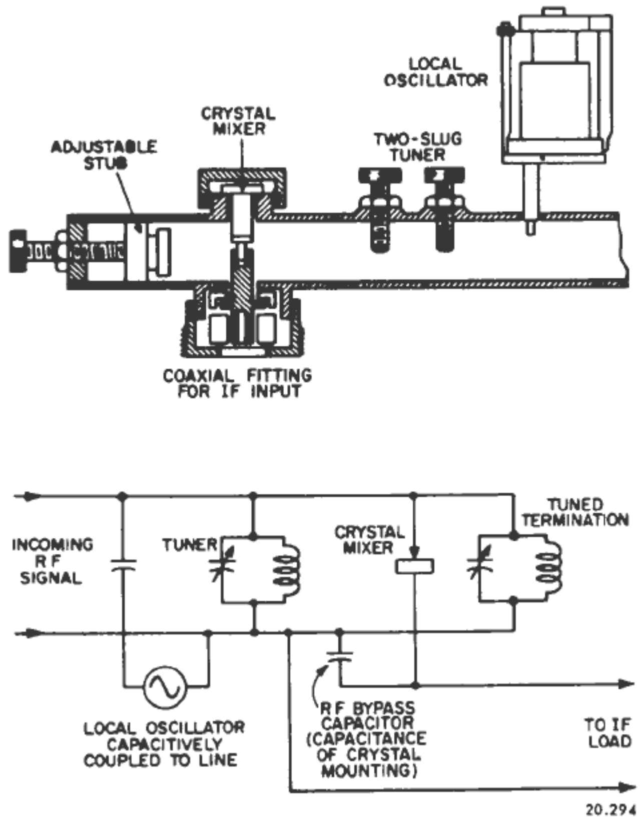


Figure 15-16.—Waveguide frequency converter which is part of the receiver equipment.

from the receiver local oscillator. The i-f energy resulting from the mixing of these two frequencies is amplified, rectified, and applied via control circuits in such a way as to tune the local oscillator. If the i-f is of the correct frequency, the resulting direct voltage maintains the local oscillator at the correct oscillator frequency. If the i-f is too low in frequency the direct voltage applied to the local oscillator causes it to shift in frequency so that the i-f will be increased. If the i-f is too high, the oscillator frequency is shifted in the opposite direction.

The stability of operation is maintained in the microwave range of frequencies by careful design; and the overall sensitivity of the receiver is greatly increased by the use of many i-f stages. Special types of tubes having low interelectrode capacitances also have been developed for use in local-oscillator and i-f stages. Re-examine the block diagrams of a radar receiver which is shown in figure 15-15. As in communications receivers, the i-f signals in a radar receiver are fed to the second detector where the signal is rectified and the i-f component is removed. The remaining modulation pattern, consisting of d-c pulses, is fed to a video amplifier. In one type of presentation the output of the video amplifier is fed to the vertical deflection plates of an electrostatic-type cathode-ray tube. The amplitude of the vertical trace formed on the screen is proportional to the strength of the received signals. Simultaneously, a sawtooth voltage is applied to the horizontal deflection plates in synchronism with the transmitted pulse. The sawtooth voltage provides a horizontal displacement that is proportional to range.

Radar video amplifiers have wide band frequency response similar to that of television video amplifiers.

POWER SUPPLY

In the functional diagram of the radar system (fig. 15-9) the power supply is represented as a single block. Functionally, this block is representative; however, it is unlikely that any one power supply could meet all the power requirements of a radar set. The distribution of the physical components of the system muy be such as to make it impractical to lump the power-supply circuits into a single physical unit. Thus, different supplies are needed to meet the varying requirements of the system and must be designed accordingly. The power-supply function is performed, therefore, by various types of supplies distributed among the circuit components of the radar equipment. Power supplies are treated in chapter 3.

CHAPTER 16

SPECIAL CIRCUITS

In order to GENERATE and RESHAPE the waves and to CONNECT one circuit to another in such a way as to cause the least disturbance to either circuit, several special circuits are needed. These circuits are listed and briefly described here, but are treated in detail in other chapters of this text and as needed in the rating books.

GENERATING CIRCUITS

A generating circuit is one that produces oscillations of a given form and frequency.

In most instances the generating circuits for radar purposes are the familiar sine wave oscillators that are discussed in chapter 8. The generated output of the sine wave oscillators may be put into shaping or reshaping circuits, such as clipping, clamping, limiting, and peaking stages to produce other essential waveshapes such as square waves, peaked waves, and sawtooth waves. Those stages require coupling stages (connecting circuits) as cathode followers and phase inverters which are explained in chapter 6.

NONSINUSOIDAL WAVES

Pure sine waves are basic waveshapes. We shall show that sine waves are the building blocks for the nonsinusoidal waves used in radar and other electronic devices. Let us return to figure 15-11 to inspect the waveshapes associated with each device, beginning with the oscillator. That waveshape is a sine wave. Next is a square wave, obtained from the overdriven amplifier in this instance. Spiked waves (also called peaked waves) are obtained from the output of the differentiating circuit. Such waves are classed as nonsinusoidal shapes to distinguish them from the well-known sine waves. You will learn that a complex periodic wave is composed of a fundamental and different harmonics. The shape of the resulting wave depends on harmonics that are present, their relative amplitudes, and relative phase relationship. In general, the steeper the sides of the waveshape, that is, the more rapid its rise or fall, the more harmonics it contains.

Composition

Any periodic wave (one that repeats itself in definite time intervals) is composed of sine waves of different frequencies and amplitudes, added together. The sine wave which has the same frequency as the complex periodic wave is called the fundamental. The fundamental corresponds to the first harmonic.

The frequencies higher than the fundamental are called harmonics, provided: The harmonics are always a whole number of times higher than the fundamental, and are designated by this integer (whole number). For example, the frequency twice as high as the fundamental is the second harmonic.

Square Wave

Figure 16-1,A, compares a square wave with a sine wave A of the same frequency. Concentrate your attention only for the moment on these (by ignoring the presence of waves B and C) and note they are considerably different. If another sine wave B of smaller amplitude, but three times the frequency, called the third harmonic, is added to A, then we get a new curve; the resultant is curve C. Now compare curve C with the desired square wave. The appearance of the resultant more nearly approaches that of the square wave. The resultant is shown again as curve C in figure 16-1,B. When the fifth harmonic is added, the sides of the new resultant are steeper than before. The new resultant is curve E in figure 16-1,B, and is carried down to figure 16-1,C. Addition of the seventh harmonic, of smaller amplitude, makes the sides of the composite curve steeper than any previous curve. Addition of more and more odd harmonics

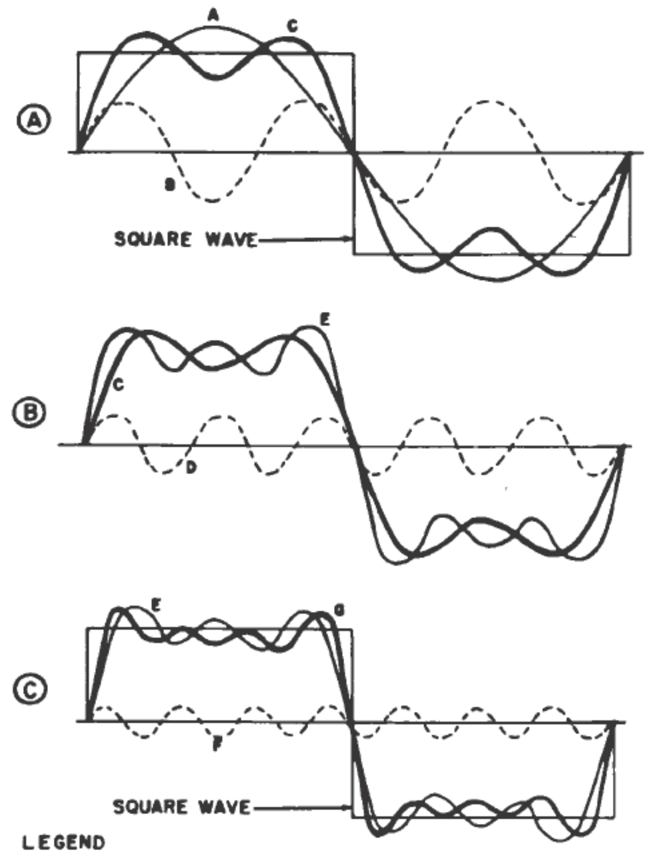


Figure 16-1.—Composition of a square wave.

brings the resulting wave nearer to the desired square wave. In the square wave composition, all the odd harmonics cross the zero reference line in phase with the fundamental. The resulting wave becomes an exact square wave if an infinite number of odd harmonics is added.

Sawtooth Wave

Similarly, a sawtooth wave is made up of different sine waves. First the second harmonic of smaller amplitude is added to the fundamental. The resultant is shown as curve C in figure 16-2,A and it is shown again as curve C in figure 16-2,B. You see that the crest of the resultant is already pushed to one side. Next, the third harmonic is added. The new resultant

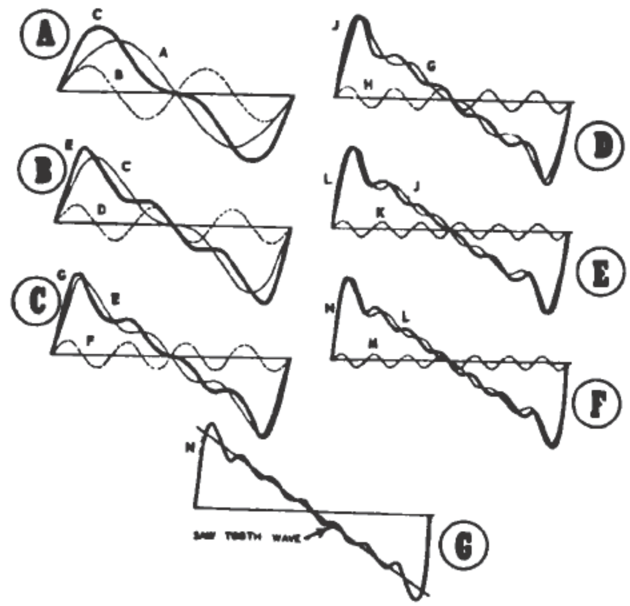


Figure 16-2.—Composition of a sawtooth wave.

is the curve E in figure 16-2,B as well as in figure 16-2,C. The peaks are pushed farther to the side. This is carried on through the other sections of figure 16-2 by adding the fourth, fifth, sixth, and seventh harmonics in turn. In the sawtooth wave composition, all the even and odd harmonics cross the zero reference line in phase with the fundamental. As each harmonic is added, the ensuing resultant more nearly resembles the sawtooth wave.

Peaked Wave

Figure 16-3 shows the trend in composition of a peaked wave. Notice how the addition of each odd harmonic makes the peak of the resultant higher and the sides steeper. The composition of a peaked wave differs from that of a square wave in that there is a different phase relationship between harmonics. In the

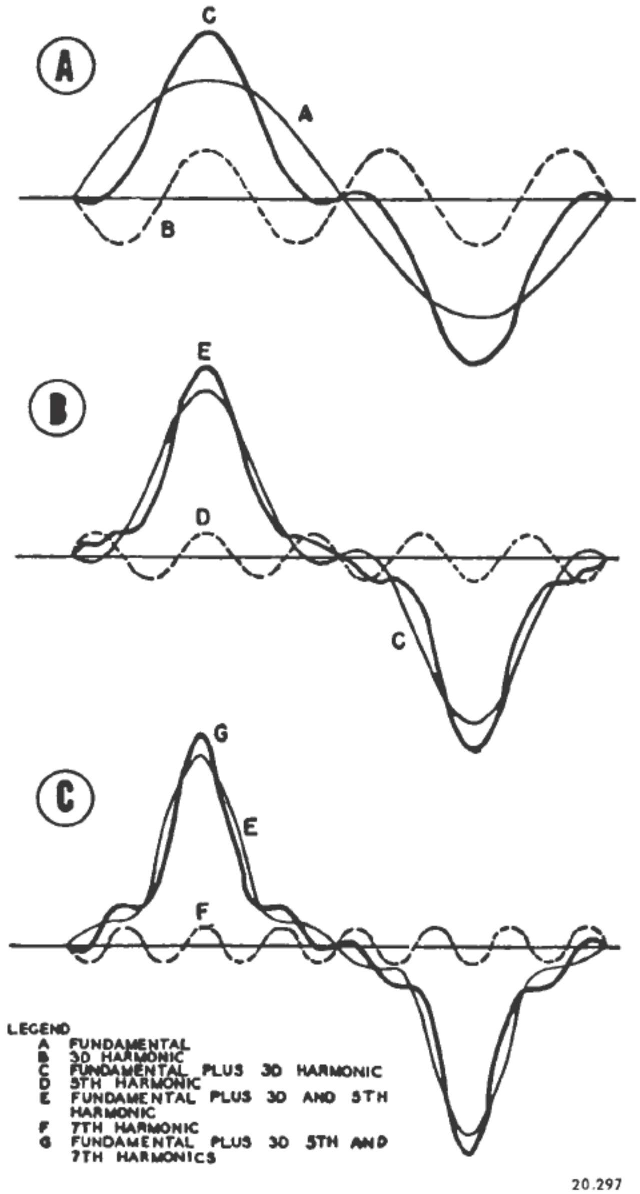


Figure 16-3.—Composition of a peaked wave.

square wave composition all the odd harmonics cross the zero line in phase with the fundamental. In the peaked wave composition the 3rd, 7th, 11th, and so forth, harmonics cross the zero line 180° out of phase with the fundamental while the 5th, 9th, 13th, and so forth, cross the zero line in phase with the fundamental.

REPETITION-RATE OSCILLATOR

The repetition-rate oscillator determines the pulse repetition rate of a radar system. The Wien-bridge oscillator is widely used because of its reliable performance under strenuous applications over long periods. The action of this oscillator is covered in chapter 17 where you learn the phase-shifting element is a frequency-selective bridge. The capacitor and resistor elements of the bridge circuit are chosen to produce a desired pulse repetition rate (PRR).

SAWTOOTH GENERATOR

The sawtooth generator is described in chapter 17. It produces waves that resemble the shape of saw teeth. This type of generator may be operated continuously to produce the sweep voltage for use with the electrostatic type of cathode-ray tube. This generator may be triggered also by a sharp voltage plus from another circuit as used in the synchroscope. Those applications are covered in chapter 17.

MULTIVIBRATOR

The multivibrator is used to produce square waves of the desired frequency. Like the sawtooth generator it may operate continuously, or it may be triggered into operation periodically by sharp voltage pulses. The multivibrator is described in chapter 17. The square waves that trigger the magnetron are produced by a multivibrator that is in turn triggered by the timing pulse.

MAGNETRON

The magnetron oscillator is also a generating circuit. It produces the high-frequency oscillations at sufficient power (during pulses) to properly "illuminate" the range area covered by the equipment. See figure 15-12.A.

LOCAL OSCILLATOR

The local oscillator in the radar receiver generates high-frequency oscillations, which, when mixed with the incoming pulse frequency, produce the intermediate frequency. This type of oscillator at superhigh frequencies (3000 mc and above) consists of a reflex velocity-modulated tube (klystron) tuned by cavity resonators. These are described inappropriate rating texts.

RESHAPING CIRCUITS

A reshaping circuit is one that takes the waveform from a generating circuit and shapes it according to the needs of the system.

LIMITING CIRCUIT

Limiting, or clipping, circuits are employed to change the shape of the wave by clipping the top or bottom of the wave, or both. This may be accomplished by operating the grid with a small bias or a large bias or by overdriving a conventional amplifier. This type of amplifier is treated in chapter 14. The effects of too little or too much bias are treated in chapter 5.

The term, limiting, refers to the removal by electronic means of one extremity or the other of an input waveform. Circuits that perform this function are called limiters, or clippers.

Limiters are useful in wave-shaping circuits where it is desirable to square off the extremities of the applied signal. A sine wave may be applied to a limiter circuit to obtain a rectangular wave; a peaked wave may be applied to a limiter to eliminate either the positive or the negative peaks from the output. In frequency-modulation receivers, where it is necessary to limit the amplitude of the signal applied to the detection system to a constant value, limiter circuits are employed. Limiters also are used as protective devices in circuits in which the input voltage to a stage must be prevented from swinging too far positive or negative.

Diode Limiting

A diode conducts only when its plate is positive with respect to its cathode. A conducting diode has a relatively low resistance, and the drop across it is low and, in general, is neglected. In the following examples of limiting, the effect of load current on the limiting action is also neglected.

The SERIES DIODE LIMITER (diode in series with the load) shown in figure 16-4, may be used for positive or negative limiting, depending on the way the diode is connected into the circuit. In each of the circuits the resistance of R is relatively large compared with the resistance of the conducting diode. The upper input and output terminals arbitrarily are considered to be positive with respect to ground when the input voltage wave is positive. On the following half cycle, the upper terminals are negative with respect to ground when the input voltage wave is negative. The circuit is similar to a half-wave rectifier.

With positive limiting, the output voltage remains at zero throughout the positive half cycle of the input because no current can flow through R. During the negative half cycle, however, the cathode is negative with respect to the plate, and the tube conducts. Except for the small voltage drop, e_p, across the tube, the output waveform follows the input waveform.

In negative limiting, the diode connections are reversed from what they are in positive limiting. The action is essentially the same except that the negative half of the waveform is limited (not passed).

An alternate method of using diodes in limiter circuits is shown in figure 16-5. The name PARALLEL DIODE LIMITING is derived from the circuit connection—that is, from the fact that the diode is connected in parallel with the load.

In positive limiting, the diode conducts during the positive half of the cycle (the plate is positive with respect to the cathode) and places virtually a short circuit across the load. Only the small voltage, e_p, appears at the output. During the negative half of the cycle the tube is an open circuit (the plate is negative with respect to the cathode), and the signal appears at the output as a negative half cycle.

In negative limiting, the diode connections are reversed with respect to their connections in positive limiting. The action is essentially the same except that the negative half of the waveform is limited.

The input voltage of a limiter can be limited (in the output) to any desired positive or negative value by holding the proper diode electrode to that voltage by means of a battery or biasing resistor. Two such circuits are illustrated in figure 16-6.

In positive limiting, the cathode is more positive than the plate by the value of E when no signal is applied to the input. As long as e_{in} remains less positive than E, the diode is an open circuit and the output waveform follows the input waveform. When e_{in} exceeds E on

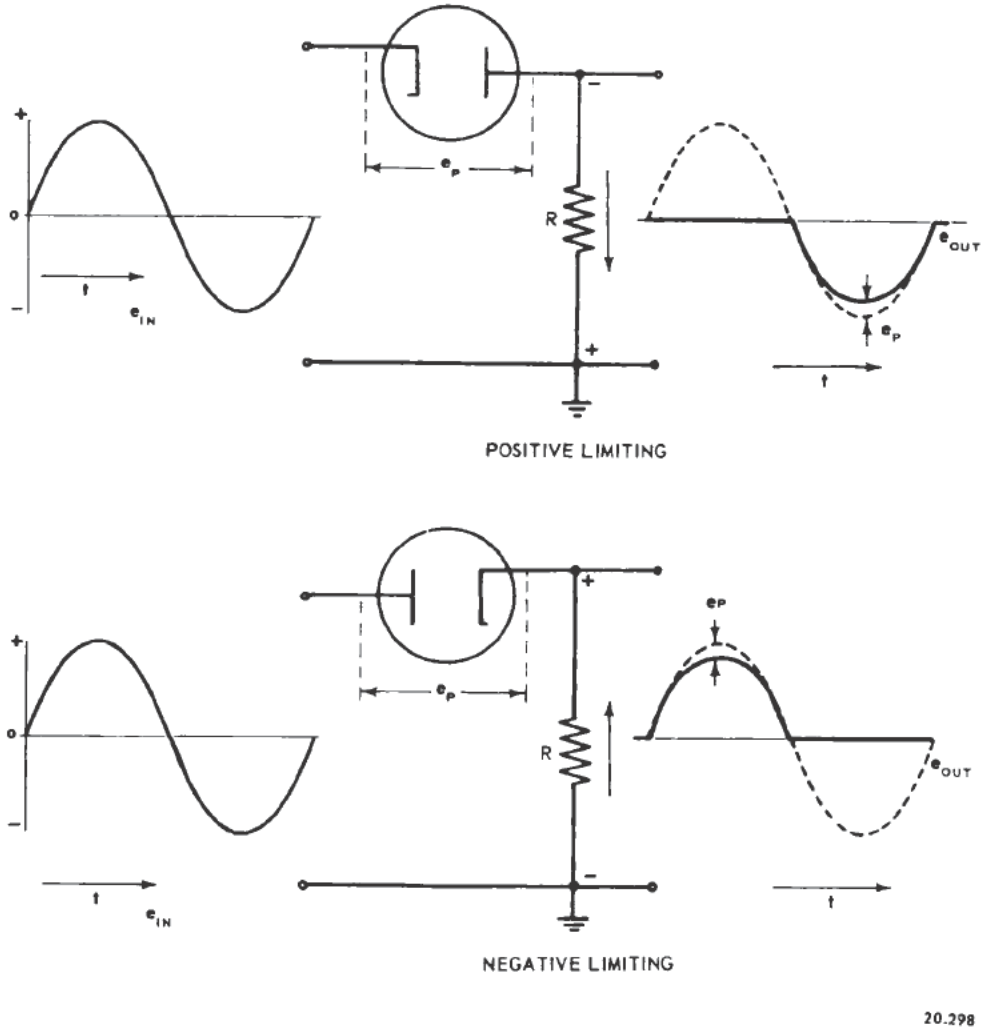


Figure 16-4.—Series diode limiting.

the positive half cycle, the diode conducts (the output is then connected effectively across E), and the output is limited to E volts above ground. The difference between e_{in} and E appears as an IR drop across R when the diode conducts. During the negative cycle the diode cannot conduct, and the output waveform follows the input waveform.

In negative limiting, the plate of the diode (under no-signal conditions) is more negative than the cathode by the value of E. Thus, as long as the input is positive or less negative than E the diode is an open circuit, and the output voltage waveform follows the input voltage waveform. When the input becomes more negative than E, the diode conducts, and the output terminals are in effect connected across the battery. The difference between e_{in} and E appears as an IR drop across R when the diode conducts.

It is sometimes desirable to pass only the positive or negative extremity of a waveform to a succeeding stage. This may be accomplished by the use of one of the circuits illustrated in figure 16-7.

The circuit illustrated in the upper part of the figure retains the negative peaks of the input waveform; that is, the negative peaks are passed to the output. During the positive half cycle the diode conducts and the output is E because the output terminals are, in effect, connected across the battery. The difference between e_{in} and E appears, of course, as an IR drop across R. This condition prevails until the negative peak arrives. During the time that the negative peak is greater than E,

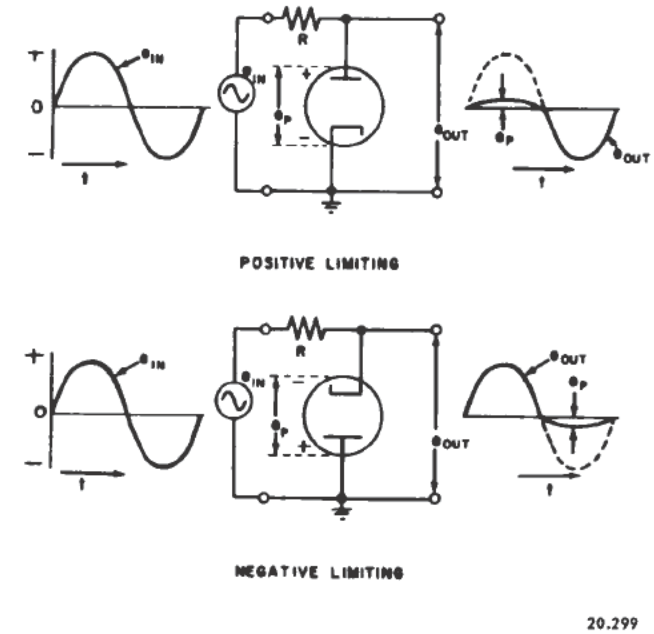


Figure 16-5.—Parallel diode limiting.

the diode ceases to conduct and the negative peak appears at the output.

The circuit in the lower part of figure 16-7 retains the positive peaks of the input waveform. During that portion of the positive half cycle, when e_{in} is less than E, the diode conducts because its plate is positive with respect to its cathode, and the output is therefore E. However, when the positive-going input signal exceeds E, the tube is cut off (the cathode is positive with respect to the plate), and the positive peak appears across the output. During the negative half cycle the diode conducts and the output is E. Again, the difference between e_{in} and E appears as an IR drop across R.

Double-diode limiters are used to limit both positive and negative amplitude extremities. A circuit connected to provide this type of limiting is illustrated in figure 16-8.

Diode V1 conducts whenever the positive-going input signal exceeds E1, thus limiting the positive output to the value of E1. This results from the fact that V1, in effect, connects the output terminals across E1; V2 is, of course, nonconducting, or an open circuit, during this time. The difference between e_{in} and E1 appears as an IR drop across R.

Diode V2 conducts whenever the negative-going input signal exceeds E2, thus limiting the output to the value of E2 during the negative half cycle. During this time, the output terminals are connected, in effect, across E2; and V1 is, of course, an open circuit.

This circuit represents a simple method of producing a square-wave output from a sine wave input voltage.

Grid Limiting

The grid-cathode circuit of a triode, tetrode, or pentode may be employed as a limiter circuit in exactly the same way as the plate-cathode circuit of the diode limiter illustrated in the upper part of figures 16-6 and 16-7. By inserting a series grid resistor (fig. 16-9,A) that is large compared with the grid-to-cathode resistance when grid current flows, essentially the entire positive half cycle of the input voltage is limited to almost zero. For example, the grid-to-cathode resistance may drop from an infinite value, when the grid is negative with respect to the cathode, to a value of the order of 1000 ohms when the grid becomes positive with respect to the cathode. If a one-megohm resistor is placed in series with the grid, the voltage drop across the 1000-ohm R_{gk} is negligible compared with that which is developed across the one-megohm resistor by the flow of grid current.

The grid limiter circuit shown in figure 16-9,A, is held normally at zero bias. During the positive portion of the input signal the grid attempts to swing positive. Grid current flows through R, developing an i_gR drop of such polarity as to oppose the positive input voltage. The larger R is with respect to R_{gk}, the smaller will be the relative voltage across R_{gk} when grid current flows. The drop across R may be considered as an automatic bias developed during that part of the input cycle when grid current flows.

Alternate circuits for limiting the positive peaks of the input voltage are shown in figure 16-9,B and C. In part B the tube is biased by the negative potential, E, supplied to the grid. No grid current flows until e_{in} rises sufficiently to equal and effectively remove the biasing voltage, E. Any further rise of e_{in} drives the grid positive with respect to the cathode, and grid current through R limits the signal on the grid by virtue of the voltage drop across R.

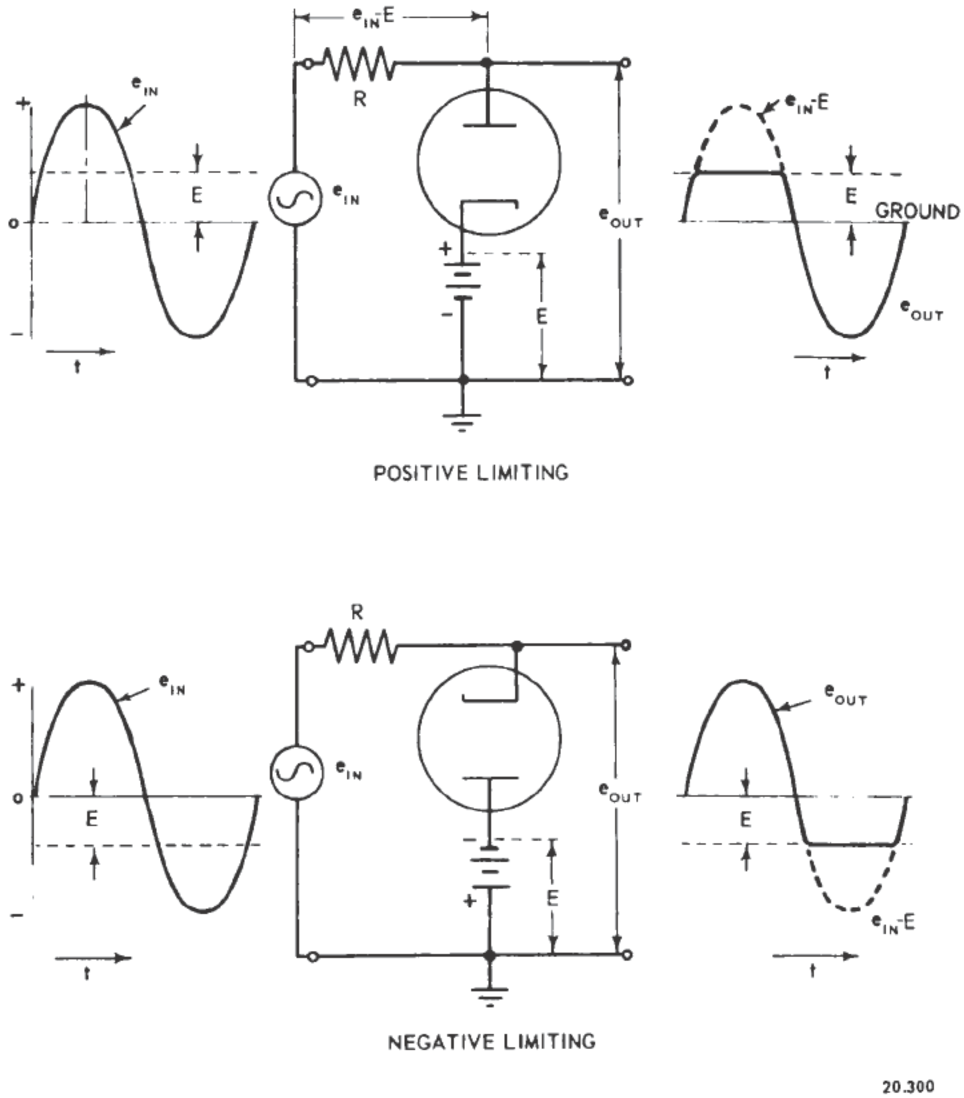


Figure 16-6.—Parallel diodes limiting above and below ground potential.

In figure 16-9,C, bias is developed between grid and cathode by the flow of PLATE current through R_k. Any positive signal, e_{in}, must drive the grid positive by an amount equal to the value of E_k before the biasing effect of R_k is removed. A further rise of the input voltage produces grid current, and this results in the limiting of the voltage at the grid due to the drop across R. C_k charges up to the value, E_k, and holds the voltage at this level over the entire cycle.

Saturation Limiting

When a series-limiting resistor (fig. 16-9) is used in the grid circuit, the grid cannot be driven to an appreciably positive voltage, and, despite the positive amplitude of the input voltage, the maximum plate current that flows is that determined by the plate supply and the resistance of the plate circuit at approximately zero bias. Thus, the minimum plate voltage is determined by the limiting action in the grid circuit. The grid voltage, plate current, and plate voltage relationships are illustrated in figure 16-10. Grid limiting occurs during the first half cycle.

The grid-limiting resistor may be omitted if the signal comes from a low-impedance, high-power source, and limiting in the plate circuit may still be realized. This is due to plate-circuit saturation and is usually referred to as saturation limiting. Plate-current saturation should not be confused with emission

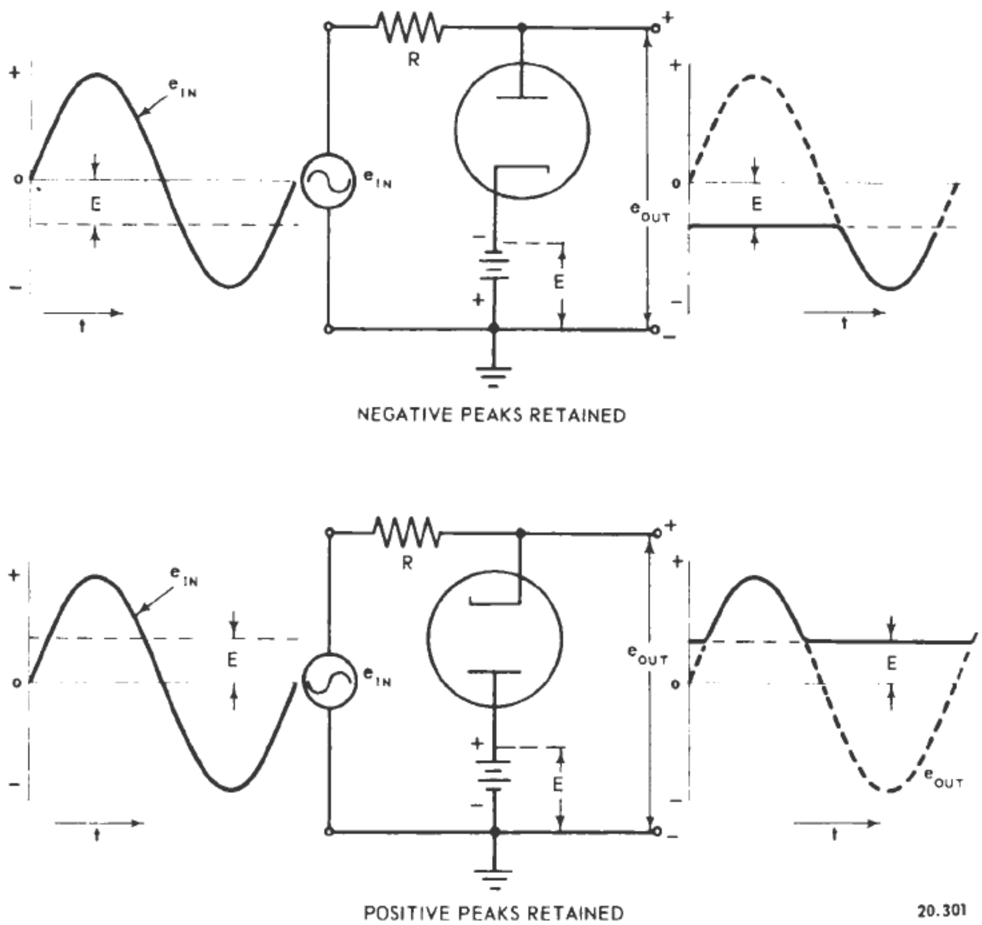


Figure 16-7.—Parallel diode limiters that pass peaks only.

saturation because in tubes using oxide-coated cathodes there is no definite saturation value of emission current. Also plate-current saturation limiting differs from grid-limiting in that the grid voltage is not limited in plate-current saturation limiting.

By using a large value of plate-load resistance R_L and a low value of plate-supply voltage

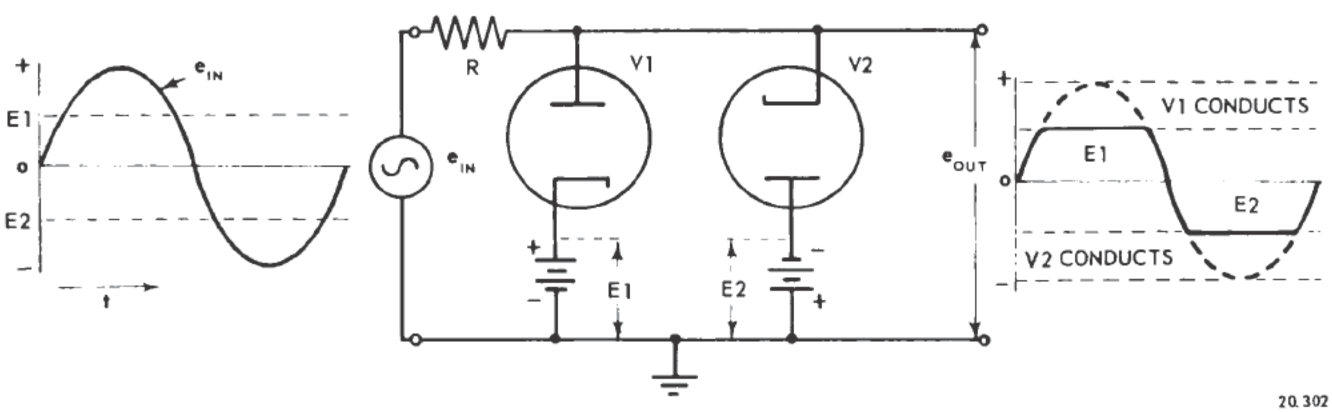
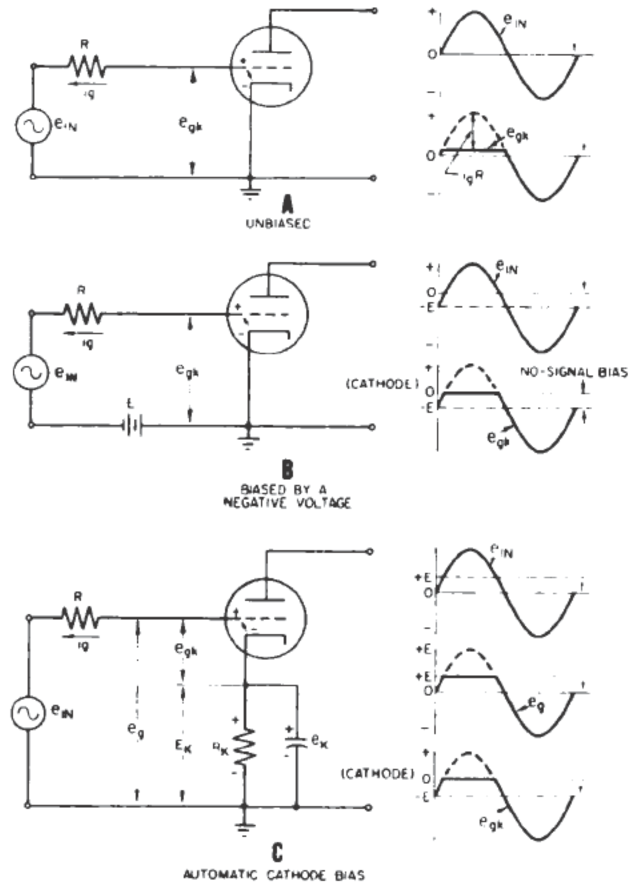


Figure 16-8.—Double-diode limiter circuit.



20.303

Figure 16-9.—Grid limiters.

E_B, saturation limiting may be produced by a relatively low amplitude of positive grid voltage. Limiting occurs in the plate circuit only. In any case, however, the plate current can never exceed the value of $\frac{E_B}{R_L}$. In an actual circuit some small positive voltage must remain on the plate to attract electrons from the cathode, and the saturation plate current never quite equals $\frac{E_B}{R_L}$. In other words, there remains across the tube a low-voltage drop when the plate current is at saturation because the plate to cathode resistance at saturation does not decrease to zero.

Cutoff Limiting

Electron current through an electron tube can flow only from cathode to plate; it cannot flow from plate to cathode. When the grid of an electron tube is driven to cutoff, the plate current is decreased to zero and remains at zero during the time the grid is below cutoff. Because no current flows in the plate circuit when the tube is cut off, no voltage is developed across the load resistance, and the plate is maintained at the full value of the plate supply voltage. Thus, a type of limiting is achieved in which the positive extreme of the plate waveform is flattened as a result of driving the grid beyond cutoff. Cutoff limiting is illustrated in the second half cycle (fig. 16-10).

The cutoff voltage, E_{co}, may be defined as the negative voltage, with respect to the cathode, to which the grid must be driven in order to prevent the flow of plate current. For any given type of tube this voltage level is a function of the plate-cathode voltage and in the case of triodes may be approximated by the expression:

$$E_{co} = \frac{E_p}{\mu}$$

where E_p is the plate-cathode voltage and μ is the amplification factor of the tube. This relationship is NOT valid in the case of tetrodes and pentodes.

Overdriven Amplifier

An amplifier circuit in which saturation limiting is employed in conjunction with cutoff limiting to produce a rectangular waveform from a sine waveform is known as an overdriven amplifier (fig. 16-10). The driving circuit for such an amplifier should have a relatively low output impedance and be capable of delivering power because considerable current is drawn during the positive swing of the grid voltage. The value of the load resistor is made as large as practicable for the plate voltage available.

Limiter Applications

In radar work, very narrow pulses often are required to start oscillators into action, to force grids above cutoff so that a tube may conduct for a short period, to force grids below cutoff so that a tube may not conduct for a short period, or to modulate radio frequencies into brief pulses. Alternately positive

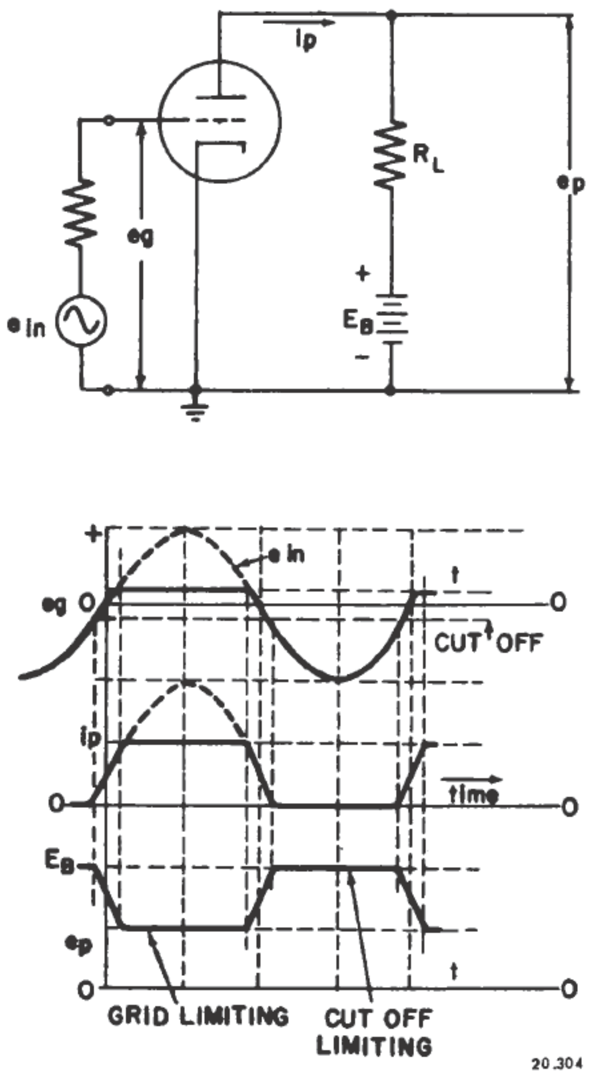


Figure 16-10.—Overdriven amplifier.

and negative pulses, obtained in various ways, may be passed through a limiting circuit to obtain pulses that are either positive or negative with respect to a reference value. This reference level may be at zero voltage or any positive or negative potential. By alternate stages of amplification and limiting, the pulse may be narrowed to any practical width desired. The waveforms of a typical series of such actions are illustrated in figure 16-11.

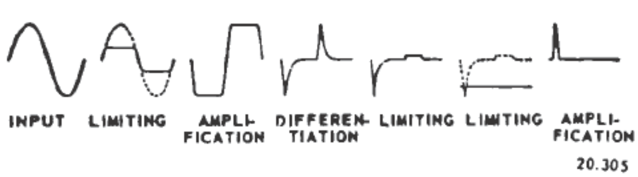


Figure 16-11.—Formation of narrow pulses by successive stages of special radar circuits.

Limiting and differentiating circuits can be used in combination to change sine waves into square waves, and then to limit the peaks of the resultant waveform. A circuit for accomplishing this sequence of events is illustrated in figure 16-12.

A sine wave is fed into the input. The series diode limiter removes the positive portion of the signal and passes the negative portion to the triode limiter. The diode can conduct only when its cathode is negative with respect to its plate—or, in this case, negative with respect to ground.

As the grid of the triode limiter becomes more negative, the plate current drops and plate voltage rises until the grid voltage reaches cutoff. The plate voltage remains at the applied voltage during the time the grid is held below cutoff. When the grid potential rises above cutoff, plate current increases, and the plate potential falls. Thus, a square wave is formed at the output of V2.

If the grid of V3 were disconnected, the square wave across the small-time constant R_1C_1 would give a peaked wave with negative and positive pulses, or spikes across R1. With the grid of V3 connected, the larger part of the positive peak is clipped off because of grid current flow in V3. The voltage applied to V3 is therefore a series of negative pulses.

Except when negative pulses are applied to the grid of V3, saturation current flows in the plate circuit, and the plate voltage is low. During the time negative pulses are applied to the grid of V3 the voltage at its plate rises to the supply voltage. The output is therefore a series of positive pulses.

