TELEVISION

2nd EDITION

V. K. ZWORYKIN
G. A. MORTON

TELEVISION

2nd EDITION

The Electronics of Image Transmission in Color and Monochrome

V. K. ZWORYKIN, E.E., Ph.D.
Director Electronics Research
RCA Laboratories

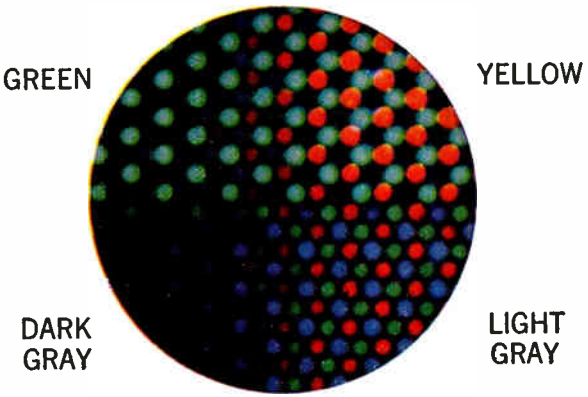
G. A. MORTON, Ph.D.
RCA Laboratories
Princeton, N. J.

JOHN WILEY & SONS, INC., NEW YORK
CHAPMAN & HALL, LIMITED, LONDON

Library of Congress Catalog Card Number: 54–11162

Printed in the United States of America

(a)



(b)

Photographs of Shadow Mask Tricolor Kinescope Screen. (a) Picture reproduced from slide scanner on early experimental tube (1951). (b) Enlargement of junctions between green, yellow, and gray fields.

PREFACE

During the fourteen years following the first edition of this book television has grown from a restricted experimental field to a nation-wide major industry. This tremendous growth has been accomplished by a parallel development in the technology of television. Although most of the fundamental principles which form the basis of present-day picture transmission and reception (with the exception of color) were outlined in the first edition, the engineering advances have been so profound as to make necessary the revision of every chapter and the complete rewriting of many. New chapters have been added discussing color and industrial television.

The organization of material in the second edition follows closely that of the first. Part 1 discusses the basic field of physics important to television, including newly added sections dealing with semiconductors, photoconductors, and recent developments in photoemitters and phosphors. The fundamentals of television are covered in Part 2, and a background of the development of television leading up to the present system is presented there. In Part 3 the components making up the television system in use today are described in detail and then integrated into the complete receiver and transmitter. Part 3 starts with a discussion of the pickup tube problem and the camera tubes, such as the image orthicon and Vidicon, which have been developed as a solution to this problem. The kinescope is next considered in detail. This section is followed by a consideration of some of the associated components, including the video amplifier, the scanning generator, and the synchronizing system. Finally, the complete television transmitter and receiver are covered in as much detail as is practicable. Part 4 presents some of the newest aspects of television. The problem of color television is treated both from the standpoint of the fundamental principles involved and its practical aspects in Chapters 17, 18, and 19. The specialized features of industrial television are presented in Chapter 20. The final chapter deals with some of the practical aspects of television broadcasting, including studio arrangement, program pres-

entation, the television station, and the integration of stations into a network by means of coaxial cables and radio relay links.

Even at the time of writing the first edition, when television was in the first stages of its development, it was impossible to cover the field completely in a single volume. Today, with the enormous growth of the art, adequate coverage is even more difficult. In preparing this edition the authors have attempted to emphasize those aspects of the field which are most characteristic of television. For that reason the camera tube and the viewing tube, together with both the physics background and the television fundamentals associated with them, have been singled out for greatest emphasis. Specific aspects of circuitry relating to the television problem, such as the video amplifier, are considered in detail. Coverage of the theoretical and engineering features of the entire television system is included, but extensive collateral reading in general radio engineering, circuit theory, and vacuum tube practice is assumed for a detailed understanding of the whole field of television.

The authors wish to take this opportunity to express their indebtedness to Dr. E. G. Ramberg. Without his cooperation this revision would not have been possible. He is responsible for much of the new subject matter presented, and, in addition, his assistance in the general preparation and editing of material has been invaluable. Chapters 18 and 19 on the problems of color television are exclusively his work.

The generous cooperation of many other staff members of the Radio Corporation of America Laboratories has also been of great assistance in this work. Specific mention should be made of Dr. G. L. Frendendall in connection with the television receiver, Drs. G. H. Brown, D. W. Epstein, R. D. Kell, and W. C. Morrison for aid with the chapters on the kinescope, the transmitter, and color television. Finally, the authors wish to acknowledge the assistance given them by the technical staff of WNBT and the courtesy of the RCA Tube Department, the *RCA Review*, *Broadcast News*, and other RCA sources in making available illustrative material.

V. K. Zworykin
G. A. Morton

Princeton, New Jersey
June, 1954

CONTENTS

Part 1. Fundamental Physical Principles

Part 2. Principles of Television

Part 4. Color Television, Industrial Television, and Television Systems

PART 1

Fundamental Physical Principles

1 Physical Electronics

1.1 Electron Physics. Television systems, in addition to employing ordinary metallic electron conduction, require as fundamental to their operation other electronically active solids. These solids include materials which exhibit the properties of photoelectric emission, photoconductivity, thermionic emission, semiconduction, secondary emission, and fluorescence (the last forming the subject of the next chapter). The present chapter discusses very briefly some of the electron physics which is the basis of the phenomena.

In some respects, electricity is not a modern discovery since there are recorded observations on static electricity which date back to antiquity. However, it was not until the eighteenth century that electrical conduction and electric current, which form the basis of its present usefulness, became known and recognized. During the nineteenth century, rapid strides were made by such men as Maxwell, Faraday, and Hertz toward unraveling the laws of electromagnetic phenomena. The last two decades of that century were particularly fruitful, culminating in the discovery of the electron by J. J. Thomson in 1897. Before the discovery of the electron itself, many observations had been made on the special electronic phenomena mentioned above. The interaction between light and electricity was noted during the first half of the nineteenth century. Early observations on photoemission, that is, the emission of electrons under the action of light, date back to observations made by Hertz in 1887 on the fact that electrical discharges were facilitated if the negative electrode of the spark gap was irradiated with ultraviolet light. The following year, Hallwachs undertook a systematic study of the effect observed by Hertz and came to the conclusion that negative electricity leaves a body which is illuminated by ultraviolet light. The work was forwarded by Elster and Geitel, who gathered data on the rate at which charge left different metals and on the relationship between the wavelength of the incident radiation and the rate of transfer of electricity. With the discovery of the electron, the nature of the phenomena was disclosed. Again, in 1905, the theory

of photoemission was advanced another major step when Einstein applied the then new quantum theory put forward by Planck to the problem of the photoelectric long wavelength threshold. Since then, the theory has advanced steadily, keeping pace with an increasing knowledge of the quantum mechanics of metallic and semiconduction.

The development of photoconductivity followed a nearly parallel course. The increase in conductivity of selenium upon exposure to light was discovered in the year 1873 by Willoughby Smith. It was recognized as related to the phenomena of semiconductivity and insulation rather than the phenomenon of metallic conduction, long before the mechanism of these different forms of conductivity was understood. It was not until quantum mechanics was very well developed that a really satisfactory model of photoconductors and semiconductors could be devised. Important contributions by Bloch, Mott, Fröhlich, and many others have led to a model which gives a fairly quantitative explanation of the experimental observations of both photoconductors and semiconductors.

Although the effect of temperature on the rate of discharge of electrified bodies has been known to science for a long time, thermionic emission cannot be said to have been discovered until 1883. In this year, Edison discovered that a negative current could be made to flow from an incandescent filament in an evacuated bulb. By the beginning of the twentieth century, a considerable amount of data had been collected pertaining to this phenomenon. In 1905, O. W. Richardson derived, with the aid of classical thermodynamics, an equation which described the observations very accurately. Later, the equation derived by Richardson was revised and reinterpreted by Dushman, Langmuir, and others, but the original exponential form was retained. With the development of quantum mechanics, the understanding of the effect was again increased greatly. Nordheim and others, with the aid of the Sommerfeld model of metals, were able to show that thermionic emission is closely related to the phenomena of conduction, photoemission, thermoelectricity, etc. This discovery, together with technical advances in making thermionic cathodes, has brought the field to its present status.

Secondary emission, that is, the emission of electrons from a circuit which is bombarded by high-velocity electrons, did not enter the field until somewhat later. The phenomenon became known by 1900 as the result of experimental work by Starke, Lenard, and others. Until about 1920, very little was done in investigating this aspect of electronics. During the last four decades, however, its practical impor-

tance has become considerably greater, and this has resulted in an impetus to the investigation of this type of emission. Important analytical work done by Fröhlich, Wooldridge, and others, using the quantum-mechanical model of electronically active solids, has led to theoretical interpretations which quantitatively account for many of the experimental observations of the phenomena of secondary emission. However, many of the complicated surfaces in use in devices which employ secondary emission as the basis of their operation are still beyond the range of theoretical treatment.

In the succeeding sections, the electronic phenomena mentioned above will be discussed without reference to the historical order of their development. For a complete history of electronically active solids, the reader is referred to standard textbooks dealing with the history of electricity.

1.2 Structure of Metals. Reading across the Periodic Table of elements from right to left, one finds that the elements become increasingly electropositive. That is, the elements in the columns on the left have an electron configuration such that they can easily lose one or more of these electrons and become positive ions. In the solid state these electropositive elements form the group of substances called metals. This group is characterized by the fact that its members are excellent conductors of both heat and electricity. Furthermore, the metals are in general both malleable and ductile, although there are some exceptions. It should be noted that these exceptions occur in the center columns of the table, and are all elements which are only slightly electropositive; in other words, these exceptions might be termed "poor metals."

X-ray analysis of the metals shows them to be crystalline. The crystal structure of these substances (at least for all the good metals) is extremely simple. All the elements of the first column of the periodic table, together with tungsten, iron, molybdenum, and many others as well, have a body-centered cubic structure. The very malleable and ductile metals such as gold, silver, copper, and aluminum are face-centered cubic; others such as zinc, magnesium, and beryllium are hexagonal close-packed.

The atoms making up the metal lattice no longer retain the loosely bound electrons associated with them in their free state. The electrons released when a metal crystal is formed do not have definite positions in the lattice, but exist rather as an electron gas which is free to move through the structure.

This theory, which assumes that the valence electrons form an electron gas, was proposed as early as 1905 by H. A. Lorentz. He assumed that the electrons which form this gas are in thermal equilibrium with the atoms of the lattice, and that they have a Maxwellian energy distribution. On the basis of this hypothesis, in order that the electrons stay within the metal, it is necessary to postulate a potential barrier at the boundaries of the metal. At a large distance from the metal, this potential is assumed to be due to the image force of the electron, but, since the ideal image force rises to infinity at the surface of the metal, it is necessary to assume a departure from this law near the boundary.

The Lorentz hypothesis explained fairly satisfactorily electrical conduction, thermionic emission, and a number of other observed phenomena, but required that the specific heat of a metal be nearly twice the observed value.

It was not until the advent of quantum mechanics and the exclusion principle that the reason for this discrepancy became evident. With the aid of the new mechanics, Sommerfeld, Pauli, and others were able to modify the Lorentz theory in such a way that it gave a very accurate quantitative account of the behavior of metals.

The difference between the old Lorentz theory and the Sommerfeld theory lies in the type of energy distribution assumed for the electron. In either case the energy E of an electron is

$$E = V_0 + \frac{p^2}{2m} \tag{1.1}$$

where V_0 is its potential energy and p its momentum. Lorentz placed no restrictions upon the values which p could assume. The newer theory, on the other hand, assumes, in accordance with quantum laws, that the three momentum components are restricted and, for a block of metal with the dimensions L_x, L_y, and L_z, can only have values satisfying the relation

$$p_x = \frac{n_x h}{L_x}, \qquad p_y = \frac{n_y h}{L_y}, \qquad p_z = \frac{n_z h}{L_z} \tag{1.2}$$

where h is Planck's constant, and n_x, n_y, and n_z are integers. The allowed energies are therefore not continuous but consist of a vast number of discrete levels. A further restriction is imposed by the exclusion principle. This rule states that only two electrons can exist in any one energy level, two electrons being allowed because of the two possible spin directions.

The effect of these two restrictions becomes clearly apparent upon consideration of what happens when the crystal is cooled to absolute zero. In the older theory, all the electrons come to rest, since they are in thermal equilibrium with the atoms making up the crystal lattice. However, this cannot happen according to the newer theory because there can be only two electrons in a state of zero energy. The remaining electrons must retain enough energy to be able to occupy other energy levels. As a result, all the lower levels are filled up, so that energies corresponding to several electron volts exist in a crystal even at zero absolute.

If, instead of being at zero absolute, the crystal is at room temperature, the mean energy of its atoms is still far below the energy held by the electrons in the highest filled level. Consequently, the probability of an electron receiving enough energy from the thermal motion of the atoms making up the crystal to raise it to an unfilled energy level is extremely small. The electrons, therefore, because they can be given almost no thermal energy by the lattice, contribute very little to the heat capacity of a metal. The greatest difficulty encountered by the older theory is thus overcome.

At very high temperature, the mean energy of the atoms making up the lattice may become greater than the energy of the highest filled level. Under these circumstances, the electrons receive thermal energy and come into equilibrium with the lattice, their energy distribution approximating the Maxwellian distribution postulated by the earlier theory. An abnormally high specific heat would exist at such temperatures. A calculation of the temperature required to increase the specific heat appreciably shows it to be beyond the vaporizing temperature of any known metal.

A more quantitative consideration, starting with the potential barrier, is necessary if these ideas are to be of practical value. The potential energy, which keeps the electrons within the material, increases with the distance from the surface, asymptotically approaching some arbitrarily chosen zero at infinity. As has been mentioned, at distances from the crystal greater than the lattice spacing this potential is the result of the simple image force. This force, as is well known, is given by the relation $F = -e^2/4x^2$, where x is the distance from the surface. As the charge nears the surface, the forces rise and would approach infinity as x approaches zero if the law were applicable at all distances. However, from the very nature of its derivation, which assumes a plane surface, it is obvious that this law will not hold for distances of less than a few atom diameters. The law which most

nearly satisfies the observed facts assumes that the attractive force is zero at the metal surface, rises rapidly to a maximum, then decreases inversely with the square of the distance from the metal surface. The force law close to the surface is extremely complicated, but for the present purpose a knowledge of its exact form is unnecessary. Figure 1.1 shows the approximate shape of the force curve.

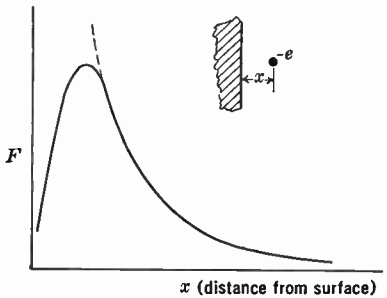


Fig. 1.1. Modified Image Forces.

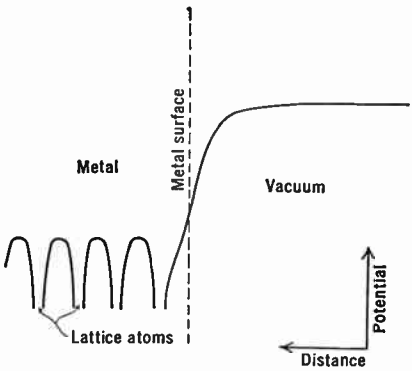


Fig. 1.2. Idealized Potential Field.

The potential at any point outside the metal is calculated by integrating the force curve. The form of the potential curve is illustrated in Fig. 1.2. Inside the metal, the potential is not uniform but consists of a three-dimensional array of potential minima, the minima corresponding to the crystal atoms. These minima are shown to the left

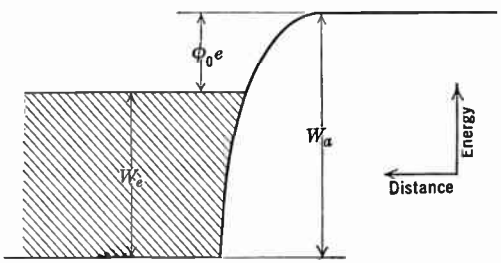


Fig. 1.3. Electron Energies in an Ideal Metal.

of the metal surface in Fig. 1.2. For purposes of this discussion, the non-uniformity of the potential within the metal can be ignored and the variable potential can be replaced by a uniform averaged potential as shown in Fig. 1.3. The value of the potential rise at the surface of a metal is extremely important in determining its emission properties. This potential, known as the "inner potential," is shown as W_a on the

diagram. It can be measured experimentally by the change in wavelength of electron waves passing through the metal surface observed in electron diffraction experiments. The values obtained by this method are in the neighborhood of 10 to 15 volts.

With these values for the inner potential, it is possible to calculate the approximate minimum distance at which the square law image force on an electron applies. This distance turns out to be of the same order of magnitude as the distance between atoms in the metal lattice.

For convenience of illustration, the kinetic energies of the electrons can be included in an energy diagram, together with the potential energy. The kinetic energy is plotted along the ordinate just as is the potential energy of the lattice. Such a diagram is shown in Fig. 1.3. The energy distribution in the electron gas, based upon the exclusion principle, was derived by both Fermi and Dirac, independently, in 1926. According to their distribution function, the number of electrons $N(u, v, w)\, du\, dv\, dw$ having velocity components u, v, w, in the range du, dv, dw is

$$N(u, v, w)\, du\, dv\, dw = \frac{\dfrac{2m^3}{h^3}\, du\, dv\, dw}{\epsilon^{(W_e-\mu)/(kT)} + 1} \tag{1.3}$$

where

$$\mu = \frac{h^2}{2m}\left(\frac{3n}{8\pi}\right)^{2/3} \tag{1.4}$$

k is the Boltzmann constant, and
n is the number of electrons per unit volume.

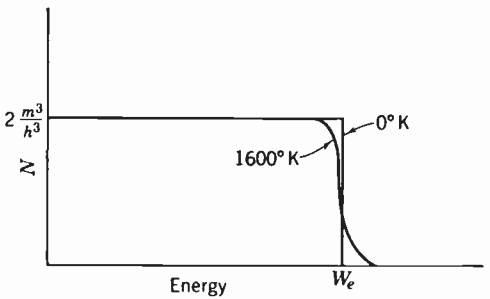


Fig. 1.4. Distribution of Electron Energies by Fermi-Dirac Statistics.

A plot of this function is shown in Fig. 1.4. Curves are given for $T = 0$ and $T = 1600°$. At $T = 0$, this function becomes

$$N(u, v, w) = 2\frac{m^3}{h^3} \quad \text{for } W_e < \mu$$

$$N(u, v, w) = 0 \qquad \text{for } W_e > \mu$$

Thus there are no electrons in energy levels above $W_e = \mu$, which is in accordance with the qualitative picture given in the preceding paragraphs.

The numerical value for $W_{e\,\max}$ can be found from the equation

$$W_{e\,\max} = \frac{h^2}{2m}\left(\frac{3n}{8\pi}\right)^{2/3} \tag{1.5}$$

However, to evaluate the expression, it is necessary to assign a value to n. This requires a knowledge of the number of atoms per unit volume and the number of electrons each atom contributes to the electron gas. If the former is calculated from the density, composition, and Avogadro's number, and a reasonable assumption is made for the latter, $W_{e\,\max}$ is found to lie between 5 and 10 volts.

At temperatures other than zero, the velocity distribution does not end sharply at the value $W_{e\,\max}$ but decreases asymptotically to zero. As the temperature is increased, the high-velocity electrons become more and more numerous. Finally, at very high temperatures the Fermi-Dirac distribution approaches a Maxwellian distribution. However, this is found, as has been mentioned, to be a higher temperature than can be reached using any known metal.

When a potential difference is established between two points in a metal, electrical conduction takes place because the field thus produced causes a drift to be superimposed upon the random motion of the electrons in the upper levels. This drift is in a direction opposite to that of the field because of the negative charge of the electron. In other words, the drift motion is opposite to the conventional direction of current flow. Although the electrons behave largely as though they were free, there is some interaction between them and the crystal lattice, which appears as the heat generated due to electrical resistance. In order to account quantitatively for electrical resistance of a metal, it is necessary to deal with the wave functions of the electrons and to consider resistance as a scattering and diffraction problem. Looked at in this way, both resistance and change of resistance with temperature can be accounted for very nicely.

1.3 Semiconductors. The metals are to be found in the left-hand column of the Periodic Table. As one moves toward the center of the

Table, the elements found there lose their valence electrons less easily when in solid form and the materials are not good conductors. In the extreme right-hand column lie the elements which have very firmly bound outer electrons and, therefore, go to make up the class of materials known as insulators. Silicon, germanium, carbon, and arsenic in the center of the Periodic Table are examples of semiconductors, whereas sulphur, iodine, and bromine are examples of good insulators. However, any model which describes semiconductors in terms of electrons which are moderately bound to their parent atoms and insulators in terms of crystal lattices with strongly bound electrons is not adequate to account for the essential behavior of these two classes of materials. As in the case of metals, the concepts of quantum mechanics must be introduced to give a quantitative interpretation of observed phenomena. A single atom, according to the quantum-mechanical model, has its extranuclear electrons in definite energy states or levels. The number of the extranuclear electrons depends upon the positive charge of the nucleus of the atom in question. Although the electrons about the nucleus of an atom cannot have their positions specified at any instant of time, they can, nevertheless, be thought of as moving in some form of orbit around the nucleus with the total energy of each specified and their position given by a probability function. Thus, a sodium atom has eleven extranuclear electrons. Of these, two are in a 1*s* energy state, two in a 2*s*-state, six in a 2*p*-state, and one is in a 3*s*-state. The energies of these states (with the exception of the lowest, 1*s*-, state), as well as those of discrete higher energy levels into which the electrons can be excited, are shown in Fig. 1.5. Electrons cannot, however, exist in the atom except in these discrete energy levels. In other words, energy transitions can occur only between these levels. It is this, of course, which gives rise to the spectral lines observed in light, ultraviolet radiation, and x-rays. If two atoms are brought close together, each energy level splits into two levels very close together. Again, if three atoms are brought together, each level splits into three; and, if some large number n atoms join together, as is the case when they form the crystal lattice of the

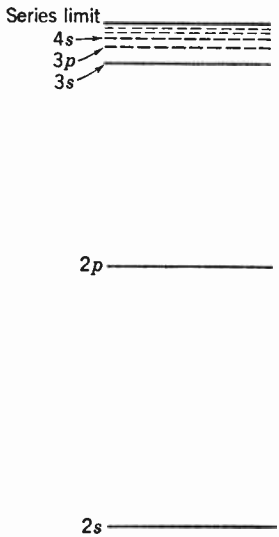


Fig. 1.5. Electron Energy Levels of a Sodium Atom.

element in question, each level divides into n very closely spaced levels. The spacing between each of the n subdivisions is so small that the original level may be thought of as giving rise to a continuous energy band rather than to a multiplicity of discrete levels. Thus a sodium crystal will have energy bands as pictured in Fig. 1.6.

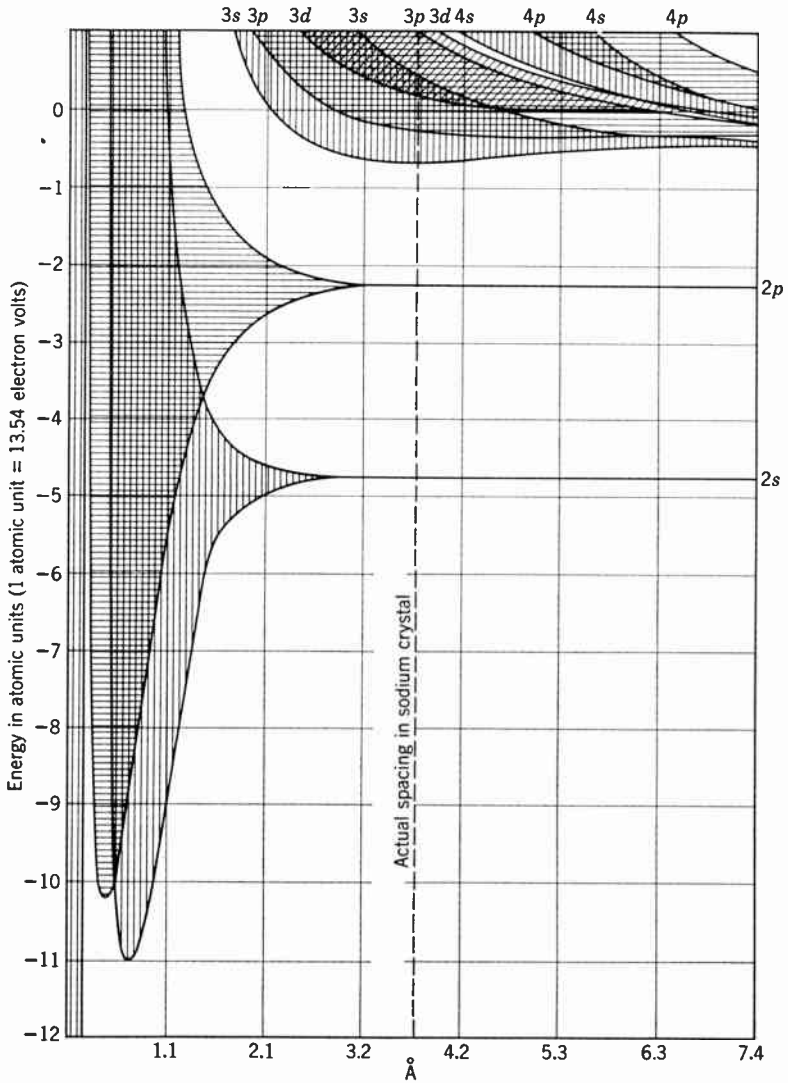


Fig. 1.6. Band Structure as Function of Atomic Spacing for Sodium. (After Slater, reference 15; courtesy of *Reviews of Modern Physics*.)

One narrow triple band which results from the splitting up of the occupied energy levels of the sodium atom is completely filled by the $6n$ $2p$-electrons belonging to the n atoms. Therefore, it would, for example, be impossible to give a small amount of additional energy to an electron in the $2p$-band without taking it away from another electron in the same band. The minimum amount of energy that could be accepted by any electron in this band without such compensation would be that required to lift it to the next vacant band. However, it will be observed that the next band is only half filled by the n $3s$-electrons. Therefore, small amounts of energy can be added to the electrons in this band in such fashion as to give the group as a whole a drift velocity in the direction of the applied field. The ability of the electron to accept small amounts of additional energy so that it can drift through the material and serve as a current carrier is the property which gives the material its metallic characteristics.

If a similar examination is made of the element sulphur, it will be found that all the bands containing electrons are completely filled. The minimum amount of energy which an electron in sulphur must accept to escape from a filled band is 5 electron volts. The material, therefore, is a good insulator since the application of an electric field across the material cannot modify the velocity distribution of the electrons so that they may serve as current carriers.

Examined in this same way, germanium is found to have all the bands occupied by electrons completely filled. Adjoining the topmost filled band there is, however, an empty band separated from it by an energy gap of only about 0.7 electron volt. This energy gap is small enough so that the thermal energy of the lattice at room temperature is sufficient to supply a large number of electrons in the filled band with sufficient energy to be excited into the conduction band adjoining it. The material, therefore, at room temperature will be an electrical conductor with an electrical specific resistivity of 60 ohm-centimeters. By way of comparison, the resistivity of a good metal is lower by 6 or 7 orders of magnitude. Germanium is, accordingly, classified as a semiconductor.

In considering the electrical properties of semiconductors, it is actually necessary to consider only the filled band, the empty or conduction band adjoining it, and the forbidden band which separates the two, as shown in Fig. 1.7. In other words, the presence of electrons in the conduction band or absence of them in the filled bands alone can contribute to the conductivity of the material. As has already been pointed out, if an electron is excited, for example, by the thermal energy

of the crystal lattice, into the conduction band, it can move under the influence of an electrostatic field, thus causing the material to be a conductor. However, if the electron is excited from a filled band, an unfilled energy level will be left in the filled band. Now a neighboring electron can move into the level vacated by the excited electron. This, in turn, leaves another vacant level which can be occupied by a second electron so that in effect the vacancy moves through the material in much the same way as an electron in the conduction band. Indeed, an analysis of the behavior of such a vacancy shows that its movement is identical with the motion of a charged particle having about the mass of an electron but carrying a positive charge in an empty conduction band. Such a vacancy, termed a "hole," is capable of serving as a current carrier just as is an excited electron.

Semiconductors whose conductivity is due to electron and hole current resulting from the excitation of electrons from the filled band to the conduction band are known as intrinsic semiconductors. Frequently, when a semiconductor or insulator crystallizes, imperfections may occur in its lattice, due either to foreign atom impurities or to lattice defects. These imperfections can introduce additional energy levels in the forbidden gap. If such a level contains an electron, the electron can be excited into the conduction band more easily than can an electron from the filled band because of the small energy difference between the conduction band and the impurity level. Such a transition is indicated by transition B shown in Fig. 1.7. An impurity of this type is called a donor level, and a semiconductor whose conductivity depends on the excitation of carriers from donor levels is known as an n-type semiconductor. Germanium, which has 4 valence electrons, crystallizes to form a semiconductor. If arsenic, with 5 valence electrons, is introduced as an impurity, the result is an n-type semiconductor. Only a minute amount of impurity leads to a marked change in the electrical characteristics of the material. For example, one part in ten million of arsenic changes the specific resistivity of germanium from 60 ohm-centimeters for the intrinsic pure material to less than 1 ohm-centimeter for the resulting n-type semiconductor. It also completely alters the variation of conductivity with temperature.

In an exactly analogous manner, impurities may introduce levels in the energy gap which are normally empty. An electron from the filled band may be excited into such an "acceptor" level, leaving a mobile hole at its vacated site. A semiconductor whose conductivity depends upon hole-type charge carriers created by electrons from the filled band being excited into vacant or acceptor levels is known as a p-type semi-

conductor. Again, starting with germanium as the semiconductor, the impurity gallium which has 3 valence electrons causes the material to become a p-type semiconductor. Like amounts of the impurity gallium or the impurity arsenic in germanium will form a semiconductor which has very similar electrical properties except that the former will be p-type and the latter n-type.

The conductivity of a semiconductor depends upon the number of current carriers free to move and upon the mobility of the current car-

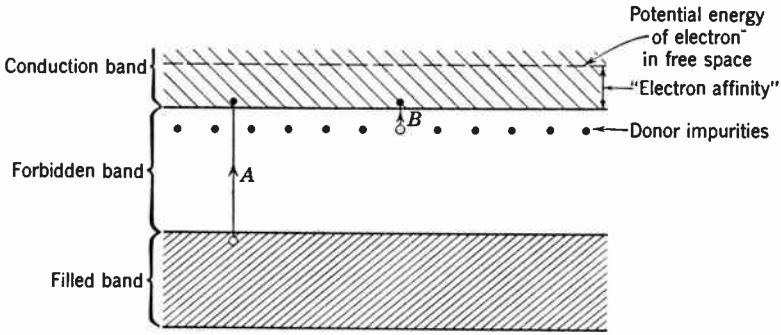


Fig. 1.7. Band Structure of a Typical Semiconductor, Showing the Conduction Band, the Filled Band, and the Forbidden Gap. A represents the excitation of an electron from the filled band, B that from a donor impurity.

riers. An electron excited into the conduction band is in constant motion due to thermal agitation, moving a short distance through the lattice, then "colliding" with a lattice irregularity in such a way as to change its direction entirely, and moving in the new direction only to make another collision. The average number of collisions made by an electron at room temperature is very large, being in the neighborhood of 10^{12} per second. If an electric field is introduced through the crystal, a drift is superimposed upon the random motion of the current carrier. Since the electron cannot experience a continuous buildup of velocity from acceleration due to the frequent collisions, its average velocity varies linearly with the applied field. The factor of proportionality between the average velocity of a current carrier and the electric field is called "mobility." It is usually given in terms of centimeter per second per volt per centimeter applied field.

Quantitatively, therefore, the conductivity of any semiconductor is given by

$$\sigma = e(n_e\mu_e + n_h\mu_h) \tag{1.6}$$

where e is the charge of an electron; n_e the number of electrons per unit volume in the conduction band; n_h the number of holes in the filled band; μ_e the mobility of an electron; and μ_h the mobility of a hole.

Both the number of current carriers and the mobility are functions of temperature. The variation of mobility with temperature is given approximately by

$$\frac{1}{\mu} = aT^{-3/2} + bT^{+3/2} \tag{1.7}$$

The first term of this expression describes the effect of impurity scattering upon mobility. At low temperatures, this term will normally dominate and leads to a mobility which decreases with decreasing temperature. At higher temperatures, the second term, with a positive exponent, dominates and leads to a decrease in mobility with rising temperature.

Thermal scattering by the crystal lattice is responsible for this term. Typical mobility curves for germanium of various purities are shown in Fig. 1.8. It will be observed that the mobility first increases as the temperature falls and then decreases. The initial increase is independent of the impurity content of the material, but the decrease at low temperatures is a direct function of the impurity content.

The quantitative determination of the number of electrons and hole carriers as a function of temperature requires the consideration of the position of the Fermi-level which describes the statistical distribution of thermal energy among the electrons of the material. For an intrinsic semiconductor, the number of carriers as a function of temperature is given by

$$n_e = n_h = 2\left(\frac{2\pi mkT}{h^2}\right)^{3/2} \epsilon^{-E_g/(2kT)} \tag{1.8}$$

where h is Planck's constant, m the mass of the electron, k Boltzmann's constant, and E_g the energy difference between the top of the filled band and the bottom of the conduction band. In this expression, the factor in brackets gives the effective number of electron states at the bottom of the conduction band. The factor 2 before this parenthesis is due to the fact that each state may be occupied by 2 electrons with opposite spins. Where impurity levels are involved, the expression for the temperature dependence is much more complex. However, the requirement of electrical neutrality of the sample fixes the position of the Fermi level for any given temperature and impurity level distribu-

tion. The number of electron and hole carriers is then given by the expression in Eq. 1.8, provided that $E_g/2$ in the exponent is replaced by the separation of the Fermi level from the lower boundary of the conduction band and the upper boundary of the valence band, respectively.

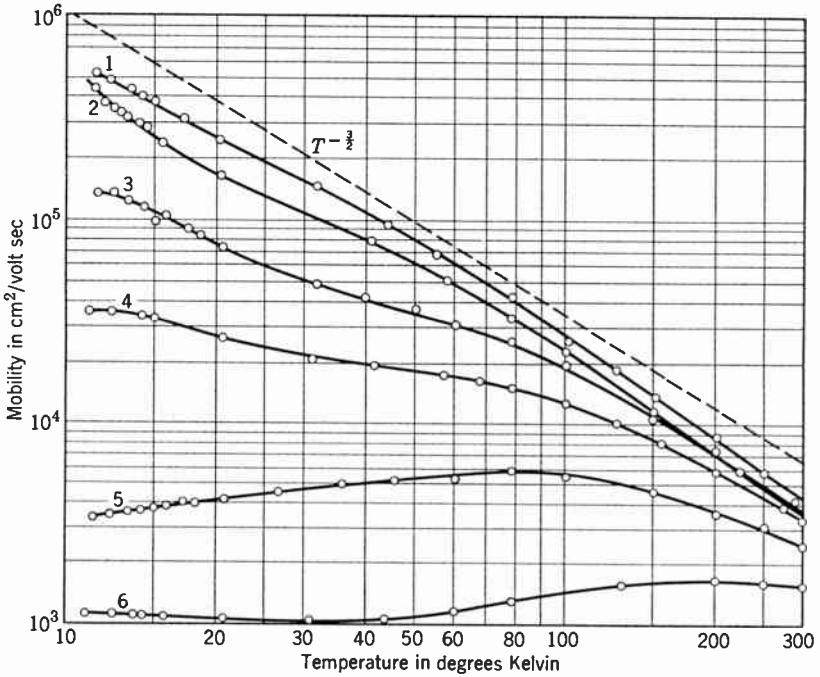


Fig. 1.8. Mobility as a Function of Temperature for Germanium (n-Type Samples 1–6, with Increasing Impurity Content). (After Debye in Conwell, reference 17; courtesy of *Proceedings of the Institute of Radio Engineers.*)

The results of measurements of the carrier concentration for the germanium samples of Fig. 1.8 are represented in Fig. 1.9, showing the logarithm of the carrier concentration as function of the reciprocal absolute temperature. An examination of the curve shows that three temperature ranges may be distinguished. At high temperatures, the number of carriers decreases with falling temperature exponentially with a slope related to the intrinsic energy gap E_g. In this temperature range, the conductivity is primarily due to carriers excited across the forbidden band to the conduction band. In an intermediate temperature range, the thermal power is insufficient to excite electrons across the intrinsic gap. However, in this range, the temperature may be high

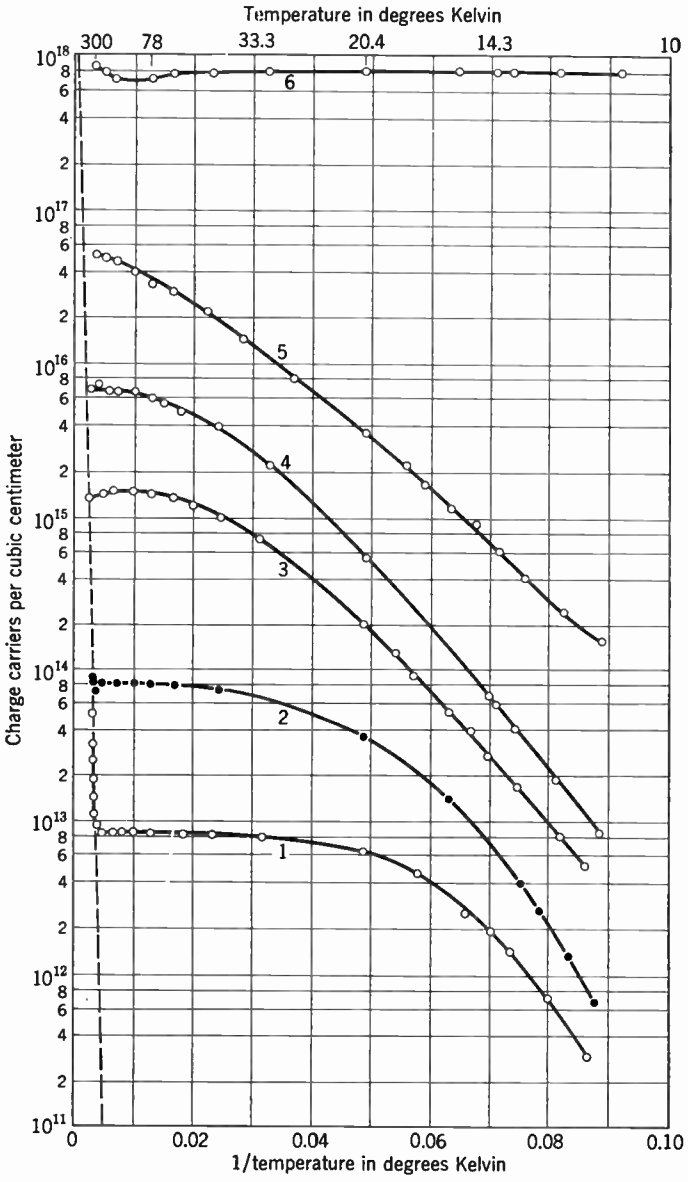


Fig. 1.9. Carrier Concentration as a Function of Temperature for *n*-Type Germanium. (After Debye in Conwell, reference 17; courtesy of *Proceedings of the Institute of Radio Engineers*.)

enough to ionize completely all the impurity levels. In this temperature range, the number of current carriers is essentially independent of

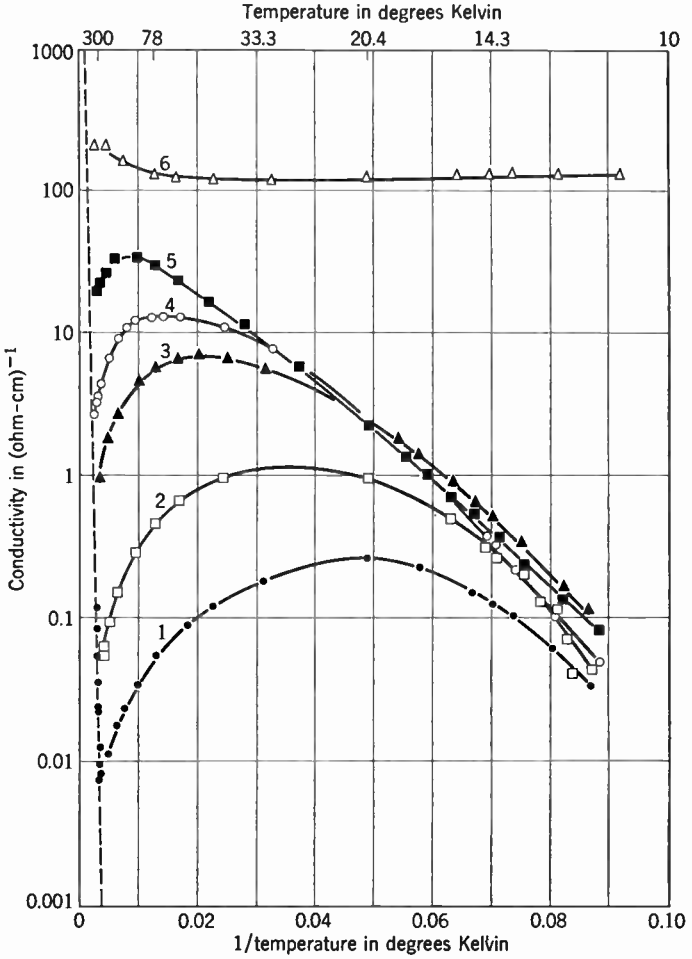


Fig. 1.10. Variation of Conductivity with Temperature for *n*-Type Germanium. (After Debye in Conwell, reference 17; courtesy of *Proceedings of the Institute of Radio Engineers.*)

temperature. At still lower temperatures, the current carriers begin to condense out into the impurity levels, and the number of carriers again falls exponentially with decreasing temperature. The decline is much less rapid than in the high-temperature range since it is related to the impurity carrier activation energies.

The conductivity as a function of temperature is similar to the curves shown in Fig. 1.9 but modified by the variation of mobility with temperature. The variation of the logarithm of the conductivity as a function of reciprocal temperature for the n-type germanium samples is shown in Fig. 1.10.

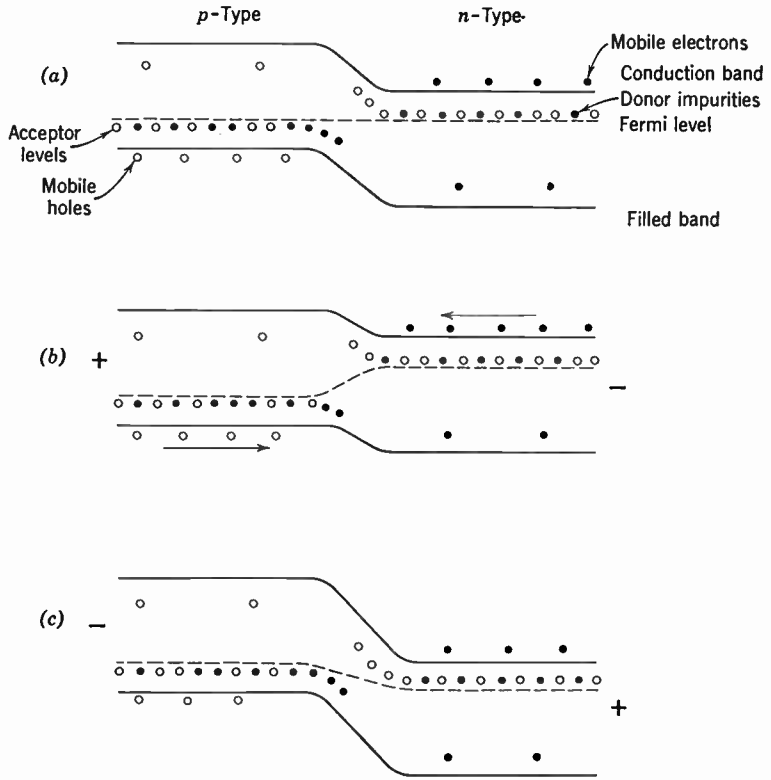


Fig. 1.11. Conduction Mechanism of a p-n Junction.

When an n- and p-type semiconductor are joined together and allowed to come into equilibrium, there is a readjustment of charge in the material in such a way as to bring the Fermi levels of the n- and p-type material together. This readjustment leads to a potential barrier at the junction of the two materials. The consequence of such a barrier is that the material can conduct the current in one direction but will have very low conductivity if the voltage is applied in the other direction. This is illustrated in Fig. 1.11. Case a shows the semiconductor when no potential is applied. In case b, the p-type

material is made positive and the *n*-type negative. Electrons from the negative electrode can enter the *n*-type material readily, and the electrons can diffuse through the barrier into the *p*-type material, where they recombine with holes. At the same time, holes can enter the *p*-type material and diffuse through the barrier and recombine with electrons in the *n*-type material. Consequently, the conductivity of the semiconductor junction is high. When the opposite potential is applied (Fig. 1.11*c*), current cannot flow since to a first approximation no electrons are formed in the *p*-type material to cross the boundary into the *n*-type material, nor are holes formed in the *n*-type material to cross the boundary into the *p*-type material. This behavior is typical of semiconductor rectifiers.

Recently, semiconductors have been brought into great prominence because of their application in transistors. This, however, is not the reason for introducing the topic in this chapter. Semiconductors are of interest primarily because photoconductors, which are becoming increasingly important in the field of television, can be described only in terms of semiconductor behavior. This more restricted group of non-metallic conductors will form the subject of the next section.

At the time of writing, transistors are not employed in any commercial television units. However, it is certain that many functions now performed by vacuum tubes will be performed in the future by transistors. Although transistor circuits will not be discussed in the succeeding pages of this book, it would not be out of place to conclude this brief section on the properties of semiconductors with a qualitative description of the transistor action of an *n*-*p*-*n* transistor.

The *n*-*p*-*n* transistor consists of two pieces of *n*-type material separated by a thin layer of *p*-type semiconductor. The arrangement is illustrated in *a* of Fig. 1.12. The left-hand *n*-type portion of the transistor is the emitter, the right-hand *n*-type portion the collector, and the central *p*-type region is the base. When a potential is applied to the material, the internal potential distribution and idealized current carrier distribution will be as shown in *b* of Fig. 1.12. If, now, the collector is made strongly positive with respect to the emitter and base, a small current flows between the emitter and collector due to the diffusion of electrons across the *p*-type barrier. This current will be greater than the current for a rectifying barrier similar to that shown in Fig. 1.11*c* because the *p*-type layer is narrow enough to permit some electron diffusion. If the emitter is made slightly negative with respect to the base (*c*), the diffusion of electrons from the emitter into the base becomes very rapid. Most of these electrons diffuse directly across

the base and into the collector with only a small part of the current passing out through the base leads. This small current is due to the recombination of electrons in the base. The result of this action is that a small change in potential between emitter and base causes a

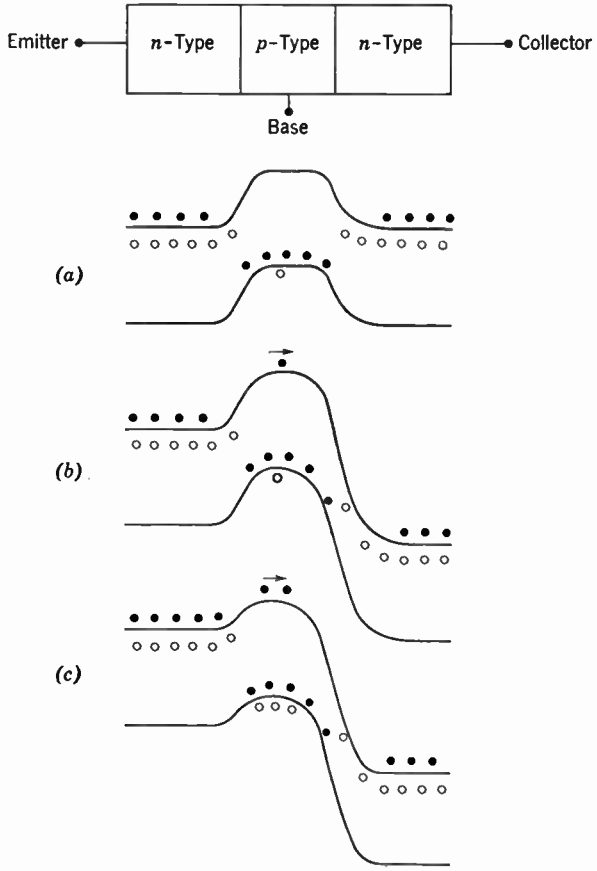


Fig. 1.12. A *n-p-n* Junction Transistor.

large change in current received by the collector. On the other hand, a large change in potential of the collector causes almost no change in current. This is exactly the condition required for amplification and is the reason why this type of device has the potentiality of replacing vacuum tubes in many of their functions.

1.4 Photoconductivity. In the preceding section, the excitation of current carriers in a semiconductor was confined to the action of the

thermal energy of the material. Other forms of energy are also capable of exciting these carriers. In particular, the material may absorb a photon of light or other radiant energy and use the energy contained therein to excite a hole or an electron. This phenomenon, known as "photoconductivity," is of considerable importance in television.

It has been recognized since the work of Planck and Einstein in the early part of the twentieth century that electromagnetic radiation, when it interacts with matter in such a way as to exchange energy, behaves as if it were corpuscular in nature. The particles making up light are known as photons. The energy associated with a photon depends upon the wavelength or frequency (ν) of the light involved and is given by the following relation

$$E = h\nu \tag{1.9}$$

where h is a Planck's constant.

Photoconduction in a semiconductor can be produced only by radiation with photon energies equal to or greater than the energy needed to generate a carrier in the material. As with thermally excited conductivity of semiconductors, the carriers may be the result of the lifting of an electron from the filled band into the vacant conduction band of the material. This corresponds to intrinsic photoconductivity. On the other hand, the excitation may be from a donor-type impurity into the conduction band or from the filled band into an acceptor-type impurity. In this case, the effect is known as impurity photoconductivity.

From the above, it is evident that the spectral response of a photoconductor is dependent entirely upon the energy intervals of the transitions involved. In an intrinsic photoconductor, the gap between the top of the filled band and the bottom of the conduction band is the determining interval. Table 1.1 lists a few intrinsic photoconductors, giving their gap width in electron volts and the long wavelength limit of the spectral response of each.

Where impurity photoconductivity is involved, the situation is more complex. Here the energy interval depends not only upon the host semiconductor but also upon the impurity which is involved in the photoconductivity. When light falls on a photoconductor, the fraction of the photons which cause transitions giving rise to photoconductivity may vary considerably from material to material. The fraction of photons which do cause such transitions is the primary quantum efficiency of the material. For many materials the quantum efficiency is low owing to the existence of transitions which can absorb photons from the incident radiation but which are not among those involved in

TABLE 1.1. GAP WIDTH AND LONG-WAVE THRESHOLD OF PHOTOCONDUCTORS

(Data from T. S. Moss, *Photoconductivity in the Elements*, Academic Press, 1952; and Zworykin and Ramberg, "Photoelectric Cells," *Encyclopedia of Chemical Technology*, 1953.)

	Gap Width E_g (electron volts)	Long-Wave Threshold λ_0 (microns)
Se (amorphous)	2.4	0.51
Se (black, crystalline)	1.5	0.80
Si	1.13	1.08
Ge	0.72	1.72
Tl_2S	1.0	1.25
PbS	0.4	3.0
PbSe	0.27	4.5
PbTe	0.25	5.0 (at 90°K)
CdS	2.4	0.51
Sb_2S_3	1.6	0.75

The threshold given above represents the point of sharp falling off of the spectral response, rather than an absolute threshold of sensitivity.

freeing bound current carriers; on the other hand, there are a great many photoconductors for which the quantum efficiency γ is close to unity. These are, of course, the ones which are of greatest interest whenever electrical effects are to be obtained from an optical stimulus. The primary quantum efficiency of a photoconductor cannot, of course, exceed unity. However, as will be outlined below, gain mechanisms can operate in photoconductors in such a way as to give an apparent quantum efficiency which is much higher than unity.

The simplest concept of a photoconductor is the following: An incident photon excites an electron from the filled band to the conduction band, leaving a hole behind. The field applied across the photoconductor causes these two current carriers to drift to the positive electrode and to the negative electrode respectively. The current which would flow in such a photoconductor, due to the incidence of a light flux F (lumens per square centimeter), will be given by

$$I = e\gamma aFLw \tag{1.10}$$

where a is the number of photons per second per lumen; γ, the quantum efficiency; L, the length of the cell; w, its width; and e, the charge on an electron.

The assumption here is that the applied voltage is sufficient to move the carriers completely out from the sample. Therefore, the voltage

does not appear in the expression for the current. This concept is much too simple to explain the behavior of any real photoconductor. In the first place, many photoconductors exhibit photocurrents many times as great as that indicated by the above relation. Furthermore, the current depends greatly upon the voltage applied across the material. Finally, the current does not cease to flow immediately upon removal of the stimulating radiation.

The next extension of the model is to assume that as each current carrier, whether it be hole or electron, moves out of the crystal, it is replaced by another flowing in from the electrodes. This replacement continues until a recombination occurs which binds the hole and/or electron so that it can no longer serve as the current carrier. In terms of this model, the current will be given by equation

$$I = e\gamma aF \frac{w}{L} \mu_a V t_c \tag{1.11}$$

In this expression, t_c is the lifetime of the carrier and μ_a its average mobility. Although the description above has been given in terms of an intrinsic photoconductor, the same principles apply when the excitation is from an impurity level to the conduction band or from the filled band to an impurity level. When this type of photoconductivity is involved, only one kind of carrier is free to move. For example, in a photoconductor with a donor-type impurity being responsible for the sensitivity to light, the action might be pictured somewhat as follows.

The incident photon excites an electron from the donor level up into the conduction band. Under the influence of the field produced by the voltage applied across the cell, the electron starts to move away from the ionized donor. As it moves toward the positive electrode, the potential about the positive donor becomes increasingly positive. This is analogous to a space charge effect. If, as the charge in question moves away from the ionized donor, a charge from the negative electrode moves toward it, the effective neutrality of the material can be maintained. As long as the ionized donor is present, the current can continue to flow. Eventually, however, the donor captures an electron, preventing further motion of electrons through the crystal. It will be apparent that an acceptor-type impurity can behave in an identical manner towards free holes in the filled band.

Before discussing the lifetime t_c of the carriers and the average mobility μ_a, let us consider further the question of the replacement of one carrier by another as the former moves out of the material or out of a specific region of the material. There are many ways in which this can

take place. Thus, crystals in which the flow of carriers is obstructed by potential barriers are of common occurrence. Assume, for example, that the current is primarily carried by electrons. If there is a small region across the cell where there is excess negative charge, there will be a potential barrier set up in this region which will prevent the flow of electrons. This is illustrated in Fig. 1.13. Such a barrier is capable of capturing a hole or positive charge should one be formed by the action of light. When a hole is captured by such a barrier, the poten-

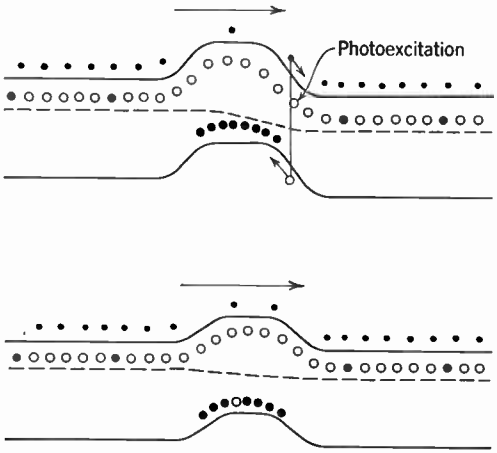


Fig. 1.13. A Photoconductor Including a Potential Barrier. Lowering of barrier by hole trapping.

tial barrier is lowered by the addition of the positive charge. This allows negative carriers to flow through this region. The flow will continue until the positive charge or hole captures an electron, re-establishing the potential barrier. For crystals which are good insulators, the flow of carriers between electrodes is prevented by space charge. At the negative electrode there will be an accumulation of electrons such that the gradient in that region will no longer allow the flow of electrons. A single hole flowing through this region will neutralize the space charge to the extent of allowing one additional negative charge to flow therein. Similarly, the positive space charge formed at the positive electrode due to the deficiency of electrons or an accumulation of holes can be neutralized by an electron, allowing a hole to flow out of this region. Again, holes or electrons entering the contact region of a photoconductive cell may serve to neutralize potential barriers at these contacts, these barriers otherwise inhibiting the flow

of electrons or holes through the material. This does not exhaust the ways in which carrier replacement may occur in photoconductors, but the examples cited above serve to indicate that the release of a primary carrier by the action of the incident radiation may permit the passage of charge which is many times greater than the single unit of charge released by the initial photon.

The average mobility μ_a appears in the expression for the current in a photoconductor. In the preceding section, the mobility and its relationship to thermal and impurity scattering were discussed. An additional effect occurs in photoconductors which makes the average or effective mobility much lower than the mobility of ideal holes or electrons. It is the existence of traps. Impurity or structural imperfections in a crystal may temporarily capture current carriers. During the time they are captured, they, of course, cannot contribute to the current flow. Because the potential energy represented by traps is small, thermal excitation of the crystal causes the carriers thus trapped to be released again and resume their drift. The average or effective mobility of the carriers, therefore, must take into account the time charge carriers spend in traps. This trapping of charge carriers is of major importance in real photoconductors. One of the objectives in making good photoconductive material is to reduce traps to the minimum.

The lifetime of a carrier is determined by the length of time the carrier or its replacement can move through the crystal before recombination occurs. The process of recombination is a complex one. In general, the probability of recombination between a free hole and an electron is very small. However, if either the hole or the electron is captured or is immobile, its capture cross-section toward a carrier of the opposite sign becomes fairly large. In real photoconductors, the capture cross-section of bound holes for electrons or bound electrons for holes ranges between 10^{-16} and 10^{-23} square centimeter. Given the capture cross-section of a center for a carrier σ and the thermal velocity of the carrier v_s, the lifetime for the carrier will be given by the following relation:

$$\tau = \frac{1}{N_c \sigma v_s} \tag{1.12}$$

where N_c is the density of centers. The lifetimes range from a few microseconds to many seconds in the range of materials used for photoconductors. The above discussion relates primarily to photoconductors which are good insulators in darkness; in other words, photoconductors

where the dark current is small compared to the photocurrent. As will be shown in Chapter 10, this type of photoconductor is important in the Vidicon pickup or camera tube and is, therefore, of primary interest in television.

1.5 Electron Emission. The theory of metals as outlined in section 1.2 can be extended to account for most of the phenomena observed in connection with the thermionic emission of electrons. On the basis of the quantum theory picture of metals, nearly every phase of electron emission from pure metals can be explained not only qualitatively but also quantitatively. The theory also has been extended to cover the case of metal surfaces on which there is a thin overlying contaminated film, but the analysis here is somewhat less rigorous. Since, however, this type of surface is very important from the standpoint of the emission of electrons, it cannot be left out of consideration. There must necessarily be gaps and indefiniteness in some of the account which follows.

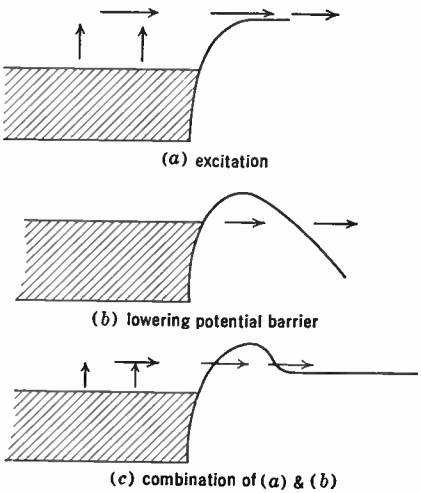


Fig. 1.14. Emission of Electrons.

With reference again to Fig. 1.3, which shows the kinetic energy of electrons superimposed upon the potential energy diagram for the metal, it is evident that one or both of two conditions must be fulfilled before electrons can escape through the surface barrier and be emitted. Either some of the electrons must be given additional energy so that their kinetic energy is at least as great as the potential barrier, or the height and width of the barrier must be reduced to such extent that the electrons can penetrate through it. Diagrammatically, these might be represented as shown in Fig. 1.14. The first diagram illustrates electrons which have been given sufficient additional energy to escape over the normal potential barrier. This energy may be supplied in the form of thermal energy taken from the lattice of a heated metal crystal, in which case the emission is known as thermionic emission. On the other hand, if the additional energy is obtained from radiant energy incident upon the metal, photoelectric emission takes place. Where the metal is bombarded by high-velocity electrons, there may be an interaction

between the electrons in the metal and the incident electrons, resulting in the former acquiring sufficient energy to escape. This phenomenon is known as secondary emission.

Diagram *b* of Fig. 1.14 shows the potential barrier depressed to such an extent that electrons can escape from the surface. This may occur when there are extremely high fields at the metal surface, the effect being known as cold emission. Whereas in classical theory emission is possible only when the kinetic energy of the electrons exceeds the potential barrier, quantum mechanics indicates a finite probability for an electron traversing a potential barrier which is higher than its kinetic energy. The probability of transmission here depends on the thickness and the height of the barrier. Very frequently both an addition of energy by a thermal or other excitation process and a depression of the barrier are responsible for the emission, as shown in diagram *c*. The depression of the barrier may, for example, be caused by a certain contamination on the surface of the metal. When this occurs, much less energy has to be added to the electrons to cause them to escape, thus making the substance a much better electron emitter. This effect will be the subject of further discussion when problems dealing with practical emitters are considered.

1.6 Shot Effect. As has been stated, energy is supplied to some of the vast numbers of free electrons in order to produce thermionic emission, photoelectric emission, etc. The fraction of electrons which receive enough energy to permit them to escape is extremely small, and only part of these move in such a direction that they are emitted. The whole process is, on this basis, an entirely random phenomenon, each electron being emitted entirely independently of any other.

Since it is a random process, the number of electrons emitted during any given interval differs by a small amount from the average number of electrons emitted in an interval of this duration. The electron current from an emitter is therefore not constant, but is subject to statistical fluctuation. As this fluctuation limits the performance of many types of electronic devices, it is important to determine its magnitude and nature.

An emitter which releases N electrons in a time T, where N is a large number, has an average rate of emission of $n_0 = N/T$ electrons per unit time. However, since the emission phenomenon is a random one, the actual number n emitted during any given unit of time will differ from n_0. According to the laws of probability, while the average difference

$$\overline{(n - n_0)}$$

must equal zero, the mean square difference in unit time is not zero but rather is given by

$$\overline{(n - n_0)^2} = \overline{\Delta n^2} = n_0 \tag{1.13}$$

Since the emission of each electron corresponds to the transfer of a charge $-e$, Eq. 1.13 can be expressed in terms of charge, giving

$$\overline{(\Delta ne)^2} = e^2 n_0$$

$$\overline{\Delta q^2} = qe \tag{1.14}$$

In other words, the mean square fluctuation of charge is proportional to the charge emitted.

In the form given above in Eq. 1.14, the relation is not very directly applicable to actual electrical circuits. In general, the circuits encountered respond to frequencies which lie in a band between definite limiting frequencies. The useful signal is in the form of current variations over the frequency band in question. The shot effect, or statistical fluctuations, introduces current variations over this same frequency band which tend to mask the useful signal. Therefore Eq. 1.13 or 1.14 must be reformulated to express the magnitude of this unwanted random "noise" in terms of a fluctuating current in the circuit containing the emitting element.

Such a relation was first worked out by Schottky in 1912. The derivation given below,* though not identical with that originally given by Schottky, yields the same information. As a starting point, the current pulse resulting from the motion of an individual electron is expressed as a Fourier integral which gives its frequency spectrum. The mean square of a large number of these current pulses is then found by summing and averaging.

The current arising from a single emitted electron can be expressed as follows:

$$i_1(t) = \frac{e}{\pi} \int_0^\infty d\omega \cos \omega(t - T_1) \tag{1.15}$$

where T_1 is the time at which the electron is actually emitted. This equation assumes that the emitted electron produces a single current pulse whose duration is extremely small compared with any interval of time for which Eq. 1.15 is to be used.

* From the unpublished work of E. G. Ramberg.

The mean square current due to N electrons emitted in the time T is, therefore,

$$\overline{i^2} = \frac{1}{T}\int_0^T dt \left\{ \sum_{v=1}^{N} \frac{e}{\pi} \int_0^\infty d\omega \cos \omega(t - T_v) \right\}^2 \tag{1.16}$$

If the indicated squaring of the sum over the N electrons is carried out, this equation becomes

$$\overline{i^2} = \frac{e^2}{\pi^2}\frac{1}{T}\int_0^T dt \int_0^\infty d\omega \int_0^\infty d\omega' \sum_{v=1}^{N} \sum_{v'=1}^{N} \cos \omega(t - T_v) \cos \omega'(t - T_{v'}) \tag{1.17}$$

This integral expression must be integrated over ω' and t. When performed, this integration leads to the equation

$$\overline{i^2} = \frac{e^2}{\pi^2}\frac{1}{2T}\int_0^\infty d\omega \sum_{v=1}^{N} \sum_{v'=1}^{N} 2\pi \cos \omega(T_v - T_{v'}) \tag{1.18}$$

which expresses the mean square current over an infinite frequency band.

Since the response of any real circuit is zero except over one or more finite bands of frequency, an expression for the current fluctuation in a limited band covering the frequencies between f_1 and f_2 is of value. For this, the limits of integration over ω instead of being 0 to ∞ are taken as $2\pi f_1$ and $2\pi f_2$. Equation 1.18 consequently becomes

$$\begin{aligned}\overline{i^2} &= \frac{e^2}{\pi^2}\frac{1}{2T}\int_{2\pi f_1}^{2\pi f_2} d\omega \sum_{v=1}^{N} \sum_{v'=1}^{N} 2\pi \cos \omega(T_v - T_{v'}) \\ &= 2\frac{e^2 N}{T}(f_2 - f_1) \\ &\quad + \frac{e^2}{\pi T} \sum_{\substack{v=1 \\ v \neq v'}}^{N} \sum_{v'=1}^{N} \frac{\sin 2\pi f_2(T_v - T_{v'}) - \sin 2\pi f_1(T_v - T_{v'})}{T_v - T_{v'}}\end{aligned}$$

The second term of this expression vanishes because of the assumed randomness of the emission. That is, since the time interval $T_v - T_{v'}$ between the emission of any two electrons is purely random, the summations represent the superposition of a vast number of sines of random angles which are as likely to have a positive as a negative value, and therefore average zero. Thus the mean square fluctuation current

over the frequency band $f_2 - f_1$, for an average emission current $i_0 (i_0 = Ne/T)$, reduces to

$$\overline{i^2} = 2ei_0(f_2 - f_1) \tag{1.19}$$

As is evident from the general nature of the derivation, Eq. 1.19 applies to nearly all types of electron emission. Actually, the relation accurately gives the fluctuations in all forms of photoelectric and thermionic emission and in cold emission if the voltage is well below the rupture point. Secondary emission is an exception, because each incident electron causes the emission of several electrons, giving a form of coherence between the emitted electrons which does not fulfill the condition of random emission upon which the derivation is based.

The statistical current fluctuation given by Eq. 1.19 represents the minimum that can be expected from any emitting surface in the absence of space charge. Other effects not directly related to the basic phenomenon of electron emission may operate to increase the total fluctuation of the current emitted.

The following problem of a phototube emitting electrons in response to a modulated light signal serves to illustrate a typical noise calculation. Let it be assumed that the photoelectric current is amplified by an amplifier which responds to a frequency band F, 10,000 cycles wide. If the photocurrent is $i_0 = I(1 + k \sin 2\pi ft)$, where k is a modulation factor which has a value 0.1, what is the minimum value of I which will give a signal-to-noise ratio of 10?

In this case the mean square signal is $\overline{i_s^2} = \frac{1}{2}k^2I^2$, and the mean square noise is $\overline{i_n^2} = 2eFI$.

The signal-to-noise ratio is, therefore,

$$R = \sqrt{\frac{\overline{i_s^2}}{\overline{i_n^2}}} = \frac{\frac{1}{\sqrt{2}} kI}{\sqrt{2eFI}}$$

or, in the example in question,

$$I = \frac{100 \times 32 \times 10^{-20} \times 10^4}{\frac{1}{2} \times 10^{-2}}$$

$$= 0.64 \times 10^{-10} \text{ ampere}$$

In the foregoing, it is assumed that the entire noise was the shot noise of the emitted electrons in the phototube. Actually noise is generated in the resistor used to couple the phototube to the grid of the first tube of the amplifier, and noise is also introduced by the first tube itself.

The effects of these elements will be discussed in Chapters 9 and 13, in which problems of amplification are considered.

1.7 Thermionic Emission. The general properties of the emission of electrons from metals having been considered, it is now proper to treat the different types of emission individually. As has already been stated, it is possible to cause electrons to be emitted by raising the temperature of a metal, this effect being known as thermionic emission. The emission of current thus obtained will depend upon the height of the potential barrier above the top of the Fermi distribution of free electrons, upon the shape of the barrier, and upon the energy distribution of the electrons.

The energy distribution of electrons can be treated quite generally by using the Fermi-Dirac distribution function, which is applicable to all metals. The height and shape of the potential barrier, however, depend not only upon the particular metal, but also upon the state of the surface of the emitter. In the discussion which follows two types of thermionic emitters will be considered: one where the base of the emitter is a pure metal with or without a contaminant modifying the surface, and the second where the body of the emitter is a semiconductor.

To calculate the electron emission, it is necessary in the first place to determine the number of electrons arriving in unit time at the surface of the emitter with sufficient kinetic energy directed toward the surface to pass through the potential barrier. This can be done with the aid of Eq. 1.3, as was shown by Nordheim, who derived the thermionic emission equation on a quantum mechanical basis. The Fermi function is integrated over the velocity components v and w taken from zero to infinity. This expression, when rewritten to express the number of electrons $N(W)\,dW$ which have a kinetic energy in the range W to $W + dW$ associated with the component of velocity u which is assumed to be directed normal to the metal surface, gives the following distribution function:

$$N(W)\,dW = \frac{4\pi m}{h^3} kT \sqrt{m/(2W)} \ln (1 + \epsilon^{-(W-\mu)/(kT)}) \qquad (1.20)$$

Then, since only those electrons having an energy W greater than W_a can escape through the boundary, the total number of electrons leaving a unit area per unit time will be

$$n_0 = \int_{W_a}^{\infty} uN(W)\,dw \qquad (1.21a)$$

For convenience in performing the indicated integration, the distribution function $N(W)$ can be expressed as

$$N(W)\,dW = \frac{4\pi mkT}{h^3}\frac{1}{u}\epsilon^{-(W-\mu)/(kT)}\,dW \tag{1.21b}$$

since over the range of integration $\epsilon^{-(W-\mu)/(kT)}$ is small. Equation 1.21*a* therefore becomes

$$n_0 = \frac{4\pi m(kT)^2}{h^3}\epsilon^{-(W_a-\mu)/(kT)} \tag{1.22}$$

Experimentally and theoretically it can be shown that some of the electrons in the energy range above W_a are reflected back. If r is taken as the reflection coefficient, the fraction of electrons escaping must be $1 - r$.

Including this transmission coefficient, combining the factor $4\pi mk^2e/h^3$ into a single constant A_0, and expressing the rate of emission of electrons in terms of current, one can write Eq. 1.22 as

$$i_0 = (1 - r)A_0T^2\epsilon^{-(W_a-\mu)/(kT)} \tag{1.23}$$

It should be noticed that this equation has practically the same form as the thermionic emission equations derived by Richardson, Dushman, and others on the basis of classical thermodynamics.

In the exponent of this equation there appears the term $(W_a - \mu)$. It represents the difference between the height of the potential barrier and the maximum kinetic energy of electrons in the Fermi distribution at absolute zero temperature, and is given the name "work function." It is common practice to express this work function in terms of electron volts, so that the exponent in the emission equations takes the form $-e\phi_0/(kT)$, or, to simplify it still further, by the substitution $b_0 = e\phi_0/k$. Also, for convenience, the coefficient A_0 and the reflection factor are often combined into a single constant A. When thus written, Eq. 1.23 becomes

$$i = AT^2\epsilon^{-b_0/T} \tag{1.23a}$$

A few representative values for the constants A and b_0 and the work function ϕ_0 are given in Table 1.2.

The thermionic emission from the metals listed is very small when the metal is heated to any reasonable temperature. For example, tung-

TABLE 1.2. THERMIONIC EMISSION CONSTANTS OF METALS *

Element	ϕ_0 (volts)	$b_0 \times 10^{-3}$ $(°K)^{-1}$	A, amperes $(°K)^{-2}$
Calcium	2.24	26	60.2
Cesium	1.81	21	162
Columbium	3.96	46	57
Molybdenum	4.44	51.5	60.2
Nickel	2.77	32.1	26.8
Platinum	6.27	72.5	1.7×10^{-4}
Tantalum	4.07	47.2	60.2
Thorium	3.35	38.9	60.2
Tungsten	4.52	52.4	60.2
Zirconium	4.13	47.9	330

* From L. R. Koller, reference 14.

sten heated to 2200°K will emit only 13 milliamperes per square centimeter. This is because for pure metal surfaces the work function is generally high. Calcium or cesium, whose work functions are low, have such low melting points that they cannot be used as thermionic cathodes.

The most completely studied composite-surface thermionic emitter consists of thorium on tungsten. In Table 1.2 the work function of tungsten alone is 4.52 volts whereas that of thorium is 3.35 volts. However, a surface consisting of tungsten having over it a monatomic layer of thorium covering about 70 percent of its surface has an effective work function of only 2.6 volts. If a contaminating layer of cesium is used instead of thorium, the work function is still further depressed, being only 1.6 volts. Unfortunately, this surface is unstable and the cesium evaporates from the base metal before temperatures can be reached which will give a high density of electron emission. Unlike the cesium surface, the thoriated surface is stable up to temperatures of 2000 to 2200°K or above, which property makes this emitter very valuable in practical electronics. The emission from a thoriated tungsten cathode at 2000°K is more than one thousand times that from pure tungsten.

The explanation of the increased emission of these complex surfaces lies in the nature of the work function. This work function may be divided into two parts, the first being due to the attraction of the lattice ions as a whole for the electrons, the second to the dipole moment of the surface. The first remains unchanged when a contaminated layer is added to the surface; the second factor may be greatly altered in either direction. When a cesium atom approaches

a tungsten surface, it enters a region of fairly high positive potential which tends to remove its valence electron, and, since the work function of tungsten is greater than the ionization potential of cesium, the atom will be actually ionized. Thus the cesium adheres to the surface in the form of a positive ion. As more cesium atoms arrive, these tend to ionize also. However, the addition of each ion tends to lower the work function of the surface so that, after a certain number have adhered, the work function is no longer sufficient to produce further ionization. Even under these conditions, partial ionization or polarization takes place. A tungsten surface, therefore, which is more or less completely covered with cesium may be thought of as being coated with a layer of cesium partial ions which cause a lowering of the effective work function ϕ_0.

For thorium, the work function of tungsten is less than the ionization potential. This does not preclude the existence of thorium ions on the surface, since both image force and work function act to produce ionization. However, unlike elements which are ionized by the work function alone, thorium evaporated from the surface does not leave in the form of ions. Beyond a certain coverage, partial ions only are formed. Both thorium ions and partial ions contribute to lowering the work function.

If, instead of an electropositive contaminating atom, an electronegative element is allowed to form a surface layer on the tungsten, the opposite effect occurs, and the work function is raised. For example, a monatomic layer of oxygen on tungsten raises the work function to over 6 volts.

A more complex layer can be formed if oxidized tungsten is exposed to cesium vapor, thus allowing the formation of a final positive layer. In this case, the work function is lower than for cesium on pure tungsten and a more stable surface is also formed. Such a surface will yield 350 milliamperes per square centimeter when the emitter is heated to 800°K.

In view of the practical importance of the thoriated tungsten emitter, it will not be out of place to discuss in greater detail the actual form of these cathodes. The base metal is tungsten to which has been added 0.5 to 2.0 percent by weight of thorium oxide. The material is shaped to the form required for the cathode, assembled in the electron tube in which it is to be used, and the tube is exhausted. To activate the emitter, it is first flashed for a few seconds at 2800°K, then aged at 2100°K. The cathode is then ready for operation.

At the flash temperature, some of the thoria reduces to metallic thorium. This metallic thorium diffuses through the base to the surface. Electron microscope examination shows that this diffusion is partly in the form of minute eruptions through the crystal grains and partly along grain boundaries. At temperatures as high as 2800°K, it is evaporated from the surface immediately upon its arrival. In other words, the rate of evaporation greatly exceeds the rate of diffusion.

As the temperature is decreased, both the rate of diffusion and the rate of evaporation decrease. However, the evaporation rate falls faster with temperature. Below 2500°K, a layer of thorium begins to form on the surface. At the aging temperature of 2100°K, an optimum coverage is obtained. At still lower temperatures, the evaporation rate reaches an extremely low value; under this condition, the cathode is stable and its operating life adequate. If the emitter loses its activation, it can usually be restored again by flashing, both to clean off possible deleterious contaminations from the surface and to reduce more thoria. The surface is then again aged at 2100°K as is done in the initial activation.

For circuits handling low power and for both television pickup and viewing tubes, the oxide-coated cathode is by far the most important. Here the thermionic emission is from a semiconductor, and the mechanism of emission is somewhat different from that of electrons from a pure metal. These emitters consist of a semiconductor, in particular barium oxide–strontium oxide, in contact with a metal such as nickel. When a metal and a semiconductor are brought into contact, equilibrium conditions require that the same electron current must flow from the metal to the semiconductor as from the semiconductor to the metal. This condition is satisfied only if the Fermi levels of the semiconductor and the metal are at the same energy level, as is indicated in Fig. 1.15*c*. This can, in general, be accomplished only by a readjustment of charges in the material such that a surface layer of charge forms at the metal-semiconductor interface, and there is an enrichment or depletion of charges inside the semiconductor near the contact surface.

The sequence in Fig. 1.15 illustrates the process. In *a* an *n*-type semiconductor is assumed to be some distance from the metal. When it is close to the metal, the charge readjustment brings the two Fermi levels into coincidence, with a consequent difference in contact potential between the two materials. The final stable condition is shown in *c*, where donor levels have lost their charges to surface states at the interface. This causes the potential of the bottom of the conduction

band and the top of the filled band to curve upward at the junction. If π is the potential difference between the bottom of the conduction band and the zero potential of space and ζ_f is the depth of the Fermi level below the bottom of the conduction band, the energy which must be given electrons at the bottom of the conduction band for emission

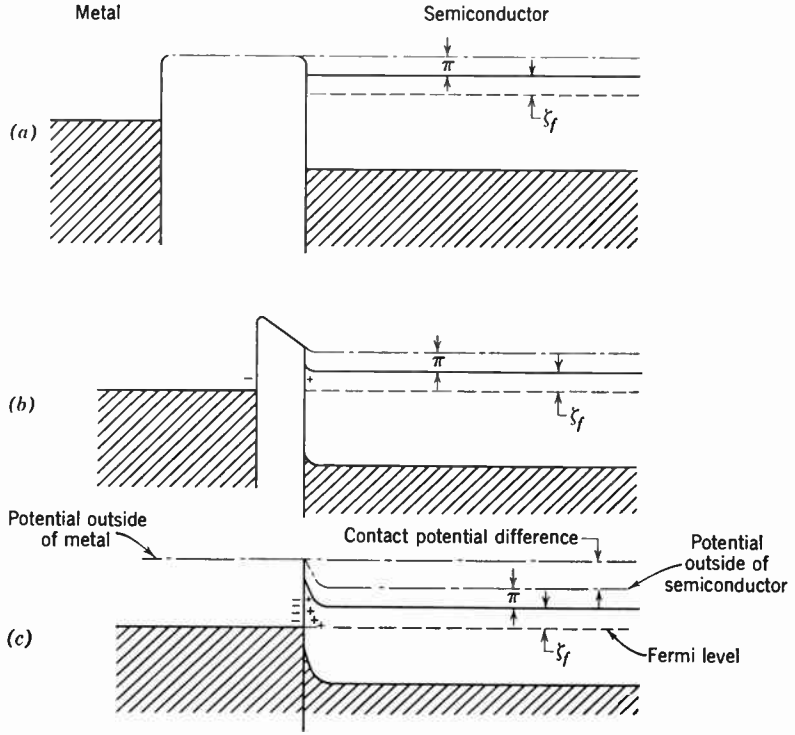


Fig. 1.15. Metal-Semiconductor Junction.

will be π. The effective thermionic work function then becomes $\pi + \zeta_f$, as indicated by the form of the equation giving the emitted current:

$$i_0 = (1 - r)ne\sqrt{\frac{kT}{2\pi m}}\,\epsilon^{-\pi/(kT)} = \frac{4\pi m(kT)^2}{h^3}e(1 - r)\epsilon^{-(\pi+\zeta_f)/(kT)} \quad (1.24)$$

Where the temperature of the emitter is high enough so that the impurities are completely ionized, this expression becomes

$$i_0 = (1 - r)Ne\sqrt{kT/(2\pi m)}\,\epsilon^{-\pi/(kT)} \quad (1.25)$$

Here n is the concentration of conduction electrons, N is that of the impurities. It will be observed that the emitted current is strongly dependent upon the impurity concentration.

The practical oxide cathode consists of a thick layer of a mixture of barium and strontium oxide on a nickel base. Frequently a small amount of a highly reducing metal such as aluminum is alloyed with the nickel. The semiconductor of these cathodes is a mixed crystal of barium and strontium oxide. The impurity which gives it an n-type character is barium. In practice, the percentage of barium mixed with

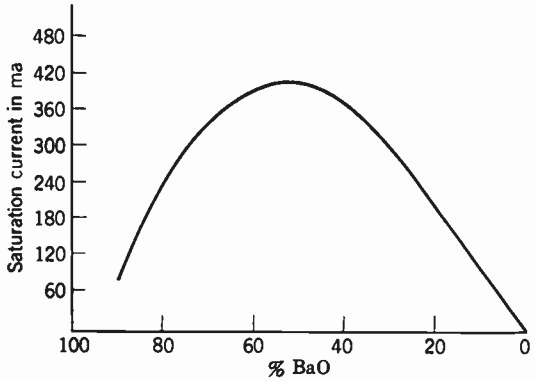


Fig. 1.16. Variation of Emission from an Oxide Cathode with Percentage of Barium Oxide.

strontium is not very critical. Figure 1.16 shows the relation between the barium oxide–strontium oxide ratio and the emission.

The material is prepared by a coprecipitation of strontium and barium carbonate to form a mixture containing about equal parts by weight of the two components. The effect of varying the relative weights of the two will be discussed later. In order to keep it from flaking off before the carbonate is reduced, the mixture is suspended in a suitable organic binder. This is applied to the nickel base of the cathode, and the whole is assembled in the electron tube in which it is to be used. After the tube has been pumped, but before it is sealed off on the vacuum system, the cathode is heated to about 1300°K for approximately one minute. This heating drives off the binder and reduces the carbonates to oxide. The temperature is then reduced to about 1100°K, and a voltage is applied between the cathode and a neighboring electrode. The temperature being left constant, the voltage is gradually increased. It will be found that the current drawn to the electrode increases and then reaches saturation. With constant

voltage and temperature, the tube is aged for a period of 10 to 20 minutes, completing the activation of the cathode.

The operating temperature of this type of cathode is from 1000 to 1050°K. A number of practical precautions must be taken in activating oxide cathodes. An oxygen-bearing film may be formed on any electrode near the cathode which is exposed to the evaporation products from it. If this electrode is subjected to electron bombardment, the film will decompose, releasing oxygen.

This is only one of the many activation schedules used in practice, the particular activation found most satisfactory depending upon the specific device in which the cathode is to be used. These variations do not, in general, increase the specific emission from the cathode material but often decrease the percentage of failures.

A properly activated surface will deliver as much as an ampere per square centimeter and has a working emission of 300 to 400 milliamperes per square centimeter at 1000°K. The work function is about one volt as determined from the variation of current with temperature.

When a cathode of this type has been operating for a period, a high-resistance layer tends to form between the nickel and the barium oxide. This layer can seriously interfere with the operation of the cathode in a number of ways. It is also found that if the emission is drawn from the cathode in the form of very short pulses with a relatively small duty cycle, currents as high as 10 or more amperes per square centimeter can be drawn from the surface. The reason for this is that under d-c operation there tends to be an electrolysis of the ionized barium donor impurities moving them away from the face of the cathode toward the nickel base. The reduction in impurity content near the emitting surface causes the observed large decrease in emission under d-c conditions.

Before the subject of thermionic emission is left, mention must be made of the variation of current with applied voltage in a normal two-electrode vacuum tube. The current-voltage relation for a diode consisting of a thermionic emitter and a collecting plate can be represented by the curve shown in Fig. 1.17. When the plate is slightly negative, a few electrons reach it owing to the fact that some have fairly high initial velocity. The Maxwellian velocity distribution of the emitted electrons determines the fraction that will arrive under any given voltage. As the voltage between the cathode and the plate is reduced to zero, more electrons reach the plate. However, the number arriving will be less than the maximum emission, owing to the retarding effect of the cloud of electrons lying between the two ele-

ments. This is known as the space-charge effect. The action of space charge can be most readily seen by considering the potential distribution between cathode and plate. This potential is the result not only of the voltage applied between the electrodes but also of the charge between them, and, to determine the potential distribution, it is necessary to integrate Poisson's equation $\Delta\phi = 4\pi\rho$, where ρ is the density of charge in the intervening space. Figure 1.18 illustrates a typical family of potential distribution curves for a given voltage difference between the cathode and anode as the specific emissivity of the cathode

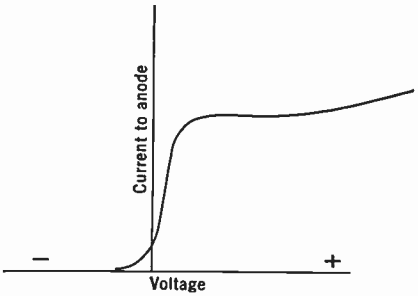


Fig. 1.17. Idealized Current Characteristics of a Diode.

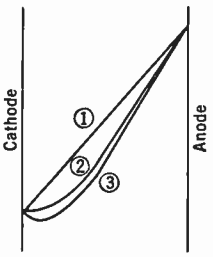


Fig. 1.18. Effect of Space Charge on Potential Distribution.

is increased. It is evident from this figure that to obtain the total emission current from the cathode, it is necessary to apply a positive potential such that a potential minimum does not exist between the two electrodes. The initial rise in current with increasing positive potential on the plate, shown in Fig. 1.17, is the result of overcoming the effect of space charge. Over this portion of the curve, the current is proportional to $\phi_0^{3/2}$. The rise in current continues until the potential minimum between the cathode and plate no longer occurs. When this condition is reached, the current is practically independent of voltage. This value of current is called the saturation current.

Actually, the current continues to increase as the voltage is raised. This is a consequence of the lowering of the potential barrier by the external field which has already been noted in an earlier section. The relation between the applied voltage and the collected current is

$$i = i_s \epsilon^{-(4.398E^{1/2})/T} \tag{1.26}$$

where i_s is the saturated current and E is the field strength at the emitter in volts per centimeter. This formula was calculated by

Schottky, and the phenomenon has become known as the "Schottky effect." For composite surface emitters and oxide-coated cathodes, this relation must be considerably modified.

1.8 Photoelectric Emission. Photoemission, i.e., the emission of electrons from matter by light or radiant energy, is of very great importance in television. It, together with the photoconductive effect, is fundamental to all classes of pickup devices. In order to account for the experimental facts of photoelectric emission, it is necessary to make use of the quantum theory of radiation just as it was for the photoconductive effect. Again, the radiation behaves as though it were composed of photons or quanta of energy each with an energy $h\nu$. When a photon interacts with an electron giving up its energy, it increases the total energy (kinetic plus potential energy) of the electrons by an amount equal to $h\nu$. When radiation falls upon a clean metal surface which is assumed to be at such a temperature that few of its electrons have energies much greater than that corresponding to the top of the filled Fermi band, some of the photons will give up their energy to free electrons. If the maximum energy any of these electrons can have before interaction with a photon is $W_{e\,\max}$, the maximum that they can have as a result of the collision is $W_{e\,\max} + h\nu$. If this energy is greater than the height of the potential barrier W_a, those electrons which happen to be moving toward the surface can escape. Thus, to obtain photoelectric emission, it is necessary that the frequency of the incident radiation be greater than the work function $W_a - W_{e\,\max} = e\phi_0$, divided by Planck's constant h. The maximum velocity of emission V_e in electron volts of the emitted electrons will, therefore, be given by the relation

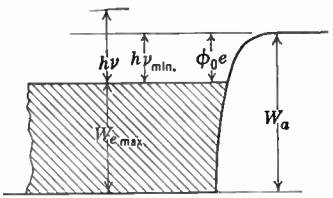


Fig. 1.19. Electron Energies Associated with Photoelectric Emission.

$$V_e = \frac{h\nu}{e} - \phi_0 \tag{1.27}$$

This relation has been verified by retardation methods and others for a great many pure surfaces and is found to be fulfilled within the accuracy of the assumptions made in deriving the expression. Chief among the assumptions are those that no electrons have energies greater than $W_{e\,\max}$ and that there are no multiple collisions between electrons and photons. Actually, some electrons are emitted with velocities greater than is indicated by Eq. 1.27, owing to the existence in the

metal of electrons with energies in excess of $W_{e\,\max}$. More exact values of the photoelectric work function are obtained by making use of the relation derived by Fowler which gives the photoelectric yield in terms of the retarding potential and temperature.

The wavelength of radiation for which $h\nu = e\phi_0$ is known as the photoelectric threshold, or long wavelength limit, and is characteristic of a given surface. Table 1.3 gives the long wavelength limit for a number of surfaces, the work function as determined by this threshold, and the thermionic work function.

TABLE 1.3. WORK FUNCTION OF METALS *

Metal	Long Wavelength Limit	Photoelectric Work Function	Thermionic Work Function
Ag	2610 Å	4.73	4.08
Al	3500 (approx.)	2.5 to 3.6	....
Au	2650	4.82	4.42
Ca	(4500)	(2.7)	2.24
Cs	6600	1.9	1.81
K	5500	1.76 to 2.25	....
Mg	3650	3.4	....
Pt	1962	6.3	6.27
W	2650	4.58	4.52
Zn	3720	3.32	....

* Data from Hughes and DuBridge, reference 11.

Like thermionic emission, the photoelectric emission is greatly increased and the long wavelength limit is raised when a surface is subjected to certain contaminations. This is particularly true when the superimposed layer is an alkali metal. Surfaces having the greatest photoelectric response are obtained, however, when the surface is a semiconductor rather than a pure metal. All the complex photoemitters known at present which have quantum efficiencies above 0.01 percent in the visible or near ultraviolet involve use of the alkali metals. Of these, cesium has given by far the best results. Most, if not all, of these complex surfaces, in addition to their high quantum yield, exhibit selective photoemission. A photoemitter is said to be a selective emitter when the quantum efficiency, instead of rising monotonically with decreasing wavelength of radiation over a wide range of wavelengths, shows maxima and minima of quantum efficiency. Furthermore, many of these emitters exhibit a difference in response depending upon whether the plane of polarization of the incident light lies in the plane of the emitter or normal to it.

The problem of obtaining a high photoelectric yield presents a number of considerations. First, the material should be capable of absorbing the radiation within a volume shallow enough to allow the escape of photoelectrons. This means that the material should be neither highly reflecting nor highly transparent. Second, the work function of the surface should be relatively low to increase the probability that a given electron-photon interaction will give an electron at least sufficient energy to cross the potential barrier. Third, collisions of the excited electrons which will lead to a dissipation of their energy before arriving at the surface should be rare. Finally, the probability of the absorption of photons by transitions which do not lead to electrons having the high kinetic energy required for escape should be small.

Conditions 1, 3, and 4 are met by materials which are neither metallic in their behavior nor true insulators. For the metals, the reflectivity is generally very high and, furthermore, only in a surface layer a few atoms thick can the absorption take place in such a way as to release photoelectrons. Semiconductors which may or may not be modified by the dispersal of colloidal metal seem to be best adapted to meet these conditions. The requirement of low work function, although less stringent than the three mentioned above, nevertheless seems to be met by the best photoemitters. The low work function usually is the result of a composite surface layer.

The two most important photosensitive surfaces from a practical standpoint are the cesium-antimony photoemitter and the cesium cesium-oxide silver surface. The first of these, namely, the cesium-antimony surface, has the highest quantum efficiency. At a wavelength corresponding to their peak of sensitivity, these cathodes may have quantum efficiencies of 10 and 20 percent or more. Under laboratory conditions, a cesium-antimony surface has been produced with as high as 30 percent quantum efficiency. The spectral response lies chiefly in the blue, the long wavelength threshold for a properly treated surface being about 6800 Å. Therefore, this type of surface, although having excellent white-light response, is not suitable where a cathode sensitive to red light is required. Figure 1.20 gives the spectral response of a typical cesium-antimony surface.

Depending upon the requirements of the tube in which it is to be used, this type of cathode may take either of two forms. The first is an opaque surface with the light falling on the same side from which electrons are emitted. The second is a semitransparent surface with

the exciting illumination falling on the back film through the supporting glass and the electrons emitted from the opposite side.

The first step in the preparation of an opaque film consists of evaporating a layer of antimony onto the backing metal. This layer is fairly thick, and the type of metal used for the underlying layer is relatively unimportant as long as its surface is very clean. Cesium is next introduced as a vapor into the cell. It is common practice to introduce an

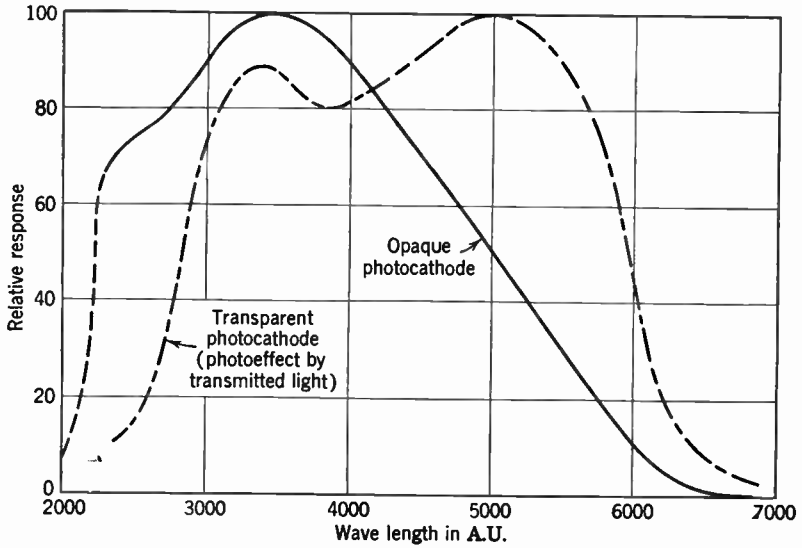


Fig. 1.20. Spectral Response of an Opaque and a Transparent Cesium-Antimony Photocathode. (Zworykin and Ramberg, *Photoelectricity*, Wiley, New York, 1949.)

excess of cesium. The cesium reacts with the antimony to form the compound $SbCs_3$. After the surface has been heated to a temperature between 150 and 200°C, the excess cesium may be distilled off by maintaining the whole tube at the above temperature with the exception of a small side tube which is made much colder so that the excess cesium distills into it.

A second procedure sometimes used as an alternative is to allow cesium to distill slowly into the tube while maintaining it at a temperature between 150 and 200°C. As the distillation progresses, the photoemission is monitored continuously. When the maximum photoemission is reached, the cesium flow is cut off.

To form the semitransparent type of photosensitive surface, a layer of antimony is deposited on the glass surface which is to form the back-

ing of the cathode. This layer should be about 30 atomic layers thick and will have an optical transmission of about 30 percent. By transmitted light, the film is very red in color. Cesium is next distilled into the tube while the tube is kept at a temperature of about 150°C. As cesium is taken up by the antimony film, the conductivity first falls. The surface also becomes more transparent. As the cesium continues to distill on the surface, the conductivity again rises. The photoelectric response which is quite low at the conductivity minimum also increases. The photo-response goes through a maximum as exposure to cesium is continued. As with the opaque cathode, the introduction of cesium may be stopped at the response maximum or it may be continued until there is a definite excess of cesium. The excess cesium is then removed by having the cathode and bulb slightly warmed and a side tube kept at a low temperature. Frequently the sensitivity can be improved by a very slight oxidation after the cesiation is complete.

Electrical contact to the film is usually made through a metal ring evaporated or otherwise deposited on the glass before the antimony is laid down. Sometimes, in addition to the metallic ring, a very thin transparent layer of metal such as palladium is deposited on the glass before the antimony is evaporated. Manganese oxide is another material that has frequently been used as an underlying layer for the antimony. However, the conductivity of a well-activated cesium-antimony surface is sufficient so that no underlying layer is required for most applications.

The activated cesium-antimony layer has the characteristics of a semiconductor. The variation of conductivity with temperature indicates carrier activation energies of 0.35 to 0.45 electron volt. The material is a thermionic emitter having a work function which ranges from 0.5 electron volt to 1.5 electron volts, depending upon such factors as the amount of oxygen which has been absorbed by the layer. The photoemission is thought to be a volume effect as should be the case with a semiconductor photoemitter. It has not been possible to work out a simple semiconductor energy diagram which will, however, account for the entire behavior of this type of photoemitter. Cesium-antimony surfaces are used in a great many practical devices. Many pickup tubes designed to have high sensitivity to white or blue light employ this surface. It is also widely used in phototubes and multiplier phototubes.

The other commonly used photosurface, the cesiated silver surface, is valuable because of its higher red response. This type of emitter is extremely sensitive in the red end of the visible spectrum, and its

sensitivity extends well down into the near infrared, the long wavelength threshold of an ordinary well-activated surface being 11,000 to 13,000 Å, whereas some surfaces have been reported to have a response extending down to 17,000 Å. The emission is selective, having a maximum at 8000 to 9000 Å, going through a minimum in the green, and then increasing again in the blue and violet regions of the spectrum. Figure 1.21 shows typical spectral response curves for this type of

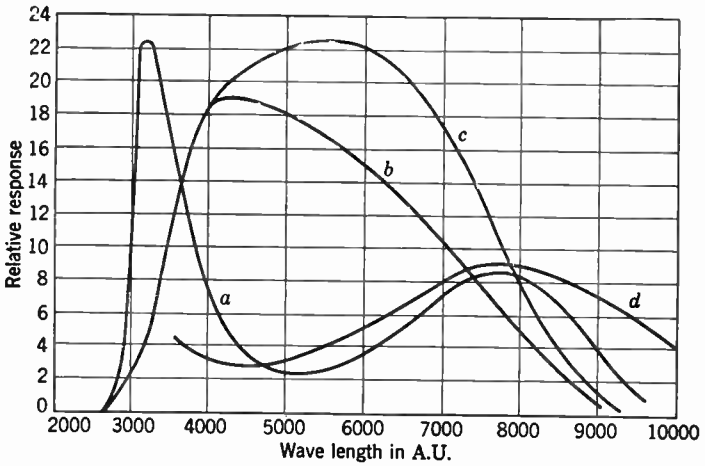


Fig. 1.21. Spectral Response of Differently Processed Silver-Cesium Oxide-Cesium Photocathodes. (Zworykin and Ramberg, *Photoelectricity*, Wiley, New York, 1949.)

photoemitter. The long wavelength end of the spectral response varies considerably with the exact activation schedule employed in preparing the cathode.

The preparation of this type of surface is as follows: A silver or a silver-coated sheet of metal is shaped to the form required for the cathode in question. As was the case for the cesium-antimony photocathode, cesiated silver cathodes may be formed on opaque metal surfaces or as semitransparent cathodes. The metal surface must be thoroughly clean, particular care being taken to remove all traces of organic material. The silver is next oxidized by admitting oxygen at a pressure of a few millimeters of mercury into the tube and passing a glow discharge between the cathode and any other convenient electrode. As the surface oxidizes, it will be found to go through a series of colors in the following order: yellow, red, blue, yellow, red, blue-green (or second blue), green. In general, the oxidation should be

stopped with the second blue or first green. Cesium is then admitted to the tube and the tube is baked for a few minutes at about 200°C. If the correct amount of cesium has been added, the color of the finished surface will be a brownish-yellow. Too much cesium is indicated by an almost white appearance of the surface; an insufficient amount of cesium usually causes the layer to appear a dark gray-green. Unlike the cesium-antimony surface, if an excess of cesium is introduced into the tube, it cannot be removed from the cathode without damaging the surface.

The sensitivity of this type of emitter can be further increased if a very thin layer of silver is evaporated over the surface of a cathode prepared as described above and the emitter again baked at 200°C. This treatment also tends to increase the long wavelength response as well as the white-light sensitivity.

The semitransparent form of the cesiated silver cathode is formed by a procedure which is somewhat similar to that outlined above. A thin film having a transmission of about 30% for white light is evaporated onto the transparent support of the cathode. This silver film is then subjected to an oxygen discharge until it becomes almost completely transparent. Cesium is next introduced and the tube is baked. Like the opaque cesiated silver cathode, this surface can be improved by evaporating a thin layer of silver over the activated film. A film prepared in this way will have a photosensitivity of 10 to 30 microamperes per lumen and a long wavelength threshold at about 12,000 Å.

Like cesium-antimony surfaces, cesium-silver photocathodes are used in diverse light-sensitive devices. They include both the ordinary hard phototube and gas phototubes, where an atmosphere of an inert gas at low pressure (100 microns of mercury) makes possible gas amplification of the photocurrent. An additional application of the cesium-silver surface is in image tubes used for infrared viewing. The infrared image tubes, which will be described in more detail in a later chapter, found considerable application in the Second World War. Red and infrared sensitive photomultipliers and pickup tubes are also included among these applications.

Useful but somewhat less efficient cathodes can be formed if rubidium takes the place of cesium, or other base materials, such as bismuth-oxygen-silver, are used with cesium. Such photosurfaces are employed when a different spectral response is required (e.g., the bismuth surface can be used to extend the red response of the cesium-antimony photoemitter, whereas rubidium, replacing cesium, decreases the infrared response of the silver-silver oxide emitter). Where only blue or

violet response is desired and when the thermionic emission must be low (e.g., for photomultipliers in scintillation counters), lithium-potassium antimony photocathodes appear to have promising properties.

1.9 Secondary Emission. A third way in which electrons can be given the energy necessary for their escape from a solid is through the interaction between them and high-velocity electrons bombarding the surface in question. The name secondary emission has been given to this phenomenon.

Secondary emission, like photoemission, is produced by an external excitation. It differs, however, in that each incident electron, unlike the incident photons which produce photoemission, may cause the release of more than one electron, the average number of electrons released by each primary being known as the secondary-emission ratio. This ratio may range from less than one electron per primary to more than 10 or 15, depending upon the surface bombarded and the conditions of bombardment.

Materials with the lowest secondary emission ratio generally take the form of very porous mat surfaces. Examples of this are carbon black and aluminum smoke on metal backing plates. Maximum secondary emission ratios of 1/2 or less may be obtained from such surfaces. Here, it is plausible that the mechanical trapping of the emitted electrons by the irregularity of the surface is the principal cause of the low secondary emission ratio.

Clean metal surfaces have a maximum secondary emission ratio in the neighborhood of unity. Whenever metals appear to have ratios as high as 4 or 5, this can be attributed to surface contamination. Calcium, beryllium, barium, and the alkaline metals are examples of these apparent exceptions. A very small amount of surface contamination will raise the secondary emission ratio to a high value. For example, a layer of oxygen alone is sufficient to double or triple the emission from such metals as magnesium, aluminum, or beryllium.

Complex surfaces similar to those used as photoemitters, consisting of silver, cesium-oxide and cesium; silver, rubidium-oxide and rubidium; or cesium-antimony are among those exhibiting the highest ratio, being capable of emitting more than 10 secondary electrons per primary at the optimum bombarding voltages. A silver-magnesium alloy, given proper oxidation and thermal treatment, also may have a very high secondary-emission ratio. Very thin layers of alkali halides on certain metals have also been observed to have very high ratios, even higher than those of the surfaces mentioned above, although such

surfaces are quite unstable and do not retain their property of high yield for more than a few minutes of bombardment by electrons.

The velocity of the bombarding electrons is important in determining the secondary emission ratio of a given surface. In general, the ratio is small for low voltages, increasing with an increase in voltage to a maximum at 300 to 600 volts and then falling slowly with further increase of the bombarding voltage. The variation of secondary emission ratio with bombarding voltage for a number of representative surfaces is shown in Fig. 1.22.

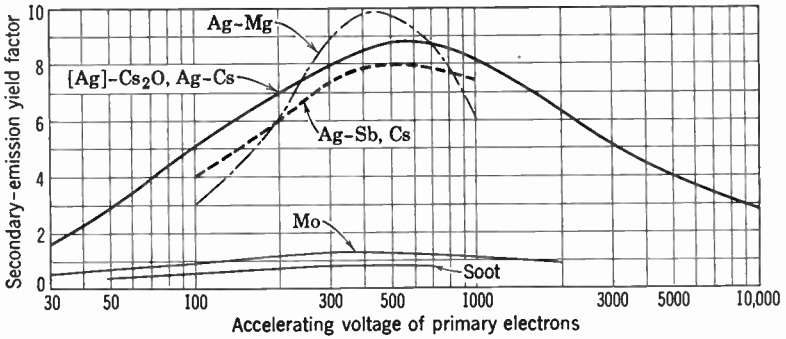


Fig. 1.22. Secondary Emission Ratio of Representative Surfaces.

Apart from the velocity of the bombarding electrons and the nature of the surface, the secondary emission ratio depends upon the angle of incidence of the primary beam. At normal incidence, the ratio is a minimum; it increases as the angle tends toward grazing incidence. The rate of increase of yield with angle of incidence is greater for high voltage primary electrons than for low. It should be pointed out that this variation with angle can be observed only for very smooth, highly polished surfaces. The surfaces used in devices employing practical secondary emitters are, in general, so rough on a microscopic scale that they exhibit little or no angular dependence of secondary emission ratio.

Observations on the velocity of emission of secondary electrons reveal that the electrons may be conveniently classified in three groups. First, there is a group of electrons which leave the surface with essentially the same velocity as that of the primary electrons. These electrons are often spoken of as reflected electrons. Except at very low bombarding voltages, this first group constitutes, in general, but a small fraction of the total emission. Second, there is a group of low velocity

electrons and, finally, there is a background of electrons with velocities distributed more or less uniformly between the two. The low-velocity group makes up the bulk of the emission for bombarding voltages in the range from about ten to several thousand volts and is the primary concern of the present discussion. A typical velocity distribution curve is given in Fig. 1.23, pointing out the three ranges of emission.

For the pure metals, where the secondary emission ratio is relatively low, the distribution rises sharply to a maximum at quite low velocities

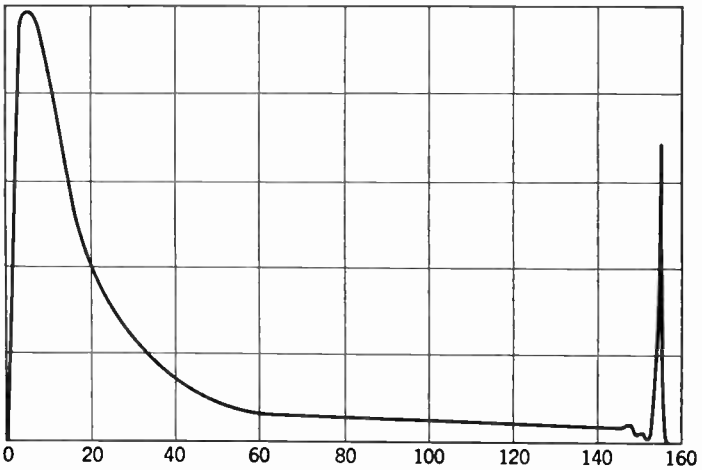


Fig. 1.23. Secondary Electron Velocity Distribution from Gold. (After Rudberg in McKay, reference 24; courtesy Academic Press, Inc., New York.)

(usually of the order of 2 volts) and then decreases slowly until it merges with the background at 20 to 40 volts. The velocity distribution for molybdenum is shown in Fig. 1.24. The velocity distribution of the more efficient emitters is much narrower than that of the pure metals. The electron-optical properties, equilibrium potential determinations, and retardation measurements indicate that more than 85 percent of the emitted electrons from a good cesiated silver emitter (ratio 8 to 10) have velocities under 3 volts. Silver magnesium surfaces in the same secondary emission range have a slightly wider velocity distribution. In general, the shape of the distribution curve of a good emitter resembles somewhat that of a Maxwellian energy curve but differs from it in detail.

A number of theories and models have been brought forward to account for the phenomenon of secondary emission. Among them are important proposals by Fröhlich, Wooldridge, and others. None of

them appears to account for the phenomenon in complete detail, particularly as applied to the more complex practical emitters.

Like the previously described forms of emission, secondary emission requires that electrons in the material be given a velocity which has a normally directed component of sufficient magnitude to permit their escape through the surface potential barrier. This energy is obtained through a collision or interaction between the in-coming electrons and electrons in the material. Obviously, since secondary emission ratios

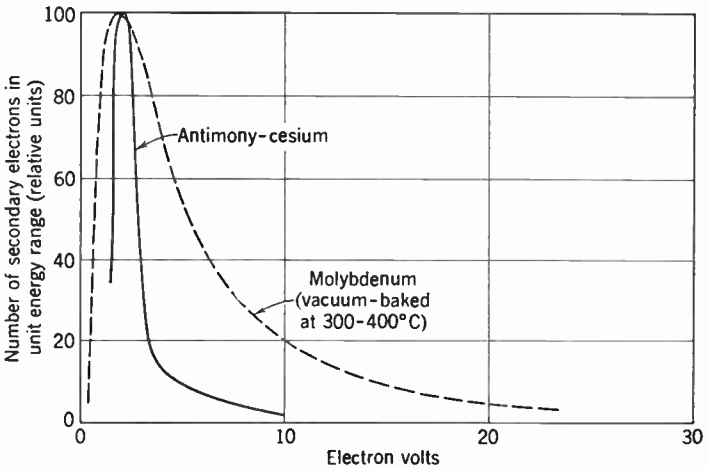


Fig. 1.24. Velocity Distribution of Secondary Electrons from Molybdenum and Antimony-Cesium. (After Kollath, reference 25.)

of 10 or more can be obtained, an incident electron must give its energy directly or indirectly to many electrons. Probably only a small fraction of the electrons which receive this energy actually find their way to the surface.

The assumption of a collision between an incident electron and the free electrons in a metal cannot account for the latter acquiring velocities in a direction which would permit their escape. This is so, of course, because in an elastic collision between two electrons momentum must be conserved, just as it is for any other type of particle. The momentum vector of the electron struck must, therefore, make an angle of less than 90 degrees with that representing the momentum of the incident electron before the collision. The experimentally observed initial velocities of the emitted electrons have components of motion opposite to that of the bombarding electrons and thus necessitate the assumption that a third element is involved in the collision. A number

of mechanisms will suffice to account for motion away from the surface, for example, multiple collisions, interactions between the lattice and the emitted electrons, or collisions between incident electrons and electrons which are bound to the lattice.

The higher secondary emission ratio of a complex (possibly semiconductor) surface over that of a pure metal may be the result of one or more of several effects. First, the complex material may provide a greater number of electrons with bindings that permit the required reversal of momentum. Second, the excited electrons may be able to move more freely and for a greater distance in such materials and thus reach the surface from a greater depth; and, finally, there may be a lowering of the surface potential barriers permitting more ready escape of the electrons. It might be pointed out in connection with this last possibility that there are many examples of materials where a higher secondary emission ratio is obtained when the surface contaminant increases the work function. For example, a pure tungsten surface whose work function is 4.52 volts has a maximum ratio of 1.5 or less. When such a surface is contaminated with a layer of oxygen, which raises the work function to over 6 volts, the secondary emission ratio may be almost doubled. Oxygen on aluminum, beryllium, or magnesium acts in a similar way. The contamination of the surface with a material which definitely lowers the work function also greatly increases the ratio. The effect in this case may be twofold, both due to an increased interaction between electrons and to the reduction in work function.

Surfaces consisting of oxidized silver treated with cesium give a very high yield. The final step in the activation of such a surface is exposure to cesium. Therefore, there is a layer of uncombined alkali present giving the characteristic low work function. If, after activation, the surface is exposed to oxygen, thus eliminating the free cesium and increasing the work function, the emission ratio drops by nearly 50 percent. The lowering of the ratio in this instance may be primarily due to the increased work function rather than any major change in the electron interaction or penetration.

The relation between the excitation of secondary electrons and the velocity of the incident electrons is, at least qualitatively, understandable. Each excited electron removes some energy from the incident electron, so that, if the latter has a given amount of energy, it can excite only a limited number of electrons in the bombarded material. Therefore, the total number excited by a bombarded electron increases with the bombarding voltage. However, the probability of excitation per unit length of path of the incident electron decreases with increas-

ing voltage. At low voltage, the incoming electron loses its energy very near the surface, and the excited electrons with velocities of the required magnitude and direction do not have far to travel in the material before escaping. In this velocity range, therefore, the yield increases with voltage. At high voltages, the incident beam penetrates farther into the emitter and, although the total number of excited electrons produced is greater, most of them originate at depths too great to permit their escape. Hence, above a certain optimum bombarding voltage, the emission ratio decreases. If the high-velocity electrons enter the surface at a small angle, the depth of origin of the excited electrons is less than that for normal incidence, and consequently the secondary-emission ratio is greater under these circumstances.

The preparation of both the silver-cesium and the cesium-antimony secondary emitting surfaces follows closely the procedure used in producing the corresponding photoemitters. For a silver-cesium oxide-cesium secondary emitter, the target to be processed may be made of silver or of some other metal plated with a heavy layer of silver. The silver should be free from contamination and thoroughly cleaned before the activation is started. The surface is first oxidized, by the same procedure employed for the cesiated silver photoemitter. After the tube has been pumped free from oxygen, cesium is admitted. The amount of cesium needed to obtain maximum yield is somewhat less than that required to produce an optimum photoemitter, the quantity required being dependent upon the degree of oxidation. The tube is next baked at 200°C to promote a reaction between the silver oxide and the cesium. The activated surface thus formed should have a brown color. The secondary-emission ratio, unlike the photoelectric response, is not increased by the deposition of additional silver. Such a procedure, in fact, reduces the secondary-emission ratio to less than half its optimum value.

The cesium-antimony secondary emitter is formed by depositing a layer of antimony upon the metal surface which is to form the dynode or target. After the antimony layer has been deposited, the surface may be exposed to air if precautions are taken to have the humidity low and the atmosphere free from organic and other contaminants. The metal coated with the antimony surface is assembled in the final tube and the tube is evacuated to a good vacuum. The out-gassing bake of the tube, after the target has been mounted in it, must be carried out at a relatively low temperature, since, if the temperature

exceeds 230°C, evaporation of the antimony layer will take place. After the tube has been well exhausted, cesium is admitted and the tube is baked in the same manner as in forming the corresponding photosensitive surface.

Silver-magnesium alloy, when properly treated, forms a third practical secondary emitter. This type of emitter is prepared as follows. A known weight of silver is melted in a suitable crucible. Next, a piece of magnesium held on a steel wire, the magnesium weighing about 10 percent of the weight of silver, is thrust into the molten silver, care being taken that it is completely submerged. The melt is then stirred long enough for thorough mixing of the magnesium and silver but not long enough to allow a large loss of magnesium from the melt. After the material is cooled, the ingot is cut to the proper size and rolled into a convenient thickness. After the surface has been cleaned and polished, the target is mounted in the vacuum tube. The tube is then heated to 400°C for a period of an hour or more. Oxygen, occluded in the melt, diffuses to the surface and forms a thin layer of magnesium oxide which is responsible for the high secondary-emission ratio. Sometimes the secondary-emission ratio is improved by high-frequency heating of the target in the presence of a small amount of oxygen. A properly activated silver-magnesium surface will have a ratio of 4 to 6 at 200 volts. Its appearance will be a golden yellow color. Such a target is not damaged when used in a tube where a photocathode is activated with cesium vapor. In fact, in general, the secondary emission ratio is improved upon exposure to cesium.

An oxidized copper-beryllium alloy also forms a useful secondary emitter. Although the secondary-emission ratio of this surface is lower than any of those discussed above, the surface, nevertheless, is very stable and finds application in demountable vacuum tubes.

Insulators and materials of very low conductivity will also emit secondary electrons. The range of the maximum secondary-emission ratio is about the same as that of conducting materials, extending from 10 to 15 for some surfaces down to less than unity for certain oxides. The ratio varies with the velocity of the incident electrons in much the same way as it does for conducting emitters.

When an insulator is bombarded with electrons, its surface must assume a potential such that the net current to or from it is zero, since electrons will not flow through the material itself to compensate for any change in number at the surface. Two potential equilibria fulfill this condition. The first is for the surface to become so negative that

the primary electrons cannot reach it. The other is for it to assume a potential which is slightly positive with respect to the element collecting the electrons, so that the secondary emission current is not saturated and the current which actually leaves the surface is just equal to the primary current. The latter is possible only when the saturated secondary-emission ratio is greater than unity. Finally, for very high collector voltages (referred to the emitting cathode), surfaces with a maximum secondary-emission ratio greater than unity will tend to assume a potential which reduces the bombarding voltage to the higher one of the two values for which the secondary-emission ratio is unity.

Because current cannot flow through a target made of insulating material, measurements of the secondary-emission ratio must be made by ballistic methods. In making these measurements, advantage can be taken of the capacitance between the surface of the target and a conducting surface backing it up.

Directly or indirectly, the secondary-emission properties of dielectrics are of considerable practical interest. The secondary-emission ratio of the glass walls of electronic devices often plays a role in their operation. The phosphors used in cathode-ray oscilloscopes and television viewing tubes to transform the energy of primary electrons into light are nonconductors. Their secondary-emission properties are, therefore, important in determining their performance, as will be discussed in the next chapter.

Various types of storage and memory tubes depend for their action on the secondary-emission ratio of insulators.

1.10 Statistics of Secondary Emission. Secondary emission is used in a variety of electronic devices to intensify an electron current. Its value for this purpose depends to a large extent upon its noise properties. Therefore, this phase of the phenomenon warrants rather detailed discussion.

Like the shot noise from a thermionic or photoelectric emitter, the noise from secondary emission is a statistical effect. If certain assumptions are made concerning the way in which electrons are emitted, the magnitude of the fluctuation can be calculated.

In striking the secondary-emissive surface, each primary electron causes the release, on the average, of σ electrons, where σ is the secondary-emission ratio. However, this does not mean that each individual incident electron causes the emission of exactly σ secondary electrons, although this possibility cannot be excluded a priori from consideration. In general, there will be a certain probability $p(z)$ that

a given primary electron will produce z electrons. On this basis, the probability that n incident electrons produce N secondaries will be

$$P(N, n) = \sum_{\Sigma z_i = N} \prod_{i=1}^{i=n} p(z_i). \tag{1.28}$$

The primary electrons, whatever their source, will be subject to some kind of fluctuation, and therefore the number, n, arriving in a unit of time must be put on a probability basis. Let $P(n)$ be the probability that exactly n electrons arrive during a given unit of time, and let $\bar{n}$ be the average number per unit time. The probability that N electrons will leave the surface then becomes

$$P(N) = \int_0^\infty P(n)P(N, n)\, dn \tag{1.29}$$

Finally, the mean square deviation from the expected number of secondary electrons $\sigma\bar{n}$ will be

$$\overline{\Delta N^2} = \sum_{N=0}^{\infty} P(N)(N - \sigma\bar{n})^2 \tag{1.30}$$

This mean square deviation in the number of electrons emitted per unit time can, of course, be related to a mean square fluctuation current over any desired frequency band by

$$\overline{i^2} = 2e^2\, \overline{\Delta N^2}(f_2 - f_1) \tag{1.31}$$

as was shown in an earlier section.

Before this expression representing the noise effective over a given frequency band can be evaluated, the various probabilities involved in its derivation must be determined. The probability $P(n)$ for the incident beam depends upon the electron source. If it is a photoelectric cathode or an emitter of similar nature, the electron arrival will be purely random; in other words, the probability $P(n)$ can be expressed by the normal law,

$$P(n) = \frac{1}{\sqrt{2\pi\bar{n}}}\, \epsilon^{-(n-\bar{n})^2/(2\bar{n})} \tag{1.32}$$

The values assigned to the probabilities $p(z)$ depend on the assumptions made with regard to the mechanism of secondary emission.

One assumption, and perhaps the simplest, is that

$$p(z) = 1 \qquad z = \sigma$$
$$p(z) = 0 \qquad z \neq \sigma$$

or, in other words, that every primary releases just σ secondaries. From this, it follows that

$$P(N, n) = 1 \qquad N = \sigma n$$
$$P(N, n) = 0 \qquad N \neq \sigma n$$

Equation 1.29 can now be written

$$P(N) = \frac{1}{\sqrt{2\pi\bar{n}\sigma^2}} \epsilon^{-(N-\sigma\bar{n})^2/(2\sigma^2\bar{n})} \tag{1.33}$$

The mean square deviation, according to the ordinary principles of probability, is

$$\overline{\Delta N^2} = \sigma^2\bar{n} \tag{1.34}$$

or, in terms of current fluctuation,

$$\overline{i_n^2} = 2e\sigma^2 i_0(f_2 - f_1) \tag{1.35}$$

where i_0 is the primary current.

This result can be obtained without resorting to the above calculation by merely considering that the expected noise is that due to the random emission of particles of charge $e\sigma$, producing a current $i = i_0\sigma$. The current fluctuations predicted by Eq. 1.35, however, do not fit the data derived from noise measurements, indicating that the assumptions used in its derivation are not valid.

A more reasonable assumption is that secondary emission itself is a random phenomenon. If the primary beam entering the target material excites a large number of electrons, only a few of which have velocities such that they are able to escape, it is not unreasonable to postulate the following two conditions:

1. The probability p of emission of any one of the m electrons excited by the beam is small compared with unity.
2. The emission of a secondary electron does not appreciably alter the probability of emission of the remaining excited electrons.

Assuming that these two conditions are fulfilled, the binomial theorem of statistics expresses the probability $p(z)$ that z electrons will be produced by a given incident electron. Therefore,

$$p(z) = \frac{m!}{z!(m-z)!} p^z (1-p)^{m-z} \tag{1.36}$$

where the expectation mp is, of course, equal to σ.

The probability of obtaining N secondary electrons from n primary electrons, which can be found with the aid of Eqs. 1.36 and 1.28, is

$$P(N, n) = \frac{(nm)!}{(nm-N)!N!} p^N (1-p)^{nm-N} \tag{1.37}$$

Since both mn and N are large and p is small compared with unity, Eq. 1.37 can, by means of Stirling's formula, be reduced to

$$P(N, n) = \frac{1}{\sqrt{2\pi n\sigma}} \epsilon^{-(N-n\sigma)^2/(2n\sigma)} \tag{1.38}$$

Again, by making use of a perfectly random primary electron beam (i.e., Eqs. 1.32 and 1.38 in Eq. 1.29), the probability of obtaining N secondary electrons from an average of $\bar{n}$ incident electrons will be

$$P(N) = \int_0^\infty \frac{1}{\sqrt{2\pi\bar{n}}} \epsilon^{-(n-\bar{n})^2/2\bar{n}} \frac{1}{\sqrt{2\pi n\sigma}} \epsilon^{-(N-n\sigma)^2/(2n\sigma)}\, dn$$

$$= \frac{1}{\sqrt{2\pi\bar{n}(\sigma+\sigma^2)}} \epsilon^{-(N-\bar{n}\sigma)^2/[2\bar{n}(\sigma+\sigma^2)]} \tag{1.39}$$

Therefore, the mean square deviation is

$$\overline{\Delta N^2} = \bar{n}(\sigma^2 + \sigma) \tag{1.40}$$

and, by the same reasoning as before, the mean square current fluctuation over a frequency band $(f_2 - f_1)$ when the primary current i_0 strikes the target is expressed by

$$\overline{i_n^2} = 2e(\sigma^2 + \sigma) i_0 (f_2 - f_1) \tag{1.41}$$

This relation is in close agreement with results obtained from efficient emitting surfaces operated at voltages below that required to give maximum yield.

Comparing Eqs. 1.19 and 1.41 will show that Eq. 1.41 consists of two parts, one giving the multiplied noise in the primary current, the other, which has the form

$$\overline{i_{n'}^2} = 2e\sigma i_0 (f_2 - f_1) \tag{1.42}$$

expressing the noise generated by secondary emission itself.

For poor emitters, or good emitters operated at high voltages, this latter portion of the noise is increased, so that generally it should be written as

$$\overline{i_n'^2} = 2e\varepsilon\sigma i_0(f_2 - f_1) \tag{1.42a}$$

where ε has a value close to unity for good emitters operated under conditions corresponding to those found in practice where secondary emission is used for current intensification.*

Where secondary emission is employed to obtain large amplification of small currents, the device takes the form of a secondary emission

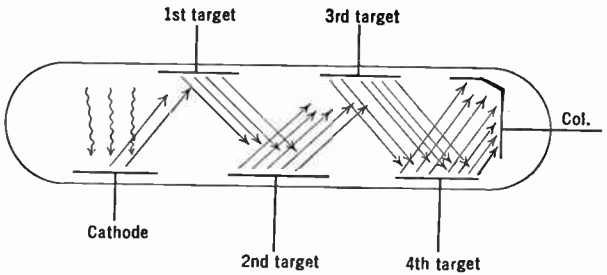


Fig. 1.25. Secondary Emission Multiplier (Schematic).

multiplier. It consists of a series of cascaded stages, each intensifying the current from the preceding stage. Diagrammatically, a multiplier phototube can be represented as shown in Fig. 1.25.

Electrons originate as photoelectrons from the cathode. They are directed by an electron optical system onto the first dynode, where they produce secondary electrons. This secondary emission is directed onto the second dynode and, in turn, generates more electrons. If the initial current in the multiplier is i_0, the output current will be

$$i = i_0\sigma^k \tag{1.43}$$

where σ is the secondary emission ratio of each dynode and k the number of dynodes. The noise contributed to the output can be calculated for each successive stage, as shown in the following table, and the total noise found by summation.

* The factor ε corresponds to p used in Zworykin, Morton, and Malter, reference 29, and to bm in Shockley and Pierce, reference 30.

TABLE 1.4

Noise from photocathode:	$i_{n0}^2 = \sigma^{2k}2e(f_2 - f_1)i_0 = \sigma^k 2e(f_2 - f_1)i$
Noise from first target:	$i_{n1}^2 = \sigma^{2(k-1)}2e(f_2 - f_1)\sigma i_0 = \sigma^{k-1}2e(f_2 - f_1)i$
Noise from second target:	$i_{n2}^2 = \sigma^{2(k-2)}2e(f_2 - f_1)\sigma^2 i_0 = \sigma^{k-2}2e(f_2 - f_1)i$
. . . . .	. . . . .
Noise from kth target:	$i_{nk}^2 = 2e(f_2 - f_1)\sigma^k i_0 = 2e(f_2 - f_1)i$

The total noise is, therefore,

$$i_n^2 = (\sigma^k + \sigma^{k-1} + \sigma^{k-2} \cdots + 1)2e(f_2 - f_1)i$$
$$= \frac{\sigma^{k+1} - 1}{\sigma - 1} 2e(f_2 - f_1)i \cong \frac{\sigma^{k+1}}{\sigma - 1} 2e(f_2 - f_1)i \qquad (1.44)$$

If Eq. 1.44 is derived using Eq. 1.42*a* instead of Eq. 1.42, the result has the form

$$i_n^2 \cong 2e(f_2 - f_1)\left(1 + \frac{\varepsilon}{\sigma - 1}\right) i\sigma^k \qquad (1.45)$$

where the number of stages is very large.

Noise measurements on secondary emission multipliers and pulse height resolution determinations with photomultipliers used for scintillation counting indicate that ε has a value in the neighborhood of 1.5.

The excellent noise characteristics of the secondary emission multiplier, together with its very high speed of response, makes it a very valuable, practical tool in many fields of electronics.

REFERENCES

1. F. K. Richtmyer and E. H. Kennard, *Introduction to Modern Physics,* McGraw-Hill, New York, 1947.
2. F. Seitz, *The Modern Theory of Solids,* McGraw-Hill, New York, 1940.
3. H. Fröhlich, *Elektronentheorie der Metalle,* Springer, Berlin, 1936.
4. D. A. Wright, *Semiconductors,* Wiley, New York, 1951.
5. N. F. Mott and R. W. Gurney, *Electronic Processes in Ionic Crystals,* Clarendon Press, Oxford, 1940.
6. W. Shockley, *Electrons and Holes in Semiconductors,* Van Nostrand, New York, 1950.
7. A. L. Reimann, *Thermionic Emission,* Wiley, New York, 1934.
8. J. H. de Boer, *Electron Emission and Adsorption Phenomena* (Cambridge University Press), Macmillan, New York, 1935.
9. G. Hermann and S. Wagener, *Die Oxydkathode,* Barth, Leipzig, 1943–44.
10. V. K. Zworykin and E. G. Ramberg, *Photoelectricity and Its Application,* Wiley, New York, 1949.

11. A. L. Hughes and L. A. DuBridge, *Photoelectric Phenomena,* McGraw-Hill, New York, 1932.
12. T. S. Moss, *Photoconductivity in the Elements,* Academic Press, New York, 1952.
13. H. Bruining, *Die Sekundär-Elektronen-Emission fester Körper,* Springer, Berlin, 1942.
14. L. R. Koller, *Physics of Electron Tubes,* McGraw-Hill, New York, 1937.
15. J. C. Slater, "Electronic Structure of Metals," *Revs. Mod. Phys.,* Vol. 6, pp. 209–280, 1934.
16. A. Rose, "An Outline of Some Photoconductive Processes," *RCA Rev.,* Vol. 12, pp. 362–414, 1951.
17. E. M. Conwell, "Properties of Silicon and Germanium," *Proc. I.R.E.,* Vol. 40, pp. 1327–1337, 1952.
18. C. Herring and M. H. Nichols, "Thermionic Emission," *Revs. Mod. Phys.,* Vol. 21, pp. 185–270, 1949.
19. A. S. Eisenstein, "Oxide-Coated Cathodes," *Advances in Electronics,* Vol. 1, pp. 1–64, Academic Press, New York, 1948.
20. L. S. Nergaard, "Studies of the Oxide Cathode," *RCA Rev.,* Vol. 13, pp. 464–545, 1952.
21. W. Schottky, "Spontaneous Current Fluctuation in Emitted Electron Streams," *Ann. Physik,* Vol. 57, pp. 541–567, 1918.
22. T. C. Fry, "Theory of the Schrot-Effect," *J. Franklin Inst.,* Vol. 199, pp. 203–220, 1937.
23. R. Kollath, "The Secondary Emission of Solid Bodies," *Physik. Z.,* Vol. 38, pp. 202–224, 1937.
24. K. G. McKay, "Secondary Electron Emission," *Advances in Electronics,* Vol. 1, pp. 65–130, 1948.
25. R. Kollath, "On the Energy Distribution of Secondary Electrons," *Ann. Physik,* Vol. 39, pp. 59–80, 1941; Vol. 1, pp. 357–380, 1947.
26. H. Fröhlich, "Theory of the Secondary Emission of Metals," *Ann. Physik,* Vol. 13, pp. 229–248, 1932.
27. D. E. Wooldridge, "Theory of Secondary Emission," *Phys. Rev.,* Vol. 56, pp. 562–578, 1939.
28. V. K. Zworykin, J. E. Ruedy, and E. W. Pike, "Silver-Magnesium Alloy as Secondary-Emitting Material," *J. Appl. Phys.,* Vol. 12. pp. 696–698, 1941.
29. V. K. Zworykin, G. A. Morton, and L. Malter, "The Secondary Emission Multiplier—A New Electronic Device," *Proc. I.R.E.,* Vol. 24, pp. 351–375, 1936.
30. W. Shockley and J. R. Pierce, "A Theory of Noise for Electron Multipliers," *Proc. I.R.E.,* Vol. 26, pp. 321–332, 1938.
31. G. A. Morton, "The Scintillation Counter," *Advances in Electronics,* Vol. 4, pp. 69–107. 1952.

2 Fluorescent Materials

Strictly speaking, the title of this chapter is a misnomer. More exactly, the phenomenon to be discussed is cathodoluminescence, and the materials which manifest this effect are known by the general name of phosphors. However, the use of the term "fluorescent materials," instead of the more exact name, is so widespread that the exact name seems almost an affectation.

2.1 Luminescence. The conversion of energy invisible to the unaided eye into visible light is a very common phenomenon. When this light exceeds that produced by a black body at the actual temperature of the emitter, it is known as luminescence. Luminescence may be classified on the basis of the means of excitation. Table 2.1 lists

TABLE 2.1. CLASSIFICATION OF LUMINESCENCE

Class	Excitation
Photoluminescence	Radiant energy; visible part of spectrum; x-rays, ultraviolet
Cathodoluminescence	Electron bombardment
Radioluminescence	Radiation from radioactive sources
Triboluminescence	Disruption of crystals, as by grinding, abrasion, etc.
Bioluminescence	Biochemical reactions
Chemiluminescence	Chemical reactions
Electroluminescence	Alternating electric field

the more important classes, together with these means. Furthermore, luminescence—irrespective of the method of excitation—is divided into fluorescence and phosphorescence. Expressed inexactly, fluorescence is luminescence which ceases almost immediately upon removal of the excitation, whereas in phosphorescence the luminescence persists. It is difficult to give a more rigorous definition of these terms because, as yet, there is no generally accepted criterion to distinguish between them. The most commonly used time limit for the duration of fluorescence after cessation of excitation is 10^{-8} second.

Cathodoluminescence, as can be seen from Table 2.1, is excited by bombarding the material with electrons. There is a vast number of substances which have the property of luminescing when thus bombarded. In fact, nearly all inorganic compounds which are nonmetallic crystals exhibit this effect to some extent, together with a great many organic compounds, nonmetallic elements, and glasses. However, for most materials the efficiency of the production of light is extremely low. They are, therefore, of no interest in the present consideration, which will be restricted to the relatively few materials, natural or synthetic, that are highly efficient phosphors.

2.2 Requirements of a Phosphor. In television viewing screens, the phosphor serves to convert the energy of a scanning beam into the light which makes up the reproduced picture.

One of the prime requisites of the screen is that it be an efficient converter of energy. In addition to meeting this requirement, it must have the following properties:

The ability to produce high, instantaneous, intrinsic brightness.
Stability and long life under electron bombardment.
Suitable electrical properties.
Low vapor pressure.
Suitable color characteristics.
A duration of phosphorescence of the same order as, but not longer than, a picture period.

The foregoing seven items are essential, but certain other properties are highly desirable. For example, the luminous output should vary approximately linearly with current over a wide range, the material should be easy to apply, and its preparation should be commercially practical.

Until quite recently very little was known about the synthesis of phosphors. Early investigators in the field of cathode-ray television were forced to rely largely on naturally occurring materials such as willemite and zinc blende. It is now possible, however, to synthesize phosphors which are far superior to the best naturally occurring material in every respect. Therefore, synthetic substances will be the subject of most of what is to follow.

2.3 The Nature of Inorganic Phosphors. The great majority of the inorganic phosphors are crystalline in nature; the fluorescent glasses, containing uranium or rare-earth oxides, are relatively inefficient and will not be considered further here. Among these crystalline solids, some exhibit strong luminescence in their pure state. This statement

must, of course, be qualified because there is no measure of absolute purity; however, the luminescence continues unchanged in these solids even when they are purified to such an extent that the spectrograph fails to reveal the presence of foreign material. Barium platinocyanide, and calcium tungstate, for example, are such substances.

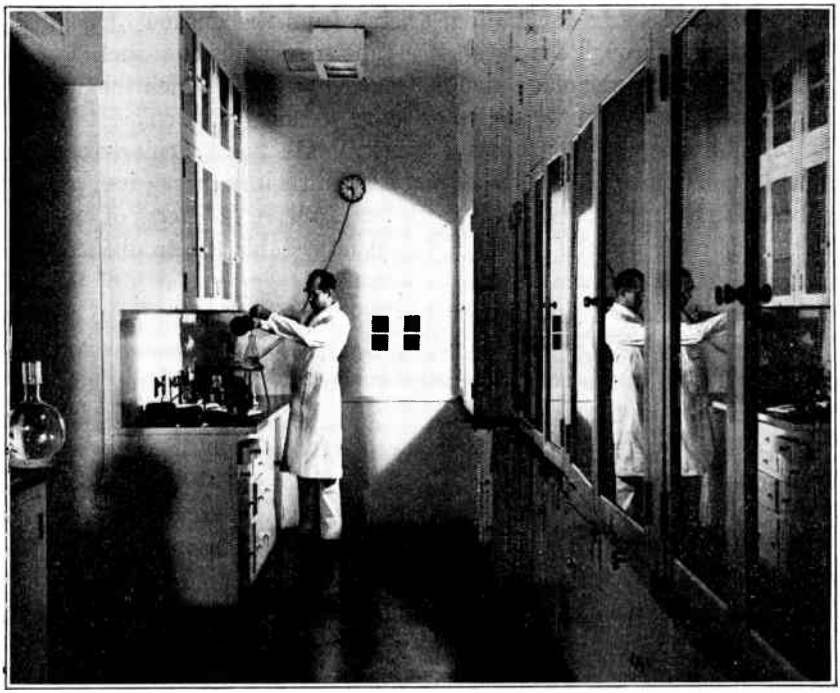


Fig. 2.1. A Modern Research Laboratory for the Investigation of Luminescent Materials (RCA Laboratories, Princeton, N. J.).

Most phosphors depend for their luminescence upon the presence of an "impurity," or activator. The activator influences not only the efficiency of the phosphor but also the color of its luminescence and the duration of phosphorescence. Therefore, by controlling the quantity and nature of the activator during the synthesis of the material, the desired properties can be obtained. The flexibility of this second class of phosphors has resulted in its wide use for practical television purposes.

Three components enter, in general, into the preparation of phosphors with activators. They are the base material, a flux, and finally the activator.

The base, which is invariably crystalline in nature, is generally colorless and conducts, if at all, as an electronic semiconductor. The sulphides and selenides, silicates and tungstates, and oxides and fluorides of the elements in the second column of the Periodic Table, principally zinc, cadmium, calcium, strontium, magnesium, and beryllium, are representative phosphor bases. They uniformly exhibit *n*-type or excess semiconduction, arising from the motion of conduction electrons through the crystal; *p*-type semiconductors such as ferrous oxide, nickel monoxide, and cuprous chloride, in which the current is carried by electron vacancies or holes, do not form efficient bases. In brief, good phosphor bases commonly are nearly colorless compounds which have their cations combined in their highest (and, preferably, only) valence states, and which tend to have an excess of cations. In addition, the material must be sufficiently stable not to disintegrate under the electron beam.

The second component, the flux, is not essential in the preparation of all types of phosphors. This material, usually an alkali or alkaline-earth halide, plays a role primarily during the crystallizing process. The flux aids in the crystallization of the substance and is largely removed subsequent to the final heating, after it has fulfilled its pseudo-catalytic function.

The addition to the base of a small quantity of certain metals, ranging from 1 atom for every 100,000 atoms of the base to 1 atom for every 100 atoms, may enormously increase its luminous efficiency. This added metal is known as an activator. The best activators are metallic elements, generally with a multiplicity of valences. For example, copper, silver, bismuth, chromium, and manganese are all effective activators, although they do not, of course, behave uniformly toward all bases. The presence of more than one activator in a base usually results in a less efficient phosphor than if the best activator is used alone; however, desirable color changes may be attained at times by the addition of a second element.

It is interesting to note that certain elements, known as "killers" or "poisons," when present even in an amount small compared with that of the activator, result in a very great reduction in the luminous efficiency of the phosphor. Iron, cobalt, and nickel behave in this way. Their action is to suppress the processes of phosphorescence and thereby reduce the total emitted energy. Minute quantities of these elements are often used in the preparation of fluorescent screens where a short time lag is desired.

Although there is a basic procedure common to most phosphors, the exact method of synthesizing a phosphor depends upon the particular material.

The base material and the flux, when used, are ground together in a clean, inert crucible. To this is added the activator required to produce the desired color, generally in the form of an aqueous solution of the salt of the metal used as activator. After being dried and reground, the mass is heated and allowed to crystallize. The phosphor may then be put through a final grinding process to comminute the crystal particles to the size required for application to the screen of the viewing tube.

2.4 General Properties of Inorganic Phosphors. Nearly every step in the preparation and subsequent handling of a phosphor has a marked influence on its properties. This influence is not confined to

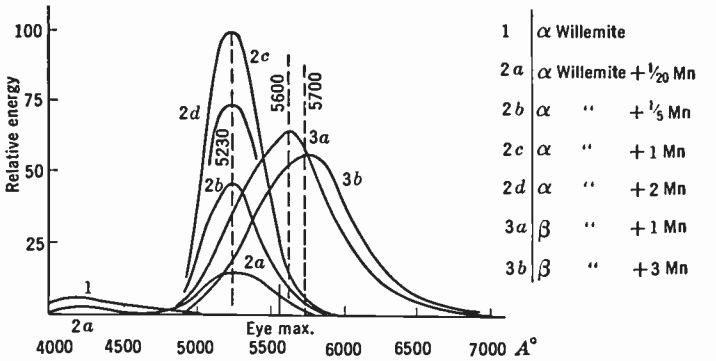


Fig. 2.2. The Effect of Preparation and Activator on the Spectral Characteristics of Zinc Orthosilicate.

the kind and quantity of activator but also the crystallization procedure, the grinding, the composition of the base, etc. Some of these effects are illustrated in the following examples.

The spectral output of zinc orthosilicate under different conditions of preparation is shown in Fig. 2.2. Curve 1 is the luminescence of the pure base. The addition of a small amount of manganese results in curve 2*a* if the material is cooled slowly. More of the manganese activator increases the luminous output without change in spectral characteristics. Eventually an optimum is reached, represented by curve 2*c*, and further addition of manganese results in a gradual decrease in efficiency. If the material, instead of being cooled slowly, is quenched rapidly from a melt, quite different results are obtained. An amount of manganese activator equal to that which previously resulted

in the material giving curve 2*c* now leads to a yellow prosphor shown in curve 3*a*. As the amount of activator is increased, the color of the luminescence shifts to longer wavelengths. Curve 3*b* is that of the material having three times the amount of manganese used to obtain curve 3*a*.

Another interesting series is that of a zinc cadmium sulphide base with a silver activator, the ratio of zinc to cadmium being varied. Figure 2.3 shows a family of curves illustrating the behavior of this

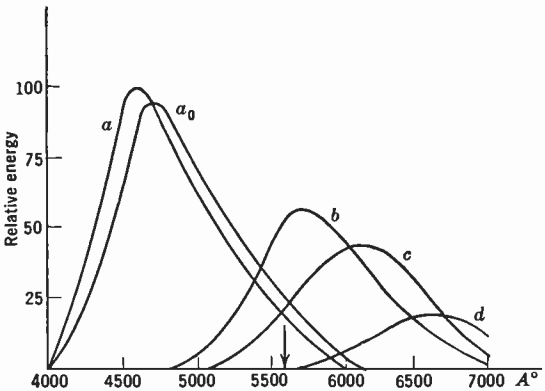


Fig. 2.3. The Variation of Spectral Characteristics of Zinc-Cadmium Sulphide with Decreasing Zinc:Cadmium Ratio.

series. Pure zinc sulphide has a blue luminescence shown by curve a_0. The addition of silver activator increases the energy efficiency and shifts the spectral output further into the blue (curve *a*). As cadmium replaces zinc in the base, the color response shifts toward the red. At the same time, there is an initial decrease in response. With further addition of cadmium, the output increases and goes through a maximum and then decreases until, for silver-activated cadmium sulphide alone, a rather weak, deep red luminescence is obtained. This is illustrated by the sequence of curves *b*, *c*, and *d*.

As in the two examples cited, a similar series of complicated phenomena is observed for nearly every phosphor group.

In addition to the crystal structure, the state of strain is an important factor in determining the luminous efficiency. For example, some freshly prepared phosphors which have high efficiency can be completely ruined if violently ground. Even a quite gentle grinding may reduce the efficiency to some extent because it will leave a strained surface layer over each grain.

The surface brightness of most phosphors is proportional to the current density of the bombarding electrons for low values of the current density. As the density is increased, the material saturates and the brightness rises less rapidly. The form of the saturation curve and the current value at which saturation begins depend upon the kind of material used and its treatment.

The relation between light output and beam voltage is rather difficult to determine for reasons that will be made clear in a later section. Measurements in the range up to 10 kilovolts indicate a light output proportional to some power of the voltage, the exponent lying between 1 and 2. At still higher operating voltages, up to 50 kilovolts, television screens of optimum thickness are found to have light outputs approximately proportional to beam voltage or, for fixed beam current, to the power input.

In a television viewing tube, the beam which serves as excitation is on a given area for a length of time equal to about 0.1 to 0.2 microsecond. The light emitted from the area continues for several hundredths of a second, as can be seen from the persistence curves shown in Figs. 2.10 and 2.14. The greater part of the light energy, therefore, is released after the beam is no longer exciting the area in question. In other words, the phosphorescence of the material provides most of the light. The total light energy released, i.e., the light intensity integrated over the time it persists, is approximately proportional to the energy supplied to the phosphor, rather than to the instantaneous power in the cathode-ray beam. Thus, if a beam of 10 microamperes and 6000 volts excites a given area for 10^{-6} second, the light averaged over $\frac{3}{100}$ second will be approximately the same as that produced by an excitation of 100 microamperes for 10^{-7} second. The accumulative effect should not be considered as a rigorous law, and it exists only over periods of time which are short compared to the decay time and for currents well below the saturation of the phosphor. As will be explained in a later chapter, the effect has been applied to the type of picture transmission involving velocity modulation.

A final consideration in this general survey is that of the effect of temperature on luminescence. The thermal state of the phosphor has a much greater effect on phosphorescence than on fluorescence. Between $-80°C$ and $300°C$, there is very little change in such phosphors as manganese-activated zinc orthosilicate and copper-activated zinc sulphide, but, if the temperature is carried much above $300°C$, the fluorescence decreases rather rapidly. The effect of a decrease in temperature on phosphorescence is to increase its duration and lower its

intensity. Calcite is an excellent illustration of this. At room temperature this substance has a brilliant orange phosphorescence which continues several minutes. If a calcite crystal is excited by cathode-ray bombardment and then cooled, the brilliancy decreases until it finally disappears. If it is again warmed up to room temperature, the phosphorescence reappears without further excitation. If it is raised to a high temperature, the brightness increases and the duration decreases. An excited calcite crystal may be kept for days at liquid air temperature, and then made to phosphoresce merely by bringing it up to room temperature. This behavior indicates that a certain quantity of energy available for phosphorescence is stored in the crystal, and that its rate of release in the form of light is a function of temperature. Thus, at high temperature, it is emitted very rapidly, giving brilliant luminescence and rapid decay, whereas at a low temperature this energy is released slowly over a long period of time.

In the next section, which discusses briefly the mechanism advanced by quantum physics to explain the phenomenon of luminescence, it will become evident why this behavior of phosphorescence is to be expected.