DIFFERENTIATOR AND INTEGRATOR CIRCUITS

Before studying differentiator (peaking) and integrator circuits you should review *Basic Electricity*, NavPers 10086A, especially the portions on $\frac{L}{R}$ and RC time constants, the universal time constant chart, the growth and decay of current in an RL series circuit, and the charge and discharge of an RC series circuit.

In *Basic Electricity*, a direct voltage was applied to the series RL and RC series circuits. In this section, periodic waveforms (sine,

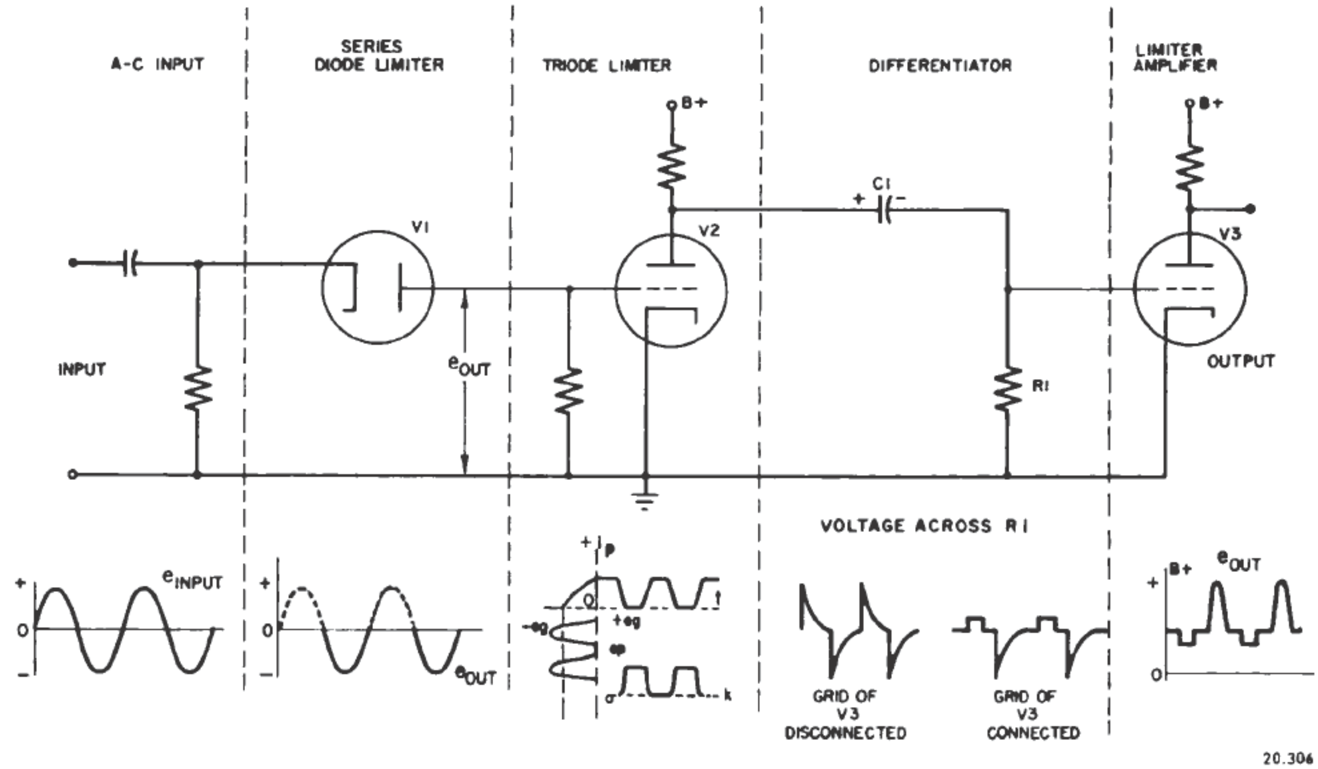


Figure 16-12.—A method of squaring and peaking a sine waveform.

square, and sawtooth waveforms) are applied to series RC circuits. The output voltage waveform across the resistor (differentiator output) and the output across the capacitor (integrator output) are illustrated and discussed in each instance.

An RC voltage divider that is designed to distort the input voltage waveshape is known as a DIFFERENTIATOR or INTEGRATOR, depending on the location of the output taps. The output from a differentiator is taken across the resistance, and the output from an integrator is taken across the capacitor. Such circuits will change the shape of any complex alternating-voltage waveshape that is impressed on them. The amount of distortion depends on the value of the time constant of the circuit as compared to the period of the input waveform. However, neither a differentiator nor an integrator can change the shape of a pure sine wave. In the following figures, both integrator and differentiator outputs are shown, but usually only one output is used in practical circuits.

Sine Wave Input

If a 1000-cycle sine wave voltage having a peak value if 100 volts is applied across an RC voltage divider that has an intermediate or short time constant, the sine wave output across R will be shifted in phase with respect to the input. The phase shift is illustrated in figure 16-13.

R is 10,000 phms and C is 0.0092 μf, and the time constant is R x C, or 10,000 x 0.0092 =92 μs. Compared with the period of the input frequency (1000 μs) the time constant, RC = 92 μs, is relatively short.

In this circuit the differentiator output voltage, e_r, is a sine wave voltage that leads the input voltage, e_{in}, by 60°. The integrator output voltage, e_c, is a sine wave voltage that lags e_r by 90° and lags e by 30°.

Square-Wave Input

If the value of the capacitor in the circuit of figure 16-13 is changed to 0.1 μf and a 1000-cycle square-wave voltage is impressed on the circuit instead of a sine wave, the outputs of figure 16-14 will be obtained.

The action of the square wave on this RC circuit is analogous to the action of the direct potential discussed in *Basic Electricity*, NavPers 10086A. A square-wave voltage of 100 volts (peak) is placed across the input of the circuit, and the capacitor alternately charges

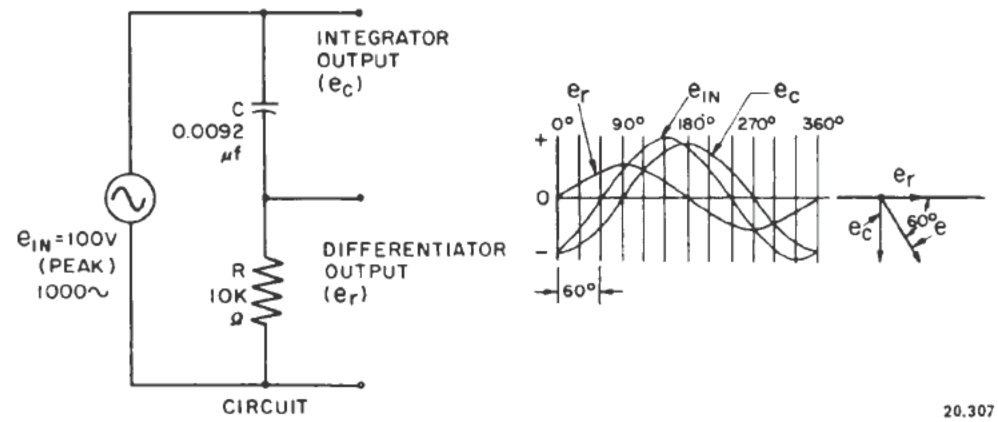


Figure 16-13.—RC integrator and differentiator action on a sine-wave voltage.

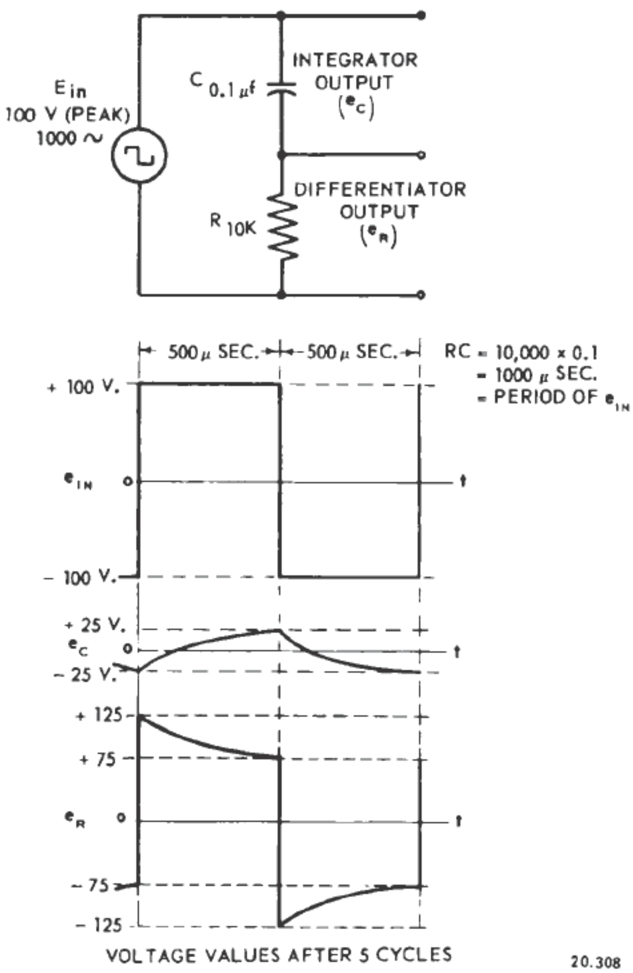


Figure 16-14.—RC integrator and differentiator action on a square-wave voltage.

positively and negatively an amount determined by the RC time constant. Because (in this case) the time constant is equal to the period of the square-wave input, the capacitor never fully charges during either half cycle, and as a result the integrator output, e_C, has a smaller amplitude than the input. The two outputs, e_C and e_r, must add at all times to the input voltage.

The output has a maximum amplitude (after the first cycle) that is greater than the input amplitude because the voltage left on the capacitor from the previous half cycle will add to the input voltage. The sum of these two voltages will appear as a voltage drop across the resistor at the instant the polarity of the input voltage changes.

Figure 16-15 shows the effect of two different values of time constant on the outputs of

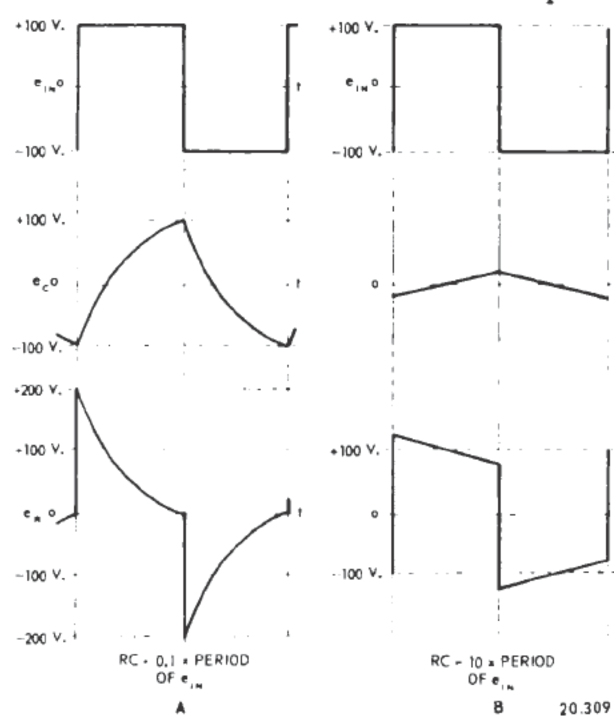


Figure 16-15.—Effect of time constant on RC differentiators and integrators.

the voltage divider. Part A has a time constant equal to one-tenth the period of the input wave. The capacitor has time to become charged during a half cycle. Such a circuit when used as a differentiator is often known as a PEAKER. As may be seen in the figure, the e_r peaks have an amplitude equal to twice the input amplitude.

Part B shows the effect of a time constant that is 10 times as long as the period of the input signal. The differentiator output voltage, e_r, is about the same as the input voltage, E, and the integrator output voltage, e_c, has a very low amplitude.

The voltage spikes in part A are very useful in triggering certain radar circuits (and also television horizontal sync circuits).

Miscellaneous Inputs

Various waveforms, other than sine waves and square waves, may be applied to short RC time constant circuits for the purpose of producing across the resistor an output voltage with an amplitude proportional to the rate of change of the input. Conversely, the shorter the RC time constant is made with respect to the period of the input wave, the more nearly the voltage across the capacitor conforms to the input.

The differentiator, e_r, outputs for two sawtooth input voltage waves are shown in figure 16-16. In both sawtooth waves the slope of the trailing edge of the voltage waveform is greater than that of the leading edge (the voltage changes more per unit of time); therefore, the amplitude of the output voltage, e_r, is greater from time 1 to 2 than it is from time 0 to 1. In the lower waveform the rate of voltage decay is increased (the interval from time 1 to 2 is shortened) with the result that amplitude of the negative peaks of the output voltage, e_r, are correspondingly increased.

CLAMPING CIRCUITS

A circuit that holds either amplitude extreme of a waveform to a given reference level of potential is called a CLAMPING CIRCUIT. The terms, D-C RESTORER and BASELINE STABILIZER, are also used. In general, these circuits may be divided into two categories: (1) DIODE and GRID CLAMPING, which clamps EITHER amplitude extreme and allows the waveform to extend in only one direction from the reference potential; and (2) SYNCHRONIZED CLAMPING, which maintains the output potential at a fixed level until a synchronizing pulse is applied, and at this instant the output potential is allowed to follow the input. At the end of the synchronizing pulse the output voltage is returned immediately to the reference level.

Before discussing clamping circuits it is desirable to review briefly the action of coupling networks. In the coupling between stages in radio and radar circuits, a coupling capacitor is generally used to keep the high positive d-c plate potential of the first tube isolated from the grid of the second tube. It is desirable that only the VARYING component of the plate potential be transmitted to the grid as a signal varying above and below some fixed reference level. If the lower end of the grid resistor is grounded, the signal varies above and below ground. If a biasing potential is employed, the signal applied to the grid varies above and below this d-c bias voltage.

For a class A amplifier, the signal applied to the grid varies above and below some fixed reference level. The biasing potential is

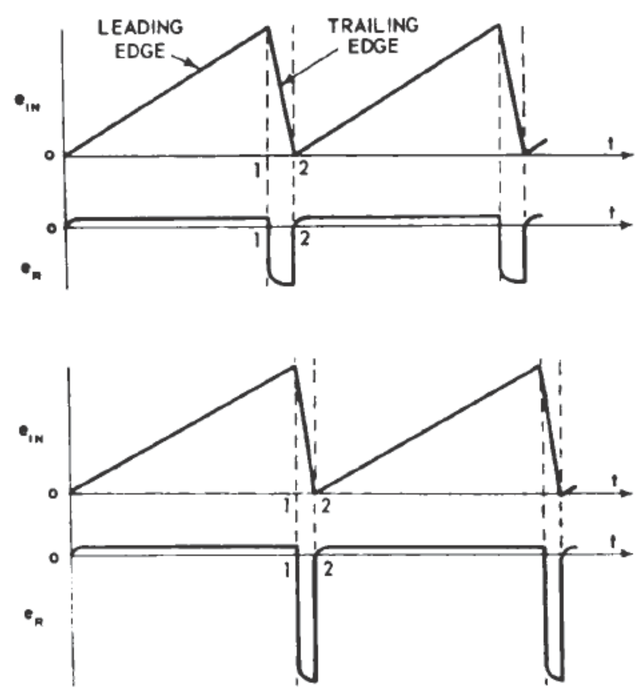


Figure 16-16.—Differentiator outputs of short time constant circuits for two sawtooth input voltage waves.

adjusted to the center of the class A range and the varying potential is kept within the limits of this range (fig. 16-17,A).

In other circuits, however, the waveform swing must be entirely above or entirely below the reference voltage instead of alternating on both sides of it (fig. 16-17,B and C). For these applications a clamping circuit is used to hold either the positive extreme or the negative extreme of the waveform to the desired level.

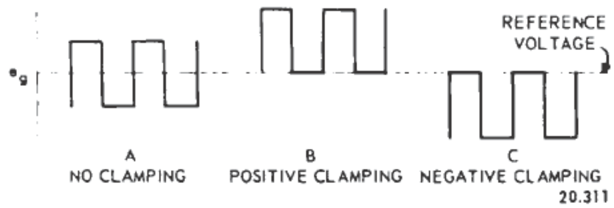


Figure 16-17.—Grid voltage variations with respect to a reference potential.

The output of an ordinary RC coupling network is alternating in character about the average voltage level of the applied waveform because the RC time constant is large compared to the period of the input signal (see fig. 16-15,B). After the coupling capacitor charges to the average voltage, any decrease in the applied voltage causes the output voltage to swing positive. If the capacitor can be made to charge to the MINIMUM applied voltage and no more, any swing has to be in the positive direction. The output voltage therefore varies between the reference voltage and some positive value, depending on the amplitude of the input signal (fig. 16-17,B). If, on the other hand, the capacitor can be made to charge to the MAXIMUM applied voltage and to remain at that level, any swing necessarily is in the negative direction; the output voltage therefore varies between the reference voltage and some negative value, depending on the amplitude of the input signal (fig. 16-17,C).

Diode Clamping

The simplest type of clamping circuit utilizes a diode in conjunction with an RC coupling circuit, one arrangement (positive clamping) of which is shown in figure 16-18. In this instance the capacitor voltage is maintained at approximately the minimum applied voltage.

To understand the action of this circuit the following points should be kept in mind:

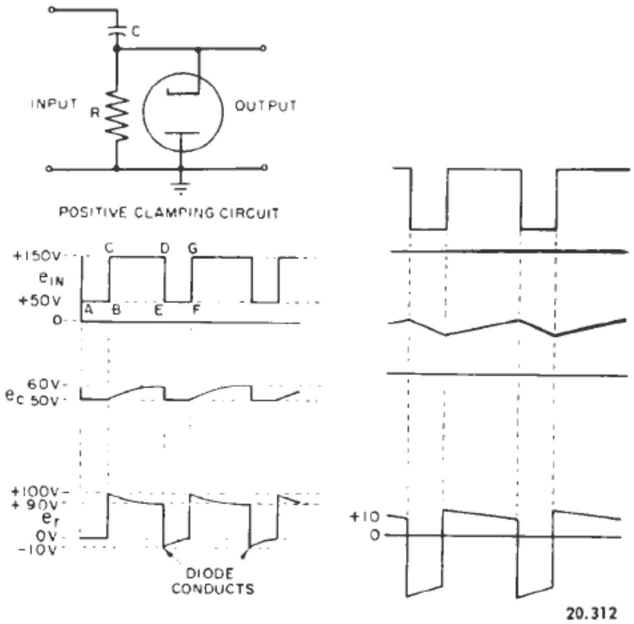


Figure 16-18.—Positive clamping circuit and voltage waveforms.

(1) If the cathode of a diode is made negative with respect to the plate (or the plate positive with respect to the cathode) electrons flow from cathode to plate and the tube becomes a LOW RESISTANCE (in effect, a short circuit); (2) if the cathode is made positive with respect to the plate, no current flows and the tube may be considered a HIGH RESISTANCE (in effect, an open circuit).

The plate-voltage variations of a circuit producing a square-wave voltage is typical of the kind of input (e_{in}) applied to the clamping circuit of figure 16-18. In this clamping circuit, capacitor C charges gradually through the high resistance, R. After a period of time, depending on the RC time constant, the charge on the capacitor reaches 50 volts, the base of the input waveform. The problem is to maintain the charge at this value in spite of the tendency of the capacitor to charge to a higher level when the applied voltage goes to +150 volts. The waveforms shown at the right will exist if no diode clamper is used.

Assuming that a steady voltage equal in magnitude to that at time A has been applied for some time, the capacitor may then be considered to be charged to 50 volts. During the time interval between A and B the charge on the capacitor is equal to the applied voltage, and no current flows through R. Then at point B the applied potential suddenly increases to

+150 volts. Because it is impossible for the charge on the capacitor to change instantaneously, the difference between the +150 volts applied and the 50 volts across the capacitor must appear across R. This difference of 100 volts becomes the output voltage, e_r.

The fact that a voltage appears across R indicates that current flows through it. This current adds to the charge on C. Generally, the RC time constant is very long and the charge added to C is small. For simplicity, assume that the 150-volt potential is applied for a time equal to $\frac{1}{10}$ RC—that is, the interval from point C to point D. Because the cathode of the diode is positive with respect to the anode (which is at ground potential), the tube is in effect an open circuit.

During a time equal to $\frac{1}{10}$ RC, the charge on the capacitor increases exponentially by 10 percent of 100 volts, or 10 volts, making the total charge on the capacitor 60 volts. During the same time the drop across the resistor decreases exponentially by 10 volts to a value of 90 volts, leaving the sum of e_r and e_c still equal to the applied potential of 150 volts.

At point D the applied voltage suddenly drops back to 50 volts. The capacitor, however, is charged to 60 volts. This would leave an output voltage (across R) of 10 volts negative with respect to ground—a condition that must be avoided. In order for the output to return to zero very quickly, the capacitor must discharge the extra 10 volts through a path having a very short RC time constant.

In figure 16-18 the cathode of the diode is connected to the high side of R, and the plate is grounded. Any output voltage that is negative with respect to ground makes the cathode negative with respect to the plate. Under this condition the diode conducts and becomes, in effect, a very low-resistance discharge path for the capacitor until the charge is again equal to the applied voltage and the output voltage returns to zero. At this time the diode becomes nonconducting.

To illustrate the operation of the positive clamping circuit further, assume that the negative-going waveform shown in figure 16-19 is applied to the input of the clamping circuit.

Because at point A the input voltage is zero, the output voltage is zero and remains so until

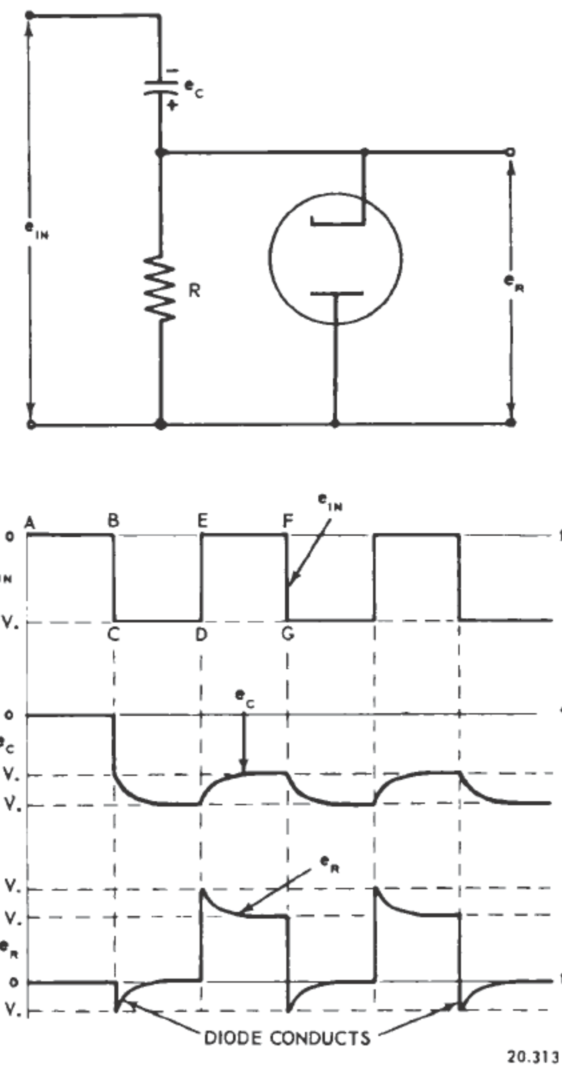


Figure 16-19.—Negative-going input signal applied to positive clamping circuit.

point B is reached. At this time the input voltage drops suddenly toward -100 volts at point C. Because the capacitor cannot change its charge instantaneously, the output voltage across R also drops suddenly toward -100 volts. When the cathode of the diode is sufficiently negative with respect to the plate, the tube conducts, charging the capacitor very rapidly through the short RC time constant of the conducting diode capacitor and reducing the voltage circuit across R to that across the conducting diode (almost zero). When the capacitor voltage becomes equal to the applied voltage, the diode becomes nonconducting. As long as the input remains at -100 volts, from point C to D, the output voltage remains at zero potential.

At point D the input voltage changes back to zero, a rise of 100 volts in the positive direction (-100 to 0). This rise produces a rise of 100 volts (0 to +100) across R because the capacitor again cannot change its charge instantaneously. The capacitor must now discharge very slowly because the diode is nonconducting, and the high-resistance path through R must be utilized.

Assuming again that the discharge time from points E to F is $\frac{1}{10}$RC, the voltage across the capacitor at point F, and thus the output voltage, decreases to 90 volts because the input is zero. At point F the input signal again drops to -100 volts (point G). Instantaneously, the output across R goes to -10 volts (input minus e_C). The diode conducts quickly, returning the charge on the capacitor to 100 volts and the output to zero. The output voltage waveform is shown in the lower part of the figure. Note that no portion of the waveform is lost after the first cycle. The function of the clamping circuit is merely to shift the waveform from above to below the zero voltage reference level (ground).

A negative clamping circuit and its associated waveforms are illustrated in figure 16-20. This diode clamping circuit is capable of causing the output voltage to vary between some negative value (-100 v in this figure) and the zero reference voltage. The only difference between this circuit and the one illustrated in figure 16-18 is in the manner in which the diode is connected. In figure 16-18, the plate is grounded and the tube conducts if the cathode is made negative with respect to the plate. In figure 16-20 the cathode is grounded, and the tube conducts whenever the plate voltage rises above ground.

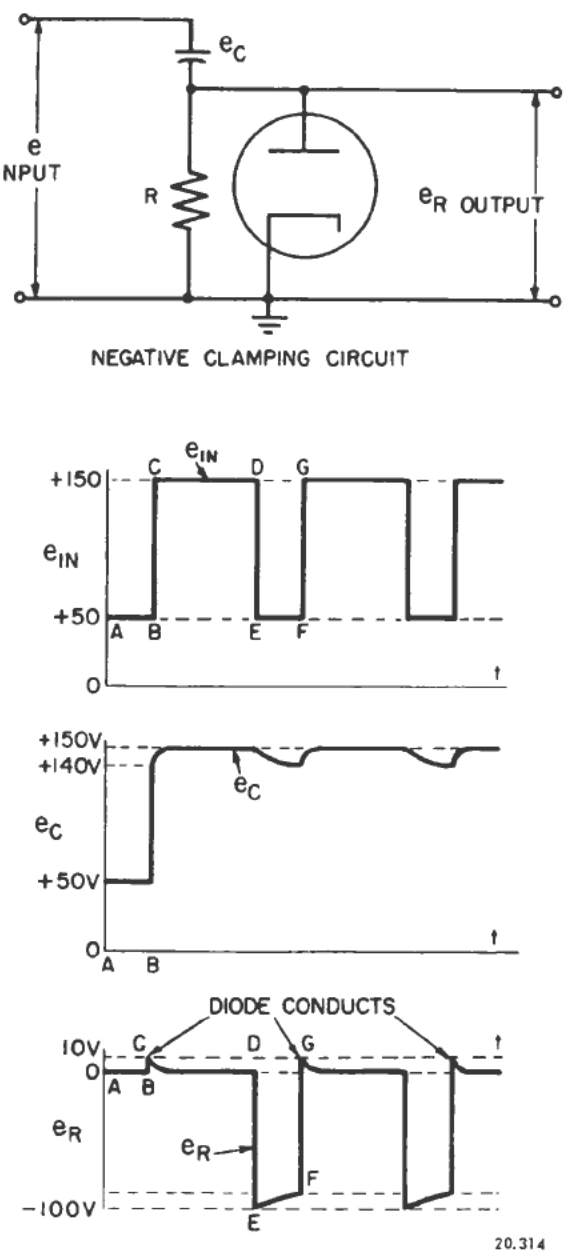


Figure 16-20.—Negative clamping circuit and voltage waveforms.

Grid Clamping

Clamping may be performed at the grid of an ordinary triode or pentode as well as in a diode. Any element of an electron tube, if made positive with respect to the cathode, attracts electrons from it. On the other hand, any element made negative with respect to the cathode repels electrons and has no current flow. Thus, the grid of a tube, connected as shown in figure 16-21, acts as the plate of a diode circuit illustrated in figure 16-20. Any tendency of the grid to go positive causes grid current to flow, charging capacitor C quickly to the applied potential through the low-resistance conducting cathode-grid circuit.

Clamping Above or Below Ground Potential

Although the circuits previously discussed clamped one extreme of the input signal to zero potential, actual circuits need not be limited to this one reference potential. Figure 16-22 illustrates the means of clamping the upper extreme of an input signal to -10 volt with

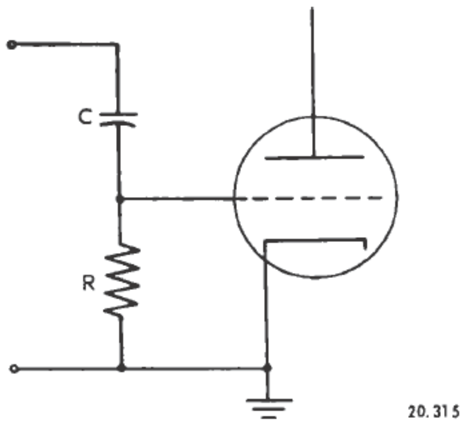


Figure 16-21.—Grid-clamping circuit.

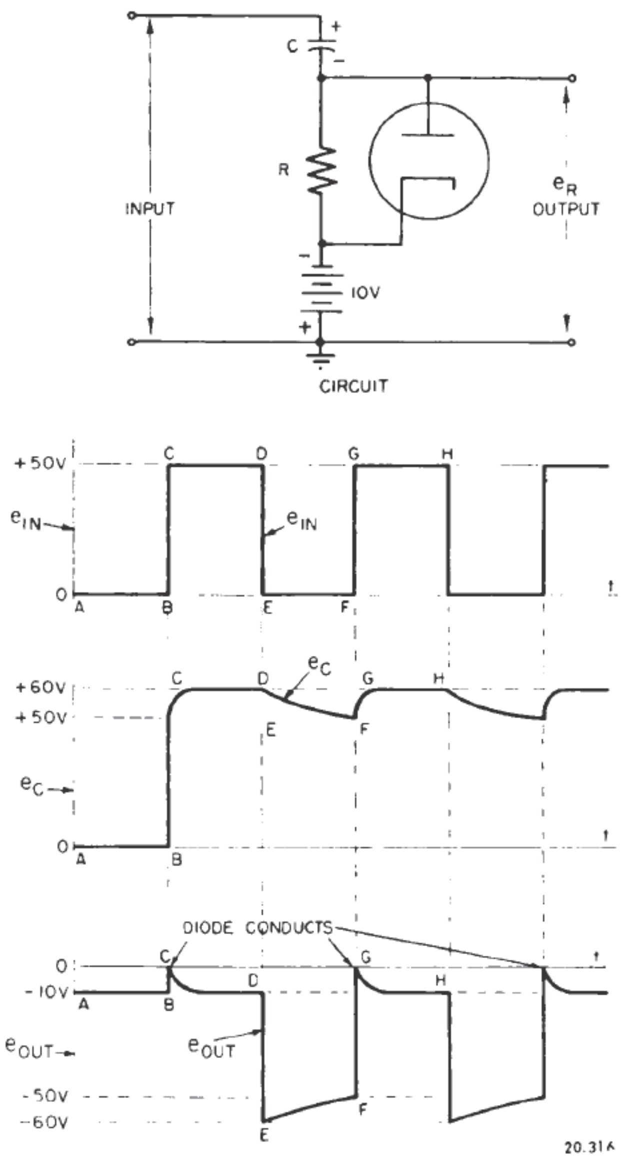


Figure 16-22.—Clamping circuit that establishes reference voltage of -10 volts.

respect to ground. The same principle may be applied to clamp either voltage extreme to any reference potential.

The circuit operates in the following manner: (1) From A to B the input is zero; and, because the output is connected across the 10-volt battery (through R), the output is negative 10 volts with respect to ground. (2) At B the +50-volt positive pulse is applied to the input, and the voltage across R rises quickly toward 60 volts. This action makes the plate positive with respect to the cathode, and the diode conducts, thus in effect shorting out R. The output potential falls to -10 volts with respect to ground as C charges quickly to 60 volts through the conducting diode. At D the input pulse is removed and C discharges through R, making the diode plate negative with respect to the cathode and removing the diode short circuit across R. This action causes the output voltage to go 50 volts more negative (from -10 to -60 volts). From E to F the output voltage remains at approximately 50 volts below the 10-volt reference potential as C discharges slowly through R. At F the positive-going input pulse is reapplied; this voltage is bucked by the remaining charge on C, but aided by the 10-volt battery, current flows upward through R, making the plate of the diode positive with respect to its cathode; current again flows through the diode, and shorts out R. The output potential is returned to -10 volts with respect to ground shortly after G. From G to H the output potential remains at -10 volts with respect to ground as the voltage across C increases quickly to 60 volts, and the drop across R decreases to zero.

Applications of Clamping Circuits

In practice, clamping usually is encountered in sweep circuits. If the sweep voltage does not always start from the same reference point, the trace itself does not begin at the same point on the screen each time the cycle is repeated, and is therefore jittery or erratic. If a clamping circuit is placed between the final sweep amplifier and the deflection element, the voltage from which the sweep signal starts can be regulated by adjusting the d-c voltage applied to the clamping circuit.

One circuit that is typical of the clamping circuits used in cathode-ray oscilloscopes is shown in figure 16-23. The beam of the cathode-ray tube, V3, is deflected by the push-pull sawtooth voltages shown at 1 and 2. The beam therefore traces a bright line on the screen when it is moved from left to right at uniform speed, starting at point A. At the end of the sweep, the beam is moved very quickly from point B back to point A. The function of the clamping circuit is to force point A to remain at the same place on the screen even though amplitude variations may occur in the applied sawtooth wave.

If diodes V1 and V2 of figure 16-23 were not connected (tubes removed from sockets) and the B voltage adjusted (for example, to +50 volts) the average potential on plate D1 would be 50 volts more positive than that of D2. This would cause the electron beam, and the spot on the screen, to be attracted to the left center of the screen in the absence of the sweep voltage. Thus, any variation in the amplitude of the sweep voltages would cause the beginning of the sweep (point A) to change position on the screen.

Diodes V1 and V2 are connected in the circuit in order to clamp the start of the sweep, (point A) in figure 16-23, to a fixed potential. The first cycle of the applied sawtooth voltage (which varies between +100 and +200 volts above ground) is shown before the diodes are connected so that the change brought about by the action of the diodes will be apparent. As may be seen in the figure, the voltage at the plate of D1 has an

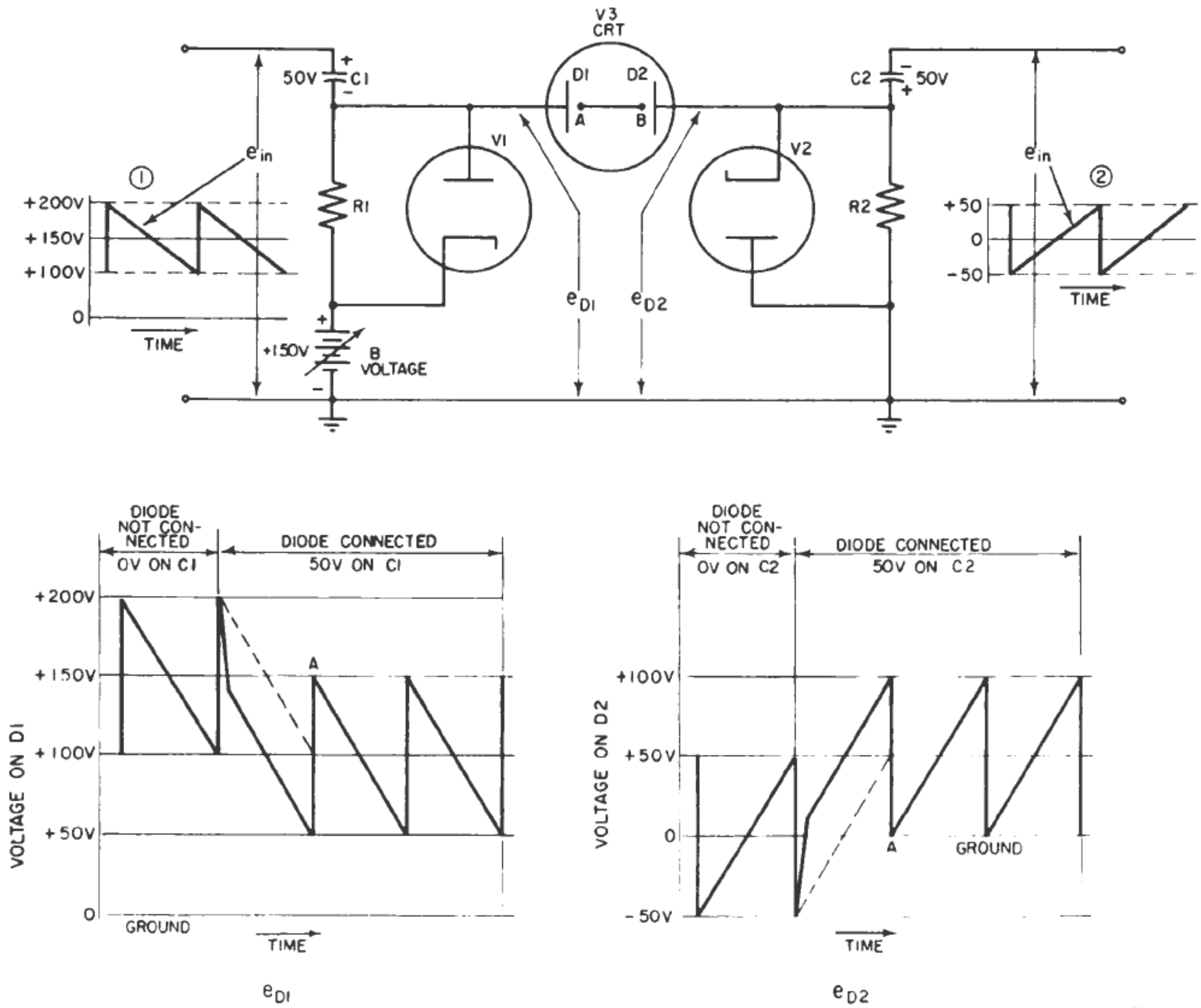


Figure 16-23.—Clamping diodes employed in electro-static sweep systems.

average value of +150 volts with respect to ground, and the average charge on C1 is zero.

When the diodes are connected in the circuit, and the B voltage adjusted so that the cathode of V1 is at +150 volts, any rise in the voltage at plate D1 above +150 volts causes VI to conduct. This results from the fact that the plate of V1 becomes positive with respect to its cathode as current flows upward through R1. When V1 conducts, C1 is rapidly charged through V1, and the voltage on plate D1 is held at +150 volts with respect to ground because the conducting diode, in effect, connects plate D1 to the positive terminal of the B supply. When the peak input voltage is +200 volts, C1 charges rapidly to 200 - 150 or 50 volts through the short RC time constant circuit composed of V1 and C1.

When the input voltage stops rising and starts to decrease, C1 stops charging and starts to discharge through R1. This action makes the plate of V1 negative with respect to the V1 cathode, and conduction ceases. The short circuit is removed from R1, and the output voltage, e_{d1}, decreases with the decrease in input voltage. The decrease in input voltage appears across R1. The output voltage is equal to 150 $-e_{r1}$. It is also equal to e_{in} -50.

For example, when the input voltage falls from 200 volts to 175 volts, e_{d1} falls from 200 -50 = 150 volts to 175 - 50 = 125 volts. When the input falls to 150 volts, e_{d1} falls to 150 - 50 = 100 volts; and when the input falls to 100 volts, e_{d1} falls to 100 - 50 = 50 volts. At this instant the input voltage suddenly increases to +200 volts. If the voltage on C1 is 50 volts, no current will flow through R1. However, if e_{in} exceeds + 200 volts, current will flow upward through R1 to increase the charge on C1; V1 will conduct and e_{d1} will be clamped to 150 volts. Thus, if the input voltage rises to 201 volts, the voltage across C1 will increase to 201 - 150, or 51 volts.

The small amount of charge that leaks off C1 during the sweep time is replenished at the start of each successive sweep through the V1 diode. The replenishment of the charge through V1 holds the start of the sweep at + 150 volts. Thus, variations in the amplitude of the sweep voltage can affect only the length of the sweep while the starting point is held fixed.

In a similar manner the average voltage to which C2 is charged is changed from zero to +50 volts by the action of diode V2. Thus, the starting potential of each sweep on plate D2 is clamped to ground potential. The input voltage in this case varies between +50 volts and -50 volts. The action of V2 is to charge C2, thus shifting the sweep voltage +50 volts so that the D2 plate voltage varies between 50 +50, or 100 volts; and 50 - 50, or 0 volts.

Note that the difference between the AVERAGE potential on D1, before and after V1 is connected, is 50 volts; and the difference between the average potential on D2, before and after V2 is connected, is also 50 volts; therefore the centering effect is the same with or without the diodes, but the B voltage in the cathode circuit of V1 must be higher (+150 volts instead of +50 volts) in the case when the clamping tubes are used.

PHASE SHIFTER

Phase shifters are known also as phase splitters and phase inverters. They are used in oscilloscopes to provide from a single source two voltages that are 180° out of phase in order to provide a push-pull output from the horizontal amplifier stage. This type of circuit, together with some of its other uses, is treated in chapter 6.

COUNTING CIRCUITS

Pulses or oscillations may be counted electronically. If the impulses that are to be counted have the same average amplitude, a counting circuit produces an output voltage proportional to the pulse repetition frequency of the input signal. Variations in amplitude of the input signal will affect the output voltage unless all variables except the pulse repetition frequency are eliminated. These variables are eliminated by shaping and limiting circuits.

Basic counting circuits may be modified for use with a blocking oscillator to produce trigger pulses, which are submultiples of the pulse repetition input frequency. The counting circuit may be connected to count either positive or negative pulses; both types of circuits are shown in figure 16-24.

When a positive pulse is applied to the input of the positive counting circuit (fig. 16-24, A), V2 conducts because its plate is made positive with respect to its cathode. When C1 charges up to the source voltage, V2, conduction stops. When the positive input pulse stops, the plate

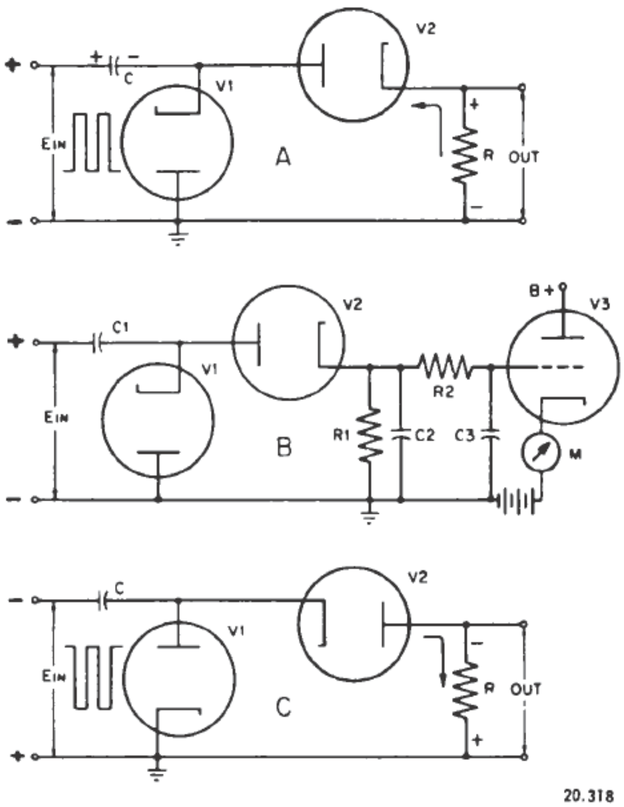


Figure 16-24.—Positive and negative counting circuits.