2.5 The Theory of Luminescence. The concept of the luminescent center, originally introduced by Lenard, plays a primary role in the interpretation of the phenomenon of luminescence. Such a center is an inhomogeneity in the crystalline base of the phosphor, normally created by the presence of an activator atom or ion either in an interstitial position within the lattice or substituting for one of the lattice ions. The center has a series of discrete electron energy levels, of which the lower ones are normally occupied. The position of these levels is a function both of the foreign atom or ion itself and of its position in the lattice. Since the base itself is a semiconductor or insulator, the phosphor has the electronic structure of an impurity semiconductor.

The occurrence of luminescence requires at least two events: the excitation of the luminescent center, either as a result of the absorption of radiation or bombardment with high-energy particles, and its return to the normal state accompanied by the emission of some fraction of the absorbed energy in the form of light. If an appreciable interval of time elapses between excitation and emission, as in phosphorescence, there must be a certain probability for the transfer of the excited system to a metastable state, from which it can return to its original condition only after an auxiliary excitation, which may be thermal in nature.

The models for the luminescence process which have been constructed to explain experimental observations differ materially for different types of phosphors; a number of typical examples are given below. Before they are considered, it may be well to note a fundamental difference between luminescence excitation by ultraviolet radiation and excitation by high-velocity electrons. It can readily be understood that, in ultraviolet excitation with a wavelength longer than the threshold of lattice absorption, conversion efficiencies approaching unity may be attained, since the centers may account for the total absorption of the material. The energy dissipation of an electron beam in the phosphor is, on the other hand, essentially independent of the concentration of centers. Conversion efficiencies of 10 percent may be reached in cathodoluminescence with an activator concentration as low as 0.01 percent. This implies that the excitation energies of lattice ions in the path of an electron must be transferred to centers at considerable distance. This transmission of energy by "excitation waves" terminating at centers * depends both on the constitution of the phosphor base and on the activator and accounts for the lack of parallelism in the properties of phosphors under ultraviolet and electron excitation. Occasionally, however, a similar mechanism is involved in ultraviolet excited fluorescence when the absorbing and emitting centers are different.

Among the various types of phosphors to be considered, the thallium-activated alkali halides represent an unusually simple example. Here, according to the model proposed by F. Seitz,† thallium ions substituted for alkali ions at the normal lattice points function as active centers. Considerably less energy is required to raise an electron of the thallium ion to an excited state than to transfer an electron from the highest filled band of the lattice to the conduction band. Such an excitation of the thallium ion is, in general, followed by a readjustment of the surrounding ions to a new position of minimum potential energy, the liberated energy being dissipated as thermal vibrations. After, or during, this readjustment the electron may return to the ground state (followed by a second ionic readjustment and thermal energy dissipation) with the emission of fluorescent radiation, the whole process consuming an interval of the order of 10^{-8} second.

In addition to this fluorescent radiation, a weak phosphorescence, proportional to the square of the concentration of the thallium ions, is

* See, e.g., Johnson, reference 6, and Fonda and Seitz, reference 3.

† See Seitz, reference 5.

also observed. This may be interpreted as resulting from the excitation of one of a pair of adjoining thallium atoms, the concentration of pairs being proportional to the square of the concentration of thallium ions. The variation of the electronic quantum levels as function of the relative position of the ion and its neighbors is shown schematically, for

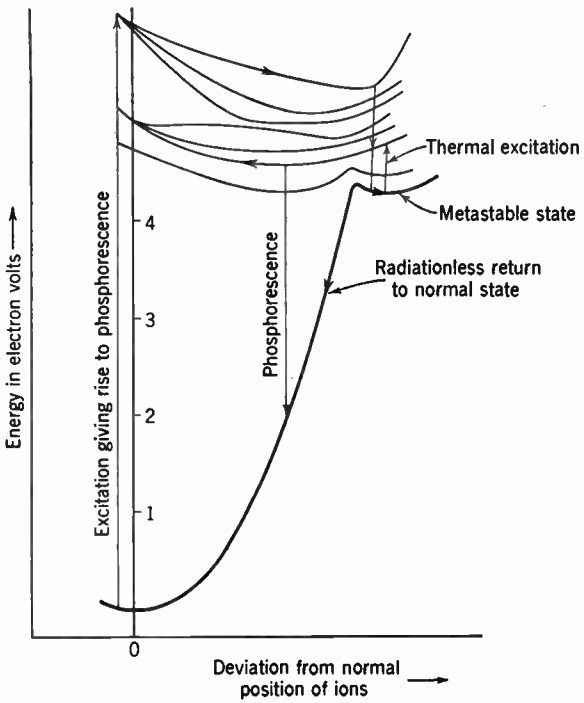


Fig. 2.4. Schematic Energy Level Diagram for Thallium Ion with Neighboring Thallium Ion in Alkali Halide Phosphor. (F. Seitz, reference 5; courtesy of *Journal of Chemical Physics.*)

a pair of adjoining thallium ions in the alkali halide lattice, in Fig. 2.4. It is seen that, in this configuration, the equilibrium position of the system in the excited state overlies a second minimum of the ground state. After the system has dropped to this metastable state, it will remain there until either the system is given enough thermal energy to raise it over the "hump" between the two minima of the ground state or an excitation to an intermediate electronic state, with an equilibrium position closer to the equilibrium position of the normal state, takes place. In the first case, the system will simply

slide down the curve for the ground state to the original condition, communicating its excess energy to the lattice in the form of thermal vibrations. In the second case, the system will return to the ground state with the emission of radiation, giving rise to phosphorescence. The average time that elapses between excitation and phosphorescent radiation is given by the mean life of the metastable state.

It will be noted that in the thallium-activated alkali halide phosphor the excited electron remains bound, throughout, to the active center. Hence, this phosphor shows no photoconductivity. The situation is quite different in the widely employed zinc and cadmium sulphide phosphors, activated with 0.01 percent of copper or silver. The active

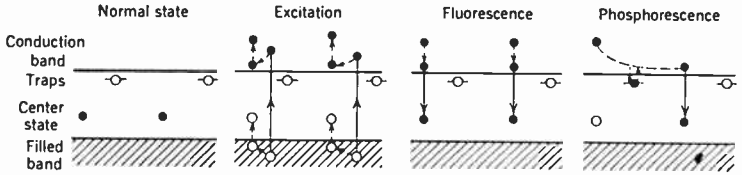


Fig. 2.5. Probable Transition Scheme Giving Rise to Luminescence in Sulphide Phosphors.

centers consist here of interstitially located atoms or, in the pure substance, interstitial zinc or cadmium. Figure 2.5 indicates the picture which has been proposed to account for the luminescence of the materials.*

As a first step, an electron may be excited either from the center or from the filled band of the lattice. If the excitation takes place at the center, the electron may return to its original state with the emission of fluorescent radiation, as in the alkali halide phosphor. On the other hand, if the lattice is excited, the electron in the conduction band and the hole in the filled band will lose their excess energy in the form of lattice vibrations and drift through the lattice as a pair, recombination being prevented by selection rules. When such a pair, or exciton, collides with a center, it transfers its energy to the center, causing ionization. Again, the possibility exists that the freed electron will drop down to the normal state of the center, with the emission of fluorescent radiation. As an alternative, it may wander off until it is caught by a trap, that is, an empty energy level lying slightly below the bottom of the conduction band. Such traps are likely to be encountered at any crystal faults.

* See, e.g., Johnson, reference 6, and Fonda and Seitz, reference 3.

Conduction electrons which fall into these traps remain there until they are raised once more into the conduction band by thermal excitation. An electron freed in this manner will drift through the lattice until it is either retrapped or drops into an empty active center. The emission of radiation—phosphorescence—will then take place during this transition.

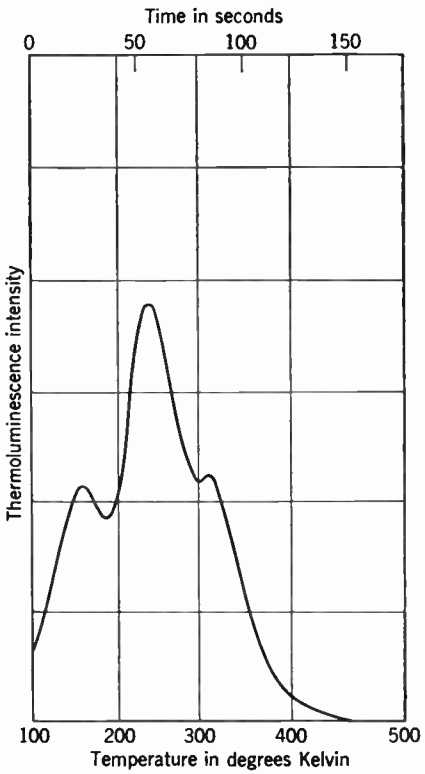


Fig. 2.6. Glow Curve for Copper-Activated Zinc Sulphide:Phosphor. (J. T. Randall and M. H. F. Wilkins, reference 8; courtesy of the *Proceedings of the Royal Society of London.*)

The predominance and duration of phosphorescence evidently depend on the concentration and the depth of the traps; the probability of thermal excitation depends exponentially on the ratio of the distance of the traps below the lower edge of the conduction band and the absolute temperature of the phosphor. As a result, the distribution of the trapping levels may be obtained by exciting the phosphor with ultraviolet light at a very low temperature and observing the variation of emission as the temperature of the phosphor is gradually increased. The "glow curve" so determined indicates the freeing of conduction electrons from progressively deeper traps in accord with the exponential dependence of the probability of thermal excitation noted above (Fig. 2.6).*

The same effect for calcite was described in the preceding paragraph. Alternately, the traps may be emptied by irradiation with infrared radiation; the so-called infrared phosphors are essentially phosphors with a large concentration of traps of such depth that they may be emptied by infrared radiation of a prescribed range. Such phosphors,

* See Garlick in Fonda and Seitz, reference 3, p. 87, and Randall and Wilkins, reference 8.

after excitation with ultraviolet light, will glow brightly under infrared stimulation even after their short-term luminescence has died down.

It is clear that, in the zinc sulphide phosphors, the electrons in the conduction band may contribute to photoconduction, drifting in the direction determined by the applied field. The second most important class of phosphors, the silicate phosphors of which zinc orthosilicate and zinc-beryllium orthosilicate activated with manganese are well-known examples, assumes an intermediate position between the alkali halide phosphors and the zinc sulphide phosphors. According to the view proposed by R. P. Johnson,* the centers are here interstitial manganese atoms; manganese ions substituted for regular lattice ions are assumed to play no role in the luminescence process. Irrespective of

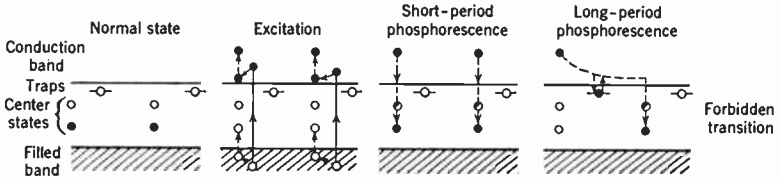


Fig. 2.7. Probable Transition Scheme Giving Rise to Luminescence in Orthosilicate Phosphors.

their origin, the centers are initially ionized. The majority of the resulting free electrons are trapped in a metastable excited state of the centers, from which they eventually drop to the normal state with the emission of radiation. The lifetime of this excited state is relatively long since the transition to the normal state is "forbidden" by selection rules.

Some of the free electrons are trapped, instead, in shallow traps from which they may be liberated by thermal excitation. These electrons, which eventually reach the metastable excited state of a center and, hence, the ground state with the emission of radiation, account for a more gradual decay of phosphorescence as well as the observed photoconductivity (Fig. 2.7).

2.6 The Theory of Luminescence (*continued*). Leaving the microphysical aspect of the problem, there are certain quantitative relations that can be deduced on a classical, statistical basis, which are helpful in a general survey of the subject.

Essentially two different types of emission processes have been described above. In one of them the excited electron remains through-

* See Johnson, reference 6.

out in the domain of the center from which it originated and eventually returns to its normal state with the emission of a photon. In the other, any one of the excited electrons may recombine with any other empty center, resulting in phosphorescence. The first process may be characterized as a monomolecular process, the second as a bimolecular process.

In the monomolecular process, the number of emission processes per second, or the decrease in the number of excited centers taking place per second, is, in the absence of excitation, simply proportional to number of excited states N:

$$-\frac{dN}{dt} = \alpha N \tag{2.1}$$

Here α is a constant, the reciprocal of the mean life of the excited state. If this expression is integrated, the number of excited centers at any time t is found to be

$$N = N_0 e^{-\alpha t} \tag{2.2}$$

Since the intensity of the emitted light is proportional to the rate at which centers recombine, the phosphorescence, as a function of time, is given by

$$L = Ce^{-\alpha t}, \qquad C = kN_0\alpha \tag{2.3}$$

k being the light content of a photon.

Under bombardment by an electron beam of current density ρ, an equilibrium is established between the rate of formation of excited centers and the rate of recombination. The number of excitations during the interval of time dt is proportional to the number of unexcited centers available and the current density of the incident beam. If, therefore, the total number of available centers is M and the number of excited centers N, the condition of equilibrium can be expressed as

$$\eta\rho(M - N)\,dt = \alpha N\,dt \tag{2.4}$$

Hence

$$N = \frac{\eta\rho M}{\eta\rho + \alpha} \tag{2.5}$$

From this, the light output as function of current density is

$$L = D\frac{\rho}{1 + \beta\rho} \qquad D = kM\eta, \quad \beta = \eta/\alpha \tag{2.6}$$

where k is the light content of a photon.

When the mechanism of luminescence involves independent excited electrons and vacant energy states, the rate at which emission can occur is proportional to the product of the number of excited electrons and the number of vacancies. If these are taken to be equal, the rate of recombination is given by

$$-\frac{dN}{dt} = \zeta N^2 \tag{2.7}$$

And the number of excited centers at time t, which is obtained by integrating Eq. 2.7, is

$$N = \frac{N_0}{1 + N_0 \zeta t} \tag{2.8}$$

From Eqs. 2.7 and 2.8, the phosphorescent decay characteristic is found to be

$$L = E\frac{1}{(1 + \gamma t)^2} \qquad E = kN_0{}^2\zeta, \quad \gamma = N_0\zeta \tag{2.9}$$

Under equilibrium conditions between excitation and emission, the following relation between L and ρ is obtained:

$$L = D\rho(1 + \rho\delta - \sqrt{\rho^2\delta^2 + 2\rho\delta}) \qquad D = kM\eta, \quad \delta = \eta/(2\zeta M) \tag{2.10}$$

For large values of the beam density ρ, this expression takes on the same form as Eq. 2.6, with 2δ replacing β.

Both of these types of decay and current characteristics are found among the phosphors. The zinc and cadmium sulphide phosphors belong distinctly to the second type, in harmony with the mechanism of phosphorescence described above; the alkali halides activated with thallium, to the first. The zinc orthosilicate phosphors exhibit an initial exponential decay, followed by a more gradual decline in the phosphorescence of the character described by Eq. 2.9. It should perhaps be stressed that exact conformity with the decay laws in Eqs. 2.3 and 2.9 and the corresponding relations between light output and current density should be expected only in exceptional cases. The equations were deduced under the assumption that the phosphorescence process is unique or, in other words, that either a single metastable state or traps of a single depth play a role in it. If, as suggested by the work of Garlick and his co-workers on the glow curves of many phosphors,*

*See Randall and Wilkins, reference 8, and Garlick in Fonda and Seitz, reference 3, p. 87.

this condition is not satisfied, the decay characteristics and the light output-current variation will necessarily be more complex.

Experimentally, the voltage variation of the light output of phosphors and the current variation are found to be practically independent of each other. It is hence permissible—at least within limited ranges of the voltage V—to express the light output L in the following form:

$$L = A \cdot f(\rho) \cdot V^n \tag{2.11}$$

In the absence of detailed knowledge regarding the process of energy transfer from the point of excitation to a near-by luminescent center, the exponent n might reasonably be expected to equal unity. This would mean that the light output, for low current densities, would be simply proportional to the energy dissipated by the beam in the phosphor, this energy being transferred to the lattice by electronic excitation and ionization. Actual measurements of the light output of zinc sulphide and zinc orthosilicate screens in the range from 500 to 10,000 volts yield values of n which range from 1.5 to more than 2.* For higher voltages, up to about 50 kilovolts, screens of optimum thickness have been found to have more nearly constant efficiency, or a light output proportional to the operating voltage.

One factor which will tend to lead to a more-than-linear increase of light output with voltage at low voltages is the presence of a "dead layer" enveloping every phosphor grain. Energy dissipated by electrons in this inactive surface layer cannot contribute to the luminescence. Since the penetration of the electrons increases rapidly with voltage, the energy lost in this manner decreases and the efficiency of conversion of electron energy into light increases. It should also be pointed out that the granular structure of fluorescent screens renders measurements of the efficiency of the basic luminescence process very difficult. The light output is materially influenced by scattering and absorption in the phosphor layer. This factor will be discussed in more detail in Chapter 11.

2.7 Phosphors for Television. Some of the principal cathodoluminescent phosphors, together with their properties, are listed in Table 2.2. The code numbers P1 to P16 have been established by the Radio and Television Manufacturers Association. The screen type P10 does not represent a luminescent screen, but a dark-trace screen; the scanning beam leaves a dark trace, whose maximum absorption is at

* See Martin and Headrick, reference 13.

TABLE 2.2. PRINCIPAL SCREEN MATERIALS USED IN CATHODE-RAY TUBES *

Code Number	Type	Chemical Composition	Color	Spectral Maximum (Å)	Persistence (seconds)	Relative Luminescent Efficiency (%)
P1	Zinc silicate	rbhdl-Zn_2SiO_4:Mn(0.25)	Green	5250	0.05	100
P2	Zinc-cadmium sulphide	ZnS(96)CdS(4):Cu(0.01)	Blue-green	5360	0.001	140
P3	Zinc-beryllium silicate	$8ZnO \cdot BeO \cdot 5SiO_2$:Mn(1.0)	Light green-yellow	5500	0.06	84
P4	Zinc sulphide and zinc-beryllium silicate	hex-ZnS:Ag(0.01) + $8ZnO \cdot BeO \cdot 5SiO_2$:Mn(1.0)	White	4500	0.06	80
P4	Zinc sulphide and zinc-cadmium sulphide	hex-ZnS:Ag(0.01) + ZnS(48)CdS(52):Ag(0.01)	White	4440	0.005	150
P5	Calcium tungstate	$CaWO_4$:W	Light violet	4300	10^{-5}	32
P7	Zinc sulphide on zinc-cadmium sulphide	cub-ZnS:Ag(0.015) on ZnS(86)CdS(14):Cu(0.0073)	Blue white	4400	3	80
P10	Potassium chloride	KCl	Dark trace	5570	5	...
P11	Zinc sulphide	hex-ZnS:Ag(0.01)	Light blue	4580	0.05	72
P12	Zinc-magnesium fluoride	$(ZnMg)F_2$:Mn(1.0)	Orange	5900	0.5	80
P13	Magnesium silicate	$MgSiO_3$:Mn(0.15)	Light red	6740	0.1	5
P14	Zinc sulphide on zinc-cadmium sulphide	cub-ZnS:Ag(0.015) on ZnS(73)CdS(27):Cu(0.004)	Pure white	4400	1	70
P15	Zinc oxide	ZnO:Zn	Light blue-white	5050	10^{-6}	60
P16	Calcium-magnesium silicate	$Ca_2MgSi_2O_7$:Ce(0.05)	Ultraviolet	3700	10^{-6}	...

* From H. W. Leverenz, "Final Report on Research and Development Leading to New and Improved Radar Indicators," Report 25481, Office of the Publication Board, Department of Commerce, Washington, D. C., 1945, and JETEC Cathode Ray Tube Committee, J6-13-1, *Description of Phosphors by Color and Persistence.*

5570 Å. In the column marked "chemical composition" the figures in parenthesis indicate percentages by weight. The measurements of the spectral maximum of the emission, the relative luminescent efficiency, and the persistence of the material were made at 6 to 10 kilovolts and the very low current density of 1 microampere per square centimeter.

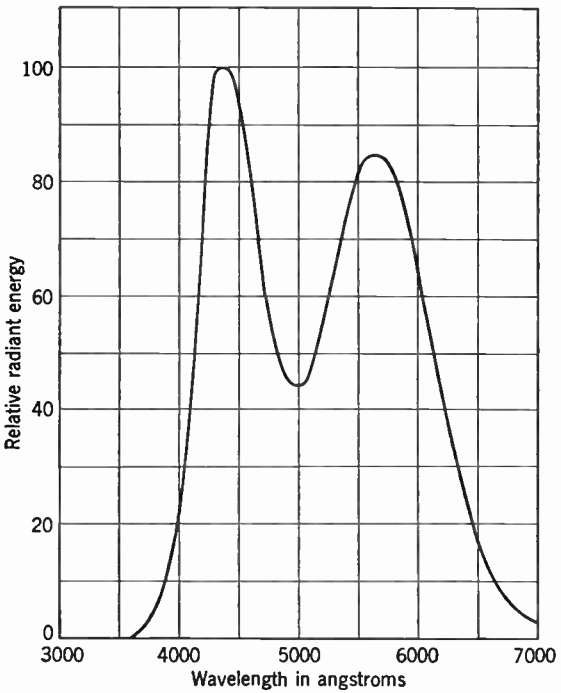


Fig. 2.8. The Spectral Characteristics of a White Fluorescent Screen Consisting of Silver-Activated Zinc Sulphide and Zinc Cadmium Sulphide.

It is found that under these conditions the absolute luminescent efficiency of the P1 phosphor screen, taken as standard, is about 6 candlepower per watt; the figures in the last column of the table make it possible to derive herefrom the absolute efficiencies of the remaining phosphors. Operated at low currents and high voltages (of the order of 15 kilovolts) and with a reflecting backing, P1 screens have yielded as much as 15 candlepower per watt. The "persistence" of the phosphor indicates the time required for the emission to drop to 1 percent of its initial value.

The two phosphors denoted by P4 in the table, formerly listed as P4 and P6,* are "white" phosphors of the type generally employed for television viewing tubes. These phosphors are obtained by mixing a light blue silver-activated zinc-sulphide phosphor with a greenish-yellow zinc-beryllium silicate or zinc-cadmium sulphide phosphor. As indicated in Fig. 2.8, the spectral distribution of the emission of this compound phosphor is very different from that of daylight, which exhibits at most a single flat maximum. The fact that both are perceived as white is a consequence of the tricolor nature of human vision. In the cascade phosphors, P7 and P14, an increased persistence is attained by depositing a blue phosphor on a yellow or orange phosphor, so that the emission of the first excites that of the second.†

In general, the practical phosphors can be divided into two classes—the sulphide (and selenide) phosphors, of which zinc sulphide and zinc-cadmium sulphide are the principal exponents, and the oxide (and fluoride) phosphors. Among the latter zinc orthosilicate, calcium-magnesium silicate, and zinc-beryllium silicate are important examples.

2.8 Sulphide Phosphors. The zinc sulphide and zinc-cadmium sulphide phosphors constitute extraordinarily efficient cathodoluminescent materials. Under normal operating conditions zinc sulphide screens activated with silver yield, at bombarding voltages of 6 to 10 kilovolts, luminous outputs of 3 to 5 candlepower per watt. A drawback of zinc sulphide screens has been, in the past, their relative instability. However, methods now employed in the technique of applying the material to television tubes and in tube exhaust have resulted in screens which have a life nearly equal to that of the most durable material and long enough so that the life of the tube is not determined by the life of the screen.

There would be no advantage in taking up the details of a factory process of preparing luminescent zinc sulphide; however, it is not out of place to describe a laboratory method of synthesizing the phosphor.

Zinc sulphate is first purified by ordinary methods until the spectroscope fails to reveal any impurities. Then, in order to make certain that no copper remains, an aqueous solution of the salt is electrolyzed, using platinum electrodes. Pure white zinc sulphide is precipitated

* According to a more recent convention, P4 denotes any phosphor for the screen of a viewing tube employed for monochrome television; P6, any phosphor employed in color television, irrespective of spectral characteristics and composition.

† See Leverenz, reference 11.

from the purified solution by bubbling clean hydrogen sulphide through it. The precipitate is washed with the purest distilled water and dried. Ten grams of this zinc sulphide are placed in an acid-cleaned quartz crucible. To a water slurry of the zinc sulphide are added 0.2 gram of pure sodium chloride as well as a silver nitrate solution which contains 0.001 gram of silver. The material is then evaporated to dryness and thoroughly ground. The phosphor is crystallized by heating to 900°C, the temperature not being at all critical. The result will be cubic ZnS:Ag, one of the best phosphors known to date.

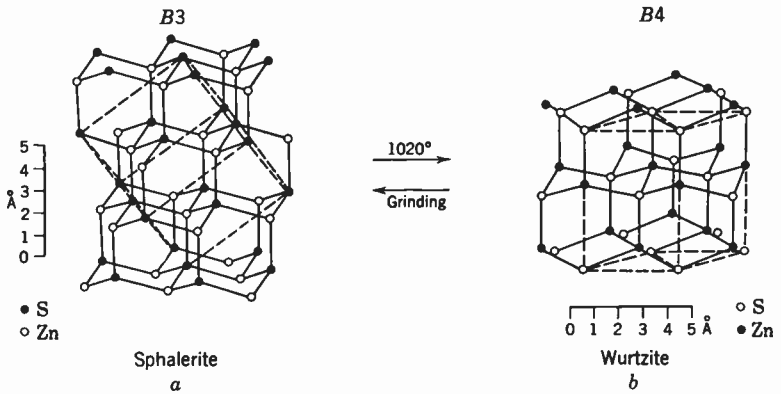


Fig. 2.9. The Crystal Structure of Sphalerite and Wurtzite.

The crystal structure of cubic zinc sulphide or sphalerite is shown in Fig. 2.9*a*. There is also an enantiotropic form of this material, wurtzite or hexagonal zinc sulphide, which crystallizes at temperatures above 1020°C, and which can be converted to sphalerite by grinding. The structure of this form is shown in Fig. 2.9*b*.

In spite of its luminous efficiency, the phosphorescent decay period of ZnS:Ag is quite short, amply so for television purposes. Its persistence characteristics are given in Fig. 2.10. From this curve, it will be seen that the light output has dropped to less than 0.1 percent in one picture period.

The light output is not strictly proportional to the bombarding current and tends toward a gradual saturation, as illustrated in Fig. 2.11. The variation with voltage is less readily obtained, as will be explained later, because of the difficulty of determining the true bombarding voltage. In general, the light output increases with nearly the second power of the bombarding voltage.

The color of the luminescence is blue, corresponding to a band extending from 4000 to 5700 Å. A spectral curve (1) of this material is

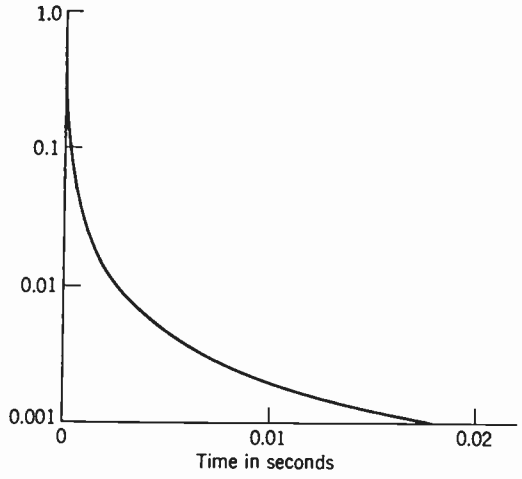


Fig. 2.10. Persistence Characteristics of a Zinc Sulphide Phosphor. (Leverenz, reference 2.)

shown in Fig. 2.12. However, this curve is merely representative, as the exact shape of the curve is very sensitive to composition, preparation, and after-treatment.

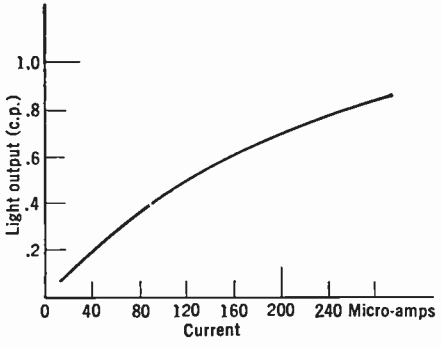


Fig. 2.11. Luminescence as a Function of Electron Current for a Zinc Sulphide Phosphor.

Zinc-cadmium sulphide is another useful sulphide phosphor. The preparation is quite similar, except that, for every gram of purified zinc sulphide, a definite amount, such as 1 gram, of cadmium sulphide

is added. Sodium chloride is used as a flux; 0.01 percent silver in the form of the nitrate may serve as activator.

From the spectral output curve (2) of Fig. 2.12, the color of this material can be seen to be yellow. Its luminous efficiency is higher

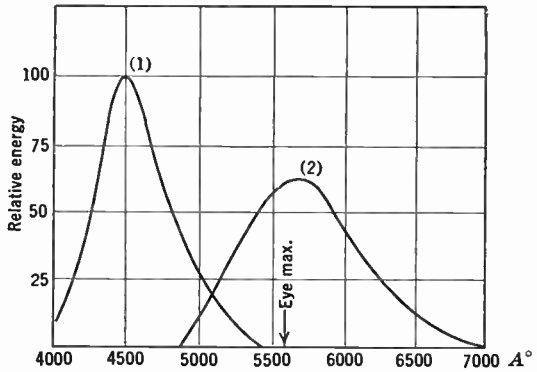


Fig. 2.12. The Spectral Characteristics of a Silver-Activated Zinc Sulphide and Zinc Cadmium Sulphide Phosphor.

than that of the zinc sulphide. The persistence characteristics of the two sulphide phosphors are very similar.

2.9 The Oxide Phosphors. Well-known representatives of this class are manganese-activated zinc orthosilicate and zinc-beryllium silicate. At moderate current densities, both materials are somewhat less ef-

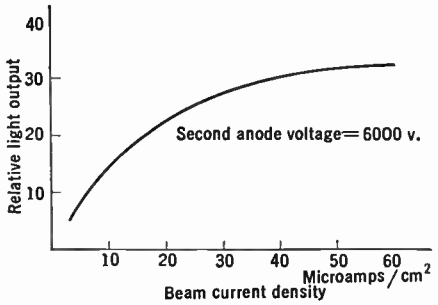


Fig. 2.13. Luminescence as Function of Electron Current for a Zinc Orthosilicate Sample.

ficient than the two sulphides just described. Their color can be made to range from green to red with only small changes in efficiency, the substitution of zinc by beryllium causing a shift of the spectral maximum toward the red end of the spectrum.

The main advantage of the silicate phosphors over the sulphide phosphors is their extreme ruggedness, which is of considerable value in television work, particularly in its experimental stages.

Small-scale laboratory preparation of a zinc orthosilicate phosphor may be carried out as follows: Twenty grams of highly purified and finely divided zinc oxide and 7.4 grams of pure silicon dioxide are ground together thoroughly. To this is added about one mole percent of manganese in solution, or 0.07 gram of the metal. The mixture is stirred, dried, and then finely ground. Lastly, it is heated in a covered

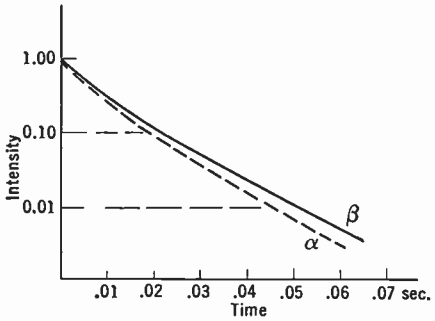


Fig. 2.14. Persistence Characteristics of Samples of α and β Zinc Orthosilicate.

platinum crucible to about 1300°C for 2 hours. If the heated material is cooled slowly, the resultant phosphor has a bright green luminescence; if it is quenched rapidly from a melt (above 1512°C), the luminescence will be yellow.

The crystal structure of this material is much more complicated than that of the sulphide. The rhombohedral form, that is, the material which is formed by a gradual cooling, has a structure which is isomorphous with phenacite, with 6 molecules in the unit cell. A second form exists but has not been completely worked out. It has been reported by G. R. Fonda to resemble cristobalite. The material may take on a third form when prepared with an excess of silica, but this is of little interest for our present purpose since it exhibits only a weak red luminescence.

The persistence characteristics for the two forms of Zn_2SiO_4:Mn are shown in Fig. 2.14. In spite of the differences in the spectral output, the decay curves are very similar. The luminous output of both these materials has dropped to about 4 percent of its initial value in one picture period.

The spectral characteristics of the two materials can be seen from Fig. 2.15. These phosphors give a characteristic broad band of radiation.

Like the sulphides, these materials exhibit a gradual saturation with current (Fig. 2.13), and a light output which rises somewhat faster than with the first power of the voltage.

Two other oxide phosphors of considerable practical value are zinc oxide and calcium tungstate. Both belong to the rather restricted class of phosphors which emit without the aid of a foreign activator. They

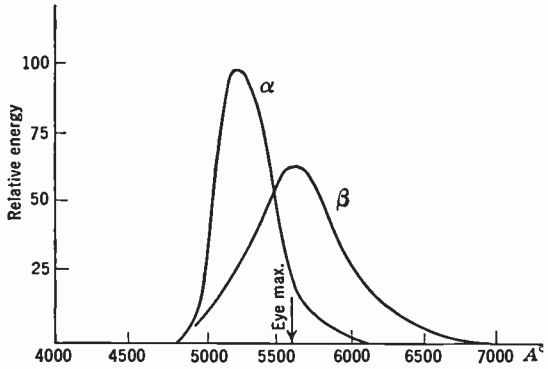


Fig. 2.15. Typical Luminescence Spectra of α and β Zinc Orthosilicate.

are somewhat less efficient than the sulphide and silicate phosphors which have been described, and luminesce with a pale bluish color. On the other hand, the extraordinarily rapid decay of their luminescence makes them suitable for oscillograph tubes designed for use where long decay cannot be tolerated. The highly actinic quality of the emission is useful for photographic purposes. Zinc sulphide activated with a minute quantity of nickel similarly has a phosphorescence of extremely short duration.

2.10 Electrical Properties of Phosphors. The specific resistance of all materials used as phosphors is extremely high. Therefore, when a phosphor is bombarded by an electron beam of the current density used in television practice, the principal way the electrons escape from the surface is by secondary emission. If the secondary emission ratio of the surface is less than unity for a given bombarding voltage, the area will become more negative until it either reaches cathode potential, thus preventing electrons from reaching the screen, or else reaches a potential such that the secondary emission ratio becomes unity.

Because of its practical importance, it is desirable to discuss the discharge process in some detail. As has been shown in the preceding chapter, the secondary emission ratio of most substances, conductors and insulators alike, rises rapidly to a maximum in the neighborhood of 300 to 800 volts, and then falls slowly as the voltage increases. This is true of all the phosphors that have been studied, and also of the glasses which are used as a backing to support the screen material. The maxima and shape of the secondary emission curves are characteristic of each substance. However, the forms of the curves are sufficiently similar to permit useful generalizations. Figure 2.16 shows a typical secondary emission curve for an insulator.

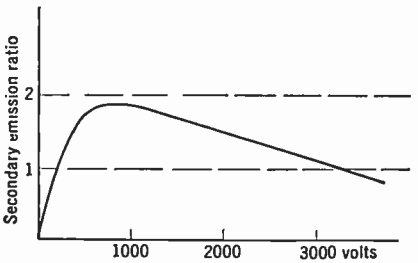


Fig. 2.16. Typical Secondary Emission Curve for an Insulator.

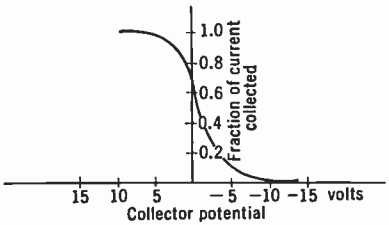


Fig. 2.17. Secondary Emission Current as Function of Collector Voltage.

Another factor determining the rate at which the electrons leave the screen is the potential difference between the bombarded surface and the electrode which collects the current from it. In order to obtain a maximum current, it is necessary that the collector be at a positive potential with respect to the screen; however, current will continue to flow to the electrode even when it is slightly negative, owing to the initial velocities of the leaving electrons. The form of the curve representing the relationship between collector potential and current ratio is illustrated in Fig. 2.17. The shape of this curve is determined by such factors as tube configuration, bombarding current, and the nature of the emitter. However, the range of potential represented by the rising portion of the saturation curve is so small compared with the bombarding voltage that the use of a generalized form does not appreciably affect the validity of the conclusions drawn.

The type of tube in which the screen material is used is shown in Fig. 2.18. The figure illustrates schematically a typical kinescope which consists of an electron gun acting as the cathode-ray source, a second anode which collects the electrons after they leave the screen,

and a screen of luminescent material. The voltage V_a is applied between the gun and second anode. However, since the additional condition that as much current must leave the screen as arrives in the beam must be fulfilled, the velocity with which the electrons strike the screen will differ, in general, from that corresponding to an accelerating potential of V_a volts. On the diagram, the true bombarding voltage is represented by V_s, which normally is less than V_a, but may rise slightly above this value.

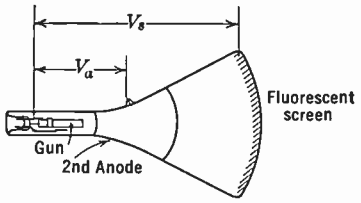


Fig. 2.18. Schematic Diagram of a Typical Kinescope.

The surface of the screen is thus bombarded by electrons with a velocity corresponding to V_s, and is subject to a collector potential $V_a - V_s$, the only parameter being V_a. Let it be assumed that the function $S(V_s)$, giving the secondary-emission ratio in terms of the true bombarding voltage, and $C(V_a - V_s)$, describing the variation of the collector current with the collector voltage, are known. The latter function, $C(V_a - V_s)$, is the ratio of the current actually collected to the saturated secondary emission at the voltage V_s in question. The current collected will be

$$i = i_b S(V_s) C(V_a - V_s) \tag{2.12}$$

Since, for equilibrium, the beam current i_b must equal the collector current, the relation

$$\frac{1}{S(V_s)} = C(V_a - V_s) \tag{2.13}$$

is necessarily fulfilled. If a plot is made of the inverse secondary emission function and of the saturation curve with its origin at the second-anode voltage V_a, the intersection of the two curves will be at the true bombarding voltage V_s. Figure 2.19 is an example of such a plot showing the determination of V_s for several values of the second-anode potential V_a.

It is instructive to consider the behavior of the screen as shown by this diagram. For second-anode voltages from 0 to A, the bombarding voltage is zero; in other words, the beam is turned back and does not strike the screen. Between A and B, where the secondary-emission ratio is greater than unity, the screen voltage is approximately that of the second anode and may actually be more positive by one or two volts. As the second-anode voltage approaches B, the screen voltage

drops below the second-anode voltage. Beyond B, the screen voltage stays constant at the value B regardless of the second-anode voltage. Thus, a curve of the true bombarding voltage as a function of second-anode voltage has the form shown in Fig. 2.20. The break-point occurs

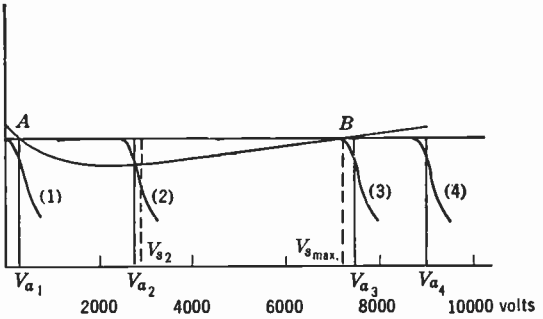


Fig. 2.19. Equilibrium Conditions of a Bombarded Insulator.

at a different second-anode voltage, depending upon the material used as screen and the treatment to which it has been subjected.

Measurements have been made of the break-point or sticking potential for glass, willemite, and zinc sulphide.* For Pyrex, such as is used in the manufacture of kinescopes, the break-point is found to lie between 2000 and 3000 volts. For willemite, it varies between 5000 and 8000 volts, depending upon the exact preparation. Measurements on zinc sulphide indicate the same range of values as for the orthosilicate.

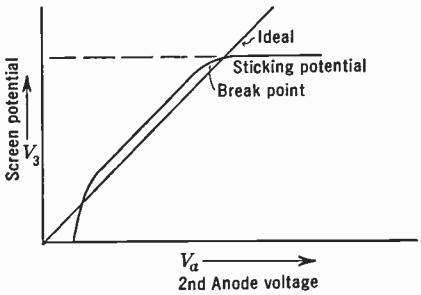


Fig. 2.20. Potential of a Fluorescent Screen Relative to the Electron Source as Function of Second-Anode Voltage.

When these materials are applied in an actual tube, conditions are much more complicated and a much wider range of values for the break-point is found. In this case, such factors as screen thickness, getter, residual gas, and age are important.

The variation of the bombarding voltage with the age of the tube is indicated in Fig. 2.21. This change may be due to the getter material,

* Nottingham, reference 12; Nelson, reference 14; and Martin and Headrick, reference 13.

which greatly increases the secondary emission ratio of the phosphor, being driven from the screen, or to small amounts of gas, etc., being freed from the walls or the gun.

Even more interesting is the apparent variation of break-point with screen thickness. Figure 2.22 shows that for very thin screens the

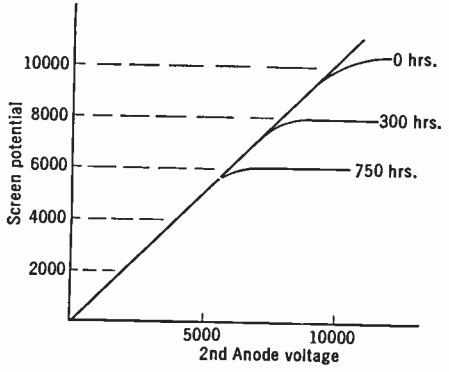


Fig. 2.21. Variation of Screen Potential Characteristics with Tube Age.

break-point is low, corresponding to that of the glass, and that for very thick screens it is similar to that of the pure phosphor. For intermediate screen thickness the break-point is no longer sharp, and the screen and second-anode potentials do not begin to differ materially

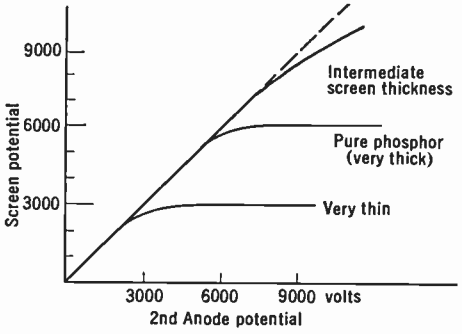


Fig. 2.22. Variation of Screen Potential Characteristics with Screen Thickness.

until very high voltages are reached. In the very thin screen, every particle of the phosphor is surrounded by a region of exposed glass which prevents its potential from rising much above the break-point of glass by its grid action on the low-velocity secondary electrons from the phosphor. The break-point of the thick screen is just that of the

phosphor itself. The conditions for the intermediate screen are more complicated, and only a conjectural explanation of its behavior can be given. There is not enough exposed glass to produce the grid action as in the thin screen, yet bombarding electrons strike both glass and phosphor. If these substances were to reach their characteristic break-points, it would mean that a potential difference of several thousand volts would exist in distances comparable to the screen thickness, which is only a few microns. Such a gradient may be sufficient to produce a breakdown which serves to discharge both phosphor and glass, and prevents establishing a true break-point.

The scanning operation adds a further complication to the picture of the electrical behavior of the screen in a tube used for television purposes. Measurements under these conditions show that, above the break-point, the screen potential tends to be closer to second-anode potential at points not under the beam than at the spot directly under the beam.* Where the second anode is 3000 volts above the break-point, a potential variation of as much as 1000 volts may exist over the screen.

If the television development had stopped with low-voltage direct-vision kinescopes, which operate below the break-point, the foregoing discussion would be of secondary importance. The fact that orthosilicate and sulphide screens may be prepared so that the screen potential continues to rise far above the break-point, even at anode voltages of 50 kilovolts, becomes significant in projection tubes. Here, the employment of very high voltages, to achieve maximum brilliance, becomes essential. An even more satisfactory solution of the problem is obtained through the deposition, on the fluorescent screen, of a very thin metallic reflecting film connected to the second anode.† This technique and its results will be described in greater detail in Chapter 11.

2.11 Conclusion. No mention has so far been made of methods of applying a phosphor to screens used in television viewing tubes. These, together with the related problems of optimum particle size, screen thickness, factors influencing contrast, etc., are reserved for a later chapter dealing specifically with the kinescope, since they pertain to a particular use of the material.

As to the future development of better luminescent materials, there is every reason to be optimistic in view of the progress that has been

* See Piore and Morton, reference 15.

† See section 11.8.

made in the past. That considerable improvement in efficiency may yet be possible is evident from a consideration of present efficiencies and the efficiency that can be obtained through ultraviolet excitation. The conversion of 1 watt of energy into radiation producing the maximum visual response will give rise to about 680 lumens of light and, if converted into radiation with the spectral response of willemite, approximately 500 lumens. The cathodoluminescence of willemite yields only 35 to 40 lumens per watt, or less than 10 percent efficiency. On the other hand, it has been reported that ultraviolet excitation of this same material in a mercury-vapor fluorescent lamp will give as much as 245 lumens per watt of ultraviolet energy or, in other words, nearly 50 percent efficiency. Although the mechanism of excitation by electrons and by light has been seen to differ, there is no basic reason why comparable efficiencies could not be attained with electron excitation. With modern, high-efficiency white phosphors, the margin for improvement is considerably less. Here, too, however, material gains in performance at high voltages and high currents, as well as in durability, may be expected.

REFERENCES

1. P. Lenard, F. Schmidt, and R. Tomaschek, "Phosphorescence and Fluorescence," *Handbuch der Experimentalphysik*, Vol. 23, parts 1 and 2, Leipzig, 1928.
2. H. W. Leverenz, *An Introduction to Luminescence of Solids*, John Wiley, New York, 1950.
3. G. R. Fonda and F. Seitz (Editors), *Preparation and Characteristics of Solid Luminescent Materials*, John Wiley, New York, 1948.
4. F. A. Kroger, *Some Aspects of the Luminescence of Solids*, Elsevier, New York, 1948.
5. F. Seitz, "Alkali Halide-Thallium Phosphors," *J. Chem. Phys.*, Vol. 6, pp. 150–162, 1938.
6. R. P. Johnson, "Luminescence of Sulphide and Selenide Phosphors," *J. Optical Soc. Am.*, Vol. 29, pp. 387–391, September, 1939.
7. N. Riehl, "Build-up and Mechanisms of Luminescent Zinc Sulphide and Other Phosphors," *Ann. Phys.*, Vol. 29, pp. 636–664, July, 1937.
8. J. T. Randall and M. H. F. Wilkins, "Phosphorescence and Electron Traps—I. The Study of Trap Distributions," *Proc. Roy. Soc.* (*London*) A, Vol. 184, pp. 365–389, November, 1945.
9. H. W. Leverenz and F. Seitz, "Luminescent Materials," *J. Appl. Phys.*, Vol. 10, pp. 479–493, July, 1939.
10. H. W. Leverenz, "Final Report on Research and Development Leading to New and Improved Radar Indicators," Report PB 25481, Office of the Publication Board, Department of Commerce, Washington, D. C., 1945.
11. H. W. Leverenz, "Luminescence and Tenebrescence as Applied in Radar," *RCA Rev.*, Vol. 7, pp. 199–239, June, 1946.

12. W. B. Nottingham, "Electrical and Luminescent Properties of Willemite under Electron Bombardment," *J. Appl. Phys.*, Vol. 8, pp. 762–778, November, 1937.
13. S. T. Martin and L. B. Headrick, "Light Output and Secondary Emission Characteristics of Luminescent Materials," *J. Appl. Phys.*, Vol. 10, pp. 116–127, February, 1939.
14. H. Nelson, "Method of Measuring Luminescent Screen Potentials," *J. Appl. Phys.*, Vol. 9, pp. 592–599, September, 1938.
15. E. R. Piore and G. A. Morton, "The Behavior of Willemite under Electron Bombardment," *J. Appl. Phys.*, Vol. 11, pp. 153–157, 1940.

3 Electron Optics

The term "electron optics" will be used in this chapter to describe that class of problems which deals with the determination of electron trajectories. The expression originated as a consequence of the close analogy between optical arrangements and the corresponding electronic systems. It was found that this analogy not only had fundamental mathematical significance, but, in many cases, could be extended to practical devices. For example, it is possible to construct electron lenses which are capable of imaging an electron source. In many instances not only is the behavior of the two types of systems the same, but also many of the mathematical methods of optics can be applied to the corresponding electron problem. There are, it is true, many systems which in no way resemble those of conventional optics. However, there is a continuous transition between these and such as have a close optical analogue. Therefore, any attempt to subdivide the field on this basis results only in confusion.

The importance of electron optics is becoming increasingly apparent with the advance of electronics. For example, in the early vacuum tubes used in radio work little attention was paid to the exact paths of the electrons between the cathode and the plate. Since then, very real improvement in efficiency and performance has been achieved by the application of electron optics to tube design. In the design of the newer devices, such as the secondary-emission multiplier, the electron gun used in television tubes, and the image tube, electron optics is essential.

The design problem usually encountered is one in which the two termini of the electron paths are specified and it is required to determine an electrode and magnetic coil configuration that will satisfy this demand. Unfortunately, a direct solution is still a good deal beyond our present mathematical means. It is not possible, except in very special cases, to determine from a given electron path the shape and potentials of the electrodes required to produce this path. In order to solve the foregoing problem it is necessary to assume an electrode configuration and then determine the resulting electron path. If this is not the re-

quired path the electrodes are changed and the trajectories recalculated. Usually this process does not have to be repeated very many times before the correct solution is reached, as the previously determined paths indicate the nature of the changes that must be made.

Restating the problem in the only form in which a solution is possible, it becomes: *Given an electrode configuration and the potentials applied, determine the electron paths in the resulting potential field.*

Even this problem has no general solution, and often can be solved only by resorting to elaborate mathematical approximations, or to the use of mechanical and graphical methods. The solution can be divided into two distinct parts, namely, the determination of the potential field produced by the electrodes, and the calculation of the electron trajectories in this field. Essentially the same procedure is used when the electrons are guided by magnetic fields.

3.1 The Laplace Equation. To determine the potential field produced by a given set of electrodes, it is necessary to solve the Laplace differential equation:

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + \frac{\partial^2 \phi}{\partial z^2} = 0 \tag{3.1}$$

with boundary conditions corresponding to the shapes and potentials of the electrodes. The solution of this equation gives the potential as a function of the coordinates, that is,

$$\phi(x, y, z)$$

The electrostatic field can be found from this potential by differentiation with respect to the coordinates. Thus,

$$E_x = -\frac{\partial}{\partial x}\phi(x, y, z)$$

$$E_y = -\frac{\partial}{\partial y}\phi(x, y, z)$$

$$E_z = -\frac{\partial}{\partial z}\phi(x, y, z)$$

From the original Laplace equation, which is satisfied by the potential function, it will be seen that the field must satisfy the differential equation:

$$\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = 0$$

It should be noticed that this equation is like an equation of continuity and may be interpreted to mean that in any volume of free space within an electrode system as many electrostatic lines of force must leave as enter. Similar equations express corresponding laws obeyed by the flow of an incompressible fluid and by electric current in a conducting medium.

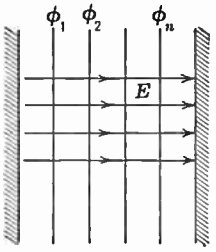


Fig. 3.1. Potential and Field between Plates.

There is no general solution for the Laplace equation, nor can any general method of attack be given. In certain special cases only can an analytic solution be obtained. Usually it is necessary to resort to series expansion or numerical integration in order to calculate a potential distribution. Both procedures are laborious in the extreme.

The simplest potential distribution is that between two infinite parallel plates, shown in Fig. 3.1. Here the function that satisfies the differential equation is

$$\phi = \phi_0 + Ax$$

The field will be found to be:

$$E_x = -\frac{\partial \phi}{\partial x} = -A$$

$$E_y = -\frac{\partial \phi}{\partial y} = 0$$

$$E_z = -\frac{\partial \phi}{\partial z} = 0$$

Other simple cases are:

Concentric spheres: $\phi = -\frac{A}{r} + \phi_0$

$$E_r = -\frac{A}{r^2}$$

Concentric cylinders: $\phi = A \ln r + \phi_0$

$$E_r = -\frac{A}{r}$$

Other cases, such as two separated spheres, a sphere and plane, a sphere between two planes, and the corresponding cylindrical systems, can also be solved.

Problems with axial symmetry, such as are illustrated in Fig. 3.2, are of considerable importance, since, as will be shown in the next chapter, this symmetry is generally found in electron lenses. When

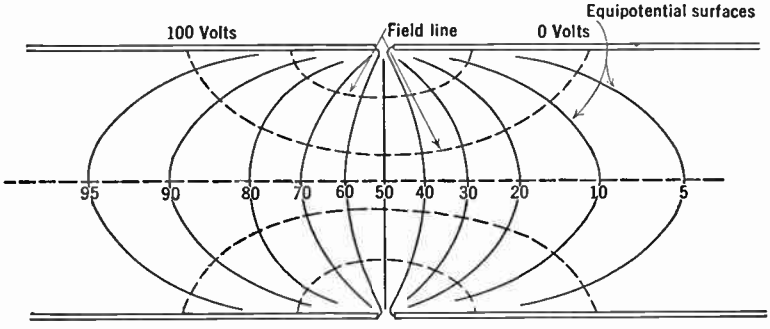


Fig. 3.2. Potential Distribution between Two Cylinders.

this symmetry is present the Laplace equation is preferably expressed in cylindrical coordinates, becoming

$$\frac{\partial^2\phi}{\partial z^2} + \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\phi}{\partial r}\right) = 0 \tag{3.2}$$

For most boundary conditions the solution of this differential equation is difficult, and no analytic solution is possible. A general method of attack is to consider the potential as a linear combination of functions in which the variables have been separated, thus:

$$\phi(r, z) = \phi_1 + \phi_2 + \cdots + \phi_k + \cdots$$

where

$$\phi_k(r, z) = F_k(z)G_k(r) \tag{3.3}$$

By substituting Eq. 3.3 in Eq. 3.2, the Laplace equation reduces to two ordinary differential equations:

$$\frac{1}{F_k}\frac{d^2F_k}{dz^2} = -k^2$$

$$\frac{1}{rG_k}\frac{d}{dr}\left(r\frac{dG_k}{dr}\right) = k^2$$

where k^2 is the separation parameter. The general solution of these equations can be written as

$$F_k(z) = a\epsilon^{ikz} + b\epsilon^{-ikz} \tag{3.4a}$$

$$G_k(r) = cJ_0(ikr) + dN_0(ikr) \tag{3.4b}$$

The solution of the Laplace equation then has the form:

$$\phi(r, z) = \sum_k A_k F_k(z) G_k(r) \tag{3.5}$$

Since k does not necessarily have discrete values, Eq. 3.5 may take the form of an integral:

$$\phi(r, z) = \int A(k) F(z, k) G(r, k)\, dk \tag{3.5a}$$

the integration being over the entire complex domain. The coefficient $A(k)$ is determined from the boundary conditions by the usual methods of evaluating Fourier coefficients.

Another class of problem of considerable importance is that in which the potential expressed in Cartesian coordinates is a function of two of these coordinates only. This type of potential field is encountered wherever the electrode surfaces can be considered as generated by moving lines which remain parallel.

The Laplace equation in this case becomes

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = 0 \tag{3.6}$$

The solution of this equation can be approached in a variety of ways. One very useful method, for example, is conformal mapping. Although this equation can be solved more frequently than that for the three-dimensional case, often no analytic solution is possible. In this two-dimensional case, practical electrode configurations are usually quite complicated, so that the mathematical complexities even of an approximate solution are prohibitive.

3.2 Electrolytic Potential Mapping. Because of the difficulties encountered in a mathematical solution of the Laplace equation, it is often expedient to resort to an electrolytic method of obtaining an equipotential map.

In essence, the method consists of immersing a large-scale model of the electrode system being studied in a slightly conducting liquid and measuring the potential distribution throughout the liquid with a probe and bridge, potentials proportional to those to be used with the system being applied to the model. Figure 3.3 shows diagrammatically an electrolytic plotting tank.

The tank used for this purpose is constructed of an insulating material so that the equipotential surfaces about the immersed electrodes meet the tank walls perpendicularly. This is necessary in order to

reduce the influence of the walls of the tank upon the field to be plotted. The size of the tank is determined by the size of the electrode models which are to be studied, and this in turn is determined primarily by the accuracy desired. In order to reduce the boundary effects, the tank must be a good deal larger than the model.

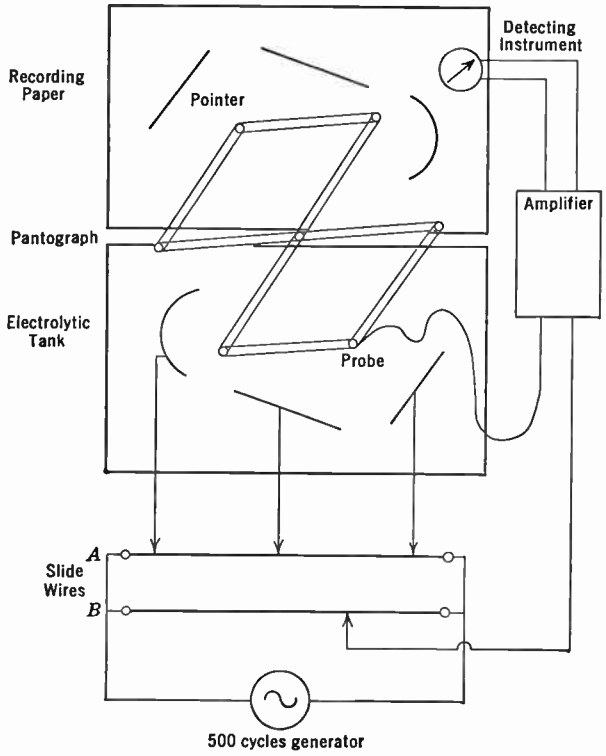


Fig. 3.3. Diagram of Potential Plotting Tank.

The electrolyte used in the tank is water to which a very small amount of soluble salt has been added. In most localities, ordinary city water contains a sufficient amount of these salts to make it amply conducting for the purpose.

Exact scale models of the electrodes, made of sheet metal and supported on insulating rods, are used in the tank. The supports should be reduced to a minimum so that they interfere with the potential distribution as little as possible. Almost every practical electrode configuration encountered in electron optics has mirror symmetry. The equipotential surfaces in space around the electrodes obviously must cross

the plane of symmetry at right angles. Because of this, as will become clear as the discussion proceeds, the models may be constructed so that they represent, to scale, half of each electrode bounded by the plane of symmetry. The model is placed in the tank in such a way that the free surface of the electrolyte coincides with the plane of symmetry. Although this is not a fundamental limitation, nearly all practical plotting tanks are limited to use with electrode systems having this symmetry.

Upon the application of the proper potentials to the model electrodes, a current will flow through the electrolyte. Since it can be assumed that the electrolyte is an ohmic conductor, the field strength at any point will be proportional to the current density. As was mentioned above, the electric current behaves like an incompressible fluid so that the current density and hence the field strength obey the equation of continuity (i.e., their divergence is zero in the absence of sources or sinks). This is merely another way of saying that the potential throughout the electrolyte obeys the Laplace equation. Thus the potential at any point in the liquid is proportional to the potential of a corresponding point in the actual electrode system.

The free surface of the electrolyte is a perfect insulating plane since no current can flow in the medium above the liquid. The equipotentials must intersect such a plane at right angles because there is zero vertical current flow, and hence the field vector normal to the surface is zero. For this reason it is possible to make use of models divided at their plane of symmetry.

The potential distribution over the plane of symmetry is measured by means of an exploring probe. This probe consists of a fine wire mounted so as to just break the surface of the liquid and is constrained to move in a horizontal plane. The potential of the probe is adjusted until zero current flows, and the potential is noted. This potential is that of the point where the probe touches the surface. For convenience, the probe is carried at the end of a pantograph linkage, so that the motion of the probe is reproduced by a stylus attached to the other end of the linkage. This stylus, or mapping pencil, moves over a plotting table. The arrangement will be clarified by reference to Fig. 3.3.

A photograph of a typical plotting tank is reproduced in Fig. 3.4. The tank itself is made of wood, coated on the inside with roofing cement to render it water-tight and shielded outside with sheet copper. It is 2½ feet wide, 8 feet long, and 2½ feet deep. Along one side is a table on which the mapping is done. Directly below the table are the

potential dividers that supply the model electrodes and probe, and behind them the amplifier whose output is connected to the null-indicating meter. The probe is attached to the pantograph pivoted at the center of the edge of the tank nearest the mapping table. A shielded lead carries the current from the probe to the amplifier.

In the example illustrated in Fig. 3.4, the probe and electrodes are supplied from a 400-cycle oscillator, instead of with direct current.

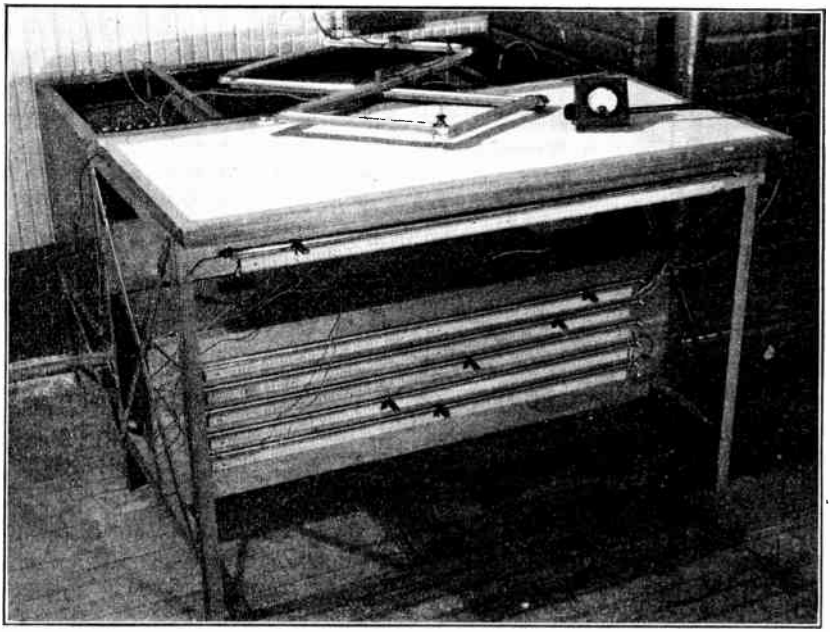


Fig. 3.4. Potential Plotting Tank.

This not only facilitates the determination of the null point, but also avoids the possibility of error due to polarization of the liquid at the probe or electrodes.

The usual procedure in mapping a potential distribution is to divide the voltage between the terminal electrodes into a convenient number of intervals, then to set the probe potential at each of these values in turn and map the path of the probe when it is moved in such a way that the current to it remains zero. The resulting map of the intersections of the equipotential surfaces corresponding to the voltage steps with the plane of symmetry of the electron-optical system is the most convenient

representation of a potential distribution for the determination of electron trajectories.

Often, in the consideration of electron lenses, to be taken up in the next chapter, it is necessary to know the axial distribution of the potential of the system, together with its first and second derivatives along the axis. The distribution can, of course, be found by direct measurement with the plotting tank. The slope of the distribution curve, plotted as a function of the axial coordinate, will give the first derivative. The second derivative can be found from the slope of the first derivative curve, but this method of obtaining it is very inaccurate. A more accurate determination can be made by means of the radius of curvature ρ of the equipotentials at the axis, the first derivative ϕ', and the relation

$$\phi''(z) = 2\phi'/\rho$$

3.3 The Motion of an Electron in a Potential Field. The potential distribution for a given electrode configuration having been obtained, the next step is the determination of the paths of electrons moving in this field.

In making this determination, it is convenient to consider the motion of an electron to be that of a charged particle of mass m obeying the laws of Newtonian mechanics, rather than to adopt the viewpoint of quantum physics and assume it to be a wave packet, as is necessary in the investigation of atomic phenomena. Furthermore, its mass will be taken as constant and equal to $9.0 \cdot 10^{-28}$ gram. Where electrons having extremely high velocities are to be considered, this assumption cannot be made, and it is necessary to correct for the increase in mass as dictated by relativity. Velocities where this correction is necessary are not encountered in the field covered by this book. The value 1.59×10^{-19} coulomb will be taken as the charge of an electron.

The force acting on an electron is the product of its charge and the field strength at the point which it occupies. Hence, the differential equations of motion are

$$m\frac{d^2x}{dt^2} = -eE_x = e\frac{\partial\phi}{\partial x}$$

$$m\frac{d^2y}{dt^2} = -eE_y = e\frac{\partial\phi}{\partial y}$$

$$m\frac{d^2z}{dt^2} = -eE_z = e\frac{\partial\phi}{\partial z}$$

In principle, in order to determine the electron path, all that is necessary is to introduce the values of the potential into the above equations, solve them, and eliminate time as a parameter.

Actually, there is no general method of carrying out this process, and it is almost always necessary to apply mathematical approximations or graphical methods to obtain a solution.

A number of practical mathematical approximations and graphical methods have been developed for the purpose of facilitating the determination of electron paths when the potential field is known. These methods, when carefully applied, are capable of yielding a high degree of accuracy.

The general three-dimensional problem is extremely difficult even by approximate methods. Fortunately, configurations requiring the solution of this general problem are rarely encountered.

In the field covered by this book, two classes of problems are of particular importance. They are:

1. Those involving electrode configurations in which the potential variation is confined to a plane.
2. Problems involving cylindrical symmetry.

The remainder of this chapter will treat the first class of problems, which are those involved in the design of the electron multiplier, deflecting plates, etc.

3.4 Electron Paths in a Two-Dimensional System. Where the potential variation is confined to a plane, the Laplace equation, as has already been pointed out, involves two coordinates only:

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = 0$$

Similarly, the laws of motion become

$$\begin{aligned} \frac{d^2x}{dt^2} &= \frac{e}{m}\frac{\partial \phi}{\partial x} \\ \frac{d^2y}{dt^2} &= \frac{e}{m}\frac{\partial \phi}{\partial y} \end{aligned} \tag{3.7}$$

This particular form of these laws is not very convenient in the present consideration, in that time enters the equations explicitly.

Taking the principle of least action as a starting point simplifies the treatment, but it should be noted that all the relations deduced below can be derived directly from the force laws of Eq. 3.7.

The principle of least action states that any particle moving between two points in a potential field will follow a path such that the integral of the momentum over this path is an extreme, either maximum or minimum. Symbolically, this principle can be written as

$$\delta \int_A^B mv\, ds = 0 \tag{3.8}$$

An electron moving in a potential field has a kinetic energy just equal to the decrease in its potential energy during its motion. If the potential is set equal to zero at a point where the electron is at rest, the following relation applies:

$$\tfrac{1}{2}mv^2 = e\phi$$

or

$$v = \sqrt{2\frac{e}{m}\phi}$$

The momentum in the action integral being represented by $\sqrt{2em\phi}$, Eq. 3.8 becomes

$$\delta \int_A^B \sqrt{\phi}\, ds = 0 \tag{3.9}$$

or

$$\delta \int_A^B \sqrt{\phi} \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx = 0 \tag{3.9a}$$

This is satisfied by a solution of the corresponding Euler differential equation:

$$\frac{d^2y}{dx^2} = \frac{1}{2\phi}\left(\frac{\partial\phi}{\partial y} - \frac{dy}{dx}\frac{\partial\phi}{\partial x}\right)\left[1 + \left(\frac{dy}{dx}\right)^2\right] \tag{3.10}$$

which is directly derivable from the variation principle. Where the numerical values of the potential as a function of x and y are known, it is possible to perform a point-by-point integration of this equation—e.g., by the method of differences—and thus determine the trajectories of electrons in this field. Although extremely laborious, this is probably the most accurate method of obtaining electron paths.

3.5 Graphical Trajectory Determination. There are graphical methods for plotting electron paths on an equipotential map which are easy, rapid, and sufficiently accurate for most practical problems. Two of these are of sufficient importance as practical design tools to be worth discussing in detail.

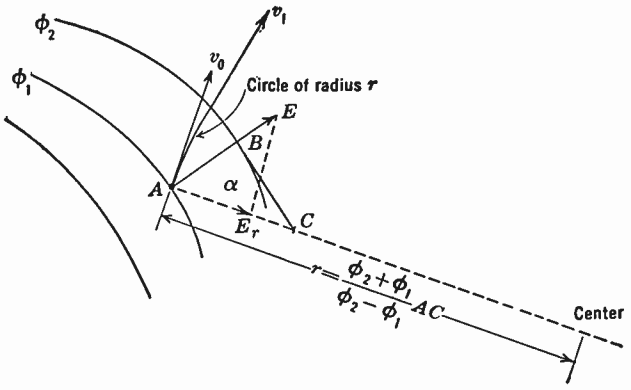


Fig. 3.5. Circle Method of Graphical Ray Tracing.

The first is a circle method. In Fig. 3.5, let it be assumed that an electron is moving in a potential field with the velocity indicated by the vector v_0. The magnitude of this velocity vector is

$$v_0 = \sqrt{2\frac{e}{m}\phi_1} \tag{3.11}$$

The electric field, E, is normal to the equipotential ϕ_1 and has a value approximately equal to

$$\frac{\phi_1 - \phi_2}{d} = E$$

One component of the field lies along the direction of motion; the other, E_r, is at right angles to this direction. The latter exerts a radial force on the electron equal to eE_r, giving rise to a centripetal acceleration:

$$\frac{v^2}{r} = \frac{eE_r}{m}$$

By solving for r and eliminating v with the aid of Eq. 3.11 this becomes

$$r = 2\frac{\phi}{E_r} \tag{3.12}$$

Accordingly, the path of the electron coincides approximately with the arc of a circle of radius r tangent to the vector v_0. Actually, this arc should be infinitesimal in length, but, since the equipotentials are close together, it may for this approximation be extended to meet the next

equipotential, ϕ_2. At ϕ_2 the velocity vector will be tangent to the arc and will have a magnitude

$$v_1 = \sqrt{2\frac{e}{m}\phi_2}$$

By repeating the procedure at the successive equipotentials, the path of the electron can be mapped.

The radius and center of the arc can be found graphically, thus avoiding the calculation indicated in Eq. 3.12. First, the approximate direction of the field vector, E, is obtained by dropping a perpendicular from A, the intersection of the path with ϕ_1, onto the equipotential ϕ_2, cutting it at B. At right angles to the line AB a line is extended until it meets the normal to the velocity vector at C. It is evident that E_r must lie along the normal to the velocity vector and that the center of the arc must also be located on this line. If the angle between E and E_r is α, then:

$$E_r = E\cos\alpha = \frac{\phi_2 - \phi_1}{AB}\cos\alpha$$

$$r = 2\frac{\phi_2}{\phi_2 - \phi_1}\frac{AB}{\cos\alpha} = 2\frac{\phi_2}{\phi_2 - \phi_1}AC \tag{3.13}$$

Somewhat greater accuracy is obtained if ϕ_2 in the numerator of Eq. 3.13 is replaced by the mean potential, giving

$$r = \frac{\phi_2 + \phi_1}{\phi_2 - \phi_1}AC \tag{3.13a}$$

With the aid of this construction, path plotting can be carried out rapidly and accurately.

When the path in question starts from a surface of zero potential it is convenient to make use of the fact that it issues normal to the surface and has an initial radius of curvature three times as large as that of the electrostatic field lines in the neighborhood of its point of origin.

The second graphical procedure is the parabola method. This method is advantageous where the curvature of the path is small, which makes the radii awkwardly large when the circle method is used. Theoretically, if the process is carried to the limit and the separation between the equipotential lines made to approach zero, either method gives the

true path. In all practical cases tested, the accuracy of the two methods is about the same.

The parabola method is based upon the fact that an electron moving in a uniform field having a velocity component at right angles to the field follows a parabolic path. It utilizes the geometric principle illustrated in Fig. 3.6*a*, namely, that the tangent to a parabola at a point at an axial distance x from its vertex meets the axis at point C, located at $-x$, or at an axial distance $2x$ from the point of tangency.

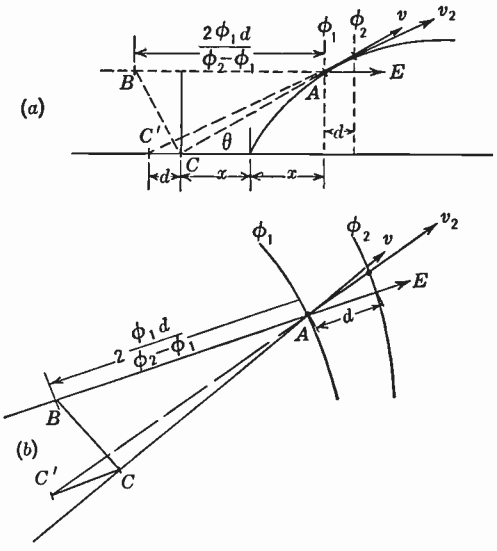


Fig. 3.6. Parabola Method of Ray Tracing.

Again, in Fig. 3.6*a*, let it be assumed that an electron at point A is moving with a velocity v, as indicated by the vector v, and that its motion is due to a uniform field. It is possible to determine the parabola giving its motion as follows: The component of velocity due to the electron having fallen through the potential field to point A is $v \cos \theta$, where θ is the angle between the velocity and field vectors. The difference of potential ϕ^* between point A and the vertex of the parabola will, therefore, be

$$\phi^* = \frac{1}{2}\frac{m}{e} v^2 \cos^2 \theta \tag{3.14}$$

If two equipotentials, ϕ_1 and ϕ_2, are separated by a distance d, the field, E, can be expressed as $(\phi_2 - \phi_1)/d$. The distance between the vertex of the parabola and the point A can consequently be written as

$$x = \frac{\frac{1}{2}\frac{m}{e}v^2 d \cos^2\theta}{\phi_2 - \phi_1}$$

$$= \frac{\phi_1}{\phi_2 - \phi_1} d \cos^2\theta \tag{3.15}$$

Equation 3.15 indicates a simple construction which will locate the point C, as shown in Fig. 3.6*a*. The field vector, E, is extended back a distance $2\phi_1 d/(\phi_2 - \phi_1)$ to B. From B a perpendicular is dropped to an extension of the velocity vector. This perpendicular will cut this vector at point C. Thus, this point can be determined from a knowledge of the vectors v and E. Furthermore, if a line parallel to E is drawn through C, and extended back a distance d to C', this new point must lie on the velocity vector for the point on the parabola where it intersects the equipotential ϕ_2. If a line is drawn through C' and A, it closely represents the path between the equipotentials ϕ_1 and ϕ_2.

To apply this construction to a general two-dimensional potential field, the procedure is as follows: Fig. 3.6*b*, the electron is at A, moving with a velocity and direction given by v. From point A, a line E representing the field direction is drawn normal to the equipotential, ϕ_1, and extends a distance d to ϕ_2. This line is drawn back from A a distance $2\phi_1 d/(\phi_2 - \phi_1)$ to point B. A perpendicular is then dropped from B onto the prolongation of v, locating point C. From C a line parallel to AB is drawn back a distance d, locating the point C'. A line through C' and A locates the position of the electron on the equipotential ϕ_2, and gives the direction of its velocity vector.

As the curvature of the path decreases, this method becomes increasingly accurate. It, therefore, is useful for determining the straighter portions of an electron trajectory, where, as has already been mentioned, the circle method becomes awkward because of the long radii involved.

The two plotting methods just discussed can be applied in any problem where the motion of the electron is confined to a plane; thus, it applies to any electrode configuration whose potential can be correctly mapped in a plotting tank of the type described.

3.6 The Rubber Model. By far the most convenient method of obtaining electron paths is by means of the rubber model. This model can be used in all problems where the potential can be expressed as a function of two rectangular coordinates, and where the electron path is

confined to the plane of these coordinates. The accuracy which can be obtained is quite high, but not quite equal to that of a path carefully plotted by the graphical methods described.

A rubber membrane, stretched over a frame, is pressed down over a model of the electrode system which is made in such a way that its plan view corresponds to the geometrical configuration of the electrodes in the $x - y$ plane while the height is proportional to the negative voltage on each electrode. The rubber is then no longer flat, nor does it follow the surfaces of the model electrodes, but rather it stretches over them in a series of mountains and valleys, touching only the top of every electrode. If care is taken that the membrane is in contact with the full length of the top edges of all the electrodes the contours of its surface are found to correspond to an equipotential map of the electrode system.

A small sphere is placed at a point corresponding to the electron source, and allowed to roll on the rubber. The horizontal projection of its path then is a map of an electron trajectory in the electrode system under investigation.

The proof that the path of the sphere correctly represents an electron trajectory is divided into two parts. First, it is necessary to show that the height of every point on the rubber model is proportional to the potential existing in the electrode system. Second, it must be proved that a sphere rolling on such a surface follows a trajectory which represents the electron motion.

In order to show that the height z of the rubber surface represents the potential distribution, it is necessary to show that the surface obeys the differential equation:

$$\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2} = 0 \tag{3.16}$$

This may readily be demonstrated if two restrictions, which are more or less fulfilled in practice, are imposed on the rubber surface. They are: (1) that the slope of the surface be everywhere small, (2) that the tension of the deformed rubber be uniform over the surface.

The most straightforward proof applies the principle of minimum energy. Since the energy in any region is proportional to the area, the area integral of the surface must be a minimum. Transforming this minimized integral into the Euler form leads to the differential equation required.

The physical significance of the shape assumed by the surface is made more apparent by the following less rigorous demonstration.

Figure 3.7 shows an element of surface area, ds, together with the forces acting on it. It is obvious that the vector sum of these forces

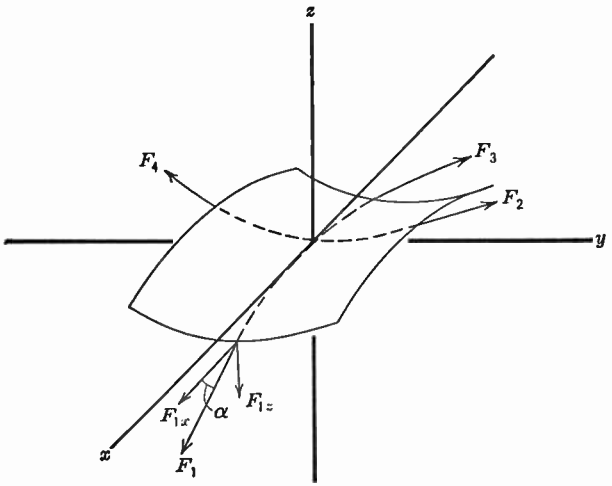


Fig. 3.7. Forces Acting on an Element of a Stretched Membrane.

must be zero since the element is in equilibrium. In order to set up the conditions of equilibrium, the four forces F_1, F_2, F_3, and F_4 must each be resolved into their components along the coordinates. Considering first F_1, it is evident from the figure that

$$F_{1x} = F_1 \cos \alpha = F_1 \frac{dx}{dx} \frac{1}{\sqrt{1 + \left(\frac{dz}{dx}\right)^2}} \tag{3.17a}$$

$$F_{1z} = F_1 \sin \alpha = F_1 \frac{dz}{dx} \frac{1}{\sqrt{1 + \left(\frac{dz}{dx}\right)^2}} \tag{3.17b}$$

where dz, dx, and α are as indicated. By the first restriction, dz/dx is small, so that

$$\frac{1}{\sqrt{1 + \left(\frac{dz}{dx}\right)^2}} \cong 1$$

and the components become

$$F_{1x} \cong F_1$$

$$F_{1z} \cong F_1 \left(\frac{dz}{dx}\right)_1$$

Similarly,

$$F_{3x} \cong F_3$$

$$F_{3z} \cong F_3 \left(\frac{dz}{dx}\right)_3$$

The two x components must be equal in magnitude and opposite in sign, hence

$$F_{3x} \cong -F_{1x}$$

$$F_3 \cong -F_1$$

and

$$F_{3z} = -F_1 \left(\frac{dz}{dx}\right)_3$$

Summing the z components, the upward force due to F_1 and F_3 is

$$F_{1z} + F_{3z} = F_1 \left[\left(\frac{dz}{dx}\right)_1 - \left(\frac{dz}{dx}\right)_3\right] = F_1 \frac{\partial^2 z}{\partial x^2} \Delta x \qquad (3.18)$$

where Δx is the length of the element in the x direction.

Applying the second restriction, that the force F_1 must equal the tension δ of the membrane times the width Δy of the element in the y direction, Eq. 3.18 becomes

$$F_{1z} + F_{3z} = \delta \frac{\partial^2 z}{\partial x^2} \Delta x \, \Delta y \qquad (3.18a)$$

In like manner, the vertical components of the other two forces are

$$F_{2z} + F_{4z} = \delta \frac{\partial^2 z}{\partial y^2} \Delta y \, \Delta x$$

Finally, since the sum of all the z force components must be zero, and δ, Δx, and Δy are not zero, the equation

$$\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2} = 0$$

must be true, proving that z satisfies the Laplace equation with the boundary conditions determined by the electrode heights.

It may be mentioned at this point that the slope of the rubber surface is everywhere proportional to the field strength, and that the force exerted on any electrode by the rubber is proportional to the capacity of that electrode.

The next problem is to show that a body moving under the action of gravity on the rubber surface moves along a path which corresponds to the electron trajectory. For simplicity, let it be assumed that the body in question slides on the surface, and that its friction is negligible.

By the principle of least action, the action integral must be stationary, or

$$\delta \int_A^B mv\, ds = 0 \tag{3.8}$$

Since the system is conservative and, hence, the sum of the kinetic and potential energies must remain constant, the momentum mv can be found as follows:

$$\text{K.E.} + \text{P.E.} = \text{const}$$

$$\tfrac{1}{2}mv^2 = -mgz$$

$$mv = \text{const}\,\sqrt{z}$$

By substitution Eq. 3.8 becomes

$$\delta \int_A^B \sqrt{z}\, ds = \delta \int_A^B \sqrt{z}\, \sqrt{1 + \left(\frac{dy}{dx}\right)^2 + \left(\frac{dz}{dx}\right)^2}\, dx = 0 \tag{3.19}$$

It has already been assumed that the slope of the rubber is small so that $(dz/dx)^2$ can be neglected compared with unity; therefore the final form of the action integral is

$$\delta \int_A^B \sqrt{z}\, \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx = 0 \tag{3.19a}$$

Except for z replacing ϕ, Eq. 3.19a is seen to be identical with Eq. 3.9a. In the previous derivation it was shown that the height, z, of the rubber is proportional to the potential ϕ. Therefore, the path of the body sliding on the rubber is geometrically similar to the corresponding electron trajectory.

If, instead of the motion of a sliding body, that of a rolling sphere is considered, the conclusions are the same, provided that the assumption is made that the radius of curvature R of the sphere is small compared

with the radius of curvature of the rubber. This is shown by deriving the total kinetic energy as follows:

Let $d\alpha$ be a small rotation of the sphere. Then the displacement, ds, of the center of the mass is given by

$$ds = R\, d\alpha$$

The angular velocity in terms of the linear velocity is thus

$$\omega = \frac{d\alpha}{dt} = \frac{1}{R}\frac{ds}{dt} = \frac{v}{R}$$

Next, the sum of the rotational and translational kinetic energies is expressed as follows:

$$\text{K.E.} = \frac{1}{2}(mv^2 + I\omega^2) = \frac{1}{2}\left(m + \frac{I}{R^2}\right)v^2$$

$$= \tfrac{1}{2}m^*v^2$$

where m^* is the effective mass.

As before, the total momentum can now be written as $\sqrt{\bar{z}}$ times a constant. Hence, Eq. 3.19*a*, expressing the principle of least action, is unchanged, and the path of the rolling sphere indicates the desired path.

Certain assumptions were made in deriving the paths on the rubber model, and the question might well arise as to how closely these assumptions must be fulfilled in order not to introduce serious errors in the final results. As a result of a large number of tests on the model, the indications are that, even if the slopes become as great as 30 to 45 degrees, and the tension in the rubber very far from uniform, the paths obtained will be sufficiently accurate for all practical purposes. In fact, the presence of friction, which has been neglected in the foregoing derivation (except the implied friction required to produce rolling), makes it advisable to use fairly steep slopes.

A practical form of the rubber model is shown in Fig. 3.8. Ordinary surgical rubber is stretched over a square frame, which is about 3 feet on a side. Usually the electrode models are made from soft metal strips, either lead or aluminum, which are cut to the correct height to represent the potential, and bent to conform with the electrode shape. The table supporting the electrodes is built of welded angle iron and has a plate-glass top. The glass top permits illumination from below, which greatly facilitates the placing of the electrodes. Since it is often necessary to press the rubber down to make it come in contact with the

more positive electrodes, the table is equipped with movable side arms to which can be clamped top electrode models.

It has been found convenient to use 3/16-inch steel ball bearings for the spheres. These have an advantage over glass spheres in that they

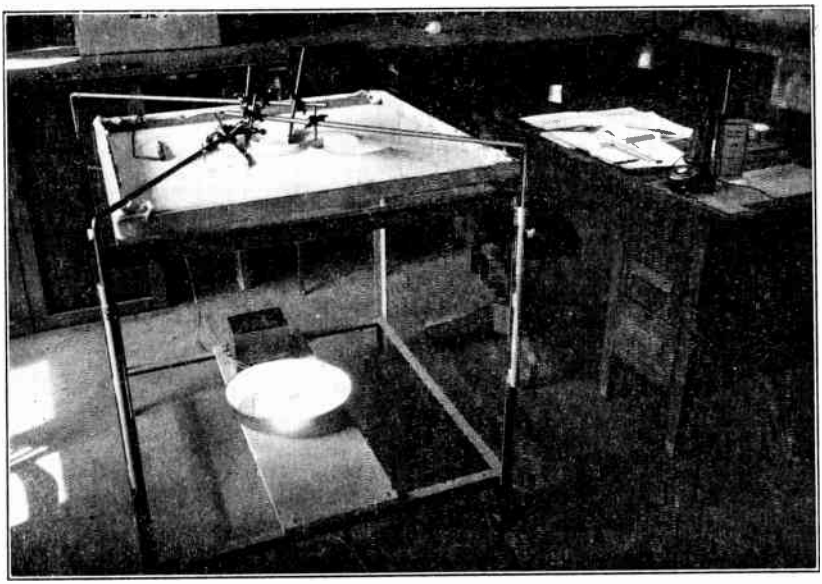


Fig. 3.8. Rubber Model for the Study of Electron Paths.

can be held at the top of the cathode electrode with a small electromagnet and released when desired, without any danger of deflecting their course, merely by cutting off the current to the magnet.

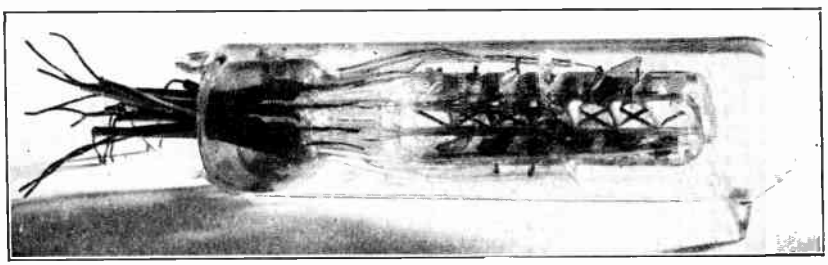


Fig. 3.9. An Experimental Electrostatic Secondary Emission Multiplier.

Where a permanent record of the path is desired, a time exposure for the duration of the motion of the sphere can be made. In this case it is better to use black rubber for the model, and to illuminate it from

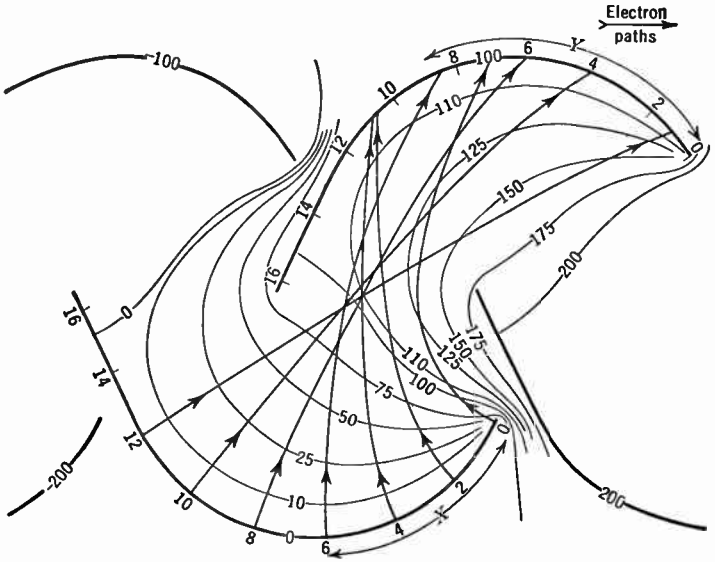


Fig. 3.10. Electron Paths in a Multiplier of the Type Shown in Fig. 3.9 as Determined from Potential Plot.

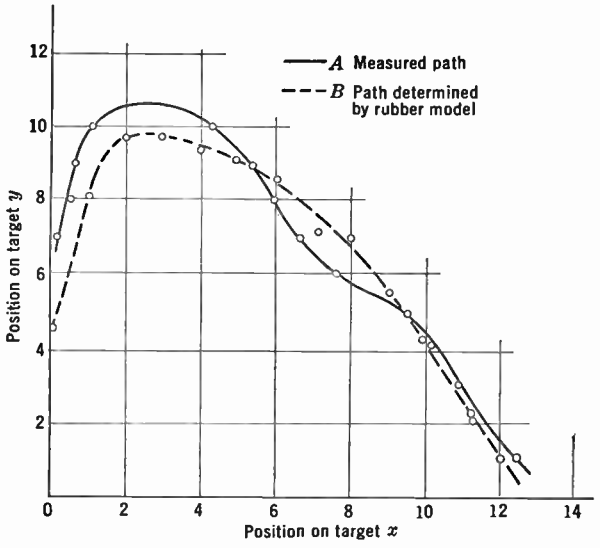


Fig. 3.11. Curves Showing Termini of Electron Trajectories in the Multiplier as Determined by Direct Measurement and from a Rubber Model.

above. Furthermore, with a pulsating light source, such as a 60-cycle cold arc, the paths will appear as dotted lines. The spacing between the dots is a measure of the velocity of the electron.

Numerical values for the accuracy that can be attained with the rubber model are difficult to give, since the error depends upon the electron path. Measurements of parabolic paths obtained on such a model were made by P. H. J. A. Kleynen, who found an error of 1 percent in the height of the parabola, and one of 7 percent in the separation between the arms of the parabola when the sphere returned to its initial potential.

Results obtained with the model when used in connection with the design of the electrostatic multiplier shown in Figs. 3.9 and 3.10 indicate its excellence. Plots were made of the initial and terminal points of a number of electrons, as indicated by the model, and again with an actual electron tube. Two curves of this type are reproduced in Fig. 3.11, showing fairly close agreement between the two methods of measurement.

REFERENCES

1. J. H. Jeans, *Electricity and Magnetism,* Cambridge University Press, 1927.
2. S. S. Attwood, *Electric and Magnetic Fields,* 3rd Ed., Wiley, New York, 1949.
3. E. Weber, *Electromagnetic Fields: Theory and Application,* Wiley, New York, 1950.
4. O. D. Kellogg, *Foundations of Potential Theory,* Julius Springer, Berlin, 1929.
5. D. Gabor, "Mechanical Tracer for Electron Trajectories," *Nature,* Vol. 139, p. 373, Feb. 27, 1937.
6. H. Salinger, "Tracing Electron Paths in Electric Fields," *Electronics,* Vol. 10, pp. 50–54, October, 1937.
7. P. H. J. A. Kleynen, "The Motion of an Electron in a Two-Dimensional Electrostatic Field" *Philips Tech. Rev.,* Vol. 2, pp. 321–352, 1937.
8. J. R. Pierce, "Electron Multiplier Design," *Bell Tel. Record,* Vol. 16, pp. 305–309, 1938.
9. V. K. Zworykin and J. A. Rajchman, "The Electrostatic Electron Multiplier," *Proc. I. R. E.,* Vol. 27, pp. 558–566, September, 1939.

4 Electron Optical Systems

The electron-optical systems to be considered in this chapter are those with axially symmetric field-producing elements. Electron lenses in the usual sense are based on configurations of this type. Furthermore, it may be stated quite generally that any electrostatic or magnetic field which has axial symmetry is capable of forming a first-order image, either real or virtual.

Electron lenses have achieved great practical importance in the last two decades. They not only find employment in a wide variety of scientific apparatus, but also play a leading role in electronic television and electron microscopy.

4.1 Applications of Electron Lenses. The earliest application of an electron lens is, unquestionably, the formation of a small electron spot on the screen of a cathode-ray tube. This spot constitutes the image of an emitting cathode, of a small aperture in a diaphragm, or of a narrow cross-section of an electron beam. The electron-optical system employed for forming the spot is designated as an electron gun. Electron guns employed in television camera and viewing tubes will be discussed in some detail in Chapter 12. They constitute an element of central importance in present-day electronic television systems.

The image tube is another relatively simple application of electron optics. As part of various types of television camera tubes, such as the image iconoscope and the image orthicon, it serves to increase the sensitivity of these tubes. In essence, an image tube is made up of an extended photocathode, a fluorescent screen maintained at a relatively high positive potential with respect to the photocathode, and an electric or magnetic electron lens which images the photocathode onto the fluorescent screen. If a light image is projected on the photocathode, a similar image is reproduced on the screen by the accelerated photoelectrons falling on it. Figure 4.1 shows an electrostatic image tube designed for observing objects illuminated by infrared radiation. The photocathode is here infrared-sensitive, and the relatively small fluorescent screen is observed through a magnifier. Image tubes, with high-

efficiency phosphor screens and operated at high voltages, may produce fluorescent images which are many times as bright as the original object.

The electron microscope constitutes the most elaborate application of the principles of electron optics. As indicated by Fig. 4.2, a modern magnetic electron microscope is a perfect analogue of a compound

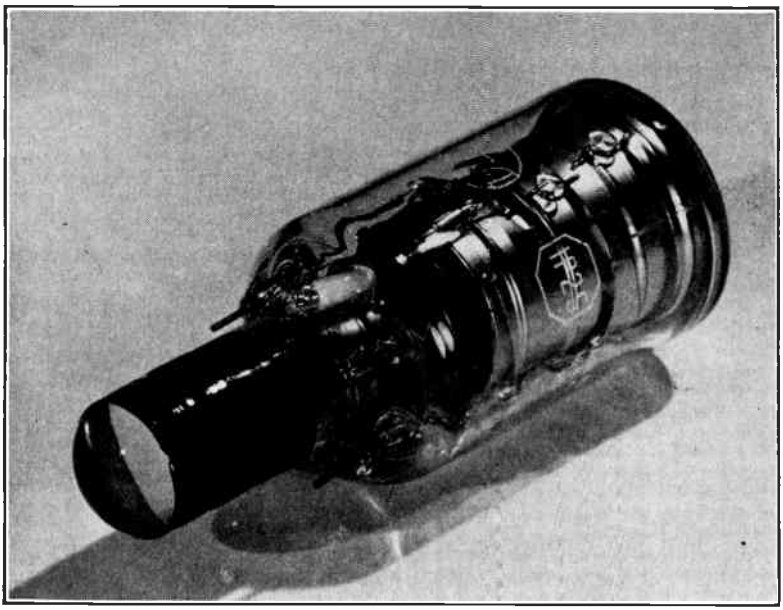


Fig. 4.1. Electrostatic Image Tube (Fluorescent Screen in Foreground).

projection microscope. The object, which is here generally either in the form of a very thin film or deposited on a thin supporting membrane, is "illuminated" by a beam of high-speed electrons. These electrons are given off by a heated tungsten hairpin filament and accelerated by a potential drop of 30 to 100 kilovolts. A short focal length magnetic "objective" focuses the beam, modified by passage through the object, into a first, or intermediate, image, which in turn is imaged by a "projector" lens into a more highly magnified final image. The last, which may have a magnification up to about 30,000, may be either viewed directly on a fluorescent screen or recorded by the electrons on a photographic plate. An additional optical enlargement of the plate so obtained, by a factor of 5 or more, may be needed to reveal all the detail recorded thereon.

Figure 4.3 shows a highly flexible type of magnetic electron microscope specifically designed for research purposes, with a magnification range of 100 to 30,000. The lenses are here excited by stabilized d-c power supplies. A much smaller, portable instrument, with permanent-magnet lens excitation and a fixed magnification of about 6000, is

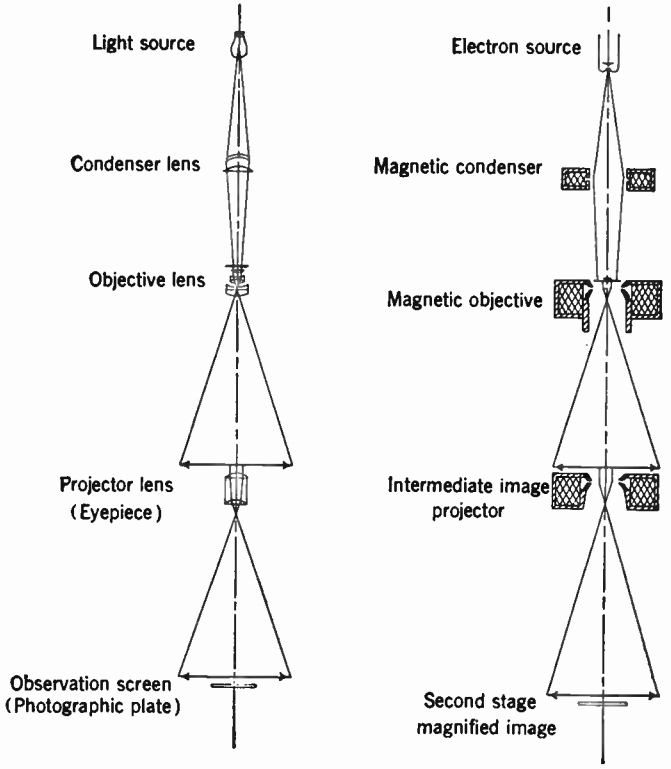


Fig. 4.2. Ray Paths in a Light Microscope and a Magnetic Electron Microscope. (Zworykin, Morton, Ramberg, Hillier, and Vance, reference 4.)

shown in Fig. 4.4. This is especially convenient for routine checks, being extraordinarily simple in operation.

Electron microscopes with similar general properties may also be constructed with electrostatic lenses, at the expense of a certain loss in ultimate resolution and magnification obtainable with a given instrument size and number of lenses. An example of a modern electrostatic electron microscope is shown in Fig. 4.5.

All these instruments, like the ordinary light microscope, find application in a broad range of fields, particularly biology, chemistry,

and metallurgy; in the last instance a thin-film replica of the etched-metal surface is prepared for examination in the electron microscope.

The shadow electron microscope, in which a very fine point source of electrons is employed to project an image of a closely spaced object on

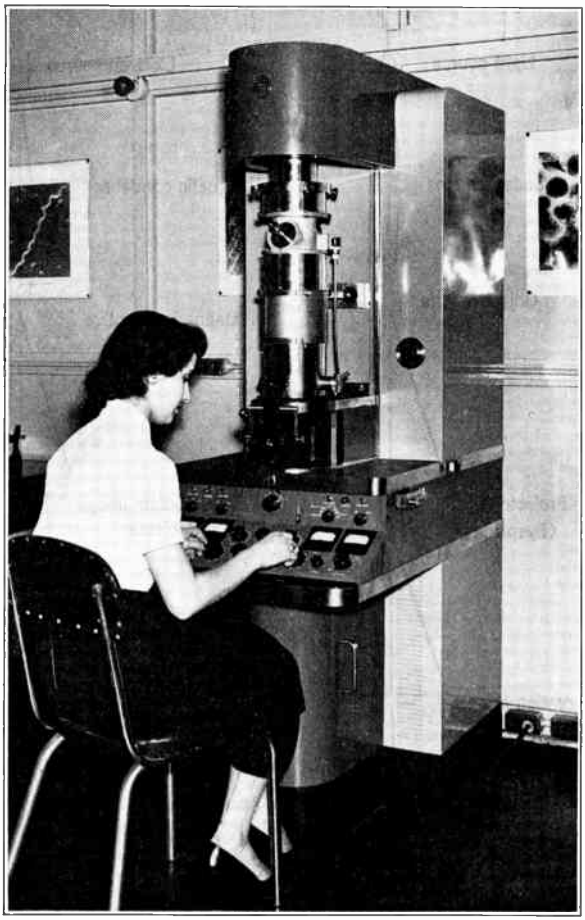


Fig. 4.3. RCA Electron Microscope, Type EMU.

a distant screen or plate, shares this property, but has attained only limited practical importance because of unfavorable intensity conditions. Most other electron microscopes are distinctly special-purpose instruments. The scanning electron microscope, in which the specimen

is explored by a fine electron probe and the image is reconstructed by television techniques, is particularly suitable for the examination of opaque surfaces; even so, it is generally less effective and more complex than the standard electron microscope employed with a replica tech-

Fig. 4.4. RCA Electron Microscope, Type EMT.

nique. Other instruments utilize thermionic, photoelectric, secondary, or field emission to form images of the emitting surfaces. Such emission microscopes are employed for studying cathode properties, changes in the crystal structure of metals at high temperatures, and the effects of adsorbed atoms and ions on electron emission. If the emitting surface is flat, the electrons are generally accelerated by a strong field and then imaged by ordinary electron lenses. On the other hand, if the cathode is cylindrical—for instance, a fine wire—or spherical or

hyperboloidal, either axial or central projection may be employed to form a magnified electron image of the surface on a cylindrical or

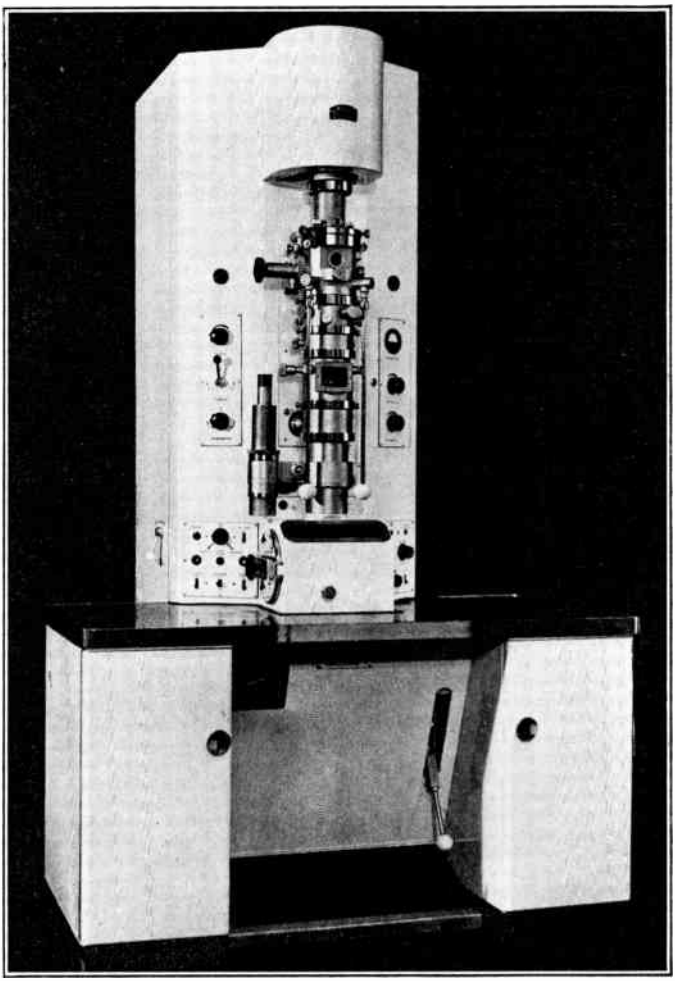


Fig. 4.5. AEG-Zeiss Electrostatic Electron Microscope. (Courtesy Süddeutsche Laboratorien.)

spherical fluorescent screen. Examples of such emission microscopes are shown schematically in Fig. 4.6. Hyperboloidal emission surfaces with a radius of curvature of the order of one micron may be obtained by etching tungsten wire in a sodium nitrite melt. Copious field emission is obtained from such a point if it is mounted in a cathode-ray

tube envelope and a difference of potential of a few kilovolts is applied between it and the fluorescent screen.

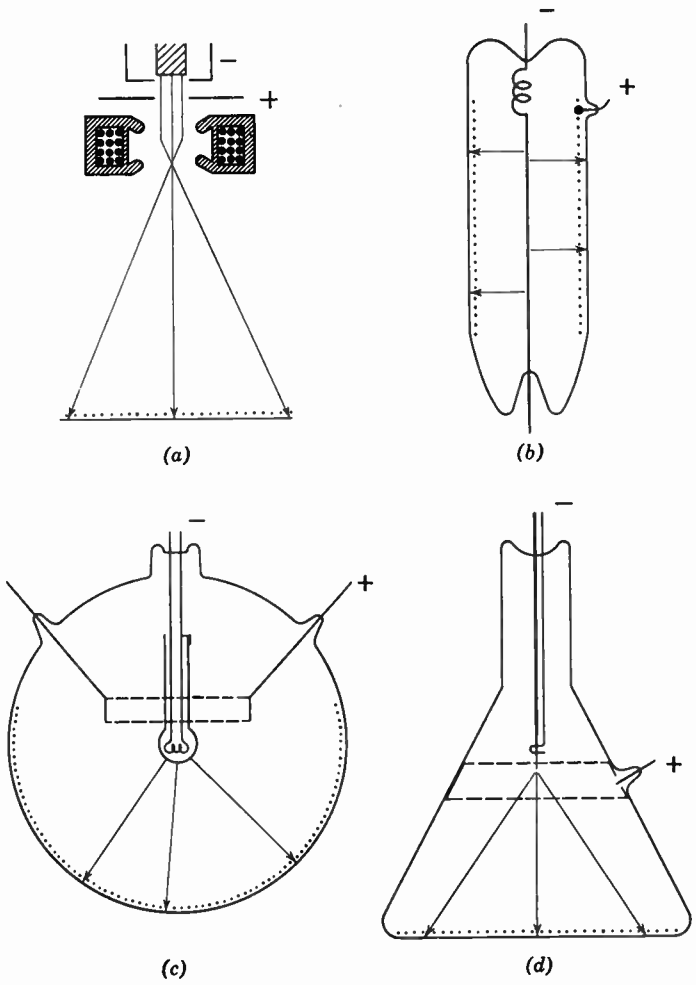


Fig. 4.6. Emission Microscopes for (*a*) Flat Cathode, (*b*) Wire, (*c*) Spherical Cathode, and (*d*) Point Cathode.

4.2 Optical Principles. When a ray of light passes through a boundary between two media in which the velocity of light differs, the ray is bent by a process known as refraction. The law governing this refraction is the well-known Snell's law:

$$n \sin \beta = n' \sin \beta' \tag{4.1}$$

where β and β' are the angles which the incident and refracted ray make with the normal to the boundary between the media having refractive indices n and n', respectively.

If the boundary separating the two media is a section of a spherical surface, it will act as a lens surface. Such a surface can be shown to have image-forming properties. By this is meant that, if the light rays from a small object enter the spherical refracting surface, the rays from any point will be bent in such a way that they converge on, or appear to diverge from, a second point known as its image point. Furthermore, the image points will be ordered in the same way as the emitting

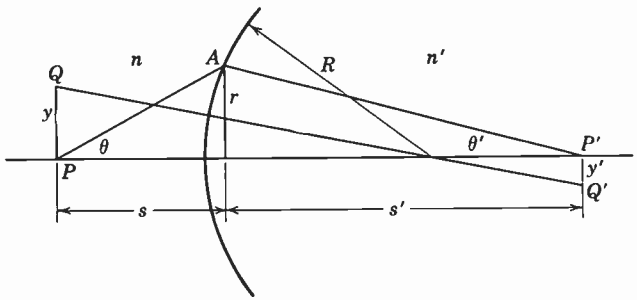


Fig. 4.7. Refraction at a Curved Surface.

or object points, so that an image is formed of the original object. Where the rays travel so that they actually converge on the image points, the image formed is said to be real; if it is necessary to extend the rays backward to the point from which they appear to diverge, the image is virtual.

To prove the existence of the image-forming property of these spherical surfaces, it is necessary to show, first, that the rays from an object point converge on an image point, and, second, that the ordering of object and image points is similar. The carrying out of this demonstration requires the imposing of two restrictions, namely, that the object be small, and that the rays make very small angles with the axis of the system. Rays which meet these requirements are known as paraxial rays, and the image theory based on these restrictions is called the first-order theory, or Gaussian dioptrics.

In Fig. 4.7, P is the object point at a distance s from the spherical boundary of radius R. The ray PA emitted at an angle θ is refracted so that it reaches the axis at s', making an angle θ' with the axis. From the geometry of the figure, it is obvious that the angles of incidence and refraction are

$$\beta = \theta + \frac{r}{R}$$

$$\beta' = -\theta' + \frac{r}{R}$$

Since, under the restrictions imposed, the angles of incidence and refraction are small, the sines appearing in Snell's law may be replaced by the angles themselves. Equation 4.1 thus becomes

$$n\beta = n'\beta' \tag{4.2}$$

and, if the expressions for β and β' are substituted,

$$n\theta + n'\theta' = \frac{r}{R}(n' - n) \tag{4.3}$$

However, $\theta = r/s$ and $\theta' = r/s'$, so that Eq. 4.3 can be written as follows:

$$\frac{n}{s} + \frac{n'}{s'} = \frac{n' - n}{R} \tag{4.4}$$

Since neither r nor θ appears in this expression, the equation proves that any ray leaving P must converge on the point P' at a distance s' from the surface. The point P' is therefore the image point of the object point P.

If next a particular ray QQ', which is normal to the refracting surface, is considered, it is evident from the similar triangles formed that

$$\frac{y'}{y} = \frac{s' - R}{R + s}$$

Combined with Eq. 4.4, this becomes

$$\frac{y'}{y} = \frac{ns'}{n's} \tag{4.5}$$

Since y'/y is the ratio of the separation of the two image points $P'Q'$ and that of the two object points PQ, it is the lateral magnification m of the image. This magnification is independent of which point on the image is chosen; therefore, the ordering of the image and object points will be similar, and the second condition for image formation is fulfilled.

Finally, substitution of the relation $s'/s = \theta/\theta'$ in Eq. 4.5 leads to the equation

$$ny\theta = n'y'\theta' \tag{4.6}$$

This equation is a statement of the Lagrange law for the first-order image formation by any number of refracting surfaces.

The formulas derived above are sufficient to determine the size and position of the first-order image formed by any number or type of coaxial curved refracting surfaces. For a pair of surfaces of radius R and R', separated by a distance which is negligibly small compared with the object and image distances and enclosing a medium of refractive index n, two successive applications of Eq. 4.4 lead to

$$\frac{1}{s} + \frac{1}{s'} = (n - 1)\left(\frac{1}{R} + \frac{1}{R'}\right) = \frac{1}{f} \tag{4.7}$$

The constant on the right-hand side of the equation is known as the refractive power, or the reciprocal of the focal length f, of the "thin lens" here considered. If the index of refraction in object space and in image space differs, that is, is n and n' respectively, instead of being unity as assumed above, the lens equation takes the form

$$\frac{n}{s} + \frac{n'}{s'} = \frac{n}{f} = \frac{n'}{f'} \tag{4.8}$$

It is thus evident, from setting either s or s' equal to infinity, that a ray parallel to the axis in object space intersects the lens axis at a

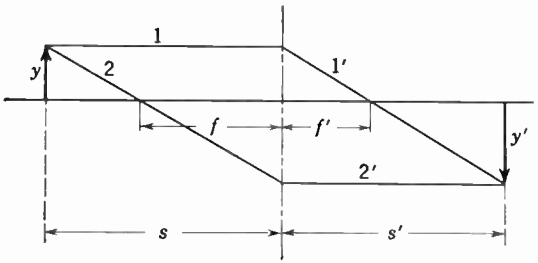


Fig. 4.8. Properties of a Thin Lens.

distance from the lens equal to f' and that a ray parallel to the axis in image space diverges from a point at a distance f from the lens in object space (Fig. 4.8). From the geometry of the figure and from Eq. 4.8 the magnification becomes

$$m = -ns'/(n's) \tag{4.9}$$

4.3 The Thick Lens. The relations developed above can generally be applied with a fair degree of accuracy to lenses whose focal length

is long compared with their thickness. If the thickness of the lens is not negligible compared to the focal length, Eqs. 4.8 and 4.9 still apply, provided that the focal lengths and object and image distances are measured from appropriate reference planes, the "principal" planes of the lens. Indeed, they apply not only to a single lens, but to a system consisting of any number of lens elements. Either the individual lens or the compound system may be described as a "thick lens."

The two principal planes of the thick lens are the conjugate planes for which the optical system has a positive magnification of unity. In other words, an object at one principal plane produces at the other a

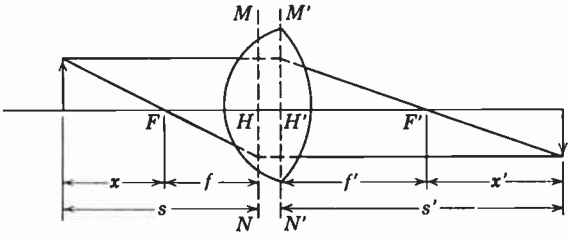


Fig. 4.9. Image Formation by a Thick Lens.

virtual (erect) image which is the same size as the object. The intersections of these planes with the axis of the system are known as the principal points. Figure 4.9 shows a thick-lens optical system, the planes MN and $M'N'$ being the two principal planes, and H and H' the principal points. Any ray of light entering the lens from object space parallel to the axis will be bent in such a way that it crosses the axis at the point F', the image-side focal point. The parallel ray and the ray through F' will, if extended, meet at the principal plane of image space, $M'N'$. Similarly, any ray through the object-side focal point F will emerge from the lens system parallel to the axis. These two rays, when extended, meet in the principal plane of object space, MN. The two focal points, together with the two principal points, are known as the cardinal points of the lens system. Once they have been located, the first-order image of any object can readily be found. The distances f and f', between the focal points and the corresponding principal planes, are known as the object-space and image-space focal lengths.

From the geometry of Fig. 4.9, it is evident that the lateral magnification m of the system is given by

$$m = -\frac{f}{x} = -\frac{x'}{f'} \tag{4.10}$$

where x and x' are the object and image distances from the focal points. From this, it follows that

$$xx' = ff' \tag{4.11}$$

The magnification and position can, of course, be referred to the principal points instead of the focal points. If s and s' are the distances of the object and image from the principal planes, their values are $s = x + f$ and $s' = f' + x'$. Thus Eq. 4.11 leads to

$$(s - f)(s' - f') = ff'$$

or

$$\frac{f}{s} + \frac{f'}{s'} = 1 \tag{4.12}$$

In the same terms, the magnification is

$$m = -\frac{f}{s - f} = \frac{s' - f'}{f'} \tag{4.13}$$

From an application of the law of Lagrange (Eq. 4.6) to the case of an object located a great distance from the lens it may be shown that

$$\frac{f}{f'} = \frac{n}{n'} \tag{4.14}$$

Substitution of this relation in Eqs. 4.12 and 4.13 leads directly to Eqs. 4.8 and 4.9 of the preceding section.

4.4 Index of Refraction in Electron Optics. From the foregoing, it is evident that the concept of the index of refraction is important in optics. In electrostatic electron optics, the potential, or rather the square root of the potential, plays the role of the index of refraction. The following simple example illustrates the similarity. Consider an electron moving from a region at potential ϕ_1 through a narrow transition region into a region at potential ϕ_2. Figure 4.10 shows the regions A and C to be field-free and at potential ϕ_1 and ϕ_2, respectively, whereas in the intervening transition space B there exists a field of magnitude $(\phi_2 - \phi_1)/d$ normal to the boundary sheets. The boundaries separating the three regions are assumed to be conducting sheets which are transparent to electrons. Such sheets are,

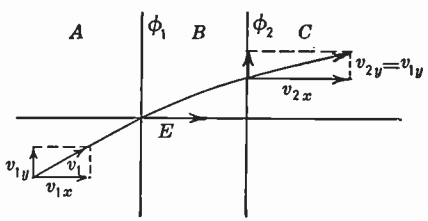


Fig. 4.10. Refraction by Potential Double Layer.

of course, purely fictitious and can be only approximated in practice; however, they are useful for illustrative purposes. The incident electron approaching the first sheet has a velocity

$$v_1 = \sqrt{2\frac{e}{m}\phi_1}$$

If it makes an angle θ_1 with the surface normal, its velocity can be resolved into two components, which are

$$v_{1x} = \sqrt{2\frac{e}{m}\phi_1}\cos\theta_1, \qquad v_{1y} = \sqrt{2\frac{e}{m}\phi_1}\sin\theta_1 \tag{4.15}$$

where x is normal to the boundary sheets and y is parallel to them. As the electron traverses the transition region, it is accelerated in the x direction by the field, while its y component of velocity remains unchanged. The total velocity of the emerging electron will, of course, be

$$v_2 = \sqrt{2\frac{e}{m}\phi_2} \tag{4.16}$$

Since the angle of emergence θ_2 is given by $\sin\theta_2 = v_{2y}/v_2$ and, furthermore, $v_{2y} = v_{1y}$, Eqs. 4.15 and 4.16 lead to

$$\sqrt{\phi_1}\sin\theta_1 = \sqrt{\phi_2}\sin\theta_2 \tag{4.17}$$

If this is compared to Snell's law (Eq. 4.1), it will be seen that $\sqrt{\phi}$ is the exact counterpart of the index of refraction n.

This correspondence is quite general, as can be proved by comparing Fermat's principle of optics:

$$\delta\int_A^B n\,ds = 0 \tag{4.18}$$

and the principle of least action for an electron in an electrostatic field:

$$\delta\int_A^B v\,ds = \text{const}\;\delta\int_A^B \sqrt{\phi}\,ds = 0 \tag{4.19}$$

4.5 Simple Double-Layer Lens. The simplest concept of an electron lens is probably that formed by two curved double layers of the type just described. The arrangement is illustrated in Fig. 4.11. The double layer on the object side is assumed to have a radius of curvature R_1 and that on the image side a radius R_2. The potential of the inner surfaces is ϕ_2; that of the outer ϕ_1.

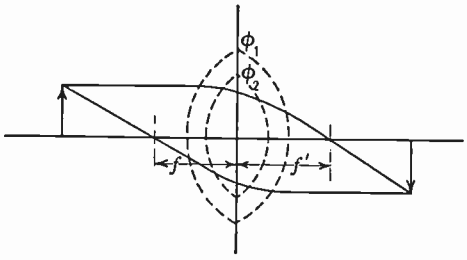


Fig. 4.11. Double-Layer Lens.

This lens corresponds exactly to an optical thin lens with an index of refraction $n_2 = \sqrt{\phi_2}$ immersed in a medium which has an index of refraction $n_1 = \sqrt{\phi_1}$. Its focal length can be calculated in elementary fashion and is given, in analogy with Eq. 4.7, by

$$\frac{1}{f} = \left(\frac{\sqrt{\phi_2}}{\sqrt{\phi_1}} - 1\right)\left(\frac{1}{R_1} + \frac{1}{R_2}\right) \tag{4.20}$$

An electron source, located at a distance s from this lens, emitting electrons with a velocity $\sqrt{2(e/m)\phi_1}$, will be imaged at a distance s'. These distances are related by the equation

$$\frac{1}{s} + \frac{1}{s'} = \frac{1}{f}$$

Double-layer lenses may be approximated with the aid of fine-mesh spherical screens. However, the approximation is crude and does not permit precise imaging.

4.6 Continuous Lenses. If the type of lens described were the only kind that existed, electron lenses would have little practical value. There is a second, fundamentally different class, and upon this the importance of electrostatic electron lenses depends. The operation of these lenses is based on the following fact. Whenever there exists a region in which there is a varying electric field having axial symmetry, this region will have properties analogous to those of an optical lens system; that is, it will be capable of forming a real or virtual image of an emitting source.

It is evident that a lens formed in this way is very different from the familiar glass lens treated in conventional optics. Instead of changing abruptly at boundaries between media of different index of refraction, the index varies continuously, both along the axis and radially.

This being so, it has been found necessary to apply different mathematical methods in calculating the lens properties of any specific field. These properties, once they have been found, can be represented by the same four cardinal points which describe an optical thick lens, namely, by two focal points and two principal points.

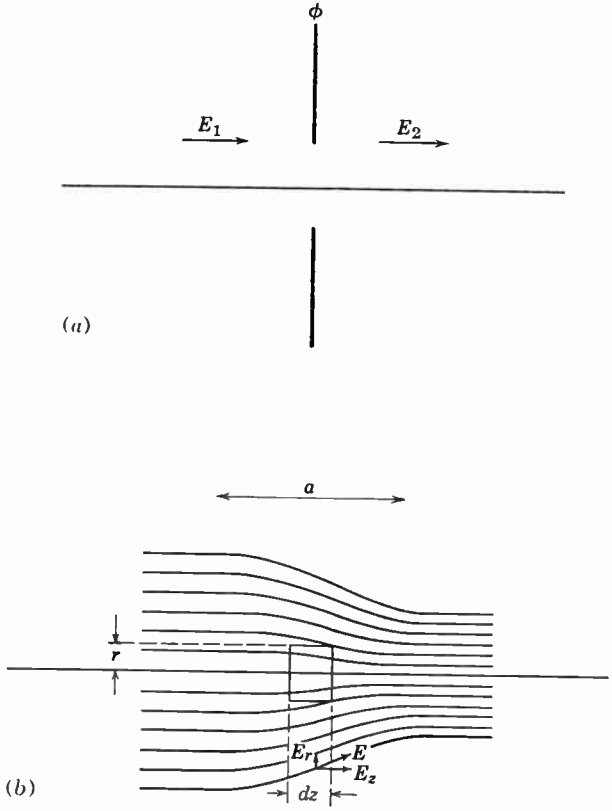


Fig. 4.12. Idealized Field Distribution in Simple Aperture Lens.

The simplest lens having continuously variable index of refraction is that formed by an axially symmetric transition region between two constant fields of different magnitude. Physically, such a region is approximated by that in an aperture having different field strengths on the two sides. This is illustrated in Fig. 4.12*a*.

In order to obtain a physical picture of the action of this lens, it is convenient to make use of the concept of lines of electrostatic force, or field lines, whose direction at any point is that of the field and whose

density is proportional to the field strength. A small volume element in the transition region of the lens is shown in Fig. 4.12*b*. In this region the Laplace equation is obeyed, and, as was pointed out in the preceding chapter, this requires that the number of field lines which enter the volume element must equal the number which leave. It is evident that, since the density of lines is different on the two ends, lines must enter or leave through the mantel surface of the volume. Figure 4.12*b* shows that this condition can be fulfilled only if the field lines are curved. Accordingly, the field lines are not parallel to the axis, and the field can be resolved into radial and axial components. The radial component of the field will deflect any electron passing through the transition region toward (or away from) the axis. This bending action increases with radial distance from the axis, as is required for image formation.

The quantitative properties of the lens can be determined as follows.* First, by applying the Laplace equation, the radial field component of the field is obtained. From the radial field component, the change in radial momentum of an electron passing through this region is calculated. This leads to the change in angle of the electron path and, hence, directly to the focal length of the system.

Consider a small cylindrical element of volume about the axis of symmetry of radius r and thickness dz. In view of the validity of Laplace's equation, the number of field lines leaving through the mantel surface must be equal to the number entering through the end surfaces, or

$$2\pi r\, dz\, E_r = \pi r^2[E_z(z) - E_z(z + dz)] \tag{4.21}$$

where E_r is the radial component of the field at a distance r from the axis, and $E_z(z)$ is the longitudinal component, averaged over the end surface at z. Replacing the parenthesis by $(\partial E_z/\partial z)\, dz$, this leads to

$$E_r = -\frac{r}{2}\frac{\partial E_z}{\partial z} \tag{4.22}$$

The change in radial momentum of an electron subjected to this field for a time equal to that required to traverse the transition region, whose width is a, is

$$\Delta(mv_r) = -\int eE_r\, dt = \int_0^a \frac{er}{2}\frac{\partial E_z}{\partial z}\frac{dz}{v_z} \tag{4.23}$$

* See Bedford, reference 15.

since $dt = dz/v_z$. Since α is small, v_z can be assumed to be constant over the transition region, and the equation may be integrated, giving

$$\Delta(mv_r) = \frac{er}{2v_z}(E_2 - E_1) \tag{4.24}$$

The change in angle is, therefore,

$$\alpha = \frac{\Delta(mv_r)}{mv_z} = \frac{er}{2v_z^2 m}(E_2 - E_1) \tag{4.25}$$

or, replacing mv_z^2 by $2e\phi$, where ϕ is the potential of the lens,

$$\alpha = r\frac{E_2 - E_1}{4\phi} \tag{4.26}$$

Since α has been assumed small enough that v_z may be regarded as a constant throughout the transition region, it is also permissible to regard r as a constant in this region, and equal to the height of incidence and emergence of the ray considered. α/r then becomes simply the sum of the reciprocal object and image distances $(1/s + 1/s')$ and hence, by the thin-lens equation 4.7,

$$f = \frac{4\phi}{E_1 - E_2} \tag{4.27}$$

Although, in view of the restrictions assumed in the derivation, Eq. 4.27 is only a first approximation, it is, nevertheless, useful for esti-

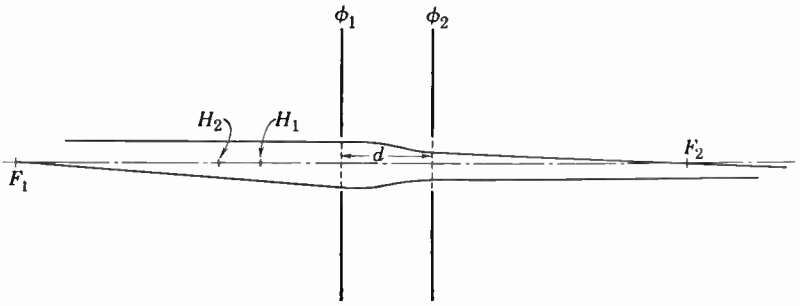


Fig. 4.13. Ray Paths in Double-Aperture Lens (Schematic).

mating the behavior of apertures and other electron-optical systems. For example, a lens often encountered in practice consists of two apertures at different potentials as illustrated in Fig. 4.13. The focal lengths of the apertures considered separately are

$$f_A = \frac{4\phi_A d}{\phi_B - \phi_A}$$

$$f_B = \frac{-4\phi_B d}{\phi_B - \phi_A}$$

This being combined into a "thick lens," due account being taken of the converging action of the field between the apertures and the difference in potential on the two sides of the lens, the focal lengths f and f' on the object and image sides of the system are

$$f = \frac{8d}{3\left(\frac{\phi_B}{\phi_A} - 1\right)\left(1 - \sqrt{\frac{\phi_A}{\phi_B}}\right)}$$

$$f' = \frac{8d}{3\sqrt{\frac{\phi_A}{\phi_B}}\left(\frac{\phi_B}{\phi_A} - 1\right)\left(1 - \sqrt{\frac{\phi_A}{\phi_B}}\right)} \tag{4.28}$$

Again, warning should be given that these results are only approximate and are strictly applicable only when the aperture diameters are very small both compared to the separation d and to the focal lengths of the system.

4.7 The Ray Equation. In most practical cases, the lens region, i.e., the region throughout which the field is varying, cannot be regarded as small, as was assumed in the derivation given in the preceding section. To determine the imaging properties of a more general electron lens, it is necessary to make use of a differential equation of the ray path, which can be integrated over the region of varying potential. If the same elementary cylindrical element, of thickness dz, about the axis of symmetry is considered as in the preceding section, the change in the radial momentum of an electron passing through the element becomes

$$d(mv_r) = \frac{er}{2}\frac{\partial E_z}{\partial z}\frac{dz}{v_z}$$

If the substitutions $v_r = dr/dt$ and $\partial E_z/\partial z = -d^2\phi/dz^2$, ϕ representing the electric potential along the axis, are made, it leads to

$$\frac{d}{dz}\left(\frac{dr}{dt}\right) = -\frac{er}{2mv_z}\frac{d^2\phi}{dz^2} \tag{4.29}$$

By a suitable rearrangement of terms and the substitution $v_z \cong v \cong \sqrt{(2e/m)\phi}$, Eq. 4.29 can be written

$$\frac{d^2r}{dz^2} = -\frac{1}{2\phi}\frac{d\phi}{dz}\frac{dr}{dz} - \frac{1}{4\phi}\frac{d^2\phi}{dz^2}r$$

or, representing differentiations with respect to the axial coordinate by primes,

$$r'' = -\frac{r'\phi'}{2\phi} - \frac{r\phi''}{4\phi} \tag{4.30}$$

This is the fundamental ray equation used in all axially symmetric electron-optical problems involving only electrostatic fields. When solved, it gives the radial distance from the axis for a paraxial electron ray in a meridional plane at every point along the axis. Furthermore, its linear form assures the fulfillment of Gaussian, or first-order, image requirements.

It is relatively simple to derive a more general path equation, which applies for all rays, whether they are paraxial and meridional or not. For this purpose, it is convenient to proceed from Newton's equations of motion in an axially symmetric electric field:

$$\ddot{r} - r\dot{\theta}^2 = \frac{e}{m}\frac{\partial\varphi}{\partial r}; \qquad \ddot{z} = \frac{e}{m}\frac{\partial\varphi}{\partial z}; \qquad \frac{d}{dt}(r^2\dot{\theta}) = 0 \tag{4.31}$$

Here, dots indicate differentiation with respect to the time t and φ denotes the electrostatic potential at any point in space. Integration of the last equation leads to

$$\dot{\theta} = \frac{r_0^2\dot{\theta}_0}{r^2};$$

$$\theta' = \frac{\dot{\theta}}{\dot{z}} = \frac{r_0^2\dot{\theta}_0}{r^2\dot{z}} = \frac{C\sqrt{1 + r'^2 + r^2\theta'^2}}{r^2\sqrt{\varphi}} = \frac{C}{r^2}\frac{\sqrt{1 + r'^2}}{\sqrt{\varphi - \dfrac{C^2}{r^2}}} \tag{4.32}$$

with

$$C = \frac{r_0^2\theta'_0\sqrt{\varphi_0}}{\sqrt{1 + r'^2_0 + r_0^2\theta'^2_0}} \tag{4.33}$$

The differential equation for r hence takes the form

$$\ddot{r} = \frac{e}{m}\left(\frac{2C^2}{r^3} + \frac{\partial\varphi}{\partial r}\right) \tag{4.34}$$

If the z-coordinate is chosen as an independent variable in place of the time, the equation becomes, since $\ddot{r} = r''\dot{z}^2 + r'\ddot{z}$

$$r'' = \frac{\ddot{r} - r'\ddot{z}}{\dot{z}^2} = \frac{1}{2\varphi}\left(\frac{2C^2}{r^3} + \frac{\partial\varphi}{\partial r} - r'\frac{\partial\varphi}{\partial z}\right)(1 + r'^2 + r^2\theta'^2) \quad (4.35)$$

If the value for θ' given by Eq. 4.32 is substituted and the modified potential

$$\varphi^* = \varphi - \frac{C^2}{r^2} \quad (4.36)$$

is introduced, the equation for r takes the form

$$r'' = \frac{(1 + r'^2)}{2\varphi^*}\left(\frac{\partial\varphi^*}{\partial r} - r'\frac{\partial\varphi^*}{\partial z}\right) \quad (4.37)$$

while θ is given by the integral

$$\theta = \theta_0 + \int_{z_0}^{z} \frac{C}{r^2}\frac{\sqrt{1 + r'^2}}{\sqrt{\varphi - \dfrac{C^2}{r^2}}}\,dz \quad (4.38)$$

For $C = 0$ the ray remains throughout in the same meridional plane ($\theta = \theta_0$) and the modified potential φ^* becomes identical with the true potential φ. C represents, but for a constant factor $\sqrt{2em}$, the angular momentum of the electron about the axis of symmetry.

If the general relation

$$\varphi = \phi - \tfrac{1}{4}r^2\phi'' + \tfrac{1}{64}r^4\phi^{\mathrm{iv}} \cdots \quad (4.39)$$

which applies for all axially symmetric potential fields, is substituted in Eq. 4.37 and only terms of the first order in r and r' are retained, the paraxial equation

$$r'' = \frac{C^2}{\phi r^3} - \frac{r'\phi'}{2\phi} - \frac{r\phi''}{4\phi} \quad (4.40)$$

is obtained which, for meridional rays, is identical with Eq. 4.30. It will be noted that the constant C is of the second order in the radial coordinate. If the dependent variable r is replaced by the complex variable $w = r \cdot e^{i\theta}$, the term with C vanishes and the equation again takes on the form of Eq. 4.30. As a result, the Gaussian imaging properties of the lens field are readily demonstrated even when skew rays are taken into consideration.

If the terms of the next-higher order—the third order in r and r'—are retained, the differential equation becomes inhomogeneous. A solution for small values of r_0 and r'_0 indicates that the rays now miss the Gaussian image points by distances which are given by a sum of terms of the third order in the initial separation from the axis r_0 and the initial slope r'_0. These "aberrations" will be considered in section 4.14.

4.8 Solution of the Ray Equation. If a pair of solutions of Eq. 4.30 is known, the first-order imaging properties of the electron lens system can be readily determined. The procedure is similar to the procedure in optics. An electron ray parallel to the axis at the object is traced through the lens. If the image falls well outside the lens, the rectilinear ray in image space is traced back to its intersection with the axis to locate the image-side focal point and to its intersection with the incident parallel ray to locate the image-side principal plane. The object-side focal point and principal plane are similarly located by tracing through the lens a ray which is parallel to the axis in image space and drawing the tangent to the ray in the object plane to its intersection with the axis and with the parallel ray. With the four cardinal points, the position and magnification of the image corresponding to an arbitrarily placed object can readily be calculated. However, it must be remembered that, in order to determine the cardinal points in this way, the image must be known to lie outside the lens.

An alternative method of determining the imaging properties of a lens, which does not require this knowledge, consists in first locating the image plane by tracing through the system a ray leaving the axial object point with a small radial velocity component and then determining the magnification by tracing a second ray leaving an off-axis point of the object parallel to the axis. The image-side focal point and principal plane are then determined as before; the object-side focal point and principal plane, with the aid of Eqs. 4.11 and 4.14. It should be noted that the cardinal points so obtained apply only for a single object position if either object or image falls within the lens.

The differential equation rarely permits an analytical solution in practical cases. As for the Laplace equation, a resort to numerical approximation or graphical methods is generally required.

A very simple and fairly rapid approximate method has been proposed by Richard Gans.* With a little care, this method is capable of an accuracy adequate for most practical cases.

* See Gans, reference 16.

The method represents the potential along the axis by a series of straight-line segments, and applies the ray equation along the segments in turn.

Over any straight segment, the second derivative ϕ'' is zero. The ray equation, therefore, becomes

$$r'' = -\frac{r'\phi'}{2\phi}$$

This can be integrated twice to give

$$r'\sqrt{\phi} = C \tag{4.41}$$

and

$$r = r_0 + \frac{2C(\sqrt{\phi} - \sqrt{\phi_0})}{\phi'} \tag{4.41a}$$

where r_0 and ϕ_0 are the radial position and potential at the beginning of the segment.

At the point where two segments meet, ϕ'' is infinite. By integrating Eq. 4.30 over the transition, the equation

$$r'_2 - r'_1 = -r\frac{\phi'_2 - \phi'_1}{4\phi} \tag{4.42}$$

is obtained. The subscript 1 indicates values before the break-point, and 2 those after the break-point.

Finally, where the segment is parallel to the axis, the solution becomes

$$r = r_0 + r'_0(z - z_0) \tag{4.43}$$

The method, because of its utility as a practical means for estimating the performance of any electron lens system for which the axial distribution is known, is worth illustrating by means of a simple numerical example.

The lens system to be considered consists of two equidiameter coaxial cylinders. The cylinders are assumed to have potentials equal to 2 and 12, respectively. Here, the units of potential are quite arbitrary, the only thing of importance, as far as the lens properties are concerned, being the voltage ratio, which in this example is 6. The axial potential distribution of the system has been calculated by means of Eq. 4.50*a* given in the next section, and is tabulated in Table 4.1, in which the distance z is given in cylinder radii.

TABLE 4.1. POTENTIAL DISTRIBUTION IN TWO-CYLINDER LENS

z	ϕ	z	ϕ	z	ϕ
−2.0	2.05	−0.6	3.72	0.7	10.62
−1.8	2.10	−0.4	4.58	0.9	11.12
−1.6	2.16	−0.2	5.71	1.1	11.45
−1.4	2.27	0.0	7.00	1.3	11.66
−1.2	2.44	0.1	7.66	1.5	11.79
−1.0	2.70	0.3	8.88	1.7	11.87
−0.8	3.12	0.5	9.88	1.9	11.93

It should be pointed out that the potential distribution for any desired voltage ratio can be obtained with the aid of the values given in Table 4.1 by adding an appropriate constant.

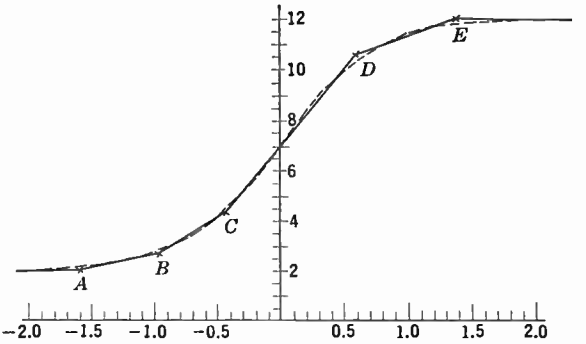


Fig. 4.14. Axial Potential Distribution and Its Line Segment Approximation for a Cylinder Lens.

The potential distribution is shown by the dotted curve of Fig. 4.14. In addition, this distribution is approximated in the figure by a series of six straight-line segments intersecting at points A, B, C, D, E. The potentials at the intersections and the potential gradients in the connecting segments are given below:

	z	ϕ	ϕ'
			1.09
A	−1.60	2.0	
			3.20
B	−0.96	2.7	
			5.95
C	−0.46	4.3	
			1.75
D	0.60	10.6	
			0
E	1.40	12.0	
∞			

An electron leaves an object point on the axis at a distance -12 lens radii from the origin, with a velocity corresponding to $\phi = 2$ and a slope of 0.096. It is required to find the conjugate image point, that is, the point at which the electron again intersects the axis.

The electron will move in a straight line until it reaches the first break-point at A. At this point, its radial distance is $r_A = 1.00$ and its slope $r'_{1A} = 0.096$. Its slope at the other side of the break-point is given by Eq. 4.42:

$$r'_{2A} = 0.096 - \frac{1.00 \times (1.09 - 0)}{4 \times 2.0} = -0.040$$

The value of C for the segment AB, given by Eq. 4.41 is:

$$C = -0.040 \times \sqrt{2.0} = -0.057$$

Therefore, by Eq. 4.41,

$$r'_{1B} = \frac{-0.057}{\sqrt{2.7}} = -0.034$$

and, from Eq. 4.41*a*,

$$r_B = 1.00 + \frac{-0.114(\sqrt{2.7} - \sqrt{2.0})}{1.09} = 0.97$$

The calculation continues in the same way for each succeeding break-point and segment. The results are given in Table 4.2. From point E,

TABLE 4.2. PARAMETERS OF APPROXIMATE PATH IN TWO-CYLINDER LENS

	r	r'_1	r'_2	C
A	1.00	0.096	-0.040	
				-0.057
B	0.97	-0.034	-0.22	
				-0.366
C	0.88	-0.18	-0.32	
				-0.657
D	0.62	-0.20	-0.14	
				-0.459
E	0.51	-0.13	-0.11	

the electron continues in a straight line, with a slope to the axis of -0.11. Therefore, it intersects the axis at a distance $0.51/0.11 = 4.6$ radii from point E, or, in other words, the image point is 18.0 lens radii from the object point. A determination of the path by the more exact method outlined below gives the distance between the object and image

point as 18.6 lens radii. Figure 4.15 shows the exact path of the electron through the system, together with the points corresponding to the radial position at the break-points as calculated by the approximate method.

A closer approximation, of course, could have been obtained if a greater number of line segments had been used to represent the potential, particularly on the low-potential side of the lens. It then becomes more convenient, however, to employ a standard method for the numerical integration of differential equations, such as that outlined below.*

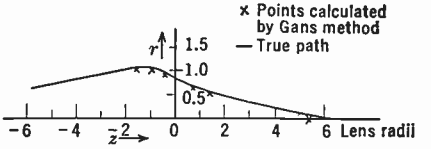


Fig. 4.15. Electron Trajectory through Cylinder Lens.

Suppose that a function y is tabulated for equal intervals h of the independent variable z in the following manner, defining the first-, second-, and third-order differences of y by

$$\Delta_1 y_i = y_i - y_{i-1}$$

$$\Delta_2 y_i = \Delta_1 y_i - \Delta_1 y_{i-1}$$

$$\Delta_3 y_i = \Delta_2 y_i - \Delta_2 y_{i-1}$$

z	y			
$z_0 - 3h$	y_{-3}			
		$\Delta_1 y_{-2}$		
$z_0 - 2h$	y_{-2}		$\Delta_2 y_{-1}$	
		$\Delta_1 y_{-1}$		$\Delta_3 y_0$
$z_0 - h$	y_{-1}		$\Delta_2 y_0$	
		$\Delta_1 y_0$		$\Delta_3 y_1$
z_0	y_0		$\Delta_2 y_1$	
		$\Delta_1 y_1$		$\Delta_3 y_2$
$z_0 + h$	y_1		$\Delta_2 y_2$	
		$\Delta_1 y_2$		$\Delta_3 y_3$
$z_0 + 2h$	y_2		$\Delta_2 y_3$	
		$\Delta_1 y_3$		
$z_0 + 3h$	y_3			

Then the integral of y over any interval h may be expressed in terms of function y and its differences as follows:

* See, e.g., Zworykin, Morton, Ramberg, Hillier, and Vance, reference 4, pp. 407–412, or E. P. Adams, reference 28, pp. 220–242.

$$\int_{z_0-h}^{z_0} y\,dz \cong h(y_0 - \tfrac{1}{2}\Delta_1 y_0 - \tfrac{1}{12}\Delta_2 y_0 - \tfrac{1}{24}\Delta_3 y_0) \quad (4.44a)$$

$$\cong h(y_{-1} + \tfrac{1}{2}\Delta_1 y_0 - \tfrac{1}{12}\Delta_2 y_1 + \tfrac{1}{24}\Delta_3 y_2) \quad (4.44b)$$

This approximation corresponds to fitting a third-order curve through four neighboring points of the function. The intervals should, in general, be chosen small enough so that the highest difference makes only a very small contribution to the integral.

The application of Eqs. 4.44 to the integration of a first-order differential equation,

$$r' = y = f(r, z) \quad (4.45)$$

is obvious. Assume, to begin with, that the equation has been integrated up to some point $z_0 - h$ and that $y = f(r, z)$ and its differences as well as r are tabulated as functions of z up to this point. Then extrapolate r to provide a guessed value $r(z_0)$ and substitute this value in Eq. 4.45 to provide a corresponding value for $y(z_0)$. With this new value of y and the corresponding differences, the increment in r is determined by Eq. 4.44*a*. If this is identical with the originally guessed increment, the guess was correct, and it is possible to proceed to the next interval. If the two increments do not agree, a new guess is made and the process is repeated until identity between the guessed increment and the integrated increment is obtained.

At the beginning of an integration, only the initial value of r, and hence, by Eq. 4.45, also that of r', are known, so that, in the integration over the first interval, only the first two terms of Eq. 4.44*a* can be used. As the integration progresses, additional differences are established, however, and the first integrations may be checked and eventually modified by applying Eq. 4.44*b*. Quite generally, the repeat operations can be minimized and the integration be carried out smoothly and speedily if sufficiently small intervals are chosen. Frequently, the differential equation can be transformed by a change of variables in such a manner that the speed and accuracy of integration is improved.

An equation of the second order, such as Eq. 4.30, which may be written in the form

$$r'' = y = f(r', r, z) \quad (4.46)$$

can be solved in quite the same manner, applying the numerical integration twice in succession. Both y and r' are tabulated with their differences, along with r, as function of z. The increments in r and r' are obtained with the aid of Eqs. 4.44 since

$$r'(z_0) - r'(z_0 - h) = \int_{z_0-h}^{z_0} y\,dz; \quad r(z_0) - r(z_0 - h) = \int_{z_0-h}^{z_0} r'\,dz \qquad (4.47)$$

The electron trajectory shown in Fig. 4.15 was calculated in this manner.

4.9 Special Lens Systems. The electron-optical system consisting of two coaxial cylinders of equal diameter forms the basis of many practical lenses. The properties of this configuration can, of course, be determined by mapping the axial potential distribution with the aid of a plotting tank and then tracing suitable rays by one of the methods just discussed. However, when high accuracy is required, a mathematical solution of Laplace's equation is to be preferred.

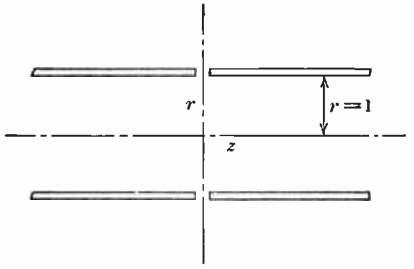


Fig. 4.16. Coordinates for Electrostatic Cylinder Lens.

Figure 4.16 illustrates, in cross section, two semi-infinite coaxial cylinders, spaced a negligible distance apart. The axis of the cylinders will be taken as the z-axis of the cylindrical coordinate system, and the origin of the system will be located at the junction of the two cylinders.

The general solution of the Laplace equation for this type of configuration was discussed in the preceding chapter. By separating the variables, it was shown that the solution could be expressed in the form of the following integral:

$$\varphi(r, z) = \int A(k)G(r, k)F(k, z)\,dk$$

where

$$F = a\epsilon^{ikz} + b\epsilon^{-ikz}$$

$$G = cJ_0(ikr) + dN_0(ikr)$$

The requirement that the potential remain finite as z increases eliminates all terms with complex k. Furthermore, the condition that it be finite along the axis requires that the coefficients of the Neumann function N_0 be zero. Finally, since $\phi - (\phi_1 + \phi_2)/2$, where ϕ_1 and ϕ_2 are the potentials of the two cylinders, is an odd function of z, the only trigometric functions to be considered are sines.

Hence, the solution can be written in the form

$$\varphi(r, z) = \int_0^\infty B(k) J_0(ikr) \sin kz \, dk + \frac{\phi_1 + \phi_2}{2} \tag{4.48}$$

The coefficient $B(k)$ can be found with the aid of boundary conditions for $r = 1$, the radius of the cylinders being taken as unit length; they are

$$\varphi(1, z) = \phi_1 \quad \text{for } z < 0$$

$$\varphi(1, z) = \phi_2 \quad \text{for } z > 0$$

The evaluation leads to

$$B(k) J_0(ik) = \frac{1}{\pi} \left\{ \int_{-\infty}^0 - \frac{\phi_2 - \phi_1}{2} \sin kz \, dz + \int_0^\infty \frac{\phi_2 - \phi_1}{2} \sin kz \, dz \right\}$$

$$B(k) = \frac{\phi_2 - \phi_1}{\pi k J_0(ik)} (1 - \lim_{z \to \infty} \cos kz)$$

If this value is substituted in Eq. 4.48, the potential is found to be

$$\varphi(r, z) = \frac{1}{\pi} \int_0^\infty \frac{\phi_2 - \phi_1}{k J_0(ik)} J_0(ikr) \sin (kz) \, dk + \frac{\phi_1 + \phi_2}{2} \tag{4.49}$$

The axial potential and its first two derivatives are

$$\phi(z) = \frac{1}{\pi} \int_0^\infty (\phi_2 - \phi_1) \frac{\sin kz}{k J_0(ik)} dk + \frac{\phi_1 + \phi_2}{2} \tag{4.50a}$$

$$\phi'(z) = \frac{1}{\pi} \int_0^\infty (\phi_2 - \phi_1) \frac{\cos kz}{J_0(ik)} dk \tag{4.50b}$$

$$\phi''(z) = -\frac{1}{\pi} \int_0^\infty (\phi_2 - \phi_1) \frac{k \sin kz}{J_0(ik)} dk \tag{4.50c}$$

These integrals cannot be solved analytically, but must be evaluated by quadrature, eventually employing Eqs. 4.44. The figures so obtained can then be utilized to integrate the ray equation by one of the techniques described in the preceding section. For many purposes, it is convenient to approximate the expression for the potential by the function: *

$$\phi(z) = \frac{\phi_2 - \phi_1}{\pi} \cdot \tanh (1.315z) + \frac{\phi_1 + \phi_2}{2} \tag{4.51}$$

* See Gray, reference 17, and Bertram, reference 18.