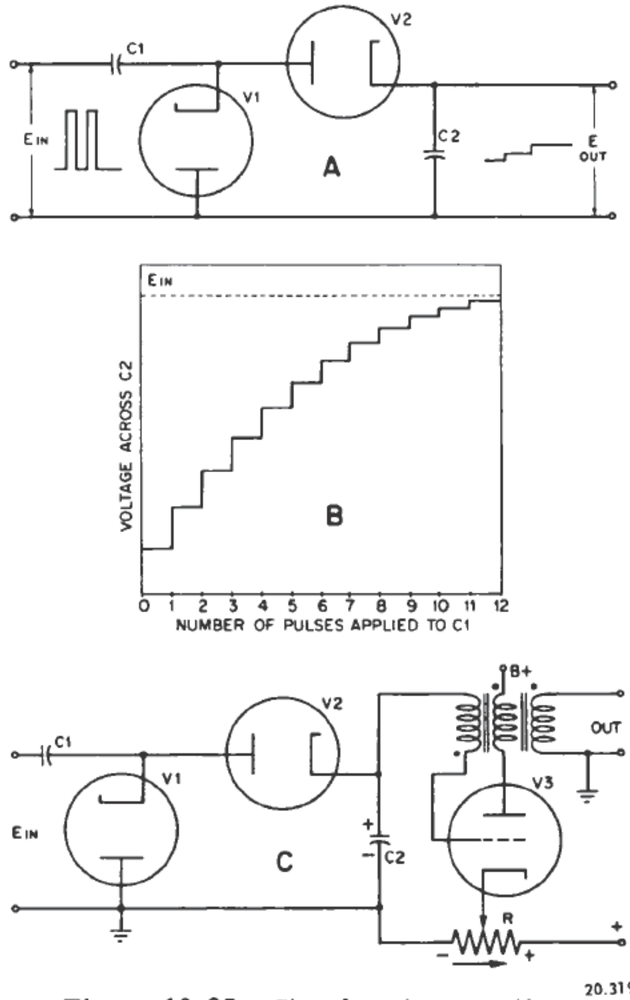


Figure 16-25.—Step-by-step counting.

of V1 is then positive with respect to the cathode, and V1 conducts, which discharges C1 to ground. This cycle is repeated each time a positive pulse is applied to the input terminals. Each time V2 conducts electrons will flow through R and produce a voltage drop across its terminals that has the polarity shown.

The average current through R increases or decreases as the pulse repetition frequency (PRF) increases or decreases. The variation in the voltage drop across R, due to changes in current, may be smoothed out by a conventional RC filter and used to control an amplifier, as shown in figure 16-24, B. The filtered counting circuit output is applied to the grid-cathode of V3, and the average plate current of V3 is indicated by the meter, M. Thus meter current indications can be considered a measure of the input, PRF.

Negative input pulses may be counted by reversing the diode connections, as shown in figure 16-24, C. The circuit operation is similar to the positive counting circuit action, except that V2 conducts when negative pulses are applied. An increase in PRF will cause the average current through the meter to decrease so that the meter indication is opposite to that of a positive counting circuit.

It is often desirable to have a counting circuit that will count a specific number of pulses and then trigger another circuit. A suitable circuit, similar to the positive counting circuit, is shown in figure 16-25, A.

For step counting, the load resistor of the positive counting circuit is replaced by capacitor C2. This capacity is relatively large in comparison to the capacity of C1. Each time V2 conducts, the charge on C2 is increased slightly as shown in figure 16-25, B. The steps are not the same height, and they decrease in height exponentially with time as the voltage across C2 approaches the final value.

As long as there is no discharge path for C2, the voltage across its terminals increases with each successive step until it is equal in amplitude to the applied pulse. When this condition is reached, V2 will no longer conduct because its plate and cathode are at the same potential. The voltage across C2 might be applied to a grid-controlled thyratron to cause the tube to conduct (fire) when the required number of pulses has been counted. The firing of the thyratron would discharge C2, and the counting cycle would start all over again.

The circuit in figure 16-25, C, may be used as a frequency divider. When used in this manner, V3 is used as a single-swing blocking oscillator that is triggered when the voltage across C2 becomes great enough to cause the tube to conduct. At other times, the tube is cut off by the bias voltage developed in the section of R that is between ground and the slider.

In operation, as soon as the charge on C2 is great enough to overcome the bias voltage, the grid of V3 swings positive with respect to its cathode, and the heavy grid current quickly discharges C2. A positive pulse at the output will appear as a submultiple of the input PRF.

The submultiple number is determined by the setting of R, which sets the bias voltage of V3 and thereby selects the number of pulses that must be applied to the input before V3 will conduct. For example, a PRF of 1000 cps may be fed to the input of the counting circuit. The bias on V3 is adjusted so that the voltage built up on C2 will overcome the bias voltage every fourth step. Therefore, the rise in voltage on the capacitor triggers the oscillator and the current flow through the oscillator then discharges the capacitor. The oscillator output pulse frequency would then be one-fourth the input frequency, or 250 cps.

CONNECTING CIRCUITS

Connecting circuits are used to connect one circuit to another in such a way that minimum interference between the circuits will result. They are also used to enable a maximum transfer of energy or to accomplish some other desired result as described below.

CATHODE FOLLOWER

The cathode follower is a degenerative electron-tube circuit in which the inverse feedback is obtained by way of an unbypassed cathode resistor across which the output is taken. It is used to prevent interference between two circuits and as such becomes a "buffer" stage. These circuits are widely used as impedance-matching devices. Cathode followers are treated in chapter 6.

ELECTRONIC SWITCH

The electronic switch is used to close, open, or change the operation of an electronic circuit. The electronic switch is very sensitive and is fast in operation. Thus, it can alternately connect one circuit to an oscilloscope, disconnect this circuit, and then connect a second circuit fast enough to present both waveforms simultaneously for a comparative study. This subject is covered in chapter 17.

CHAPTER 17
ELECTRONIC TEST EQUIPMENT

CATHODE-RAY OSCILLOSCOPE

CATHODE-RAY TUBE

The cathode-ray tube (CRT) is a special type of vacuum tube in which electrons emitted from the cathode are shaped into a narrow beam and accelerated to a high velocity before striking a phosphor-coated viewing screen. The screen fluoresces or glows at the point where the electron beam strikes and thus provides visual waveforms of current and voltage. The CATHODE-RAY OSCILLOSCOPE is a test instrument which uses the cathode-ray tube. The waveforms can be positioned on the screen so they appear stationary because the electron beam repeatedly reproduces the pattern in the same location, provided the horizontal motion of the electron beam is kept in step with the frequency of the signal.

The comparison is made against waveforms located on the schematic diagrams. Scope patterns periodically taken at the test points are compared with these waveforms. Differences between the optimum waveform and the scope pattern indicate that corrective action is needed. By using the oscilloscope in this manner, difficulties may be pin-pointed to a specific circuit or portion of a circuit in a short time.

The tube is also used as the visual indicating device for radar, sonar, radio, direction finders, loran, and television.

The beam of electrons has practically no weight or inertia but follows a straight line unless diverted by an electric or a magnetic field. Cathode-ray tubes are of two types according to the method of deflecting the electron beam—(1) electrostatic, and (2) electromagnetic. Almost all types of test-instrument oscilloscopes use the electrostatic type of CRT because of the need for high sweep frequencies in both the vertical and horizontal directions. Certain radar, sonar, and TV sets use electromagnetic type CRT's. Focusing or narrowing of the beam is achieved by either electrostatic or electromagnetic methods. Electrostatic deflection bends the beam by an electric field produced by a deflection voltage between parallel plates inside the CRT. Electromagnetic reflection bends the beam by means of a magnetic field produced by a deflection current in a coil around the neck of the tube.

An electrostatic type of CRT is illustrated in figure 17-1. The cathode, when heated by its enclosed filament, releases free electrons. A cylindrical grid surrounds the cathode and controls the beam intensity as electrons pass through the end opening of the grid. The control is achieved by varying the negative voltage on the grid and is called INTENSITY or BRIGHTNESS CONTROL. After leaving the grid, the electron stream passes through two or more cylindrical focusing anodes which narrows the beam. The first anode concentrates the free electrons and the second anode increases their speed. The entire assembly is called the ELECTRON GUN. The electrons emerge from the electron gun at high speed.

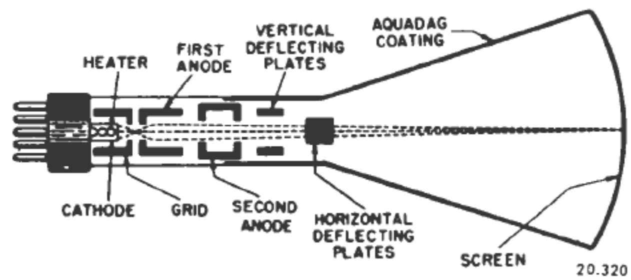


Figure 17-1.—Construction of cathode-ray tube using electrostatic deflection and focusing.

The grid helps to narrow the beam but cannot focus it to a sharp point on the viewing screen. The two anodes aid in the focusing action, as shown in figure 17-2.

Both cylindrical anodes are positive with respect to the cathode but the second anode is positive with respect to the first anode. Thus, an electric field is established between the

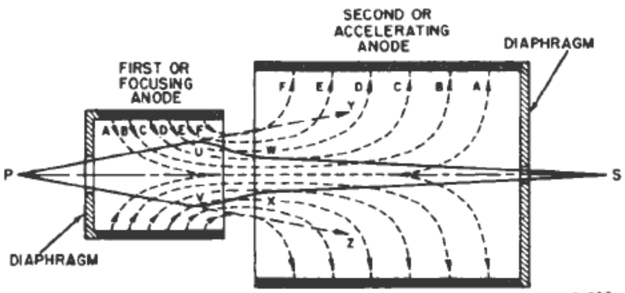


Figure 17-2.—Electrostatic focusing.

anodes as shown in figure 17-2. The electrons which are emitted from the cathode are attracted by the first anode. Some of the electrons pass through the hole in the end of the first anode and into the field between the two anodes. The purpose of the diaphragm is to prevent all electrons except those making a small angle with the axis of the beam from passing through the hole in the diaphragm. This serves to keep the beam narrow. The electrons entering the curved electric field between the anodes are subjected to inward-directed forces thereby focusing the beam. As the beam passes parallel to the lines of force, the electrons are accelerated to a very high speed. Thus, the net result of the forces influencing the beam of electrons is a high speed inward-directed beam converging at point S on the screen. The repelling force of like charges tends to scatter the electrons but they are accelerated to such a high speed that the scattering action is not effective in defocusing the beam. Nevertheless the mutual repulsion between electrons in relation to the speed of the electrons determines the sharpness with which a beam may be focused on the screen. The diaphragm on the accelerating anode is used to stop all wide angle electrons from hitting the screen.

The focus of the electrostatic type of cathode-ray tube is generally controlled by varying the voltage between the first anode and the cathode. This voltage varies the force exerted on the electrons and tends to narrow the beam. Thus, if the screen is observed when the first anode voltage is varied, the beam may be brought to a bright sharp spot.

Focusing in an electromagnetic cathode-ray tube is achieved by a coil encircling the outside neck of the tube. The coil may be moved along the neck to a limited extend to initially focus the beam, but the usual method, after the focus coil is in the proper position, is to vary the current flowing through the coil with a variable resistor. The term "focus control" is derived from this function.

Without lateral deflection, the electron gun produces only a small spot of light on the viewing screen. With deflection, the trace of the spot forms a line on the screen. The electrostatic-type of cathode-ray tube uses two pairs of deflection plates mounted at right angles to each other, as shown in figure 17-3. The vertical deflection plates (YY') deflect the beam vertically and the horizontal deflection plates (XX') deflect it horizontally. Both pairs usually function simultaneously. The beam is attracted by the positive plate and repelled by the negative plate as the electrons pass between them. One plate of each pair may be grounded. To deflect the beam, a positive or negative voltage is applied between the other plate and ground, thus establishing an electric field between the plates. The deflecting force varies with the deflection voltage across the plates and with the field intensity.

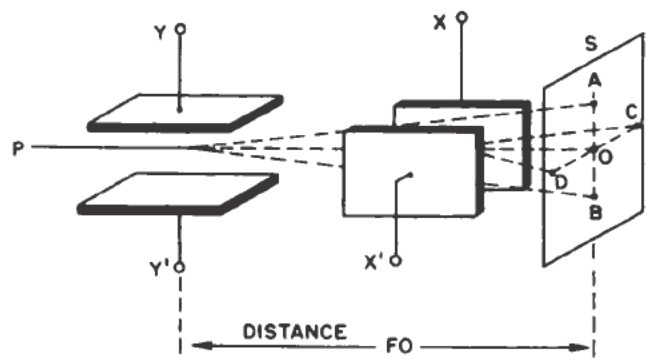


Figure 17-3.—Deflecting plates for electrostatic cathode-ray tube.

If plate Y is positive with respect to Y' the beam is deflected upward, striking the screen at A. If plate Y is negative with respect to Y' the beam is deflected downward, striking the screen at B. If there is no deflection voltage across the plates, the beam will strike the screen at 0. The amount of deflection varies with the deflection voltage across the plates. Note the relationship between length of line and voltage. This characteristic makes it practical for a scope to serve as a voltmeter because the length of line (so produced) is a measure of the applied voltage.

Another practical use of the scope is for determining polarity. Again refer to figure 17-3. If plate X is positive with respect to X' the beam will be deflected horizontally, and will strike the screen at C. If X is negative with respect to X' the beam will strike the screen at D. Both pairs of plates are mounted near the output end of the electron gun, with the vertical deflection plates farthest from the screen. Two centering controls (horizontal and vertical) enable the operator to move the spot as he desires, to any place on the screen. Each control customarily consists of a variable resistor that serves as a voltage divider to enable the centering action.

The deflection angle of the electron beam is the total angle through which the beam may be moved or diverted. A cathode-ray tube with a 50° deflection angle can deflect the beam at any angle that equals (or is less than) 25° from the center line.

The length of time that the screen glows, or flouresces, at the point where the electron beam strikes depends on the material of the phosphor coating on the screen, and is known as SCREEN PERSISTENCE. Some cathode-ray tubes have a long persistence screen and others have a short one, depending on their use. The screen phosphors are designated by the letter "P" followed by a number. Most radar indicator cathode-ray tubes employ P1, P4, or P7 screen phosphors. The P1 and P4 phosphors have medium persistence and give off green and white light respectively. The P7 screen has long persistence and gives off yellow light.

All flourescent materials have some phosphorescence, or afterglow, but the duration of the afterglow varies with the material, as well as with the amount of energy in the beam causing the emission of light. For oscilloscopes that are to be used for observing nonrepeating phenomena or periodic phenomena that occur at a low repetition ratio, a screen material on which the image will linger is desirable. Where the image changes rapidly, prolonged afterglow is a disadvantage, because it may cause confusion on the screen.

The eye retains an image for about one-sixteenth of a second. Thus in a motion picture, the illusion of motion is created by a series of still pictures flashed on the screen so rapidly that the eye cannot follow them as separate pictures. In the cathode-ray tube the beam is repeatedly swept across the screen and the series of adjacent spots appear as a continuous line. Thus the wave shape of an a-c voltage can be observed on the screen when the a-c voltage is applied to one pair of deflection plates and simultaneously a second voltage of appropriate characteristics is applied to the other pair of plates.

The conventional way of representing voltage or current of sine waveform is shown in figure 17-4,A. The voltage to be observed is applied across the vertical deflection plates and simultaneously a saw-tooth voltage is applied across the horizontal deflection plates. The saw-tooth voltage moves the beam from left to right at constant speed to form the time scale along OX; then it returns the beam rapidly to the starting position at the left and repeats the operation. The saw-tooth voltage is so named because when plotted against time it resembles a saw-tooth, as shown in figure 17-4.

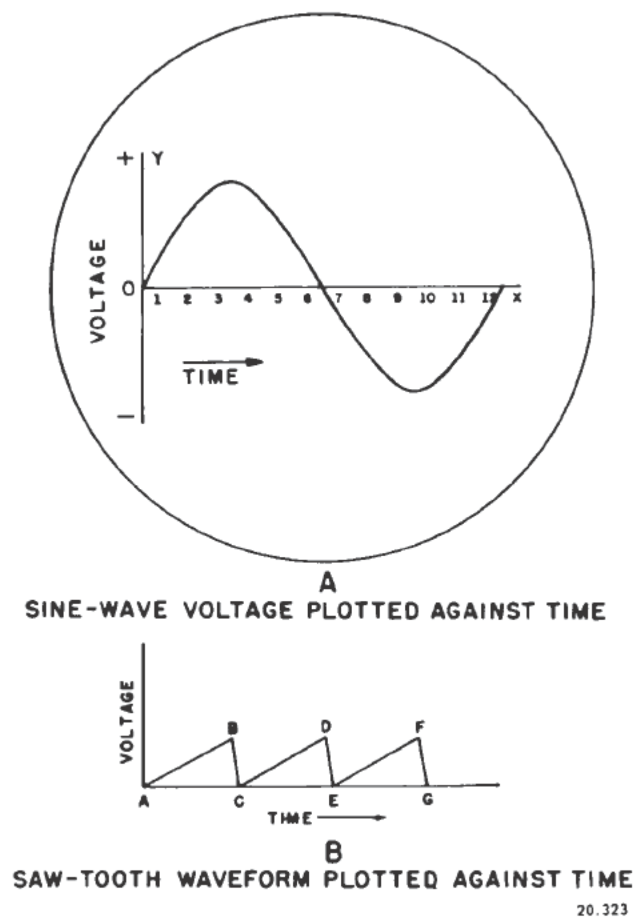


Figure 17-4.—Sine-wave and saw-tooth voltage waveforms.

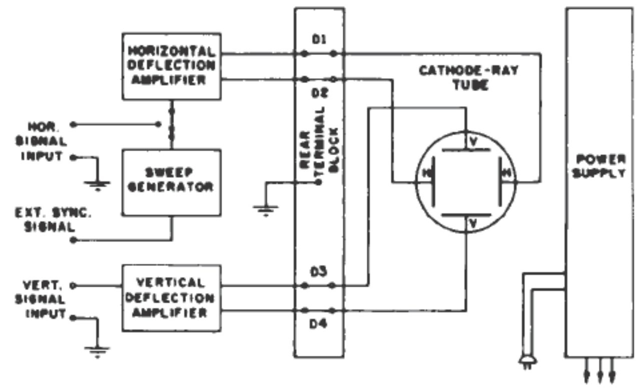
B. As the voltage increases from A to B the beam is swept from 0 to 12 (fig. 17-4,A). As the voltage falls from B to C in figure 17-4,B, the beam is quickly returned to its starting position and the process is repeated.

If an a-c voltage of sine waveform is placed across the vertical deflection plates with no horizontal deflection, a single vertical line appears on the screen. The varying rate of change of the voltage is hidden because the vertical movements retrace themselves repeatedly on the same vertical line. Similarly, if a sweep voltage of saw-tooth waveform is applied to the horizontal deflection plate in the absence of vertical deflection, a horizontal line is formed and the rate of change of the voltage is obscured. However, when both voltages are introduced at the same time, the vertical motion of the beam is spread out across the screen to form a sine curve, like that shown in figure 17-4,A.

OSCILLOSCOPE CIRCUITS

A block diagram of a cathode-ray oscilloscope is shown in figure 17-5. The horizontal deflection amplifier is a high-gain R-C coupled class-A wide-band voltage amplifier that increases the amplitude of the horizontal input voltage and applies it to the horizontal deflection plates. The sweep generator supplies a saw-tooth voltage to the input of the horizontal amplifier through a switch that provides an optional external connection. The vertical deflection amplifier increases the amplitude of the vertical input voltage before applying it to the vertical deflection plates. The input to the vertical amplifier appears in magnified form on the viewing screen as a graph of the current or voltage waveform being examined. A rear terminal block provides direct electrical connections to the deflection plates. The direct connections are used, for example, when examining d-c potentials, or high-frequency signals that would be attenuated excessively by the amplifier circuits. The power supply provides all d-c voltages for the tubes, including a high d-c potential for the cathode-ray tube.

A schematic diagram of an elementary cathode ray oscilloscope is shown in figure 17-6. The cathode-ray tube employs electrostatic focusing and deflection. V1 is the vertical amplifier, V2 is the horizontal amplifier, and V3 the



20.324

Figure 17-5.—Block diagram of a cathode-ray oscilloscope.

sweep generator. R1 is a manual vertical gain control, R2 is a manual horizontal gain control, S3 is the coarse frequency adjustment for V3, and R10 is the fine frequency adjustment. An external synchronizing signal may be applied from an external source to the grid of V3 when S2 is in the EXT. SYNC. position. The synchronizing signal is obtained from the plate of V1 when S2 is in the INT. SYNC. position. R3 provides manual control of the sync signal amplitude.

The low voltage power supply includes a conventional full wave rectifier, V4; secondary windings L4, L5, and L6 of the power supply transformer; and the pi-filter (C17, L3, and C18). The cathode of V4 is positive with respect to ground. The output voltage (400 volts) is applied to the plate circuits of V1, V2, and V3. The voltage divider (R14, R15, and R19) is connected across the output of V4 and supplies +170 volts to the left side of the centering controls (R17 and R18).

The high voltage power supply includes V5 and secondary windings L6, L7, and L8 of the power supply transformer. The output voltage of V5 is negative with respect to ground and is applied across the voltage divider (R20, R21, R22, and R24). C19 filters the output voltage. The tap between R20 and R21 supplies -170 volts to the right side of R17 and R18. Since -170 volts and +170 volts are applied respectively to the right side and the left side of R17 and R18, the exact electrical centers of the two potentiometers are at ground potential or zero volts.

R17 and R18 are positioning controls that provide manual adjustment of the low d-c voltages that may be applied across the two pairs of

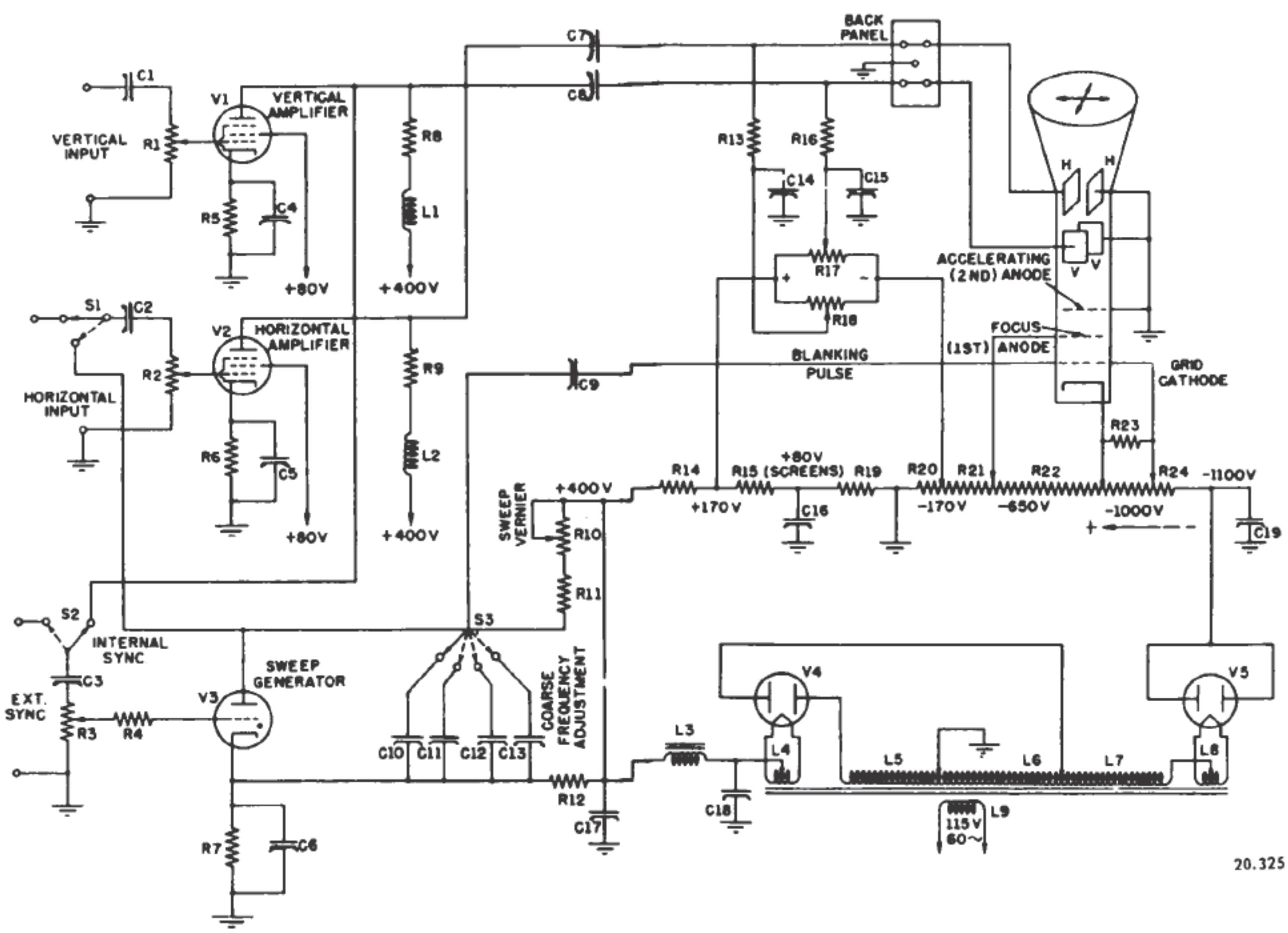


Figure 17-6.—Schematic diagram of an elementary cathode-ray oscilloscope.

deflection plates. The spot is approximately centered on the screen when the contact arms of R17 and R18 are positioned on their electrical centers thus applying ground potential to their respective deflection plates. Moving the contact arm of R17 to the right makes one vertical plate more negative and thus repels the beam and moves the spot vertically a certain distance on on the screen. Conversely moving the contact arm of R17 to the left of the zero position makes the vertical plate more positive, and thus attracts the beam and moves the spot in the opposite direction.

The cathode of the cathode-ray tube is connected at -1,000 volts with respect to ground. The second anode in the cathode-ray tube is grounded, and the first anode is negative with respect to the second anode, but both anodes are positive with respect to the cathode. This arrangement provides the necessary focusing and accelerating voltages for the electrons in the beam to form a bright spot on the screen, and at the same time prevents defocusing the spot by holding the average voltage across the deflection plates close to the potential of the second anode. The arrangement also introduces a safety factor by removing high voltage from the deflection plates and the associated input terminals on the rear panel.

Capacitors C1, C2, and C3 block any external d-c voltage components from the grids of V1, V2, and V3. Similarly, capacitors C7 and C8 block the d-c components of plate voltage from the cathode-ray tube deflection plates and at the same time couple the a-c components to them. C9 couples a blanking pulse to the grid of the cathode-ray tube, which blanks out the return trace of the sweep generator. During the time the sweep voltage rises in a positive direction, C9 charges at a constant rate through R24 and the C-R tube bias is reduced accordingly. As the sweep voltage suddenly falls and snaps the electron beam back to the left side of the screen, C9 rapidly discharges through R23, driving the cathode more positive, and biases the C-R tube below cutoff so that the return trace is invisible.

The synchronizing voltage applied to the grid of V3 stabilizes the screen pattern, as described in chapter 8.

APPLICATIONS

The cathode-ray oscilloscope is generally used to observe voltage waveforms in testing electrical circuits. The electrostatic cathode-ray tube employs voltage sources rather than current, to deflect the electron beam. For this reason the electrostatic type of cathode-ray tube is used in test oscilloscopes. The electromagnetic cathode-ray tube is a current-operated device. It is used in certain applications other than general testing, where its properties make it more suitable than the electrostatic tube.

To obtain an accurate representation of the voltage waveform, a few precautions must be observed. For the protection of both the operator and the oscilloscope, the approximate magnitude of the voltages in the circuit under test must be known. Dependable data can be obtained from the oscilloscope only if its sensitivity and its frequency characteristics are known. To make certain that the waveform will not be distorted, it is essential to understand how and why distortion takes place and that precautions be taken to minimize it.

The input to most oscilloscopes is between an input terminal and ground. The input terminal is coupled to the amplifier grid through a capacitor whose voltage rating rarely exceeds 450 volts. Therefore, unless the approximate magnitude of the voltage under test is known, damage to the oscilloscope through breakdown of the input capacitor may occur.

In some cases it may be necessary to observe waveforms in circuits where the voltage is much greater than that which the components within the oscilloscope can withstand. An external voltage divider may be used in such instances to reduce the voltage to a value that will not damage the equipment. In any case it is important that the oscilloscope be adequately grounded—a precaution that must be taken for the protection of the operator, because a failure of some part of the voltage divider can raise the potential of the whole oscilloscope to a dangerous level if the case is not solidly connected to ground.

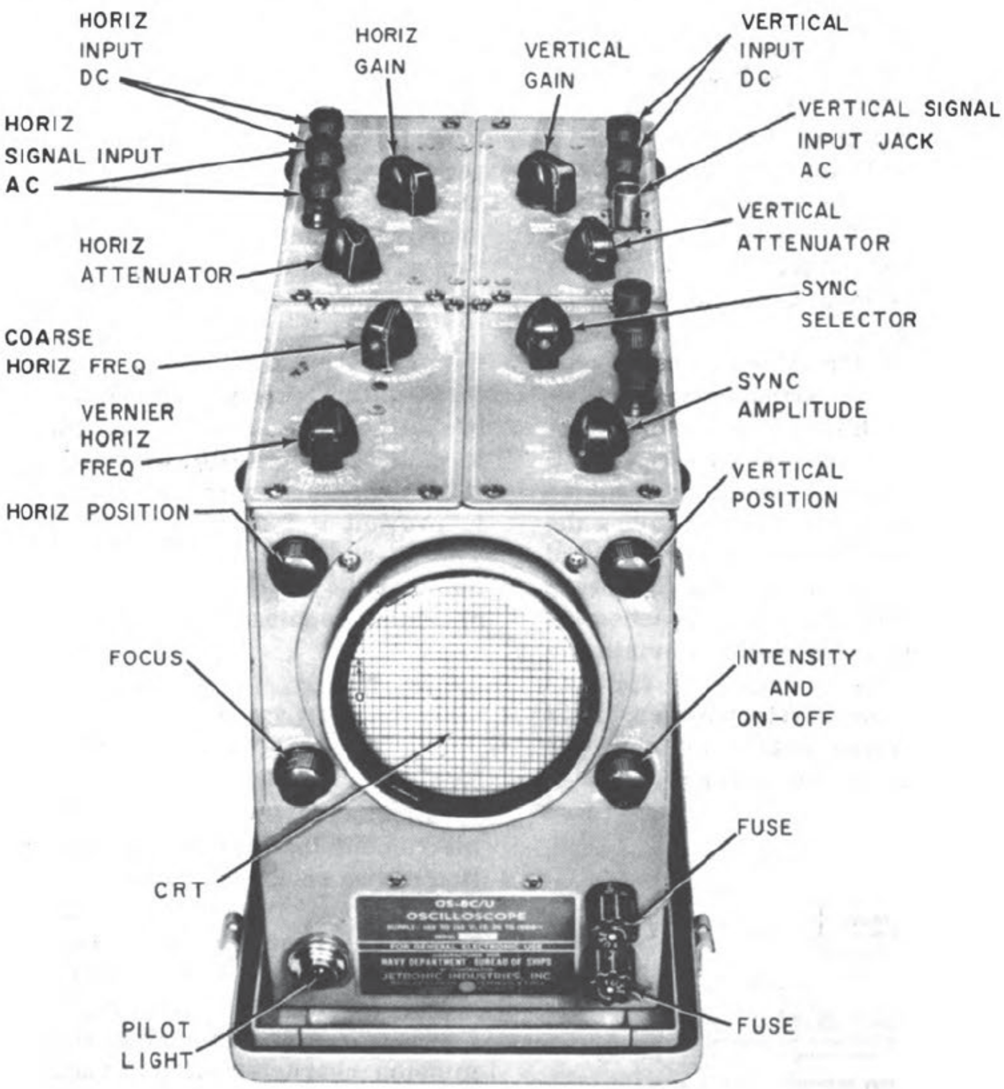
If a capacitance voltage divider is used, a wise precaution is to shunt each capacitor with a high resistance to maintain the proper voltage distribution across each capacitor.

The range of sweep frequencies in a given oscilloscope is usually indicated on the control panel. The sweep frequency generator in this example,—the Model OS-8C/U scope has a frequency range of 3 to 50,000 cps, as indicated on the dial for the switch that controls the coarse horizontal frequency. The frequency range that the vertical and horizontal amplifiers are capable of amplifying properly is given in the applicable manufacturer's instruction book. In this example, the vertical amplifier has a bandwidth of from 30 cps to 2 mc and the horizontal amplifier has a bandwidth of from 25 cps to 100,000 cps. Generally, only the best oscilloscopes use amplifiers that will amplify voltages whose frequency is below 30 cps or above 100,000 cps. Oscilloscopes that do not cover as wide a range of frequencies as the one shown in figure 16-7 may be satisfactory for most uses, but distortion is likely to occur when saw-tooth or rectangular waveforms of a high recurrence rate are investigated.

The deflection sensitivity may be expressed as the distance in millimeters that the spot is moved on the screen when 1 volt is applied across a pair of deflection plates. The deflection sensitivity of the vertical deflection plates in the oscilloscope shown in figure 17-7 is 0.528 millimeters per volt. Expressed in volts per inch of deflection, the deflection becomes $\frac{25.4}{0.528}$, or 48 volts per inch. The deflection sensitivity of the horizontal plates in this example is 0.379 millimeters per volt, or 67 volts per inch.

The deflection sensitivity may also be expressed as the input voltage to the amplifier (horizontal or vertical) for a deflection of 1 inch of the spot on the screen. In this case the amplifier gain control is adjusted to a suitable value that is arbitrary (for example, mid scale). In the example of figure 17-7, both the horizontal and vertical deflection sensitivity are 0.1 volt rms for 1 inch peak-to-peak deflection.

To avoid pick-up of stray signals the leads from the circuit under test should be as short as possible, and they should be shielded. The cathode-ray tube itself is shielded by the aquadag



20.326

Figure 17-7.—Oscilloscope controls and input connections of the OS-8C/U oscilloscope.

coating on the inside of the tube and a metal shield on the outside. A common side of the oscilloscope circuits is grounded and should be connected to a ground point in the circuit under test and to a good external ground connection.

Several causes of distortion are possible in the production of a cathode-ray tube display. Some of these causes are:

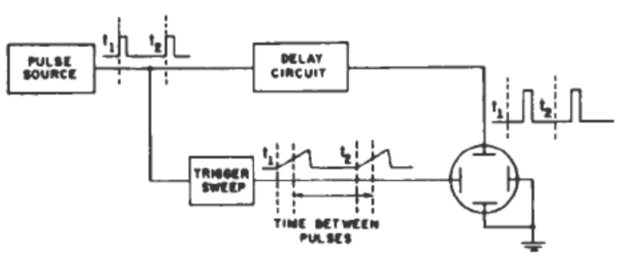
1. exceeding the bandwidth limitation of the deflection amplifiers;
2. a defective sweep generator;
3. excessive fly-back time;
4. excessive synchronizing voltage;
5. oscilloscope loading of a high impedance test circuit;
6. oscilloscope capacitance shunting of video amplifier test circuit;
7. use of a variety of oscilloscopes and test leads on one equipment; and
8. improper shielding of test leads.

SYNCHROSCOPE

COMPONENTS

A synchroscope is an improved type of cathode-ray oscilloscope with its horizontal sweep inoperative until initiated by a signal-burst, or pulse, from a trigger circuit. Oscilloscope AN/URM-24 is a representative, portable, field-type synchroscope used in bench-testing of radar and communication equipment. The sweep is arranged so that the electron beam rests with zero intensity at one side of the tube until the receipt of a trigger pulse, at which time the sweep begins and the beam is intensified by a square wave applied to the control grid. After passing across the tube the electron beam is reduced in intensity and quickly returned to the origin where it remains at zero intensity until the next trigger pulse.

A block diagram of a simplified synchroscope is shown in figure 17-8. A signal pulse triggers the sweep circuit, which has a saw-tooth waveform in which the time of one sweep is small compared with the time between successive sweeps. An alternate method of triggering the sweep is by a separate timing generator (not shown in the figure). In either case the sweep is initiated before the signal pulse reaches the deflection plates. The delay circuit permits the sweep circuit to be initiated before the application of the pulse to the vertical plates. This action causes the entire pulse, including the leading edge, to appear on the screen.



20.327

Figure 17-8.—Block diagram of a simplified synchroscope.

Synchroscopes have several calibrated sweeps, for example, the sweep may be 50 microseconds, 200 microseconds, or 1,000 microseconds in length. Such instruments are useful in observing the waveform of very short pulses like those from a radar equipment, or in observing the time interval between pulses and the duration of a pulse. With a suitable r-f coupling device and demodulator, a synchroscope can be used to observe the standing-wave ratio in a waveguide between a radar transmitter and antenna in tuning and checking the equipment for optimum performance.

ELECTRONIC SWITCHING

Occasions arise for simultaneously displaying the behavior of a pair of waveforms. An example of this need is for comparison of signal performances within stages of an amplifier. This may be done with a pair of scopes. However with the use of an electronic switch the two waveforms can be displayed simultaneously on the same screen.

An electronic switch is a device that utilizes the properties of gas-filled or high-vacuum tubes for rapidly closing, opening, or changing the operation of an electronic circuit. It may be used to GATE an amplifier circuit; that is, to cause it to function during a given period of time and to prevent it from functioning at other times. The on-and-off periods of operation may be the same or they may be different, depending on the operating requirements.

MULTIVIBRATOR USED AS ELECTRONIC SWITCH

Electronic Switch TS-433B/U is a portable instrument that permits simultaneous observation of two recurrent patterns on a cathode-ray oscilloscope. It contains a multivibrator with amplifier whose simplified circuits may be compared with figure 17-9,A. A square wave voltage of variable frequency and amplitude is available for use as a test signal in studying the transmission characteristics of vacuum-tube amplifiers and other circuits.

A multivibrator may be used as in figure 17-9,A, to cut an amplifier tube, V1, on and off at the multivibrator frequency. The waveforms shown in figure 17-9,B, indicate the manner in which V1 is gated. The multivibrator output, e_k, composed of essentially square waves, is developed across the cathode resistor of V3. Tube V3 conducts periodically and develops a positive gate voltage, e_k, between the cathode and ground of V1. The gate voltage cuts off V1 during the time V3 conducts. If the input to V1 is a series of regularly spaced positive voltage spikes, as shown in figure 17-9,B, the gate voltage will permit two pulses to pass through V1 and will block off two pulses. If the gate voltage existed for a longer period, three or more pulses could be passed and an equal number suppressed.

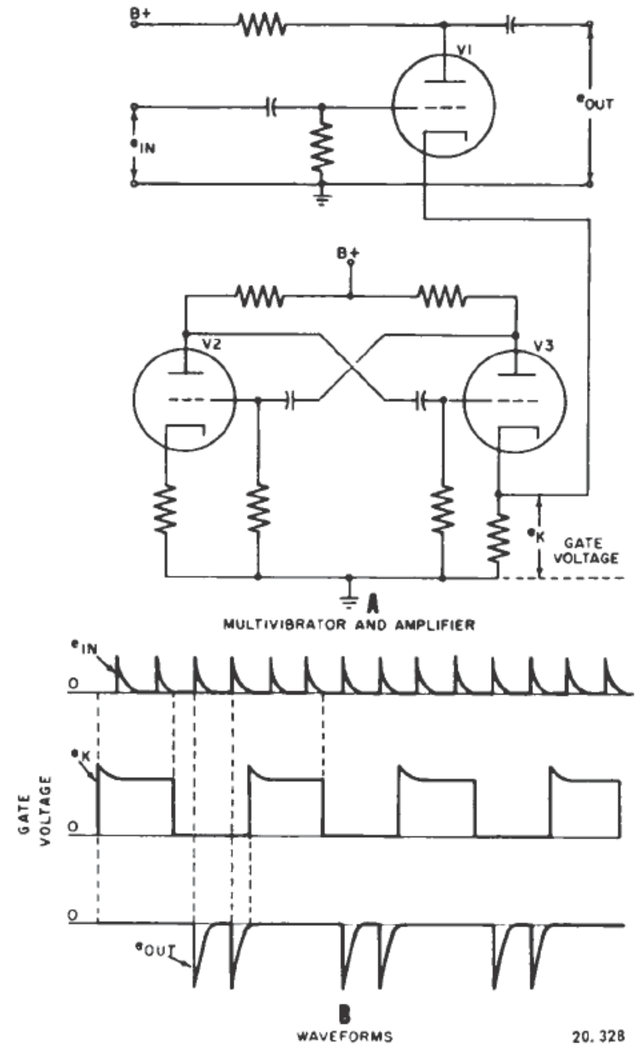


Figure 17-9.—Multivibrator electronic switch used to provide a gate voltage, and waveforms.

The switching action of a multivibrator may also be used to show, for purposes of comparison, two or more signals on a cathode-ray oscilloscope apparently at the same time. Although the signals are not actually present at the same time, they appear to be so because of persistance of vision of the human eye as well as persistence of the CRT screen.

An arrangement for presenting two signals at the same time is shown in figure 17-10. Tubes V1 and V2 with their associated circuits make up a conventional multivibrator. Tubes V3 and V4 are distinct amplifier stages having a separate input and a common output via C5 to the vertical deflection plates of an oscilloscope.

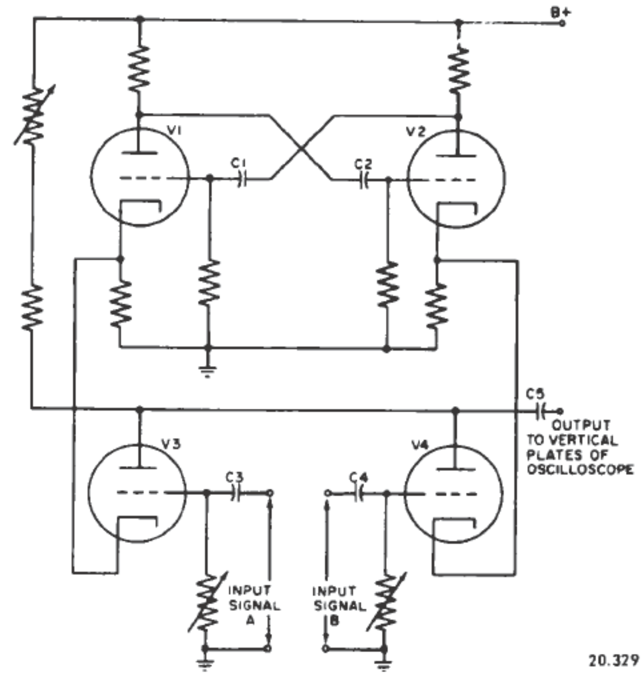


Figure 17-10.—Multivibrator electronic switch for use with a cathode-ray oscilloscope.

The circuit operates as follows: In order that the signal will be stationary on the oscilloscope, assume that input signals A and B have the same frequency, and that input signal A is used to synchronize the multivibrator. When V1 conducts, V3 is cut off, and the amplified signal from V4 appears on the oscilloscope. When V2 conducts, V4 is cut off, and the amplified signal from V3 appears on the oscilloscope. It is assumed that the time of one positive pulse from the multivibrator is equal to that of one or more periods of input signal A.

By connecting the output of the electronic switch to the vertical amplifier on the oscilloscope and by proper adjustment of the level controls on the electronic switch (not shown in the figure) the two signals are made to appear on the same horizontal axis on the oscilloscope screen. By adjusting the external synchronizing signal control on the oscilloscope the two signals are made to start at the same point on the horizontal axis. The external synchronizing signal may originate with either signal A or signal B.

ELECTRON-RAY TUBES

The electron-ray tube is used as a signal indicator in electronic test equipment such as capacity checkers and signal tracers. It is also used as a tuning indicator in radio receivers.

The electron-ray tube, or MAGIC EYE, contains two sets of elements, one of which is a triode amplifier and the other a cathode-ray indicator. The plate of the triode section is internally connected to the ray-control electrode (fig. 17-11,A) so that as plate voltage varies with the applied signal, the voltage on the ray-control electrode also varies. The ray-control electrode is a flat, metal strip so placed relative to the cathode that it deflects some of the electrons emitted from the cathode. The electrons that strike the anode, or target, cause it to fluoresce, or give off light. The deflection caused by the ray-control electrode prevents electrons from striking part of the target; thus a wedge-shaped shadow is produced on the target. The size of this shadow is determined by the voltage on the ray-control electrode. When this electrode is at approximately the same potential as the fluorescent anode, the shadow disappears.

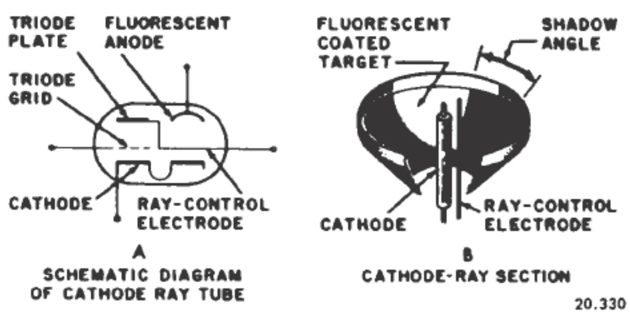


Figure 17-11.—Electron-ray tube.

If the ray-control electrode is less positive than the anode, a shadow appears, the width of which is dependent upon the voltage on the ray-control electrode. If the tube is calibrated, it may be used as a voltmeter when rough measurements will suffice. However, the principal uses of the magic-eye tube are as a tuning indicator, in receiving sets and as a balance indicator in bridge circuits.