In this example, as in many others, the solution of the ray equation can be expedited by the substitution

$$c = -r'/r \tag{4.52}$$

The function c is the convergence of the ray. It is thus named because it is the reciprocal of the distance along the axis from the point at which c is determined to the intersection of a rectilinear extension of the ray with the axis. With the convergence as a dependent variable, the ray equation (4.30) becomes

$$c' = c^2 - \frac{c\phi'}{2\phi} + \frac{\phi''}{4\phi} \tag{4.53}$$

The evaluation of the equation must be continued until the ray is substantially a straight line, or until it passes through the image plane.

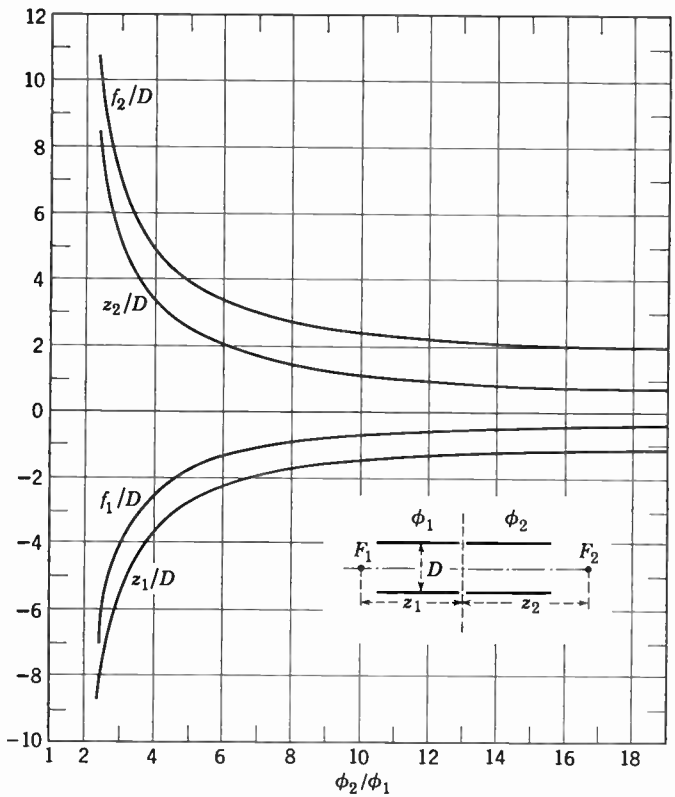


Fig. 4.17. Focal Properties of the Equidiameter Cylinder Lens. (Zworykin, Morton, Ramberg, Hillier, and Vance, reference 4.)

The curves reproduced in Fig. 4.17 locate the four cardinal points of the coaxial equidiameter cylinder lens for various ratios of the applied potentials ϕ_1 and ϕ_2. All distances are expressed in terms of cylinder diameters.

Besides the coaxial cylinder lens using cylinders of the same diameters, an infinite number of other lenses can, of course, be formed by cylinders of unequal diameters. The position of the principal planes is affected, to some extent, by the choice of the diameter ratio. There

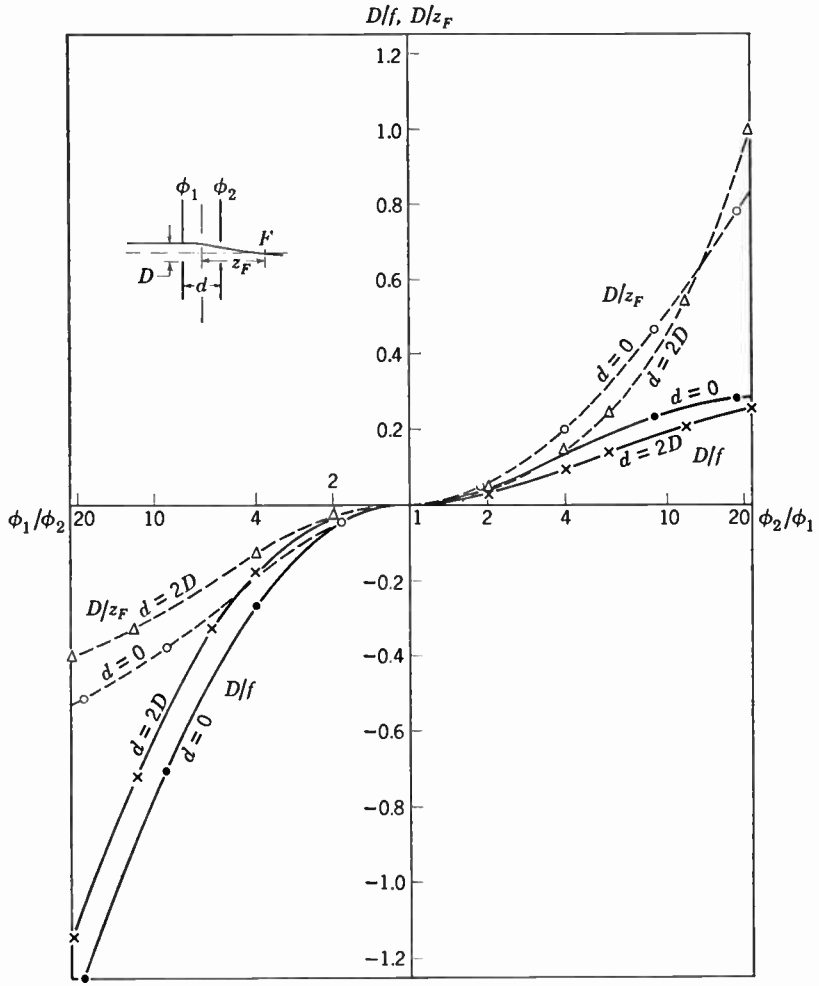


Fig. 4.18a. Focal Properties of Equidiameter Aperture Lens for $d/D = 0$, $d/D = 2$.

is rarely any advantage in choosing cylinders of different diameters from the standpoint of electron optics. Coaxial cylinder lenses will be discussed in more detail in relation to the electron gun.

4.10 Aperture Lenses. Another important class of lenses consists of those formed by a pair of apertured conductors at different potentials. They have found application both in electron guns and in certain types of electron imaging devices.

The simplest case of this type has already been discussed. If the

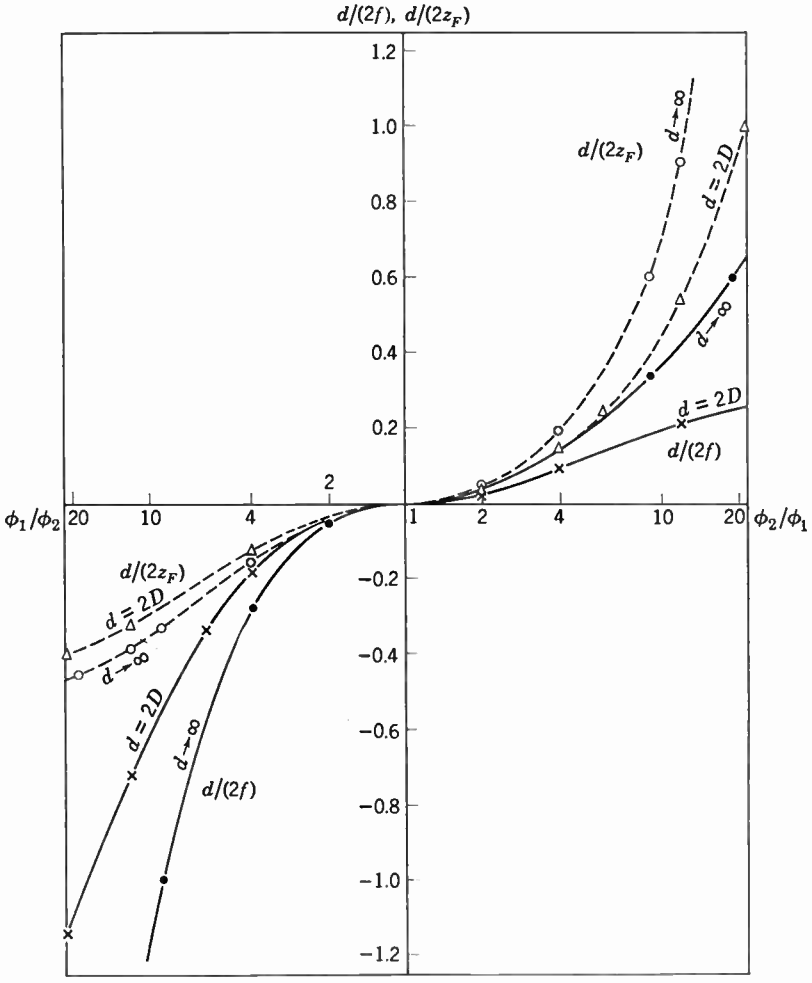


Fig. 4.18*b*. Focal Properties of Equidiameter Aperture Lens for $d/D = 2$, $d/D = \infty$.

separation d of the apertures is very large compared to the aperture diameter, the focal lengths of the lens are given by Eq. 4.28. Setting $\phi_A = \phi_1$, and $\phi_B = \phi_2$, both the principal planes lie beyond the low-voltage aperture, at distances

$$h = -\frac{d}{2} - \frac{4d\phi_1}{3(\phi_2 - \phi_1)}, \qquad h' = -\frac{5d}{6} - \frac{4d\phi_1}{3(\phi_2 - \phi_1)} \tag{4.54}$$

from the plane of symmetry of the lens.

In the other extreme case, that is, a separation which is negligible compared with the aperture diameter, the potential distribution along the axis takes on the simple form

$$\phi = \frac{\phi_2 + \phi_1}{2} + \frac{\phi_2 - \phi_1}{2} \frac{z}{\sqrt{(D/2)^2 + z^2}} \tag{4.55}$$

Here, D is the aperture diameter. With the aid of this expression the cardinal points can be determined, for any value of ϕ_2/ϕ_1, by integrating the ray equation.

For other values of d/D, the potential distribution must first be found with the aid of the electrolytic tank or by a method of successive approximations. Figures 4.18*a* and *b* show the position of the cardinal points of a lens formed by two equidiameter apertures for the three conditions $d/D = 0$, $d/D = 2$, and $d/D = \infty$. The divergence between the values for the last two cases is seen to be most pronounced for large values of ϕ_2/ϕ_1, or high refractive powers.

4.11 Cathode Lens Systems. In many practical electron-optical systems, the electrons enter the system with essentially zero velocity. An example is to be found in the electrostatically focused image tube (which, in connection with the image iconoscope, is discussed in Chapter 10). The electron-optical properties of such a lens system are radically different from those of the lenses discussed so far.

The potential at the point where the electron enters the system is zero. In consequence, the electron ray originates in a region of zero index of refraction. Because of this, the first focal length is zero, and the first principal plane is located at the object.

If the electrons were actually emitted from the object, that is, the cathode, with zero velocity, the image plane would be completely indeterminate. However, the assumption of a small radial initial velocity suffices to determine the position and, with it, the magnification of the image. In practice, both the radial and axial components of the initial

velocity are far from negligible, and are responsible for certain aberrations discussed in section 4.14.

One of the simplest lenses of this type consists of two coaxial cylinders of equal diameter, one infinite in extent, the other terminated by a flat plane. The terminal plane is assumed to be the electron object. In a conventional image tube, for example, this plane would be a photocathode. The cathode surface is conducting and is electrically connected to the cathode cylinder, these elements being at zero potential. For convenience, the lens will be said to be located at the junction of the two cylinders, although actually it extends from the cathode to a distance of several cylinder radii beyond the junction. The geometry of the lens can be seen from Fig. 4.19.

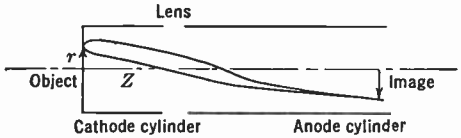


Fig. 4.19. Cathode Lens.

The calculation of the potential distribution proceeds exactly as in the previous example of the two-cylinder lens. It is expedient here to use cylindrical coordinates with their origin at the cathode. The general solution of the Laplace equation, which can be evaluated from the boundary conditions of the lens, leads to the following integrals, giving the potential along the axis together with the first and second derivatives:

$$\phi(z) = \frac{2\phi_1}{\pi}\int_0^\infty \frac{\cos ku \sin kz}{kJ_0(ik)}\,dk \tag{4.56a}$$

$$\phi'(z) = \frac{2\phi_1}{\pi}\int_0^\infty \frac{\cos ku \cos kz}{J_0(ik)}\,dk \tag{4.56b}$$

$$\phi''(z) = -\frac{2\phi_1}{\pi}\int_0^\infty \frac{k \cos ku \sin kz}{J_0(ik)}\,dk \tag{4.56c}$$

As before, these integrals must be evaluated by quadrature.

The numerical values of the potential thus obtained can, with the aid of the ray equation, be used to obtain the lens properties. In this instance, the convergence function c becomes infinite at $z = 0$ and thus cannot be used. However, the substitution of the function

$$b = \frac{1}{2z} - \frac{r'}{r}$$

leading to

$$b' = b^2 - b\left(\frac{1}{z} + \frac{\phi'}{2\phi}\right) + \frac{\phi''}{4\phi} + \frac{1}{2z}\left(\frac{\phi'}{2\phi} - \frac{1}{2z}\right) \tag{4.57}$$

will facilitate the solution by reducing the ray equation to a first-order differential equation.

Numerical determinations of two electron paths will locate the image and give its magnification. In Fig. 4.20, the potential and its first two derivatives are shown, together with two electron trajectories. It should be noted that the latter are independent of the applied poten-

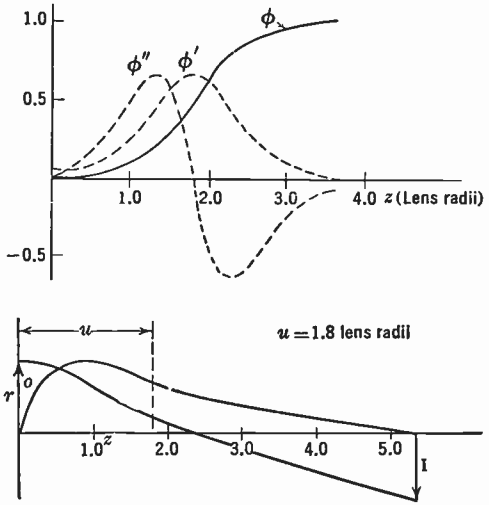


Fig. 4.20. Axial Potential Distribution and Electron Trajectories in a Cathode Lens.

tial, ϕ_1, and are a function of the cathode-to-lens distance, u, only. The magnification is approximately half the ratio of image distance to object distance, both measured from the junction of the two cylinders. The magnification is unity for $u = 1.8$ lens radii.

It is often desirable to arrange the lens in such a way that the focal length and image position can be varied electrically. This can be done by making the cathode cylinder of resistive material so that an axial potential gradient can be established on it between cathode and lens. In practice, it is convenient to divide the cathode cylinder into a number of rings having equal potential steps between them. Experimen-

tally, this is found to simulate the effect of the resistive cathode cylinder qualitatively and quantitatively.

The performance of this system can be calculated by the methods already given. The results of such a determination are of value in many design problems.

The magnification of the system just described is a function of the image and object distance alone and cannot be changed in a given tube configuration. It is possible, by incorporating a third element, to make a lens system which has variable magnification. The additional electrode may be in the form of an aperture or a short cylinder placed between the anode and cathode cylinders. An electron-optical system of this type is illustrated in Fig. 4.21. For a given tube structure and overall voltage, the magnification is varied by changing the potential ϕ_3 on the extra electrode and refocusing the image by the potential ϕ_2 across the cathode cylinder.

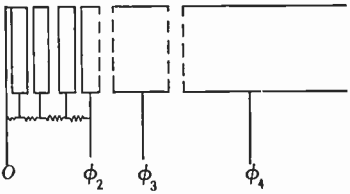


Fig. 4.21. An Electron Lens System Arranged to Give Variable Magnification.

Before the subject of cathode lens systems is left, some mention should be made of the procedure required when they are to be investigated by the Gans method described in section 4.8. In order to determine the image position, it is necessary to assume an initial radial velocity. In

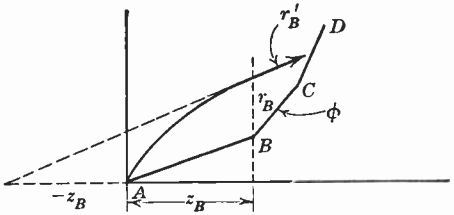


Fig. 4.22. Determination of Electron Trajectories near Cathode in Gans Method of Ray Tracing.

Fig. 4.22 the axial potential distribution of a cathode lens system is represented by a series of straight-line segments. The electron is emitted from the object point, A, with a small radial initial velocity, into a region of uniform field as represented by the first segment, AB. The trajectory of the electron is a parabola. If, at the end of the first segment, its radial distance from the axis is r_B, then, from the properties of the parabola, the slope of its trajectory will be $r_B/(2z_B)$, where z_B

is its axial position. Therefore, the electron will enter the first breakpoint on a path described as follows:

$$r'_1 = \frac{r_B}{2z_B}$$

$$r = r_B, \quad \phi = \phi_B \tag{4.58}$$

From this point on, the procedure is the same as for any lens system. The choice of the radial separation r_B is arbitrary, and does not affect the determination of the image position.

4.12 The Magnetic Lens. Thus far the discussion has been limited to electrostatic lens systems. Magnetic lenses, however, rank as at least equal in importance. From a theoretical standpoint, they are very much more complicated, because the principal electron paths are not confined to a plane.

The simplest magnetic lens consists of a uniform magnetic field parallel to the axis of the system and pervading all the region between the anode and the cathode. Such a field is produced in the interior of a long solenoid.

An electron which leaves the emitter parallel to the axis is subject to no force from the magnetic field. The time required by this electron to traverse the distance between anode and cathode is given by the integral

$$t_1 = \int_0^L \frac{dz}{v_z} \tag{4.59}$$

where L is the total distance and v_z the axial velocity. If, in addition, the particle has a radial velocity component, v_r, it will experience a force

$$F = Hev_r$$

where H is the magnetic field strength. Since the direction of this force is at right angles to the transverse velocity and to the field, the electron path will be curved. The projection of this path on a plane normal to the axis (Fig. 4.23) is a circle, the radius ρ being given

Fig. 4.23. Electron Paths in a Uniform Magnetic Field.

by the relation between the inward acceleration and the centripetal force, i.e.:

$$\frac{mv_r^2}{\rho} = Hev_r; \qquad \frac{v_r}{\rho} = \frac{e}{m}H$$

The circumference of this circle will be $2\pi\rho$, and, therefore, the time t_2 required to traverse the circle will be

$$t_2 = \frac{2\pi\rho}{v_r} = \frac{2\pi m}{eH} \tag{4.60}$$

This time is independent of the initial radial velocity. If t_2 in Eq. 4.60, or any multiple thereof, is equal to t_1 in Eq. 4.59, all electrons leaving a point on the source will come together at a point in the image plane, regardless of their initial radial velocity component. It should be noticed that the image, unlike that formed by an electrostatic lens, is erect.

The investigation of electron motion in a more general axially symmetric magnetic field calls for the introduction of a new quantity, the vector potential A of the magnetic field. It is related to the axial and radial magnetic field components, h_z and h_r, by *

$$h_z = \frac{1}{r}\frac{\partial(rA)}{\partial r}; \qquad h_r = -\frac{\partial A}{\partial z} \tag{4.61}$$

and is equal to the ratio of the magnetic flux N through a circle about the axis through the reference point divided by the circumference of the circle:

$$A = \frac{N}{2\pi r} = \frac{1}{r}\int_0^r h_z r\, dr \tag{4.62}$$

The equations of motion in a general axially symmetric electric and magnetic field then become

$$\ddot{r} - r\dot{\theta}^2 = -\frac{e}{m}\left(r\dot{\theta}h_z - \frac{\partial\varphi}{\partial r}\right) = -\frac{e}{m}\left(\dot{\theta}\frac{\partial(rA)}{\partial r} - \frac{\partial\varphi}{\partial r}\right) \tag{4.63a}$$

* Lower-case letters are here employed to distinguish the magnetic field off the axis from the field $H(z) = h_z(z, 0)$ on the axis. The general relation between magnetic field and vector potential is, in vector notation,

$$\mathbf{h} = \text{curl } \mathbf{A}$$

In an axially symmetric field, only the azimuthal component of $\mathbf{A}$, $A_\theta = A$ differs from zero.

$$\frac{1}{r}\frac{d}{dt}(r^2\dot{\theta}) = -\frac{e}{m}(\dot{z}h_r - \dot{r}h_z) = \frac{e}{m}\left(\dot{z}\frac{\partial A}{\partial z} + \frac{\dot{r}}{r}\frac{\partial(rA)}{\partial r}\right) \tag{4.63b}$$

$$\ddot{z} = \frac{e}{m}\left(r\dot{\theta}h_r + \frac{\partial\varphi}{\partial z}\right) = -\frac{e}{m}\left(r\dot{\theta}\frac{\partial A}{\partial z} - \frac{\partial\varphi}{\partial z}\right) \tag{4.63c}$$

Integration of the second equation leads to

$$r^2\dot{\theta} = \sqrt{\frac{2e}{m}}\left(C + \sqrt{\frac{e}{2m}}rA\right) \tag{4.64}$$

$$\theta' = \frac{\dot{\theta}}{\dot{z}} = \frac{\left(C + \sqrt{\frac{e}{2m}}rA\right)}{r^2\sqrt{\varphi}}\sqrt{1 + r'^2 + r^2\theta'^2}$$

$$= \frac{\left(C + \sqrt{\frac{e}{2m}}rA\right)\sqrt{1 + r'^2}}{r^2\sqrt{\varphi - \left(\frac{C}{r} + \sqrt{\frac{e}{2m}}A\right)^2}} \tag{4.65}$$

with

$$C = -\sqrt{\frac{e}{2m}}r_0A_0 + \frac{r_0^2\theta'_0\sqrt{\varphi_0}}{\sqrt{1 + r'_0{}^2 + r_0^2\theta'_0{}^2}} \tag{4.66}$$

This leads to an equation for r of the same form as for the purely electric field (Eq. 4.37):

$$\ddot{r} = \frac{e}{m}\left\{\frac{2}{r}\left(\frac{C}{r} + \sqrt{\frac{e}{2m}}A\right)^2 + \frac{\partial\varphi}{\partial r} - \frac{1}{r}\sqrt{\frac{2e}{m}}\left(\frac{C}{r} + \sqrt{\frac{e}{2m}}A\right)\frac{\partial(rA)}{\partial r}\right\}$$

$$r'' = \frac{\ddot{r} - r'\ddot{z}}{\dot{z}^2} = \frac{(1 + r'^2 + r^2\theta'^2)}{2\varphi}\left\{\frac{2}{r}\left(\frac{C}{r} + \sqrt{\frac{e}{2m}}A\right)^2 + \frac{\partial\varphi}{\partial r}\right.$$

$$\left. - \frac{1}{r}\sqrt{\frac{2e}{m}}\left(\frac{C}{r} + \sqrt{\frac{e}{2m}}A\right)\left(\frac{\partial(rA)}{\partial r} - r'r\frac{\partial A}{\partial z}\right) - r'\frac{\partial\varphi}{\partial z}\right\}$$

$$= \frac{(1 + r'^2)}{2\varphi^*}\left\{\frac{\partial\varphi^*}{\partial r} - r'\frac{\partial\varphi^*}{\partial z}\right\} \tag{4.67}$$

with the modified potential φ^* defined by

$$\varphi^* = \varphi - \left[\frac{C}{r} + \sqrt{\frac{e}{2m}}A\right]^2 \tag{4.68}$$

In view of the fact that the magnetic field, like the electric field, obeys the Laplace equation, the vector potential may be written in the form

$$A = \frac{r}{2}H - \frac{r^3}{16}H'' + \cdots + \frac{(-1)^n H^{(2n)}}{n!(n+1)!}\left(\frac{r}{2}\right)^{2n+1} + \cdots \tag{4.69}$$

If this series, along with that for φ (Eq. 4.39), is substituted in Eq. 4.68 and only the terms of lowest order in r and r' are retained, the paraxial equation

$$r'' = \frac{C^2}{\phi r^3} - \frac{r'\phi'}{2\phi} - \frac{r\phi''}{4\phi} - \frac{eH^2 r}{8m\phi} \tag{4.70}$$

is obtained. In addition, Eq. 4.65 leads to the following expression for the angle of rotation of the electron:

$$\theta = \theta_0 + \int_{z_o}^{z}\left\{\frac{C}{r^2\sqrt{\phi}} + \sqrt{\frac{e}{8m\phi}}H\right\}dz \tag{4.71}$$

If attention is confined to electrons which, outside of the lens, move in meridional planes, so that their angular momentum, and hence C, vanishes, it becomes evident from Eq. 4.71 that the magnetic field rotates the image through an angle

$$\chi = \int_{z_o}^{z_i}\sqrt{\frac{e}{8m\phi}}H\,dz \tag{4.72}$$

where z_i is the axial coordinate of the image plane. Furthermore, for paraxial meridional electrons passing through a purely magnetic lens field,

$$r'' = -\frac{eH^2 r}{8m\phi} \tag{4.73}$$

For a thin magnetic lens, where the radial coordinate can be regarded as practically constant within the refracting field, Eq. 4.73 can be integrated to yield

$$r'_o - r'_i = \frac{er}{8m\phi}\int_{z_o}^{z_i}H^2\,dz \tag{4.73a}$$

If s and s' are the object and image distances, the left side of Eq. 4.73 becomes identical with $r/s + r/s'$. A comparison with the general lens equation 4.8, with $n = n'$, yields for the focal length f,

$$\frac{1}{f} = \frac{e}{8m\phi}\int_{z_o}^{z_i} H^2\, dz \tag{4.74}$$

In addition, the magnetic field of any electron lens rotates the image through an angle*

$$\chi = \int_{z_o}^{z_i} \sqrt{\frac{e}{8m\phi}}\, H\, dz \tag{4.75}$$

A comparison of the expression for the angle of rotation (Eq. 4.75) and that for the refractive power of a thin magnetic lens (Eq. 4.74) shows that, if a lens is made progressively thinner, keeping its refrac-

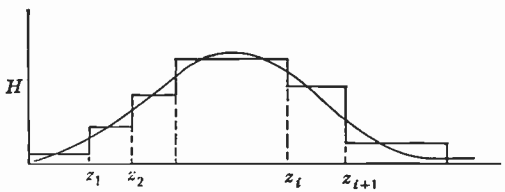


Fig. 4.24. Approximate Representation of Axial Magnetic Field.

tive power constant, the rotation produced by the lens approaches zero. A very thin magnetic lens, like an electrostatic lens, produces an inverted real image of a distant object. For the very long lens (uniform field), on the other hand, the angle of rotation becomes just equal to π and completely compensates the inversion of the image produced by the focusing process.

If the lens cannot be regarded as very thin, Eq. 4.72 must, in general, be integrated numerically. An alternative method, analogous to that proposed by Gans for an electrostatic lens, has been suggested by E. G. Ramberg.

To illustrate this method, it is assumed that the axial distribution of a field has been measured or estimated, leading to the smooth curve in Fig. 4.24. The actual distribution is approximated by a series of step segments as shown in the figure.

* This is simply the rotation of the path of "meridional" electrons ($C = 0$) about the axis, as given by Eq. 4.71.

The solution of Eq. 4.73 over any segment, for example, that between z_i and z_{i+1}, is

$$r = r_i \cos \sqrt{eH^2/(8m\phi)}\,(z - z_i) + r'_i \frac{\sin \sqrt{eH^2/(8m\phi)}\,(z - z_i)}{\sqrt{eH^2/(8m\phi)}} \tag{4.76}$$

where ϕ is the potential in the segment; H, the magnetic field; r, the radial distance of the ray; and $z_i < z < z_{i+1}$. Both r and r' are continuous at the break-points, so that the solutions for the individual segments given by Eq. 4.76 join smoothly.

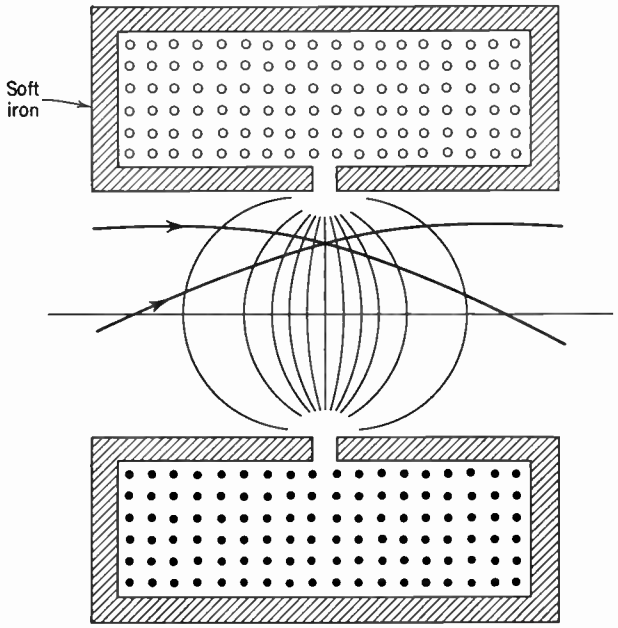


Fig. 4.25. Short Magnetic Lens. (Zworykin, Morton, Ramberg, Hillier, and Vance, reference 4.)

One field distribution, which permits analytic integration of the differential equation and, at the same time, is closely approximated by many magnetic lenses encountered in practice, is given by *

$$H = \frac{H_0}{1 + (z/a)^2} \tag{4.77}$$

* See Glaser, reference 29.

where H_0 is the maximum field, at the center of symmetry of the lens, and $2a$ is the width at half value of the field distribution. The focal length of this field is given by

$$\frac{1}{f} = \frac{1}{a} \sin \frac{\pi}{\sqrt{1+k^2}} \cong \frac{\pi}{a}\left(\frac{k^2}{2} - \frac{3k^4}{8} + \cdots\right) \tag{4.78}$$

the distance of the focal plane from the center of symmetry, by

$$\frac{1}{z_F} = \frac{1}{a} \tan \frac{\pi}{\sqrt{1+k^2}} \cong \frac{\pi}{a}\left(\frac{k^2}{2} - \frac{3k^4}{8} + \cdots\right) \tag{4.79}$$

Here, the parameter k^2 is given by

$$k^2 = \frac{ea^2 {H_0}^2}{8m\phi} \tag{4.80}$$

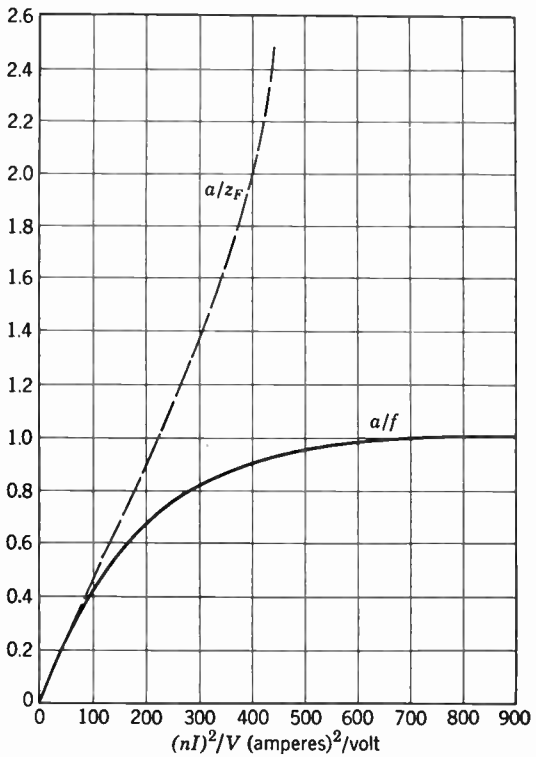


Fig. 4.26. Variation of Refractive Power (a/f) and Reciprocal Focal Distance (a/z_F) with Lens Excitation for the Magnetic Field $H_0/(1 + (z/a)^2)$.

In particular, if the lens field is produced by an armored coil with an air gap, such as that shown in Fig. 4.25, with a number of ampere turns nI, $aH_0 = 4nI/10$ and

$$k^2 = \frac{0.0035(nI)^2}{V} \tag{4.81}$$

where V is the operating voltage.

Figure 4.26 shows the variation of the refractive power and reciprocal focal distance of the lens field (Eq. 4.77) as function of the ratio $(nI)^2/V$. It indicates, as is incidentally obvious from Eqs. 4.78 and 4.79, that such a magnetic lens behaves, for moderate field strengths, like a thin magnetic lens with the lens position at the plane of symmetry. For stronger fields, the principal planes, which are crossed in the sense of the image-side principal plane lying closer to object space than the object-side principal plane, move apart.

4.13 Image Defects. The theory thus far developed has dealt with images formed by rays very close to the axis of the lens. In other words, it deals with rays making angles with the axis which are sufficiently small that only the first terms of the sine expansion need be used. However, because the first-order theory deals with such a small portion of the lens and object, it tells nothing about the quality of the image.

The theory has been extended to take into account the next term of the expansion. This third-order theory, developed by Ludwig von Seidel and bearing his name, leads to a series of five correction terms which vanish when the oblique rays are deflected in the same manner as the paraxial rays. When one or more of these terms differ from zero, the oblique rays from the object do not converge in the image points indicated by the laws governing the paraxial rays and the image is unsharp or deformed; the system is then said to have aberrations. Corresponding to the five terms in the third-order theory, there are five aberrations from which a system may suffer. They are (1) Spherical aberration. (2) Astigmatism. (3) Coma. (4) Curvature of the image field. (5) Distortion of the image.

In addition to these five third-order aberrations, there are two arising from variations in the initial velocities of the electrons as they leave the object. They are known as the chromatic difference in image position and chromatic difference in magnification, in analogy with the optical chromatic aberrations caused by the variation of the index of refraction with the wavelength of light.

These aberrations play an important role in electron imaging and, hence, merit consideration.

Spherical aberration is present when rays leaving the object at the axis and passing through the outer portions of the lens field do not converge at the paraxial image point. This is illustrated in Fig. 4.27*a*.

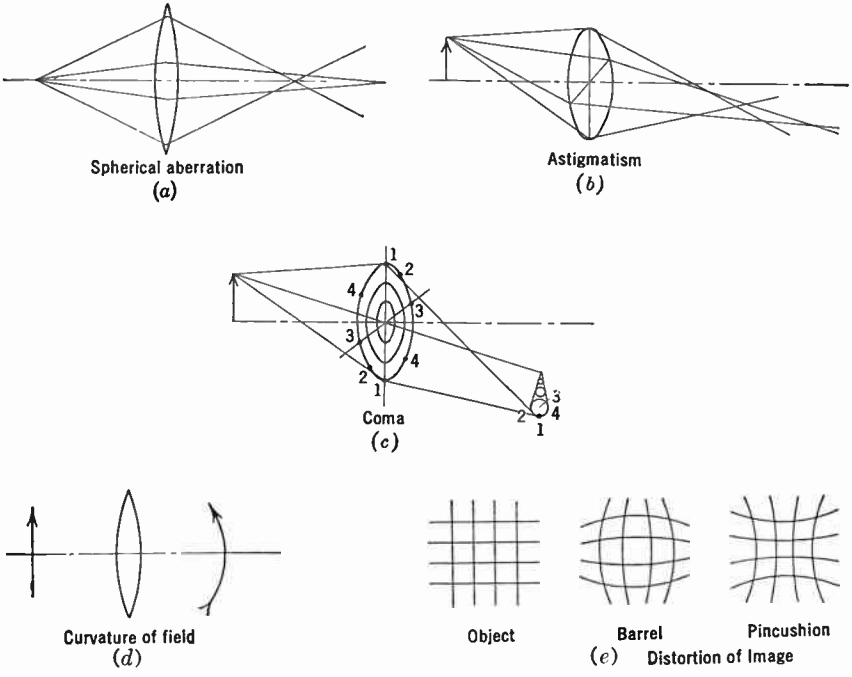


Fig. 4.27. Nature of the Five Third-Order Aberrations.

Astigmatism. The images of object points lying at a distance from the axis of the system suffer from an additional defect resulting from the fact that rays in a plane containing the axis converge on a different point from those in a plane perpendicular to this plane. This defect, known as astigmatism, is shown in Fig. 4.27*b*.

Coma. Even in the absence of the two aforementioned aberrations, the image of a point off the axis may not be sharp, but will be a comet-shaped area whose vertex coincides with the first-order image point. Because of the form of the image, this defect is known as coma (*c*).

Curvature of the Image Field. Again leaving chromatic defects out of consideration, every point of the object will be sharply imaged after correction has been made for spherical aberration, coma, and astig-

matism. However, the image points may not lie in a plane unless correction for the curvature of the image field has been made (*d*).

Distortion of the Image. Even in the absence of the above four aberrations, a fifth aberration may be present, consisting of a nonuniformity of the magnification or a twist of the image (*e*).

4.14 Aberrations. The foregoing classification may be derived from a consideration of the symmetry conditions of image formation; hence it is applicable to optical and electron lens systems. The exact derivation of these aberrations is beyond the scope of this chapter, but the dependence of the aberrations on the parameters of the imaging rays can be shown in the following simplified way.

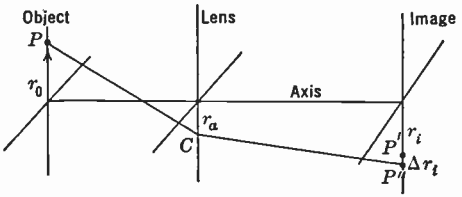


Fig. 4.28. Object, Lens, and Image Planes of an Electron-Optical System.

Figure 4.28 represents a plane through the lens system which includes the axis. The lens itself is confined to a narrow region as indicated on the figure. An object point, P, is located at a distance r_o from the axis and the corresponding Gaussian image point, P', at r_{io}. In order that the image be faithful, r_{io} must be proportional to r_o. This condition is fulfilled by the first-order image. Next, consider the ray PCP'', which originates at the object point at r_o and meets the image plane at r_i. It is evident that r_i is a function of r_o and r_a, and that by expansion this dependence can be expressed as a power series with constant coefficients. This expression of dependence will be symbolized by

$$r_i \rightarrow r_o, r_a, r_o^2, r_a^2, r_o r_a, r_o^3, r_o^2 r_a, r_o r_a^2, r_a^3, \cdots$$

The first-order image, by definition, is such that, when r_o and r_a are small, the ray PCP'' will meet the Gaussian image point. Therefore, the aberration which is given by

$$\Delta r_i = r_i - r_{io}$$

is not dependent upon the first power of r_o or r_a. Furthermore, the axial symmetry of the system makes it impossible for r_i to depend on any term of even order. Thus, the aberration consists of terms of the third,

fifth, . . . order only. If attention is restricted to the terms of the third order, it is possible to write

$$\Delta r_i \rightarrow r_o^3, r_o^2 r_a, r_o r_a^2, r_a^3$$

The five aberrations are classified according to the terms of this expansion appearing in them. As the coefficients of these terms become zero, the corresponding aberration vanishes.

Spherical aberration is that corresponding to r_a^3; it is, therefore, proportional to the cube of the diameter of the ray bundle of electrons as they pass through the lens region. Since r_o does not appear in this expression, the aberration exists for object points on the axis to the same extent as for those off the axis.

The term with $r_o r_a^2$ indicates the extent of coma present. Since r_o appears in this term, coma vanishes for a point on the axis. Both curvature of the image field and the astigmatism are the result of non-negligible coefficients of terms in $r_o^2 r_a$. Like coma, these are field aberrations.

The final term in r_o^3 gives rise to the distortion of the image.

It will be clear from a consideration of such aberrations as coma and astigmatism that a representation in a plane containing the axis is not sufficient to describe them, since both require a consideration of rays which do not lie in this plane. A coordinate corresponding to a transverse displacement must be introduced in a complete analysis. However, this approximate survey will serve to indicate the nature of the third-order image defects, which are shown in detail in Fig. 4.29.

The procedure for calculating the image defects of an actual system is very laborious. The usual method is to calculate two paraxial rays which establish the location and magnification of the image in the manner which has already been described and to utilize these and the known field variation within the lens to evaluate integrals representing the aberration coefficients. For more complete discussions of the analytical methods of dealing with these aberrations, the reader is referred to books on electron optics.*

It can be shown that it is impossible to eliminate spherical aberration completely in electrostatic, magnetic, or combined lens fields of the usual character. In other words, an aplanatic lens does not exist in present-day electron optics. Furthermore, the spherical aberration for

*See, in particular, the papers by Scherzer and Glaser in Busch and Brüche, reference 2, pp. 33 and 24, and Chapters 16 and 17 of Zworykin, Morton, Ramberg, Hillier, and Vance, reference 4.

a combined electric and magnetic lens cannot be made appreciably less than for either type alone. Immersion lenses having a photocathode as object, such as are used in the image tube, in general require the focusing of extremely narrow ray bundles, so that this aberration will be small. Spherical aberration, however, is a very important defect in the electron gun and is one of the limiting factors in producing a small

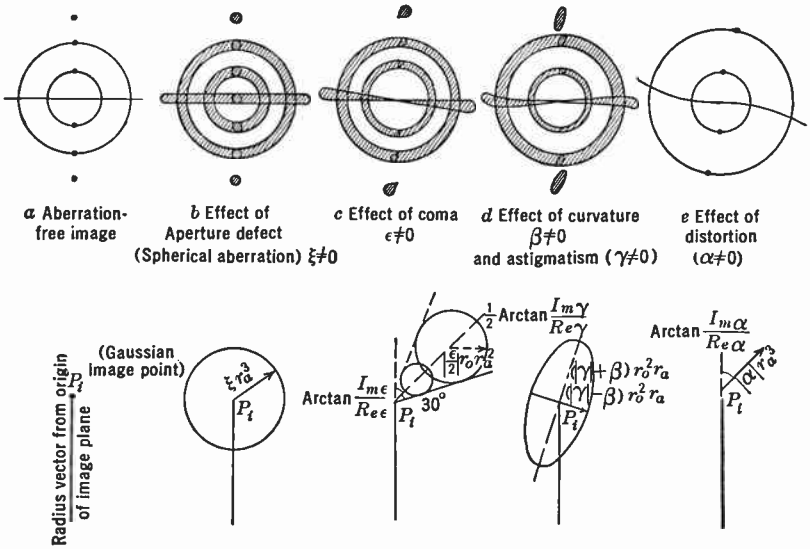


Fig. 4.29. The Effects of the Individual Geometric Aberrations on an Electron Image.

spot. In this connection this aberration has been studied for coaxial cylindrical lenses by Epstein and by Gundert.* A general method for the experimental determination of spherical aberration has been described by Spangenberg and Field.†

In order of magnitude, the spherical aberration of the most commonly employed electron lenses is comparable with that of uncorrected glass lenses. Figure 4.30 shows the variation of the spherical aberration constant with refractive power for the coaxial cylinder lens and a simple magnetic gap lens. These curves apply specifically to the

* See Epstein, reference 19, and Gundert, reference 27.

† See reference 21 or Zworykin, Morton, Ramberg, Hillier, and Vance, reference 4, p. 618.

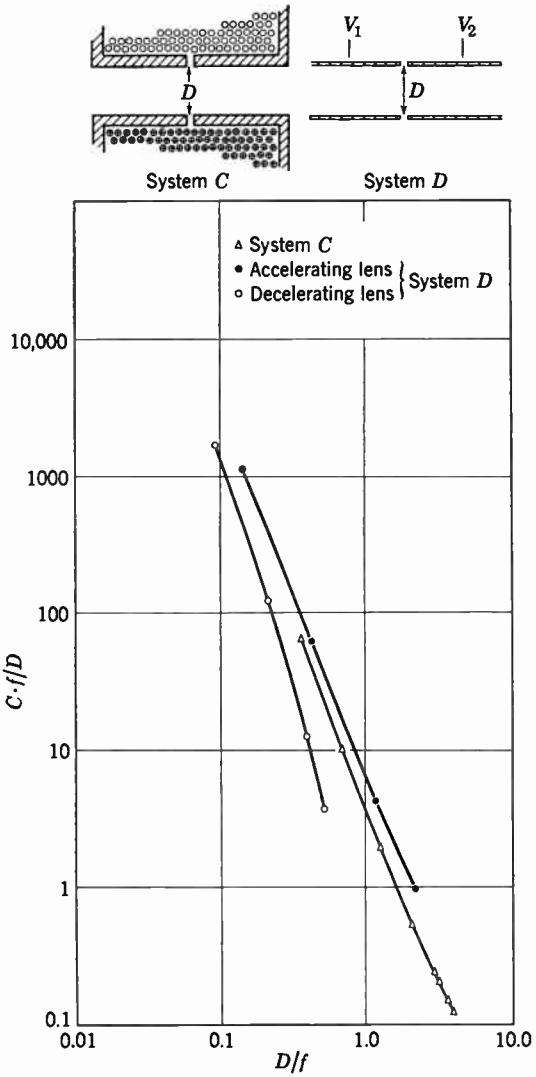


Fig. 4.30. Spherical Aberration as Function of Refractive Power for Electrostatic Equidiameter Cylinder Lens and Magnetic Gap Lens. (After Ramberg, reference 25.)

case of large magnification M. The constant C is then defined by the relation

$$\Delta r_i = MCf\alpha^3$$

for an object point on the axis; f here denotes the object-side focal length; α, the angle of inclination (aperture angle) of the ray at the object point. The refractive power of the lenses is increased, of course, by increasing the ratio of the applied potentials in the case of the electrostatic lens and increasing the exciting current in the case of the magnetic lens. If the potential of the object-side cylinder of the electrostatic lens is made lower than that of the image-side cylinder, it

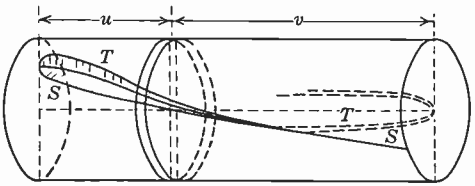


Fig. 4.31. Curvature of Field and Astigmatism in a Flat-Cathode Electrostatic Image Tube.

represents an accelerating lens; in the converse case, a decelerating lens.

Curvature of field and astigmatism, as was shown, have the same dependence upon object position and radius of the lens aperture. No practical lens having a plane object has been found by either theoretical or experimental methods which is free from both these defects, except those consisting of a uniform electric and magnetic field. Furthermore, the image plane is invariably concave toward the lens, the tangential image normally having the greater curvature. These defects depend on the first power of the aperture and on the square of the distance of the object point from the axis. They are hence, despite the small aperture of the electron beams, very important in the image tube, where the object field is large. The formation of the sagittal and tangential images in a flat cathode image tube is shown in Fig. 4.31. It has been found possible to correct these two defects in the image tube to a large extent by curving the cathode. Since these aberrations are field aberrations which vanish for points on the axis, they do not enter into the design of the electron gun.

Coma can be made to vanish in a system where the electric field is symmetrical about a plane midway between the object and image, and

the magnetic field antisymmetric. In the image tube, where the electron has zero velocity at the cathode, this type of correction obviously is impossible as the electron would have to have zero velocity at the image surface as well. On the other hand, the small aperture of the imaging pencil makes this defect unimportant. Furthermore, since coma vanishes on the axis, it plays a small role in the design of electron guns.

Distortion, the last of the five Seidel aberrations to be considered, can be divided into two types: radial distortion and rotational distortion. The former alone is present in the electrostatic lens. When it has the form of magnification increasing with radial image distance, it leads to pincushion distortion, the type most prevalent in electron lenses. The converse variation of magnification leads to barrel distortion. Both rotational and radial distortion may be present in magnetic systems. Rotational distortion arises from a variation of the rotation of the image point about the axis with its radial position. Both types of distortion may be nullified by appropriate design of the lens fields.

It can readily be demonstrated that the refractive power of continuous electron lenses invariably decreases with an increase in the initial velocity of the electrons. Achromatic electron lenses hence do not exist. On the other hand, the introduction of an electron mirror into the imaging systems permits complete correction of a cathode lens for two different initial velocities, as well as the correction of first-order chromatic aberration for other electron lenses.

The magnitude of the chromatic aberration (chromatic difference in image position) for cathode lenses depends on the ratio of the initial velocity (in volts) to the overall voltage, and upon the configuration of the lens. For simple image tube types, the diameter of the circle of aberration can readily be estimated. The three configurations illustrated in Fig. 4.32 have the following values for Δ, the diameter of the circle of aberration:

(a) simple accelerating field: $\Delta = 4L\sqrt{V/\phi}$,

(b) uniform magnetic and electric fields superposed: $\Delta = 2L(V/\phi)$,

(c) image tube with electric or magnetic lens: $\Delta = 2m(V/F)$, which, for a short magnetic lens, with the whole field applied between lens and object, becomes

$$\Delta = \frac{2m}{2m+1} L \frac{V}{\phi}$$

In the foregoing equations, the symbols have the following significance: L is the length of the tube; m the magnification of the image; V the initial kinetic energy in electron volts; F the field strength at the cathode; and ϕ the overall applied voltage.

The dependence of this aberration on the field strength at the cathode is apparent from the relations above. It is evident that high field strengths are required to reduce chromatic image defects.

Chromatic difference of magnification (and rotation) distorts an image point into a short line segment proportional to the distance of

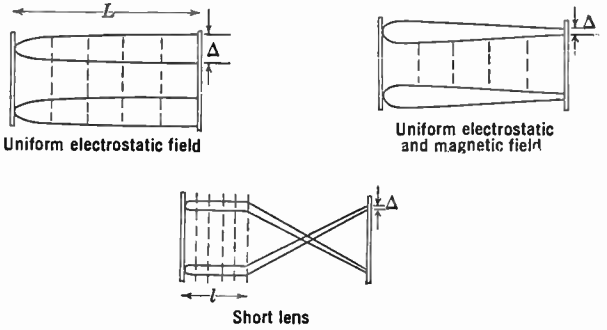


Fig. 4.32. Chromatic Aberration for Three Simple Image Tubes.

the image point from the axis. It can be nullified by measures similar to those effective for coma and distortion.

The preceding treatment of electron optics and, in particular, the discussion of aberrations have necessarily been very much simplified and abbreviated since, although important in the development of a television system, electron optics represents a very small part of the whole field.

REFERENCES

1. E. Brüche and O. Scherzer, *Elektronenoptik,* Julius Springer, Berlin, 1934.
2. H. Busch and E. Brüche, *Beiträge zur Elektronenoptik,* J. A. Barth, Leipzig, 1937.
3. I. G. Maloff and D. W. Epstein, *Electron Optics in Television,* McGraw-Hill, New York, 1938.
4. V. K. Zworykin, G. A. Morton, E. G. Ramberg, J. Hillier, and A. W. Vance, *Electron Optics and the Electron Microscope,* John Wiley, New York, 1945.
5. O. Klemperer, *Electron Optics* (Second Edition), Cambridge University Press, 1953.
6. H. Busch, "On the operation of the concentrating coil in the Braun tube," *Arch. Elektrotech.,* Vol. 18, pp. 583–594, June 1927.

7. M. Knoll and E. Ruska, "Geometric Electron Optics," *Ann. Physik,* Vol. 12, pp. 607–640, 641–661, 1932.
8. J. Picht, "Theory of Geometric Optics for Electrons," *Ann. Physik,* Vol. 15, pp. 926–964, 1932.
9. E. Brüche, "Fundamentals of Geometrical Electron Optics," *Z. tech. Physik,* Vol. 14, pp. 49–58, 1933. See also *Z. Physik,* Vol. 78, pp. 26–42, 1932.
10. O. Scherzer, "Theory of Electric Electron Condensing Lenses," *Z. Physik,* Vol. 80, pp. 193–202, 1933.
11. W. Glaser, "Geometrical Optics of Electron Rays," *Z. Physik,* Vol. 80, pp. 451–464, 1933.
12. V. K. Zworykin, "Electron Optics," *J. Franklin Inst.,* Vol. 215, pp. 535–555, 1933.
13. O. Scherzer, "Errors of Electron Lenses," *Z. Physik,* Vol. 101, pp. 593–603, 1936.
14. W. Rogowski, "Aberrations of Electron Images," *Arch. Elektrotech.,* Vol. 31, pp. 555–593, 1937.
15. L. H. Bedford, "Electron Lens Formulas," *Proc. Phys. Soc. (London),* Vol. 46, pp. 882–888, 1934. See, also, C. J. Davisson and C. J. Calbick, *Phys. Rev.,* Vol. 38, p. 585, 1931; and Vol. 42, p. 580, 1932.
16. R. Gans, "Electron Paths in Electron Optics," *Z. tech. Physik,* Vol. 18, pp. 41–48, 1937.
17. F. Gray, "Electrostatic Electron Optics," *Bell System Tech. J.,* Vol. 18, pp. 1–31, 1939.
18. S. Bertram, "Determination of the Axial Potential Distribution in Axially Symmetric Electrostatic Fields," *Proc. I.R.E.,* Vol. 28, pp. 418–420, 1940.
19. D. W. Epstein, "Electron Optical Systems of Two Cylinders," *Proc. I.R.E.,* Vol. 24, pp. 1095–1139, 1936.
20. G. A. Morton and E. G. Ramberg, "Electron Optics of an Image Tube," *Physics,* Vol. 7, pp. 451–459, 1936.
21. K. Spangenberg and L. M. Field, "Some Simplified Methods of Determining the Optical Characteristics of Electron Lenses," *Proc. I.R.E.,* Vol. 30, pp. 138–144, 1942.
22. J. Dosse, "Optical Data of Strong Electron Lenses," *Z. Physik,* Vol. 117, pp. 722–753, 1941.
23. E. G. Ramberg, "Simplified Derivation of General Properties of an Electron Optical Image," *J. Optical Soc. Am.,* Vol. 29, pp. 79–83, 1939.
24. H. Voit, "Third-Order Electron-Optical Aberrations," *Z. Instrumentenk.,* Vol. 59, pp. 71–82, February 1939.
25. E. G. Ramberg, "Variation of Axial Aberrations of Electron Lenses with Lens Strength," *J. Appl. Phys.,* Vol. 13, pp. 582–594, September 1942.
26. W. Henneberg and A. Recknagel, "Chromatic Errors of Electron-Optical Systems," *Z. tech. Physik,* Vol. 16, pp. 230–235, 1935.
27. E. Gundert, "The Aperture Defect of Electrostatic Electron Lenses," *Die Telefunkenröhre,* No. 19–20, pp. 61–98, March 1941.
28. E. P. Adams, *Smithsonian Mathematical Formulae and Tables of Elliptic Functions,* Smithsonian Institution, Washington, D. C., 1922.
29. W. Glaser, "Exact calculation of magnetic lenses with the field distribution $H = H_0/[1 + (z/a)^2]$," *Z. Physik,* Vol. 117, pp. 285–315, 1941.

PART 2

Principles of Television

5 Fundamentals of Television

The primary function of a television system is to produce on the observer's viewing screen a satisfactory reproduction of the scene before the pickup camera. The basic principles of the television systems now in use will be developed along very general lines in this chapter, and an attempt will be made to state what is meant by "satisfactory reproduction." The latter is a formidable task because of the wide variety of material which is handled by television and because the problem cannot be solved on a technical basis alone, but must also take into account physiological and psychological considerations. Although the experience of more than a decade of commercial television and a background of many years of motion-picture practice supply many of the answers that are needed, many questions remain which cannot be answered.

Since television is not a matter of transferring the observer's eyes to the camera position, but rather one of producing a picture which best represents the subject before the camera, there are actually two parts to the problem. First, what constitutes a good picture, and, second, what must be the relationship between the elements of the picture and those of the subject to give the best representation.

The discussion of the quality of the picture required is not concerned with its artistic aspects, but only with the purely technical factors involved. Some of these factors are size, aspect ratio, definition, contrast, brightness, and spurious shading, both coarse- and fine-grain, which may be superimposed on the subject matter.

In the projection of the original scene on a plane, such as may be produced by a camera, the brightness is a continuous function of the vertical and horizontal position (h, w) within the picture area and of time (t). Obviously, the brightness distribution cannot be expressed as a function of one variable (e.g., time t) only. Consequently, it is fundamentally impossible to transmit the complete picture over a single one-dimensional channel (or a finite number of such channels) as is required in television transmission. It is equally true, however, that

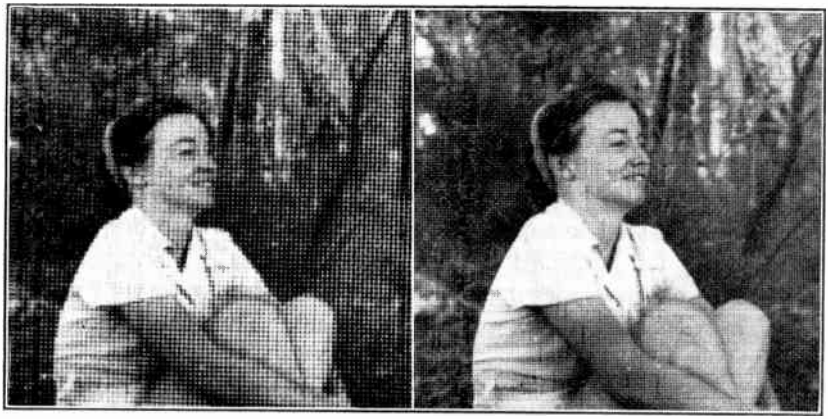
such a picture contains more information than can be physiologically accepted by the observer's eyes. Two physiological limitations of the eye, namely, its finite resolving power and the persistence of vision, set the upper limit to the rate at which information can be received by the observer. The present methods of television are possible only through the exploitation of these two properties of human vision.

If a picture is considered as made up of small elements of area, each uniform in brightness, the limited resolving power of the eye means that there is a limit beyond which any further decrease in element size or increase in element number does not improve the picture. When the element size and number reaches this limit, the picture becomes indistinguishable from the original truly two-dimensional continuum. This can be seen readily from Fig. 5.1. It might be pointed out that this process of breaking a picture up into a finite number of elements is made use of in all half-tone printing processes. Thus, it can be concluded that to transmit a satisfactory stationary picture, it is necessary only to transmit a finite number of elements of information (i.e., magnitudes representing the brightness of each element).

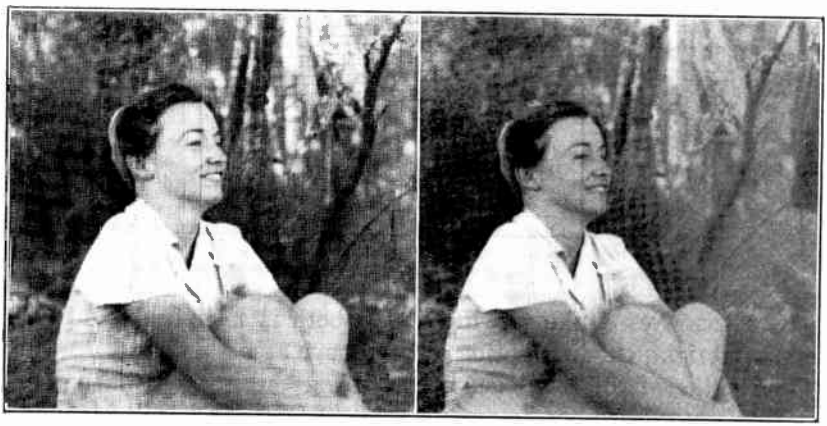
Because of the persistence of vision, it is possible to produce the illusion of continuous motion by a series of rapidly superimposed stationary pictures. It is well known that, if pictures taken of an object in a consecutive series of positions are observed in sequence, at a rate of more than twelve or sixteen pictures per second, the image gives the impression of smooth motion. This is, of course, the familiar phenomenon utilized in producing motion pictures.

If then, the information required to form each individual picture is transmitted in a short enough time so that persistence of vision can be used to give the impression of continuous motion, a satisfactory moving picture can be produced. Thus, it can be concluded that to reproduce a moving scene with all the detail of structure and motion that can be accepted by human vision requires the transmission of information at a finite rate only, and, therefore, can be accomplished over a single communication channel.

5.1 Basis of Television Transmission. The problem of the transmission of a series of pictures suitable for giving a moving picture, each picture being subdivided into a finite number of picture elements, does not have a unique solution. Of the several possible methods, that involving the process of scanning has been chosen as the most practical. This process consists of moving an exploring element or spot over the image to be transmitted in a periodically repeated path cover-



75^2 picture elements 150^2 picture elements



225^2 picture elements 375^2 picture elements

original

Fig. 5.1. Relation between Picture Perfection and Number of Picture Elements.

ing the entire image area. The exploring element is so constructed that it generates a signal giving information as to the brightness (either instantaneous or averaged over a time equal to or less than the time of persistence of vision) of the picture element on which the exploring element momentarily rests. This signal is transmitted over the communication channel to a picture-reproducing spot whose brightness is controlled by the information supplied by the signal. The reproducing spot moves over the viewing screen in a path similar to, and in synchronism with, that of the exploring element. Thus, the reproducing

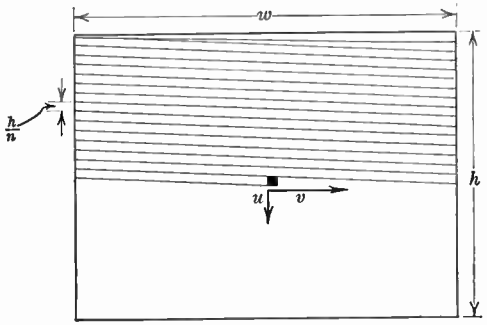


Fig. 5.2. Scanning Raster.

spot reconstructs at the viewing screen, both in magnitude and position, the brightness distribution on the image area.

The paths of the moving spots covering the image area and the viewing screen need not be continuous. In fact, the form of scanning which has been universally accepted as most satisfactory for television is one in which the spots move in a series of straight parallel lines covering the picture area as shown in Fig. 5.2. There is, of course, a theoretical infinity of paths which might be chosen to cover the picture area, for example, radial scanning, spiral scanning, sinusoidal scanning, etc., to name a few. These other forms of scanning find application in radar practice, oscillography, and other special fields, but in what follows, unless otherwise stated, straight-line scanning will be assumed.

Before enlarging on the implications of the scanning process of the preceding paragraph as applied to television, it might be of interest to examine further the statement that scanning methods are not the only solution to picture transmission over a single finite communication channel. Another possibility, again assuming the picture subdivided into a finite number of elements, is to assign a different frequency to each element in such a way that the frequency characterizes the posi-

tion of the element and the amplitude of a component of frequency gives the brightness of that element. This might be accomplished in principle by constructing n oscillators, each tuned to a different frequency and each controlled in amplitude by the output of a photocell located at one of the n picture elements of the image area. The n signals of different frequency are combined to form a single complex wave output, which is transmitted to the receiver. The receiver consists in principle of n tuned circuits to match the oscillators. Each tuned circuit includes a detector which controls a light source forming a picture element of the viewing area. Each circuit is tuned to the

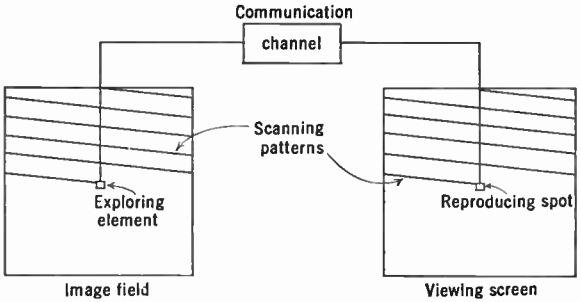


Fig. 5.3. Functional Representation of a Television System.

frequency of the corresponding oscillator at the transmitter. In this way the brightness of every picture element of the viewing screen reproduces the brightness of its counterpart at the image field of the camera.

A study of the alternatives has shown them to be less satisfactory than the straight-line scanning process used in present-day television. The complete system for this type of television transmission and reception can be represented by the five basic elements shown in Fig. 5.3. These elements are (1) Scanning pattern or raster on the image field. (2) Scanning raster on the viewing screen. (3) Exploring element. (4) Reproducing spot. (5) Communication channel.

Roughly speaking, the geometric correctness of the reproduced picture is dependent upon the similarity and synchronism of the scanning rasters at the pickup and receiver; the tonal values (i.e., contrast ratios, etc.) upon the communication channel; the range of brightness at the receiver upon the reproducing spot; and the resolution upon all five factors. Spurious signal and *noise* also depend upon all five of these factors. Actually, this generalization is too broad to be of much value in considering a physical system of television, particularly be-

cause under it a portion of the communication channel is included in both pickup means (camera tube) and viewing device (kinescope). However, for purposes of analyzing the transmission and reconstruction of a picture, this separation of functions is useful, and, once the analysis has been completed, the results can be fitted into the physical system.

In electronic television, as used today, the exploring element and image scanning raster are in the pickup device (i.e., camera tube), for example, the iconoscope or image orthicon, where the exploring spot is formed by cathode-ray beams deflected over the raster by means of a suitable magnetic deflecting system. The reproducing spot and viewing screen raster are in the cathode-ray viewing tube, e.g., the kinescope, where an electron beam also serves as the scanning element. The communication chanel extends into the pickup tube and includes the chain of video amplifiers, the radio transmitter, the medium through which the signal is propagated, the receiver with its amplifiers and, finally, the control element in the viewing tube.

The communication channel must be capable of accommodating a very wide band of frequencies in order to transmit all the information required to reconstruct the picture. That this is so can be seen from the following very rough consideration. However, it is a point of great importance and will be examined in greater detail later in the discussion.

Suppose that the picture being scanned has n elements and the repetition rate is N pictures per second. Then, if alternate elements are black and white, the signal frequency will be $f = nN/2$—roughly the maximum frequency which will ever have to be transmitted. All other frequencies down to the picture frequency N will be required for general image reproduction. Thus, the channel width must be approximately $\Delta f = nN/2$.

To return briefly to the non-scanning type of picture transmitters mentioned early in this section, it might be thought that a narrower band of frequencies might be possible. For example, if the n frequencies needed could be less than one cycle per second apart, the bandwidth required would be $\Delta f < n$. However, this would be possible only for a stationary picture. It can be readily shown with the aid of Fourier transforms that a train of waves of periodicity $1/f$ and duration T seconds actually includes a band of frequencies $\Delta f \simeq 1/(2T)$. Since, to obtain the illusion of continuous motion, each frequency must be capable of changing in $T = 1/N$ seconds, the total bandwidth must be $\Delta f \simeq n/(2T) = nN/2$.

Indeed, it can be shown in general that to transmit nN pieces of information per second, the frequency bandwidth is of the order of nN. This is known as Hartley's law.*

The transmission of a coded picture is in a sense an exception to Hartley's law. However, this apparent exception occurs only because of the different nature of the information transmitted. To illustrate coded transmission, consider the very simple picture consisting of $n = 16$ elements illustrated in Fig. 5.4, where each element may be either black or white. Again, the repetition rate is N pictures per second. Evidently all possible pictures are given by the number $2^{16} \sim 66{,}000$. By assigning a number or amplitude to each configuration, it would be possible by the transmission of a single amplitude N times per second, or a bandwidth of $\Delta f = N$, to convey all the information needed to reconstruct the picture. This assumes that the single amplitude can be transmitted with sufficient accuracy to permit the discrimination of some 66,000 steps at the receiver. In principle, coding of this type could be used where, instead of simply black and white elements, the picture consisted of elements which had any desired number of discrete steps of brightness. Furthermore, it is not necessary to transmit the information only once each frame. For example, a series of amplitudes might be transmitted, each giving information about two picture elements, so that the channel bandwidth required would be half that needed for normal transmission. Since this form of transmission is not used at present in practical television, it will not be discussed further here.

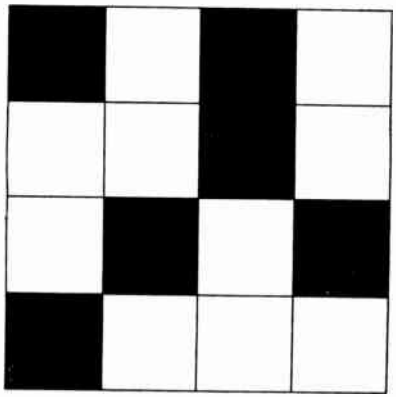


Fig. 5.4. Coded Transmission: A Single Signal with 2^{16} Distinguishable Steps of Amplitude Can Serve to Indicate Which of the 16 Squares Are Black and Which Are White.

5.2 General Considerations of Picture Quality. Before continuing the discussion of the reproduction of pictures by television, it would be well to determine, in as far as possible, what technically constitutes a satisfactory picture for the purpose. To do this, it is necessary to examine critically the requirements of size, brightness, resolution, contrast, and spurious signal (including noise) of a picture suitable for

* See Hartley, reference 3.

regular viewing in the average home. It should, however, be recognized that there is a very great latitude in meeting most of the requirements, and that because of technical limitations of all present-day television systems, many of the requirements are met with minimal rather than optimum figures.

Size. The size of the picture most satisfactory for home viewing depends upon what should be considered the normal viewing distance. Experiment and moving-picture experience show that the height of a picture whose aspect (dimensional) ratio is $V:H = 3:4$ should subtend a vertical angle of not more than about 15 degrees. Such an angle permits obtaining a more or less simultaneous impression of the entire content of the picture, and a detailed examination of all points simply with motion of the eyeballs, without motion of the head. This viewing angle means that the ratio of picture distance to height should be not less than about 4. In other words, if 6 feet is assumed to be the closest viewing distance, the picture should be about 19 inches in height V and 25 inches in its horizontal dimension H. However, it is not a sharp optimum, and both smaller and larger pictures have been found to be entirely acceptable.

Brightness. It is as difficult to assign an optimum brightness to the picture as to determine a best size. The picture should be bright enough so that it can be readily watched in a moderately lighted room. The brightness also should be great enough so that all the information that is in the picture can be perceived without strain. Illumination engineers have found that for reading, or for close work, an illumination of 10 foot-candles is desirable. This means that a white surface (e.g., a piece of paper) will have a brightness of 8 to 9 foot-lamberts. However, since a television receiver is generally operated where relatively high ambient illumination tends to reduce picture contrast, a higher kinescope brightness is desirable. It has been estimated that the average brightness should be not less than 10 foot-lamberts in order to give the required contrast range and detail visibility. The peak brightness should be able to reach values of 30 to 50 foot-lamberts.

The range of brightness over which the eye can sense light is enormous. It is of the order of 10^8 if both photopic and scotopic (dark-adapted) vision are included. The brightness range for normal photopic vision is much smaller, being of the order of 10,000. However, rarely does any one scene cover a range of more than 1000. Inasmuch as the reflectivity of a very good black pigment, or of black velvet, is about 1 percent, the brightness range of pictures almost never exceeds 100,

and in photographic work a range of 30 is considered good. A brightness ratio of 100 should be entirely satisfactory, and even a ratio of 30 not too low for good reproduction with a television receiver. A receiver operated with moderate ambient light, arranged so that no lights shine on the viewing screen, can easily exceed the smaller of these values and may reach the larger. The brightness of the screen due to reflected light (i.e., the brightness of black areas) will be a few tenths of a foot-lambert, and the highlights may reach 30 foot-lamberts.

Contrast. The contrast of an object with its background is generally defined as

$$\text{Contrast} = \frac{B_o - B_s}{B_s}$$

where B_o is the brightness of the object and B_s that of the background. If B_o and B_s differ by a small amount, it is found that for a wide range of values for B_s, if the relationship

$$\Delta B/B = \text{constant}$$

is satisfied, the same sensation of increase of brightness is obtained. This is known as Fechner's law. The extent to which it is satisfied is

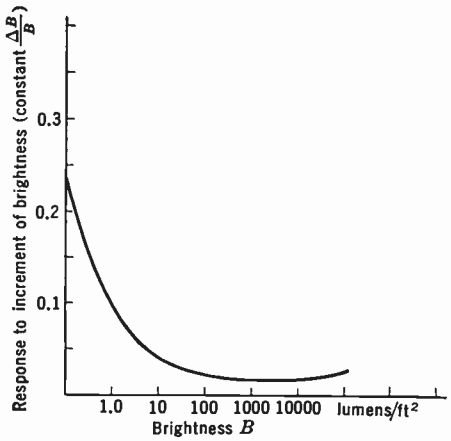


Fig. 5.5. Visual Response to Incremental Increases in Brightness.

shown in Fig. 5.5. From a brightness of a few foot-lamberts to many thousand, it will be seen that $\Delta B/B$ is practically independent of brightness. To the extent to which Fechner's law holds, the visual

response to brightness will be proportional to the logarithm of the brightness, as is illustrated in Fig. 5.6. The concepts of contrast, and contrast ratio between reproduction and original, are important in discussing the relation between camera and viewing screen.

In comparing the contrast properties of two images, for example, that of a transmitted and received image, it is useful to construct a logarithmic plot of the brightness of corresponding points. When the response of the channel is linear and the gain unity, the brightness of

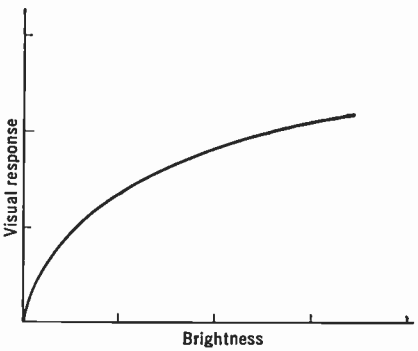


Fig. 5.6. Visual Response as a Function of Brightness.

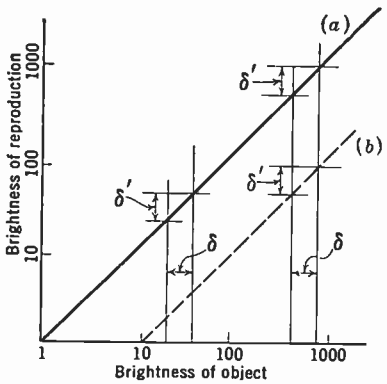


Fig. 5.7. Brightness and Contrast.

transmitted and received image points plotted in this way lies on a straight line at 45 degrees to the abscissa, as shown in curve a of Fig. 5.7. A change in gain does not alter the slope of the line, but merely moves it up or down, as is the case with curve b representing a gain of $\frac{1}{10}$. A given contrast, irrespective of the average brightness, between two areas is represented by a constant interval. For example, the interval δ in the figure represents a contrast of 0.2 in the image. Since both lines a and b make an angle of 45 degrees, a given δ of the transmitted image results in an equal interval δ' along the ordinate representing the received image brightness. Therefore the contrast in the two images will be equal.

When the curve representing the received and transmitted brightness is a straight line making an angle either more or less than 45 degrees with the abscissa, as shown by curves b and c, Fig. 5.8, the contrast in the two pictures will no longer be the same. However, the relative contrast of various parts of the transmitted image will be the same as

for those in the received image. That is, if the contrast between two objects in one portion of the transmitted image is twice that of two objects in some other portions, the contrast between the first pair of objects will be twice that of the second in the received image. This relationship is illustrated diagrammatically in the figure. If the line representing the response makes an angle greater than 45 degrees, the contrast in the received image is greater than that in the transmitted image, and the received image appears crisp or harsh, whereas if the angle is less than 45 degrees, the rendition is soft or flat.

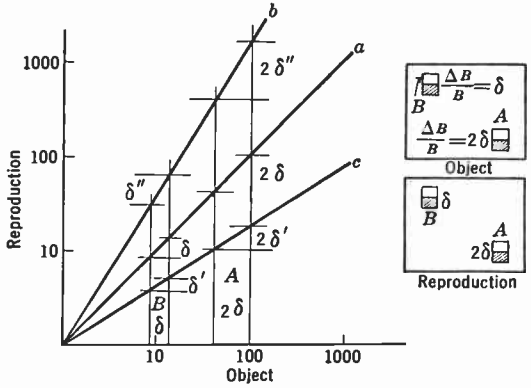


Fig. 5.8. Relative Contrasts in Object and Reproduction.

In photographic practice the slope of the response line has been given the name gamma (γ). Thus for identical contrasts γ is unity; if it is greater or less in the reproduction, γ is greater or less than one. The concept of γ is very useful in photography because in the process of development the approximate linearity of the logarithmic plot between the reproduction transparency and the brightness of the original is preserved, but the slope or γ of the curve changes with the various factors, such as time and temperature, influencing development. This same nomenclature proves useful in the field of television, although in the physical communication channel there is no element which automatically preserves the linearity of logarithmic response.

Resolution. The definition, that is, the finest detail which exists in a picture, is an extremely important factor in determining its quality. In general, a picture which is not sharp is unsatisfactory, and is both difficult and tiring to watch. Obviously, the apparent sharpness of a given picture depends upon the distance between it and the observer.

A picture which may look sharp at a distance may look badly blurred when viewed at close range. In other words, it is a matter of the angle subtended at the eye. The sharpness also depends upon the brightness of the picture; that is, a blurred picture may look sharp in a dim light. Finally, with a given brightness, a high-contrast picture will show greater detail than a low-contrast picture.

Since sharpness is one of the requisites of a good picture, the next step is to gain a quantitative idea of the definition required to satisfy the eye in this respect. One measure of the resolving power of the eye is the angular separation between two points which are so close together that they can only just be distinguished as two objects. For a good picture, taking into account ordinary brightness and detail contrast, probably 1½ minutes of angle is a good working figure. This point will be considered in more detail in section 5.5. On this basis a picture which subtends 15 degrees vertically at the eye should have some 600 picture elements in that dimension. A moving picture or a television picture does not have to have quite this definition, and with 450 or 500 picture elements on a vertical line, the eye will be completely satisfied. A further discussion of the psychology of picture "sharpness" is presented in the final section of this chapter.

Spurious Signal. The amount of spurious shading or brightness variation which can be tolerated in a picture is rather difficult to specify. For convenience, spurious effects will be divided into three different types. The first is a gradual variation (not necessarily monotonic) in brightness over the picture which will be designated as "shading." The second is a sharp discontinuity in brightness, and the third a fine-grained random variation (the variations approaching picture element size in their dimensions) in brightness which for reasons which will become apparent later will be termed "noise."

False discontinuities in brightness of the second type are extremely objectionable in a picture. A change in brightness which is entirely unnoticeable if it is gradual from one side of the picture to the other will be very prominent if it is in the form of a brightness step. In fact, the mechanism of seeing is such that the eye is more sensitive to this type of discontinuity than to any other effect, either wanted or spurious, in a picture. This is because, as the eye examines any scene, it is in slight continuous motion, which permits a single perceptor center (rod or cone) of the eye to compare adjoining elements of the image. In reproducing a television picture, it is necessary to insure that this type of spurious effect is absent. The sensitivity of the eye to distortion of this form presents a serious obstacle to the application of the type of

coding mentioned in the previous section, in which discrete steps of brightness are employed.

The amount of shading permissible in a picture depends to a large extent upon the way in which the variation is distributed. If the variation is continuous from one side of the picture to the other, a variation of 2 to 3 percent is almost unnoticeable for ordinary subject matter. However, if the change goes from minimum to maximum in half the width of the picture, the change in brightness should be less than 1 percent; and if the change is in less than one-tenth of the picture width, only about 0.1 percent variation is permissible.

The two types of spurious effects described above are nonfundamental, and both can be reduced below any prescribed level by suitable electronic, electrical, or optical techniques. The third effect, the brightness fluctuation characterized as noise, is fundamental and cannot be reduced below a definite limit under given conditions. To illustrate this, consider a small element of area of h units of length on a side, which forms part of a picture. If the picture is luminous and the element has a brightness B_A, the number of lumens from the area will be

$$L = h^2 B_A$$

Now, if the area of the aperture of the observer's eye is A at a distance d, the light flux entering his eye will be

$$L' = \frac{A}{\pi d^2} L = \frac{A h^2 B_A}{\pi d^2}$$

But each lumen of "white" light consists of 1.3×10^{16} photons of visible radiation per second. Therefore, entering the observer's eye there will be

$$\frac{dN}{dt} = \frac{A h^2 B_A}{\pi d^2} 1.3 \times 10^{16} \text{ photons per second}$$

or, if T second is the persistence time of vision,

$$N = 1.3 \times 10^{16} \frac{A h^2 T B_A}{\pi d^2}$$

Since the emission of photons is a purely random effect, there will be a statistical variation in the number whose root-mean-square deviation is equal to the square root of the number. Hence

$$\sqrt{\overline{\Delta N^2}} = \sqrt{1.3 \times 10^{16} \frac{A h^2 B_A T}{\pi d^2}}$$

Obviously, if two elements A and B have the same area and nearly the same brightness, it will be fundamentally impossible to say definitely which is the brighter if $(B_A - B_B)$ is not larger than

$$\sqrt{B}/\sqrt{1.3 \times 10^{16} A h^2 T/(\pi d^2)}$$

Actually, the photosensitive process in the eye is so constituted that a factor σ representing the quantum efficiency of vision must be introduced. This increases the brightness difference necessary to insure differentiation. Experiment indicates that for the eye to recognize a difference in brightness (i.e., the smallest perceptible contrast step), the difference in the number of photons stimulating the eye must be three to five times the root-mean-square fluctuation in this number. Thus

$$\left(\frac{\Delta B}{B}\right)_{\min} \approx 5/\sqrt{\frac{1.3 \times 10^{16}\, A h^2 T \sigma B}{\pi d^2}}$$

It is quite evident that a relationship of this form will apply to any sort of pickup device, if h, T, and σ are properly interpreted.

The fluctuation in brightness just described exists in any picture. However, additional fluctuation may exist in a television picture (or in any other picture for that matter) arising from the nature of the pickup, transmission, and reproduction of the image which is very similar to the above fundamental "noise." It is a consequence of electrical fluctuations fundamental to all circuits and of "shot noise," which was discussed in Chapter 1.

This additional noise can be handled in much the same way as that described above. Assume that the desired picture has a brightness range of 100, and let the unit b of light flux from a picture element be such that

$$B_{\max} = 100b$$

corresponds to the maximum brightness. The root-mean-square fluctuation in brightness can be expressed in these units. If $\Delta B_{\text{rms}} = b$ for a single element whose average brightness is $B = nb$, it would, as was pointed out, be impossible to state with any degree of certainty whether B had brightness nb, $(n + 1)b$, or $(n - 1)b$, etc. However, if a group of say $10 \times 10 = 100$ elements is considered, the total flux from them is $100nb$ and the root-mean-square deviation $10b$. The fractional deviation has been reduced by a factor of 10.

If it were necessary to distinguish 100 steps in brightness with the certainty factor 5 mentioned in an earlier paragraph for every picture element, the maximum root-mean-square deviation permissible would be

$$\Delta B = b/5$$

The signal to noise ratio defined as

$$\frac{B_{\max}}{\Delta B_{\text{rms}}}$$

would have to be 500. However, the eye cannot possibly distinguish 100 steps in brightness in a single picture element. Actually 5 brightness steps would be adequate. Hence, a signal-to-noise ratio of

$$\frac{B_{\max}}{\Delta B_{\text{rms}}} = \frac{100b}{20b/5} = 25$$

would be adequate. With this ratio, an area of some 20 elements on a side, if considered as a unit, would have a root-mean-square deviation small enough to distinguish 100 steps in brightness. It can be concluded that for a high-quality picture the signal to noise ratio as described should be 25 to 50.

The foregoing qualifications are essential for a satisfactory picture. Other requirements must be met also, but these are characteristic of current television practice and not of a fundamental nature. Hence they will be treated in separate sections.

5.3 Picture Quality in Relation to Television Transmission. The previous section reviewed the principal factors which are essential for satisfactory picture quality. Each of the factors must be examined to determine what requirements they place upon the basic television system outlined in section 5.1.

The dependence of geometric correspondence between image and reproduction upon the similarity and synchronism of the two scanning patterns is evident without further explanation. A slight distortion can be caused by faults in the communication channel, such as variable time delay of the signal, but these faults produce such a serious destruction of tonal values that when this latter aspect is corrected their geometric effect is entirely negligible. Therefore, for all practical purposes, geometric distortion arises only from scanning defects. The problem of maintaining synchronism and fidelity of scanning rasters will be the subject of a later chapter.

It is evident that some relationship must exist between the brightness of the elements of the image at the camera and the brightness in the reproduction at the viewing device which will yield the most satisfactory representation of the scene before the camera. The simplest relationship which might be postulated is one for which the brightness of corresponding elements in image and reproduction are proportional. This would be achieved if, for example, the voltage output of the exploring element is proportional to the image brightness, the channel has the characteristics of a linear amplifier, and the brightness of the reproducing spot is proportional to the applied voltage. In practice, neither the reproducing spot nor the exploring element is linear in its characteristics, so that it is frequently necessary to distort the amplitude response of one or more of the amplifiers in the transmission channel to obtain the desired contrast characteristics.

It has been pointed out that, in order to maintain the illusion of continuous motion, the separate pictures showing the successive stages of the changing scene must follow one another rapidly. Sixteen frames per second is sufficient for most pictures, although a few require a higher rate. The motion-picture industry has adopted 24 frames per second as a standard. This rate, possibly higher than absolutely necessary, insures smooth motion for all types of scenes. For television, in this country, a rate of 30 frames per second has been chosen, not because the high rate is needed for continuity of motion, but because it bears a simple relation to the frequency of commercial power. The reason why this is desirable will be taken up later. Actually, 30 field traversals per second, although fulfilling all the requirements of continuity of motion, is not high enough to avoid flicker effects. To meet this situation, interlaced scanning, a special form of straight-line scanning explained in the next section, is used, making unnecessary any greater frame frequency.

In the preceding section it was concluded that the definition of a picture satisfactory for television viewing should have some 500 picture elements in the vertical direction. Since the picture aspect ratio is 3:4, there should be about 650 elements horizontally. The first figure means that the scanning raster ought to consist of at least 500 lines. Present television standards call for 525 lines with about 7 percent unused to permit the return of the scanning spot to the starting position and to allow for a synchronizing signal, as will be discussed in later chapters.

Vertical and horizontal resolution bear distinctly different relationships to the mechanism of transmission. Vertical resolution depends

almost entirely upon the number of scanning lines making up the raster and upon the size of the scanning spots. The nature of the communication channel plays no direct part in its determination. The horizontal resolution, on the other hand, depends upon the frequency response of the channel and again upon the size and shape of the exploring and reproducing spots. It is related to the number of lines only because the number of lines and the frame frequency determine the time available for scanning a single line.

It is self-evident from the mechanism that has been postulated for picture transmission that the size of the exploring and reproducing spots affects the picture definition, and that small spots are required

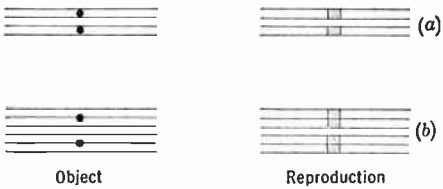


Fig. 5.9. Vertical Resolution for Various Object Positions.

for high definition both horizontally and vertically. In the vertical direction, it is obvious that two points cannot be resolved unless separated by two or three lines as indicated in Fig. 5.9. To a first approximation the resolution corresponds to the separation between lines.

The dependence of horizontal resolution upon the overall communication channel was pointed out in section 5.1. Here, it was shown that high definition can be achieved only if the frequency bandwidth of the communication channel is great. The necessity of small exploring and reproducing spots has just been mentioned. This requirement is analyzed in detail in sections 5.5 and 5.6.

5.4 Flicker. Even before the advent of motion pictures, the behavior of the visual sense towards an object which is periodically illuminated had been investigated. Since their advent the subject has become one of considerable practical importance, and consequently has been rather completely examined. The general conclusions are of interest to the television engineer, although the details, because of the rather special nature of the intermittency involved in television, are not particularly pertinent.

It was found that, although the eye is unaware of discontinuity of motion at frequencies above about 15 cycles per second, it nevertheless could detect flicker at very much higher frequency. The threshold

frequency for awareness of flicker is a function of brightness of the object, color of the light, relative duration of light and dark, and the position on the retina where the light image falls. Figure 5.10 illustrates the results of a series of tests of flicker threshold as a function of brightness for different relative amounts of light and dark.* The length of the illuminated period is given in degrees, the entire period of light and dark equaling 360 degrees. The light is averaged over the period, to correspond to the reaction of the eye to intermittent light. These tests were made with white light and with the image covering

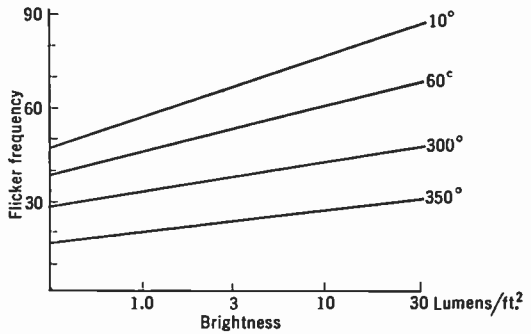


Fig. 5.10. Relation of the Threshold Frequency of Flicker to Brightness.

a large area of the retina. Other tests have shown that the threshold frequency is a maximum for light whose color is such that it produces the maximum visual sensation and decreases toward the red and blue, and, also that the sensation of flicker at a high frequency is less at the edges of the retina than at the center.

The scanning process introduces a type of flicker which is sufficiently different from that described above so that the results of these tests cannot be used, a priori, to formulate the requirements necessary to insure the elimination of flicker from the television viewing screen.

Reconstruction of the televised image on the viewing screen is accomplished by the motion of a bright modulated spot traversing the series of straight parallel lines making up the scanning pattern. Therefore, except for the very short time which may be required for the spot to return to its original vertical position (known as the return time), some point of the screen is always bombarded by electrons. However, if a given small area of the screen is considered, it obviously emits light intermittently for a very short interval at field frequency. For most

* See Engstrom, reference 4.

practical cathode-ray television viewing tubes the duration of the interval for which the screen emits light is a function of the decay characteristics of the phosphor used on the screen, rather than the actual length of time the scanning spot is on the area under consideration. For the most commonly used phosphors a bombarded area remains luminous for a time long compared to a line scan. Thus, as far as flicker is concerned, the process of picture reconstruction may be considered as being the result of a bright horizontal strip moving vertically across the screen at field frequency. The leading edge of this strip has the maximum brightness, the luminosity decaying gradually towards the trailing edge. In order to determine the reaction of the visual senses to this type of flicker, tests were made using the projected image of continuously moved film whose frames had the density distribution shown in Fig. 5.11. The results of these are shown in Fig. 5.12. The general qualitative resemblance of these results with those of Fig. 5.10 is evident. From the information given in these tests it can be concluded that the scanning field frequency must be about 50 cycles per second or higher to avoid all sensation of flicker on a bright picture.

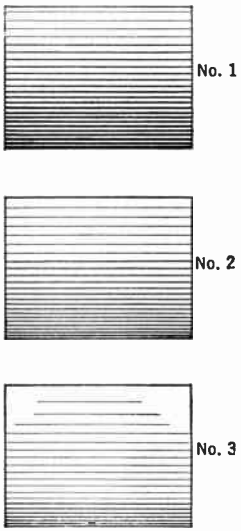


Fig. 5.11. Test Films Used to Determine Threshold of Flicker for a Scanning Pattern.

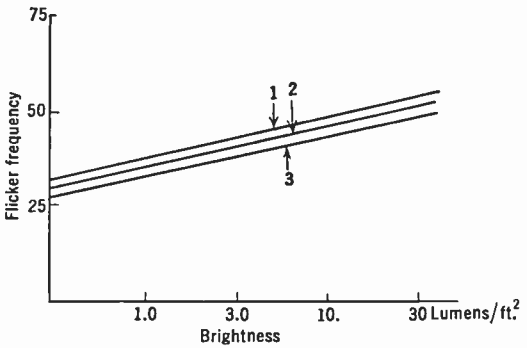


Fig. 5.12. The Threshold Frequency of Flicker for a Scanning Pattern.

This frequency is much higher than is needed for continuity of motion, and therefore requires an unnecessarily large frequency band

of the communication channel. Economy of bandwidth can be retained without reducing the effective field frequency by means of interlaced scanning.

Interlaced scanning is a form of straight-line scanning. The spot, instead of moving across the horizontal lines in the sequence of 1, 2, 3, 4 ..., covers the odd numbered lines first, then the even-numbered lines, i.e., 1, 3, 5, ...; 2, 4, 6, ... This process of scanning is illustrated in Fig. 5.13. It will be seen that the scanning beam has to complete two scanning patterns, displaced by the width of one line, in order to cover the entire field. Therefore, the field frequency is twice the frame frequency.

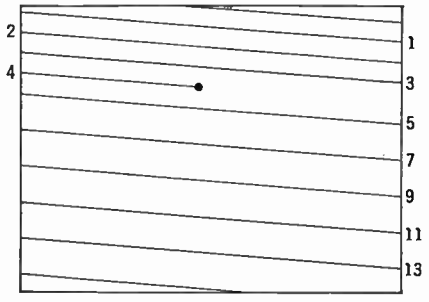


Fig. 5.13. Interlaced Scanning.

This form of scanning is not restricted to covering every second line in the first vertical sweep and the intervening lines on the second sweep. The scanning may be interlaced in such a way that the first sweep covers every third or every fourth line, etc.; that is, the sequence may be

$$1, 4, 7, 10, \ldots; \quad 2, 5, 8, \ldots; \quad 3, 6, 9, \ldots$$

or

$$1, 5, 9, 13, \ldots; \quad 2, 6, 10, \ldots; \quad 3, 7, 11, \ldots; \quad 4, 8, 12, \ldots$$

or, in general,

$$1, k+1, 2k+1, \ldots; \quad 2, k+2, 2k+2, \ldots;$$
$$3, k+3, 2k+3, \ldots; \quad \ldots; \quad k, 2k, \ldots$$

Interlaced scanning, as applied in most practical systems, employs a frequency which is twice the repetition frequency of the picture. Tests show that at optimum viewing distance, namely, a distance such that the line structure is just below the limit of resolution, the field frequency of interlaced scanning and frame frequency of sequential scanning have identical relationship to the perception of flicker. Therefore, a field frequency above 50 cycles with interlaced scanning gives a flicker-free reproduction at normal picture brightness. However, if the individual lines can be resolved, a slight interline flicker or weave may be perceptible in a bright picture. In the United States, for

reasons connected with the 60-cycle standard of commercial power, a repetition rate of 30 frames per second and a field frequency of 60 cycles per second have been adopted; in England and in continental Europe, where 50-cycle power is standard, frame and field frequencies are 25 and 50 cycles, respectively. At the high picture brightness levels available in modern receivers the higher field frequency (60 cycles per second) has been found decidedly advantageous for flicker suppression.

5.5 Resolution. Occasionally a picture which is essentially continuous is encountered, but most of those viewed are grainy in nature. Pictures of the latter type must be divided into two classes: those in which the grain is regularly disposed, and those in which the individual grains are scattered at random. Another classification divides them into those in which the individual grains may have any shading from light to dark, and those in which only black grains and white grains are used and picture contrast is obtained by the relative density of the two kinds. These two classifications are mutually inclusive.

The scanned picture cannot be brought into the fold of either grainy or continuous pictures, because along any line it is essentially continuous, while at right angles to this it is grainy in nature with a regular distribution of elements.

The question of resolution is relatively simple in a picture of regular grain, but becomes more complex for a continuous picture. Considering the former, it is evident that the grain, or picture elements, should be sufficiently small that the eye does not resolve the individual units. Numerous tests have shown that the average eye cannot resolve two points separated by a distance subtending less than 1 minute at the eye. Exceptional eyesight under ideal illumination can resolve as little as one-half minute, but eyes which cannot resolve less than 2 minutes of angle are fairly common. Actually this cannot be interpreted to mean that, where a regular pattern of picture elements is used, elements spaced at 1 minute can reproduce a picture whose limiting resolution is 1 minute of angle, because the components of the picture are randomly distributed. However, a regular pattern of elements with spacings which subtend an angle of about 1 minute of angle is capable of forming images which reach the limit of visual acuity for all but very special patterns (i.e., patterns having regularities which "beat" with the element regularity).

Measurements made by E. W. Engstrom,* using projected pictures whose structure simulates that of a television picture, give valuable

* See Engstrom, reference 4.

information relative to visual acuity, image structure, and picture size. His conclusions are based upon the results of a number of tests by several observers on pictures having structure and structureless images. The findings from some of these tests, in terms of scanning lines per inch and viewing distance, are shown in Fig. 5.14. The general conclusion from these data is that the optimum viewing distance is such that the structure subtends an angle of about 1½ minutes. At this distance the structure becomes unobtrusive or invisible, and the detail

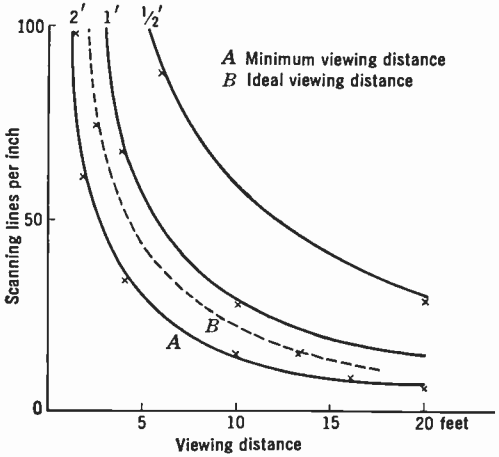


Fig. 5.14. Relation between Picture Structure and Viewing Distance.

appears to be the same as that of a structureless reproduction of a similar image viewed from the same distance.

It has been stated that an equal horizontal and vertical resolution is to be desired, at least to the best of present knowledge. Since in the vertical direction the transition from black to white occurs in a distance equal at most to the separation between the centers of two lines, that is, in a distance $l = h/n$, this same separation should be attainable in the horizontal direction. If the image of a small illuminated point is transmitted, the resultant reconstructed image will be a bright spot somewhat larger than the original point. Its width, in order to meet the resolution requirements for the horizontal width, should be not more than $2l$. In order to determine the channel width needed to fulfill this requirement, it is necessary to make use of a Fourier integral. (*Note:* A Fourier series is actually required if the spot is stationary owing to the periodicity of the signal, but is more cumbersome to apply and lends nothing to the accuracy of the present calculation.)

The scanning beam in passing over the illuminated part gives rise to an impulse of very short duration which can be represented as shown in Fig. 5.15. The frequency spectrum is given by

$$F(f) = \frac{2A}{\tau}\int_0^{\tau} \cos 2\pi f t \, dt = 2A \frac{\sin 2\pi f\tau}{2\pi f\tau} \tag{5.1}$$

where 2τ is the duration of the impulse. If τ is small, this represents an essentially constant band out to a very high frequency, the frequency limit increasing as the duration of the pulse decreases. The

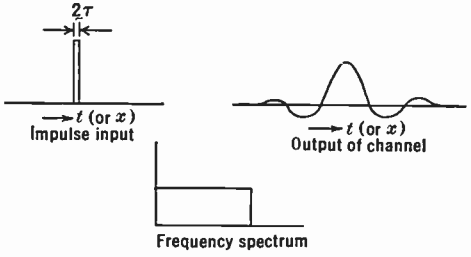


Fig. 5.15. Reproduction of a Bright Bar by a Sharply Cutoff System.

communication channel, however, transmits only a band of frequencies extending from approximately zero to a cutoff frequency f_c. Therefore, the frequency spectrum of the pulse at the receiving end of the channel consists of a band of constant amplitude $2A$ extending from zero to f_c. The integral representing the resultant pulse is

$$P(t) = 2A\int_0^{f_c} \cos 2\pi f t \, df = \frac{A}{\pi}\frac{\sin 2\pi f_c t}{t} \tag{5.2}$$

Figure 5.15 illustrates the form of the pulse. The length of time between the point of maximum amplitude and the first minimum can be found by determining the derivative of $P(t)$ with respect to t and equating it to zero:

$$\frac{dP(t)}{dt} = 2Af_c\frac{\cos 2\pi f_c t}{t} - \frac{A}{\pi}\frac{\sin 2\pi f_c t}{t^2} = 0 \tag{5.3}$$

$$2\pi f_c t_0 = \tan 2\pi f_c t_0 \tag{5.3a}$$

$$f_c = \frac{4.4934}{2\pi}\frac{1}{t_0}$$

The half width l of the reproduced image is equal to the product of the velocity of the reproducing spot and the time t_0. Therefore,

$$t_0 v = l = h/n$$

and, since

$$v = Nnw$$

it follows that

$$t_0 = \frac{h}{w}\frac{1}{n^2 N}$$

and from this the cutoff frequency of the channel will be

$$f_c = \frac{4.493}{2\pi}\frac{w}{h}n^2 N = 0.715\frac{w}{h}n^2 N \tag{5.4}$$

In this expression w/h is the aspect ratio of the picture, which may be taken as 4/3.

Equation 5.4, when evaluated for a 525-line picture having a repetition rate of 30 frames per second, gives a cutoff frequency of 7.9 megacycles.

Actually the assumptions made in the evaluation were much too simple to lead to an accurate estimate of the bandwidth. Two very important factors that were not considered at all are the properties of a physical spot, and the shape of the transmission band as it exists in reality. Before these factors can be added to the analysis of the bandwidth requirements the theory of scanning must be examined more deeply.

5.6 Theory of Scanning. The analysis of a picture into an electrical signal and the synthesis of this signal back into an image by the scanning are quite complicated. A very complete treatment of this subject was given by Mertz and Gray,* making use of harmonic analysis. The following discussion is based upon their treatment, but is necessarily seriously abridged because of lack of space; therefore an interested reader will do well to refer to the original article.

The scanning of a stationary picture is a process of repetition; consequently the signal produced can be represented by the Fourier series

$$E(t) = \sum_{k=0}^{\infty} A_k \cos\left(\frac{2k\pi t}{T} + \theta_k\right) \tag{5.5}$$

* See Mertz and Gray, reference 5.

where k is the order of the term, A_k the amplitude of the kth Fourier component, and T the frame period. It will be noted that the number and frequency of the Fourier components do not depend upon the light distribution in the picture, but only upon the scanning process. The coefficients A_k are, however, functions of the light distribution in the image being scanned.

The variable t can be expressed in terms of the position of the exploring spot with the aid of the relation

$$\frac{t}{T} = \frac{x}{wn}$$

The series can therefore be written in the form

$$E(x) = \sum_{k=0}^{\infty} A_k \cos\left(2\pi \frac{kx}{wn} + \theta_k\right) \tag{5.6}$$

This can best be visualized by considering that the scanning process consists of moving the exploring spot over an array of identical images, as shown in Fig. 5.16.

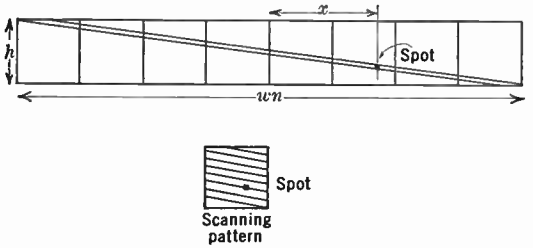


Fig. 5.16. One-Dimensional Representation of Scanning.

The one-dimensional series just given, which is required by the channel, does not provide a means of determining whether this type of series can contain components of such a form that the picture can be reconstructed.

In order to represent completely a two-dimensional image, it is necessary to use a double series. If $B(x, y)$ is the brightness at any point, x, y, the image can be described by the following series:

$$B(x, y) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} A_{kl} \cos \left[2\pi \left(\frac{kx}{w} + \frac{ly}{h} \right) + \theta_{kl} \right] \tag{5.7}$$

This series represents an array of sinusoidal variations over the field, of a wide range of frequencies and directions. Examples of some of the components are shown in Fig. 5.17. Since in the scanning process both x and y are functions of time alone, the series given in Eq. 5.7 can be written in terms of this variable

$$E(t) = \sum_{k} \sum_{l} A_{kl} \cos \left[2\pi \left(\frac{kv}{w} + \frac{lu}{h} \right) t + \theta_{kl} \right] \tag{5.8}$$

where $vt = x$ and $ut = y$.

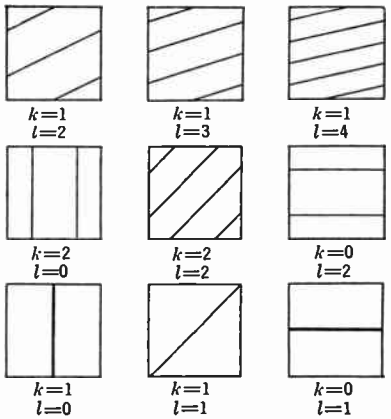


Fig. 5.17. Components of the Fourier Series Used in Picture Analysis.

However, there exists, from the nature of the scanning pattern, a relation between the horizontal velocity v and the vertical velocity u, namely:*

$$u = \frac{h}{w} \frac{v}{n}$$

and, furthermore,

$$v = wnN$$

* If the analysis here given is applied to interlaced scanning, N must be interpreted as the field frequency and n as the number of lines per field (e.g., 262.5).

Therefore, the frequencies f which appear in the argument of the cosine terms are

$$f = \frac{v}{w}\left(k + \frac{l}{n}\right) \tag{5.9}$$

$$= nN\left(k + \frac{l}{n}\right) \tag{5.9a}$$

Several interesting points may be noticed in this relation. When k is zero it represents a family of frequencies which are multiples of the frame frequency. All frequencies for which l is zero are multiples of the line frequency nN. The resulting spectrum may be thought of as groups of frequencies about the various multiples of line frequency, the members of each group being separated by multiples of frame frequency. Another feature of this relation, which will be considered at greater length presently, is that f can have the same value for several values of k, l.

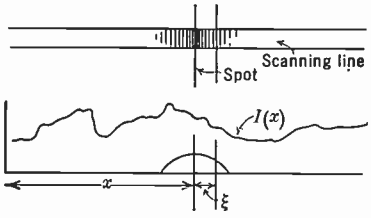


Fig. 5.18. Scanning with a Finite Aperture.

So far both the exploring element and the reproducing spot have been considered as points. Actually, of course, these spots have a finite size which is comparable with the resolution of the device. The effect of the size is a filtering action on the higher frequencies. Returning again to the one-dimensional series given in Eq. 5.6, the quantitative effect of a finite exploring aperture can readily be shown. Figure 5.18 represents the line being scanned, together with the finite aperture whose response, which varies over its length, is shown by the small rounded curve.

The picture intensity I along the scanned line is a function of the distance along the line. However, the amount of signal picked up from any point depends not only on the intensity I, but also upon the particular region of the spot that lies on that point. Let x be the coordinate locating the center of the exploring spot, and ξ that locating a position within the spot itself, relative to the center of the spot. For simplicity it will be assumed that the exploring element is symmetric about the center and its response is $S(\xi)$.

The intensity of illumination at point $(x + \xi)$ is*

$$I(x + \xi) = \sum_{k=-\infty}^{\infty} A_k \exp\left[\frac{2\pi ik(x + \xi)}{wn}\right] \tag{5.10}$$

whereas the signal from the total area of the exploring spot at point x is

$$E(x) = \int S(\xi) I(x + \xi)\, d\xi \tag{5.11}$$

Therefore, if Eq. 5.10 is substituted in Eq. 5.11, the signal becomes

$$E(x) = \sum_k \int S(\xi) A_k \exp\left(\frac{2\pi ik(x + \xi)}{wn}\right) d\xi$$

$$= \sum_k \int S(\xi) \exp\frac{2\pi ik\xi}{wn} A_k \exp\frac{2\pi ikx}{wn}\, d\xi \tag{5.12}$$

or

$$E(x) = \sum_k A_k Y(k) \exp\frac{2\pi ikx}{wn} \tag{5.12a}$$

where

$$Y(k) = \int S(\xi) \exp\left(\frac{2\pi ik\xi}{wn}\right) d\xi = \int S(\xi) \cos\left(\frac{2\pi k\xi}{wn}\right) d\xi \tag{5.13}$$

The last equation assumes that the spot sensitivity is symmetrical, $S(-\xi) = S(\xi)$.

The factor $Y(k)$, called the aperture admittance, appears in the series in the same form as electrical admittance, since it is a function of frequency only in the one-dimensional case. The function is illustrated in Fig. 5.19 for a uniformly sensitive round spot a, square spot b, and c, one in which the sensitivity is given by $S = e^{-\pi r^2/b^2}$.

* Although the cosine series is convenient for physical interpretation, some mathematical simplification is effected by the use of the exponential form of the Fourier series. Therefore the latter will be used, returning to the former only where practical interpretation makes it necessary. The relation between the forms can be seen from

$$2a \cos(x + \phi) = (a\epsilon^{i\phi})\epsilon^{ix} + (a\epsilon^{-i\phi})\epsilon^{-ix}$$

Therefore

$$\sum_{k=0}^{\infty} a_k \cos\left(\frac{2\pi k(x + \zeta)}{wn} + \phi_k\right)$$

becomes

$$\sum_{k=-\infty}^{k=\infty} A_k \exp\left(2\pi i \frac{k(x + \zeta)}{wn}\right)$$

where

$$A_k = (\tfrac{1}{2})a_k\epsilon^{+i\phi_k}; \qquad A_{-k} = (\tfrac{1}{2})a_k\epsilon^{-i\phi_k}$$

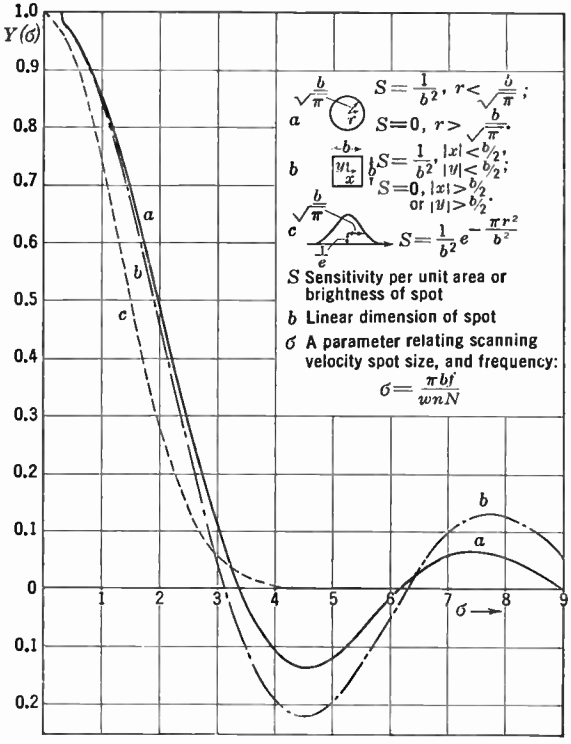


Fig. 5.19. Aperture Admittance of Various Scanning Spots.

For the more complete two-dimensional analysis resulting in Eq. 5.8, the aperture admittance becomes

$$Y(k, l) = \iint S(\xi, \eta) \exp 2\pi i \left(\frac{k\xi}{w} + \frac{l\eta}{h}\right) d\xi \, d\eta \tag{5.14}$$

and

$$E(x, y) = \sum_k \sum_l Y(k, l) A_{kl} \exp 2\pi i \left(\frac{kx}{w} + \frac{ly}{h}\right) \tag{5.15}$$

Thus as a function of time the signal is

$$E(t) = \sum_k \sum_l Y(k, l) A_{kl} \exp \left[2\pi i n N \left(k + \frac{l}{n}\right) t\right] \tag{5.16}$$

If the sensitivity of the exploring spot has symmetry about both ξ and η axes, $Y(k, l)$ becomes

$$Y(k, l) = \iint S(\xi, \eta) \cos 2\pi \left(\frac{k\xi}{w} + \frac{l\eta}{h}\right) d\xi \, d\eta \tag{5.17}$$

A very commonly occurring case is that in which the sensitivity can be broken into two factors which are each dependent on one variable only,

$$S(\xi, \eta) = S(\xi) \cdot S(\eta)$$

in which case the admittance becomes

$$Y(k, l) = Y(k) \cdot Y(l) \tag{5.18}$$

where

$$Y(j) = \int S(\delta) \cos 2\pi \left(\frac{j}{p}\right) \delta \, d\delta \tag{5.19}$$

The signal arriving at the receiving end of the system, neglecting for the moment the frequency response of the channel, is, of course, $E(t)$. This signal modulates the brightness of the reproducing spot which scans the viewing screen. The reproducing spot is finite in size, and not necessarily of uniform brightness over its area. Let $S'(\xi, \eta)$ be the function representing its brightness with respect to a point fixed on the spot, the primed symbols being used to designate properties of the receiving system.

The instantaneous brightness of a point on the screen will be

$$S'(\xi, \eta)E(t) = S'(x - vt, y - ut)E(t) \tag{5.20}$$

Let $B'(x, y)$ represent the brightness at a point x, y in the reproduced image; by the same reasoning as before, this may be expressed by a double Fourier series, as follows:

$$B'(x, y) = \sum_{k'} \sum_{l'} A'_{k'l'} \exp 2\pi i \left(\frac{k'x}{w} + \frac{l'y}{h}\right) \tag{5.21}$$

To compare the received picture with that transmitted, the Fourier terms of the two pictures may be compared. The coefficients $A'_{k'l'}$ of the image on the reproducer are

$$A'_{k'l'} = \frac{1}{wh} \int_{-w/2}^{+w/2} \int_{-h/2}^{+h/2} B'(x, y) \exp \left[-2\pi i \left(\frac{k'x}{w} + \frac{l'y}{h}\right)\right] dy \, dx \tag{5.22}$$

If Eq. 5.20 is used to express the instantaneous brightness, with the variables changed to the ξ, η coordinates and the order of integration with respect to t and η reversed as a matter of convenience, the coefficients become (taking due account of line overlap)

$$A'_{k'l'} = \frac{1}{hw}\int_{-\infty}^{+\infty}\int_{-\frac{h}{2u}-\frac{\eta}{u}}^{\frac{h}{2u}-\frac{\eta}{u}}\int_{-\infty}^{+\infty} S'(\xi, \eta)\exp\left[-2\pi i\left(\frac{k'\xi}{w}+\frac{l'\eta}{h}\right)\right]$$
$$\exp\left[-2\pi i\left(\frac{k'v}{w}+\frac{l'u}{h}\right)\right] tE(t)\,d\xi\,dt\,d\eta \quad (5.23)$$

Considering the integral in time, which is, omitting constant terms and substituting for $E(t)$,

$$\int \exp\left(2\pi i\left[\frac{(k-k')v}{w}+\frac{(l-l')u}{h}\right]t\right)dt$$

it is evident that the only terms contributing to the coefficients $A'_{k'l'}$ are those for which the relation

$$\frac{kv}{w}+\frac{lu}{h}=\frac{k'v}{w}+\frac{l'u}{h}$$
$$k+\frac{l}{n}=k'+\frac{l'}{n} \quad (5.24)$$

is fulfilled.

One condition satisfying Eq. 5.24 is

$$k = k'$$
$$l = l'$$

This gives rise to a direct correspondence between the coefficients in the transmitted picture and the received image. The amplitudes of the corresponding coefficients are related as follows:

$$A'_{kl} = A_{kl}Y(k, l)Y'(k, l) \quad (5.25)$$

where, for symmetrical spots,

$$Y'(k, l) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} S'(\xi, \eta)\cos\left[2\pi\left(\frac{k\xi}{w}+\frac{l\eta}{h}\right)\right]d\xi\,d\eta \quad (5.25a)$$

The factor $Y'(k, l)$ bears the same relation to the receiving spot as $Y(k, l)$ bears to the exploring element. In other words, it is the aperture admittance of the reproducing spot.

No account has been taken in Eq. 5.25 of the transmission characteristic of the communication channel. Since it is a function of frequency, it obviously has an effect on the relative amplitude of cor-

responding coefficients. The frequency response of the channel may be represented by the function $G(f)$, which, because of the relation between f and k, l, can be expressed as $G(nN[k + l/n])$. Therefore, including this in Eq. 5.25, the coefficients at the transmitted and received pictures are given by

$$A'_{kl} = A_{kl} G\left(nN\left[k + \frac{l}{n}\right]\right) Y(k, l) Y'(k, l)$$

From the standpoint of horizontal resolution the admittance may be treated as a function of the frequency of the signal components. The relation between the coefficients in the transmitted and reproduced picture then is

$$A'_f = A_f G(f) Y'(f) Y(f)$$

It will be noticed that the spot admittances have the same form as the channel response, and therefore, by making the latter proportional to the reciprocal of the spot admittances, the effect of the decreasing admittance with increasing frequency can be compensated. In practice this compensation is only partial. Compensation of this type can be used for the horizontal frequency components even with an unsymmetrical spot.

The condition $k = k'$, $l = l'$ is not the only one which satisfies Eq. 5.24. Therefore, some extraneous components will be present in the reproduced picture which do not exist in the transmitted image, as well as the desired corresponding Fourier components. These extraneous components will be

$$A'_{k'l'} = A_{kl} G\left(nN\left[k + \frac{l}{n}\right]\right) Y'(k', l') Y(k, l)$$

where

$$k + \frac{l}{n} = k' + \frac{l'}{n}, \quad \begin{cases} k \neq k' \\ l \neq l' \end{cases}$$

These extraneous components give the line structure of the reproduced image. Because of them the vertical resolution is restricted to a step process, and a bright point moving in a vertical direction moves in finite increments and may change in size periodically as it progresses. In addition, these components may cause one or more false patterns to accompany the normal pattern. This is very strikingly demonstrated in the transmitted reproduction of a zone plate shown in Fig.

5.20. The process of generating such false patterns is often graphically, but not quite accurately, referred to as beats between the scanning pattern and the image.

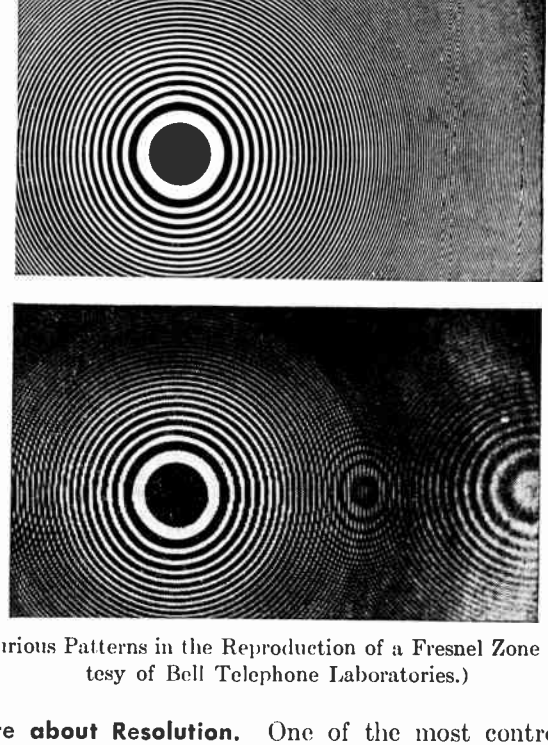


Fig. 5.20. Spurious Patterns in the Reproduction of a Fresnel Zone Plate. (Courtesy of Bell Telephone Laboratories.)

5.7 More about Resolution. One of the most controversial and difficult problems of television reproduction is correlating the subjective sharpness of a picture with the objective characteristics of resolution. This is required in order to obtain quantitative values for the vertical and horizontal resolution of an image which can be considered acceptable. Vertical and horizontal resolution in turn has, as has already been shown, a bearing on the number of lines forming the scanning raster and upon the frequency bandwidth of the transmission channel. In an earlier section it was shown that the structure of a 525-line picture is sufficiently fine so as to be unobtrusive when observed from a distance such that the image subtends an angle at the eye which is large enough to satisfy most viewers. This number of lines in the scanning raster has been adopted as standard for commercial television broadcasting in the United States. Experience of

ten or more years seems to indicate that it is satisfactory to the television audience.

The maximum vertical resolution is obtained when the scanning apertures at the transmitter and reproducer are made very small. However, such a procedure does not give an optimum both because it makes the horizontal line structure unduly prominent and because it enhances spurious images due to the scanning pattern. The existence of the spurious images resulting from scanning was indicated in the preceding section by the existence of extraneous Fourier components shown to be present in the received picture. The effect is particularly objectionable if the spot size is smaller than the line separation. For example, if the exploring spot at the transmitter is less than the vertical line separation, the reproduction of a bright line nearly parallel to the scanning lines will be a row of separate bright beads clearly visible even at distances for which the line structure cannot be resolved. This and other spurious effects due to line structure can be minimized by making the distribution of the spot such that the response and brightness at the transmitting and receiving ends respectively are as nearly as possible uniform over the entire field. A field meeting this requirement will be termed a "flat field." Rectangular spots having uniform distribution whose heights are equal to the line separation will, of course, give a flat field.

In most cathode-ray systems, the spot has radial symmetry and a distribution such that the brightness or maximum response is in the center. To obtain a flat field with such a spot, it is necessary to overlap the scanning lines.

The actual distribution of brightness at the receiver and response at the transmitter across a scanning line approximates an error function. It can also be represented by a cosine squared distribution with a fair degree of accuracy, this distribution having certain advantages in mathematical manipulation. The minimum size of a spot giving this distribution which will yield a flat field requires a 50 percent overlap. The distribution for this type of spot is illustrated in Fig. 5.21.

The next step in analyzing the definition which can be obtained with the system is to determine what vertical resolution can be obtained in a picture formed with a 525-line pattern and a cosine squared spot having a 50 percent overlap. This vertical resolution will then be compared with the horizontal resolution that can be obtained when the frequency band is limited to the presently accepted standard channel width. Finally, the subjective quality of sharpness of such a picture

will be discussed, in order to obtain a quantitative orientation of the merits of the picture which can be transmitted within the limits imposed by the present standard television signal.

The difficulty in evaluating the resolution of a television picture in the two directions lies, of course, in the fact that the picture is essentially anisotropic in structure. It is, therefore, necessary to select some arbitrary structural feature of a generalized image as the basis of comparison. Of the several methods of evaluating resolution, two will

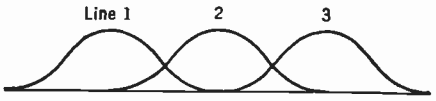


Fig. 5.21. Overlapping Cosine-Squared Spot Which Yields a Flat Field.

be discussed. One of these, due to Wheeler and Loughren,* makes use of the "width of confusion." It is defined as the width of a uniform bright bar whose brightness is equal to the peak brightness of the reproduction of a very narrow horizontal line of light and whose total flux is equal to that of the reproduction. Figure 5.22 is given to further clarify the meaning of the conception "width of confusion." The narrow rectangular peak a is the distribution of light in the original image, brightness being plotted vertically. The broad, rounded curve b

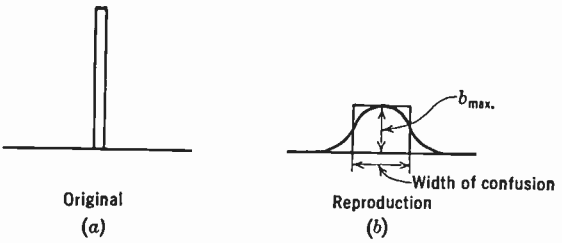


Fig. 5.22. Width of Confusion.

shows the brightness variation across the reproduced line. Superimposed on curve b is shown the rectangle having equal height and area. The width of this rectangle is the width of confusion of the reproduction.

When the narrow line used as the test original is exactly parallel to the scanning line, the width of the reproduction will vary as the original is moved between successive lines. By placing the original line at a small angle with respect to the scanning lines, the average width of

* See Wheeler and Loughren, reference 6.

confusion will be independent of the position of the test line. The width of confusion will, of course, vary along the reproduced line and must be averaged in determining the vertical resolution.

If the brightness across the lines which make up the scanning raster at the reproducer varies as the cosine squared and there is a similar variation in sensitivity across the exploring lines of the pickup device, the average vertical width of confusion can be shown to be 1.414 times the separation between lines for the 50 percent overlap required for a flat field.

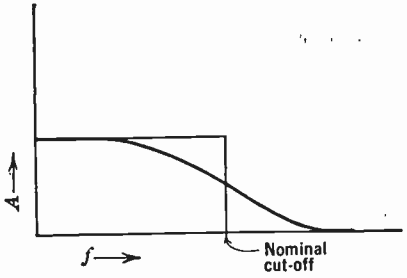


Fig. 5.23. Attenuation of the High Frequencies by the Communication Channel.

To apply the concept of the width of confusion in the horizontal direction, it is necessary to consider the frequency spectrum of the original image and the way it is altered by the aperture defects and channel limitations. The frequency spectrum of a sharp pulse is essentially constant at all frequencies. The upper frequencies will be attenuated by the admittance of the two exploring spots involved and by the channel response so that the reproduced pulse will be the result of band frequencies whose amplitudes decrease gradually to zero with increasing frequency, as shown in Fig. 5.23.

The width of confusion can be shown by the theory of Fourier transforms to be equal to a half period of the nominal cutoff frequency of the passband. The nominal cutoff frequency is defined as the upper frequency limit of a band having constant amplitude equal to the low-frequency amplitude of the passband and an area which is the same as that of the passband. The nominal cutoff frequency is indicated for the bandpass characteristics of the entire channel in Fig. 5.23.

Since the vertical width of confusion is 1.414 times the vertical line separation, the nominal cutoff required to give an equal horizontal width of confusion must be

$$f_c = \frac{1}{2\sqrt{2}} \frac{w}{h} n^2 N$$

$$f_c = 3.90 \cdot 10^6 \text{ cycles}$$

for a 525-line picture with a repetition rate of 30 frames per second.

Under the present standards adopted for television transmission, the nominal cutoff frequency of the communication channel is only slightly more than 3.5 megacycles. Further attenuation is caused by the two scanning spots still further reducing the effective nominal cutoff of the overall system. Therefore, on the basis of the foregoing described criteria, the horizontal resolution will be somewhat less than the vertical resolution.

A second criterion for comparing vertical and horizontal resolutions has been proposed by Kell, Bedford, and Fredendall.* This method uses the width of blurring of the reproduction of an abrupt transition from light to darkness. This unit step is used as a test element in preference to the checkerboard pattern referred to earlier in this chapter or the line of light used by Wheeler and Loughren because such unit steps may be thought of as the building blocks out of which real pictures may be made. Thus the unit brightness step is considered quite analogous to the Heaviside unit function used in electric circuit theory. In estimating the vertical resolution, the light distribution across the line is considered to vary with the cosine squared and, furthermore, the overlap is taken as 50 percent to give a flat field.

The test pattern is placed on the pickup device in such a way that the white-to-black junction is nearly but not quite parallel to the scanning lines. The vertical distribution of light intensity across the junction varies with position along the junction because of the changing relationship between the scanning lines and the test edge. Figure 5.24 shows the intensity distribution for four representative positions along the test edge. It is assumed that the eye will see an arithmetic average of the transition. Curve v is such an arithmetic average. This average curve is a measure of the vertical resolution and is to be compared with the horizontal resolution measured in a comparable way. To determine the horizontal transition curve, the junction of the brightness step is placed vertically on the sensitive screen of the pickup device. The sharpness which can be obtained in the distribution along a horizontal line is determined by considering the attenuation of the high-frequency components in the original light pattern by the frequency characteristics of the two scanning spots and the passband of the communication channel.

The conclusions that can be reached by such a calculation are summarized in Fig. 5.25. Here, the ordinate gives the horizontal and vertical resolutions in arbitrary units, and the abscissa gives the

* See Kell, Bedford, and Fredendall, reference 10.

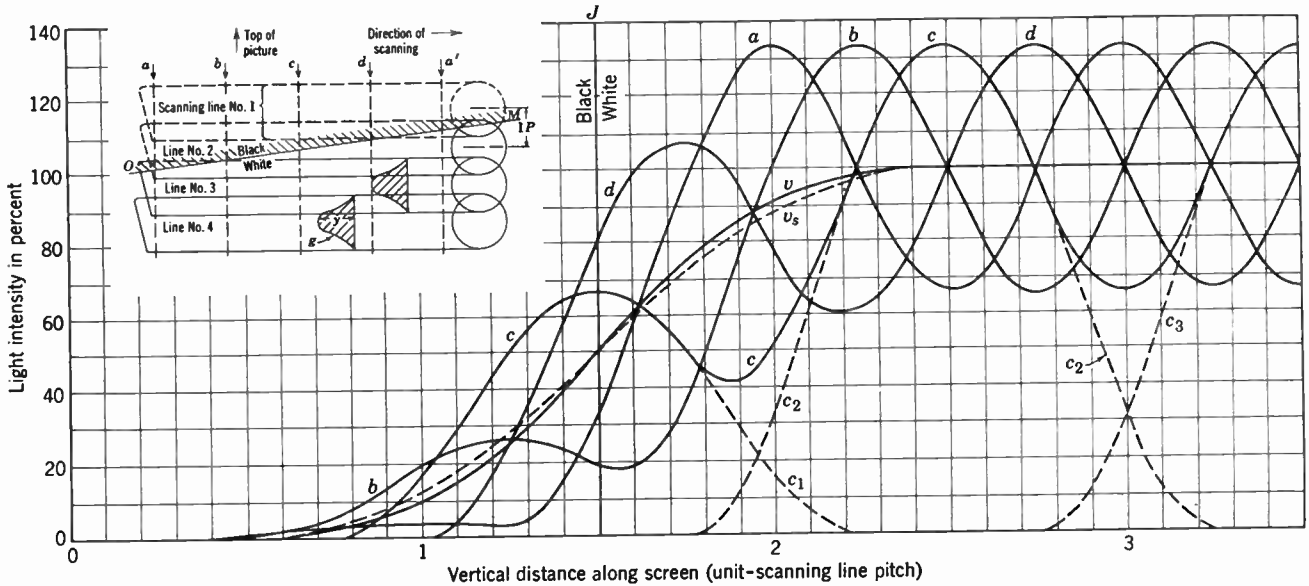


Fig. 5.24. Intensity Distribution across the Reproduction of a Horizontal Edge in Relation to the Center of a Scanning Line of the Pickup System. Curve (v) represents an arithmetic average of the distribution. (Kell, Bedford, and Fredendall, reference 10.)

number of scanning lines. The three horizontal resolution curves are for the three different channel frequency response shapes indicated in the inset drawing, all having a total bandwidth of 4½ megacycles. For the horizontal resolution curve $H_1(m = 1)$, the point of equal horizontal and vertical resolution is very near the 525 lines of the

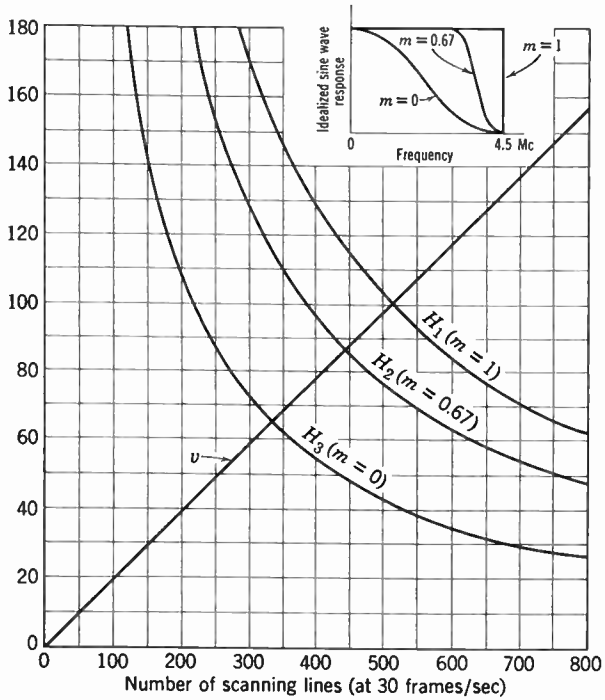


Fig. 5.25. Relative Values of Vertical and Horizontal Resolution as Function of Number of Scanning Lines for Three Different Frequency Responses. (Kell, Bedford, and Fredendall, reference 10.)

present television standards. However, when a more reasonable channel response shape is assumed (e.g., curve H_2), the conclusion is that the horizontal resolution is only about 75 percent of the vertical resolution.

The term "resolution" as employed in the preceding paragraphs relates to an objective property of the reproduced image. It would be of value to obtain a measure of the subjective effect of this property. The term "sharpness" of the picture will be used to designate the subjective or psychological effect of the objective property of high resolution. It is very difficult to obtain a quantitative correlation

between sharpness and resolution. Indeed, it is even difficult to specify what is meant by such a correlation. A very interesting treatment of this question has been presented by M. W. Baldwin * of Bell Telephone Laboratories. In this treatment not only is the quantitative method of evaluating the sharpness of a picture worked out but also a measure is obtained of the subjective effect of unequal vertical and horizontal resolution.

The unit of measure which was employed to compare the sharpness of two pictures was termed the "liminal unit." It was obtained in the following way.

Two pictures of slightly different objective resolution were shown to a large number of observers, and they were asked to designate which appeared to be the sharper. The difference in resolution was small enough so that there was not unanimous agreement as to which was the sharper. The two pictures were said to differ by one liminal unit if 25 percent of the observers guessed wrong and 75 percent right. The assumption here is that if 25 percent of the observers made an incorrect choice, then another 25 percent made a correct choice simply on the basis of a guess rather than real discrimination. Thus, two images which differ in sharpness by one liminal unit have a resolution difference which is only perceptible to 50 percent of the observers. By comparing a large number of pictures of different resolution, a liminal scale of sharpness can be established. Having established a quantitative measure of sharpness, the next step was to obtain a curve giving the sharpness of a picture in terms of the number of picture elements (figures of confusion) which it includes. The picture had an aspect ratio of 4 to 3 and was viewed from a distance of 30 inches, which was four times its vertical height. The curve obtained is shown in Fig. 5.26. This curve is based on more than a thousand observations. It is important to note that the subjective sharpness rises rapidly at first and then flattens out so that a point of diminishing returns is soon reached as the number of picture elements is increased.

The total number of picture elements is directly proportional to the effective frequency band over which the picture can be transmitted. This frequency band is indicated on a separate scale in Fig. 5.26. The position of the 525-line picture is also indicated. It will be quite evident that a very large increase in frequency band would be required to give a significant increase in subjective sharpness. Using the same type of criteria, the effect of differences in horizontal and vertical resolution was measured. The results are shown in Fig. 5.27. In each

* See Baldwin, reference 13.

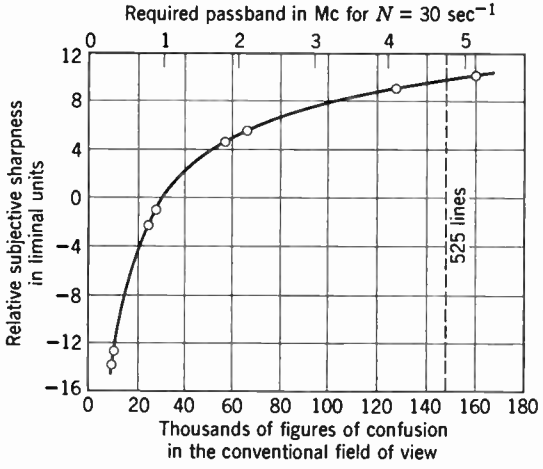


Fig. 5.26. "Sharpness" as Function of Total Number of Picture Elements. (Baldwin, reference 13; courtesy of *Proceedings of the Institute of Radio Engineers* and *Bell System Technical Journal*.)

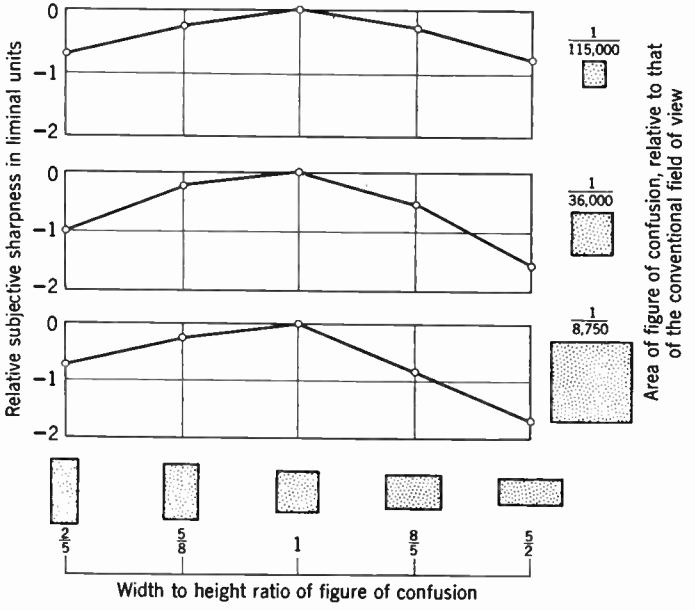


Fig. 5.27. Sharpness as Function of Ratio of Horizontal to Vertical Resolution for Equal Total Numbers of Picture Elements. (Baldwin, reference 13; courtesy of *Proceedings of the Institute of Radio Engineers* and *Bell System Technical Journal*.)

of the three curves shown, the area of the circle of confusion was constant over the entire area, but the ratio of its height to width was changed in the manner indicated at the bottom of the figure. It will be evident from this family of curves that as long as the total number of picture elements making up the picture remains constant, the shape of the picture element makes very little difference. In other words, if the product of the vertical and horizontal resolutions is constant, the fact that they are not equal is inconsequential.

Although the preceding measurement gives meaning to the concept of sharpness of a picture in terms of angular resolution, it still leaves open the question of the total amount of detail required in a picture. Laboratory measurements of the total number of picture elements required are difficult to make because of the impossibility of duplicating the conditions, both physical and psychological, under which the average broadcast receiver is used. Some information about this point can be obtained from the last ten years of television broadcast experience.

When television broadcasting was started under present standards, it was not known what direction demands for improvement would take. Many workers in the field felt that the demand would be for pictures having a larger number of picture elements, even if this necessitated a wider transmission channel. However, the demand has not been along these lines, but instead has been for larger and larger viewing areas without very much demand for a larger information content in the picture. In fact, the public reaction to picture quality is curiously reminiscent of its reaction toward high-fidelity sound. Possibly some as yet unformulated psychological law is involved here which states that there is an upper limit to the rate at which any recreational instrument should supply information to the observer's brain. However that may be, it appears for the present that the amount of information supplied by a 525-line picture and a 4½-megacycle transmission band is adequate.

REFERENCES

1. J. W. T. Walsh, *Photometry,* Constable, London, 1926.
2. J. C. Wilson, *Television Engineering,* Pitman, London, 1937.
3. R. V. L. Hartley, "Transmission of Information," *Bell System Tech. J.*, Vol. 7, pp. 535–563, 1928.
4. E. W. Engstrom, "A Study of Television Image Characteristics," Part I, *Proc. I.R.E.,* Vol. 21, pp. 1631–1651, December, 1933 and Part II, *Proc. I.R.E.,* Vol. 23, pp. 295–310, 1935.
5. Pierre Mertz and Frank Gray, "Theory of Scanning," *Bell System Tech. J.,* Vol. 13, pp. 464–515, 1934.

6. H. A. Wheeler and A. V. Loughren, "The Line Structure of Television Images," *Proc. I.R.E.*, Vol. 26, pp. 540–575, 1938.
7. A. V. Bedford, "A Figure of Merit for Television Performance," *RCA Rev.*, Vol. 3, pp. 36–44, 1938.
8. S. W. Seeley, "Effect of the Receiving Antenna on Television Reception Fidelity," *RCA Rev.*, Vol. 2, pp. 433–441, 1938.
9. I. G. Maloff, "Gamma and Range in Television," *RCA Rev.*, Vol. 3, pp. 409–417, 1939.
10. R. D. Kell, A. V. Bedford, and G. L. Fredendall, "A Determination of Optimum Number of Lines in a Television System," *RCA Rev.*, Vol. 5, pp. 8–30, 1940.
11. R. R. Law, "Contrast in Kinescopes," *Proc. I.R.E.*, Vol. 27, pp. 511–524, 1939.
12. J. C. Wilson, "Channel Width and Resolving Power in Television Systems," *J. Television Soc.*, Series 2, Vol. 2, pp. 397–419, 1938.
13. M. W. Baldwin, "The Subjective Sharpness of Simulated Television Images," *Proc. I.R.E.*, Vol. 28, pp. 458–468, 1940, and *Bell System Tech. J.*, Vol. 19, pp. 563–586, 1940.
14. C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Tech. J.*, Vol. 27, pp. 379–423 and 623–656, 1948.

6 Transmission and Reproduction of High-Definition Pictures

The preceding chapter put forward certain requirements that must be met by an idealized skeleton television system in order to produce a high-definition picture. These requirements are not out of keeping with what can be physically realized with the resources available. The purpose here is to outline the actual embodiment of a high-definition

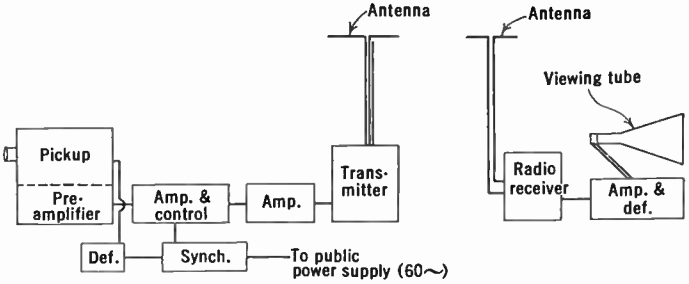


Fig. 6.1. Elements of a Television System.

system, or more exactly, high-definition systems in general, for the discussion will center chiefly around those elements which are common to most systems.

Briefly, the system may be subdivided into the following elements: (1) Pickup device. (2) Video amplifier. (3) Radio transmitter. (4) Radio receiver. (5) Receiver amplifier. (6) Viewing device. (7) Synchronizing equipment. Schematically they can be represented as shown in Fig. 6.1. Because the system functions as an entity, this subdivision is not without its disadvantages. However, for the sake of simplicity and clarity it has been adopted.

6.1 The Pickup Device. The pickup device is the element in the system which converts the light image into a train of electrical impulses making up the video signal. As has already been explained, this involves scanning the picture area with a suitable exploring element. There are a number of ways in which the conversion of a light image

into the video signal may be accomplished. The means for this conversion may be either electronic or mechanical.

The electronic pickup devices may be divided into two categories: storage systems and nonstorage systems. In the former, exemplified by the iconoscope, the orthicon, the image orthicon, and the Vidicon, a charge image, which is a replica of the optical picture of the scene being televised, is stored by the photoelectric effect on a target. The charge is removed point by point from the mosaic by a fine pencil of electrons which scans the stored image. The released charge forms the picture signal. Nonstorage systems, such as the dissector tube and the flying-spot system, make use of a quite different principle. In the dissector an electron image is formed by focusing the electrons from a photocathode on which the picture being televised is projected. This electron image is moved across a small opening so that it is scanned by this aperture. The current entering the aperture is amplified by a secondary-emission multiplier and constitutes the video signal. In the flying-spot system, employed primarily for the transmission of motion pictures, the scanning spot tracing a pattern on the fluorescent screen of a cathode-ray tube is imaged on the object; the light of the scanning spot which is reflected or transmitted by the object falls on a multiplier phototube which generates the picture signal. It is seen that in all the above systems the scanning depends on the deflection of an electron stream. The deflection is accomplished by means of suitably directed and varied electrostatic or magnetic fields. The voltage or current producing the fields is generated by a deflection generator, which in turn is controlled by synchronizing equipment.

In general, mechanical pickup devices do not lend themselves to high-definition television work. They have retained their usefulness longest in connection with the transmission of moving pictures. For this a Nipkow disk, preferably running in vacuum, rotates in the plane of the image, apertures in the disk moving across the image to produce the scanning pattern. Behind the disk is the light-sensitive element, usually in the form of a multiplier phototube, which converts the light passing through the apertures into the video signal. The rotation of the disk is controlled by a signal from the synchronizing equipment.

These pickup devices will be described in greater detail in the next chapter; the above brief account merely indicates the role this equipment plays in the system as a whole.

6.2 The Amplifiers. The signal output from any one of the foregoing pickup devices usually represents only a minute amount of power—commonly not more than a few thousandths of a microwatt

delivered at a voltage level of the order of a few millivolts. Before this signal can be used to modulate a broadcast transmitter of even moderate power, it must be amplified to a level of several hundred watts.* Hence the total power gain of the chain of amplifiers raising the signal level ranges from 10^{10} to 10^{12}. This total gain is attained with 15 to 20 stages of amplification, with an average effective power gain per stage between 4 and 5. The average gain per stage is low since a considerable portion of the amplifier gain is canceled by attenuation losses in coaxial cables, losses incident in controlling and shaping the signal, etc.

The video amplifier, like the amplifiers employed for audio amplification, makes use of thermionic vacuum tubes. However, because of the much more exacting requirements imposed by the amplification of the video signal, the means used to cascade the tubes are somewhat different. A discussion of the design of this type of amplifier will be deferred until Chapter 13, and only the general purpose and nature of the video amplifier will be considered in this section.

The frequency band required to give the desired resolution has been shown to be approximately 4.5 megacycles wide. Over this range of frequencies the amplifiers must have uniform gain and constant time delay. An exception to the requirement of uniform gain occurs when it is necessary to correct for the aperture defect of the scanning spots; in this case the gain at high frequencies is made greater than that at low frequencies in order to compensate for the reduction in amplitude of the high-frequency components which results from the finite spot size.

The signal which is generated by the pickup device does not contain the complete information necessary for the reconstruction of the picture. In addition to voltage variations representing the light and shade in the picture, it is necessary to furnish the information required to insure synchronism between the scanning pattern at the transmitter and the receiver. This is supplied in the form of properly shaped pulses, timed so that they occur after the end of each scanning line and before the beginning of the next. Similarly, a pulse or series of pulses is added between frames or field repetitions to insure vertical synchronization.

The synchronizing impulses are added to the picture signal in the amplifier chain. This operation is performed by rendering the ampli-

* In some television transmitters the carrier is modulated at a considerably lower level than this. Low-level modulation has, however, material drawbacks for the maintenance of transmitter adjustment.

fier inactive, ahead of the point of injection, for a period equal to or slightly in excess of that occupied by the actual synchronizing signal, and injecting the impulse during this interval. The purpose of this "blanking" process is to prevent any possible signal from the pickup device from interfering with the synchronizing signal. During the blanking period the voltage level in the amplifier is brought to a value which corresponds to maximum white or, what is more usual, maximum black of the picture, thus providing a reference point from which the

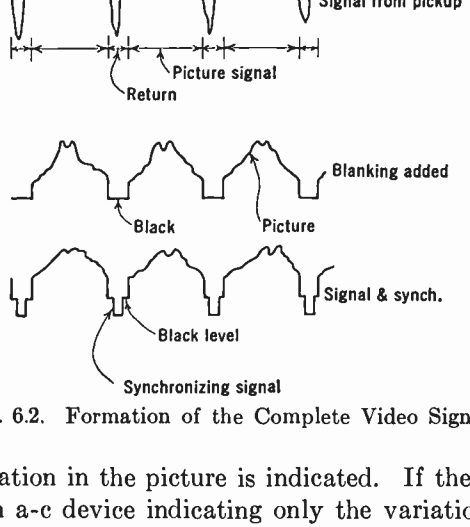


Fig. 6.2. Formation of the Complete Video Signal.

average illumination in the picture is indicated. If the pickup device is essentially an a-c device indicating only the variation in light and dark over the picture, but not giving information as to the average illumination, as, e.g., an iconoscope, a special phototube and d-c amplifier may be necessary to establish this value. The various steps in the formation of the complete signal are indicated in Fig. 6.2.