Analysis of Tube Action

The width of the shadow angle depends upon the relation of the voltage between the ray-control electrode and ground compared to the voltage between a point on the electric field gradient and ground in the vicinity of the ray-control electrode, as indicated in figure 17-12,A.

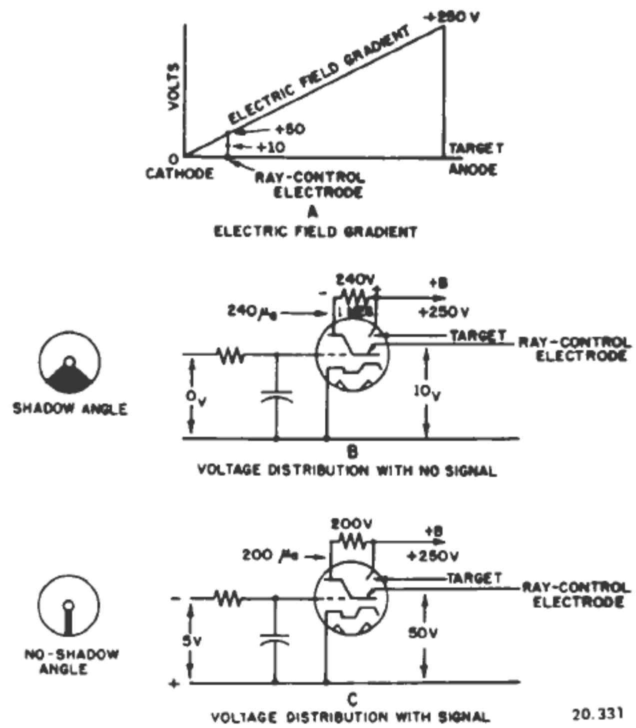


Figure 17-12.—Analysis of magic-eye tube action.

With no signal applied to the grid of the triode section, plate current is 240 μa (fig. 17-12,B). The voltage on the ray-control electrode is equal to the plate supply voltage less the drop through the 1-megohn resistor, or 250 - 240 = 10 volts. The electric field gradient is assumed to vary as a straight line starting at the cathode with zero potential and terminating at the anode with a potential of +250 volts with respect to the cathode. A point on the electric field gradient in the vicinity of the ray-control electrode has a potential of +50 volts with respect to ground. Thus the ray-control electrode is negative with respect to the field at this point by an amount equal to -(50-10), or -40 volts. The negative charge repels electrons and the shadow angle is established.

In figure 17-12,C, a 5-volt signal is developed between grid and ground of the triode section of the magic-eye tube. The plate current is reduced to 200 μa and the potential of the ray-control electrode is equal to 250-200, or +50 volts with respect to ground. Since the potential of a point on the electric field gradient in the immediate vicinity of the ray-control electrode is also +50 volts with respect to ground, there is no difference in potential between the control electrode and the field. Thus, the control electrode does not repel electrons and the shadow angle closes, indicating the signal voltage applied to the triode grid.

ABSORPTION WAVE METER

The names, wave meter, frequency meter, and echo box, are used for a device that is a calibrated electric resonator, the resonant frequency of which can be adjusted to known values. Each is used to measure either the wavelength, or the frequency of a radio wave or electric oscillation. When calibrated in units of centimeters or meters, the device is designated as a wave meter. Calibration such as this is convenient when a system of Lecher wires (see ch. 12) is used for the calibrating source. The modern trend is to calibrate the device in units of cycles-per-second and designate it as a frequency meter.

Test Set TS-480/U is designated as a portable frequency meter and is used to measure frequencies from 0.5 mc to 150 mc. It may also be used in neutralizing amplifiers, indicating stray r-f fields, and determining harmonic and parasitic oscillations. The electrical circuit of this tester is the simple series arrangement of B, C, and L shown in figure 17-13. A set of five detachable coils permits the selection of the correct inductance; L, for establishing resonance.

Absorption-type wave meters operate on the principle that the energy in an L-C-R circuit is absorbed by an adjoining L-C-R circuit. The amount of energy that is absorbed reaches a maximum value (coils fixed) when each circuit is resonant to the frequency in the tuned circuit under test, like the one shown in figure 17-13. Maximum current is indicated in the wave meter circuit by the indicator bulb glowing brightly. Therefore, the wave meter is held in the vicinity of the circuit under test and the dial of the variable capacitor, C, is adjusted until the greatest brilliancy is noted on the indicator bulb. The value of resonant frequency is then determined from the dial on the variable capacitor.

Absorption wave meters are not reliable for accurate measurements because they tend to detune the circuit under test. Since absorption wave meters tend to detune self-excited oscillator circuits when coupled closely, care must be used to achieve the greatest degree of accuracy by making the final tuning adjustment when the wave meter is farthest from the tuned circuit under test, yet close enough to produce a visible (faint) glow on the indicator bulb.

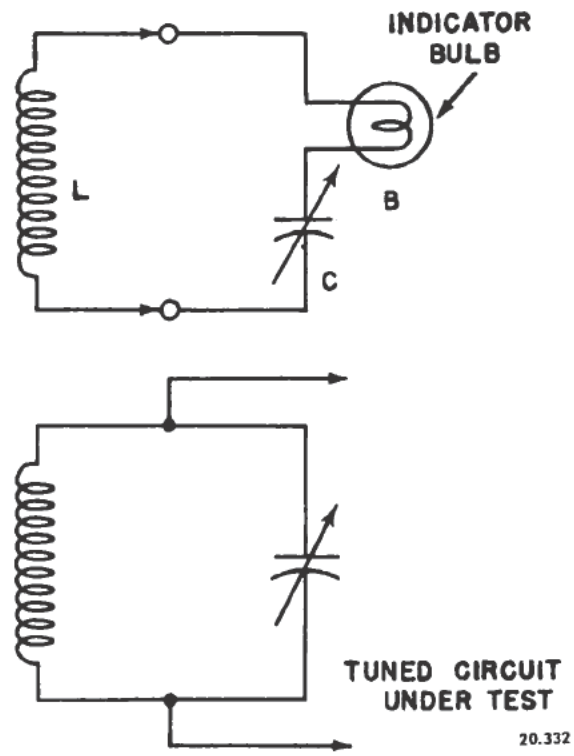


Figure 17-13.—Absorption wave meter.

A well-designed tester uses coils and associated parts with high Q-factor ratings because the accuracy with which frequency can be determined with a wave meter depends heavily on the Q of the wave meter circuit and precision with which the capacitor dial can be calibrated.

Wave meters can be operated as either absorption or reaction devices. For the latter purpose, a meter would be arranged in the source of r-f. One practical way for arranging such a meter is shown in figure 17-14. When the wave meter of figure 17-13 is tuned to resonance in the vicinity of L1 of figure 17-14, the reaction is registered on the meter, M, shown in the latter diagram. Because the absorbed power is removed from the r-f oscillator circuit, this meter shows a decrease and the needle will plunge or dip downward. This accounts for the name, grid-dip frequency meter that is known simply as a grid-dip meter.

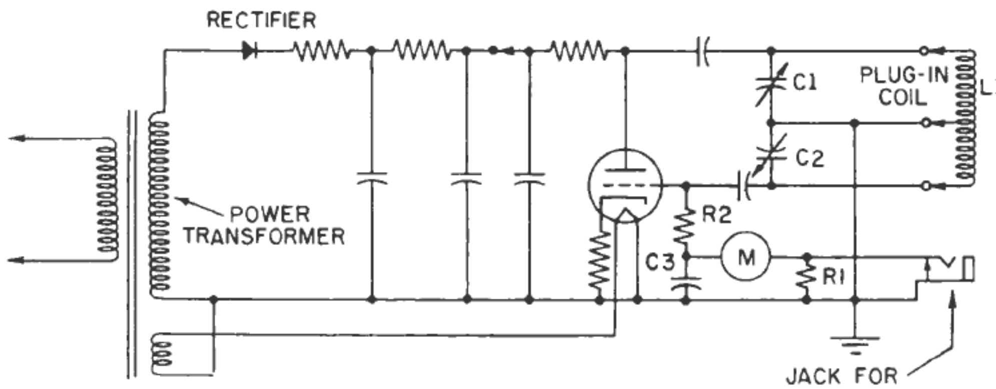


Figure 17-14.—Schematic diagram of a grid-dip meter.

GRID-DIP METER

The grid-dip meter is a versatile and popular test instrument, especially in the laboratory where development and research are conducted. A simplified circuit of the grid-dip device is shown in figure 17-14. Here, a Colpitts oscillator is used. A Colpitts oscillator uses two capacitors, C1 and C2, connected in series across the tank coil, L1, with the junction of the capacitors connected to the grounded end of the cathode resistor. Two coupling capacitors connect the tuned circuit to the plate and grid, respectively. The meter is in series with the grid-leak resistors; the resistors being selected to produce a time constant with the grid capacitor C3. The grid capacitor becomes charged to a degree that cuts off oscillations until its charge leaks away through the meter and grid-leak resistor, R1. Oscillations are restored for a period that depends on the time-constant value determined by R2 and C3. The net result is that the r-f oscillations are modulated with an a-f that is determined by the time- constant value. A jack is provided across resistor R1 for listening with a headphone to the a-f signal which changes in intensity. Some operators can detect small aural changes more readily than the amount of dip registered on the meter.

With this tester it is possible, among other things, to determine the resonant frequency of an antenna system, to detect harmonics, and to check relative field strengths; for example, in rotating a television antenna for maximum signal strength. The meter may also be used as an absorption frequency meter when the oscillator is not energized.

CIRCUIT OPERATION OF GRID-DIP METER

Basically the grid-dip meter is a calibrated oscillator which meters the grid current in the oscillator circuit. With the oscillator functioning, energy is coupled from the tuned circuit (composed of C1, C2, and L1) to the circuit under test. The circuit under test is supplied a small amount of energy via tank coil L1 of the meter. Except for the field of L1, the circuit under test is deenergized. The capacitors are rotated to the point where the oscillator tank frequency is equal to the resonant frequency of the circuit under test. At resonance the grid current decreases as indicated by the dip in the grid meter. The energy absorbed from L1 by the circuit under test decreases the a-c component of plate voltage, thus causing a decrease in feedback energy from the plate to the oscillator grid. The grid voltage is driven less positive and the grid current decreases.

APPLICATIONS OF GRID-DIP METER

In order to determine the resonant frequency of an antenna system, coil L1 of the meter is brought close to the antenna when the latter is deenergized. The proper point along the antenna is the point corresponding to a high-current point when the antenna is energized. If the antenna has an open center, a jumper is used as a temporary short so that the test may be made. The meter is tuned until the dip of the grid current indicates resonance. Harmonics may be indicated in the same manner.

Standing waves on an open transmission line may be checked by removing the plate power supply and operating the tube as a diode detector. This action is similar to that of an absorption wave meter.

FREQUENCY STANDARDS

WWV PRIMARY FREQUENCY STANDARD

When a major naval operation is planned, surface ships, submarines, carrier-based planes, and other units may be required to maintain radio silence until contact with the enemy is made. Separate movements are then synchronized by preset-communication frequencies. One of the most important assignments of a technician is to have radio receivers and transmitters accurately adjusted to the frequencies assigned by the task-force communications officer. This type of assignment involves the use of a secondary frequency standard of high accuracy, such as the AN/URM-82 or the LM-21, the LM-18, the LM-15 or LM-11. (Caution: Other frequency meters in the LM series are not acceptable as secondary standards.)

If the secondary meter is out of order (for example, as the result of a defective reference crystal), it becomes necessary to replace the crystal and verify the accuracy of the new reference crystal. The primary frequency standard is radio transmitter station WWV, operated by the National Bureau of Standards. Every frequency meter should be checked weekly against the WWV transmissions to assure that the reference crystal is still reliable. Standard-frequency transmissions are useful also as standards for measuring field intensity and audio frequencies.

The schedule of services offered by station WWV is published in Radio Navigational Aids, H. O. 205, issued by the USN Hydrographic Office. For current information refer to this publication. Revisions to the schedule of services are reported from time to time in nearly all naval electronic magazines.

SECONDARY FREQUENCY STANDARDS

The LM frequency meter is a secondary frequency standard of high accuracy. The simplified block diagram shown in figure 17-15 shows the basic components of the LM-18 meter and indicates how they function. This meter is fundamentally a stable, self-excited heterodyne oscillator of the electron-coupled type covering the range from 125 kc to 20 mc, and has a separate crystal-controlled reference oscillator. Zero beats are provided at several reference points between the two oscillators.

As indicated in figure 17-15,A, the crystal oscillator serves to check the heterodyne oscillator frequency. A small trimmer capacitor is connected in parallel with the main tuning capacitor and serves to correct the frequency of the heterodyne oscillator at the nearest crystal check point. The beat frequency between the two oscillators is detected, amplified, and fed to the headphones for aural indication. At zero beat, the heterodyne oscillator frequency is correct for the dial setting as indicated in the calibration book, for the selected crystal check point.

For transmitter adjustment, the heterodyne oscillator, after calibration, is combined with the r-f signal input from the transmitter, as shown in figure 17-15,B. Zero beat results when the transmitter is adjusted to the frequency of the oscillator. Aural indication is accomplished in the same manner as calibration of the heterodyne oscillator.

The equipment may serve as a signal source for alignment and calibration of receivers, as indicated in figure 17-15,C. After calibration to the nearest crystal check point, the output of the heterodyne oscillator is fed to the receiver. An r-f attenuator is provided for adjustment of the output signal level. The receiver is tuned to zero beat with the heterodyne oscillator, as indicated by the headphones in the output circuit of the receiver. By circuit switching, the audio amplifier serves as a 500-cycle modulator. For a-m input signals, receiver calibration is performed by tuning for maximum audio output.

The output of the LM meter is too low for some maintenance uses for which less accurate signal generators with higher output power are provided. Since the oscillator may be set within 1 part in 10,000 for the range from 2 to 20 mc, careful use of the LM meter should result in close agreement of all transmitters and receivers set to any one frequency.

Other secondary frequency standards used by the Navy include the AN/USM-29 and TS-186/UP. The AN/USM-29 is a portable instrument used in calibrating frequencies of radio transmitters and receivers. It is used in field and depot maintenance. The AN/USM-29 is identical

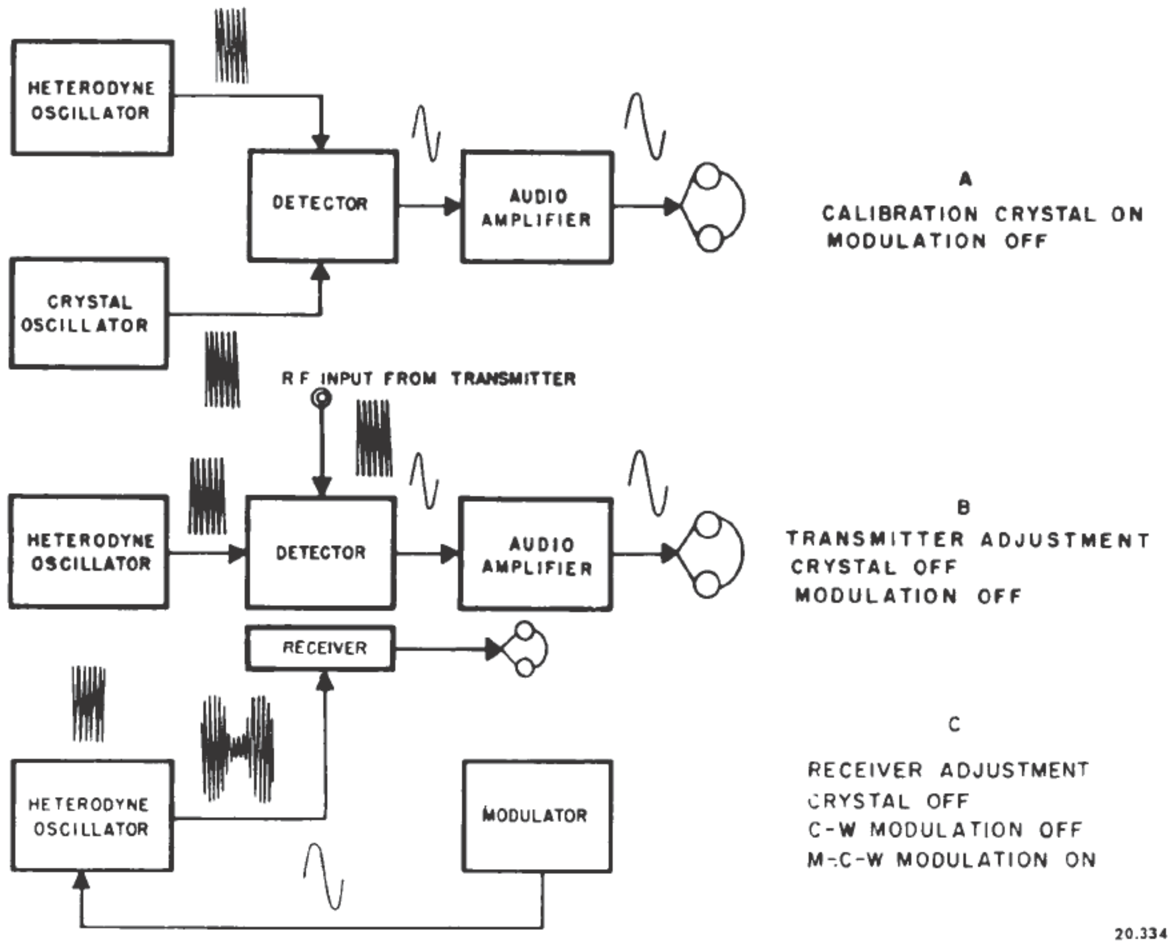


Figure 17-15.—Simplified block diagram of the LM-18 meter.

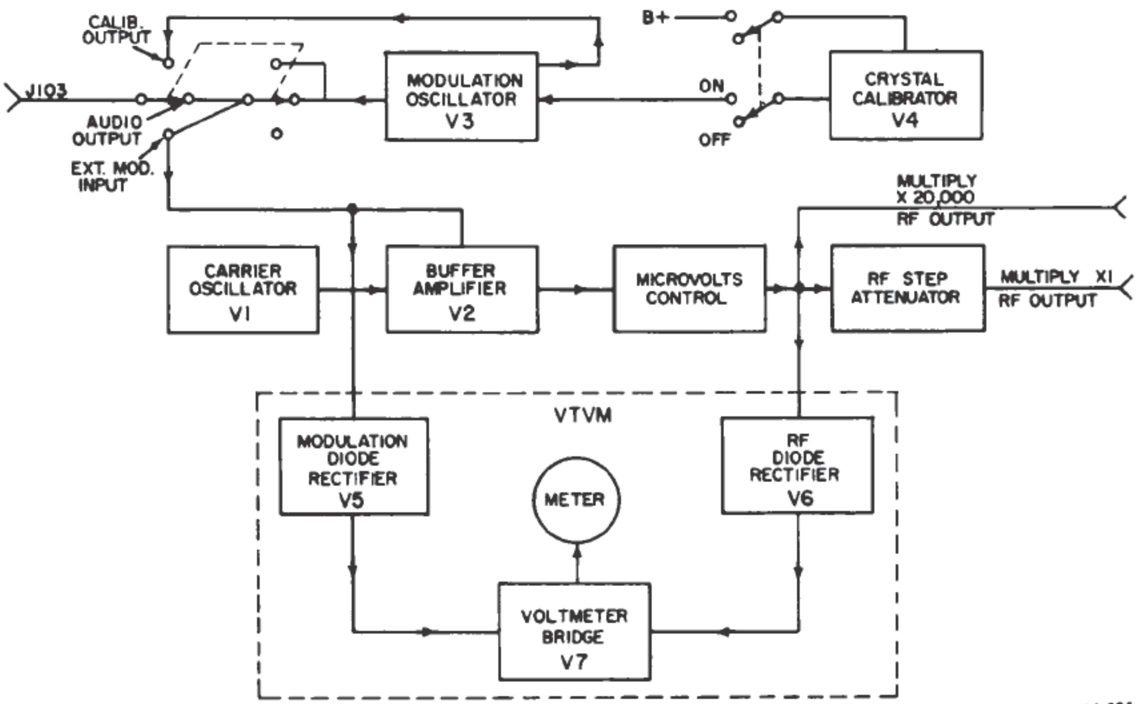
to Frequency Meter FR-47/U, except for additional accessories, and supersedes the Crystal Calibrator Equipment, Model LR. The AN/USM-29 has a fundamental frequency range from 15 kc to 30 mc, with an overall accuracy of 0.01 percent, and its harmonics to the tenth are acceptable for checking frequencies from 30 mc to 300 mcs. The Test Set TS-186/UP covers the range from 100 to 10,000 mc having an overall accuracy of 0.01 percent.

R-F SIGNAL GENERATOR

The output power of a secondary frequency standard or frequency meter is too low to energize directly an output meter when radio equipment is being serviced. Signal generators, such as Navy types AN/URM-25 and AN/URM-26 have sufficient power for general service work. Some signal generators include vacuum-tube voltmeters circuits, modulating circuits, and others. The various test sets (for example, TS-382C/U and TS-535/U) use auxiliary circuits for specialized test work.

Basically an r-f signal generator is a device which produces r-f voltages from a well-shielded oscillator. It can be set to generate an unmodulated or audio-tone-modulated signal for test purposes at any frequency for aligning or servicing r-f stages in receivers or transmitters. If the calibration of the equipment is sufficiently accurate, and if a detector circuit with one or more audio stages is included, the device may be used to determine unknown frequencies by the zero-beat method.

To afford a better understanding of the construction and function of an r-f signal generator, the Navy portable model AN/URM-25B signal generator is described briefly. A simplified block diagram of this generator is shown in figure 17-16. The arrangement of parts and their functions are similar to those of a radio



20.335

Figure 17-16.—R-f signal generator AN/URM-25B, block diagram.

frequency transmitter. A carrier frequency oscillator, V1, generates a variable r-f signal which is applied to the control grid of a buffer amplifier. V3 is the audio modulation oscillator. It is a standard Wien-bridge oscillator described in chapter 8, which generates an audio voltage with a choice of two frequencies (400 or 1000 cps) which is applied to the grid of the buffer amplifier, V2, when desired. Otherwise, V3 may be switched to an OFF postion or switched to deliver the a-f signal to the output jack, J103.

The output of the buffer amplifier, V2 (either modulated or unmodulated) is fed to the r-f attenuator system, consisting of a step attenuator with dual potentiometer (microvolts control) for delivering calibrated amounts of output signal to two jacks, J101 or J102.

The range of r-f frequencies covered is from 10 kilocycles to 50 megacycles in 8 bands. The r-f oscillator tube V1, also shown in figure 17-17, is a type 6J6 dual triode with both triode-sections connected in parallel. It uses the conventional Hartley oscillator described in chapter 8. Any one of 8 inductance coils, represented by L2, may be selected by switching it in parallel with the tuning capacitor, C7. The trimmer capacitor, C10, permits precise calibration during initial calibration. Each of the individual inductance coils is provided with its own trimmer.

Bias for the V1 oscillator tube is produced by the oscillator grid current which charges the grid blocking capacitor, C6. This capacitor receives its charge during the portion of the r-f cycle when the grid is positive. The grid resistor, R3, in parallel with C6 allows this capacitor to discharge during the reversal of the r-f cycle. The net result is a bias on the grid which is proportional to the amplitude of the r-f voltage across the grid tank circuit.

A vacuum-tube-voltmeter (VTVM), in figure 17-16, receives rectified signals from either the modulation diode, V5, or from the r-f diode, V6. The a-c voltages produced from low and high frequencies are converted to rectified currents that are filtered and applied to a balanced bridge network, as described later for the a-c voltmeter of figure 17-30. The VTVM permits accurate selection of voltage values for injection into circuits to be tested.

A-F SIGNAL GENERATOR

An a-f signal generator that serves as a dependable source of alternating voltage, with

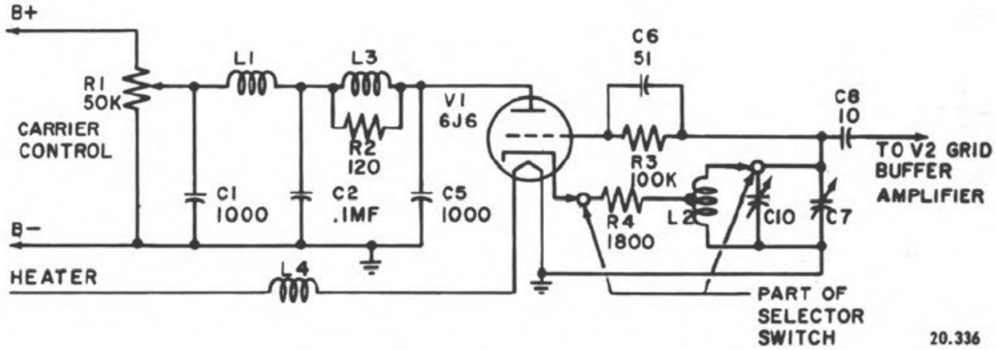


Figure 17-17.—Hartley type of r-f carrier oscillator in AN/URM-25B.

waveforms of known shape, from a few volts to a few millivolts, is a useful device for testing audio frequency (a-f) amplifiers.

An a-f signal is sometimes available from a tester that serves another basic purpose. The AN/URM-25B instrument described earlier is an example where the basic instrument is an r-f generator but a-f is also available from the output jack, J103, shown in figure 17-16. Here the choice of a-f is restricted to either 400 or 1000 cps with sine-wave waveform. Rather than being restricted to a single frequency output, an a-f generator is more useful when it can deliver a-c voltages ranging from 20 cps to 200,000 cycles per second at amplitudes which may be varied as desired from zero to 10 volts. The TS-239/UP audio frequency generator is capable of delivering these frequencies with a high degree of accuracy.

The TS-535/U is also classed as a special a-f generator, designed to produce sine waves of known frequencies and amplitudes in the upper audio and supersonic bands, ranging between 7 kc and 160 kc. The output is either c-w or m-c-w. The modulation is obtained from either the internal 400 cps (fixed) oscillator or an external source.

One of the uses of the TS-535/U is the testing of sonar receivers. Its internal a-f calibrator oscillator, with precise frequency of 5000 cps, may be used as a frequency calibrator of other external generators.

The accuracy of any signal generator depends on careful design. Let us examine an a-f oscillator that produces a single frequency of 5000 cps for calibrating purposes in the TS-535/U. The circuit is shown in figure 17-18 where the 6J5 tube supplies energy to a tuning fork in order to maintain the vibration of the fork. The plate circuit of this tube is connected to one of the fork driving coils, L1, while the grid circuit is connected to the other driving coil, L2. The core of the pair of coils is a permanent horseshoe magnet, the ends of which are set close to the tines of the tuning fork.

A tuning fork functions somewhat like a quartz crystal in that it vibrates at a single

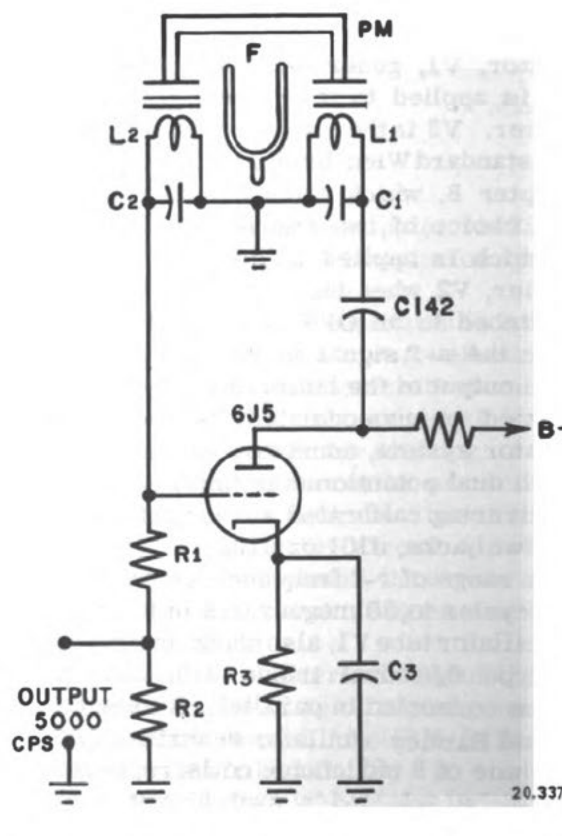


Figure 17-18.—Fork oscillator of TS-535/U.

frequency (in this case 5 kc) and maintains that frequency within very close limits (0.01%). The tuning fork used in this instrument is of bimetallic construction, designed so that the fork has a very low temperature coefficient, comparable to that of good quality quartz plates.

The operation begins when plate currents flows and capacitor, C142, discharges. This action produces a changing current through the coil, L1, which causes the tine of the fork to move. The movement of the fork changes the reluctance of the magnetic circuit in the grid coil, L2, with the result that a voltage is produced and applied to the grid circuit of the 6J5 tube. The grid coil, L2, is connected between grid and ground so that the generated sine-wave voltage on the grid is amplified and appears 180 degrees out-of-phase in the plate circuit from where it passes to the plate coil, L1, changing the current further to repeat the cycle.

Once a tine of a tuning fork is put into vibration, its neighboring tine also vibrates. You may wish to verify this with a dinner fork, or kitchen fork—a fork with two or four tines may be used. Pluck one of the tines—listen—also touch a neighboring (unplucked) tine to verify that it vibrates too. This action is called mechanical resonance.

Each L-C circuit is associated with one of the mechanical tines of the fork, F. Therefore, in addition to the voltage induced in L2 (explained before) there is another voltage induced in L1. We can now comprehend that the multiple-resonance action takes place in four places in figure 17-18, even when the 6J5 tube is removed, provided any one of the tines (or electrical L-C circuits) is disturbed. Re-insertion of the tube will reinforce the original resonance by the amplifying action of the tube. Sufficient energy appears across R2 for use as a precise 5000 cps signal.

Wien Bridge A-F Oscillator

An oscillator that uses a Wien-bridge circuit (ch. 8), in order to maintain a stable output, must use R-C components that maintain constant electrical properties of resistance and capacitance, respectively.

The TS-382A/U audio oscillator (front panel shown in figure 17-19,A) is an example of an instrument that has similar counterparts made by other manufacturers, designated as TS-382B/U, TS-382C/U, and TS-382D/U. The oscillator circuit is shown in figure 17-19,B, and then re-drawn as a Wien-bridge in figure 17-19,C, to show that the phase-shifting element of the circuit is a frequency selective bridge. However, it is simpler to use the circuit shown in figure B for purposes of discussion, since the feedback paths are shown more clearly.

Tube V101 is the oscillator tube. Tube V102 acts as an amplifier and inverter. You will recall from chapter 8 that the system amplifies voltages of a very wide range of frequencies. Voltages of any frequency, or of any combination of frequencies, can cause oscillation. The bridge circuit is used, therefore, to eliminate feedback voltages of all frequencies except the single frequency desired in the output.

A degenerative feedback voltage is provided by the voltage divider, consisting of R114 and the heat-sensitive resistor (that is actually a 3-watt lamp) R115. There is no phase shift across this voltage divider, regardless of frequencies, with the result that the amplitude of the negative feedback voltage (coupled through capacitor C108) maintains a constant output for all frequencies within the range of the equipment.

APPLICATION OF A-F OSCILLATOR

An audio oscillator is used in testing and trouble shooting amplifiers, audio sections of radio receivers, and radio transmitters, recorders, and filters. The overall performance of such equipment is determined by plotting their response curves. The variable-frequency a-f generators may also be used to supply audio modulation to r-f signal generators. Such a-f gear is also useful for checking sonar transducers, sonar networks, and servomechanisms.

RADIO-INTERFERENCE FIELD-INTENSITY METER

Field strength meters covering the various frequency ranges have been developed for locating r-f interference. An example of an h-f and v-h-f radio-interference field-intensity meter is the TS-587A/U.

A field intensity meter is essentially a portable radio receiver with an attached meter to indicate the strength of the received signal. It is useful in locating the source of an interfering signal on own ship; for testing the

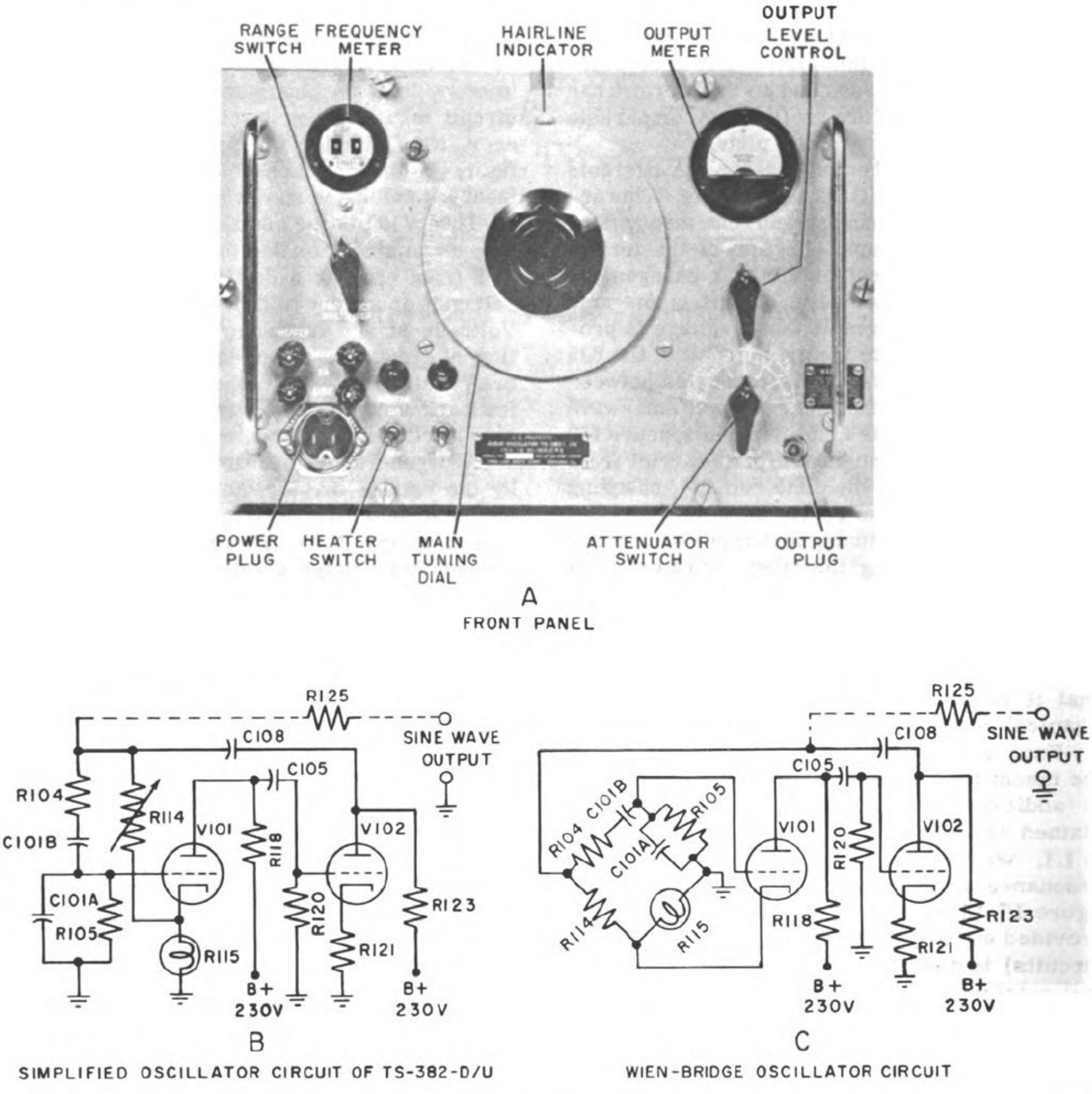


Figure 17-19.—Audio oscillator TS-382D/U.

effectiveness of measures for eliminating interference; for seeking out the sources of radiation on own ship that violate radio silence; and for plotting field patterns for directive antenna arrays.

Figure 17-20 illustrates one use of the AN/URM-47 interference locator, and figure 17-21 is a block diagram of this equipment, which is representative of the class. Other models have been developed, such as the AN/PRM-1 and AN/URM-17, that have wider frequency ranges and greater sensitivity.

The block diagram of the AN/URM-47 (fig. 17-21) indicates that it is essentially a superheterodyne radio receiver. The circuit is conventional except for the addition of the r-f attenuator and the meter circuits. The noise generator is used for calibration purposes. It

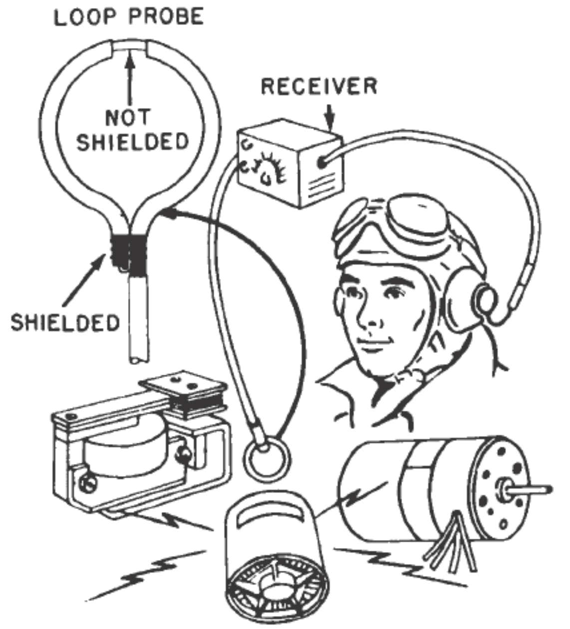


Figure 17-20.—Location of radio noise.

duplicates the strength of signal under test. Because the former is known, the latter becomes known by comparison.

Control-grid bias for the 2nd and 3rd i-f tubes (having remote cutoff characteristics) is supplied by the a-g-c system. This feature enables the voltage developed across the diode load resistor (in the 2nd detector) to be stabilized at a value approximately proportional to the logarithm of the input signal. This action is necessary in order to provide an indicating meter scale with a range extending from 10 to 1,000 microvolts, with uniform reading accuracy over the entire range.

A diode is used to supply the a-g-c voltage, and a portion of this voltage is fed to the d-c amplifier which has the indicating meter in the plate circuit. Thus, while the meter reads the rectified a-g-c voltage, it is effectively reading signal voltage because, as mentioned previously, the a-g-c voltage is proportional to the logarithm of the signal voltage. A battery is provided in the plate circuit of the diode to buck out the indicating meter current under zero input conditions. In this way, the meter is set for zero under conditions of no-signal input. Another diode rectifies the signal to provide an audio output for the headphones.

MEASUREMENT AND LOCATION OF INTERFERENCE

In locating a source of radio interference, tune in the radio noise on a field-intensity meter. Identify the signal with earphones as in figure 17-20. An electrostatically shielded loop probe (shown enlarged in fig. 17-20) may be used as an antenna to locate sources of noise in machinery. Moving the probe in the direction of the source (or some conductor radiating the noise energy) causes the signal strength, as indicated by the meter or earphones, to increase. Moving the probe away from the source causes the signal strength to decrease. Inspection of all rotating equipment usually is necessary to locate interference on shipboard, and a final check should be made by starting and stopping the suspected device.

Occasionally v-h-f transmitters indirectly cause interference in lower frequency bands. For example, recently a search radar was causing interference with all TBS and all other radio reception on an aircraft carrier. The annoyance was caused by two loose steel-wire stays near the radar antenna. Energy picked up from the antenna caused an arc which was reradiated at a lower frequency.

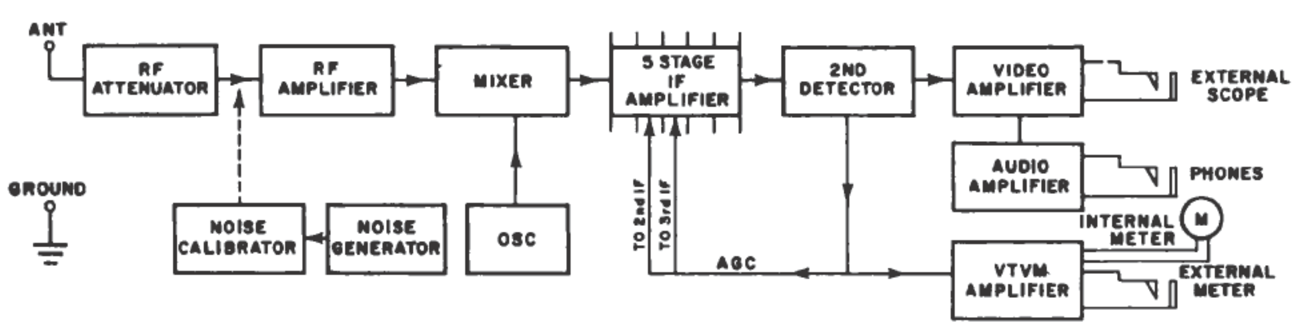


Figure 17-21.—Block diagram of model AN/URM-47 interference locating equipment.

In many commercial broadcast receivers the antenna is coupled to the mixer, thus permitting radiation in the antenna from the local oscillator. The use of such receivers might produce radiation that could be detected by the enemy. Therefore all Navy receivers approved for shipboard use must have the oscillator separated from the antenna circuit by sufficient preselection and shielding to reduce radiation from the antenna to less than 400 micromicrowatts. The exact method of measuring this radiation is set forth in specifications provided by the Bureau of Ships. Electric razors and other sources of interference not only cause difficulties in radio reception but also interrupt radio silence and render a ship vulnerable to attack.

It is the responsibility of the electronics officer to make certain that no equipment on the ship is emitting a radio signal on which an enemy direction finder can be trained.

SPECTRUM ANALYZER

A spectrum analyzer is an electronic test equipment which provides a visual indication of the frequency component (spectra) of an amplitude-modulated radio wave. The radio wave may be modulated by keying, by voice, by radar pulses, and so forth. In every case the resulting waves include a carrier frequency and its associated upper and lower sideband components. The pattern on the screen of a cathode-ray tube is a graph of signal voltage versus frequency. Ordinates represent peak voltage, and abscissas represent frequency.

A simplified block diagram of one form of spectrum analyzer is shown in figure 17-22. The input signal is converted in the mixer to the difference frequency (IF) between the input signal and local oscillator. Any modulation present in the input signal is transferred to the intermediate frequency amplifier, removed from it at the detector, amplified by the video amplifier, and applied to the vertical deflection plates of the cathode-ray indicator tube. The action is similar to that of a conventional superheterodyne receiver except that the local oscillator is frequency modulated. In other words, the frequency of the local oscillator is constantly being varied at a rate and to an extent that is determined by the sweep generator. The sweep generator has a saw-tooth waveform and sweeps the spot across the screen at a rate that is proportional to the rate of change of frequency of the local oscillator. Thus the horizontal position of the spot at any instant is proportional to the frequency of the applied signal at that instant and the vertical position of the spot is proportional to the amplitude of the signal.

Spectrum analyzers are frequently used to study the radio frequency spectrum produced when the carrier is amplitude modulated by a succession of rectangular pulses as in radar signals. A characteristic pattern of the pulse spectrum of a radar signal is shown in figure 17-23.

The local oscillator of the spectrum analyzer superheterodyne receiver is swept in frequency at a rate that is proportional to the radar pulse

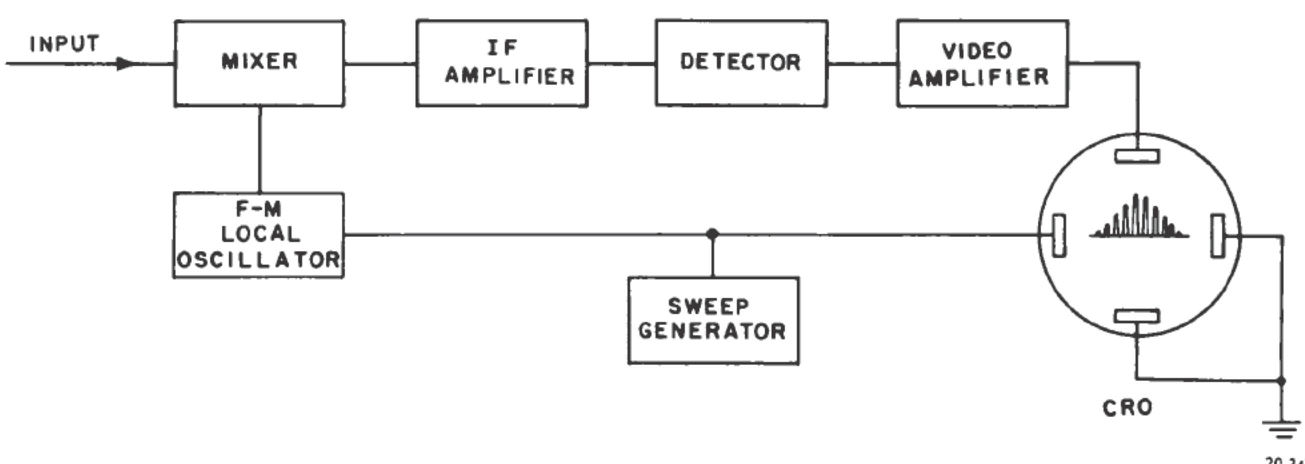
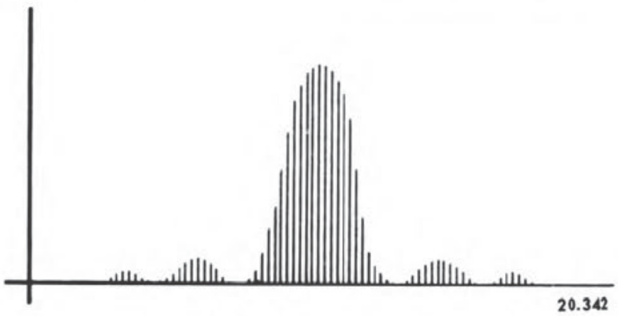


Figure 17-22.—Simplified block diagram of one form of spectrum analyzer.

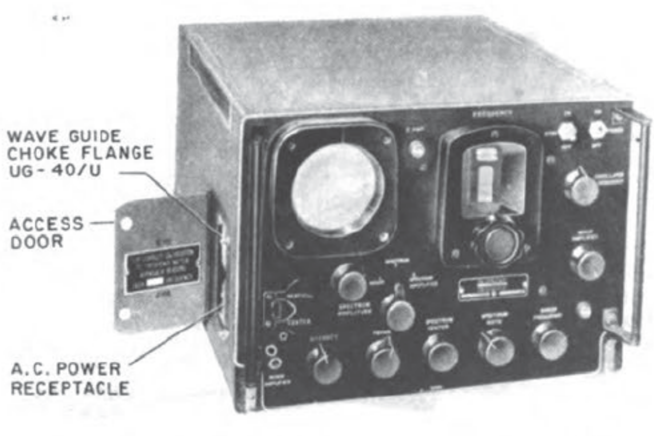


20.342

Figure 17-23.—Radar pulse spectrum.

recurrence frequency. The pattern is an envelope formed by the succession of pulses that are received during the time the spot is swept across the screen. In order to provide sufficient detail in the screen pattern the sweep interval is made long enough to allow at least 50 pulses to occur for each sweep interval. Thus, the sweep speed should not exceed one-fiftieth of the pulse-recurrence frequency. To avoid flicker, a long-persistence screen is used.

The TS-148/UP spectrum analyzer is a representative electronic test equipment used with radar and beacon equipment and provides a visual indication of the spectra of radio frequency oscillators within a range of 8,470 to 9,630 megacycles. A pictorial view of this spectrum analyzer control panel is shown in figure 17-24. This analyzer is useful for observing spectra of pulsed magnetrons; measuring magnetron frequencies; tuning waveguides in a radar transmitter; checking frequency meters, TR boxes, and echo boxes; measuring



20.343

Figure 17-24.—Spectrum analyzer control panel, TS-148/UP.

the bandwidths of resonant cavities and pulse widths; and determining the distribution of useful transmitted power.

CAPACITANCE-INDUCTANCE-RESISTANCE BRIDGES

Capacitance, inductance, and resistance are measured for precise accuracy by alternating current bridges which are composed of capacitors, inductors, and resistors in a wide variety of combinations. These bridges operate on the principle of the Wheatstone bridge, in which an unknown resistance is balanced against known resistances. The unknown resistance is calculated in terms of the known resistance after the bridge has been balanced. One type of capacitance bridge circuit is shown in simplified form in figure 17-25. When the bridge is balanced by adjusting the two variable resistors, there is no a-c voltage developed across the input of the indicator tube, V1, and the shadow angle is maximum. Any slight unbalance produces an a-c voltage, which, in turn, develops a grid-leak bias and lowers the plate current of V1, reducing the shadow angle.

The following relations exist when the bridge is balanced:

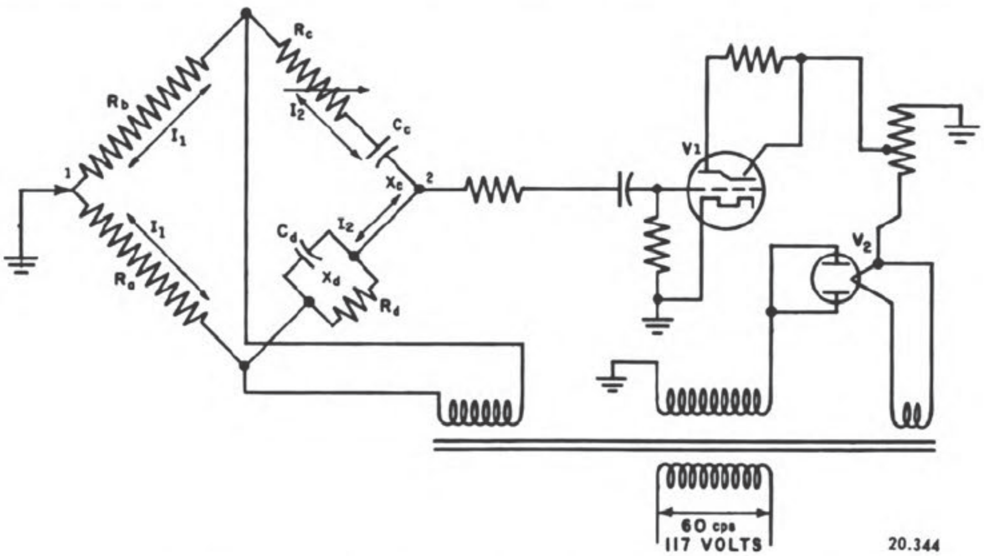
$$\frac{C_d}{C_c} = \frac{R_b}{R_a} - \frac{R_c}{R_d} \qquad (17\text{-}1)$$

and

$$\omega^2 = \frac{1}{R_d R_c C_d C_c} \qquad (17\text{-}2)$$

where R_a, R_b, R_c, and R_d are the resistances indicated in the figure; C_c is the standard capacitance; and C_d the unknown capacitance. $\omega = 2\pi f$, where f is the frequency of the voltage applied across the bridge.

In the basic Wheatstone bridge circuit using d-c voltages and simple resistances the balance is obtained when the voltage drops across the ratio arms are equal. In the a-c capacity bridge it is not sufficient to have equality of voltage drops in the ratio arms, but in addition the phase angle between current and voltage in the two arms containing the capacitors must be equal in order to obtain a balance. When a balance is obtained, the current in R_a is equal to that in R_b and the current in C_c is equal to the current in the parallel circuit of C_d and R_d.



20.344

Figure 17-25.—Simplified schematic of capacity checker.

The capacitance-inductance-resistance bridge, type ZM-11/U, shown in figure 17-26, is widely used to measure C, L, and R values in addition to special tests, such as the turns ratio of transformers and capacitor quality tests. This instrument is self-contained except for a source of line power, and has its own source of 1,000-cps bridge current with a sensitive bridge balance indicator, an adjustable source of direct current for electrolytic capacitor and insulation resistance testing, and a meter with suitable ranges for leakage current tests on electrolytic capacitors.

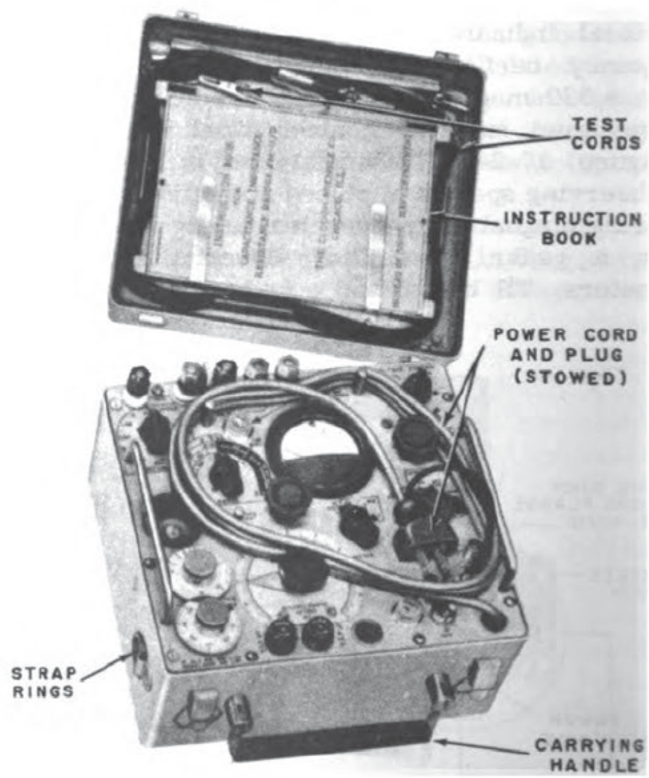
TUBE TESTERS

Two types of tube testers are in general use. One type, the EMISSION-TYPE TESTER, indicates the relative value of an electron tube in terms of its ability to emit electrons from the cathode. The second and more accurate type, is the MUTUAL-CONDUCTANCE (or TRANSCONDUCTANCE) tube tester. This tube tester not only gives an indication of the electron emission, but also indicates the ability of the grid voltage to control the plate current.

CIRCUITS FOR TUBE TESTS

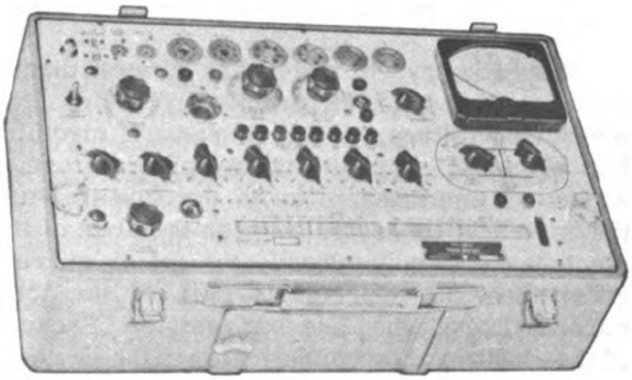
Tube Tester TV-3B/U, shown in figure 17-27, is a portable tube tester of the dynamic mutual-conductance type designed to test and measure the mutual conductance of electron tubes of the receiving types and many of the smaller transmitting types.

A multimeter section, using the same indicator, is also incorporated in the equipment to

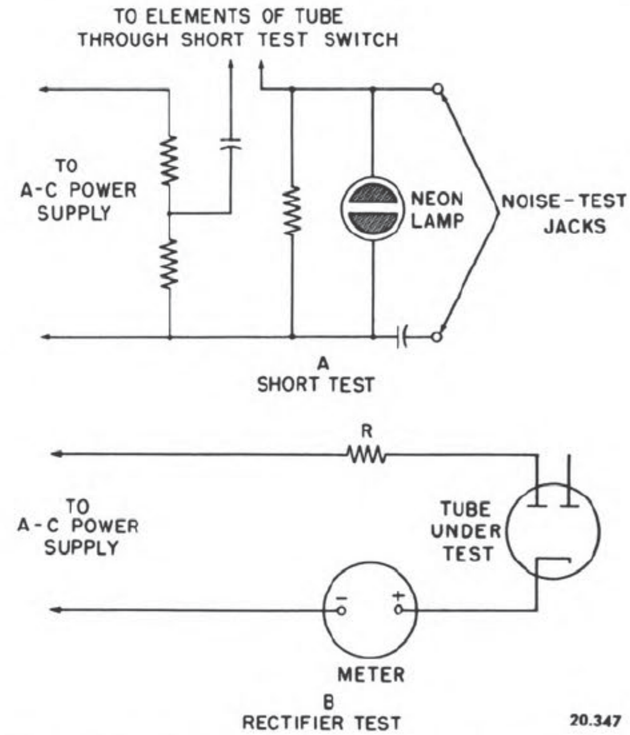


20.345

Figure 17-26.—Capacitance, inductance, resistance bridge, ZM-11/U.



20.346
Figure 17-27.—Tube tester TV-3B/U.



20.347
Figure 17-28.—Simplified short-and rectifier-test circuits.

permit measurements of a-c and d-c volts, d-c miliamperes, and resistance and capacitance in a number of ranges.

Line voltage applied to the primary of the power-supply transformer is adjusted by a variable resistor in series with the primary power leads. The line adjustment switch connects the meter so that the meter deflection is proportional to the magnitude of the applied line voltage.

The shorts switch connects the various tube elements to a voltage source in series with a neon lamp so that it glows if there is a short between the elements. The simplified circuit is shown in figure 17-28,A.

The noise-test jacks, shown in figure 17-28, A, may be connected to the antenna and ground posts of a radio receiver for the noise test. The short-test switch is turned through the various positions as the tube under test is tapped gently. Any intermittent disturbances between the electrodes cause momentary oscillations that are reproduced by the loudspeaker.

Rectifier and diode detector tubes are tested for emission as shown in the simplified circuit of figure 17-28,B. The tube being tested rectifies the alternating current and causes a pulsating direct current to flow through the meter. The current indicated by the meter is proportional to the electron emission of the tube. Rectifiers of the cold cathode type (such as the OZ4) require an a-c supply of 330-volts; whereas, diode detectors like the 6AL5 require only about 20 volts. In each instance, if there are two or more plates in the tube, each is tested separately.

The simplified circuits for the mutual-conductance test are shown in figure 17-29. The proper d-c grid voltage for the tube under test is supplied by V2. The a-c signal voltage is developed in L5 of the power transformer and acts in series with the grid bias.

The plate voltage for the tube under test is supplied by V1. In the absence of a grid signal the a-c voltage on the d-c meter will cause no deflection because the current in the two sections of R are equal and flow in opposite directions. However, when an a-c signal is applied to the grid of the tube under test, the plate current alternately increases and decreases through resistor R in phase with the grid signal voltage. Thus, the currents through the two sections of R become unbalanced and the voltage across the meter is equal to the difference in voltage across the two sections of R. The alternating deflecting force on the meter is thereby unbalanced and the indication is no longer zero. The meter indication is proportional to the average increase in plate current and is calibrated in micromhos.

The normal screen voltage of 130 volts is excessive for testing certain tubes. In such

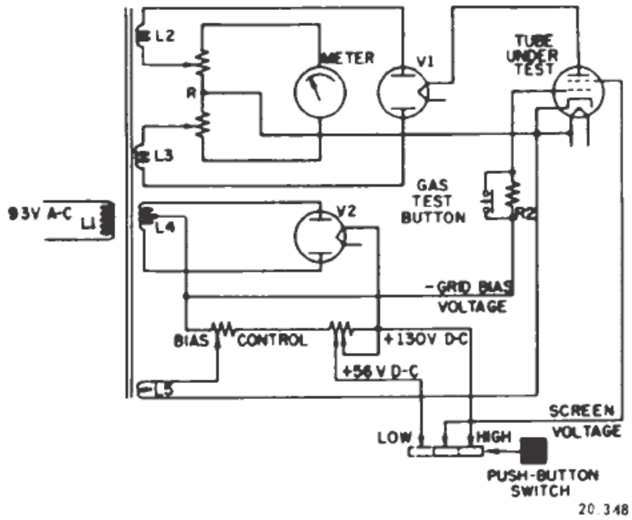


Figure 17-29.—Simplified mutual-conductance and gas-test circuits.

cases the screen may be connected to a 56-volt source by the push-button switch as indicated in figure 17-27.

Also, a simplified gas-test circuit is included in figure 17-29. Depressing the gas-test push-button inserts resistor R2 in series with the grid. If the tube is gassy, reversed grid current will flow through R2. The drop across R2 is opposed to the grid bias voltage and plate current will increase.

MULTIMETER SECTION

The multimeter section of the tube tester includes a standard nonelectronic multirange volt-ohm-milliammeter. The theory of operation of these meters is included in training manuals for basic electricity.

VOLT-OHM-AMMETER (ELECTRONIC)

The electronic volt-ohm-ammeter incorporates several measuring instruments within one enclosure. The ohmmeter and ammeter sections are similar to those described in manuals for basic electricity. The theory of operation of the ohmmeter and ammeter does not involve the electron tube and for that reason is not discussed in this training manual. The voltmeter section is described in this chapter because it involves the operation of the basic triode amplifier.

VOLTMETER ERRORS

The ordinary voltmeter has several disadvantages that make it practically useless for measuring voltages in high-impedance circuits. For example, suppose that the plate voltage of a pentode amplifier is to be measured. When the meter is connected between the plate of the electron tube and ground, the meter current constitutes an appreciable part of the total current through the plate load resistor. Because of the shunting effect of the meter on the pentode, the plate voltage decreases as the current through the plate load resistor increases. As a result, an incorrect indication of plate voltage is obtained.

Before the voltmeter is connected, the plate current is limited by the effective resistance of the plate circuit and the plate voltage. If the tube has an effective resistance of 100,000 ohms, the plate load a resistance of 100,000 ohms, and the plate power supply is constant at 200 volts, then the plate current is $\frac{200}{200,000}$, or 0.001 ampere. The plate voltage is 0.001x100,000, or 100 volts.

Assume that the voltmeter used to measure the plate voltage of the tube has a sensitivity of 1,000 ohms per volt and that the range is from 0 to 250 volts. The meter will than have a resistance of 250,000 ohms. This resistance in parallel with the tube resistance of 100,000 ohms produces an effective resistance of 71,400 ohms in series with the plate load resistor. The total resistance across the B supply is therefore 171,400 ohms and the current through the plate load resistor is $\frac{200}{171,400}$, or 0.00117 ampere. Across the plate load resistor the voltage drop is 0.00117x100,000, or 117 volts and the plate-to-ground voltage on the tube is 200-117, or 83 volts when the meter is connected, thus causing an error of 17 percent. The lower the sensitivity of the meter the greater this error will be.

A meter having a sensitivity of 20,000 ohms per volt and a 250-volt maximum scale reading would introduce an error of about 1 percent. However, in circuits where very high impedances are encountered, such as in grid circuits of electron tubes, even a meter of this sensitivity would impose too much of a load on the circuit.

ELECTRON-TUBE VOLTMETER

Another limitation of the D'Arsonval a-c rectifier type voltmeter is the shunting effect at high frequencies of the relatively large meter rectifier capacitance. This shunting effect may be eliminated by replacing the usual metallic oxide rectifier with an electron-tube amplifier in which the plate circuit contains the d-c meter, and the voltage to be measured is applied to the grid circuit. Such a device is called an ELECTRON-TUBE VOLTMETER. Voltages at frequencies up to and greater than 100 megacycles can be measured accurately with this type of meter.

THE INPUT IMPEDANCE IS LARGE, and therefore the current drawn from the circuit whose voltage is being measured is small and in most cases, negligible.

Simplified diagrams of the a-c and d-c electron-tube voltmeter sections of Navy model AN/USM-34 volt-ohm-ammeter are shown in figure 17-30.

The operation of d-c amplifiers of the type used in electron tube voltmeters is discussed in chapter 6 and may be reviewed for a better understanding of the operation of electron-tube voltmeters.

The a-c voltage to be measured is applied to the a-c probe (fig. 17-30,A). It is rectified by V1 and filtered by the R-C network in the probe.

The meter circuit is a balanced bridge network. When the input voltage betwen the probe and ground is zero, the bridge is balanced and the voltages across the two arms containing the plate load resistors of V2 are equal. Thus, the d-c meter indicates zero. If a voltage is applied between the probe and ground, the bridge becomes unbalanced and current flows through the meter. The meter is calibrated in rms volts. The input impedance is very high. At the lower frequencies the input capacitance is negligible, but as the frequency increases the input capacitance introduces an additional load on the circuit under test and causes an error in the meter reading.

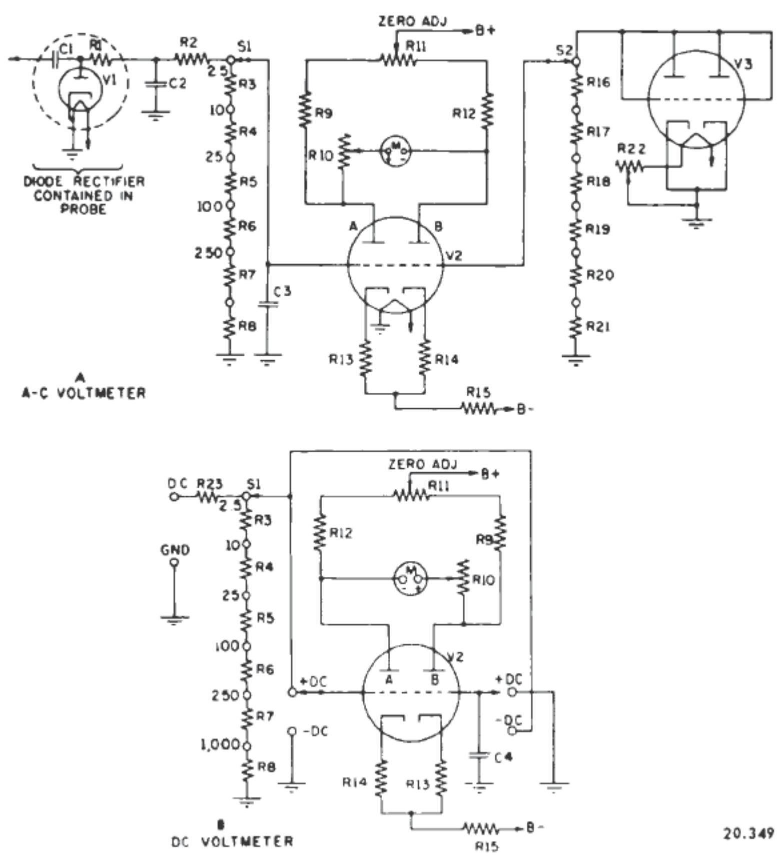


Figure 17-30.—Electron-tube voltmeter.

The d-c electron-tube voltmeter circuit is shown in figure 17-30,B. The d-c voltage to be measured is applied between the d-c input terminal and ground. The d-c input voltage is therefore applied through R23 to the divider network feeding the grid of V2A. The grid of V2B is grounded. The meter is connected across a normally balanced bridge so that the application of the d-c voltage unbalances the bridge and causes the meter to deflect. The calibration is in d-c volts. Bias is obtained for V2A and B through the voltage drop across R13, R14, and R15. The cathodes are positive with respect to B- by an amount equal to the bias. Thus the grids are correspondingly negative with respect to the cathodes.

In figure 17-30,A, diode V1 causes a contact potential to be established across the voltage divider network connected to the grid of V2A. This voltage would unbalance the bridge. Therefore a similar contact potential is introduced across the grid of V2B from V3 and its associated voltage divider to balance the bridge before the a-c voltage to be measured is applied to the diode probe.

In figure 17-30,B, no diode probe is used, hence no contact potential is established so that V3 and its associated voltage divider network are omitted from the circuit.

CHAPTER 18

INTRODUCTION TO COMPUTERS (PART I)

INTRODUCTION

As modern computer technology advances, the Navy is finding it possible to apply computers to a greater variety of tasks. Problems and situations which were once considered complex and time consuming are now quickly and accurately solved and/or evaluated by computers.

The Naval Tactical Data System (NTDS), composed of transistorized computers, pictorial displays, and digital communications equipment, is now being installed in several combatant ships.

NTDS greatly increases the ability of task force personnel to track and evaluate high speed targets by automatically computing and processing battle intelligence, solving and displaying combat problems, and communicating information and orders between units of a task force.

GENERAL

Today, a good mathematician, or for that matter, a good physicist or engineer, makes few routine longhand mathematical calculations to solve his problems. Instead, he uses slide rules, small office calculators, and sometimes giant electronic computers (fig. 18-1), all of which are broadly classified as computers. At first glance, then, it might seem that a computer is merely a "machine for doing mathematics."

This is only partly true. A computer certainly is a "mathematics machine," but it is also something more. It is a record-keeping device; it is a sorting device; and it is even being used experimentally as a language translating device.

DEFINITION

A practical method of deriving a definition of the word computer would be to identify common characteristics of two computers whose physical characteristics differ from each other as greatly as possible.

Figure 18-2, for example, shows two computers that satisfy the requirement of being different from each other. One is the simple adding machine, and the other is the automatic language translator. The adding machine handles numbers; the automatic translator handles words. What is common to both of these? Data, of course. Both machines take data, or information, at the input, process it, and then feed it out in a more useful form. Both computers can be subdivided into three basic sections: input, processing, and output.

From the function of each basic section, the following definition has been derived: A computer is any device capable of accepting information, applying prescribed processes to this information, and supplying the results of these processes. This definition fits both adding machine and language translator, as well as all other computers.

The following information will briefly describe several applications of a computer.

COMPUTER APPLICATIONS

A computer, like any other machine, is used because it does certain jobs better and more efficiently than humans. Specifically, it receives more information and processes it much faster than man can. Within any specified period, a giant computer can do more arithmetic than all of the people in the United States can do with pencil and paper in the same period.

Many people have the notion that computers are used to solve problems that cannot be solved by man. This is not so. A computer can do only what its designer builds into it. If its designer cannot solve a given problem, then the computer will not be capable of performing the logical operations necessary for its solution. If man can only approximate an answer to a problem then this is only what a computer will do, but

20.350
Figure 18-1.—The IBM 705 electronic data processing system.

the computer will find the approximate answer to the problem much faster. Speed becomes important when weeks or months of pencil and paper work can be replaced by a computer solution requiring only a matter of minutes.

Computers are used to solve problems which cannot be solved economically by man. Even when the problem has no exact solution, a computer is used when the time saved offsets the cost of using the machine. Because of these advantages—high speed and high capacity—the Navy is finding many uses for computers. The following examples point up a few uses to which computers are put.

1. Solving design problems. The mathematical abilities of the computer are very useful to design engineers. The wing design of a supersonic aircraft, for example, depends on many factors. The designer describes each of these factors mathematically in the form of an equation and feeds it in the proper form to the computer. The computer then processes this information and supplies a useful answer.

2. Solving problems in personnel, fiscal, and stock control management. Personnel, supply, and accounting departments find computers extremely useful. Sometimes, both speed and high capacity combine to make the computer a super clerk. Inventories are reported and managed through the aid of computers; invoices are checked for accuracy and automatic replenishment is ensured. Supply and demand can be balanced and control points can order replenishment from areas of over supply to areas where there is a shortage.

3. Processing information from missile ranges and tracking stations. At a missile firing range, for example, fifty or more streams of information may be coming in simultaneously.

Each of these streams may carry hundreds, thousands, or even millions of information units every second. To classify and route this data

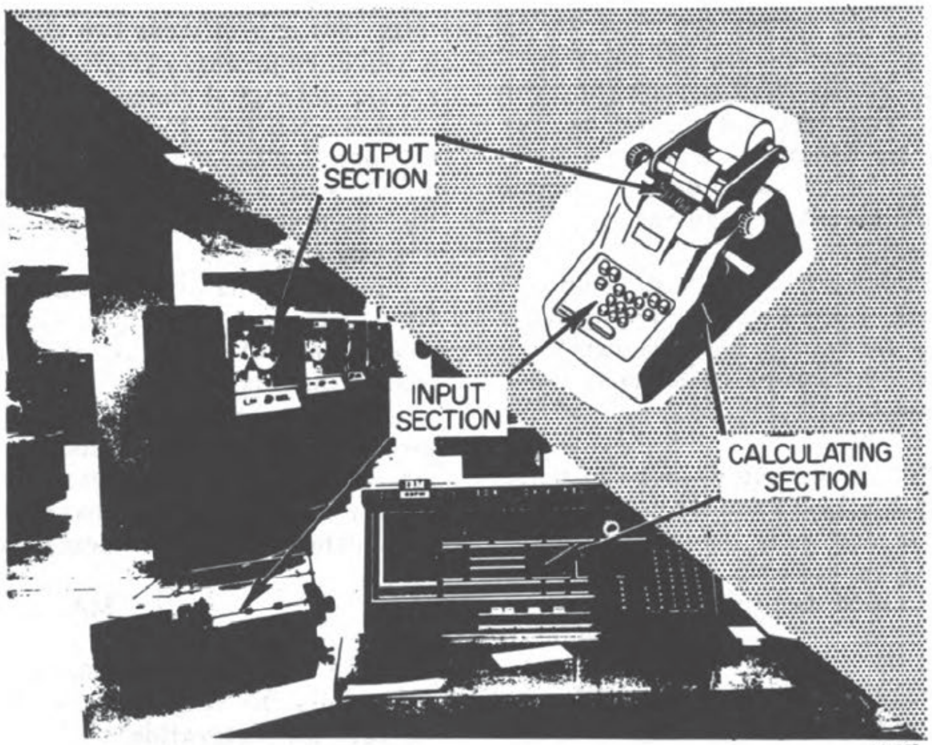


Figure 18-2.—Computer comparison.

to the proper inputs of a computer is in itself a job for another computer. Unless such a computer clerk were available, missile launchings would not give us half the information they do.

4. Forming a part of military systems. Computers form a part of many military systems, including communications, fire control, tactics, strategy, and logistics.

In some systems computers retrieve and evaluate signals too well hidden in noise to be detected by a human operator. They are extremely useful in telemetering and other systems whose signals, once missed, are lost forever. In some systems, the problem is how to hit a target; in others, the problem is what to do and when to do it; and in others, the problem is what to send and where to send it. More specific computer applications include automatic piloting, automatic navigation, and automatic landing.

All of these applications contain factors which require the computer to furnish accurate solutions or evaluations quickly, either for a fixed condition or for changing situations.

5. Automating processes. The computer is a useful tool for manufacturing and inspecting products automatically. A computer can run a turret lathe, a milling machine, and many other machine tools more rapidly and accurately than a man. It inspects the part as it is being made, and adjusts the machine tool as needed. If an incoming piece is defective the computer rejects it and starts the machine on the next piece.

6. Training by simulation. It is often expensive, dangerous, and impracticable to train a large group of men under actual conditions to operate an aircraft or submarine, drop bombs on a target, shoot down enemy planes without being shot down, control a satellite, or operate a space ship. A computer can simulate all of these conditions for the trainee, react to his actions, and show him the results of his actions. Thus, he can get many days of on-the-job experience without personal risk and without risking the destruction of expensive equipment.

COMPUTER CAPABILITIES AND LIMITATIONS

The usefulness of a computer can be better understood when its basic capabilities and limitations are known. These are listed below.

Capabilities

REPETITIVE OPERATION.—A computer can perform similar mathematical operations over and over again for hundreds, thousands, millions, or trillions of times, without becoming tired or bored and therefore careless.

SPEED.—A computer works with enormous speeds, which are limited only by the ingenuity of the designer. Modern computers can solve a problem many millions of times faster than a skilled mathematician.

FLEXIBILITY.—A computer can be made as flexible or as specialized as desired. It can be a special purpose computer designed to solve only one type of problem, or it can be a general purpose computer designed to solve many types of problems.

ACCURACY.—In most applications, the right computer can provide an answer that is as accurate as desired. Furthermore, the computer can be designed to check its own answers.

Limitations

THINKING.—It is not yet known just exactly what "thinking" is. According to the commonly understood meaning of the word, a computer is a machine and cannot "think." It automatically follows orders and cannot consider any new information that is not fed to it, no matter how important. Thus, an automatic pilot will fly a damaged plane with its wounded crew directly into enemy territory, because that is what it was ordered to do. A man, however, can think for the computer and feed new data (aircraft damage) into it so that it may adjust and act accordingly.

INTUITION.—A computer has no intuition. It does its job exactly as directed. On the basis of so-called intuition, a man may suddenly find the answer to a problem without working out the details. A computer cannot do this. It can only proceed as ordered.

BASIC COMPUTER TYPES

Computers can be classified into two basic categories: analog and digital. Every computer falls into one of these two classes, unless it is a hybrid computer. A hybrid computer has both analog and digital characteristics.

One good way to investigate the difference between analog and digital computers is to find a simple problem, and then feed it to both types. By watching how they get the answers, one can see how they work.

Consider this problem: If a missile launched from a cruiser has an acceleration of 2 units and a mass of 75 units, what is the force that pushes the missile?

Recalling some basic high school physics, Newton's second law of motion is used to get the answer. The law states that the force on a body is equal to the mass of the body times its acceleration. In mathematical form,

$$F = MA,$$

where F represents the force on the body, M represents the body's mass, and A represents the body's acceleration.

The problem is solved simply by substituting the known values in the equation and then multiplying as shown below.

$$F = MA$$

$$F = 75 \times 2 = 150 \text{ units of force}$$

Now by applying this problem to both computers, it is easily seen how each type approaches the problem.

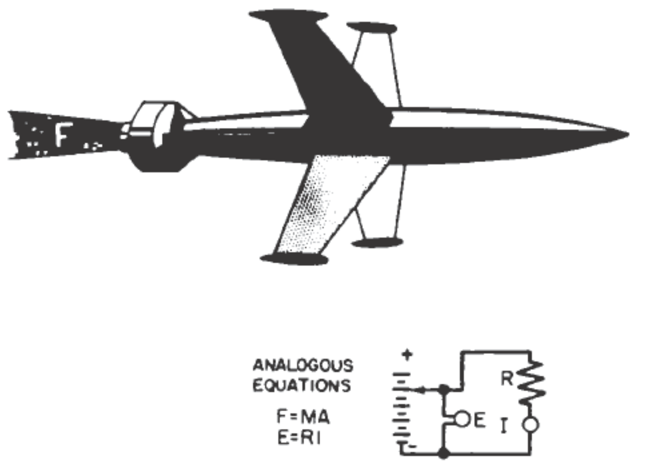
Analog Approach

A comparison of Newton's law and Ohm's law will show that the forms of the equations are identical.

Newton's law . . . . $F = MA$

Ohm's law . . . . $E = RI$

Accordingly, an electrical circuit described by Ohm's law is "analogous" to a moving body described by Newton's law. The advantage of this relationship between the two systems is that one can be used to give information about the other. By setting the given values in the circuit shown in figure 18-3, the voltmeter will automatically produce the value of force.



12.81

Figure 18-3.—The analog approach to measurements.

In the analogous electric circuit the voltage, E, is proportional to the force, F; the resistance, R, is proportional to the mass, M; and the current, I, is proportional to the acceleration, A. Thus, the product RI is proportional to the product MA.

Therefore, the electrical circuit can be used to compute the force on the accelerating missile. For this reason, the circuit when used in this manner is called an analog computer.

Within its physical limitations an analog computer can be used to multiply any two numbers instantly. Used in this way it is purely a mathematical machine.

Thus, an analog computer is really two devices in one: It is a simulating machine, for in a mathematical sense it is a model or analog of some other system. In the preceding example, an electrical system was used to simulate a mechanical system. It is also a mathematical machine in that any two quantities could be multiplied by setting the current and resistance at the desired values, and then reading the product of the two on the voltmeter.

Digital Approach

A digital computer can also solve this problem, but quite differently. Instead of looking for some analogous physical system, the equation F = MA is simply looked upon as a multiplication problem and the actual numbers are used immediately. In digital form, the problem is

$$75 \times 2 = ?$$

To solve it an office adding machine, which is a digital computer, can be used. The digits 75 are punched twice and the answer is read in the form of three digits, 150.

Thus, a digital computer is simply an arithmetic machine. It receives individual digits, performs simple arithmetic on them, and produces an answer in the form of individual digits.

Specific Differences

INPUT.—The input to the analog computer is never an absolute number; it is an approximate position on a continuous scale (for example, the position of a potentiometer arm) as shown in figure 18-4,A. There will always be some error present no matter how small. On the other hand, the number fed into the digital computer is precisely the one desired.

PROCESSING.—The analog computer works with continuous values, while the digital computer works with discrete individual digits (fig. 18-4,B).

OUTPUT.—The output of an analog computer, once more, is never an absolute number; it is an approximate position on a continuous scale. The digital computer, however, produces a specific series of digits for an output (fig. 18-4,C).

General Differences

ACCURACY.—The analog computer used in the previous problem is far from accurate. The problems of potentiometer and voltmeter accuracy, as well as the problem of loading, prevent it from being a practical instrument. It was used only for example's sake. By improving the design, the accuracy of the analog computer can be increased, but even the most expensive analog computer cannot be perfectly accurate because its components include potentiometers, galvanometers, transistors, and electron tubes. Although these components can be built with high accuracy, they can never be fully accurate; thus the computer's accuracy

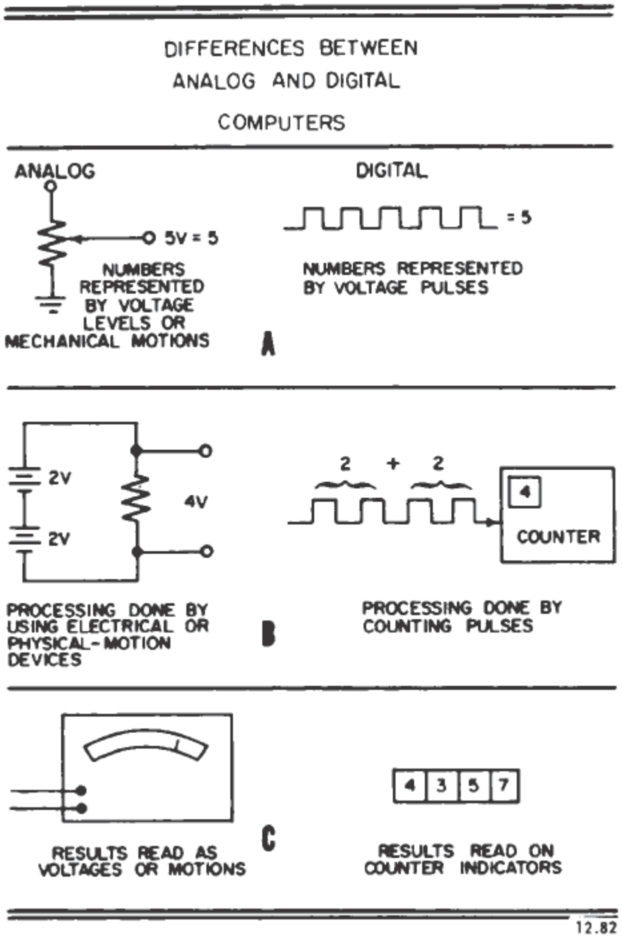


Figure 18-4.—Basic differences between analog and digital computers.

cannot be higher than its least accurate component.

The output of a digital computer, on the other hand, can be made as accurate as the designer wishes. For example, given enough time, a digital computer can produce the value of π to as many places as desired.

SPEED.—Most analog computers will produce an answer as soon as a problem is fed into it. This ability is vital in military operations such as fire control, where there is not much time to compute the position of a flying target.

On the other hand, a digital computer does not work instantaneously. It produces an answer some time after the problem is fed to it. But don't conclude that this makes it a slow machine. The "some time after" may be a few millionths of a second. However, because the digital computer must do arithmetic, there is always a time lag between problem and solution.

PRESENTATION.—The two types of computers present their answers differently. Once the analog computer was set up in the previous problem it was possible to do much more than get the product of 75 and 2. By varying the voltage, the answers to a number of problems can be attained: 75 x 2, 75 x 2.1, 75 x 2.2, 75 x 2.3—in fact, 75 times every number on the ammeter dial. Furthermore, by replacing the voltmeter with a recording voltmeter as shown in figure 18-5, a graph can be plotted showing the continuous relationship between the products and one of the factors. Therefore, the missile force for a continuous spread of acceleration values is indicated.

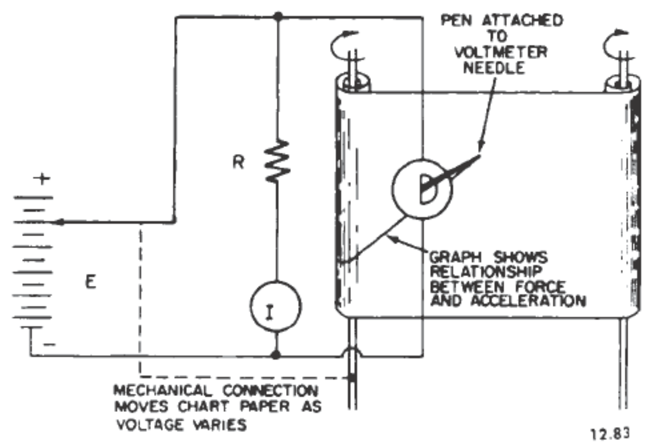


Figure 18-5.—The recording voltmeter as a method of continuous calculation-recording.

It is much more difficult to get all of these answers with a digital computer. It can be done, however, but it would require a separate multiplication for each answer. It would be necessary to multiply 75 x 2 and record the answer; and so on. Of course, large digital computers can do this automatically, but even then it takes more time. Also, the digital computer cannot possibly give the answer to every single input value. There would always be a gap between any two answers, no matter how close. Thus, the digital computer can never provide a continuous relationship between two quantities.

SUMMARY

Basically, the analog computer deals with a continuous system, while the digital computer works with discrete numbers. While the analog computer is faster than the digital, it is much less accurate. Also, the analog computer can easily present the overall picture of an entire system, but the digital computer can give only the answer to one specific arithmetic problem at a time.

The purpose of this entire section has been to give a general discussion of the differences between the analog and digital computers. In later sections each computer type will be explained in more detail.

ANALOG COMPUTER

As soon as man discovered the relation between two or more physical quantities, like mass and acceleration, he had a potential analog computer. When he used one of these factors to find out more about the other, the potential became the actual.

CHARACTERISTICS OF ANALOG COMPUTERS

Basic Principles of Analog Computation

Although man has been using analog computers since the dawn of civilization, he has not fully understood the basic principle involved until quite recently. This principle may be stated as follows: "Different physical systems can often be described by the same mathematical equations."

A physical system is defined as any system that exists in nature (fig. 18-6). It may consist of a nut and bolt, an entire building, an electron tube, an entire radar installation, a turning propeller, an entire ship, the solar system, or an analog computer.

When speaking of a specific physical system, it is important to describe that system exactly. This is where mathematics comes in. Its precise nature makes it ideal for describing physical systems. This mathematical description, which is called the mathematical model,

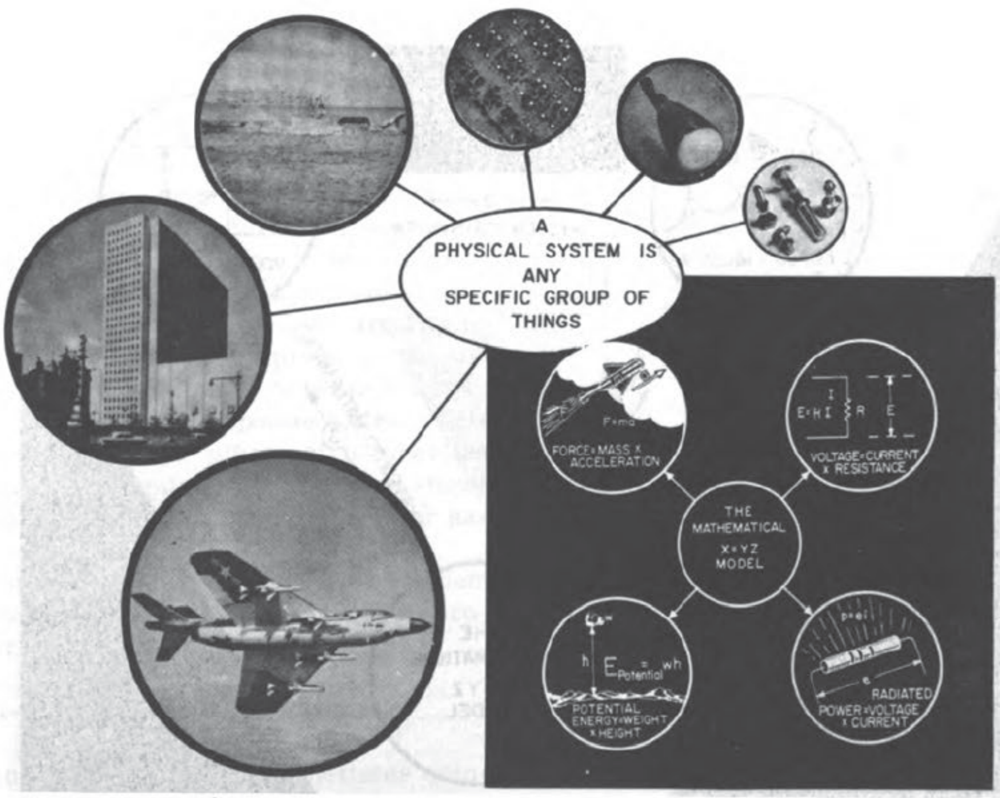


Figure 18-6.—Analogies between physical and mathematical systems. 20.352

often applies to more than one physical system. Systems which are described by the same mathematical model are called analogous systems.