For convenience, the amplifier at the transmitter may be considered to consist of three parts. At the pickup device there is a preamplifier which receives the very weak signal, which may range from a fraction of a millivolt to about 200 millivolts, and raises it to a level of 0.1 to 1 or 2 volts. This amplifier is made to have as low a noise level as possible, since the input at this point is commonly not greatly in excess of the statistical fluctuations in the necessary coupling resistors and vacuum tubes. As will be shown later, in order to obtain a maximum signal-to-noise ratio, it may be necessary to make the gain of the first few stages as high as possible. This brings in the need of rather com-

plicated coupling arrangements to maintain constancy of frequency response.

The output of the preamplifier is sufficient to drive a cable (usually a low-loss coaxial) which connects it with a second amplifier chain, variously called the monitoring amplifier or the control amplifier. This unit raises the signal to a level of 1 to 10 volts, and it is in this amplifier that the synchronizing signal is added. Also the gain, the d-c level, etc., are controlled in this amplifier, and any special signals which may be necessary are added.

The output of this amplifier is fed to the modulation amplifier which raises the signal level to that required by the modulator of the radio transmitter.

This subdivision of the amplifier chain into three, and only three, physical units is, of course, a simplification which does not necessarily exist in practical transmitters; at times the physical layout makes more than three units advisable. Functionally these three units can generally be identified.

6.3 The Video Transmitter. The design problem presented by the radio transmitter to be used for television broadcasting entails many features not encountered in other fields. This is a consequence of the tremendous bandwidth of frequencies which must be radiated if high definition is to be attained.

Like a sound transmitter, the transmitter in question is made up of a carrier generator which produces a high-frequency oscillation of constant amplitude, and a modulator which varies the amplitude of this radio-frequency carrier in such a way that it is at every instant proportional to the signal applied to the modulator, in this case the video signal. The modulated radio frequency carrier is then fed to the transmitting antenna, from which it is radiated. This is shown schematically in Fig. 6.3.

Obviously, since the amplitude of the radio-frequency output of the modulator is not constant, the signal no longer consists of a single frequency as does the unmodulated carrier. The existence of these several frequencies can be demonstrated by a brief consideration of the process of modulation. For this purpose let the carrier frequency be f and the modulating frequency p. A factor must also be included which represents the amount by which the amplitude of the carrier is varied, and which is proportional to the amplitude of the modulating frequency. This factor k, when expressed as a percentage, is known as the percent modulation. It varies from zero, when the amplitude of the modulating frequency is zero, to unity (or 100 percent) when the amplitude

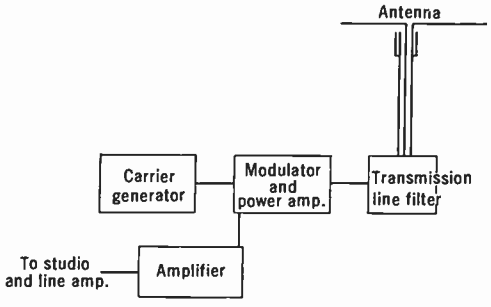


Fig. 6.3. Block Diagram of a Television Transmitter.

is equal to that of the carrier. In terms of these symbols the modulated signal is

$$A(1 + k \cos 2\pi pt) \cos 2\pi ft$$

which, by the rules of trigonometry, can be transformed to

$$A \left\{ \cos 2\pi ft + \frac{k}{2} \cos 2\pi (f + p)t + \frac{k}{2} \cos 2\pi (f - p)t \right\} \qquad (6.1)$$

As was shown in the preceding chapter, the video frequency p may have any value from zero (approximately) to 4.5 megacycles. Therefore, the band of frequencies which must be transmitted is $2p_{\max}$ cycles wide—i.e., 9 megacycles wide—with its center at the carrier frequency. These additional frequencies above and below the carrier are known as the upper and lower sidebands, respectively.

In order to accommodate this wide band of frequencies the carrier frequency itself must be very high, i.e., in the very-high or ultra-high-frequency region, if more than one broadcasting station is to function at any one time. Actually, frequencies in the range of 54 to 88 and 174 to 216 megacycles, as well as between 470 and 890 megacycles, are used for television broadcasting, whereas frequencies near 4000 and 6000 megacycles are used for television relay purposes. The channels allotted by the Federal Communications Commission (FCC) to television broadcasting (1953) are indicated in Fig. 6.4. Certain inherent difficulties, which militate against the use of much higher frequencies for television broadcasting, will be reviewed in Chapter 15.

It will be noticed in Fig. 6.4 that the channel width assigned for each television transmitter is 6 megacycles. In view of the 9 megacycles apparently demanded by Eq. 6.1, it would seem that this would seriously limit the definition of the pictures which could be broadcast. Actually this is not so, because it can be shown that all the information

in the video signal is present if the frequencies transmitted include the carrier and all frequencies above the carrier, that is, a band from f to $f + p_{max}$. Similarly, if the carrier and all frequencies below it are transmitted, the signal contains the full information. Therefore, the bandwidth transmitted needs to be only slightly greater than the 4.5 megacycles. For practical reasons connected with the means adopted to eliminate the unwanted sideband, the frequencies of this sideband which are close to the carrier are included in the signal. Furthermore, the sound which supplements the picture is also transmitted in this channel, albeit with a separate carrier. These signals fill the 6-megacycle band, as shown in Fig. 6.5.

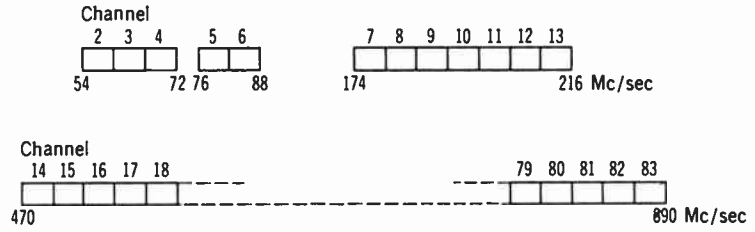


Fig. 6.4. Portion of Frequency Spectrum Allocated to Television Broadcasting.

The propagation characteristics of the ultra-short-wave signal are somewhat different from those of broadcast or ordinary short-wave signals. This difference arises from the fact that this radiation has much more in common with light than the lower frequencies have. The useful range of ultra-high-frequency transmission is essentially determined by the horizon; in other words, the radiation does not follow the curvature of the earth's surface. Because of this the service area of any one transmitter is limited to a radius of 25 to 50 miles for practical transmitting antenna heights. This is not entirely without its advantages, inasmuch as it avoids all difficulties from reflections at the various ion layers in the upper atmosphere which are responsible for the fading encountered at the longer wavelengths. If the properties of very-high-frequency radiation resembled those of the lower frequencies, fading would be so severe because of the great bandwidth that the quality of pictures coming from distant points would be seriously impaired. Furthermore, these same unusable distant signals would interfere with local reception.

Owing to the quasi-optical properties of the radiation, it can be very easily directed by antenna arrangement. The signal can, therefore, be placed where it is of most value; for example, it can, by means

of a suitable transmitting antenna, be confined to a horizontal plane, so that the energy is not wasted in the vertical direction where it cannot be used. Similarly, the receiving antenna can be made directive and consequently insensitive to radiation coming from any direction other than that of a particular transmitter.

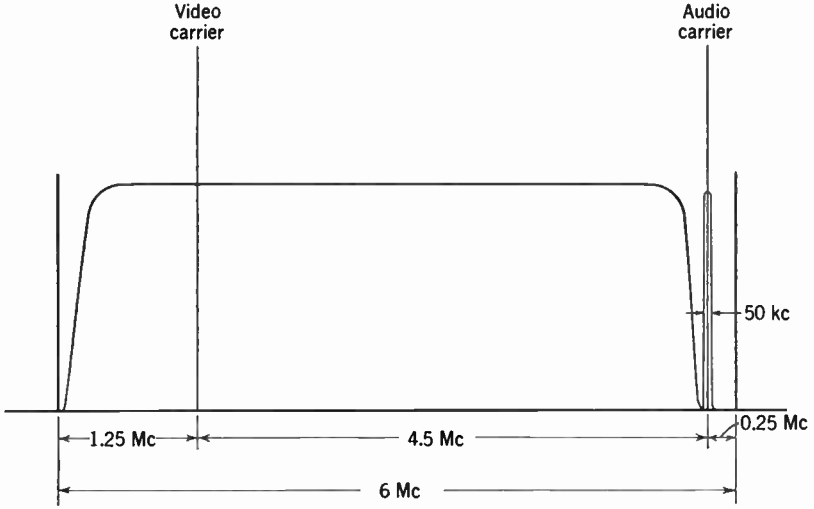


Fig. 6.5. Utilization of the 6-Megacycle Channel.

6.4 Single-Sideband Transmission and Reception. It was stated, without proof, that a carrier and all the sideband frequencies either above and below the carrier carry the full information contained in the video signal. Since there is a one-to-one correspondence between the video frequencies and the different frequencies of the sideband and carrier, and also a proportionality between the sideband amplitudes and video-frequency amplitudes, this statement does not seem unreasonable.

It is not immediately obvious that the video signal can be obtained from carrier and single sideband by the ordinary processes of detection. Therefore, the next consideration will be that of the action of a detector on the single-sideband signal. As has already been shown, a carrier f_0 modulated with a single frequency p leads to a signal having the following form:

$$e = A\left(\cos 2\pi f_0 t + \frac{k}{2}\cos 2\pi(f_0 + p)t + \frac{k}{2}\cos 2\pi(f_0 - p)t\right) \quad (6.2)$$

At the receiver, since the entering signal passes through various band-pass filters, there is an attenuation which is not independent of frequency. The same circuit elements introduce phase shifts which are a function of frequency. If the amplitude factor is $B(f)$ and the phase shift $\phi(t)$, the signal reaching the detector is

$$e = A\left[B(f_0)\cos\left[2\pi f_0 t + \phi(f_0)\right] + \frac{kB(f_0+p)}{2}\cos\left[2\pi(f_0+p)t + \phi(f_0+p)\right] + \frac{kB(f_0-p)}{2}\cos\left[2\pi(f_0-p)t + \phi(f_0-p)\right]\right] \quad (6.3)$$

When the signal is rectified by a linear detector and the output put through a low-pass filter to eliminate frequencies much above the maximum video frequency $p_{\max}$, the resulting signal is the envelope of this expression. If Eq. 6.3 is written in the form

$$e = V_s \cos\left(2\pi f_0 t + \phi^*\right)$$

the factor V_s is for all practical purposes the envelope.*

Therefore, the detector output has the form

$$V_s = A\left[B^2(f_0) + \frac{k^2}{4}\left\{B^2(f_0+p) + B^2(f_0-p) + 2B(f_0+p)B(f_0-p)\cos\left[4\pi pt + \phi(f_0+p) - \phi(f_0-p)\right]\right\} + kB(f_0)\left\{B(f_0+p)\cos\left[2\pi pt + \phi(f_0+p) - \phi(f_0)\right] + B(f_0-p)\cos\left[2\pi pt + \phi(f_0) - \phi(f_0-p)\right]\right\}\right]^{1/2} \quad (6.4)$$

If, as they should be in a well-designed filter system for double-sideband operation, the following conditions are fulfilled:

$$B(f_0+p) = B(f_0-p) = B(p)$$
$$\phi(f_0+p) - \phi(f_0) = \phi(f_0) - \phi(f_0-p) = \theta(p)$$

Eq. 6.4 reduces to

$$V_s = A\left[B^2(f_0) + \frac{k^2B^2(p)}{2}\left\{1 + \cos\left(4\pi pt + 2\theta\right)\right\} + 2kB(p)B(f_0)\cos\left(2\pi pt + \theta\right)\right]^{1/2} \quad (6.5)$$

* The phase shift ϕ^* is here chosen so as to eliminate time dependence with the carrier frequency from the coefficient V_s. It is readily seen that the time deriva-

or

$$V_s = A[B(f_0) + kB(p) \cos (2\pi pt + \theta)] \tag{6.5a}$$

This equation shows that no distortion is introduced by the detector.

Single-sideband operation can be represented with Eq. 6.4 by placing either one of the sideband attenuation factors $B(f_0 + p)$ or $B(f_0 - p)$ equal to zero.* As far as the calculation is concerned no distinction is necessary between the suppression of the sideband at the transmitter or at the receiver.

Equation 6.4 becomes, with $B(f_0 - p) = 0$:

$$V_s = A\left[B^2(f_0) + \frac{k^2B^2(f_0 + p)}{4} + kB(f_0)B(f_0 + p) \cos [2\pi pt + \phi(f_0 + p) - \phi(f_0)]\right]^{1/2} \tag{6.6}$$

It is immediately evident that if the modulation factor k is large, this equation represents a distorted output. If k is small, the equation can be expanded by the binominal theorem, giving

$$V_s = A\left[B(f_0) + \frac{kB(f_0 + p)}{2} \cos (2\pi pt + \phi(f_0 + p) - \phi(f_0))\right]$$

$$= A\left[B(f_0) + \frac{kB(p)}{2} \cos (2\pi pt + \theta)\right] \tag{6.7}$$

using the notation of Eq. 6.5*a*. This equation has the same form as Eq. 6.5*a* and indicates that the detector introduces no distortion.

The fidelity of response can be seen to depend upon the attenuation characteristic of the filter in both Eqs. 6.5*a* and 6.7. In order to have a response independent of the video frequency p, the attenuation factor $B(p)$ must be constant. Furthermore, unless the phase factor θ is proportional to p, the time delay will depend upon frequency, with consequent distortion of the reproduced picture.

If one sideband, for example, the lower, is only partially suppressed, that is, if frequencies from $f_0 + p_{max}$ to $f_0 - p_1$ are received at the detector and those from $f_0 - p_1$ to $f_0 - p_{max}$ are rejected, the resulting video signal will have twice the amplitude over the range of 0 to p_1 as from p_1 to p_{max}. This is illustrated in Fig. 6.6*b*.

tives of e and V_s are identical at the points of contact of the two curves [$e = V_s$, $\cos (2\pi f_0 t + \phi^*) = 1$], so that V_s is indeed the envelope of e.

* See Poch and Epstein, reference 8.

On the other hand, if the band includes the upper sideband from $f_0 + p_{max}$ to $f_0 + p_1$ uniformly, then tapers linearly to zero between $f_0 + p_1$ to $f_0 - p_1$, as shown in Fig. 6.6*c*, the output will be undistorted, at least for a low percentage modulation.

As is apparent from Eq. 6.6, truly undistorted linear detection is possible only when the percentage modulation is small. For a modulation fraction which is not negligible compared with unity, higher-order terms in the expansion of Eq. 6.6 cannot be dropped. Under

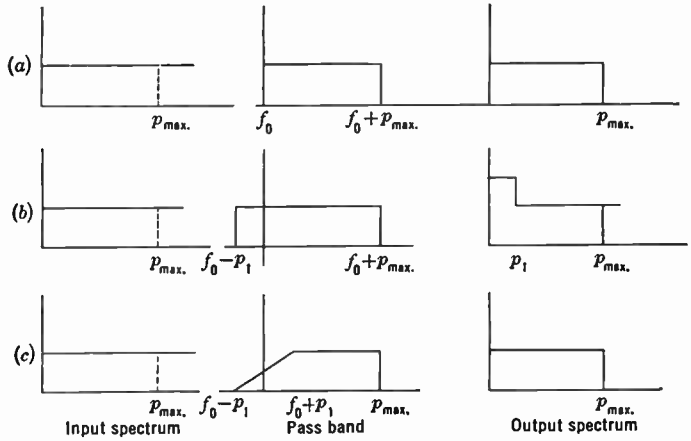


Fig. 6.6. Effect of Attenuating a Sideband on the Character of the Transmitted Frequency Spectrum.

these conditions the detector output consists not only of the original video frequency but also of harmonics of it. At 100 percent modulation the second harmonic is about 12 percent and the third harmonic about 3 percent. In audio work this amount of distortion would be very serious. However, the reconstructed picture is very insensitive to distortion of this nature, so that the presence of this amount does no visible harm.

So far the discussion has been limited to the behavior of the system toward a single frequency. Of greater interest is the response of the two types of systems to a square pulse. This requires the use of Fourier integrals.

The signal whose response is to be investigated, shown in Fig. 6.7*a*, consists of a square pulse whose duration is 2τ. Its spectrum can be expressed as

$$G(\omega) = k\frac{\sin \omega\tau}{\omega} \tag{6.8}$$

where ω is $2\pi p$, p being the video frequency. Hence the Fourier integral describing the pulse is

$$\frac{2k}{\pi}\int_0^\infty \frac{\sin \omega\tau}{\omega}\cos \omega t\, d\omega \tag{6.8a}$$

This pulse modulates a carrier whose frequency is $f_0 = \nu/(2\pi)$, giving rise to a signal having the following form:

$$\cos(\nu t + \phi) + \frac{k}{\pi}\int_0^\infty \frac{\sin \omega\tau}{\omega}\{\cos[(\nu+\omega)t+\phi] + \cos[(\nu-\omega)t+\phi]\}\, d\omega \tag{6.9}$$

This represents a carrier and both upper and lower sidebands extending over an infinite range on either side of this carrier. In the course

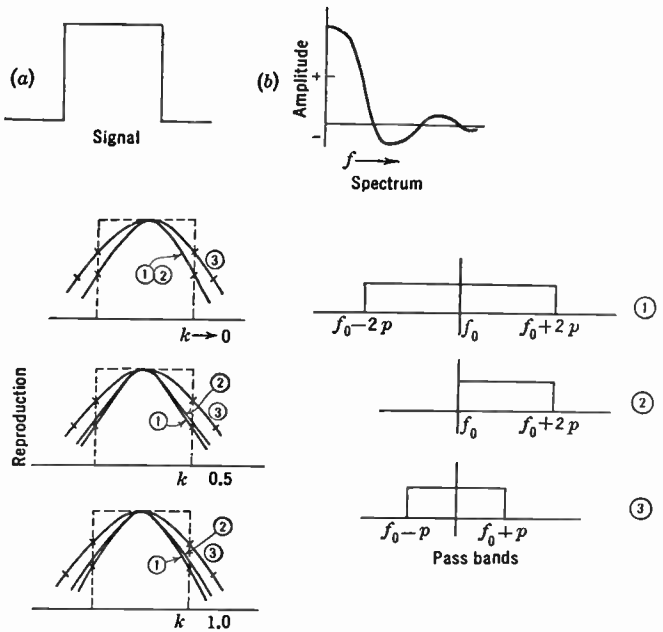


Fig. 6.7. Comparison of Single- and Double-Sideband Transmission of a Square Pulse.

of transmission and reception, the radio-frequency signal passes through circuit elements which act as filters, limiting the extent of the sidebands before it reaches the detector. Before it is possible to determine the shape of the video pulse leaving the detector, it is necessary to make certain assumptions about the characteristics of the passband. Three

conditions will be considered; they are illustrated in Fig. 6.7. In the last two, the passbands have equal widths, one passing frequencies from $f_0 - p$ to $f_0 + p$, the other from f_0 to $f_0 + 2p$. Over these bands the attenuation factor $B(f)$ is constant, being zero for all other frequencies. The phase factor will be assumed to be a linear function of frequency $\phi(f) = \alpha p + \beta$. The first filter admits both sidebands, having twice the width of the others, that is, it passes the range of frequencies from $f_0 - 2p$ to $f_0 + 2p$.

The detector output after the signal has been attenuated by the bandpass filter is determined, as before, from the envelope of the modulated signal. For the two filters passing double-sideband signals, the detector output is:

$$V_s = 1 + \frac{2k}{\pi}\int_0^{\Omega} \frac{\sin \omega\tau}{\omega} \cos \omega t'\, d\omega$$

$$= 1 + \frac{k}{\pi}\left(\mathrm{Si}[\Omega(t' + \tau)] - \mathrm{Si}[\Omega(t' - \tau)]\right)^* \tag{6.10}$$

where $2\Omega = 4\pi p_{\max}$ is the total width of the passband, and $t' = t + \alpha/(2\pi)$.

With single-sideband reception the results are not quite so simple. The signal reaching the detector is

$$\cos(\nu t + \phi_0) + \frac{k}{\pi}\int_0^{2\Omega} \frac{\sin \omega\tau}{\omega} \{\cos(\nu + \omega)t + \phi\}\, d\omega \tag{6.11}$$

or

$$\cos(\nu t + \phi_0) + \frac{k}{\pi}\left(\int_0^{2\Omega} \frac{\sin \omega\tau}{\omega} \cos \omega t'\, d\omega\right) \cos(\nu t + \phi_0)$$

$$- \frac{k}{\pi}\left(\int_0^{2\Omega} \frac{\sin \omega\tau}{\omega} \sin \omega t'\, d\omega\right) \sin(\nu t + \phi_0)$$

$$= \left[1 + \frac{k}{2\pi}\mathrm{Si}[2\Omega(t' + \tau)] - \frac{k}{2\pi}\mathrm{Si}[2\Omega(t' - \tau)]\right] \cos(\nu t + \phi_0)$$

$$+ \frac{k}{2\pi}\left[\mathrm{Ci}[2\Omega(t' + \tau)] - \mathrm{Ci}[2\Omega(t' - \tau)] - \ln\left(\frac{t' + \tau}{t' - \tau}\right)\right] \sin(\nu t + \phi_0) \tag{6.12}$$

* $\mathrm{Si}(x)$ is the "sine integral" of x: $\mathrm{Si}(x) = \int_0^x \frac{\sin y}{y}\, dy$. And $\mathrm{Ci}(x)$ is the "cosine integral" of x: $\mathrm{Ci}(x) = -\int_x^{\infty} \frac{\cos y}{y}\, dy$.

Here $\phi_0 = \beta$. This signal has an envelope which is given by

$$V_s = \sqrt{\left\{\left[1 + \frac{k}{2\pi}\,\mathrm{Si}[2\Omega(t' + \tau)] - \frac{k}{2\pi}\,\mathrm{Si}[2\Omega(t' - \tau)]\right]^2 + \left[\frac{k}{2\pi}\left\{\mathrm{Ci}[2\Omega(t' + \tau)] - \mathrm{Ci}[2\Omega(t' - \tau)] - \ln\left(\frac{t' + \tau}{t' - \tau}\right)\right\}\right]^2\right\}} \tag{6.13}$$

and the phase shift

$$\theta = \pm \tan^{-1} \frac{\mathrm{Ci}[2\Omega(t' + \tau)] - \mathrm{Ci}[2\Omega(t' - \tau)] - \ln\left(\dfrac{t' + \tau}{t' - \tau}\right)}{\dfrac{2\pi}{k} + \mathrm{Si}[2\Omega(t' + \tau)] - \mathrm{Si}[2\Omega(t' - \tau)]} \tag{6.13a}$$

The responses corresponding to the pulses passing through the three passbands:

(1) $2\Omega = 8\pi p$—double sideband,
(2) $2\Omega = 4\pi p$—single sideband,
(3) $2\Omega = 4\pi p$—double sideband,

are shown for k small, $k = 0.5$, and $k = 1.0$ in Fig. 6.7. It will be seen that even for the worst case, that of $k = 1$, the reproduction for single-sideband operation does not differ greatly from that obtained with double-sideband operation, using twice the overall bandwidth. For square-law detection these two coincide even more closely for all modulations. It is interesting to note that for a step function with single-sideband transmission, where the lower sideband is cut off sharply at the carrier frequency, the envelope does not converge. If, however, the passband tapers off gradually through the region of the carrier, as shown in Fig. 6.8, this is no longer true.

The conclusions which can be reached from this brief introduction to single-sideband operation are that if the cutoff of the suppressed sideband is not too sharp, and is symmetrical about the carrier, a faithful reproduction of the original video signal can be obtained with linear detection, and that the effective definition will be nearly that of double-sideband operation, using twice the overall radio-frequency bandwidth.

A diagram illustrating the ultra-high-frequency spectrum of a single-sideband transmitter was given in Fig. 6.5. The lower sideband only is suppressed, the gradual cutoff beginning at carrier frequency. If this form of signal reached the final detector, there would be a distortion

of the picture in the form of accentuated low frequency. Therefore, the bandpass characteristics of the radio- and intermediate-frequency stages of the receiver are so adjusted as to give a cutoff which tapers off gradually through the region of the carrier, as shown in Fig. 6.8. With this characteristic the requirements necessary for faithful reproduction with one sideband are met. The reasons for not shaping the transmitter characteristics to give the spectrum needed at the detector

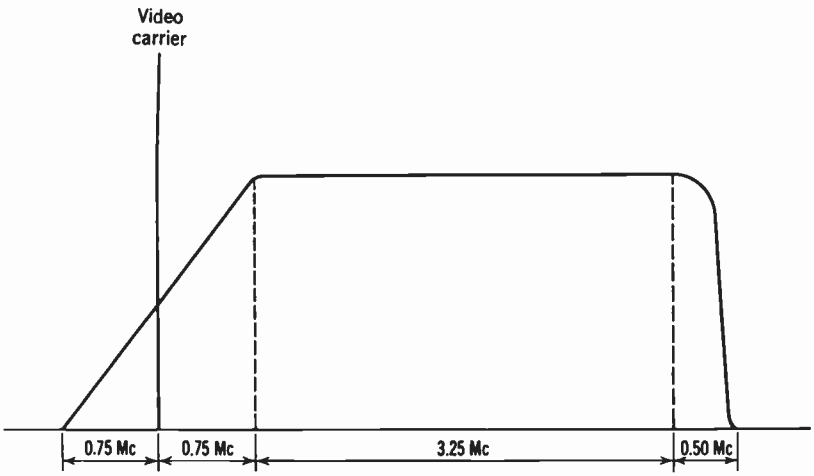


Fig. 6.8. Passband of a Receiver Meeting High-Definition Requirements.

are purely practical and connected with the difficulty of designing and constructing a filter capable of handling the large amounts of power, yet having the required attenuation and phase properties.

6.5 Radio Receiver. The signal received by the antenna is conveyed to the radio receiver by some form of balanced line. This line may merely be a twisted pair, if the location of the antenna is such that a relatively short line can be used, and if the signal strength is high. On the other hand, if the feed line is long, and particularly if the field strength at the antenna is small, it may be necessary to use a two-wire, air- or plastic-insulated, transmission line or even a coaxial cable.

The receiving equipment itself provides for converting the radio-frequency signal into a video signal and amplifying it sufficiently so that it can be used to operate the viewing device. It also separates the synchronizing impulses from the picture signal, employing the former to control the scanning pattern of the viewing device.

As with the conventional radio receiver, several alternative types of television receivers are possible. The two most common types employ tuned radio-frequency and superheterodyne circuits, respectively.

A tuned radio-frequency receiver has five or six tuned stages of ultra-high-frequency amplification, a detector, and several stages of video amplification. Because of the difficulty of accurately tuning the many radio-frequency stages, it is rarely used where reception from more than one transmitter is desired. Figure 6.9 shows in block diagram this type of receiver.

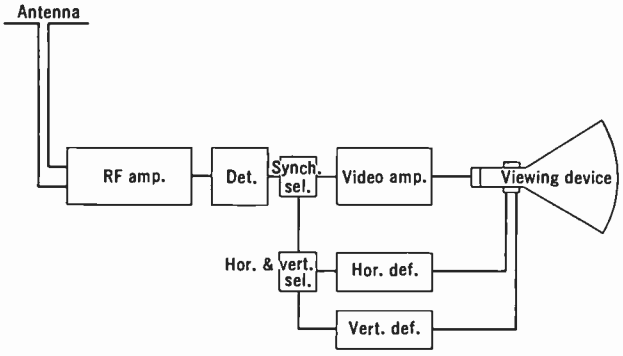


Fig. 6.9. Television Receiver Employing Radio-Frequency Amplification.

The superheterodyne receiver is made up of a radio-frequency element, a first detector and oscillator, an intermediate amplifier, and a second detector followed by a video amplifier. The radio-frequency element may merely be a broadly tuned circuit whose chief function is to match the antenna feeder to the first detector circuit, or it may be more complicated, containing, for example, a stage of radio-frequency amplification. The radio frequency output of this first stage is mixed with the output of the local oscillator and rectified by the first detector, giving rise to the intermediate frequency signal. The mixing is merely an addition to the incoming signal and the locally generated frequency, or a modulation of one by the other, producing a beat effect between these two waves. For example, if the incoming signal consists of the carrier alone, and so has the form $A_0 \cos 2\pi f_0 t$, and the local wave is $A_1 \cos 2\pi f_1 t$, the envelope of the resultant wave will be

$$\sqrt{A_0^2 + A_1^2 + 2A_0A_1 \cos 2\pi(f_0 - f_1)t} \tag{6.14}$$

The sideband frequencies which are present beat with the locally generated oscillation in a similar way. When the mixed signal is then rectified by an ideal square-law detector, the output contains an intermediate carrier whose frequency is $(f_0 - f_1)$ and sidebands associated with it which correspond exactly to the sidebands present in the original signal. The intermediate carrier and its sidebands are amplified by the intermediate amplifier, the other frequencies present in the output of the detector being removed by the bandpass characteristics of the interstage coupling of this amplifier. Rectification of the mixed

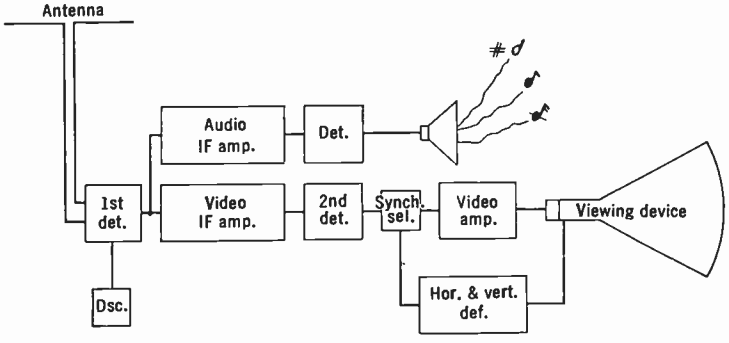


Fig. 6.10. Block Diagram of a Superheterodyne Television Receiver.

signal by a linear detector also leads to the intermediate carrier and sidebands, but there are present, in addition, terms resulting from the addition of various sideband frequencies. However, if the amplitude of the locally generated oscillation is large compared with the incoming signal, the amplitude of the unwanted combinations of sideband frequencies will be small compared with those of the fundamentals. For practical reasons it is common to use a linear detector and a powerful local oscillator rather than a square-law detector.

After amplification, the intermediate carrier and sidebands are again rectified, this time by the second detector, which is also usually a linear detector. The output of this element is a signal which corresponds exactly to the video signal generated by the pickup device together with synchronizing impulses and any correcting signals inserted at the monitoring amplifier. A superheterodyne receiver is illustrated by the block diagram given in Fig. 6.10.

6.6 Receiver Video Amplifier. The level of the video signal leaving the final detector is, in general, not sufficient to actuate the viewing device. In consequence, it must be amplified by several stages of video amplification before it can be used. These stages are similar to

those between the pickup equipment and the transmitter, being resistance-coupled and compensated to increase the high-frequency response. Because of the nature of the coupling, the amplifier will pass only the a-c components of the video signal. Therefore, it is usually necessary to provide some arrangement to convert the signal to a d-c signal, in order to establish the average illumination of the picture being transmitted. By "d-c" signal is meant one for which the voltage relative to some fixed reference voltage corresponds to the absolute brightness of the point under the exploring spot at the transmitter.

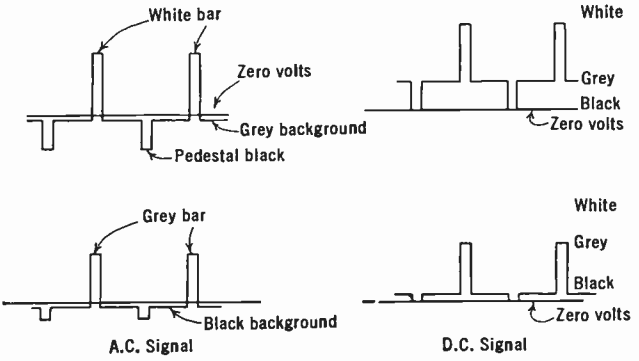


Fig. 6.11. The Video Signal with and without the D-C Component.

The diagrammatic oscillograms of Fig. 6.11 illustrate the distinction between the two types of signal. The d-c level of the video signal can be reinserted by making use of the pedestal or of the synchronizing impulse, because these bear a definite relation to either maximum dark or light of the picture. A diode rectifier or its equivalent in connection with the final stage of the amplifier, as will be described in Chapter 16, is the usual means for reinserting the d-c component of the signal.

6.7 The Viewing Device. The viewing device converts the video signal into the final visible reproduction, which is the ultimate objective of the system.

Two operations are necessary to perform the reconstructing of the image. The first consists of modulating a small but intense spot of light with the video signal; the second, of causing this spot to move in such a way that it sweeps out the scanning pattern on the viewing screen.

Broadly, there are two classes of viewing devices, namely, electronic and mechanical. Of these, the former are almost exclusively used at present, particularly in home receivers. Both systems have been suc-

cessfully utilized as a means of producing large pictures for public viewing, although here, too, electronic devices alone find practical application at the present time.

A survey of electronic and mechanical viewing devices will be presented in Chapter 8. Further description of these devices and their operation will be deferred until then.

6.8 Synchronization. As was pointed out in the preceding chapter, the geometric fidelity of the reproduction depends upon the exact correspondence in position of the scanning spot at the transmitter and receiver. This correspondence requires that the periodicity and phasing of the horizontal and vertical scanning motions at the two termini be alike. It has not been found practicable to attempt to maintain this condition with independent control of the scanning pattern at the pickup and viewing device. Instead, suitable signals are transmitted which indicate the beginning of each frame (or field repetition in the case of interlaced scanning) and the beginning of each line. These synchronizing signals are part of the complete video signal, and occur during time not utilized by the picture signal itself, that is, during the interval in which the scanning spot is returning to its original position, after completing a line or field traversal. In terms of the frequency spectrum of the picture signal discussed in Chapter 5, the frequency components of the synchronizing signal are multiples of line and frame frequency, their phase and amplitude being such that they have no effect on the picture except around the edges, where they form a narrow border which cannot be used for the image and which, in cathode-ray terminal tubes, has no real existence. Instead of this multiplicity of frequencies it might seem possible to use, for purposes of synchronization, a single frequency, falling, for example, in one of the vacant regions of the picture spectrum. However, while no picture component of this frequency appears in the spectrum, the presence of the signal would be reproduced in the form of a spurious moving shading.

In a simple Nipkow disk mechanical system, the rotation of a single element only need be governed, and for this one kind of impulse alone is necessary. The usual procedure is to drive the disk with two motors: one a variable-speed motor which supplies most of the power and is adjusted to run at nearly the required speed, the other a synchronous motor driven by the synchronizing impulse itself, which serves to lock the disk into the correct speed and phase.

More often two types of synchronizing impulses are needed, one to control the horizontal spot displacement, the other the vertical motion. Mechanical film scanning requires synchronization of the vertical scan-

ning produced by the motion of the film, and of the disk or drum giving line scanning. Similarly, where two mirror drums are used for picture reproduction, one for each direction of scanning, both frame and line synchronization are necessary.

Scanning in the case of electronic terminal tubes is produced by deflecting an electron stream periodically in two mutually perpendicular directions by means of suitably varying magnetic or electric fields. The current or voltage producing these fields is supplied from two deflection generators, one operating at line frequency, the other at frame (or field) frequency. Each generator is controlled by its own synchronizing impulse; therefore, the complete signal must include two types which can be distinguished from one another by some form of selector circuit.

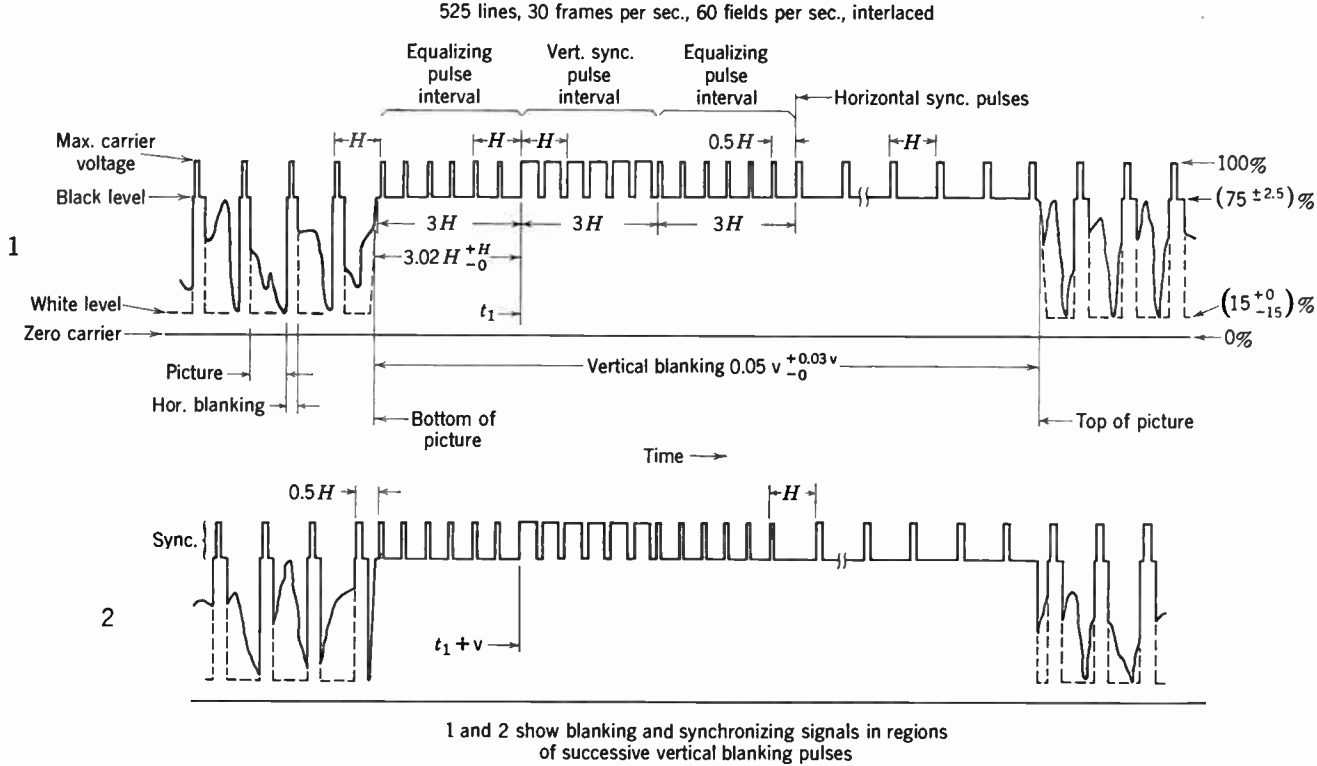
It is general practice to make the generator producing the synchronizing impulses at the transmitter the fundamental timing unit of the entire system. Thus the synchronizing signal governs not only the scanning pattern at the reproducer, but also that at the pickup. When possible, the synchronizing signal generator is itself tied in with the public electric power supply for the service area of the transmitter in order to minimize the effect of stray 60-cycle fields on the picture.

The exact shape of the impulses for horizontal and vertical synchronization depends upon how the signal is applied at the deflection generator, and upon the circuits which separate the two components. This phase of the problem will be considered in a later chapter.

Experience has prompted the almost universal adoption of synchronizing impulses which are "blacker than black," that is, the blanking level corresponds to black in the video signal, and the impulses extend below this level in the direction of black.

Some of the conditions which must be satisfied by the synchronizing signal are:

1. The wave shape of the pulse must be such as to permit accurate control of the scanning pattern.
2. The pulse must not interfere with the picture.
3. The signal must be of such shape and amplitude that synchronization is not easily interrupted by interference.
4. The line and frame pulses must be easily separable.
5. Where possible, the signals should aid in some other picture operation, such as extinguishing the beam in the cathode-ray viewing tube or providing automatic gain control.



1 and 2 show blanking and synchronizing signals in regions of successive vertical blanking pulses

Fig. 6.12. FCC Standard Television Synchronizing Signal.

The form of the synchronizing impulse which has been adopted as the present standard in this country by the Federal Communications Commission is shown in Fig. 6.12. How this signal meets the listed requirements will be discussed in Chapter 14, devoted to the problems of scanning and synchronization.

REFERENCES

1. F. Schröter (Editor), *Fernsehen,* Springer, Berlin, 1937.
2. F. E. Terman, *Radio Engineering,* McGraw-Hill, New York, 1947.
3. D. G. Fink, *Principles of Television Engineering,* McGraw-Hill, New York, 1940.
4. D. G. Fink (editor), *Television Standards and Practice,* McGraw-Hill, New York, 1943.
5. R. D. Kell, "Description of Experimental Television Transmitting Apparatus," *Proc. I.R.E.,* Vol. 21, pp. 1674–1691, 1933.
6. G. L. Beers, "Description of Experimental Television Receiver," *Proc. I.R.E.,* Vol. 21, pp. 1692–1706, 1933.
7. "An Experimental Television System," *Proc. I.R.E.,* Vol. 22, November, 1934: Part I by E. W. Engstrom, pp. 1241–1245; Part II by R. D. Kell, A. V. Bedford, and M. A. Trainer, pp. 1246–1265; Part III by R. S. Holmes, W. L. Carlson, and W. A. Tolson, pp. 1266–1285.
8. W. J. Poch and D. W. Epstein, "Partial Suppression of One Side-Band in Television Reception," *Proc. I.R.E.,* Vol. 25, pp. 15–31, January, 1937; also *RCA Rev.,* Vol. 1, pp. 19–34, January, 1937.
9. J. E. Smith, B. Trevor, and P. S. Carter, "Selective Sideband and Double Sideband Transmission of Telegraph and Facsimile," *RCA Rev.,* Vol. 3, pp. 213–238, October, 1938.

7 Video Pickup Devices

7.1 Mechanical Systems: the Nipkow Disk. The earliest means for obtaining scanning suitable for television transmission was the Nipkow scanning disk. In its simplest form it is a perforated disk, having apertures spaced at equal angular intervals on a spiral. A Nipkow disk having a spiral of one turn is shown in Fig. 7.1.

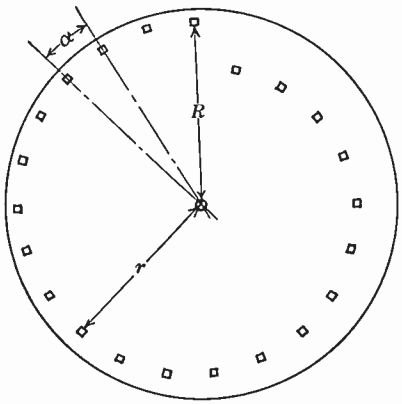


Fig. 7.1. Nipkow Disk.

The basic arrangement of a Nipkow disk in a direct pickup device for a mechanical television system is illustrated in Fig. 7.2. The scene being transmitted is imaged on the disk by means of lens L. Behind the disk is a light-sensitive element such as a photosensitive secondary-emission multiplier, or a phototube, which generates a current proportional to the light passing through the apertures. The image area occupies only a small portion of the disk's surface, its width being equal to the chord separating the apertures. The top and bottom of the image coincide with the outer and inner apertures. One and only one opening at any given instant permits light to reach the light-sensitive element. As the disk turns, successive apertures sweep across the image area, scanning it with a pattern of nearly straight parallel lines.

With a spiral of one revolution as shown, the disk must rotate at frame frequency N. Furthermore, the angular separation between the holes must be $\alpha = 360/n$ degrees, where n is the number of lines making up the scanning pattern. The spacing between these perforations is, therefore, $2\pi r/n$, r being the radial distance from the center of the disk to the point in question. This spacing is the width of the image on the disk. Assuming an aspect ratio of $1/b$ for this image, its height must be $2\pi br/n$. The perforations for maximum efficiency are square, and

so positioned that the edges of consecutive lines just touch. Obviously the height of each square opening is therefore $2\pi br/n^2$. It should now be apparent that r can be taken as equal to the radius R of the disk, since over the entire pattern all the perforations lie within R and $R\left(1 - \frac{2\pi b}{n}\right)$, and n for high definition is large.

The usefulness of this pickup arrangement depends upon the video signal which can be obtained with a given amount of light. In Fig. 7.2,

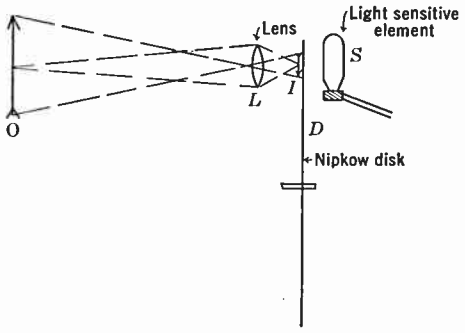


Fig. 7.2. Direct Pickup Using Nipkow Disk.

the light reaching the photosensitive element S, placed behind the disk, and generating the video signal, can be estimated as follows. On a very bright midsummer day, the illumination in an outdoor scene may reach as high as 10,000 foot-candles, although 3000 foot-candles might be taken as an average for summer daytime illumination. If a reflection coefficient of 1/2 is assumed, the brightness B of the scene will be

$$B = \frac{10{,}000}{2\pi} \cong 1500 \text{ candles per square foot}$$

The incident illumination of the image projected on the disk depends upon the photographic aperture or f-number A of lens L. For purposes of computation, a value of $A = 2$ for the aperture is assumed. The illumination under these conditions is

$$I = \frac{\pi B}{4A^2} = \frac{10{,}000}{2\pi}\,\frac{\pi}{4 \cdot 2^2} \cong 300 \text{ lumens per square foot}$$

It should be noted that absorption in the lens is not considered, and the field angle is assumed to be small. Practical considerations limit the maximum diameter of the scanning disk, for it must be maintained

accurately in synchronism at a very high angular velocity. A circumference of 12 feet, though large, is not beyond physical realization. The light flux passing through an aperture of a disk of this size will be

$$F = I\left(\frac{2\pi br}{n^2}\right)^2 = 4 \times 10^4 \cdot \frac{b^2}{n^4} \text{ lumen}$$

If, now, a picture of the quality under discussion is being transmitted, with 525 lines making up the scanning pattern, the amount of light reaching the phototube will be about 3×10^{-7} lumen. Even if the sensitivity of the light-sensitive element is 50 microamperes per lumen, the signal current will be only 1.7×10^{-11} ampere. Whether or not this amount of signal is sufficient to give an acceptable picture depends upon the noise fluctuation that accompanies the signal. The absolute minimum of noise is determined by the statistical fluctuation of the photocurrent itself. This lower limit might be approximated if the light-sensitive element were a good secondary-emission multiplier. This limiting noise, over the required 4.25-megacycle video band, is

$$\overline{i_n^2} = 2eIf = 32 \times 10^{-20} \times 1.7 \times 10^{-11} \times 4.25 \times 10^6$$

or

$$\overline{i_n} \text{ (rms)} \cong 5 \times 10^{-12} \text{ ampere}$$

In other words, the signal-to-noise ratio, even under the ideal illumination assumed, is only about 3. This ratio is far too low to give what can be considered a picture of high entertainment value.

The light reaching the light-sensitive element can be increased to some extent by means of a more complicated arrangement, as, for example, a disk having a spiral of two or more turns, or one in which lenses replace the apertures. However, this increase is not sufficient to make the system practical for direct pickup.

Although these calculations indicate that the scanning disk is not suited for this service, it has found a place in other types of pickup.

7.2 Flying Spot Scanning. For studio pickup of a very small area, e.g., the close-up of a head, a Nipkow disk used in a somewhat different way has been found fairly successful. The pickup equipment is arranged as shown schematically in Fig. 7.3. Instead of an image of the scene being projected on the disk, the disk is illuminated with a high-intensity arc, or other powerful light source, so that the light through it forms a narrow moving beam. This beam is focused on the object being televised, to produce a spot of light which scans its surface. The light reflected back from the object is received in a battery of photo-

tubes surrounding it. These tubes produce the video signal. The advantage of imaging the flying light spot on the scene and picking up the reflected light from all points continuously is twofold. The surrounding phototubes can be made to subtend a very large solid angle about the illuminated point, thus making use of a fairly large fraction of the reflected light. Also, the average illumination needed on the scene is only $1/n^2$ that which would be needed to obtain an equal video signal by the first-mentioned system, even if the light-gathering power of the objective imaging the scene on the disk is

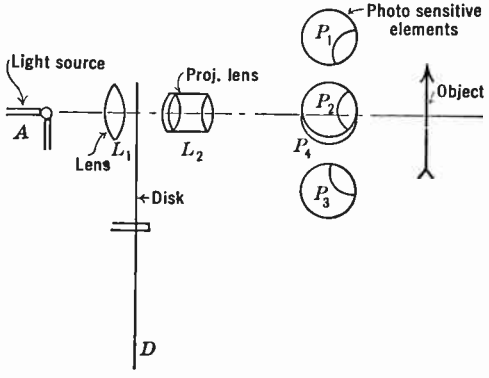


Fig. 7.3. Flying Spot Scanner.

assumed to be equivalent to that of the phototubes. In order to maintain the large light-gathering power of the phototubes, it is necessary to restrict the field to be reproduced to a relatively small area in the immediate vicinity of the light-sensitive elements.

7.3 Film Scanning. In a very similar way, the Nipkow disk can be used to pick up the image from a moving-picture film. Two arrangements for this type of pickup are shown in Fig. 7.4. The image from the film is projected onto the scanning disk in the first. In the other, the light source is imaged onto the disk, and the opposite side of the disk is in turn imaged onto the film, so that the apertures produce a small moving spot of light which sweeps over the picture. In each case, the light passing through the disk and film is picked up and converted into the video signal by a light-sensitive element. The choice between these two arrangements rests on practical considerations of mechanical construction, rather than any inherent superiority in optics. For the direct pickup of actual objects, vertical displacement of the scanning spot is obtained through arranging the series of apertures in a spiral. This is not necessary for film pickup, as the motion of the

film itself can be utilized. The perforations in the disk are arranged in a circle, and the scanning spot moves across a single line, rather than over a pattern of parallel lines. The film is moved with constant speed and at such a rate that a single frame passes the position of the scanned line in one picture period. Therefore, as illustrated in Fig.

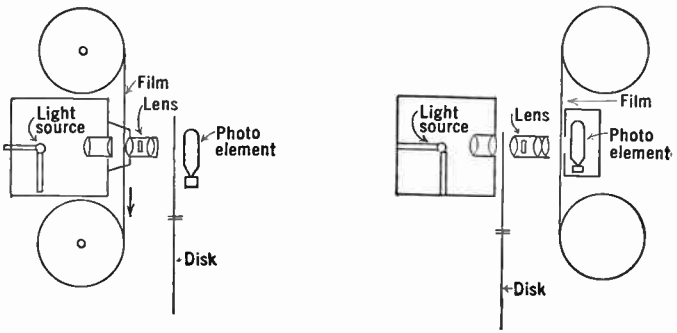


Fig. 7.4. Two Forms of Simple Disk Film Scanners.

7.5, relative to the film the scanning spot sweeps out an ordinary pattern of parallel lines.

So far no mention has been made of interlaced scanning. Many expedients are possible. For film transmission, odd-line interlacing

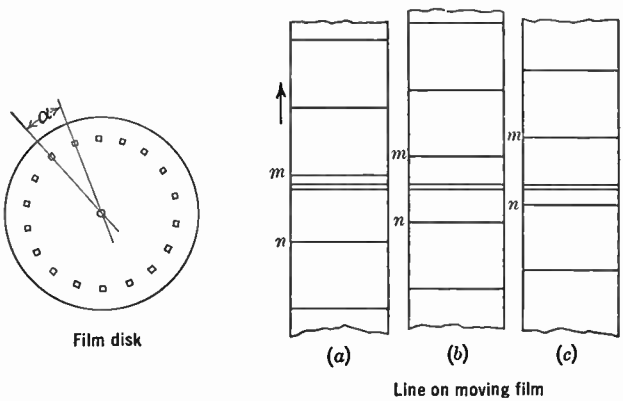


Fig. 7.5. Principle of Film Scanning.

is obtained if the disk contains the same number of apertures as the complete pattern has lines, and the film passes it at the rate of two frames per revolution. A double spiral of holes will give interlacing for studio pickup.

To return to film scanning, the disk can be designed so as to make more than one revolution per picture period. If α is the angular separation between apertures, and n' the number, the relation between the number of revolutions per second, N', and the frame frequency is

$$\alpha n N = 2\pi N'$$

$$\alpha n' = 2\pi$$

$$\frac{n}{n'} N = N'$$

When n' is less than n, a higher rate of revolution is required; however, the diameter of the disk can be reduced, with a consequent reduction in the net mechanical stresses.

7.4 Modifications of Scanning Disk. As an alternative to the disk described, a drum perforated with a helical array of apertures can be used to obtain the scanning pattern required for the direct pickup from a physical object. Similarly, a single row of perforations around such a drum will serve as a film scanner.

A modification of the simple disk is shown in Fig. 7.6. Instead of the tiny apertures heretofore described, the holes are fairly large, and each is fitted with a lens. Figure 7.6 shows a lens disk arranged for film pickup. Each lens in the disk is a micro-objective of very short focal length and large numerical aperture. These objectives form very much reduced images of the relatively large aperture on the film. The tiny bright image is the exploring element which sweeps out the scanning pattern. Again a phototube intercepts the light passing through the film and generates the video signal. Of course, the lens disk has its counterpart in the lens drum. A lens drum used for practical television transmission by Telefunken is shown in Fig. 7.7. This drum has a light gain of the order of thirty times over a simple perforated drum of equal size. The drum is equipped with a single row of holes for film transmission and a helix of two revolutions for direct pickup. The complete scanner incorporating this drum is illustrated in Fig. 7.8.

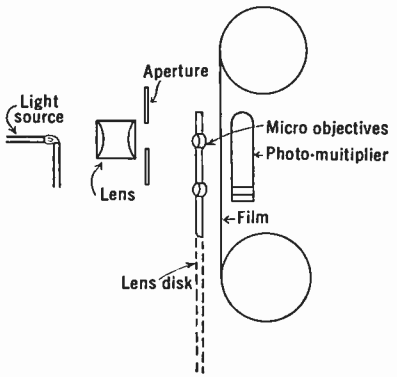


Fig. 7.6. Film Scanner with Lens Disk.

Fig. 7.7. Lens Drum (Telefunken). (Courtesy of J. Springer.)

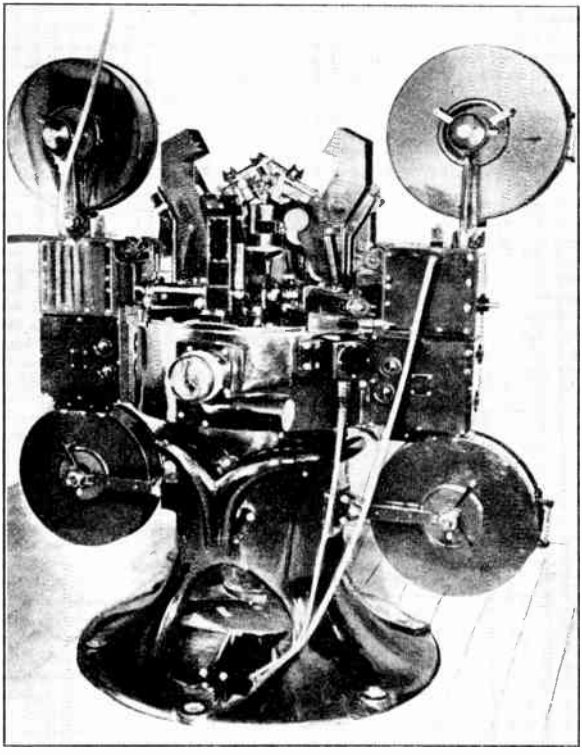


Fig. 7.8. Complete Film and Direct Pickup Scanner (Telefunken). (Courtesy of J. Springer.)

A very successful film transmitter, built and used by the Fernseh Company, is shown in Fig. 7.9. A 375-line picture with a 25-cycle frame frequency is obtained from it. In this machine an image of the continuously moving film is projected onto an aperture disk, and the light passing through the latter is picked up by a photoelectric sec-

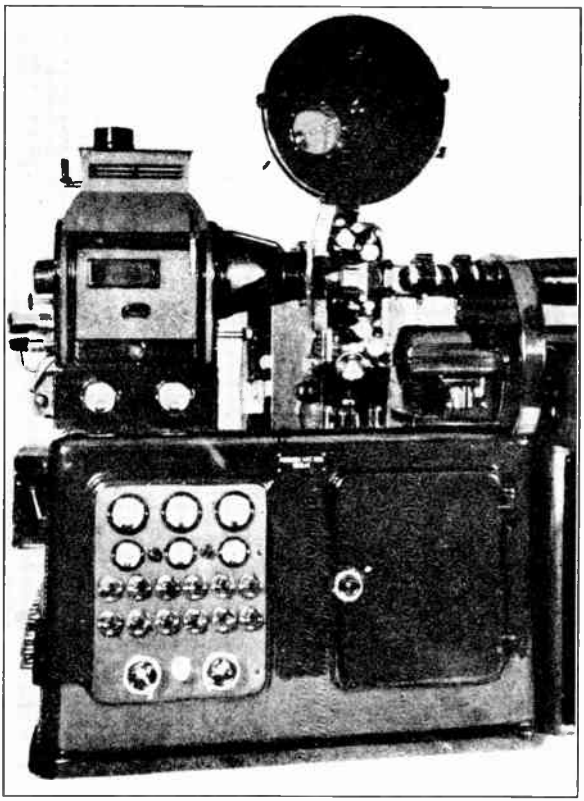


Fig. 7.9. Lens Disk Film Scanner (Fernseh). (Courtesy of J. Springer.)

ondary-emission multiplier. The disk itself is very carefully balanced, and rotates in a partially evacuated housing in order to reduce the power required to drive it and thus facilitate synchronization.

The video-telephone system which once operated in Germany between Berlin, Leipzig, and Munich made use of a lens-disk flying spot scanner, giving a 180-line pattern. Large photoelectric multipliers arranged around the window through which the moving spot is projected transform the reflected light into the video signal. This scanner is shown in Fig. 7.10.

Other forms of mechanical scanning have been used experimentally with varying degrees of success. The mirror drum, illustrated in Fig. 7.11, exemplifies one form. A modification of this drum has curved

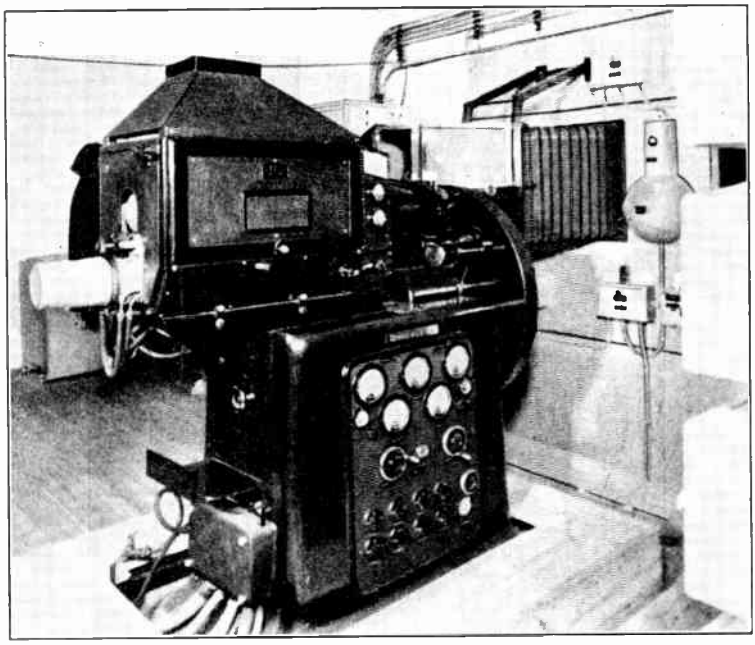


Fig. 7.10. Flying Spot Scanner for a Video-Telephone (Fernseh). (Courtesy of J. Springer.)

mirrors instead of flat. Prisms take the place of mirrors in another, very similar, arrangement. These scanners, together with others not mentioned, have shown some promise as a means of effecting mechani-

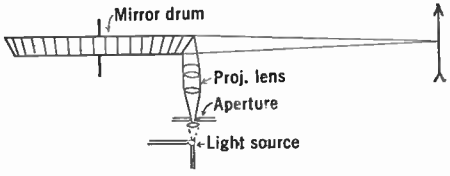


Fig. 7.11. Schematic Diagram of a Mirror Drum Scanner.

cal television pickup, but they have not been developed to the same extent as have the disk and drum scanners described in the earlier pages of this chapter.

7.5 Intermediate Film Pickup. Although, as the foregoing estimate indicates, the mechanical scanner, because of its intrinsic lack of sensitivity, is excluded from the field of direct pickup, this, by itself, is not so serious a limitation as it appears to be at first glance. An intermediate moving-picture film provides the means for overcoming the obstacle. The scene to be televised is photographed by means of a moving-picture camera, rapid development of this film then follows, and finally a mechanical film scanner converts the registered image into the video signal. A diagrammatic view of this type of trans-

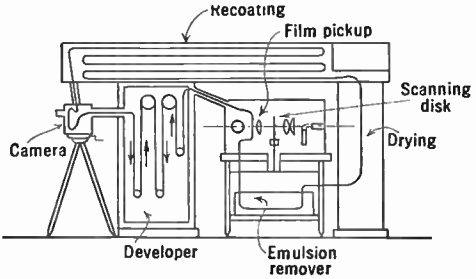


Fig. 7.12. Continuous Intermediate Film Scanner.

mitter, including camera, processing equipment, and scanner, is shown in Fig. 7.12.

One disadvantage of this system is the delay between the actual taking of the picture and converting it into video signal, due to the time required to process the film. Special methods of developing and fixing the film, however, have reduced this time to less than a minute. Of course, it is unnecessary to dry the film before it is used in the scanner. Also, a negative is as satisfactory as a positive for transmission because the reproduced picture can be changed from negative to positive merely by changing the polarity of the signal. (*Note:* This inversion takes place with every stage of amplification so that an amplifier of an even number of stages gives a picture having the same sense as that at the input, whereas an odd number of stages inverts light and dark.) Therefore, the delay that would be incurred by printing or reversing the film is avoided. The most serious practical objections to the system are the complicated and bulky equipment required and the cost of operation.

7.6 Electronic Pickup Systems: Electronic Scanning. The first published suggestion of a cathode-ray pickup arrangement appears to have been made by A. A. Campbell Swinton in a letter to *Nature* as early

as June, 1908. As a historic landmark this letter is worthy of quoting verbatim:

DISTANT ELECTRIC VISION

Referring to Mr. Shelford Bidwell's illuminating communication on this subject published in *Nature* of June 4, may I point out that though, as stated by Mr. Bidwell, it is wildly impracticable to effect even 160,000 synchronised operations per second by ordinary mechanical means, this part of the problem of obtaining distant electric vision can probably be solved by the employment of two beams of kathode rays (one at the transmitting and one at the receiving station) synchronously deflected by the varying fields of two electromagnets placed at right angles to one another and energised by two alternating electric currents of widely different frequencies, so that the moving extremities of the two beams are caused to sweep synchronously over the whole of the required surfaces within the one-tenth of a second necessary to take advantage of visual persistence.

Indeed, so far as the receiving apparatus is concerned, the moving kathode beam has only to be arranged to impinge on a sufficiently sensitive fluorescent screen, and given suitable variations in its intensity, to obtain the desired result.

The real difficulties lie in devising an efficient transmitter which, under the influence of light and shade, shall sufficiently vary the transmitted electric current so as to produce the necessary alterations in the intensity of the kathode beam of the receiver, and further in making the transmitter sufficiently rapid in its action to respond to the 160,000 variations per second that are necessary as a minimum.

Possibly no photoelectric phenomenon at present known will provide what is required in this respect, but should something suitable be discovered, distant electric vision will, I think, come within the region of possibility.

Although this letter did not describe in detail how such an electronic scanner could be made to operate to convert the light image into the desired video signal, a later paper by the same author, presented before the Roentgen Society in 1911, and repeated in part in *Wireless World*, April, 1924, discussed in some detail a pickup tube embodying cathode-ray scanning. The tube and circuit are diagrammed in Fig. 7.13. The transmitting tube is a Crookes tube with the cathode B connected to the negative end of a 100,000-volt d-c power source, while the anode is connected to the positive end. The anode is perforated with a fine aperture through which passes a narrow pencil of cathode rays. Continuing with the description given by A. A. Campbell Swinton in *Wireless World and Radio Review:* *

. . . the cathode rays fall on a screen J, the whole surface of which they search out every tenth of a second under the influence of the magnets D and E. Further, it is to be remarked that as the two magnets D and D', and the two magnets E and E' are energised by the same currents, the move-

* See A. A. Campbell Swinton, reference 7.

ments of the two beams of cathode rays will be exactly synchronous and the cathode rays will always fall on the two screens H and J on each corresponding spot simultaneously.

In the transmitter, the screen J, which is gas-tight, is formed of a number of small metallic cubes insulated from one another but presenting a clean metallic surface to the cathode rays on the one side, and to a suitable gas or vapour, say sodium vapour, on the other. The metallic cubes which compose J are made of some metal, such as rubidium, which is strongly active photoelectrically, in readily discharging negative electricity under the influence of light, while the receptacle K is filled with a gas or vapour, such

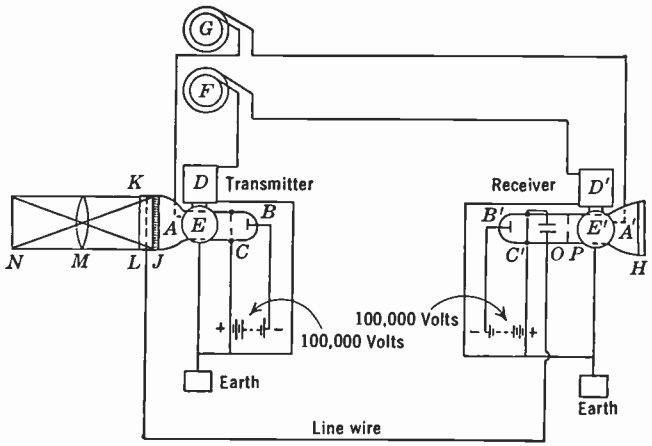


Fig. 7.13. Electronic Television System Proposed by Campbell Swinton.

as sodium vapour, which conducts negative electricity more readily under the influence of light than in the dark.

Parallel to the screen J is another screen of metallic gauze L, and the image to be transmitted of the object N is projected by the lens M through the gauze screen L on to the screen J, through the vapour contained in K. The gauze screen L of the transmitter is connected through the line wire to a metallic plate O in the receiver, past which the cathode rays have to pass. There is, further, a diaphragm P fitted with an aperture in such a position as, having regard to the inclined position of B', to cut off the cathode rays coming from the latter, and prevent them from reaching the screen H, unless they are slightly repelled from the plate O, when they are able to pass through the aperture.

The whole apparatus is designed to function as follows: Assume a uniform beam of cathode rays to be passing in the Crookes tubes A and A', and the magnets D and E and D' and E' to be energised with alternating current, as mentioned. Assume, further, that the image that is desired to be transmitted is strongly projected by lens M through the gauze screen L on to the screen J. Then, as the cathode rays in A oscillate and search out the surface of J, they will impart a negative charge in turn to all of the metallic cubes of which J is composed. In the case of cubes on which no light is

projected, nothing further will happen, the charge dissipating itself in the tube; but in the case of such of those cubes as are brightly illuminated by the projected image, the negative charge imparted to them by the cathode rays will pass away through the ionised gas along the line of the illuminating beam of light until it reaches the screen *L*, whence the charge will travel by means of the line wire to the plate *O* of the receiver. This plate will thereby be charged; will slightly repel the cathode rays in the receiver; will enable these rays to pass through the diaphragm *P*, and impinging on the fluorescent screen *H*, will make a spot of light. This will occur in the case of each metallic cube of the screen *J*, which is illuminated, while each bright spot on the screen *H* will have relatively exactly the same position as that of the illuminated cube of *J*. Consequently, as the cathode ray beam in the transmitter passes over in turn each of the metallic cubes of the screen *J*, it will indicate by a corresponding bright spot on *H* whether the cube in *J* is or is not illuminated, with the result that *H*, within one-tenth of a second, will be covered with a number of luminous spots exactly corresponding to the luminous image thrown on *J* by the lens *M*, to the extent that this image can be reconstructed in a mosaic fashion. By making the beams of cathode rays very thin, by employing a very large number of very small metallic cubes in the screen *J*, and by employing a very high rate of alternation in the dynamo *G*, it is obvious that the luminous spots on *H*, of which the image is constituted, can be made very small and numerous, with the result that the more these conditions are observed, the more distinct and accurate will be the received image.

Furthermore, it is obvious that, by employing for the fluorescent material on the screen *H* something that has some degree of persistency in its fluorescence, it will be possible to reduce the rate at which the synchronised motions and impulses need take place, though this will only be attained at the expense of being able to follow rapid movements in the image that is being transmitted.

It is further to be noted that as each of the metallic cubes in the screen *J* acts as an independent photoelectric cell, and is only called upon to act once in a tenth of a second, the arrangement has obvious advantages over other arrangements that have been suggested, in which a single photoelectric cell is called upon to produce the many thousands of separate impulses that are required to be transmitted through the line wire per second, a condition which no known form of photoelectric cell will admit of.

The basic similarity between this type of tube and the mechanical scanners is evident from the description of its mode of operation. Instead of there being a single photocell upon which light is projected from the various picture elements in succession, there are many photocells which are in operation only during the time the scanning beam, which functions as a commutator, is actually on them. The photocurrent which makes up the video signal is collected on an anode which is common to all the cells, but since only one cell is active at any one instant, the current collected corresponds to that from the successive elements as the beam sweeps over them. The limiting sensitivity can

be calculated in the same way as that for the disk scanner. If A is the area of the array of elements and there are n^2/b elements in all, corresponding to an n-line picture with an aspect ratio of $1/b$, the area of each element a will be Ab/n^2. Now, assume an average illumination I falling on the picture area, and an effective photosensitivity p. The average current picked up by the collector element L (since one element, and only one element, is always under the beam) is given by

$$i_s = Ip\frac{bA}{n^2}$$

Under the conditions previously assumed with a screen area of 20 square inches and a photosensitivity of 20 microamperes per lumen, the signal current is 2.3×10^{-9} ampere. If this current could be collected with an ideal multiplier, the accompanying noise would be about 5.6×10^{-11} ampere. The signal-to-noise ratio, therefore, is sufficient to give a picture of the desired quality. The ratio, which decreases with the square root of the illumination, becomes too low for practical purposes when the illumination falls below an average value of 1000 foot-candles. Even at that value, noise would be plainly visible in the picture.

In computing the limiting sensitivity of the mechanical pickup arrangement it was assumed that the photocurrent was amplified by means of a "noiseless" multiplier. Such an arrangement is not always practicable. If, instead of the secondary-emission multiplier, a thermionic amplifier of conventional design is employed, the limiting sensitivity is greatly curtailed. This is because the statistical fluctuation introduced by the amplifier and coupling arrangement is much greater than that present in the photocurrent. The signal current output from the pickup device is passed through the coupling resistor, as shown in Fig. 7.14, and the IR drop across it actuates the grid of the first radio amplifier tube. The magnitude of this voltage depends upon the value of the coupling resistor, and this in turn is limited by the capacity to ground of the combined output element of the pickup tube, coupling element, and amplifier grid. Details of this limitation will be taken up in Chapter 13. With the electronic device just described, if a resistor of 20,000 ohms is used, the signal voltage is

$$e_s = 0.46 \times 10^{-4} \text{ volt}$$

whereas the noise voltage, as will be shown in the chapter referred to, is

$$\overline{e_n} \text{ (rms)} = 3.7 \times 10^{-5} \text{ volt}$$

Under these circumstances the signal-to-noise ratio is below the useful limit. A somewhat more complicated coupling arrangement may increase this ratio to some extent, but not greatly. The idea of replacing mechanical moving parts by an electron beam, however, represents a very important step forward.

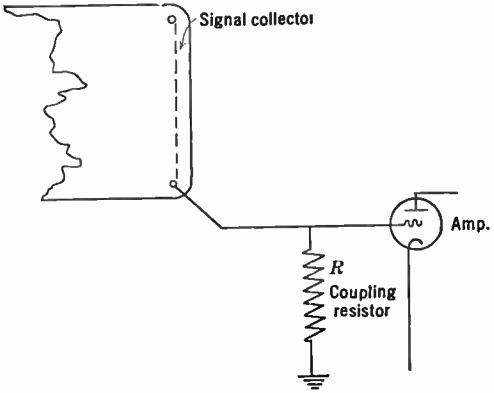


Fig. 7.14. Simple Means for Coupling a Pickup Tube to the Video Amplifier.