Analog Triangle

If a schematic diagram of the connection between a physical system, its mathematical model, and some analogous system is drawn, a triangle will be formed as shown in figure 18-7. Each side of this triangle represents a basic application of the analog principle.

SIDE 1—MATHEMATICS-TO-SYSTEM.—This application—the use of mathematics to help in understanding a physical system—is simply the standard pencil-and-paper method of using mathematics. For example, the equation F = MA is used to describe the force acting on an accelerating body.

SIDE 2 — SYSTEM - TO - SYSTEM. —Because the voltage developed across a resistor (E = RI) and the force acting on a moving body (F = MA) are described by the same type of equation, the resistor and the moving body are analogous systems. For this reason, the operation of one system can be simulated by using the other. As shown earlier, if the magnitude of the voltage developed across the resistor is known, the amount of force acting on the moving body can be determined easily. This technique is called simulation.

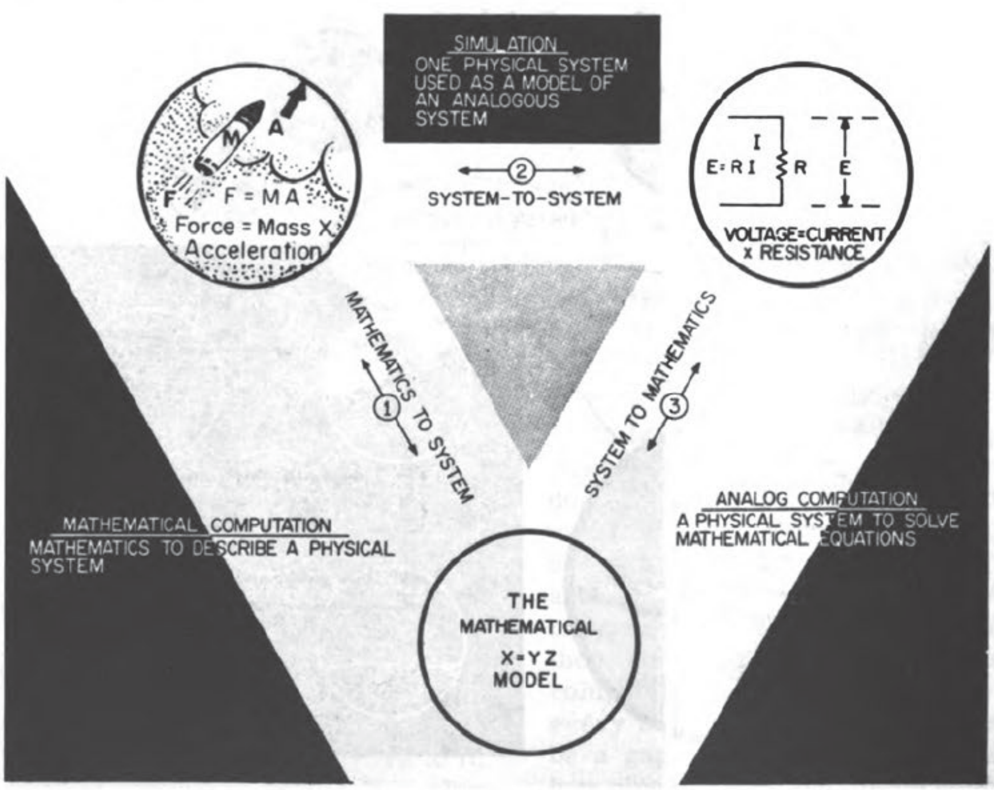
SIDE 3 — SYSTEM - TO - MATHEMATICS.—Just as one can learn about a physical system through the mathematics that describes it, one can also solve mathematical equations through the physical system described by them. Examples of this principle are given here.

Analog Arithmetic Elements

The equation;

$$A + B = x,$$

describes the operation of a scale for measuring weight in pounds. A and B are the weights placed on the scale, while x represents the scale's marker, or readout, as shown in figure 18-8.



20.353

Figure 18-7.—The analog triangle.

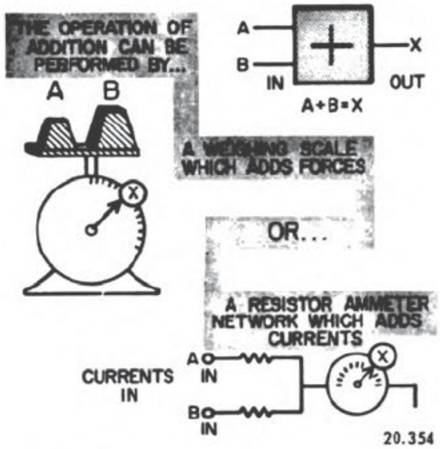


Figure 18-8.—Analog additions—direct read-out methods.

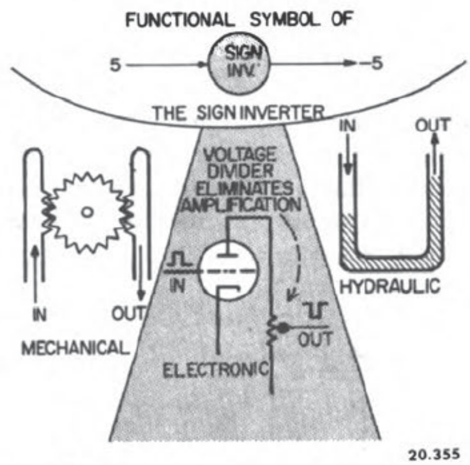


Figure 18-9.—Sign inversions—the basis for all analogs.

The same equation can also be used to represent the operation of two electric currents that enter the same ammeter. A and B represent the two input currents, while x represents the reading on the ammeter.

Both of these devices may be called natural adding elements. For the sake of convenience, adding elements are represented schematically by drawing a box with a number of inputs and one output, and placing a "+" symbol in it.

By applying a well known rule of algebra, an adding element can be used to subtract. That is, to subtract one number from another, change the sign of the subtrahend, and add. With a sign-inverting device, an adding element can be used for subtraction. Many such sign-inverting devices are available. An amplifying triode circuit, for example, produces an inverted voltage at its output. A lever's output end moves in a direction opposite to that of its input. These, and all other sign-inverting elements, are represented by a circle as shown in figure 18-9. Note that the sign inverter has only one input and one output.

As shown in figure 18-10, an adding element has been combined with a sign inverter to produce a subtracting element.

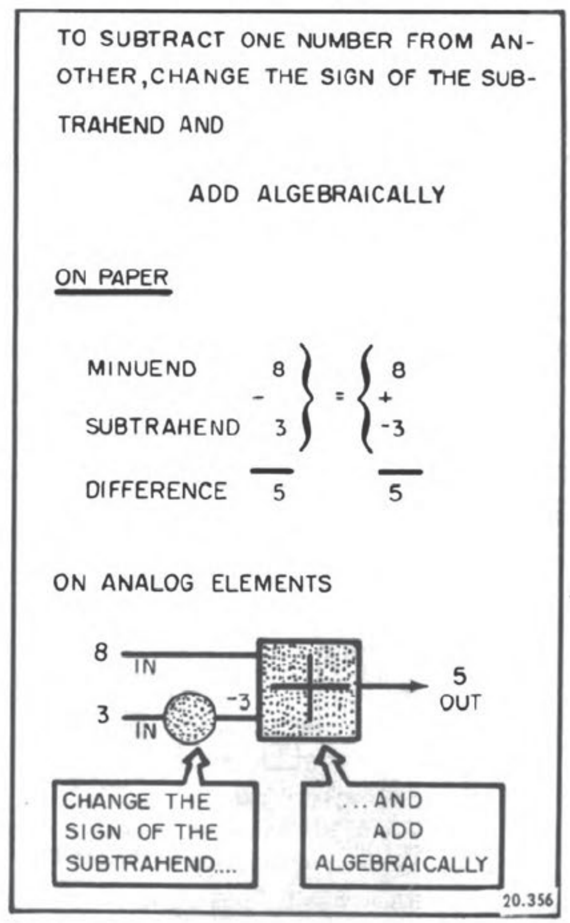


Figure 18-10.—Subtraction on paper and through analog elements.

The Special-Purpose Analog Computer

Basic analog elements are sometimes combined for a specialized application. Used in this way, the combination is called a special-purpose analog computer.

Assume, for example, that it is desired to obtain a continuous record of the total amount of fuel available in two fuel tanks. This can be done by placing some kind of liquid level-measuring device in each tank, and feeding the outputs of these two devices to an adding element, as shown in figure 18-11. The adding elements output can be fed to a strip-chart recorder, so that a continuous, permanent record of the total amount of fuel in gallons will be available at all times.

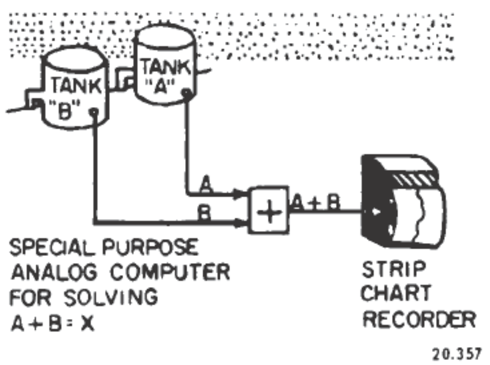


Figure 18-11.—Adding quantities and reading-out on analog components, such as this strip chart recorder, for permanent records.

Now to go one step further, assume that the fuel costs \$2.00 per gallon, and that it is desired to obtain a continuous record of the total dollar value of fuel available. In other words, we multiply the output by two. This can be done simply by adding a multiplying element to the adding element, as shown in figure 18-12.

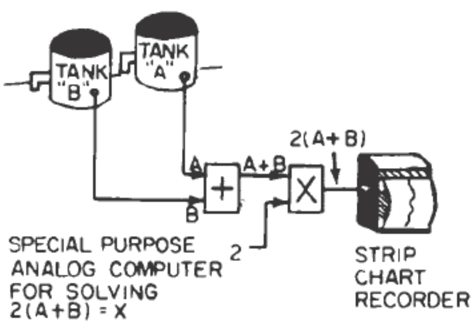


Figure 18-12.—Adding and multiplying using a special purpose analog computer.

Proceeding further, assume that the fuel is used to run 25 machines and that the dollar value of the fuel available for each machine is desired; that is, the output is to read "so many dollars worth of fuel available per machine."

This can be accomplished by dividing the total dollar value of the fuel by 25. The new special-purpose computer, with the dividing element installed, is shown in figure 18-13.

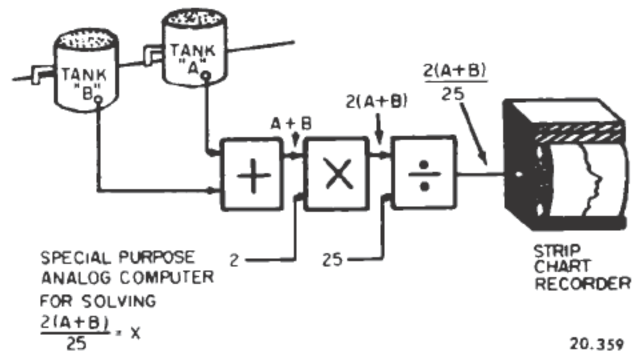


Figure 18-13.—Solves a problem in addition, multiplication and division in one step.

By adding more basic elements to this combination, mathematical problems of increasing complexity can be solved. Before going any further, however, this basic principle should be clear: A special-purpose analog computer is simply a combination of basic computing elements that have been selected to do one specific job.

PROBLEM: Design the special-purpose analog computers, in block diagram form, that will solve each of the following problems:

a. $p - q = x$

b. $s(p - q) = y$

c. $\dfrac{s}{p + q} = z$

d. $s\dfrac{p + q}{r + t} = c$

See figure 18-14 for solutions.

General-Purpose Analog Computer

The general-purpose analog computer simply employs an extension of the special-purpose analog computer principle. Because it usually contains more basic elements than may be required for accomplishing any one particular job, the general-purpose computer can solve a variety of problems. A highly simplified block

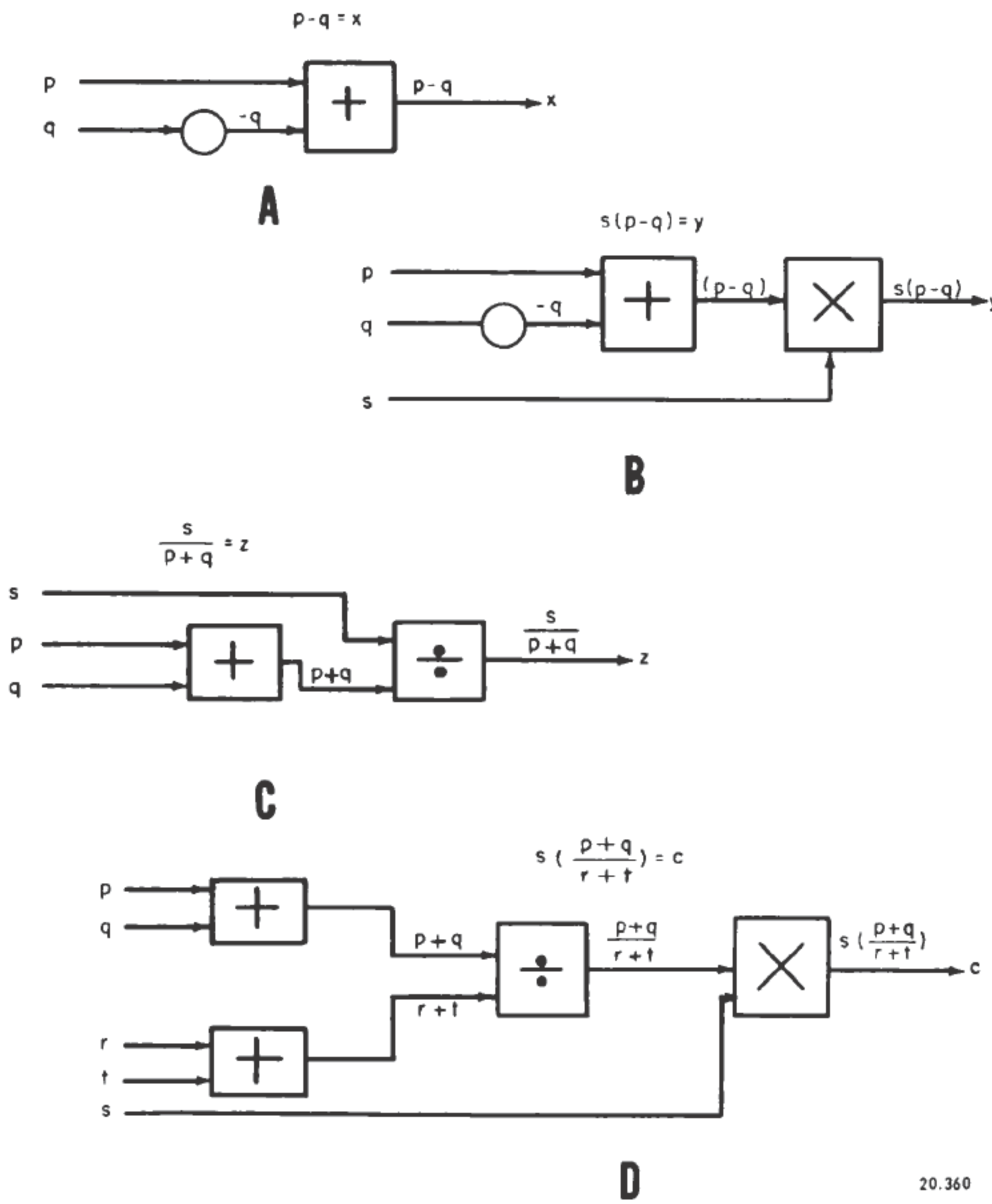


Figure 18-14.—Solutions to problems.

diagram of such a computer is shown in figure 18-15.

Note that all of the basic elements are available. Using them is simply a matter of connecting them in the desired way. To solve the problem,

$$\frac{(A + B)\,(C + D)}{(E - F)},$$

begin by setting up each of the expressions within the parentheses. Thus, feed inputs A and B to one adder, and C and D to another. Inputs E and F require a little more work. Since F must be subtracted from E, F is fed to a sign inverter, obtaining -F. Now E and -F can be fed to a third adder to obtain the difference between these two for an output.

The next step is to combine parenthetical expressions. Thus, to multiply (A+B) by (C+D), feed the outputs of the top two adders to a multiplier, obtaining an output of (A+B)(C+D).

Finally, divide the numerator by the denominator by feeding the outputs of the multiplier [(A+B)(C+D)] and the third adder (E-F)

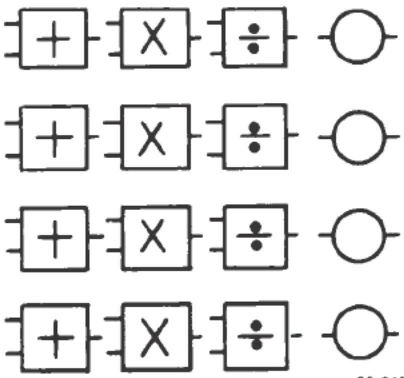


Figure 18-15.—The general-purpose analog computer.

to a divider, making sure that (E-F) enters the divisor input.

Now the output of the dividing element can be called x for it is the solution to our problem. The complete connection is shown in figure 18-16.

PROBLEMS: Connect the general-purpose computer in figure 18-15 to solve each one of the following problems.

a. $\frac{(A + B) - (C + D)}{(E + F)} = X$

b. $-\frac{(A + B)(C + D)(E + F)}{(G + H)} = Y$

See figure 18-17 for the solutions to these problems.

As was illustrated in this section, the design of special purpose and general purpose analog computers in block form is rather simple. In the next section, those devices used by the analog computer to perform the various mathematical operations will be discussed.

COMPUTING DEVICES

The purpose of this section is to introduce to the reader a few basic electrical, electronic, and electromechanical devices used in analog computers to perform the following mathematical operations: addition, subtraction, multiplication, and division. It should be remembered that the methods presented in this chapter are basic and many variations of these methods will be found.

Those devices which are mechanical in nature will not be discussed in this text. For information concerning these devices refer to *"Basic Fire Control Mechanisms,"* ORDNANCE PAMPHLET 1140.

Summation: Addition and Subtraction

The algebraic summation of two quantities can best be illustrated with a simple circuit such as that shown in figure 18-18, A. Assuming that T_1 and T_2 are ideal transformers, the output voltage E_o will depend directly upon the magnitude and polarity of input voltages E_1 and E_2. The operation of the circuit is summarized in the table shown in figure 18-18,B.

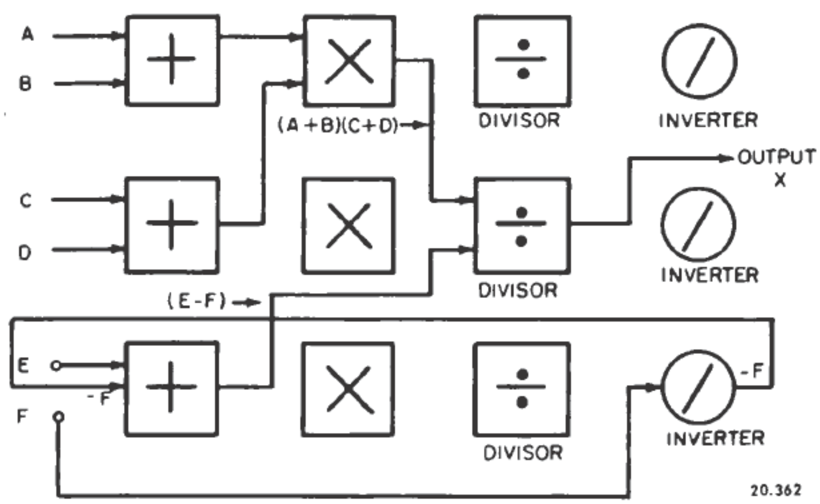


Figure 18-16.—Connections for solving addition, subtraction, multiplication and division with the general-purpose analog computer.

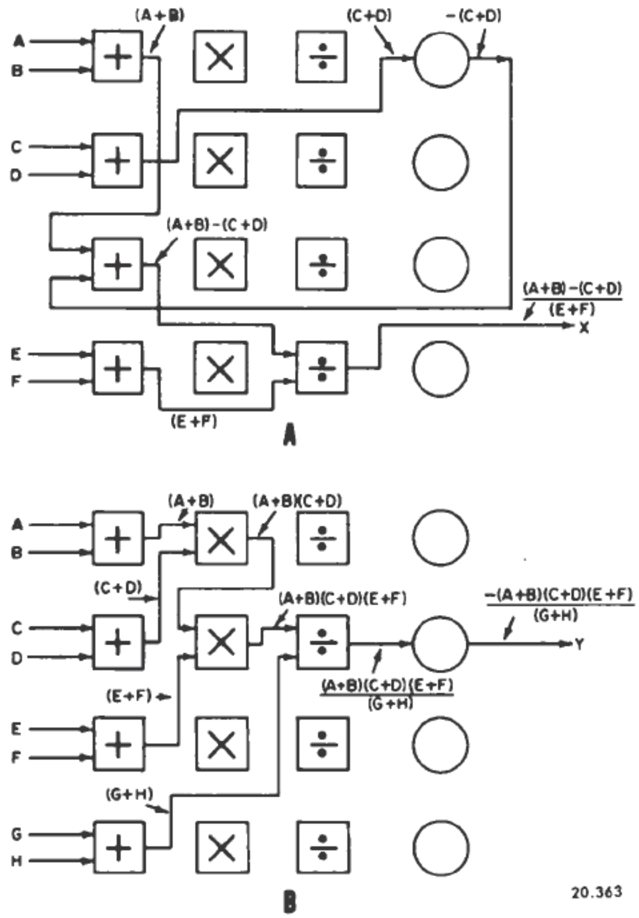


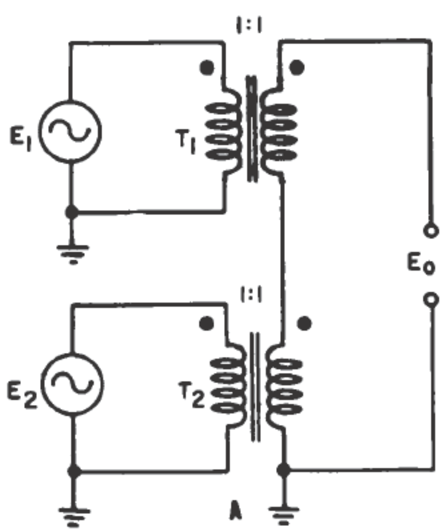
Figure 18-17.—Solution to problems.

Addition and subtraction of d-c voltages can be accomplished with a circuit such as that shown in figure 18-19, A. The output voltage, E_O, depends upon the polarity of the d-c sources, and the position of each potentiometer arm. See the table in figure 18-19,B, for a summary of circuit operation.

Although an analog computer solves a problem instantaneously, there is always some error involved. With proper design, however, this error can be reduced to a very small value and neglected. As evidence of this fact, examine the parallel summation network in figure 18-20,A. Neglecting R_O, the current i in each branch is approximately equal to the input voltage, e, divided by the resistance, R, of that branch.

Because the total current through R_O is the sum of the branch currents, and they in turn are proportional to the branch voltages, the output voltage across R_O will be proportional to the sum of the input voltages, as indicated in the table in figure 18-20,B. The multiplication factor, K, required to raise the output voltage to the sum of the input voltage, is thus seen to be 10^6. This action may be obtained, for example, by feeding it to a 3-stage cascade connected amplifier whose overall voltage gain is $-10^6/1$ (fig. 18-21) and whose stage gain is $-100/1$.

The table shown in figure 18-20,B, lists the results. Because e_O is a fraction of $e_1 + e2$



A

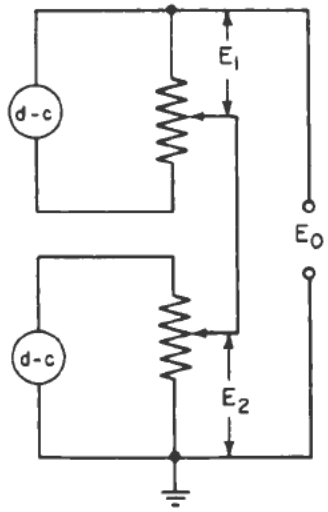
$E_O = E_1 + E_2$

$E_1 + E_2 = E_O$		
IN PHASE		
E_1	E_2	E_O
+10	+5	+15
-10	-5	-15
180° OUT OF PHASE		
E_1	E_2	E_O
+10	-5	+5
-10	+5	-5

B

20.364

Figure 18-18.—A. Summation circuit.
B. Summary of circuit operation.



A

$E_1 + E_2 = E_0$		
E_1	E_2	E_0
+20	+15	+35
-20	-15	-35
+20	-15	+5
-20	+15	-5

B

20.365

Figure 18-19.—A. D-c summation circuit.
B. Summary of circuit operation.

$+ e_3$, it must be multiplied by some factor such that

$$e_0 \text{ (mult. factor or K)} = e_1 + e_2 + e_3.$$

The amplifier output E_0 is given by

$$E_0 = -Ae_0$$

where A is the gain of the amplifier.

The symbol labeled -A (fig. 18-21) represents an electron tube amplifier, which may contain several stages of amplification. The letter A may be either positive or negative, depending on whether or not the amplifier produces a phase reversal from input to output.

Multiplication

The process of multiplication can be easily accomplished by using devices such as potentiometers, synchros, autotransformers, or amplifiers. For example, if one desired to multiply a constant by any number ranging from 0 to ± 1, the circuit (assuming no losses and negligible error) shown in figure 18-22 would suffice. In this case, the constant quantity is +10 volts. It must be remembered, however, that the voltages involved are analogs of certain mathematical quantities, and that one volt could represent any number of units. Assume in figure 18-22 that 1 volt represents 1 unit.

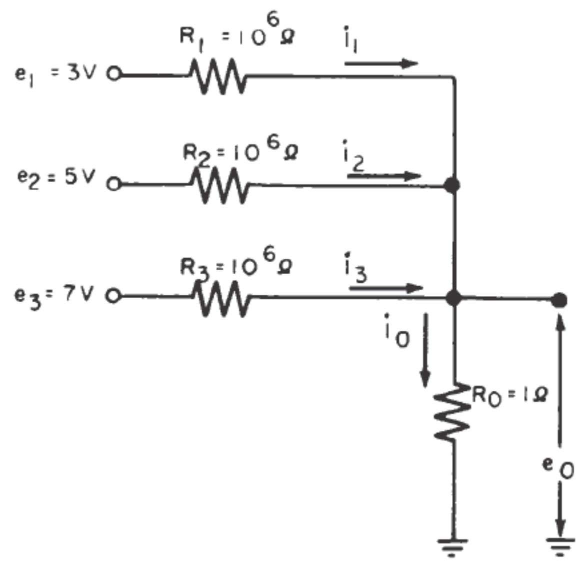
In order to perform the operation indicated by $e_0 = (10)(+0.35)$ simply set the potentiometer dial on +0.35, and read e_0 with a voltmeter having a high input impedance. Assuming no losses, the voltmeter should read +3.5 volts. Notice that the positive voltage indicates a positive product. If the potentiometer dial were set to -0.35, e_0 would be equal to -3.5v. The negative sign in this case indicates a negative product. Thus, +10 volts can be multiplied by any value between 0 and ± 1.

The obvious disadvantage of the potentiometer type multiplier is the fact that a quantity cannot be multiplied by a factor greater than ± 1. However, amplifiers can be used in conjunction with a potentiometer to increase the range of multiplication. To illustrate this point, examine the multiplier shown in figure 18-23. By adding an amplifier having a gain of 10, the range of multiplication is increased from 0 to 1 to 0 to 10. To multiply E by +6.5, set the potentiometer dial to +0.65, and read the voltage from terminal 2 to ground. By adding more amplifiers, the range of multiplication can be increased.

In figure 18-24, two potentiometers are used to solve the general equation $e_0 = E \cdot X \cdot Y$ where X and Y are both variables.

A variety of multiplying circuits have been devised, employing electron tube amplifiers. However, every linear electron tube amplifier

A



$$i_1 = \frac{e_1}{R_1 + R_O} \approx \frac{e_1}{R_1} = \frac{3}{1 \times 10^6} = 3 \times 10^{-6} a$$

$$i_2 = 5 \times 10^{-6} a$$

$$i_3 = 7 \times 10^{-6} a$$

$$i_O = (3 + 5 + 7) 10^{-6} a$$

$$e_O = i_O R_O = (3 + 5 + 7)\ 10^{-6} \times 1 = 15 \times 10^{-6} V$$

MULTIPLICATION FACTOR, $K = 10^6$, WHERE $Ke_O = e_1 + e_2 + e_3$

B

R_O	R_1	R_2	R_3	e_1	e_2	e_3
$1\ \Omega$	1 meg	1 meg	1 meg	3V	5V	7V

i_1	i_2	i_3	i_O	e_O	K	Ke_O
$3 \times 10^{-6} a$	$5 \times 10^{-6} a$	$7 \times 10^{-6} a$	$15 \times 10^{-6} a$	$15 \times 10^{-6} v$	10^6	15V

20.366

Figure 18-20.—A. Parallel summation network.
B. Table of values.

can be used as a multiplier by itself. Normally, the output voltage e_O is given by

$$e_o = \mu e_g \frac{R_L}{r_p + R},$$

as discussed in chapter 7 of this training course. Where R_L is very large compared with r_p, the simpler formula $e_O = \mu e_g$ is a good approximation of the output voltage. By inspection, it can be seen that

$$e_O = \mu e_g$$

is analogous to

$$y = KX.$$

In order to compensate for slight variations in the amplification factor, μ, analog computers frequently use amplifiers employing degenerative feedback.

Division

The familiar potentiometer is also widely used as a voltage divider. When division by a

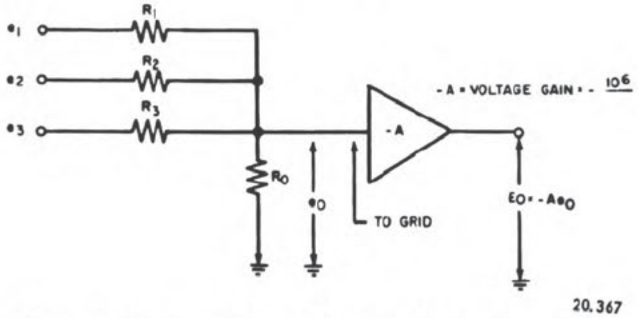


Figure 18-21.—Operational schematic of an adding network with amplifiers followup.

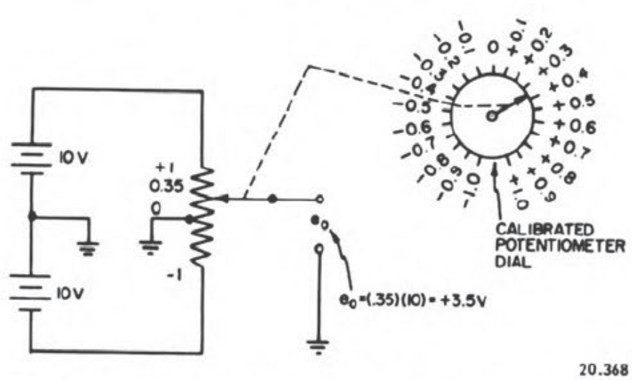


Figure 18-22.—Multiplying a constant by a variable.

constant is required, the calibrated potentiometer dial is rotated to the desired setting, the voltage to be divided is applied across the entire potentiometer, and the quotient is taken from the arm to ground. When division by a variable is required, the potentiometer is servo driven.

As an example of division, refer to figure 18-22 and divide 10 by 2. Since the potentiometer dial is calibrated for multiplication, set the reciprocal of 2 or 0.5 on the dial and read the answer at e_o. The output voltage e_o should equal 5 volts. That is,

$$\frac{10}{2} = 10 \times \frac{1}{2} = 10 \times 0.5 = 5 \text{ volts.}$$

BASIC ORGANIZATION OF THE ANALOG COMPUTER

Preceding sections have discussed how a few devices can be utilized to perform mathematical operations. However, any one of these items used alone is of limited value. To organize an analog computer, a collection of these devices must be arranged in such a manner that will allow them to perform a desired series of mathematical operations as part of the process of solving a problem. For example, to solve a simple equation, it may be necessary to multiply, divide, add, and subtract. A machine that could solve this problem would consist of units capable of performing multiplication, division, addition, and subtraction.

When organizing a basic computer, one must determine the arrangement and the number of each device to be used. Also, for an arrangement of appreciable size, one should provide a system of checking for proper operation.

The following discussion applies to the organization of an electronic analog computer.

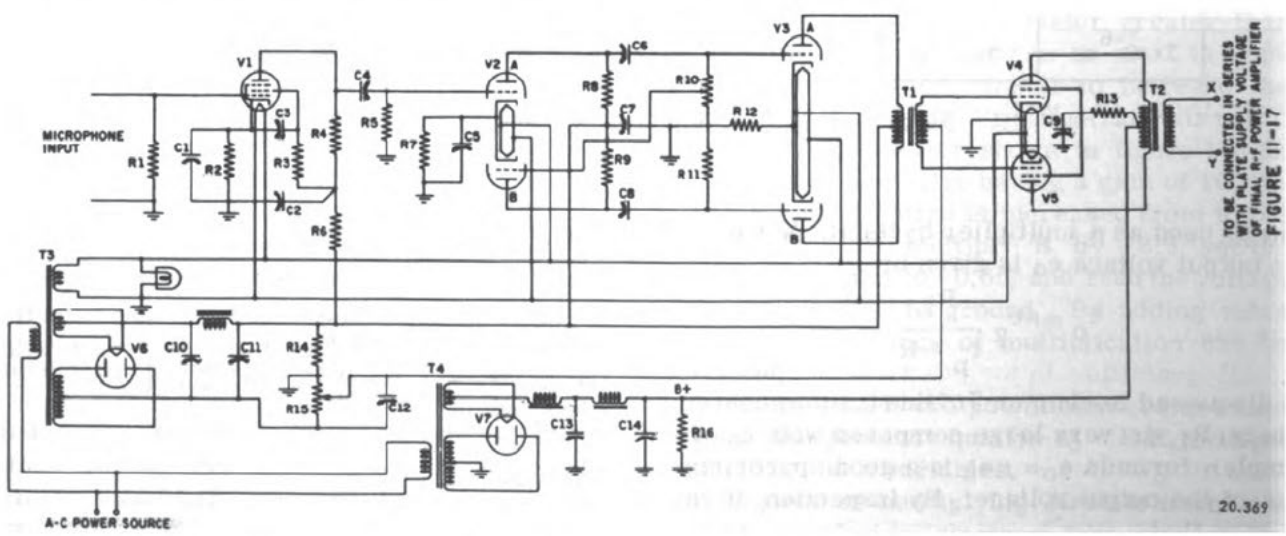


Figure 18-23.—Multiplier.

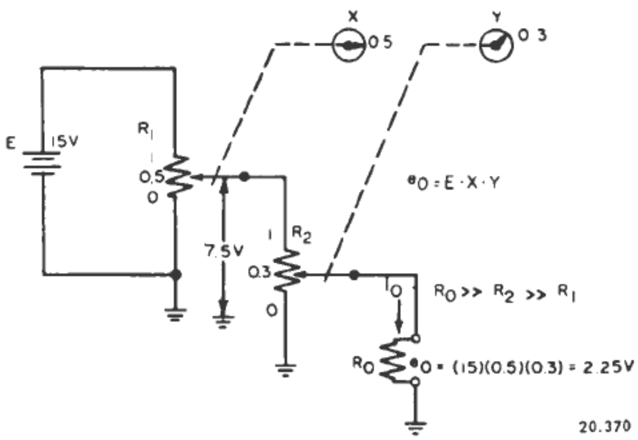


Figure 18-24.—Multiplying a constant by two variables.

The operational amplifier might be called the heart of the electronic analog computer. A proper choice of components and input, output, and feedback network connections will allow the amplifier to perform such mathematical operations as integration, sign inversion, addition, subtraction, or multiplication by a constant. However, in spite of their versatility, there are mathematical operations that these amplifiers cannot perform. To multiply two varying quantities together, or to generate mathematical functions of problem variables, servo multipliers and resolvers or the newer electronic multipliers and diode function generators may be used. Functions may also be generated by means of servo-driven function-card potentiometers.

The mathematical expression of a problem normally is an equation that contains constants and numerical coefficients in addition to the variables. Often, initial values of the variables are specified as well as the limits or boundaries within which the problem is to be solved. In order to introduce these values into the machine, potentiometers and regulated sources of voltage are used.

All quantities in the practical computer are represented by voltages. Since the operational amplifier has a restricted range of voltage over which it operates satisfactorily in its function as a mathematical tool, quantities involved in a problem can be represented by voltages that will remain within this range. This introduces the need for scaling. For this operation scaling potentiometers and calibrated voltage-divider networks are normally used.

The operational area of the analog computer will contain the operational amplifiers. The number of amplifiers will be determined by the complexity of the problem to be solved. A smaller number of servomechanisms, or perhaps of electronic multipliers and diode function generators will be needed. A relatively large number of potentiometers will be used.

Since it is desired to be able to change the manner in which these units are interconnected, their inouts and outputs must be readily available. This can be done by means of pinjacks or banana plugs near each operational element, or these points can be brought out to one central location and wired to a patch bay. For even greater ease and flexibility, a removable plugboard could be mounted into the patch bay. This board could be removed for "off-line" rearrangement of leads and components as desired. During this rearrangement, of course, the computer could be solving a problem under the control of another plugboard.

It may be necessary to introduce such voltage waveforms as square waves, sine waves, and sawtooth waves of varying amplitudes and durations during the solution of a particular problem. To do this, it is necessary to have signal generators capable of producing these signals. An oscilloscope might be of use as an aid in setting up problems and observing various portions of the problem being solved. In some cases the oscilloscope, with a camera, might be used to record part or all of the solution.

It is necessary to control the computer and to check it for proper operation. Switches and relays will be required to initiate the actions of the various components, and then to return them to their quiescent, or zero-value, state before the computer is run again. Some systems perhaps one employing neon lights, can be used to indicate amplifier overloads. Meters and calibrated voltages must be available to check the operation of the units.

Another factor to be considered is the manner in which the solution to the problem is to be made available to the operator. Frequently used devices are chart recorders, plotting boards, and oscilloscopes. The chart recorder usually plots the amplitude of a particular signal as it varies in time. If one is interested in the manner in which one unknown quantity varies with respect to another, an X-Y coordinate plotting board could be used. On either of these graphical recorders, facilities for making calibration marks associated with the variables involved in a specific problem will be needed. If only a general visual idea of the behavior of the problem is desired, observation of results on an oscilloscope might be entirely satisfactory. As stated previously, a permanent record could be made by photographing the face of the oscilloscope.

Since quantities in the computer are represented by the use of voltages proportional to their values, the reference voltages used in the equipment certainly should be of known constant value. Therefore, extreme care must be taken to maintain well regulated voltages and power supplies.

In brief, the electronic analog computer will consist of the following units: (1) an operational area bay, (2) a servo and function generation bay, (probably including the coefficient and initial-value potentiometers), (3) a patch bay and possibly plugboards, (4) a control and test bay, (5) a power bay, and (6) necessary input and output equipment.

If an analog computer is to be used for a particular application or for a special purpose, and its manner of operation will remain the same, it is not necessary to provide such complete flexibility of connection and control. However, the basic organization remains the same.

CHAPTER 19

INTRODUCTION TO COMPUTERS (PART II)

In chapter 18 we compared the ways in which two types of computers (analog and digital) solved a particular problem. You saw that electric analog computers use continuous voltages or currents to represent quantities or numbers, while digital computers, on the other hand, work with digits represented by discrete electrical pulses. Analog computers have a sliding scale of accuracy; digital computers count things to the nearest digit. The scale of a slide rule, for instance, can be read with greater accuracy if it is expanded in size. A digital computer's accuracy is limited only by the number of significant figures it can handle.

The advantages of analog computers include relative simplicity and economy of construction. Digital computers tend to be expensive and are rather hard to assemble and check out. But they are capable of far greater accuracy and flexibility than analog computers.

Achievement of speed by means of electronic calculations is a characteristic of electronic digital computers. The following simple illustrations show the progress that has been made in this area.

Example: Multiply 1234567890 by 5678901234. A thousand such calculations would take (1) by hand (pencil and paper), 1 week at 24 hours a day; (2) by mechanical desk machine, 1 day of 24 hours; (3) by small electronic calculator, 1 minute; and (4) by large electronic calculator, 1 second.

NUMBER SYSTEMS

A study of numbers is essential to a study of digital computers. Digital computers used by the Navy employ a very simple number system. In fact, it is so simple that it has only two digits: 1 and 0. You'll find that this system is a very powerful and highly reliable system. Digital computers are intended to be super-reliable and therefore are designed on the "YES or NO" theory. Something happens or it doesn't happen. A number is either 1 or 0, it cannot be a fraction of one or more than one, or negative. A function is THERE, or it is NOT THERE. A circuit is energized or not energized. A diode is conducting or not conducting. A voltage pulse is present or not present. How simple can you get? And with simplicity you get its running mate—reliability.

Numbers and numerals do not mean the same thing. Numerals are symbols that represent numbers. Numbers never change, but numerals may take many forms. Numeral forms and names change from country to country.

The needs of primitive peoples are simple and they probably have the simplest arithmetic in the world: "one, two, plenty" is as far as it goes. Most cultures, however, have a somewhat more sophisticated method of counting. This is the one-to-one system. A primitive tribesman, for example, might count time by throwing a stone into a pot for each passing day. If the vessel can hold only 30 stones, a full pot marks a month. Observe that one stone represents one day—a one-to-one relationship.

Sticks, marks on the earth, bones, and many other objects were used as tools for this one-to-one counting method. One tool, however, was much more convenient than any of the others. Discovered early in man's history, this counting tool is such an obviously practical one that you have used it yourself. You have it available all the time. This tool is, of course, your ten fingers. The Romans liked this counting tool so much that they based their entire number system on the use of the fingers. Even their symbols (numerals) are simply streamlined pictures of fingers in various positions. The symbol V represents an open palm of five fingers; X represents two hands crossed, a total of ten fingers. Other examples are shown in figure 19-1. Many Roman concepts about numbers have been kept to the present time. The word digit comes from the Latin word for finger—digitus. Calculus comes from the Latin word for pebble. Pebbles were used in business transactions to count.

Speaking of pebbles, primitive man had a counting device which used pebbles. He placed 10 pebbles on the perimeter of a clay wheel. Each stone represented one unit, and a complete revolution of the wheel represented 10 units.

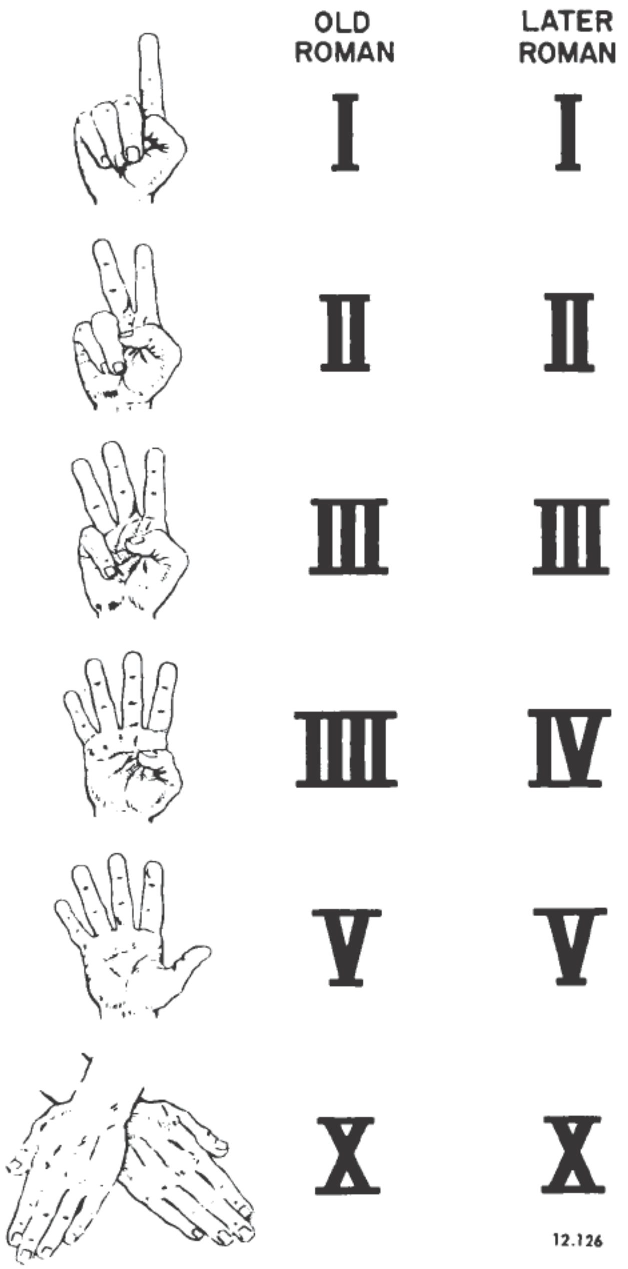


Figure 19-1.—Roman counting system using ten fingers.

By placing a single mark on the wheel, rotating it, and then counting the number of times the mark passes a fixed index, a simple sysyem of counting by tens was created. The number 77 in this counting system could be described as 7 marks and 7 pebbles. This wheel may have been the first decimal digital computer. The decimal system with its 10 numerals, 0 through 9, is convenient and natural for man. But it might not be convenient for 12-fingered Martians. Actually the Martians, if they have 12 fingers, could have a better number system than we have. And for this reason, functions of multiplication and division could be much more easily performed with a system based on 12, since this would be divisible by four numbers—2, 3, 4, and 6—instead of only 2 and 5 as in the decimal system.

The need for simplicity and reliability of representing numbers led the Chinese, 4000 years ago, to invent a 2-valued system of numerals for representing numbers. Reliability is best served by a 2-valued system, since the chance for error increases as the number of digits increases.

Reliability of ARITHMETIC operations may be greatly increased by decreasing the number of rules to be observed in any operation, as well as decreasing the number of symbols in the numbering system. For instance, in the decimal system there are 10 rules of addition to be memorized, whereas in the binary system there are only 3 rules to learn. In the decimal system there are 55 rules of multiplication to learn, whereas in the binary system there are only 3 rules. Subtraction and division can be disregarded, since it can be reasoned intuitively (that is, without positive proof) that a person can learn to do "binary arithmetic" with the aid of only 6 rules, whereas he must know 100-odd rules to do the same thing with decimal arithmetic. You can see that the reliability of a machine which has to perform only 6 operations is certain to be much greater than that of a machine which must perform 110 different operations.

Digital computers are made up of basic units such as diodes, relays, magnetic cores, or flip-flop circuits. All of these basic units or elements have one thing in common—they have only two possible states. They are either ON or OFF, conducting or not conducting, or energized or deenergized. All of these devices are used to count and to control operations. Figure 19-2 shows some of the binary elements used in digital computers. Another basic unit which we personally can use as a counter is the finger. It can do only one of two things: it can be raised or lowered. No other finger positions can be considered to have any meaning. If it is raised, we say that it has value; if it is lowered, it has no value. It is therefore a binary element. We are fortunate in that we ordinarily have ten of these binary units which provide us with a readymade binary digital computer. By using only the fingers of two hands we can count to any number between zero and 1023. We can do this by assigning values to the fingers of each hand. We call

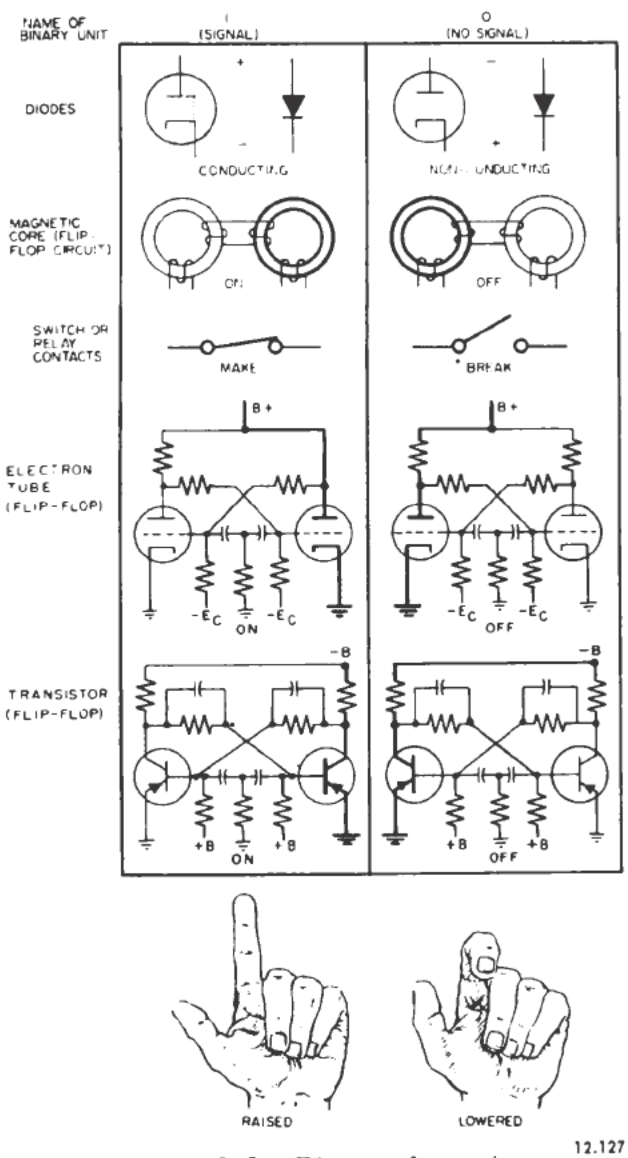


Figure 19-2.—Binary elements.

this number assignment process "coding". Let the thumb of the right hand represent one, then the second finger represents twice the value of the first, or 2. The third finger represents twice the value of the second, or 4. Continue assigning each finger of the right hand twice the value of the one before it, as shown in figure 19-3. Then assign each finger of the left hand a number as indicated in the drawing. Note that each value we assigned to a finger is 2 times as great as the one preceding it. A quick look at the number sequence shows that the terms in the sequence are powers of the number 2. Thus,

$$1 = 2^0$$
$$2 = 2^1$$
$$4 = 2^2$$
$$8 = 2^3$$
$$16 = 2^4$$
etc.

This arrangement is often called the 8 4 2 1 Code, after the four lowest quantities. Other names for this code is the Pure Binary Code, or the Pure Binary Number System. A number that is represented by this code is a binary number.

WEIGHTING VALUES

It is more convenient to use symbols instead of fingers to represent binary numbers. A raised finger represents 1, and 0 is represented by a lowered finger. Remember, we can do this because the finger is a binary unit; it can have two states. The value of each finger position is shown in the table in figure 19-4. This value is called the digit position weighting value, weighting value, or weight for short.

Consider the top binary number, 0000010101. We can find its value simply by adding the

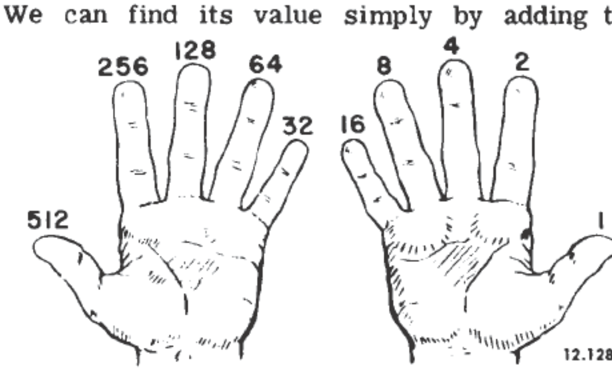
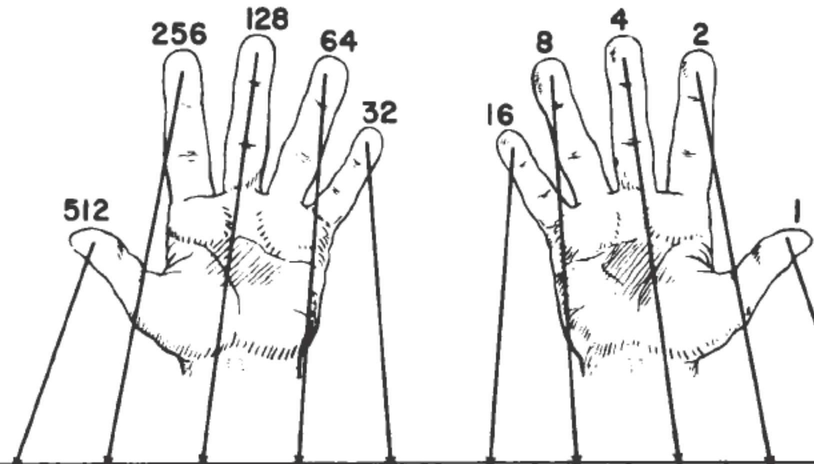


Figure 19-3.—The 8 4 2 1 code.



POWERS OF 2	2^9	2^8	2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0		
WEIGHTING VALUE	512	256	128	64	32	16	8	4	2	1		DECIMAL VALUE
	0	0	0	0	0	1	0	1	0	1	=	21
	0	0	0	0	0	0	0	0	0	0	=	0
	1	1	1	1	1	1	1	1	1	1	=	1023
	1	0	0	0	1	1	1	0	0	0	=	568
	0	0	0	0	0	0	0	1	0	1	=	5
	0	0	1	0	0	0	0	1	0	1	=	133
	0	0	0	0	0	0	0	0	0	1	=	1

12.129

Figure 19-4.—The digit position weighting value.

weighting value of each position that has 1 below it. Adding from RIGHT to left, we obtain:

$$21 = 16 + 4 + 1$$

Thus,

$$0000010101 = 21$$

The values of the other binary numbers in the table are determined in the same manner. Just for drill, why don't you check them out on your fingers. Figure 19-5 will give you some answers.

Here's another way of considering the binary code. As you can see in figure 19-3, each digit (1 or 0) in a binary number has a different digit position weighting value above it. We can multiply each digit by its weighting value, add all the products, and note the result. Using 0000010101 again, we obtain: $(0 \times 512) + (0 \times 256) + (0 \times 128) + (0 \times 64) + (0 \times 32) + (1 \times 16) + (0 \times 8) + (1 \times 4) + (0 \times 2) + (1 \times 1) = 16 + 4 + 1 = 21$. This is the same value that we obtained before.

One more point must be noted. By convention, weighting values are always arranged in the same manner. The highest one is on the extreme left, and the lowest one is on the extreme right. Thus, position weighting values begins at 1, and increases from RIGHT to LEFT. This convention possesses two very practical advantages. First, it enables us to eliminate the table. We do not have to label each binary number with weighting values, because we know that the digit at the extreme right is always multiplied by 1, the one to its left is always multiplied by 2, etc. Second, we can often eliminate some of the zeros. Notice that the zeros, whether they are to the left or the right, never add to the value of the binary number. Zero times any number is zero. Thus, for practical purposes, only the 1's have to be multiplied by their weighting values, and added to each other. Some zeros are required, however. The zeros to the right of the highest valued 1 serve as place-keepers, or spacers, to retain the 1's in their correct

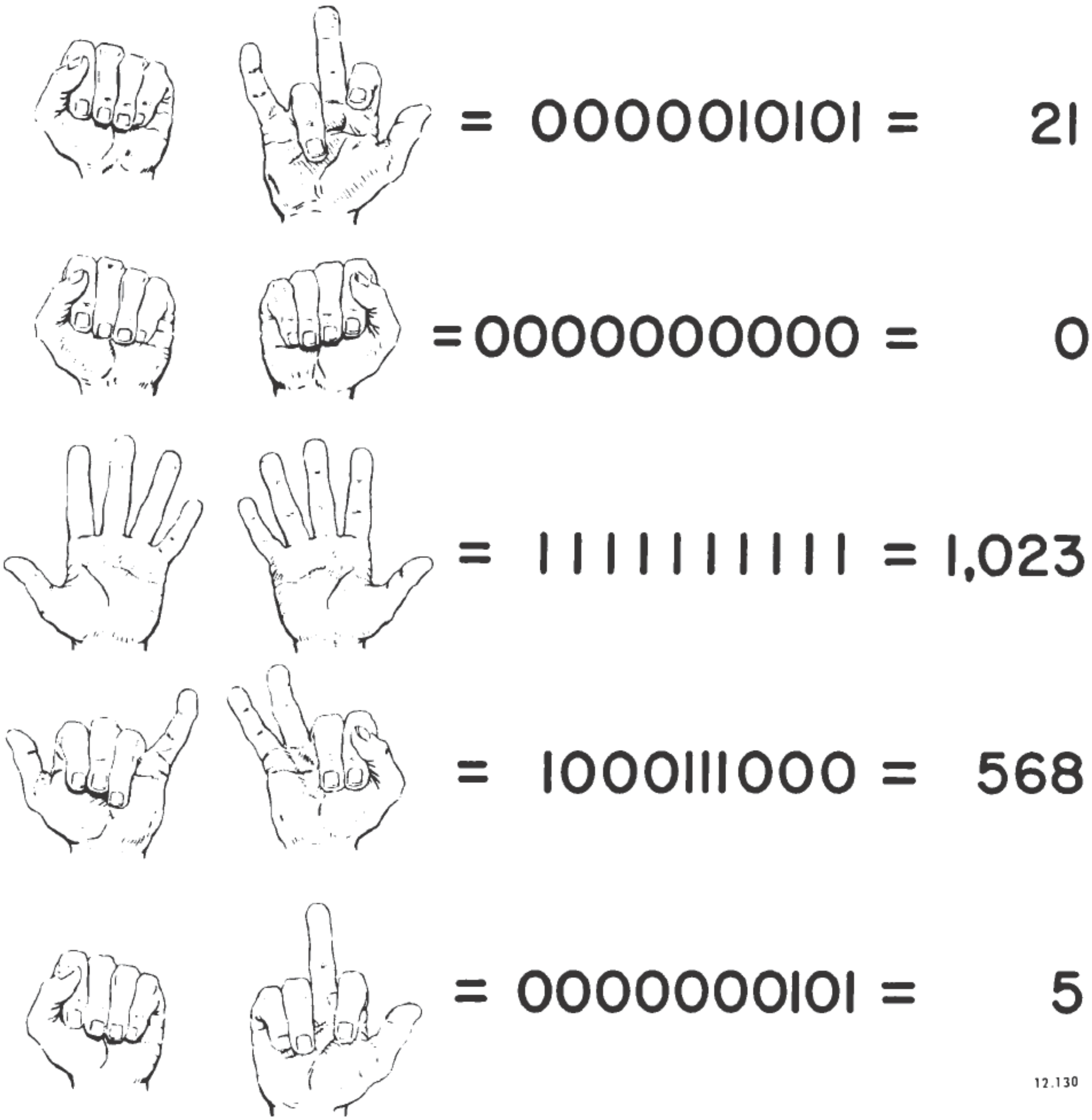


Figure 19-5.—Finding the value of the binary number.

positions. However, the zeroes to the left provide no information about the number, and so they can be eliminated. Thus, 0000010101 can be written as 10101.

Significant Digits

The correct name for the left-most 1 in any binary number is the Most Significant Digit. This is often abbreviated as MSD. It is the "most significant" because it is multiplied by the highest digit position weighting value. The Least Significant Digit (LSD) is, as you have probably guessed, the extreme right digit. Unlike the MSD, it may be a 1 or a zero. The LSD has the lowest weighting value, namely 1. Figure 19-6 illustrates the meaning of MSD and LSD. These terms possess the same meanings in both the binary and decimal systems.

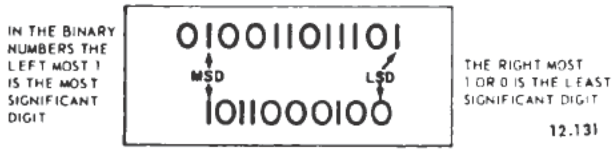


Figure 19-6.—The most significant digit (MSD) and the least significant digit (LDS).

CONVERTING FROM DECIMAL TO BINARY

Obviously it is necessary to convert (by hand or mechanically) from the decimal to the binary system when making an input to a binary digital computer. Also it will be necessary to convert binary digital results to decimal form to produce a usable output.

Here, and in NavPers 10069, *Mathematics*, you learned that our decimal system of Hindu-Arabic characters is to the base 10. There are 10 digits in this system. Each place in a decimal number represents a power of 10 starting with 10^0. The digits give the frequency of the power of 10 for the particular place.

For example:

$(4{,}208)_{10}$ ←Indicates number is to the base ten

$$
\begin{aligned}
8 \times 10^0 &= 8 \times 1 &&= 8 \\
0 \times 10^1 &= 0 \times 10 &&= 00 \\
2 \times 10^2 &= 2 \times 100 &&= 200 \\
4 \times 10^3 &= 4 \times 1000 &&= 4000 \\
& && \overline{(4208)_{10}} \text{ or } 4208
\end{aligned}
$$

You just learned a moment ago that the binary system is to the base 2, and that there are only two digits in this system. Each place in a binary number represents a power of two, starting with 2^0.

For example:

MSD LSD

$(101011)_2$ ←Indicates number is to the base two

Decimal Equivalent

$$
\begin{aligned}
1 \times 2^0 &= 1 \times 1 &&= 1 \\
1 \times 2^1 &= 1 \times 2 &&= 2 \\
0 \times 2^2 &= 0 \times 4 &&= 0 \\
1 \times 2^3 &= 1 \times 8 &&= 8 \\
0 \times 2^4 &= 0 \times 16 &&= 0 \\
1 \times 2^5 &= 1 \times 32 &&= 32 \\
& && \overline{(43)_{10}} \text{ or } 43
\end{aligned}
$$

In the previous example, besides showing you how each place in a binary number represents a power of two, we also converted a binary number to its decimal equivalent. Now let's convert a decimal number to a binary number. We will convert the decimal number, 173, to its binary equivalent.

Thus,

$$
\begin{aligned}
(173)_{10} &= 128 + 45 \\
&= 2^7 + 32 + 13 \\
&= 2^7 + 2^5 + 8 + 5 \\
&= 2^7 + 2^5 + 2^3 + 4 + 1 \\
&= 2^7 + 2^5 + 2^3 + 2^2 + 2^0 \\
&= (1\ 0\ 1\ 0\ 1\ 1\ 0\ 1)_2 \\
& \boxed{2^7\ 2^6\ 2^5\ 2^4\ 2^3\ 2^2\ 2^1\ 2^0}
\end{aligned}
$$

This process of converting a decimal number to a binary number is known as the subtraction method. First, find the highest power of two that can be subtracted from 173. In this case, it is 2^7 or 128. You successively subtract the highest power of two that can be taken from the remainder. The powers of two that can be subtracted are noted by a "one." Zeros occupy the spaces not used.

Decimal to binary conversion, as you have seen, is an easy process. In the following section you will learn to add binary numbers. If you memorize a few simple rules, you will find binary addition is also a simple process.

BINARY ADDITION

Binary digital computers can add, subtract, divide, and multiply. In the interest of brevity, we will discuss only binary addition. Now let's consider the following pencil-and-paper addition of two binary numbers.

```
 1    carried
  10 = 2
  11 = 3
 ---   -
 101 = 5
```

This addition is easily accomplished by observing these simple and logical rules:

0 and 0 = 0

0 and 1 = 1

1 and 0 = 1

1 and 1 = 0 and carry 1

In the above example we first add the two digits in the first place column to get the result of 1. Then we add the two digits in the second place column, using the last rule listed, with a result of 0 and carry 1 to the next column on the left.

Here's another more complex example:

```
  1
  1 1 1 1    } carried
 ---------
      1 0 1 =  5
      1 1 1 =  7
    1 1 1 1 = 15
 1  1 0 1 1 = 27
```

Study these examples and make up a few of your own. Make sure you understand the binary addition process and then try converting each of the binary numbers to decimal form, adding them in a decimal column with the binary addition as proof. After a little practice you may find that you can add binary numbers faster than you can decimal numbers. So much for the addition of binary numbers using the paper-and-pencil method. In the next section you will see how adding is done by a simple binary digital computer.

Binary Addition Using a Device

Fundamentally, a digital computer is an electronically operated machine or system that receives data and processing instructions, then performs arithmetical and logical operations to produce required answers. It performs these operations by its ability to rapidly count electrical pulses which represent numbers (binary digits). A pulse as commonly used in digital computer practice is a d-c voltage that rises very sharply from zero to maximum value, falls equally sharply to zero, lasts only a few microseconds or less, and may repeat 500 times or more per second. These pulses are the same basic type used as timing pulses in a radar. In fact, the pulse generating and wave shaping circuits in a digital computer are similar to those used in other types of radar systems.

One of the basic circuits in digital computers is the FLIP-FLOP, also called bistable multivibrator, Eccles-Jordan trigger circuit, and other terms too numerous to mention. The flip-flop circuit, if you will remember from your study of chapters 8 and 9, has a circuit arrangement of two triodes (or transistors) and other components. Here we will review its characteristics when a negative pulse is applied to the input (fig. 19-7).

When the flip-flop is off (fig. 19-7), V1b is not conducting and its plate voltage is nearly equal to the +B supply voltage (small current flows through R4). The neon indicator is out. The voltage division across R2, R6, and R4 makes the V1A grid-voltage across R2 only slightly more than the drop across R_K, and V1A is forward biased and conducting (reset condition). The V1A plate voltage is low due to the relatively large drop across R5. The voltage division across R5, R7, and R1, makes the V1b grid voltage across R1 considerably

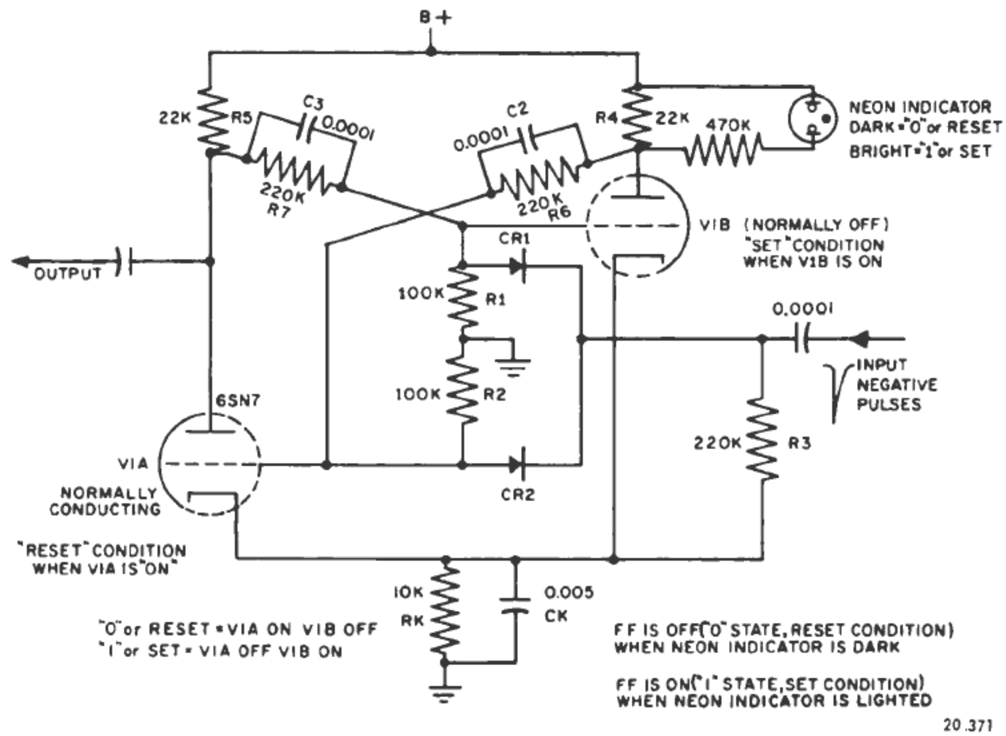


Figure 19-7.—Bi-stable multivibrator (flip-flop).

less than the drop across R_K and biases V1b to cutoff.

The neon indicator will glow when a short negative input pulse triggers V1B into conduction ("1" or "set" state). The negative pulse momentarily impresses a voltage via CR2 on the V1A grid which cuts off V1A. The rise in V1A plate voltage is impressed on the V1B grid via R1 of voltage divider R5, R7, R1, and triggers V1B into conduction. The accompanying decrease in V1B plate voltage (due to increased drop across R4) causes a decrease in the V1A grid voltage across R2 of voltage divider R4, R6, R2 below the magnitude of the voltage across R_K to bias V1A to cutoff after the negative pulse across R2 has ceased. The multivibrator will remain in the "1" or "set" state (V1B on) until the next negative input trigger is applied.

Diodes CR1 and CR2 prevent positive input pulses from getting to the grids and also direct the negative trigger pulses to the conducting triode. For example, a negative trigger applied to the multivibrator when it is in the "zero" or resting state (V1A on, reset condition) will encounter diode CR2 in a conducting condition due to forward bias developed across R2 as a result of voltage divider action across R2, R6, and R4. At the same time CR1 is back-biased by an amount equal to the excess of the voltage across R_K over that across R1. Thus, the negative trigger is initially blocked from V1B; it is immediately effective on V1A and cuts off V1A. Diode CR1 is back-biased until the input pulse nears its peak; the effect of the negative pulse added to the already negative V1B grid (with respect to cathode) is small compared to the positive pulse from the V1A plate. The rise in V1A plate voltage comprises the positive pulse that triggers V1B into conduction as previously noted. Thus, the first negative input pulse causes the flip-flop (FF) to change from the "reset" ("0") condition (V1A on and V1B off) to the "set" ("1") condition (V1B on and V1A off). Because the output is coupled to the V1A plate the initial trigger causes the output pulse to go positive. This pulse is not a useful output to a second stage identical to the first because the diodes block it.

A second trigger applied to the input of the FF will cause it to go from the "set" or "1" condition (V1B on and V1A off) to the "reset" or "0" state (V1B off and V1A on). The accompanying decrease in the V1A plate voltage comprises a useful output pulse that will not be blocked by the diodes in the following stage.

Thus, we can say that the FF has a useful output when going from set (1) to reset (0), but no useful output when going from reset (0) to set (1).

Summarizing this action, the flip-flop circuit is BISTABLE;. this means that when it is energized it has two stable states. In one it has an output; in the other it doesn't. When you feed it a single pulse, it changes from one state to the other. Feed it a second pulse, and it will change back to the original state. Because it does this, you can arrange a series of flip-flops so that they will add the pulses and show the sum in binary form.

Assume that four flip-flop circuits like that of figure 19-7 are connected in cascade to form a four-digit binary counter, as in figure 19-8. Neon indicators may be used to indicate a "1" or "set" condition when they are glowing; or a "0" or "reset" condition when they are off. The four indicators may be read from left to right as a four-digit binary number. The input pulses to be counted are applied to the right-hand flip-flop, which represents the least significant digit.

Assume that all four flip-flops are in the zero (reset) condition (upper triodes off), and that a negative pulse appears at the input. FF_1 changes state, reset to set, and its indicator goes on; but FF_1 produces no output, and the other three flip-flops continue to indicate zero. The four indicators may be red as a binary number: 0001. With the second input, FF_1 returns to zero, and provides a negative-going input to FF_2. FF_2 changes to the "one" state, but produces no output. The binary indication is now 0010 (the binary equivalent of decimal 2), indicating that two inputs have been counted.

With the third input, FF_1 changes to the "one" (set) state. But it produces no output. FF_2 therefore remains in the "one" state. and the binary indication is now 0011 (decimal 3). On the fourth input. FF_1 returns to zero (set to reset) and produces an output. FF_2 returns to zero (set to reset) and also produces an output. FF_3 changes to the "one" state (reset to set), but produces no output. The binary indication is now 0100 (decimal 4). And so on.

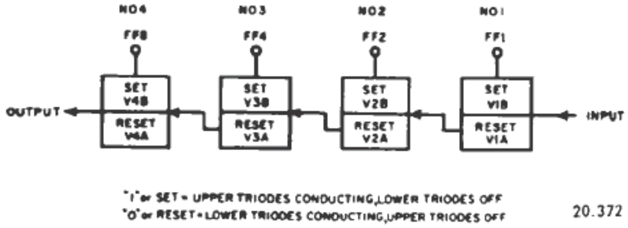


Figure 19-8.—Binary counter using bi-stable flip-flop.

After fifteen inputs, the neon tubes will indicate the binary number 1111 (decimal 15). On the sixteenth input, FF_1 changes to zero and produces an output; FF_2 therefore changes to zero and produces an output; FF_4 therefore changes to zero and produces an output, which changes FF_8 to zero. The counter therefore returns to its original state (0000) after sixteen inputs. If a fifth flip-flop were added, receiving its input from FF_8, the counter could count up to 11111 (decimal 31), and return to the original state on the 32nd input. And so on.

To repeat, in zero condition, all the flip-flops are "OFF" or in the reset condition (See part A of fig. 19-9). Now let three pulses go into the flip-flops, which are arranged in series (part B of fig. 19-9). The first pulse turns No. 1 from nonconducting, OFF, to conducting, ON. This action energizes the neon lamp associated with the first flip-flop. The second pulse to No. 1 causes it to go from "1" to "0" and to produce an output to No. 2, switching it ON. The third pulse switches No. 1 ON again.

After three pulses, the indicator lamps for flip-flops Nos. 1 and 2 are ON, while Nos. 3 and 4 are dark. If a glowing lamp equals 1 and a dark lamp equals 0, the binary indication is 0011. This corresponds to 3 in the decimal system.

Now let 5 more pulses go into the flip-flops. The first pulse of this group of 5 produces outputs from Nos. 1 and 2, since they are ON, and switches No. 3 to ON, but in going through, it turns Nos. 1 and 2 to OFF. The second pulse of the 5 merely turns No. 1 ON (no output), so that the indication is 0101 which is 5 in binary notation. You can follow this same sequence, with the assistance of your fingers and your knowledge of the powers of two, to 8, at which point the flip-flops will indicate 1000, as shown in part C of figure 19-9.

This example illustrates the counting function possible with a series of flip-flops. With modifications and elaborations of this and the addition of other circuits it is possible to perform

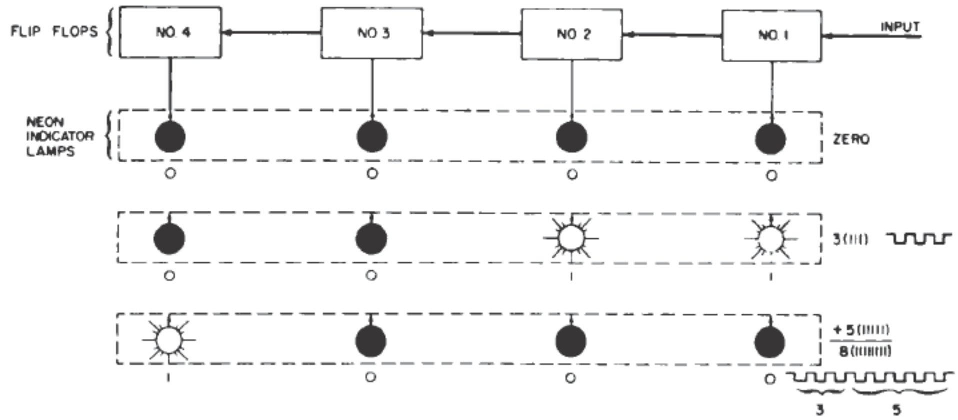


Figure 19-9.—Adding function of flip-flop. 12.132

all the fundamental arithmetic operations. Note too that the flip-flops are capable of storing information. In other words, they have memory. In this case they remembered the binary word for 3 until the pulse train representing 5 came along. These two numbers were added and the binary sum was stored. In the digital computer the unit which stores a number, then adds another number to the stored number, and then stores the sum of the two, is called an accumulator.

BASIC LOGIC CIRCUITS

Digital computers are used to make logical decisions (but not to reason) about matters which can be logically decided upon. Some logical decisions that a computer may be required to make are (1) when to perform an operation; (2) what operation to perform; and (3) how (of several ways) to perform it. Keep in mind that machines never reason why. They operate from instructions prepared by a man interpreting them to determine what decisions are to be made.

In order for the computer to make logical decisions or perform logical operations the proper information must be given it. The information given to the computer is called the "program." The program is a series of coded instructions which are stored in the computer's memory and they tell the computer how to store, locate, process, and present the data it receives. Solutions to various problems are solved basically by means of logical truths.

LOGICAL TRUTH differs from ordinary truth in that we not only deal with facts, but also in suppositions based on facts. For example, the statement "sugar dissolves in water" is an ordinary truth. Here, however, certain things are understood—the amount of sugar is much smaller than the amount of water; if we tried to mix a whole bag of sugar with a teaspoonful of water, all of the sugar would not dissolve. The computer cannot operate by understanding such limiting conditions, unless these conditions are given to it so that it may form a logical truth pattern to work from.

Many logical truths are used by people everyday without their realizing it. Most of our simpler logical patterns are distinguished by words such as: AND, OR, NOT, if, else, then. In mathematical terms these logical patterns may be indicated by plus, minus, times, divided by, or combinations of these terms.

If a set of statements is made, it is possible to set up a table that will determine what is known as the truth values of these statements. In setting up this table, it is necessary to use some of the logical connectives, such as AND, OR, NOT, etc. In the end we would have a table that would list the truth or falsity of each of the conditions of the statements that we were considering.

It is possible to set up mathematical relations of the truth values of statements and perform mathematical operations on these statements. This study of mathematical logic is called the algebra of logic.

In 1847, a mathematician named George Boole demonstrated that when logical statements are written in certain symbolic forms, these symbols can be manipulated very much like algebraic symbols.

There are many different types of symbolic logic, but for our purposes only one type will be described. This is the type that is used in most computers, and we will refer to it as either symbolic logic or Boolean algebra.

In the arithmetic with which you are familiar, there are four basic operations: adding, subtracting, multiplying, and dividing. In Boolean algebra there are only three basic operations: AND, OR, NOT. If these phrases do not sound mathematical, it is only because of the fact that logic began first with words, and not until much later was it translated into mathematical terms by Boole. You will find that the three basic logical operations are very similar to the words used to describe them.

For convenience, the familiar arithmetic symbols are used for two of the Boolean algebra operations.

Thus,

AB means A AND B

A + B means A OR B

Notice that the OR operation is indicated by the addition symbol, while the AND operation is indicated by the multiplication symbol. If desired we can use any of the multiplication symbols, such as the parentheses or the dot, to indicate the AND operation.

Thus,

(A + B)(C) means A OR B, and C

A · B means A AND B

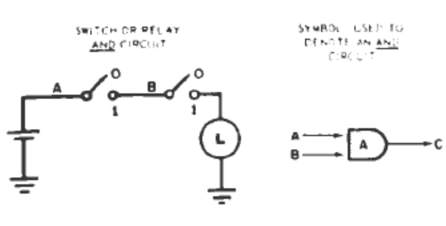
The NOT operation is indicated by a bar over the letter.

Thus,

$\overline{A}$ means A NOT, while A$\overline{B}$ means A AND B NOT

AND Circuit

We are going to illustrate the three basic logical operations by using simple electrical circuits. Look at figure 19-10. Here, you see a circuit designed to light a lamp. In this instance, series switching is used to represent an AND circuit. The open condition of each switch may be signified by 0. The closed condition of each switch may be indicated by 1. The absence of current in the lamp is signified by 0 and the presence of current by the symbol 1. A truth table may be used to list the combinations of conditions and their results. A truth table is a shorthand method of describing a logic circuit's complete function. If you study the table shown in the figure for a moment, you can see that the AND circuit represents MULTIPLICATION. In this circuit if a switch is open, it will be shown in the truth table as 0; if it is closed it will be shown in the table as 1. A 0 in the output column indicates no-flow of current and 1 shows that current is flowing in the circuit. By referring to the table, you find that there is just one condition of the switches that allow current to flow through the lamp (output). This condition exists when switches "A" AND "B" are both closed.



TRUTH TABLE

INPUT		OUTPUT
A	B	F
0	0	0
0	1	0
1	0	0
1	1	1

12.133

Figure 19-10.—The AND circuit.

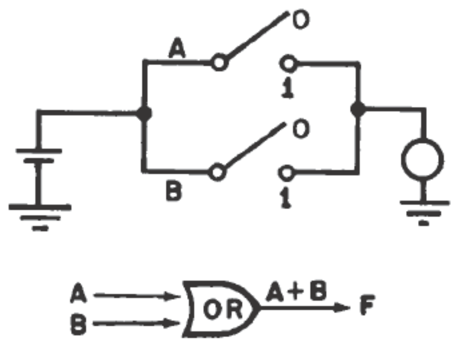
OR Circuit

The OR circuit represents addition. If it were not necessary to perform the carry operation in binary addition the OR circuit could be used by itself. However, this is not the case, so OR circuits must be used in combination with AND and NOT circuits to perform addition of numbers that require a carry.

We can represent an OR circuit electrically by using two single-pole knife switches connected in parallel as shown in figure 19-11. There is an output signal at F if an input signal is placed at A OR B. In this case, placing a signal at a point in the circuit means closing a switch. We have used only two inputs to keep the illustration simple; however, there can be any number of input signals to an OR circuit. Any one or more of these input signals can produce an output signal.

NOT Circuit

The last of the basic logic circuits that we will briefly cover is the NOT circuit. The requirements of a NOT circuit are that a signal injected at the input produce no signal at the output, and that no signal at the input



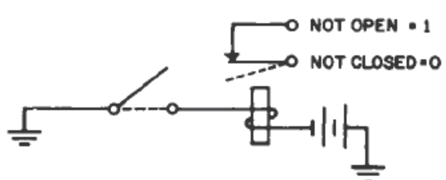
INPUT		OUTPUT
A	B	
0	0	0
0	1	1
1	0	1
1	1	1

12.134

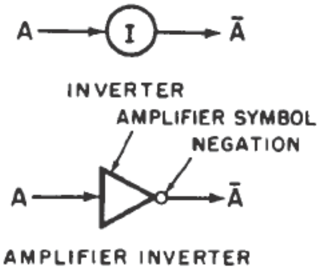
Figure 19-11.—The OR circuit.

produce a signal at the output. A switch and a relay, for example, can be connected as shown in figure 19-12 to form a simple NOT circuit. If we identify a closed contact as a signal and an open contact as no signal you can easily see that this connection fulfills the binary requirements of a NOT circuit. A signal (closed contact) at its input produces NO signal (open contact) at its output, and vice versa. The truth table shown in figure 19-12 illustrates the shorthand method of describing the NOT circuit's function. The symbol ''A'' (read ''A NOT'' or ''bar A'') means the opposite of A. This opposite or inverted condition of A points out the reason why the NOT circuit is sometimes called an INVERTER.

Another familiar example of a NOT or inverter is a simple triode vacuum tube circuit. As you know, phase inversion takes place in it. With a positive signal applied to the control grid, the output is a negative signal.



INPUT	OUTPUT
A	$\bar{A}$
0	1
1	0



12.135

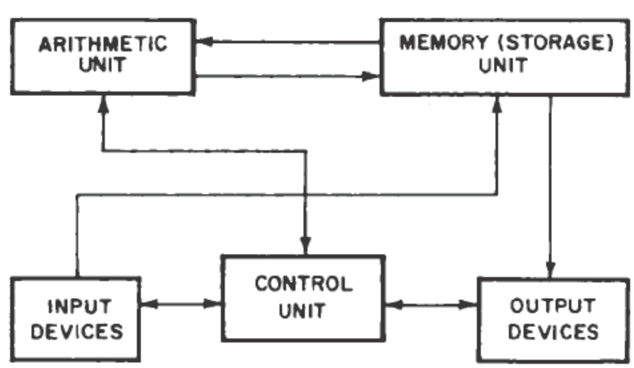
Figure 19-12.—The NOT circuit.

ORGANIZATION OF A BASIC DIGITAL COMPUTER

Figure 19-13 shows a block diagram of an elementary digital computer. Most digital computers contain the basic sections shown in the diagram. If a machine is to be called a computer, it must be capable of performing some type of arithmetic operations. For the purpose of organization, the elements of a digital computer that will meet this requirement will be classified as the ARITHMETIC unit.

We must be able to control the operation of the arithmetic unit so that it can accomplish the desired task. Consequently, a CONTROL unit is necessary.

Since the arithmetic unit performs a mathematical operation, it may be required to store a partial answer while the unit computes another part of the problem. This stored partial answer can then be used in another section to solve another part of the problem. It may also be desirable for the control unit and the arithmetic unit to have information available for their use and for use by other units within the machine. To meet this requirement, the computer must contain elements that are capable of memorizing or storing information. These elements are included in the MEMORY unit.



12.136

Figure 19-13.—Organization of a typical digital computer.

The digital computer is built for the purpose of serving humans in some manner. Therefore, there must be some method of transmitting our needs to the computer and of receiving the results of the machine's computations. The components that perform these functions are called the INPUT-OUTPUT units.

The name "arithmetic unit" does not imply that the circuits that perform arithmetic operation are physically lumped together. In fact, it is possible that a particular circuit may operate during one moment as part of the arithmetic unit while during the next moment that circuit may function as part of the control unit.

It is extremely important that the machine operate at high speed. In order to accomplish the required tasks within a reasonable length of time, the computer must possess a certain minimum speed of operation.

ARITHMETIC UNIT

Basically, the arithmetic unit of the digital computer is the operational unit. It performs the actual work of computation and comparison. It can do its job by counting pulses, or by using logic circuits. Modern computers employ components such as electron tubes, crystal diodes, and transistors. Switches and relays have been used in the past, and they are acceptable as far as their ability to perform computations are concerned. However, because of their greater speed and smaller size, modern computers tend to make use of electronic components wherever possible. Therefore the arithmetic unit of the present day computer is made up of such operational elements as binary counting circuits, and logical AND, OR, and NOT circuits.

CONTROL UNIT

The control unit of the digital computer is the administrative, or switching section, of the computer. It receives the incoming information that enters the machine, and decides what is to be done and when to do it. It tells the arithmetic unit what to do and where to find the necessary information. It decides when the arithmetic unit has completed following its instructions. The control unit decides what to do with the results, and tells the arithmetic unit what to do next. The control unit contains the switching and gating circuits that are required for its operation, and it also contains the basic timing generators.

The control unit contains many circuits found in the arithmetic unit. Some of the arithmetic and control circuits may be so closely interrelated that it is difficult to tell one from the other. With the exception of a few completely automatic operations, everything that happens in the machine is noted by the control unit. This unit might possibly be distinguished from the arithmetic unit by the fact that it contains a greater number of electrical connections, as well as a greater number of switching or gating circuits. However, in general, the circuits of the control unit are very similar to those in the arithmetic unit.

MEMORY UNIT

The memory unit, sometimes called the storage unit, provides information to both the control unit and the arithmetic unit. The terms "memory" and "storage" may be used interchangeably. The Instruction book for a certain digital computer may refer to the physically separate magnetic tape units as storage units and to the magnetic cores or drums that are inside the machine as memory units. This does not necessarily mean that there is any real difference between the definition of memory and that of storage as applied to digital computers. Here, we will define this storage function as memory.

The requirements of the memory unit may vary considerably within one machine. When the arithmetic unit is multiplying two numbers, each containing more than one digit, it performs a series of additions for each digit within the multiplier. The memory unit "holds" that information until the required addition has been performed for the next digit, and the unit then adds these two results. The word "hold" implies that the memory unit has the ability to remember. It is not incorrect to say that the section of a computer that holds information until the arithmetic unit is ready for it is the memory unit. The memory unit may merely consist of a series of capacitors that store a charge for the required period of time, or it may consist of a unit, such as an accumulator, that stores data in the form of static d-c voltages at the output of the flip-flop circuits. The memory could be stored on a length of magnetic tape that is wound on a reel or on the magnetic surface of a revolving drum or disc. Whatever this unit may be, it stores information until it is needed.

In other computations, the arithmetic unit may be required to refer to numbers that have been used previously. Since one of the requirements is speed of operation, it would be better if we could avoid having to reinsert this value each time it is to be used. Consequently, it is made available for the arithmetic unit within the machine by means of local storing. Also it is possible to store all of the information and instructions needed to perform the desired computations.

Information can be stored in many ways other than those that have been mentioned. The equipment may use gas-filled tubes, punched paper tape, magnetic cores, photographic film, or even relays and switches.

INPUT DEVICES

Input devices insert the values with which we are concerned, together with the instructions for their use, and the relay them to the control unit. The output devices record, in some manner, the results of computer operations. These results may be recorded in permanent form or they may be used to initiate some type of direct physical action.

Input unit requirements are complex. For a small computer, a manually operated keyboard may be considered sufficient. Other computers may use cards in which holes have been punched according to a prearranged sequence. The number and location of the holes in the card convey the desired information to the control unit. In some systems, instructions may be prewired on a removable plugboard that is attached to the computer. Cards may also contain typed or written information intended for purposes other than providing data input to the computer. The punched paper tape, which is similar in principle to the punched card, is used to feed information into the computer. This method does not provide as great flexibility as the punched-card system. If a more rapid means of input is required, magnetic tapes may be used.

OUTPUT UNITS

Any one of the previously mentioned input devices could be used to take the information out of the machine and record it. The output section translates the answers into appropriate forms—such as holes in punch cards or tapes, magnetic patterns on tape, and characters produced by electric typewriter. The output section contains tranducers, card punch or tape punch units, magnetic recording heads, and other devices as required.

In conclusion, computers can be constructed from many basic components, some of which you are already familiar with. They can be designed to perform many kinds of jobs. However, the organization of the computer must fulfill the basic fundamental requirements of arithmetical ability, control, memory capacity, and necessary input and output equipment.

APPENDIX I

TRAINING FILM LIST

Training films that are directly related to the information presented in this training course are listed below. Under each chapter number and title the training films are identified by Navy number and title and are briefly described. Other training films that may be of interest are listed in the United States Navy Film Catalog, NavPers 10000 (revised).

Chapter 1

OPERATING PRINCIPLES OF THE ELECTRON TUBE

MA-2031A — Vacuum Tubes—Elementary Electron Theory and the Diode Tube. (16 min.—B&W—Sound—Unclassified—1942.) USAF TF1-470. Shows, by animation, how the electron theory is applied in the diode tube and how half wave and full wave rectification is obtained with this tube.

MA-2031B — Vacuum Tubes—The Triode and Multipurpose Tubes. (14 min.—B&W—Sound—Unclassified—1942.) USAF TF1-471. Parts of the tube are pointed out and grid bias, amplification, and several varieties of the triode are shown and explained. Both straight photography and animation are employed.

SA-2519A — Vacuum Tubes—Part 1—Fundamentals. (126 frames—B&W—Silent—Unclassified—1943.) USAF FS1-177. Describes the following: Edison effect; electron emission; operation of the diode, triode, and tetrode tubes; pentode special purpose tube; and, the cathode ray tube.

SN-650 — Vacuum Tubes. (37 frames—B&W—Silent—Unclassified—1942.) Describes theory of operation of vacuum tubes and function in the radio circuit.

SN-1540G — Radio Technicial Training—Special Purpose Vacuum Tubes. (19 min.—75 frames—B&W—Sound—Unclassified—1943.) Explains construction and use of various vacuum tubes; pointing out that size, uses, and models are multiple.

ME-6092E — Engineering Electronics—No. 3—The Triode Amplification. (14 min.—B&W—Sound—Unclassified—1945.) USOE OE-177. An elementary film which discusses parts of the triode (grid, cathode, and anode), and explains effect upon amplified output of variations in grid current. Use of tube as DC or AC voltage amplifier is considered.

Chapter 2

INTRODUCTION TO TRANSISTORS

MN-8479A Transistors—P-N Junction Fundamentals. (11 min.—B&W—Sound—Unclassified—1957.) Introduces the theory and mechanisms of semi-conductor diode and transistor action. Discusses the fundamental principles that apply to all transistors and junction rectifiers.

MN-8479B Transistors—Triode Fundamentals. (10 min.—B&W—Sound—Unclassified—1957.) Shows how junction transistors, or triodes, consist of three sections with two P-N junctions separating them. Discusses the fundamentals of this arrangement as an amplifying device.

MN-8479C Transistors—Minority Carriers. (10 min.—B&W—Sound—Unclassified—1958.) This film introduces the principle of minority carriers, showing how they produce a small reverse current under normal conditions. It demonstrates the limitations imposed on transistor behavior by minority carriers when the transistor is heated or loaded.

Chapter 3

POWER SUPPLIES FOR ELECTRONIC EQUIPMENTS

MN-8611B Radio Set AN/GRC-14—Theory of Operation. (19 min.—B&W—Sound—Unclassified—1959.) This film shows the theory of operation and the circuits of the Radio Set AN/GRC-14(AN/MRC-55) transmitter. The set includes a transmitter, two receivers, a power supply and remote-control units. All of these units employ familiar, conventional circuits except the transmitter.

MA-8862C Radio Set AN/GRC-26—Part 3—Operation Power Plant and Starting Operation. (4 min.—B&W—Sound—Unclassified—1958.) US Army TF11-2491. This film discusses the features and function of the PE-95 Power Plant of the Radio Set AN/GRC-26, showing how it is started and how it operates to feed power to the equipment in the shelter.

Chapter 4

TUNED CIRCUITS

SN-649 Tuning—Radio. (27 frames—B&W—Silent—Unclassified—1942.) Describes theory of tuning radio circuit.

SN-647 Inductive Reactance. (33 frames—B&W—Silent—Unclassified—1942.) Explains basic theory of inductive reactance and application to radio instruments.

SN-648 Capacitive Reactance. (29 frames—B&W—Silent—Unclassified—1942.) Explains basic theory of capacitive reactance and application to radio instruments.

MN-1540A — Radio Technician Training—Capacitance—Part 1 and Part 2. (15 min. each part—B&W—Sound—Unclassified—1943.) Demonstrates flow of electrons through circuit and charging and discharging condensers. Variations of charge on condenser in relation to time and the behavior of capacitance with alternating current are discussed. (Accompanied by film strip SN-1540AA)

MN-1540B — Radio Technician Training—Inductance—Part 1 and Part 2. (18 min. each part —B&W—Sound—Unclassified—1943.) Shows by means of straight photography and animation how magnetic force reacts around a coil. Explains self-inductance, how to increase value of it, and how to increase the inductance of coil. (Accompanied by film strip SN-1540AB)

MN-1540C — Radio Technician Training—RCL—Resistance, Capacitance—Part 1 and Part 2. (14 min.—Part 1—21 min.—Part 2—B&W—Sound—Unclassified—1943.) Explains current and voltage in relation to time; voltage and current curves; how current leads voltage and voltage leads current in different instances, and relation of current and voltage. Shows graphic method measuring voltage at source, addition of phase components, and effect of impedance on resonance. (Accompanied by film strip SN-1540AC)

MN-1540N — Radio Technician Training—Vectors. (12 min.—B&W—Sound—Unclassified—1945.) Film defines a vector as a line indicating direction and quantity, as a force or movement, and explains how vectors may be added to produce a resultant of different quantity and perhaps different direction. A number of analogies from carnival entertainments emphasize concepts involved.

MA-8273 — Tuned Circuits. (28 min.—B&W—Sound—Unclassified—1954.) US Army TF11-1831. Shows theory and application of resistive, capacitive and inductive circuits, with DC and AC, series, series resonant and parallel resonant tuned circuits.

MN-8018A — Basic Electricity—Inductance in AC Circuits. (7 min.—B&W—Sound—Unclassified—1959.) This film gives the definition of inductance and explains its cause and effect in an AC circuit. It defines and illustrates inductive reactance, factors that affect inductive reactance in an AC circuit, and the phase relationship between current and voltage as a result of the presence of inductance in the circuit.

MN-8018B — Basic Electricity—Capacitance in AC Circuits. (5 min.—B&W—Sound—Unclassified—1959.) This film defines capacitance and demonstrates how a capacitor works. It discusses the physical factors that affect capacitance and the effect of capacitance in an AC circuit. It also defines capacitive reactance, its effects in an AC circuit, factors that affect capacitive reactance, and phase angle.

MN-8018C — Basic Electricity—AC Series Circuits. (4 min.—B&W—Sound-Unclassified—1959.) This film shows the elements of an AC series circuit, and demonstrates by sine waves and circuits, the effects upon and phase relationship between current and voltage. Using an LCR circuit, it examines the effects of voltage shows what the generator sees when XL exceeds XC and when XC exceeds XL. It also demonstrates a series resonant circuit using a LCR circuit where XC and XL are equal.

MN-8018D Basic Electricity—AC Parallel Circuits. (5 min.—B&W—Sound—Unclassified—1959.) This film shows the elements of an AC parallel circuit. It examines the effects of current and shows what the generator sees in the following circuits: LC parallel circuit, where XL exceeds XC; an LCR parallel circuit, where XC exceeds XL; and an LCR parallel circuit, where XC equals XL.

MA-8473 Radio Sets AN/PRC 8, 9 and 10. (26 min.—B&W—Sound—Unclassified—1955.) US Army TF11-2180. Explains frequency, operating channels, tuning, operation, maintenance typical installations, tactical employment and use of auxiliary equipment.

MA-8273 Tuned Circuits—(28 min—B&W—Sound—Unclassified—1954)

Chapter 5

INTRODUCTION TO ELECTRON TUBE AMPLIFIERS

SA-2519C Vacuum Tubes—Part 3—Amplification Fundamentals. (125 frames—B&W—Silent—Unclassified—1943.) USAF FS1-196. Explains amplification and shows practical applications. Shows difference between AF and RF amplification and describes these classes of operation: vacuum tube method; obtaining grid bias; voltage coupling methods; push-pull operation of tubes.

MN-8086C Radio Transmitting Sets AN/URT 2, 3, 4—The RF Amplifier and Modulator. (7 min—B&W—Sound—Unclassified—1955.) Shows how the frequency is amplified in the RF amplifier and modulated in the Low Level Radio Modulator for both voice and various types of keying modulation.

ME-6092E Engineering Electronics—No. 3—The Triode Amplification. (14 min.—B&W—Unclassified—1945.) USOE OE-177. An elementary film which discusses parts of the triode (arid, cathode, and anode), and explains effect upon amplified output of variation in grid current. Use of tube as DC or AC voltage amplifier is considered.

Chapter 6

ELECTRON TUBE AMPLIFIER CIRCUITS

SA-2519C Vacuum Tubes—Part 3—Amplification Fundamentals. (125 frames—B&W—Silent—Unclassified—1943.) USAF FS1-196. Explains amplification and shows practical applications. Shows difference between AF and RF amplification and describes these classes of operation: vacuum tube method; obtaining grid bias; voltage coupling methods; push-pull operation of tubes.

MA-1631 Radio Receivers—Principles and Typical Circuits—Part 1. (17 min.—B&W—Sound—Unclassified—1942.) USAF TF1—472. Explains crystal theory and covers amplification, tuned radio frequency superheterodyne, and audio-frequency amplification. Both straight photography and animated drawing are employed.

Chapter 7

AUDIO POWER AMPLIFIERS

SA-2520 Radio Receivers. (70 frames—B&W—Silent—Unclassified—1943.) USAF FS1-197. Amplification, selectivity, and fidelity are explained. Shows use of pentode tubes, coupling methods and manual gain control, push-pull Class A phase inversion, inverse feed-back loud speaker, superheterodyne sets.

Chapter 8

OSCILLATORS

MN-1540I Radio Technician Training—Oscillators. (13 min.—B&W—Sound—Unclassified—1945.) Picture presents an explanation of basic principles of electronic oscillation. Starting with a simple tank circuit and adding refinements of triode amplifier and grid leak, it is demonstrated that oscillations may be started and continued in the circuit. Frequency of oscillation may be altered by changing size of coil or capacitor in the circuit.

MN-8814D Electronics—Special Circuits and Devices—Free Running Blocking Oscillator Circuit—Operation. (7 min.—B&W—Sound—Unclassified—1959.) This film shows simplified version of action of blocking oscillator.

MN-8814E Electronics—Special Circuits and Devices—Triggered Blocking Oscillator Circuit—Operation. (7 min.—B&W—Sound—Unclassified—1959.) This film shows action and conditions necessary for triggered blocking oscillator.

MN-8086B Radio Transmitting Sets AN/URT 2,3,4—The RF Oscillator. (17 min.—B&W—Sound—Unclassified—1955.) Shows how the AN/URT series of transmitters generate and modulate carrier frequencies. Describes how the various circuits multiply, add, and subtract frequencies to produce the final carrier frequency, and how the frequency is amplified in the RF Amplifier and modulated in the Low Level Radio Modulator for voice or key modulation.

Chapter 9

TRANSISTOR CIRCUITS

MN-8479D Transistors—Low Frequency Amplifiers. (15 min.—B&W—Sound—Unclassified—1958.) This film shows how transistors are used to amplify low frequencies in common base, common emitter, and common collector circuits. How the transistor functions is explained in detail for a common base amplifier and a common emitter amplifier. Refined common emitter circuits are also shown.

MN8479E Transistors—High Frequency Operation—Amplifiers and Oscillators. (14 min.—B&W—Sound—Unclassified—1959.) This film describes how transistors operate in high frequency amplifiers and oscillator circuits. It shows the influence of transit effects in the base, collector capacitance, and base resistance on high frequency performance.

MN-8479F Transistors—Switching. (14 min.—B&W—Sound—Unclassified—1959.) This film shows examples of switching circuits in transistorized computers, explaining briefly the concept of digital computation and how transistors are used. The film includes a more detailed explanation of how a simple transistor switch works, minority carrier storage in the base, and how delaying effects of this storage are overcome.

Chapter 10

MODULATION AND DEMODULATION

SN-651 Detection—Radio. (33 frames—B&W—Silent—Unclassified—1942.) Describes how radio wave is detected and isolated.

MA-4975 Basic Principles of Frequency Modulation. (30 min.—B&W—Sound—Unclassified—1944.) US Army TF11-2069. Compares transmission and reception by amplitude modulation and frequency modulation. Demonstrates advantages of FM in overcoming static and how this is done.

Chapter 11

TRANSMITTERS

MN-8862E Radio Set AN/GRC-26—Part 5—Preparing Transmitter for Operation. (5 min.—B&W—Sound—Unclassified—1958.) US Army TF11-2564. This film depicts the preliminary steps necessary to prepare the radio transmitter of Radio Set AN/GRC-26 for operation on a demonstration frequency of 4280 kc.

MN-8611B Radio Set AN/GRC-14—Theory of Operation. (19 min.—B&W—Sound—Unclassified—1959.) This film shows the theory of operation and the circuits of the Radio Set AN/GRC-14 (AN/MRC-55) transmitter. The set includes a transmitter, two receivers, a power supply and remote-control units. All of these units employ familiar, conventional circuits except the transmitter.

MN-6756A The Model TDZ Radio Transmission—Tuning. (26 min.—B&W—Sound—Unclassified—1950.) Shows TDZ radio transmission equipment, and presents subject of tuning by first charging frequency channel and then adjustment of dials on the panel board.

MN-8086A Radio Transmitting Sets AN/URT 2,3,4—Introduction. (15 min.—B&W—Sound—Unclassified—1955.) Shows the units that comprise the three transmitters, the capabilities of each and how they are operated in a normal automatic condition including the setting up of ten channel frequencies. Procedures for semi-automatic and manual tuning are also shown.

MN-8086B Radio Transmitting Sets AN/URT 2,3,4—The RF Oscillator. (17 min.—B&W—Sound—Unclassified—1955.) Shows how the AN/URT series of transmitters generate and modulate carrier frequencies. Describes how the various circuits multiply, add, and subtract frequencies to produce the final carrier frequency, and how the frequency is amplified in the RF Amplifier and modulated in the Low Level Radio Modulator for voice or key modulation.

MN8086C Radio Transmitting Sets AN/URT 2,3,4—The RF Amplifier and Modulator. (7 min.—B&W—Sound—Unclassified—1955.) Shows how the frequency is amplified in the RF Amplifier and modulated in the Low Level Radio Modulator for both voice and various types of keying modulation.

MN-8086D Radio Transmitting Sets, AN/URT 2,3,4—Automatic Frequency Selection. (14 min.—B&W—Sound—Unclassified—1955.) Shows how the telephone-type dial controls the automatic frequency-selection circuit for switching the R-F Oscillator to the pre-set frequency.

MN-8086E Radio Transmitting Sets, AN/URT 2,3,4—Automatic Bandswitching. (7 min.—B&W—Sound—Unclassified—1955.) Shows how the Radio Frequency Amplifier divides the frequency range into six different bands and how the bandswitch motor automatically selects the band that includes the carrier frequency.

MN-8086F Radio Transmitting Sets, AN/URT 2,3,4—Automatic Amplifier Tuning. (18 min.—B&W—Sound—Unclassified—1955.) Shows how the automatic amplifier tuning circuits tune the Intermediate Power Amplifier and the Power Amplifier.

MN-8086G Radio Transmitting Sets, AN/URT 2,3,4—Automatic Antenna Tuning. (18 min.—B&W—Sound—Unclassified—1955.) Shows how the RF tuner and capacitor assembly automatically tunes the antenna to the frequency that has been selected for transmitting. This is the last of four automatic tuning operations that take place when a new transmitting frequency has been selected.

MN-8086H Radio Transmitting Sets, AN/URT 2,3,4—Troubleshooting. (12 min.—B&W—Sound—Unclassified—1955.) Shows how to determine the general location of trouble in the AN/URT series of transmitters. Describes the use of various lights and meters to locate trouble, the built-in test oscilloscope for testing the operation of the internal oscillators and finally the external test equipment.

MA-1706 Radio Transmitters—Principles and Typical Circuits. (18 min.—B&W—Sound—Unclassified—1942.) USAF TF1-476. Points out parts of several types of radio transmitters, describes functions, and illustrates action of the amplifier, microphone, and crystal oscillator. Both straight photography and animation are used.

SN-2773A Naval Aircraft Radio—Introduction to Airborne Communication Equipment. (52 frames—B&W—Silent—Equipment—Unclassified—1944.) Shows quantities of gear needed in fighters, torpedo bombers, patrol bombers, and LTA craft. An introduction to the SN-2773 series, B through P.

Chapter 12

TRANSMISSION LINES

MN-8814A Electronics—Special Circuits and Devices—Waveguides. (18 min.—B&W—Sound—Unclassified—1959.) This film presents a brief theory of propagation of energy down a waveguide.

MN-2562A Transmission Lines—Maintenance of the Coaxial Line. (28 min.—B&W—Sound—Unclassified—1944.) Description of the construction and principles of replacing and repairing coaxial transmission lines.

MN-2562B Transmission Lines—Waveguides—Part 1—Manufacture of Replacement Sections. (14 min.—B&W—Sound—Unclassified—1944.) Instructions for bending and twisting waveguides on both a large and portable press. Trimming, soldering, brazing, and shipping are also covered.

MN-1540K Radio Technician Training—Standing Waves on Transmission Lines. (23 min.—B&W—Sound—Unclassified—1945.) By means of laboratory demonstrations, animated diagrams, and diverse analogies, results, causes, and prevention of standing waves in radio high frequency transmission lines are discussed.

Chapter 13

ANTENNAS AND PROPAGATION

SN-1074 Directive Antenna—Radar No. 4. (51 frames—B&W—Silent—Unclassified—1942.) Explains the principle of the directional antenna, how ultrashort waves can be reflected and focused into a beam, how directional antennas can be used for both transmitting and receiving, and shows how different type directional antenna are rigged on a PBY.

MA-1704 Radio Antennas—Creation and Behavior of Radio Waves. (11 min.—B&W—Sound—Unclassified—1942.) USAF TFI—474. Demonstrates how electrodynamic and electrostatic magnet fields are formed, and illustrates behavior of sky and ground radio waves. Animated diagrams are used almost exclusively.

MA-1705 Radio Antennas—Fundamentals. (13 min.—B&W—Sound—Unclassified—1942.) USAF TF1-475. Shows how a standing wave is created; explains voltage and current strength at various areas along the antenna; shows how antenna length is determined, and demonstrates handling trailing antenna. Both straight photography and animation are used.

MA-6963 Field Wire Line Construction. (24 min.—B&W—Sound—Unclassified—1950.) US Army TF11-1632. By means of animation, shows characteristics of propagated radio waves at various frequencies. Action and characteristics of both ground waves and sky waves are shown, and advantages and disadvantages of each discussed. Effect of ionospheres on the sky waves of higher frequencies is discussed in some length. Refraction of radio waves entering the ionosphere and effect of skip distance are shown and discussed for different frequencies and angles of transmission. Effects of favorable and unfavorable ionospheric conditions are shown.

MN-9329A Radio Antenna Fundamentals—Part 1. (27 min.—B&W—Sound—Unclassified—1952.) USAF TF1-4801. Photography and animated charts technically describe radio transmission system. Cycle of wave length is followed and miracle of standing wave described, along with other radio antanna reception components.

MN-9329B Radio Antenna Fundamentals—Part 2. (16 min.—B&W—Sound—Unclassified—1952.) USAF TF1-4802. A highly technical film with animated charts and photographs covering radio transmission system in ground and air operations. Non-directional radiation and various antennas are discussed.

Chapter 14

ELEMENTARY COMMUNICATIONS RECEIVERS

MA-1631 Radio Receivers—Principles and Typical Circuits—Part 1. (17 min.—B&W—Sound—Unclassified—1942.) USAF TF1-472. Explains crystal theory and covers amplification, tuned radio frequency superheterodyne, and audio-frequency amplification. Both straight photography and animated drawing are employed.

SN-652 Audio Frequency Amplification. (25 frames—B&W—Silent—Unclassified—1942.) Describes theory and practice of audio wave amplification.

SN-653 Radio Frequency Amplification. (18 frames—B&W—Silent—Unclassified—1942). Describes theory and practice of detected radio wave amplification.

SN-654 Reproducers. (29 frames—B&W—Silent—Unclassified—1942.) Describes change of sound waves to electrical impulses, to radio waves and back to sound waves.

Reviews principles of electromagnetism and applies them to reproducers. Shows schematic makeup of a simple head phone and its operation. The dynamic type loud speaker is shown.

MN-1679B Radio Operator Training—Superheterodyne Receivers. (14 min.—B&W—Sound—Unclassified—1944.) Discusses manual procedure and principles involved in detailed operation of receivers; demonstrates typical errors and how to guard against them. (Accompanied by film strip SN-1679AB)

Chapter 15

INTRODUCTION TO RADAR

AND

Chapter 16

SPECIAL CIRCUITS

MN-8814B Electronics—Special Circuits and Devices—Cavity Resonators. (8 min.—B&W—Sound—Unclassified—1959.) This film presents a brief theory of the action of energy in a cavity resonator.

MN-8814C Electronics—Special Circuits and Devices—Saw Tooth Oscillator Circuit—Operation. (8 min.—B&W—Sound—Unclassified—1959.) This film illustrates the action of saw-tooth voltage generator with neon and thyratron tubes.

MV-9332 Microwave Oscillators. (18 min.—B&W—Sound—Unclassified—1949.) USAF TF1—4660. Demonstrates the factors governing the choice of an oscillator from a radar system designed to operate in the microwave region. The magnetron and klystron are the two oscillators capable of meeting the requirements. Theory of operation and uses are explained.

Chapter 17

ELECTRONIC TEST EQUIPMENT

MA-7812A Circuit Testing with Meters and Multimeters—Theory. (35 min.—B&W—Sound—Unclassified—1951.) US Army TF11-1666. Describes the basic principles of meters and multimeters and illustrates their use in the operation and maintenance of communications equipment. Largely through the medium of animation photography, the steps in building a meter are delineated. Included are voltmeter, vacuum tube voltmeter, ohmmeter, tube-tester, and wattmeter. The general purpose multimeter is described as a combination of several meters. In radio, telephone, and teletypewriter troubleshooting, the meter is the most important and useful tool available.

MA-7812B Circuit Testing with Meters and Multimeters—Practical Applications. (33 min.—B&W—Sound—Unclassified—1951.) US Army TF11-1667. Meters and multimeters are indispensable in the operation, maintenance and repair of electronic equipment. The equipment described includes: Volt-Ohm-Milliameters, the Wheatstone Bridge, decibel meter and the tube tester. A step by step procedure for using these meters is shown, followed by a description of how they are used in testing transformers, capacitors, resistors, telephone loop circuits, etc. It is emphasized that the technicians should read the operating manuals for this equipment, and keep them handy for the reference purposes. Rigorous safety practices should be followed at all times. The film ends by stressing care of meters to preserve their accuracy.

MN-8687B Reading Multimeter Scales. (6 min.—B&W—Sound—Unclassified—1956.) US Army TF11-2392. Multimeter scales must be read correctly for effective use of the instrument in radio repair. All multimeters have similar scales. Using a typical multimeter, this film demonstrates how to read the scales to measure direct current, DC voltage, AC voltage, and resistance.

MA-8688 Use of Signal Generator AN/URM-25D. (7 min.—B&W—Sound—Unclassified—1957.) US Army TF11-2441. The signal generator is an instrument designed to generate AC signals suitable for test purposes. It is used for trouble shooting or aligning signal communication equipment. This film explains and demonstrates the correct use of Signal Generator AN/URM-25D.

MN 2104B The Cathode Ray Oscilloscope. (23 min.—B&W—Sound—Unclassified—1944.) Explains wide application of the cathode ray in making instantaneous graphs of the wave form of an electric current. Shows use of vertical amplifiers, horizontal amplifier, sweep generator which furnishes the time base cathode ray tube, and power supply. Illustrates the use of the course and fine sweep circuit to synchronize signals so that wave form is stationary.

MN-2104A The Cathode Ray Tube—How it Works. (15 min.—B&W—Sound—Unclassified—1943.) Demonstrates construction and function of each part of the cathode ray tube and how it produces visual images on a screen. Explains electrostatic deflection, electro-magnetic deflection, and how varied currents affect position of the spot of light on the scope.

SN-657 Calibrating the Type LM Crystal Frequency Indicator. (42 frames—B&W—Silent—Unclassified—1942.) Demonstrates calibration of type LM Crystal Frequency Indicator.

SN-658 Making Frequency Measurements with the CFI. (39 frames—B&W—Silent—Unclassified—1942.) Describes steps in measuring radio wave frequency with the LM-CF1.

MA-8862D Radio Set AN/GRC-26—Part 4—Setting up Receiver as Frequency Standard. (15 min.—B&W—Sound—Unclassified—1958.) US Army TF11-2563. This film portrays the step-by-step procedure to set up Receiver "A" as a frequency standard. A demonstration frequency of 4280 kc is used.

MN-8086H Radio Transmitting Sets AN/URT 2,3,4—Troubleshooting. (12 min.—B&W—Sound—Unclassified—1955.) Shows how to determine the general location of trouble in the AN/URT series of transmitters. Describes the use of various lights and meters to locate trouble, the built-in test oscilloscope for testing the operation of the internal oscillators and finally the external test equipment.

MN-1540Q Radio Technician Training—Signal Generator Operation. (9 min.—B&W—Sound—Unclassified—1945.) Use of signal generator in receiver alignment is subject of the picture. Use of modulation switch, adjustments needed to secure various frequencies (for example 456kc and 87MC) and use of alternator switch is shown. Voltmeter gives visual check of adjustments while earphones provide audio alignment check.

MN-1540P Radio Technician Training—Tube Tester Operation. (9 min.—B&W—Sound—Unclassified—1944.) Shows testers designed to check (1) cathode emission, and (2) dynamic mutual conductance of tube. Emphasize use of instruction book supplied with tester and of the tube manual. Testers are practically fool proof. Simply turn index scale to number of tube being tested and follow lines to operate appropriate control of push button.

MN-1540R Radio Technician Training—Audio Oscillator Operation. (9 min.—B&W—Sound—Unclassified—1945.) Explains operation of audio oscillator and demonstrates use. Performance of audio oscillator is shown by checking an amplifier with the audio oscillator and "A" scope. Illustrates and explains all knob turning.

MN-1540S Radio Technician Training—Volt-Ohmmeter Operation. (15 min.—B&W—Sound—Unclassified—1944.) Demonstrates use of various types of volt-ohmmeters, (including the electronic meter) and gives cautions to be followed, such as using the large scale first, (R x 1000; R x 10; R ranges available) and connecting the voltmeter in parallel (Scales; 600, 300, 30, 3 volts).

MA-8369 Calibrating and Tuning Radio Set AN/PRC-10. (9 min.—B&W—Sound—Unclassified—1955.) US Army TF11-2181. Teaches the step-by-step procedure for calibrating and tuning Radio Set AN/PRC-10.

Chapter 18

INTRODUCTION TO COMPUTERS (PART I)

(No applicable training films.)

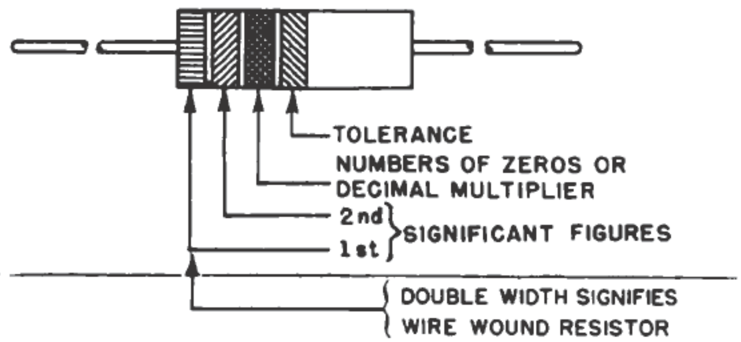
Chapter 19

INTRODUCTION TO COMPUTERS (PART II)

(No applicable training films.)

APPENDIX II

ELECTRONIC COLOR CODING AND SYMBOLS

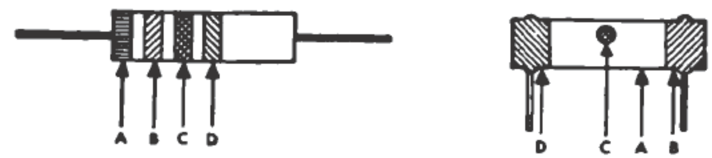


COLOR	SIGNIFICANT FIGURE OR NUMBER OF ZEROS	DECIMAL MULTIPLIER	RESISTANCE TOLERANCE
			PERCENT ±
BLACK	0	---	---
BROWN	1	---	---
RED	2	---	---
ORANGE	3	---	---
YELLOW	4	---	---
GREEN	5	---	---
BLUE	6	---	---
VIOLET	7	---	---
GRAY	8	---	---
WHITE	9	---	---
GOLD	---	0.1	5(J) *
SILVER	---	---	10(K) *
NO COLOR...	---	---	20(M) *

* SYMBOL DESIGNATION ALTERNATE FOR COLOR

20.373

Figure A-1.—Standard resistor color code.

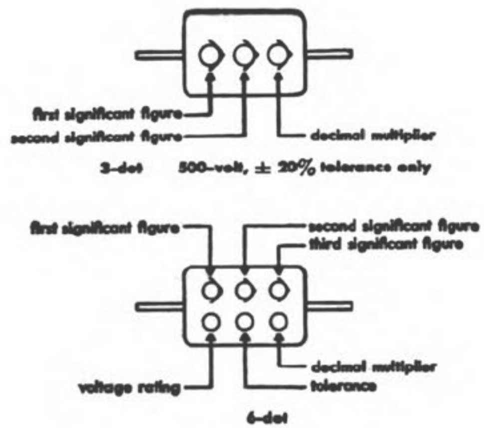


axial leads	color	radial leads
Band A	indicates first significant figure of resistance value in ohms.	Body A
Band B	indicates second significant figure.	End B
Band C	indicates decimal multiplier.	Band C or dot
Band D	if any, indicates tolerance in percent about nominal resistance value. If no color appears in this position, tolerance is 20%.	Band D

Note: Low-power insulated wire-wound resistors have axial leads and are color coded similar to the left-hand figure above except that band A is double width.

20.374

Figure A-2.—Color code for axial- and radial-lead fixed composition resistors.



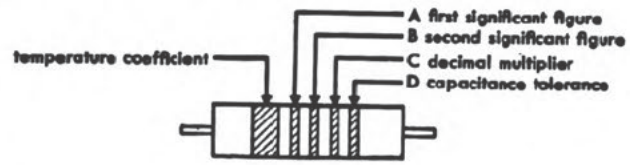
Examples

type	top row left	top row center	top row right	bottom row left	bottom row center (tolerance multiplier)	bottom row right	description
6 dot	red	green	brown	none	none	none	250 μμf ± 20%, 500 volts
	brown	black	black	blue	green	brown	1000 μμf ± 5%, 600 volts
	brown	red	green	gold	red	brown	1250 μμf ± 2%, 1000 volts

20.375

Figure A-3.—3-dot and 6-dot color code for fixed mica capacitors.

color	significant figure	multiplier	capacitance tolerance in % $c > 10\ \mu\mu f$	capacitance tolerance in $\mu\mu f$ $c \lesssim 10\ \mu\mu f$	temperature coefficient parts/million/° C
black	0	1	±20	2.0	0
brown	1	10	±1		−30
red	2	100	±2		−80
orange	3	1,000			−150
yellow	4	—			−220
green	5	—	±5	0.5	−330
blue	6	—			−470
violet	7	—			−750
gray	8	0.01		0.25	+30
white	9	0.1	±10	1.0	−330 ±500



Examples

wide band	narrow bands or dots A	B	C	D	description
black	black	red	black	black	2.0 μμf ± 2 μμf, zero temp coeff
blue	red	red	black	green	22 μμf ±5%, −470 ppm/° C temp coeff
violet	gray	red	brown	silver	820 μμf ±10%, −750 ppm/° C temp coeff

20.377

Figure A-5.—Color code for fixed ceramic capacitors.

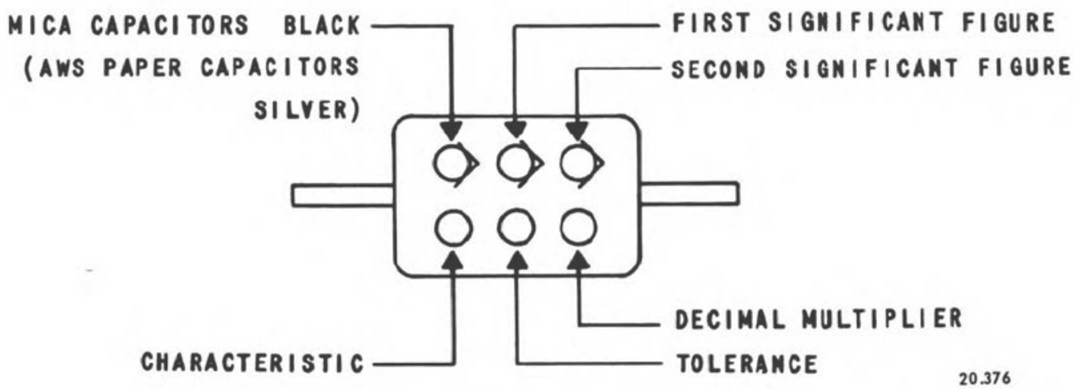


Figure A-4.—Alternate method color code for fixed mica capacitors.

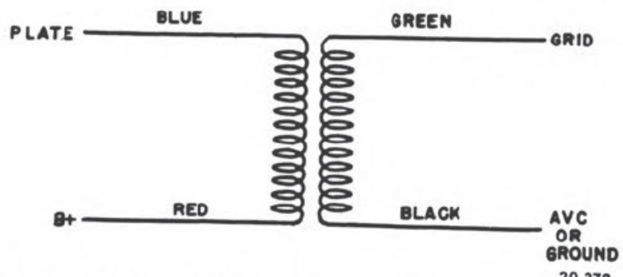


Figure A-6.—Color code for i-f transformers.

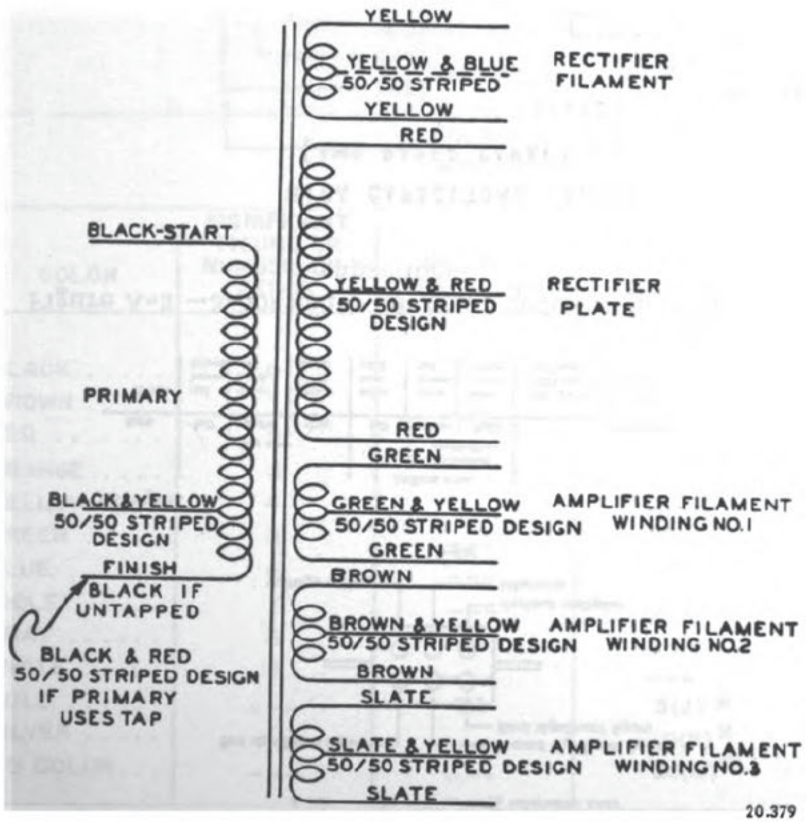


Figure A-7.—Color code for power transformers.

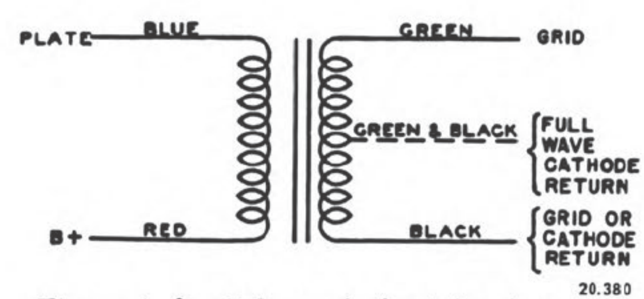


Figure A-8.—Color code for interstage audio transformers.

Standard colors used in chassis wiring for the purpose of circuit identification of the equipment are as follows:

Circuit	Color
Grounds, grounded elements, and returns.	Black.
Heaters or filaments, off ground	Brown.
Power supply B plus	Red.
Screen grids.	Orange.
Cathodes	Yellow.
Control grids	Green.
Plates	Blue.
Power supply, minus	Violet (purple).
A-c power lines	Gray.
Miscellaneous, above or below ground returns, a-v-c, etc. . .	White.

For other electrical and electronic symbols refer to Military Standard, MIL-STD-15A.

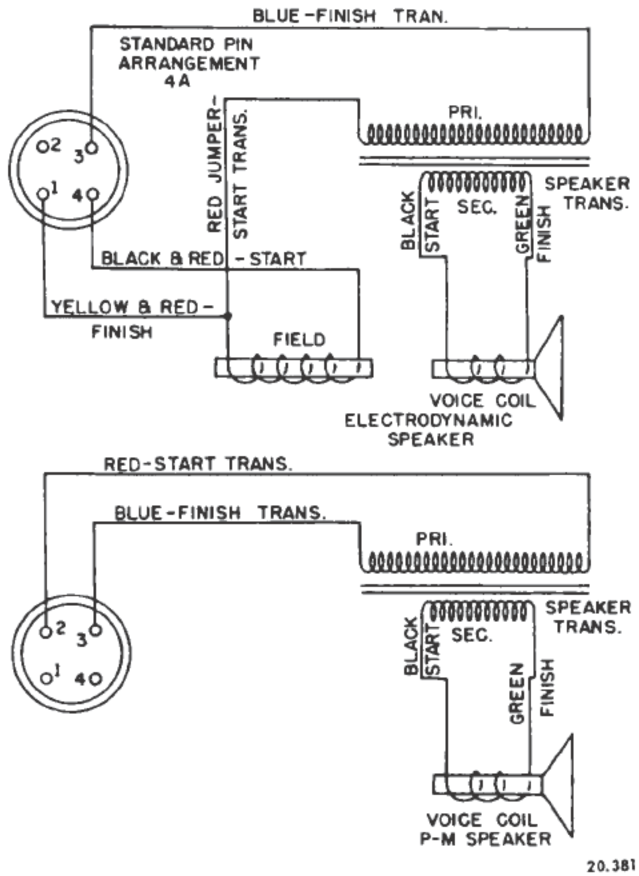


Figure A-9.—Standard pin arrangement and color code for electrodynamic and P.M. speakers.

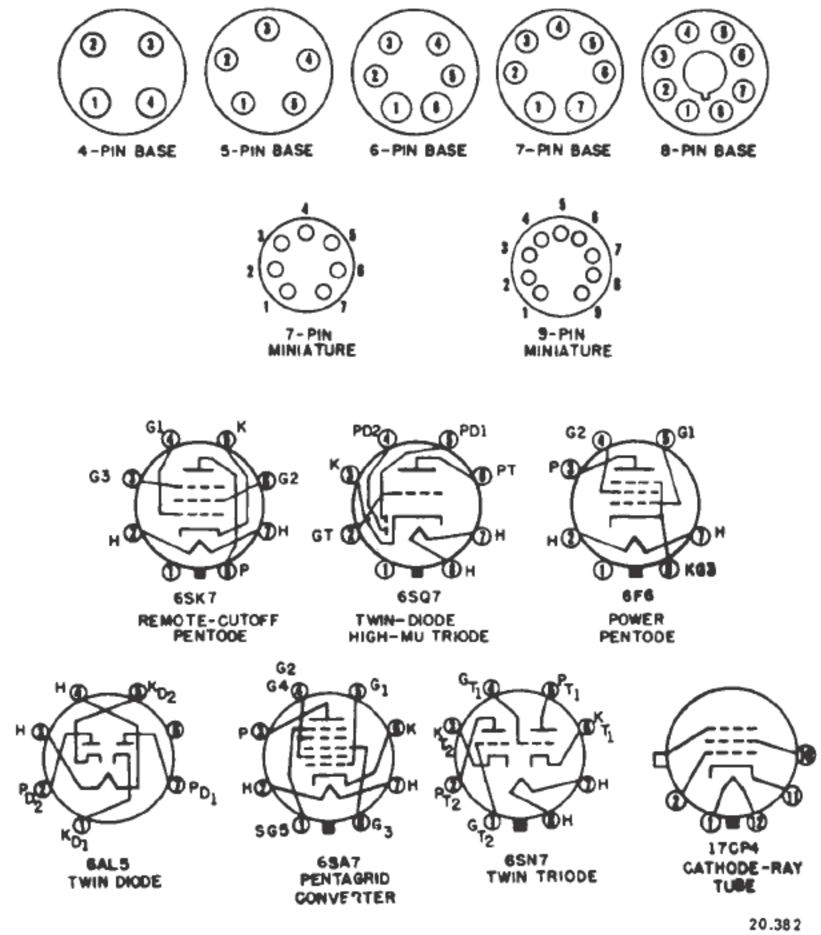


Figure A-10.—Common electron-tube bases showing arrangement of pins as viewed from the bottom.

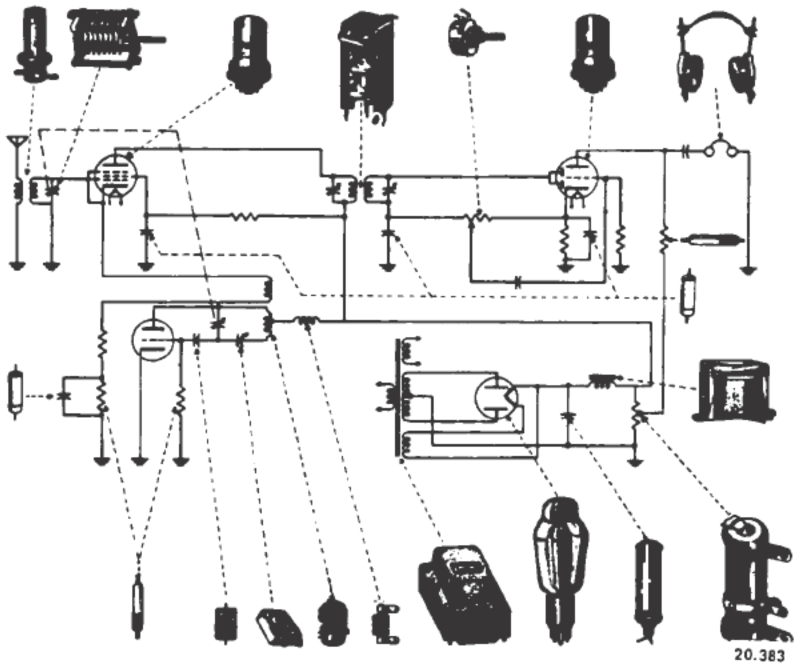


Figure A-11.—Commonly used circuit components and their corresponding symbols.

Numerical Magnitude =	Number of 10 factors	Prefix	Symbol
1 000 000 000 000 =	10^{12}	tera	T
1 000 000 000 =	10^{9}	giga	G
1 000 000 =	10^{6}	mega	M
1 000 =	10^{3}	kilo	k
100 =	10^{2}	hecto	h
10 =	10	deka	dk
0.1 =	10^{-1}	deci	d
0.01 =	10^{-2}	centi	c
0.001 =	10^{-3}	milli	m
0.000 001 =	10^{-6}	micro	u or μ
0.000 000 001 =	10^{-9}	nano	n
0.000 000 000 001 =	10^{-12}	pico	p

Figure A-12.—Table of numerical magnitudes used with electronic units.

INDEX