7.7 Dissector Tube. Another and quite different type of electronic scanning was developed by Philo T. Farnsworth.* The principle of this pickup device is shown in Fig. 7.15. The optical image is projected onto the photocathode, from which are released electrons whose density

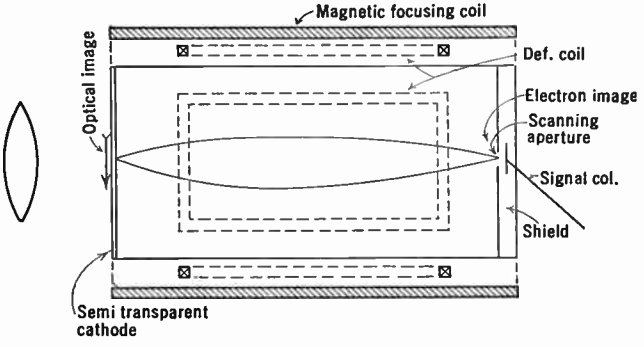


Fig. 7.15. An Early Farnsworth Dissector Tube.

is distributed in accordance with the distribution of light intensity. Thus, an electron image is formed. These electrons are accelerated by a potential difference between the photocathode and the shield at the

* See Farnsworth, reference 8.

other end of the tube (or suitable metallic coatings on the wall, etc.) and move toward the shield. A magnetic field produced by a long coil focuses the electron image on the shield. This shield is perforated in the center by a tiny aperture behind which is an electrode that collects the electrons passing through the aperture. Two pairs of coils, perpendicular to the axis of the focusing field and to each other, deflect the image as a whole, either vertically or horizontally. By supplying one pair with a sawtooth current at line frequency, and the other with a similar current wave at field frequency, the entire image is moved in such a way that the aperture sweeps out a scanning pattern relative to the image itself. In other words, moving the image across the scanning aperture is exactly equivalent to moving the aperture over the image. The current to the collecting electrode has an instantaneous value equal to the photoemission from the picture element resting on the aperture, and therefore varies as the aperture scans the image, to give the video signal. Just as in the previous pickup arrangement proposed by Campbell Swinton, the signal current has an instantaneous value equal to the photocurrent from a single element.

The electron optics involved in this device, which goes under the name of dissector tube, is of no little interest. The problem is one of imaging the emission from an extended photocathode. However, it is not identical with that of the image tube mentioned in Chapter 4, because in the latter all image points must be simultaneously in focus, whereas in the dissector tube only those points in the immediate vicinity of the scanning aperture need be accurately sharp. This makes it possible to use suitable components of the deflecting voltage or current to correct for any curvature of the image field when necessary. These components may be supplied either to the magnetic lens or to the accelerating voltage.

The dissector tube used with a conventional amplifier and coupling system, such as was described in the first-mentioned electronic pickup tube, would yield a signal-to-noise ratio too low to produce a picture of high entertainment value, except under exceptional lighting conditions, or in conjunction with a moving-picture projector. However, this type of tube lends itself readily to the use of a secondary-emission amplifier. As first used in this type of tube, the multiplier took the form shown in Fig. 7.16. Instead of the shield shown in Fig. 7.15, a narrow hollow metal tube contains the aperture which scans the image. This permits using an opaque photoemitter instead of the semi-transparent or screen cathode, the light being only slightly obstructed by the metal tube. Inside the tube, directly behind the aperture, is a

small secondary emitting element T which is made a few volts negative with respect to the tube itself. Electrons entering the aperture strike this element, producing secondary electrons which are collected by the walls of the metal tube. The net current resulting from the difference between the secondary-emission current and the entering

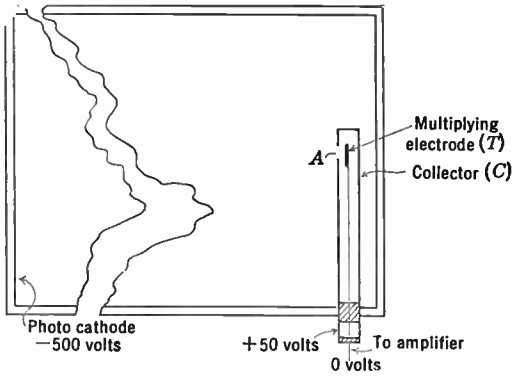


Fig. 7.16. Simple Multiplier in Farnsworth Dissector Tube.

electron stream is the signal current. With a good emitter for the target T, the current will be eight or more times the incoming photocurrent. Further secondary-emission multiplication was obtained in later tubes by incorporating a minute dynamic secondary-emission

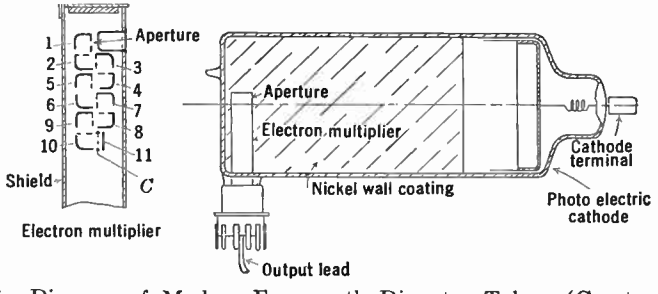


Fig. 7.17. Diagram of Modern Farnsworth Dissector Tube. (Courtesy Farnsworth Television, Inc.)

multiplier in the tube containing the aperture. Recent forms of this tube use a multi-stage static multiplier. A tube of this type is shown in Figs. 7.17 and 7.18. In practice, these tubes are found to give an excellent television image when the illumination is adequate. Because of the fairly high incident illumination needed, they are rather better

suited for film reproduction than direct pickup; however, they also still find employment in industrial television applications where illumination presents no problem.

Electronic scanning in the viewing tube, as successfully demonstrated by Boris Rosing in 1907, and in the pickup tube as proposed by Campbell Swinton in 1908, were the first two steps leading up to the modern electronic high-definition system. The third important step was the introduction of the storage principle.*

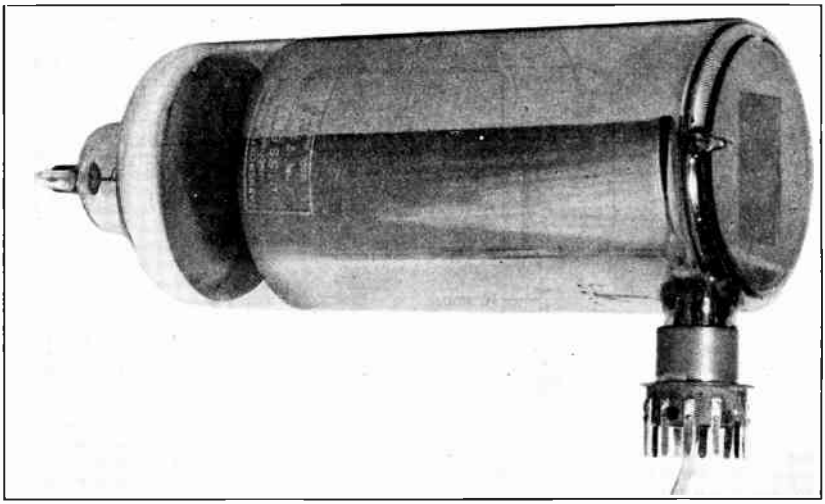


Fig. 7.18. Photograph of Modern Farnsworth Dissector Tube. (Courtesy Farnsworth Television, Inc.)

7.8 Storage Principle—the Iconoscope. In the systems described heretofore, the photoemission from each picture element is used only during the very short interval of time it is actually covered by the scanning spot or aperture. In contradistinction, under the storage principle, photoemission continues during the entire time, being accumulated in the form of charge at each image point. The entire accumulated charge is removed once each picture period, from each of the picture elements in sequence, as the scanning spot passes over them. The result is an increase in the effective photocurrent by a factor equal to the number of picture elements.

This principle, together with that of scanning with a cathode-ray beam, forms the basis of the iconoscope. There are many possible

* See Zworykin, reference 9.

modifications of this tube. Figure 7.19 illustrates one form whose operation is straightforward and therefore easily explained. This iconoscope consists of an electron gun G, a mosaic screen S, and two collecting elements A_2 and C, all enclosed in a glass bulb which is thoroughly evacuated. The details of the mosaic are shown in the figure. A conducting mesh screen is coated with a thin dielectric layer, leaving the openings of the mesh free. These interstices are filled with plugs of a photoelectrical material. Silver plugs activated with cesium are often used. The mosaic is mounted in the tube in such a position that

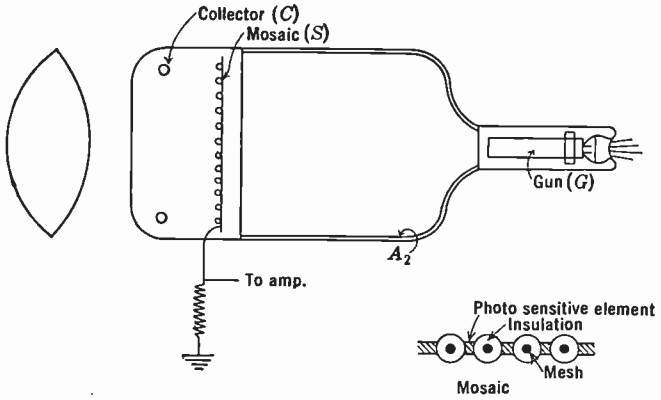


Fig. 7.19. Diagram of Iconoscope Employing a Two-Sided Mosaic.

the photosensitized side of the plugs faces the collector C, and the opposite side faces the gun.

The scanning beam, when it rests on any element, produces secondary electrons, and, if the emission is saturated, the yield exceeds the primary beam. Since the element is insulated, it tends to assume a positive equilibrium potential with respect to the collector A_2 such that the current reaching the element just equals the current leaving. This equilibrium potential is independent of the initial potential of the element.

When a portion of the mosaic is illuminated, elements in that area emit photoelectrons which go to the positive collector C. By virtue of this emission, the elements accumulate a positive charge. When the scanning beam in the course of sweeping out the pattern passes over a positively charged element, it is returned to its original equilibrium potential. Each illuminated element therefore goes through the cycle of accumulating a positive charge during the entire picture period and releasing it each time the scanning beam sweeps over it. Since each

element is coupled by capacity to the metallic screen which forms the foundation of the mosaic, any change in charge on the former induces a corresponding change in the latter. The change in charge produces a current in the lead connected to the coupling impedance and amplifier.

The sensitivity of this system can be estimated in a manner similar to that used to determine the behavior of the preceding systems. If I is the average light intensity of the optical image, A the area of the mosaic, and n^2/b the number of elements, the charge accumulated by an element during one picture period is

$$q = \frac{IApb}{n^2N}$$

This charge is released in an interval equal to that required by the beam in traversing the charged element, namely, $t_e = b/n^2N$. Therefore, the instantaneous current is

$$i_e = \frac{q}{t_e} = IAp$$

Since each element behaves in this way, the average signal current is equal to this value. Assuming as before a 20,000-ohm coupling resistance, the signal-to-noise ratio, for an incident illumination of 10,000 foot-candles, and a mosaic whose area and photosensitivity are 20 square inches and 20 microamperes per lumen, respectively, is

$$\frac{e_s}{e_n \text{ (rms)}} = \frac{IpAR}{3.7 \times 10^{-5}} = 4.5 \times 10^5$$

In other words, an incident illumination of about 1 foot-candle would be needed to give the required signal-to-noise ratio. This estimate is, of course, based on an ideal tube, and does not take into account various inefficiencies occurring in an actual tube.

7.9 Iconoscope; Single-Sided Mosaic Type. The iconoscope as used in practice operates on the same principle but has a somewhat different embodiment. The tube is diagrammed in Fig. 7.20. Instead of the two-sided mosaic built up on a mesh base, as in the preceding tube, the mosaic consists of a very thin mica sheet coated with conducting film on the side away from the gun, and with a myriad of tiny photosensitized silver elements on the other. Each of these elements is insulated from the others, and as a result of incident light stores a charge due to the emitted photoelectrons. The scanning beam returns the elements to their equilibrium potential with respect to the collector

A_2, just as in the tube using the two-sided mosaic. The actual cycle of operation which is described in detail in Chapter 9 is somewhat more complicated than that outlined above. This is because secondary electrons not only go to collector A_2, but also return to other parts of the mosaic.

The efficiency of this type of tube is fairly low, that is, only 5 or 10 percent, as compared to an ideal tube. This is in a large measure due to the fact that the photoemission from the mosaic is unsaturated.

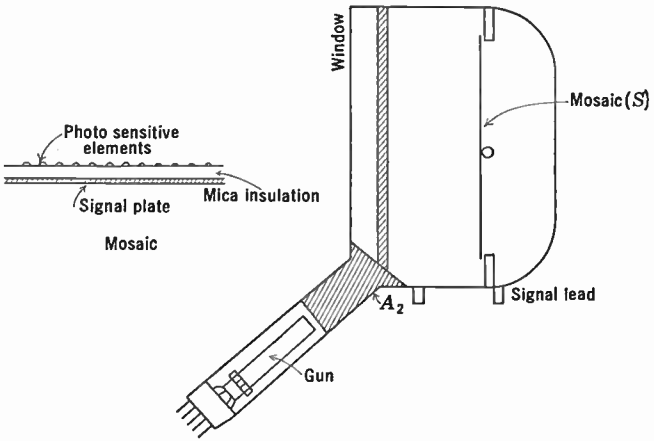


Fig. 7.20. Normal Iconoscope with a Single-Sided Mosaic.

Such a condition is to be expected since the potential of the mosaic element can never be greatly different from that required for secondary-emission equilibrium with respect to collector A_2. However, the photoemission, particularly for small light intensities, is more efficient than might at first be expected because of the effect of the secondary-emission electrons which return to the mosaic and tend to drive elements not directly under the beam slightly negative. A detailed analysis must also take into account the short-circuiting effect of these same returning electrons. The reason for the use of this type of mosaic instead of the two-sided form which permits saturated photoemission is purely technological. The difficulties encountered in trying to construct a two-sided mosaic suitable for the reproduction of a high-definition picture are enormous and present a serious obstacle to its practical production.

In spite of its inefficiency, the advantage gained by the use of the storage principle is so great that a tube of the type illustrated in Fig. 7.20 gives excellent performance, both for direct and film pickup. The experimentally determined sensitivity is sufficient so that a tube having

a mosaic area of 100 square centimeters and working with an $f/2.7$ lens will transmit a scene illuminated by 50 to 100 foot-candles with a signal-to-noise ratio of greater than 30, and will give a picture which is fully recognizable with less than 10 foot-candles illumination.

With all its high sensitivity, the iconoscope is not without its faults. One of the most serious is a spurious signal which is the result of an unsaturated secondary emission required in its operation. This spurious signal takes the form of shading over the reproduced picture and must be compensated for by special circuit networks in the video amplifiers immediately following the pickup tube.

A large share of more recent pickup tube development has been devoted to overcoming these faults of the iconoscope by changes in construction and operation. To mention only the steps which have resulted in tubes which have found, and are still finding, practical use in television broadcasting, there are:

1. The image iconoscope and Multicon (in America), Superemitron (in England), and Eriscope (in France), in which enhanced sensitivity is obtained, by the use of a more efficient continuous photocathode and secondary-emission amplification at the mosaic, through the placing of an image tube section ahead of the mosaic.

2. The orthicon (in America) and the Cathode-Potential-Stabilized (C.P.S.) Emitron (in England), where the elimination of spurious signals and the saturation of photoemission and secondary emission are attained by the employment of low-velocity scanning, though at the expense of instability at very high light levels.

3. The image orthicon, which incorporates an image tube section, two-sided mosaic, low-velocity scanning, and an electron multiplier for the signal current and thus manages to combine a sensitivity closely approaching the theoretical optimum for a storage system with perfect stability at all light levels and relative freedom from spurious signals.

The construction of all these tubes and details of their operating characteristics will be described in Chapter 10. They have so far outstripped in performance the mechanical pickup systems discussed in the earlier sections of this chapter that these systems retain only historical and didactic interest.

7.10 Photoconductive Pickup Tube. Effective pickup tubes, similar in construction to the iconoscope (Fig. 7.20) or the orthicon, may be obtained by replacing the photoemissive mosaic by a layer of photoconductive material. A tube of this type, or Vidicon, with low-velocity

scanning like the orthicon, is represented diagrammatically in Fig. 7.21. In darkness the resistivity of the photoconductive material is high enough (10^{12} ohm-centimeters or better) so that the low-velocity scanning beam can maintain the surface exposed to it substantially at the potential of the gun cathode; the time constant of the target (product of resistance and capacitance) is materially greater than the time between successive scannings ($\frac{1}{30}$ second). When the target is illuminated, on the other hand, the resultant photocurrent locally reduces

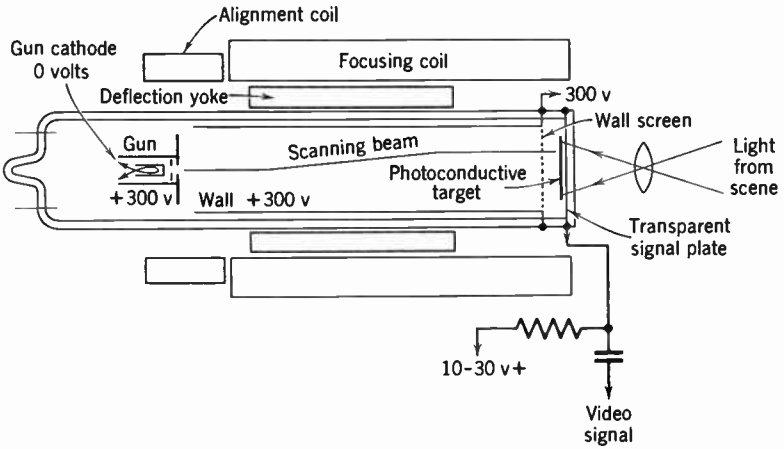


Fig. 7.21. The "Vidicon" Photoconductive Pickup Tube.

the voltage drop across the photoconductive layer, making the surface exposed to the scanning beam more positive. At the instant at which an illuminated picture element is scanned, just enough electrons are deposited by the scanning beam to replace the charge removed in the preceding frame period by photoconduction. The corresponding pulse in the capacitatively coupled signal plate circuit constitutes the picture signal.

It is clear from the foregoing description that the Vidicon operates as a true storage tube. As compared with the ordinary orthicon, it possesses the advantage of considerably greater photosensitivity: Selenium, as well as various sulfides, selenides, and oxides, which possess the requisite high resistivity, exhibit photocurrents which may amount to many electrons per incident light quantum; there exists thus, essentially, an internal signal multiplication which reduces the importance of additional secondary-emission multipliers and image tube stages. Also, the greater target sensitivity permits the use of

smaller targets. These factors combine to make the Vidicon ideally suited for equipment which is compact in size and simple in operation.

7.11 Electronic Flying Spot Scanning. For film and slide reproduction the flying spot scanning process described in section 7.3 now universally employs an electronic flying spot tube instead of the cumbersome combination of Nipkow disk and light source. This is simply a cathode-ray tube in which the scanning pattern is traced by an unmodulated beam on a short-persistence screen. The scanning pattern is imaged on the film or slide, so that the light reaching the photocathode of a multiplier phototube (Fig. 7.22) is just that transmitted

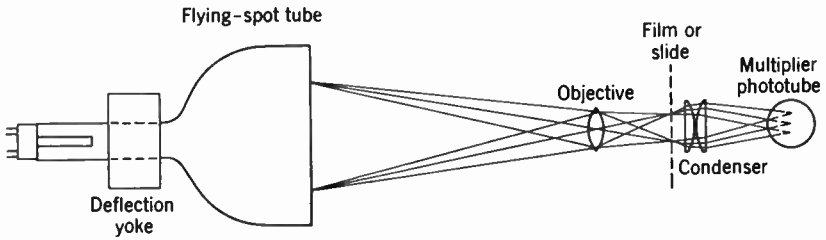


Fig. 7.22. Flying-Spot Pickup System for the Transmission of Still Pictures.

by a single picture element of the transparency. This presupposes that the light emission of the screen drops to a negligible fraction within a picture element period after passage of the beam, a condition satisfied by certain phosphor screens (e.g., a zinc oxide screen).

It can readily be seen that enough light is provided by the flying spot tube to yield picture signals adequately free from noise. A typical flying spot tube may operate at a voltage $V = 20{,}000$ volts with a beam current $i = 150$ microamperes and be provided with a 5-inch zinc oxide screen with a conversion efficiency σ of the order of 1 candlepower per watt. If m denotes the demagnification of the pattern, A the f-number of the imaging system, and p the sensitivity of the multiplier photocathode in amperes per lumen, the photocathode signal current i_s becomes, for 100 percent transmission,

$$i_s = \frac{iV\sigma m^2 p}{(2A)^2} = 2 \cdot 10^{-7} \text{ ampere}$$

for $m = \frac{1}{4}$, $p = 20 \cdot 10^{-6}$ amperes per lumen, and $A = 2$. If, in the highlights, we assume a total intensity reduction (by absorption and reflection) by a factor of 10, we still retain a signal current $i_s = 2 \cdot 10^{-8}$ ampere. The corresponding noise current is

$$i_n = \sqrt{32 \cdot 10^{-20} \cdot 4.25 \cdot 10^6 \cdot 2 \cdot 10^{-8}} = 1.6 \cdot 10^{-10}$$

The signal-to-noise ratio is thus better than 100, or quite sufficient to yield an excellent picture.

Like the scanning disk and the dissector, the flying spot tube represents a nonstorage system. The transmission of standard motion-picture film, with 24 frames per second, at the television frequency of 30 frames and 60 field per second, demands, hence, either extremely rapid pull-down (in vertical return time, approximately equal to 1

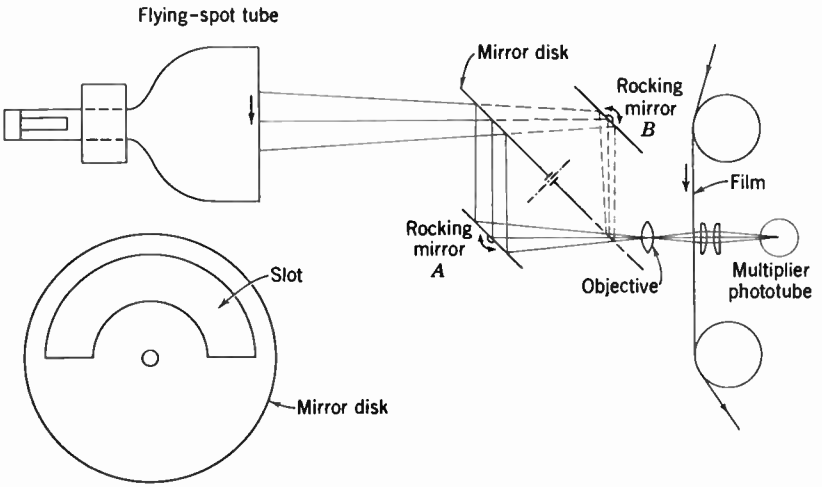


Fig. 7.23. Film Pickup System with Optical Compensation.

millisecond), alternatingly $\frac{1}{20}$ ($\frac{3}{60}$) and $\frac{1}{30}$ ($\frac{2}{60}$) second apart, in an intermittent film projector, or an optical compensation system in conjunction with continuous film motion.

In the second instance, exemplified by the arrangement shown in Fig. 7.23, the scanning pattern on the flying spot tube is compressed, so that its height is just three-fifths the height of the normal pattern; the film motion at 24 frames per second is then just sufficient to have a scanning pattern of normal proportions formed on the film. The mirror system serves to impart slight deflections to the light beam such that motion-picture frames are scanned alternately three and two times in succession. The large mirror in the center is a disk rotating in its own plane with frame frequency, divided into halves, one of which reflects, while the other transmits light freely. The other two mirrors are rocked through small angles during the alternating field periods

when light does *not* reach them from the tube. If we indicate the angular displacement of each mirror by the resulting displacement of the scanning pattern on the film in terms of the pattern height h, we find as successive stationary positions of the mirrors A and B, e.g.:

$$A\ -2h/5;\quad B\ 0;\quad A\ 2h/5;\quad B\ -h/5;\quad A\ h/5;\quad B\ -2h/5;$$
$$A\ 0;\quad B\ 2h/5;\quad A\ -h/5;\quad B\ h/5;\quad A\ -2h/5;\quad \text{etc.}$$

The small rocking motions with a total repetition period of $\frac{1}{6}$ second are imparted to the mirrors by an intermittent, cam-controlled drive coupled to the rotation of the mirror disk.

7.12 Velocity Modulation. Before this survey of pickup devices is closed, another system based on fundamentally different television principles from those considered heretofore must be described. This system utilizes "velocity modulation." All the systems discussed in the preceding two chapters employ a scanning pattern which is independent of the contents of the picture transmitted. Picture reconstruction is effected by varying the brightness of the spot moving over this pattern. With velocity modulation, the brightness of the reproducing spot remains constant, and variation in light intensity over the picture is obtained by varying the velocity of the spot as it moves over the viewing screen. Over bright areas of the picture the velocity of the spot is low, while low light intensities are produced by an extremely rapid motion of the spot. It is quite obvious that, since there must be a geometrical correspondence in the positions of the exploring spots of the transmitter and receiver, both spots must vary in velocity in the same way, that is, the deflecting means of the pickup and reproducing devices must be tied together by the communication channel. The fact that the scanning velocity must be altered rapidly almost completely excludes all mechanical systems from this type of transmission. However, any of the cathode-ray devices can be made to serve. The advantages of this system are twofold: first, theoretically at least, all problems of synchronizing are removed; and, second, the beam at the reproducer is seen at maximum intensity at all times, thus giving a high screen brightness. However, the contrast range is rather limited because it is not possible to move the scanning beam fast enough to reduce greatly the light on the viewing screen. Furthermore, the resolution is not uniform over the picture but is greater in bright areas than in dark. There are also other practical drawbacks; for example, an amplitude distortion in the transmitting or reproducing units results

in a geometrical distortion of the image in addition to false brightness values.

An iconoscope and associated circuits, arranged for pickup as a velocity modulator, are shown in Fig. 7.24. A video amplifier is connected to the signal plate of the iconoscope in the usual way, and the output is supplied to the transmitting unit. In addition, the signal is also applied to the grid of the amplifier tube T_1 in such polarity that the signal corresponding to a dark area makes the grid of this tube more positive, thus increasing the plate current. The horizontal deflecting plates of the iconoscope are connected across the condenser C_1 so that the deflection of the beam is proportional to the charge in the capacitor. This condenser is charged by the plate current of T_1; therefore, the rate of deflection, i.e., velocity of the spot, is proportional to the signal, being great for the signal from a dark area and small for one corresponding to a bright area. The thyratron G_1 is also placed across the condenser, and, when its voltage exceeds a predetermined value, this tube becomes conductive, discharging C_1 and returning the beam to its initial horizontal position. The large discharge current flowing through the resistance R_1 drives the grid of T_2 momentarily positive, allowing a current pulse of definite magnitude to flow into C_2. The condenser C_2 is connected to the vertical deflecting plates, and its value is so chosen that each current pulse produces a vertical displacement of the beam equal to one line spacing. When the charge in C_2 reaches a certain value, it is in turn discharged by the thyratron G_2. It is interesting to note that, if the mosaic of the iconoscope is fluorescent under the beam, a luminous image of the picture being transmitted will be formed.

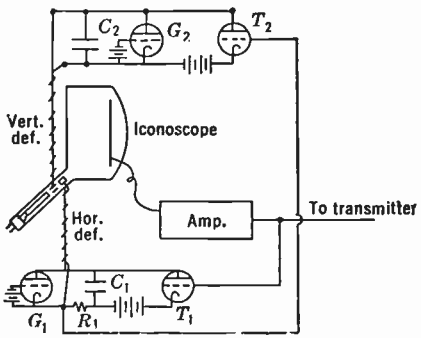


Fig. 7.24. Velocity Modulation Pickup Using an Iconoscope.

The scanning spot at the receiver must merely follow the motion of the spot at the transmitter. By using deflecting circuits similar to those at the pickup, and supplying the control tube with the video signal, this motion can be obtained. Figure 7.25 shows the elements of the circuit for the receiver. The horizontal deflecting circuit differs from that of Fig. 7.24 only in that the discharge of the condenser C_1 is

triggered off by the signal from the transmitter. A special signal may also be required to generate the vertical return.

The video signal generated by the transmitter differs from that produced by intensity modulation. However, the bandwidth required for

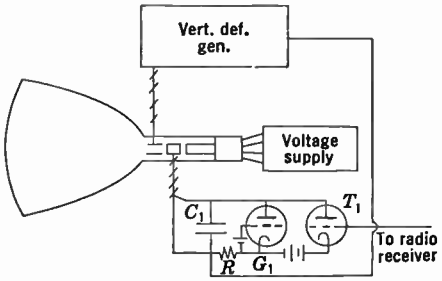


Fig. 7.25. Reproducer of a Velocity Modulation Television System.

a given average resolution is about the same for either system. The actual form of the signal and its spectrum can be determined by means of Fourier analysis, as was done in Chapter 5 for intensity modulation. Figure 7.26 shows the signal for a few simple light-intensity distributions.

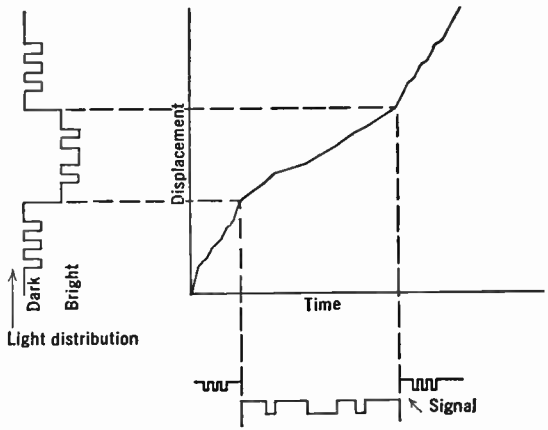


Fig. 7.26. Video Signal Obtained with Velocity Modulation.

Velocity modulation is not a newcomer in the field. As early as 1911, Rosing suggested it as a means of varying the light distribution on the screen of a cathode-ray viewing tube, although he failed to realize that the scanning at the transmitter has to be similar to that

at the receiver if geometrical distortion is to be avoided. Before the general use of cathode-ray pickup tubes, this type of transmission proved successful as a means of pickup from moving-picture film.* The arrangement by which this was accomplished is shown in Fig. 7.27. The spot from a Braun tube was imaged on the film to be transmitted. The light passing through the film was picked up by a phototube. The output of this tube was amplified and supplied to the control tube (equivalent to T_1, Fig. 7.24) of the horizontal deflection. Therefore, the velocity of the light spot scanning the film is a function of density distribution of the film. The picture being transmitted is, of course,

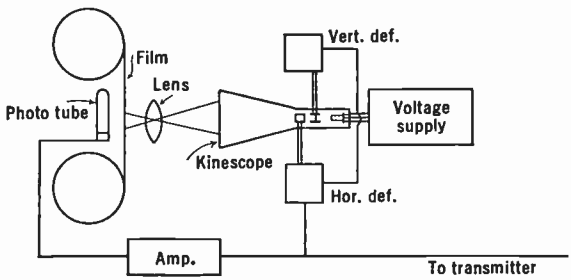


Fig. 7.27. Flying-Spot Film Scanner for Velocity Modulation.

reproduced on the screen of the cathode-ray tube which furnishes the scanning light spot, as well as on the screen of the viewing tube at the receiver.

As has been pointed out, a high contrast ratio cannot be obtained by this system of transmission. To overcome this deficiency, experimental pickup systems combining velocity and intensity modulation, using the latter to reduce the scanning beam current in the viewing tube in order to produce black areas, have been set up. This does not greatly reduce the gain in brightness characteristic of velocity modulation, while permitting high values of contrast.

In spite of its theoretical interest, velocity modulation has not been considered of much practical importance, the gain in brightness at the viewing tube being more than offset by the added complications in equipment, the nonuniformity of resolution, and the difficulty in avoiding geometric distortions due to nonlinear channel characteristics or interference.

This chapter is intended to outline a few of the more important pickup systems. Many pickup systems which, though different in

* See Bedford and Puckle, reference 13.

many important respects, are similar in basic principles were unavoidably omitted. Some of these systems have given very good service in experimental television tests.

Because of the very great theoretical and demonstrated practical advantage of the storage-type pickup device, particularly in the field of direct pickup, its construction, theory, and performance must be examined in more detail. This further consideration will be reserved for Chapters 9 and 10.

REFERENCES

1. A. Dinsdale, *First Principles of Television,* John Wiley, New York, 1932.
2. J. H. Reyner, *Television, Theory and Practice,* Chapman and Hall, London, 1934.
3. F. Schröter (Editor), *Handbuch der Bildtelegraphie und des Fernsehens,* Julius Springer, Berlin, 1932.
4. J. C. Wilson, *Television Engineering,* Pitman, London, 1937.
5. F. Schröter, *Fernsehen,* Julius Springer, Berlin, 1937.
6. George Newnes, Ltd., *Television Today,* Vols. I and II, London, 1936.
7. A. A. Campbell Swinton, "The Possibilities of Television," *Wireless World and Radio Review,* pp. 51–56, April 9, 1924.
8. P. T. Farnsworth, "Television by Electron Image Scanning," *J. Franklin Inst.,* Vol. 218, pp. 411–444, October, 1934.
9. V. K. Zworykin, "The Iconoscope," *Proc. I.R.E.,* Vol. 22, pp. 16–32, January, 1934.
10. V. K. Zworykin, G. A. Morton, and L. E. Flory, "Theory and Performance of the Iconoscope," *Proc. I.R.E.,* Vol. 25, pp. 1071–1092, August, 1937.
11. H. Iams and A. Rose, "Television Pickup Tubes," *Proc. I.R.E.,* Vol. 25, pp. 1048–1070, August, 1937.
12. P. K. Weimer, S. V. Forgue, and R. R. Goodrich, "The Vidicon—A New Photoconductive Television Pickup Tube," *Electronics,* pp. 70–73, May, 1950.
13. L. H. Bedford and O. S. Puckle, "A Velocity Modulation Television System," *J. Inst. Elec. Eng.,* Vol. 75, pp. 63–82, July, 1934.

8 Picture-Reproducing Systems

8.1 Mechanical Scanning. The division of pickup devices into two classes has its counterpart in a similar subdivision of viewing devices, classifying them into mechanical and electronic systems. Because of its historical priority the class of mechanical systems will be considered first.

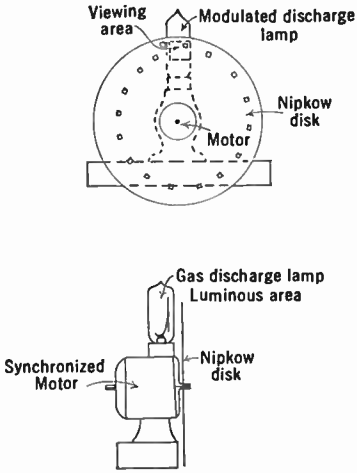


Fig. 8.1. Simple Nipkow Disk Viewer.

The principal components of a mechanical reproducer are the scanning mechanism, the light source, and the means for controlling the light intensity. In the earliest systems the last two elements were combined, that is, the light source was a gaseous glow discharge and its brightness was varied by exciting it directly with the video signal. A Nipkow disk with a spiral row of apertures equal in number to the number of lines making up the picture effected the scanning. The glow lamp was designed in such a way that the discharge was spread uniformly over an area equal to the picture area. This made possible the very simple arrangement shown in Fig. 8.1. Later the use of a crater lamp and suitable optical systems greatly increased the brightness of this type of reproducer.

The frequency response of the glow-discharge type of light source is very much too limited to make this arrangement feasible for high-definition purposes. Furthermore, when a large number of lines are used the amount of light which can be obtained through a simple aperture disk, even with a very intense arc source, is very small, as has already been pointed out.

In order to meet the difficulty resulting from the limited light passed by the scanning disk, a number of optically more efficient scanners

have been devised. The lens disk and lens drum improve the optical efficiency, but they are both expensive to make and critical in adjustment. Mirror and prism drums or screws have been found to meet the necessary optical requirements without the disadvantages of the lens disk and lens drum. Several alternative forms of the simple mirror drum are shown in Fig. 8.2. Each of the drums shown has segments equal in number to the number of lines in the picture and is driven at

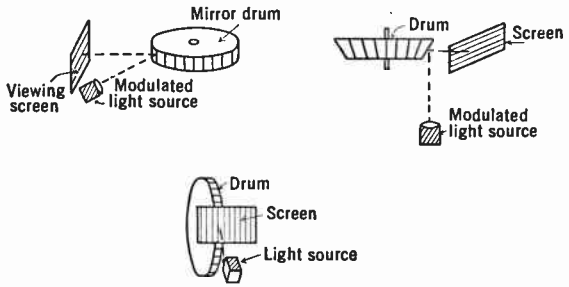


Fig. 8.2. Modifications of the Mirror Drum.

frame frequency. Unless the drum is very large indeed, the mirror segments for high-definition work must be very narrow. Consequently, some increase in optical efficiency can be obtained by introducing combinations of cylindrical lenses instead of the simple spherical lens shown. This lens or combination of lenses images an aperture at the light source onto the viewing screen.

An interesting variation on the mirror drum is shown in Fig. 8.3. The segmented drum is stationary and is made up of internally placed

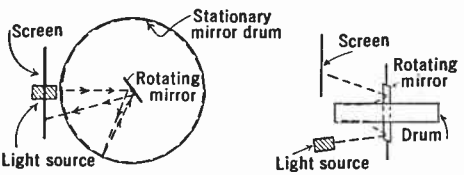


Fig. 8.3. Scanning System Employing a Stationary Mirror Drum.

mirrors. A two-sided plane mirror in the center revolves at frame frequency and produces the scanning action. The size of the mirror drum becomes very large for anything approaching high-definition work. To effect a reduction in size a polyhedral mirror can be substituted for the plane revolving mirror.

The mirror screw is another mechanical scanner for picture reproduction. An array of striplike mirrors is arranged along a shaft in

such a way that their narrow dimension includes the axis of the shaft and their faces are each inclined at a small fixed angle with respect to the preceding mirror. In appearance the assemblage resembles a screw, thus giving the arrangement its name. Figure 8.4 shows diagrammatically such a scanner. The number of mirrors is equal to the number of lines used, and the whole structure is rotated at frame frequency. The picture is viewed directly on the screw, so that the ob-

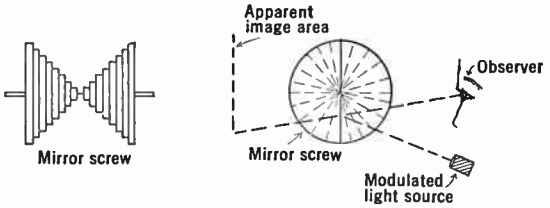


Fig. 8.4. Mirror Screw.

server sees the image in space, apparently located behind the scanner, since it is made up of the direct reflection of the modulated light source as seen in the spinning mirror screw.

Another modification of the mirror-drum principle involves two drums, one rotating at high speed to give line scanning, the other at the much lower speed necessary to produce the displacement of the

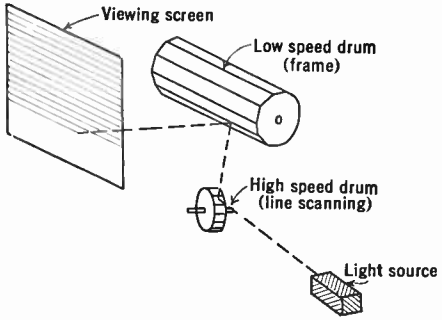


Fig. 8.5. Television Reproducer Using High- and Low-Speed Mirror Drums.

line. When cylindrical lenses are employed this system can be made optically quite efficient. The two drums make it possible so to reduce the size of the equipment that it is no longer totally impractical for high-definition work. A schematic diagram of the arrangement is shown in Fig. 8.5.

8.2 The Kerr Cell. One of the major problems in connection with mechanical viewing systems is the modulation of a light beam of suffi-

cient intensity to give a usable picture. The direct modulation of a gas discharge or arc is immediately excluded, because no known source of this type can respond to variations of the frequency necessary for high-definition pictures. Mechanical light valves (excluding supersonic wave motion in elastic media), such as oscillating mirrors or shutters, suffer from the same limitation. Even the most nearly inertia-free systems of strings or reflectors will not respond to frequencies much above 50 kilocycles, whereas high definition requires nearly 100 times this frequency.

There are several electro-optical effects which suggest themselves as a means of performing the function of light modulation. Upon investi-

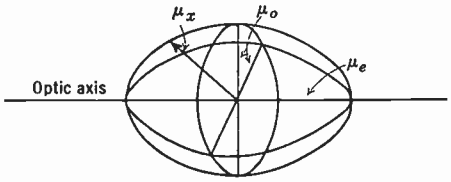


Fig. 8.6. Ellipsoid of Refractive Index for a Uniaxial Birefringent Material.

gation, only the Kerr effect has, in the past, been found to offer anything resembling a practical solution. Therefore, the discussion of electro-optical light valves will be limited to those based on this phenomenon.

Certain organic liquids, when subjected to an electric field, lose their isotropic optical properties and become birefringent. This is known as the Kerr effect. When a beam of polarized light passes through a birefringent material, the velocity with which it traverses the substance depends upon the direction of the electric vector of the incident light. In other words, the index of refraction of the material depends upon the direction of propagation and plane of polarization of the light. The index of refraction may be represented as an ellipsoid, as shown in Fig. 8.6. If this ellipsoid is one of revolution, there is one direction through the material for which the index is independent of the plane of polarization. This is known as the optic axis, and the material may be classified as uniaxial. Substances which exhibit the Kerr effect are uniaxially birefringent under the influence of the field, with the optic axis in the direction of the field.

The mechanism of the effect may be pictured as follows: The molecules of the material (a liquid usually) are polar, that is, each molecule contains an equal positive and negative charge which can be displaced relatively in some preferential direction. Along the direction in which

this displacement can take place, the effective "dielectric constant" is greater than in any other direction. In the absence of a strong constant electric field, the molecules are randomly oriented and the medium is isotropic. An electric field causes a preferred orientation to be superimposed on the statistical randomness of the molecules. This makes the material anisotropic.

If a polarized beam of light passes through an isotropic medium it remains unchanged, both in degree and plane of polarization. For a birefringent material this is not true. A beam of light, polarized in a plane at 45 degrees with respect to the optic axis and directed normal

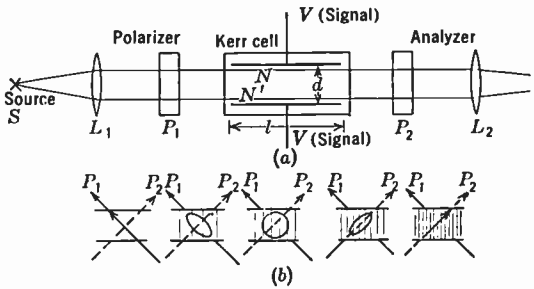


Fig. 8.7. Principle of the Kerr Cell Light Valve.

to the axis, is divided into two components, each of which traverses the medium at slightly different velocity. Upon emerging the two components will, in general, differ in phase, and the light will no longer be plane polarized. If, however, the delay between the two components is half a wavelength, the emerging light will be polarized but in a plane normal to the original plane. The extent of the delay depends upon the medium, the applied voltage, and the length of path. This difference expressed as a phase shift is

$$\phi = 2\pi BlE^2$$

where E is the field strength in the liquid, l the length of the light path, and B a constant, depending upon the medium, known as the Kerr constant.

One way of using this effect in a light valve is shown in Fig. 8.7*a*. Light from the source S is polarized by the element P_1, which may be a Nicol prism, a Polaroid sheet, or similar means, in a plane making an angle of 45 degrees to the electrodes NN' in the cell. The element P_2 is the analyzer and is similar in construction to the polarizer P_1. It passes light polarized in a plane at right angles to that of the light emerging from P_1. The heart of the valve is the cell, which is filled

with a liquid exhibiting the Kerr effect and contains two plane parallel electrodes NN' to which the modulating voltage V is applied.

When no potential is applied to the electrodes, a negligible amount of light passes through the system since in effect it is one consisting of crossed polarizers. However, when a potential is applied between the plates, making the liquid anisotropic, light can pass through the analyzer. The amount of light increases with potential until, when the light is completely polarized in the plane of transmission of the analyzer P_2, half the light entering will be transmitted (half being rejected by the polarizer P_1). A further increase in potential again decreases the light. The light emerging from the cell is represented diagrammatically in Fig. 8.7*b*. Quantitatively, the amount of light emerging I is given by

$$I = \frac{I_0}{2} \sin^2 \frac{\phi}{2}$$

where I_0 is the light entering the system. This, in terms of the cell constants, is

$$I = \frac{I_0}{2} \sin^2 \pi \frac{BlV^2}{d^2}$$

$$= \frac{I_0}{4}\left(1 - \cos 2\pi \frac{BlV^2}{d^2}\right)$$

where V is the potential between plates and d the spacing between them. This expression has its maximum value when

$$V_{\max} = d\sqrt{\frac{1}{2Bl}}$$

An idealized curve showing the variation in light with applied voltage is shown in Fig. 8.8. Unfortunately, the Kerr effect is independent of neither temperature nor wavelength of light. Therefore, not only must the cell be kept at constant temperature, but also some correction must be made for the dispersion of the cell. A mica half-wave plate can be made to serve the latter purpose.

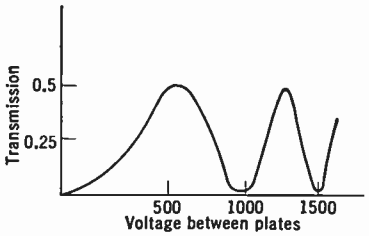


Fig. 8.8. Light Transmission Characteristic of a Typical Kerr Cell.

The Kerr constant B for a number of liquids is given in Table 8.1. From the magnitude of the Kerr constant

TABLE 8.1. KERR CONSTANT OF LIQUIDS

Substance		$B \times 10^6$ (esu)	Resistivity	Dielectric Constant
Carbon disulphide	CS	0.36	13×10^{15}	2.6
Chloroform	$CHCl_3$	0.32	5×10^7	5.0
Acetone	C_3H_6O	1.6	8×10^6	21.0
Nitrobenzene (purified)	$C_6H_5NO_2$	41.0	10^{10}	36.0

it is evident that, even with a nitrobenzene cell, the spacing between the plates must be very small or the voltage needed across the plates will be enormous. For example, a cell 5 centimeters long will have a maximum voltage required for the change from complete extinction to maximum transmission:

$$V = d\sqrt{\frac{10^6}{2 \cdot 40 \cdot 5}} \times 300$$

It can be seen that, with $d = 1.0$ millimeters, about 1500 volts are required. If in operation the cell is biased to 1100 volts, the peak control voltage is some 400 or 500 volts. The narrow spacing and great length of the cell mean that the optical aperture of the system will be very seriously limited. To overcome this, Karolus suggested a multiple cell. A cell of this type is illustrated in Fig. 8.9. The capacity, together with the high-control voltage, makes this cell rather impracticable because of the very great power required to drive it. Power at video frequency is both difficult to obtain and expensive. In fact, this fault excludes the multiplate cell from practical high-definition work.

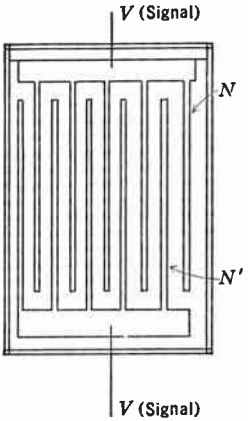


Fig. 8.9. Karolus Multiple Plate Kerr Cell.

The Wright cell is a somewhat more practical solution. Figure 8.10 illustrates this type of cell. If an image of the scanning aperture is focused on the center of the cell, the system may be made fairly efficient optically.

As has been pointed out, half of the light is lost at the first polarizer. Several systems have been suggested to avoid this loss of light, only one of which will be discussed. The arrangement is shown in Fig. 8.11. Instead of the initial polarizer a calcite crystal divides the incoming light into two parts polarized at right angles and slightly displaced, owing to the two indices of refraction. A second calcite replaces the

analyzer. This still further separates the two beams, so that they fall onto the mask on either side of the slit. If the Kerr cell between the plates causes a shift of 90 degrees in the plane of polarization of each ray, the two rays will be bent in the second crystal in such a way that they meet upon emerging and pass through the slit. At lower control

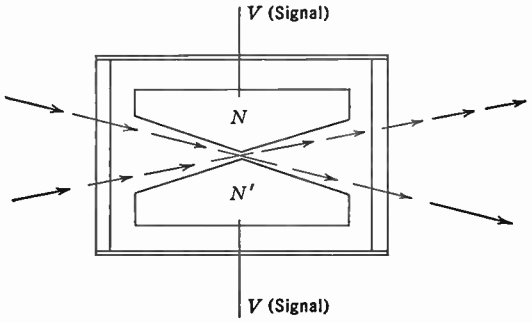


Fig. 8.10. Wright Wedge Kerr Cell.

voltages, when the cell only elliptically polarizes the light, only a fraction will pass through the slit.

If there were no alternatives, the Kerr cell would offer a possible means of obtaining a high-definition picture. However, the power

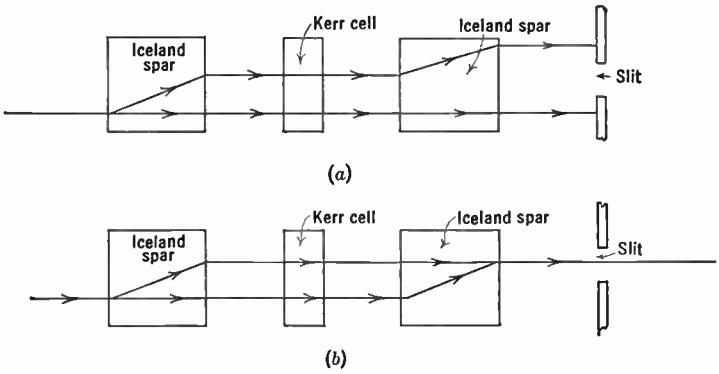


Fig. 8.11. Modified Kerr Cell Light Valve Utilizing Nearly All Incident Light.

required from the video system, the limitations it imposes upon the optical arrangement, together with other less important objections such as nonlinearity of control, render the system less suited to the present requirements than other arrangements, in particular, electronic viewing devices.

8.3 Supersonic Light Valve. Another, radically different, light valve has also been used in television reproduction. The action of this valve is based upon the diffraction of light by supersonic waves in a transparent medium. The method of exciting supersonic waves in liquids by a piezoelectric crystal was developed by Langevin about 1917, and, furthermore, Brillouin at an early date had considered mathematically the scattering of light by thermal motion treated as superimposed elastic waves. It was not until 1932, however, that the diffraction of light by supersonic waves was observed experimentally.

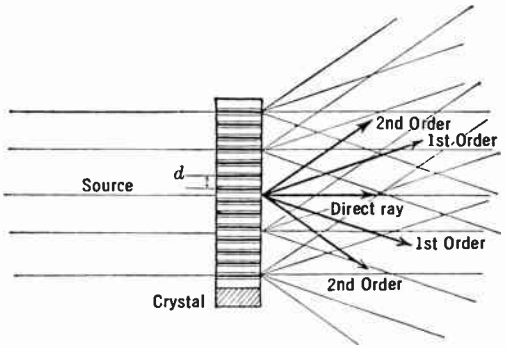


Fig. 8.12. Diffraction of Light by Supersonic Waves.

The phenomenon was shown independently by Debye and Sears in this country and by Lucas and Biquard in France. Their observations indicated that, if plane waves were excited in a trough containing a liquid such as carbon tetrachloride, benzene, kerosene, or glycerin, a light beam parallel to the wave fronts would be efficiently diffracted. The experimental arrangement is shown in Fig. 8.12. The waves in the liquid are made up of alternately compressed and rarefied regions, and consequently the index of refraction of the liquid varies sinusoidally in correspondence with these compression waves. If the velocity of propagation of the elastic wave is v, and the excitation frequency f, the separation between successive high-index regions (i.e., wavelength) is $v/f = d$. The light which passes through the regions of high index is delayed slightly more than that through the lower-index portions, so that the emerging light has a phase varying periodically normal to its direction of propagation, or, considering surfaces of constant phase, these surfaces are corrugated. On the basis of this picture of the action of the cell it is possible, with the aid of Huygens' principle, to calculate the intensity and position of the diffracted beams and the intensity

of the direct or central beam. The positions of the diffracted beams are found to be given by the relation

$$n\lambda = d \sin \theta$$

where n is the order of the diffracted ray, λ the wavelength, and θ the angle measured from the direction of the central beam. The intensity of the light going into the diffracted rays depends upon the amplitude of the supersonic waves, and, with sufficient amplitude, practically all the light can be thrown from the central ray into the scattered beams. It is interesting to note that the position of the orders of diffracted light is unchanged by any change in amplitude of the elastic waves.

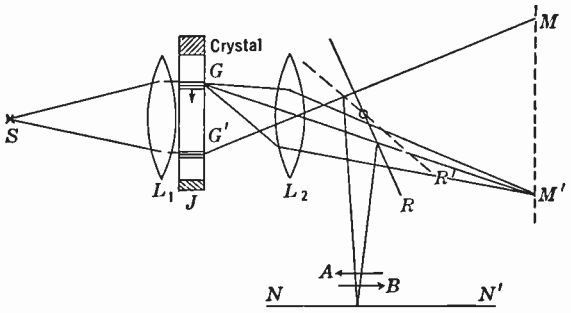


Fig. 8.13. Optics of the Supersonic Light Valve; Method of Arresting the Motion of the Image of a Picture Element.

A very ingenious application of this effect to the problem of television reproduction was made by Jeffree in 1936. A supersonic light valve based on this principle forms the nucleus of the Scophony system.* The cell consists of a fairly long trough filled with a suitable liquid and having a quartz piezoelectric crystal at one end. At the other end there is a layer of absorbing material such as glass wool so that the compression waves are not reflected back. When the crystal is made to oscillate, traveling waves move down the trough at a velocity v which depends upon the density and elastic constants of the liquid. In practice a liquid giving a velocity of about 1000 meters per second is used, while the normal frequency of the crystal is in the neighborhood of 10 megacycles, resulting in a wavelength of $10^5/10^7 = 0.01$ centimeter. A cell of this type is shown in Fig. 8.13.

If the crystal is driven for a few cycles, a wave group will move through the trough at the propagation velocity v. In the figure, G and G' indicate two positions of such a group as it moves up the trough.

* See Lee, reference 8.

The trough is illuminated from the source S, which in conjunction with the lens L_1 produces a parallel beam of light through the cell. Lens L_2 images the cell into the plane MM', but the mirror R intercepts the light from this lens and forms the actual image on the screen NN'. By an optical arrangement which will be explained below, only the light diffracted by a disturbed region is permitted to pass through the system. Therefore, if the mirror is held stationary the image of the wave group G will appear as a bright area moving across the screen NN' in the direction indicated by the arrow A. However, if the mirror rotates in such a direction that the image of a stationary point on the cell moves

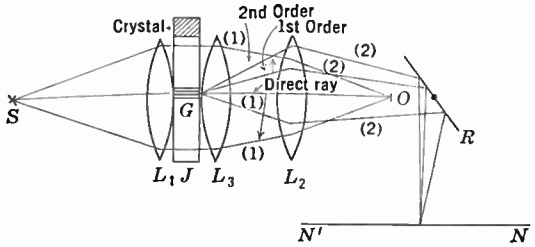


Fig. 8.14. Optics of the Supersonic Light Valve; Method of Controlling Light Transmission.

with equal velocity in the opposite direction across the screen, as indicated by the arrow B, the bright image of the wave group will appear stationary.

In order to remove the direct beam and use only the diffracted light produced by periodic variations in the index of refraction of the liquid, the optical arrangement may be as shown in Fig. 8.14. The combination of lenses L_1 L_2 L_3 images the source S onto the shield O. Undisturbed rays 1 indicate this image formation. If the fluid is undisturbed, all the light entering the system is stopped by this shield, so that the screen is entirely in darkness. A disturbed region in the cell, such as G, causes some of the light to be thrown into diffracted rays. These rays (represented by rays 2 in the diagram) are not stopped by the shield, but instead are brought together by lenses L_3 and L_2 on the screen NN' described above.

Used in a television system, an oscillator drives the crystal at resonance, usually, as has already been mentioned, at a frequency around 10 megacycles. The amplitude of this oscillation is modulated with the video signal in such a way that the signal corresponding to a black area leads to zero amplitude, whereas a bright area gives large amplitudes. If the instantaneous distribution of amplitudes in the cell is

considered, it will be seen that they record the video signal for a period of time equal to l/v, where l is the length of the cell. By making the length of the cell within the optical system equal to the product of the velocity of propagation and the line period, the distribution corresponding to a complete line will be included. Therefore, the optical system images a light distribution corresponding to an entire picture line on the screen. As the distribution moves through the cell, the light image is held stationary by the motion of the mirror. In an

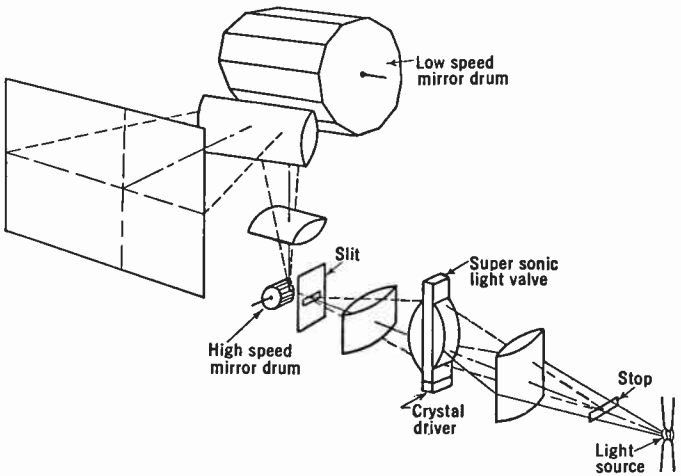


Fig. 8.15. Schematic Diagram of a Scophony Television Receiver Employing a Supersonic Light Valve.

actual television reproducer the moving mirror takes the form of a small mirror drum, so that as one face carries the image off the screen an adjoining face brings another image onto the screen. A second mirror drum provides scanning motion at right angles to the direction of the lines.

The complete optical system of a reproducer using this system is shown in Fig. 8.15. In this arrangement the shield has been moved up to the light source and a slit is located at the position previously occupied by the shield. However, it will be seen that the action is the same as described above. The small line scanning mirror drum must run at a very high speed. For example, if this drum has 20 faces it must be driven at more than 47,000 revolutions per minute for a 525-line picture. Therefore, it must be very accurately balanced and have the minimum practical moment of inertia. It will be noticed that cylindrical lenses rather than the usual spherical lenses are used

throughout the optical system. Such elements greatly reduce the size of the equipment necessary to obtain a given optical aperture.

The supersonic light valve gives in one sense a storage of the information of the picture elements over a period of a line. However, it does not lead to the full light gain that would be expected from an unrestricted storage system. The limitation is due to the fact that light must pass through the cell in an essentially parallel beam. As a consequence of this limitation, it is evident that a light valve which has the same area as the supersonic cell, but which controls the transmission simultaneously over its entire area, can be made to produce

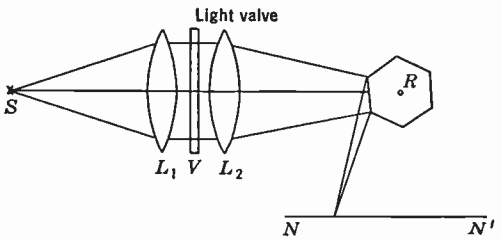


Fig. 8.16. Large Area Light Valve.

the same illumination on the viewing screen. Such a system is illustrated diagrammatically in Fig. 8.16. Only one mirror drum is shown, although two are actually required. For the sake of economy of size in the moving parts, cylindrical lenses would be used. Similarly, a much smaller cell which would control light entering from all directions could be made to produce the same results.

However, in the absence of these ideal valves, the supersonic cell, in spite of its complications, has shown itself to be by far the most efficient arrangement to be used in connection with the mechanical scanner for the reproduction of television pictures. The amount of light that can be obtained is sufficient, when a high-intensity moving-picture type of arc serves as a light source, so that the image formed on a 5 by 6 foot screen can be viewed without discomfort.

8.4 Light-Valve Cathode-Ray Tubes. In present-day television the reproduced image is almost exclusively presented on a phosphor screen cathode-ray tube. Before discussing this type of reproducing device another application of the cathode-ray tube for picture presentation will be treated, namely the cathode-ray-operated light valve. This type of device is attractive because the light intensity is not directly limited by the power capabilities of the scanning electron beam. Here the light is modulated with a two-dimensional light valve in which the

transmission of each picture element is controlled with a cathode-ray beam.

Several methods suggest themselves for this purpose. Of the four which will be described briefly below, only two have progressed to the point of practical utilization and only one of these in the field of television proper.

An obvious technique, which has not been realized in practice, consists of the employment of the longitudinal analogue of the Kerr effect.

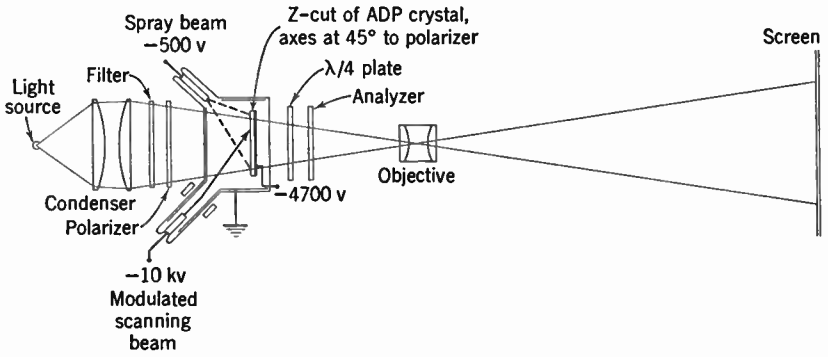


Fig. 8.17. Principle of Television Projector Employing Crystalline Light-Valve Plate.

If certain uniaxial crystals, such as ammonium dihydrogen phosphate ($NH_4H_2PO_4$ or ADP), are subjected to an electric field parallel to their optic axis, the velocities of propagation of the ordinary and extraordinary rays, which are identical in the unexcited state, become unequal. The retardation produced is proportional to the applied voltage and independent of the thickness of the crystal. For green light (5460 Å) a retardation of 180 degrees requires a voltage of 9400 volts.

A crystal plate of ADP might be employed as light valve in a television projector as indicated in Fig. 8.17. Light from a mercury arc provided with a filter for the green line is plane polarized and directed on the crystal plate mounted in a cathode-ray tube. The back of the plate is provided with a transparent conductive coating which is biased by 4700 volts with respect to the anode coating (here indicated as grounded) of the tube. The front of the crystal plate is scanned by a high-velocity beam modulated with the picture signal. The beam tends to drive the surface of the crystal negative by an amount proportional to the modulating signal. Between scannings a low-velocity spray beam returns the surface to anode potential ("shuts the light

valve"). Beyond the tube is placed a quarter-wave plate so oriented as to compensate the retardation produced by the crystal plate for zero modulation. The analyzer, crossed with respect to the polarizer, the projection lens, and the viewing screen complete the optical assembly.

The arrangement shown is by no means unique. The d-c bias on the crystal plate and the compensating quarter-wave plate serve primarily to shift the operating point of the system into the range of maximum sensitivity. Some of the inherent weaknesses are obvious. Thus, for a high light efficiency, the signal power must be very great and the voltage variations over the plate so high that they may be difficult to maintain and will distort the scanning pattern. Furthermore, as with the supersonic light valve, only small apertures may be employed, since the retardations depend on the direction of passage through the crystal. Finally, since the magnitude of the retardation depends also on wavelength, a monochromatic source must be employed.

A second system, which is free from many of the complications of the crystal light valve, but possesses others in their stead, uses a suspension of opaque flakes suspended in a liquid to control the passage of light. Graphite flakes suspended in oil may serve as example. The flakes, which are randomly oriented in the absence of a field, tend to line up with their flat surfaces parallel to the field direction when a field is applied. Their projected area on a plane normal to the field, which determines the effective absorption of a light beam parallel to the field by the suspension, is accordingly reduced. Quantitatively, the change in transmission is proportional to the square of the applied field. As the field is removed, the randomness of the particle orientation is rapidly restored by thermal agitation.

An experimental arrangement for utilizing the suspension as television light valve is shown schematically in Fig. 8.18. The scanning surface here is a mica plate, 0.010 to 0.020 inch thick, which forms one of the cell electrodes and also a part of the cathode-ray tube envelope. The back electrode of the cell is a second mica plate with a transparent conducting coating. The thickness of the suspension itself is 0.020 to 0.050 inch.

The sequence of charge and discharge is similar to that in the preceding example. The charge is accomplished by a high-velocity scanning beam modulated with the video signal, which charges the scanned mica surface to a negative potential. A low-velocity beam returns the surface again to anode potential. Either a separate spray beam or the scanning beam, reduced in velocity, defocused, and unmodulated,

may be used for this purpose. The reduction in velocity is most simply accomplished by lowering the potential of the anode and back electrode (and hence also that of the scanned surface). More specifically, if the initial anode potential is $V_a > V_b$, V_b being the upper unity yield

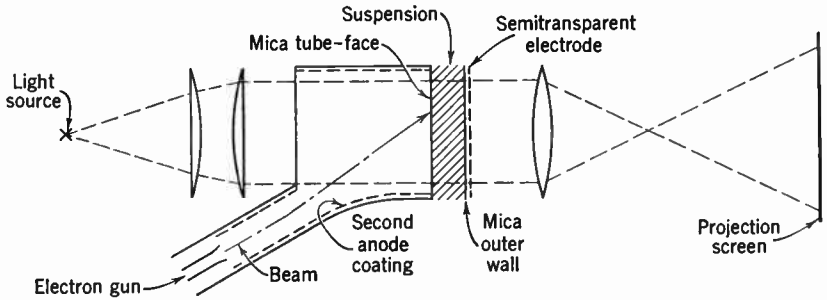


Fig. 8.18. Suspension Light-Valve Projector. (Donal, reference 19; courtesy of *Proceedings of the Institute of Radio Engineers.*)

point on the secondary-emission yield curve of the scanned surface, the voltage of anode and back electrode may be shifted to V_c (Fig. 8.19). Here $V_d < V_c < V_b$, where V_d is the lower unity yield point of the scanned surface, and, preferably, also $V_c - V_d > V_a - V_b$. Under

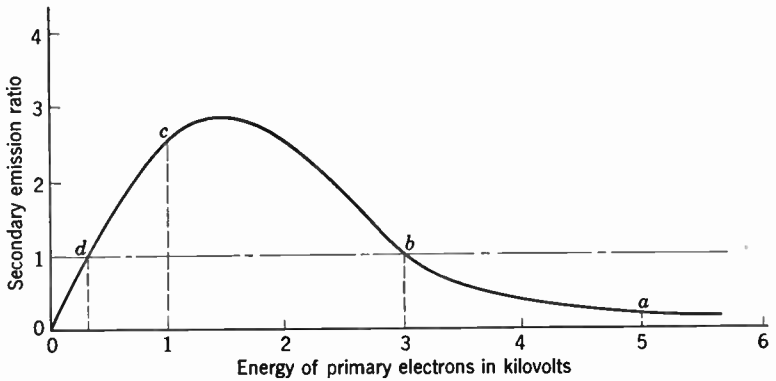


Fig. 8.19. Secondary Emission of Target at Operating Points of Light-Valve Projector. (Donal, reference 19; courtesy of *Proceedings of the Institute of Radio Engineers.*)

these conditions the low-velocity beam returns the surface to anode potential regardless of the magnitude of the charging signal.

It might be inferred from the above that, if the image is scanned, after each high-velocity scanning, with a low-velocity beam of ade-

quate strength, the image is completely wiped out after every frame. This is true, however, only if the suspension is a perfect insulator. In

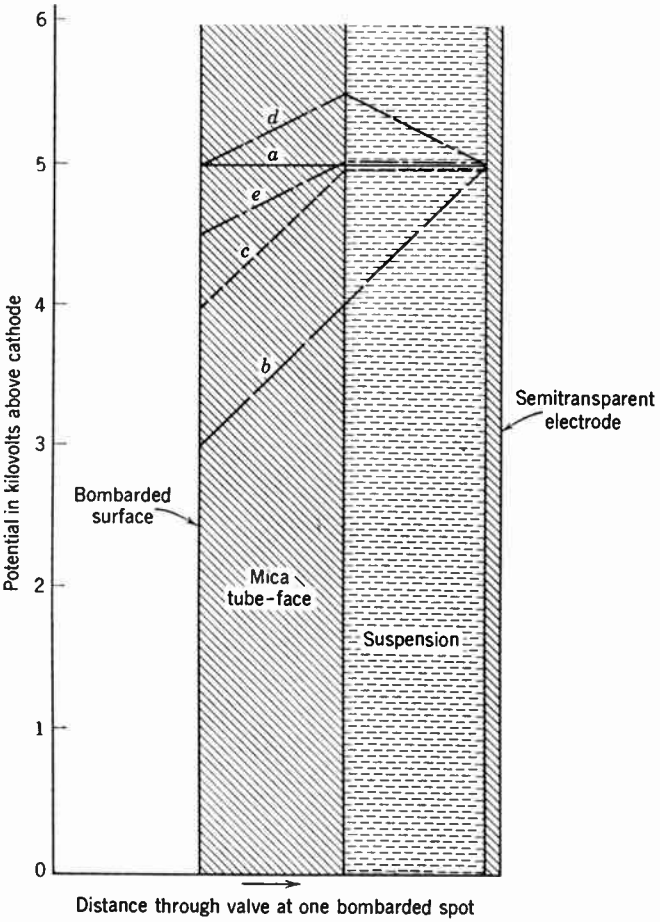


Fig. 8.20. Potential Variation in Target Plate and Suspension at Various Stages of the Scanning Process. (*a*) Before charging. (*b*) Immediately after charging. (*c*) After polarization following charging. (*d*) Immediately after discharging. (*e*) After polarization following discharging. (Donal, reference 19; courtesy of *Proceedings of the Institute of Radio Engineers*.)

practice, the cell is polarized by conduction and hence becomes opaque approximately 0.01 second after charging. As shown in Fig. 8.20, discharge actually produces a reverse field across the cell of lesser strength, resulting in a second flash of light of reduced intensity. The original

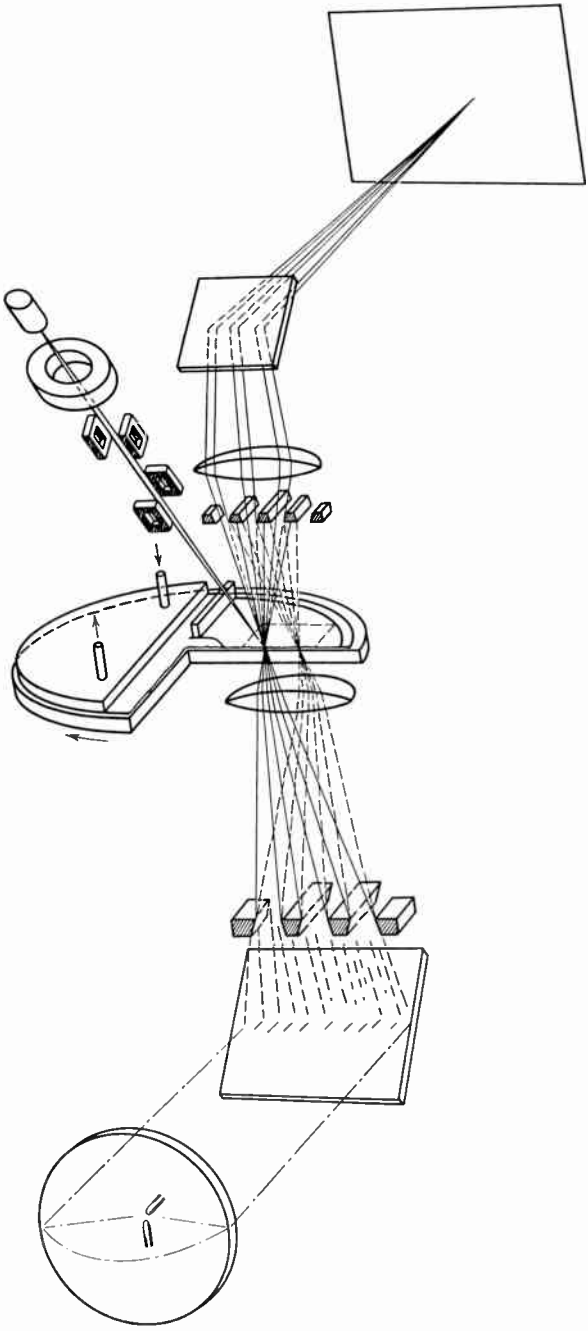


Fig. 8.21. Principle of the Eidophor Projector. (Thiemann, reference 22; courtesy of *Bulletin de l'association suisse électriciens.*)

image is thus completely wiped off only after a number of cycles. Actually the decay is rapid, however, since the transmission of the cell varies as the square of the field within it.

Although the possibility of forming television images with the suspension valve has been demonstrated, it has not found practical application in television so far. On the other hand, the Eidophor system developed by Fischer and Thiemann at the Technische Hochschule in Zurich has been found capable of projecting large-screen theatre television pictures. Only a brief description can be given here of this system whose realization involved the overcoming of formidable technical difficulties.

The essential elements of the Eidophor system are shown in Fig. 8.21. Its characteristic feature is the imaging of a brightly illuminated field containing a series of parallel bars through the light valve onto a second plane containing a set of complementary bars. The light valve itself is imaged, through the second mask, on the viewing screen. It consists of a glass disk coated with a layer of viscous liquid, the Eidophor (= image bearer), which is scanned by the modulated beam. If the surface of the liquid is flat, as it is if the charge deposited by the beam does not vary from point to point, no light whatsoever is transmitted. On the other hand, if the beam is modulated, the variation of charge over the surface gives rise to electrostatic forces which give it a wavy contour. This distorted surface thus acts as a grating which diffracts a portion of the incident light so that it may pass through the openings in the second mask (Fig. 8.22). The intensity of the transmitted light depends on the amplitude of the surface distortion, the control characteristic, on the relative dimensions of the lattice constant (i.e., the frequency of modulation) and the grid spacings in the masks.

The mechanical and electrical characteristics of the Eidophor liquid must be chosen so that the distortions imparted to the surface decay in approximately one frame period. A more rapid decay results in low optical efficiency, since the valve is opened by the beam for only a small fraction of the total time; and a slower decay, in picture lag and an inadequate reproduction of moving objects. Furthermore, gross charge variations over the picture area are to be avoided, since they would result in slowly decaying low-frequency deformations. Accordingly, a scanning beam of uniform intensity is employed and the signal, modulating a carrier frequency which considerably exceeds the maximum

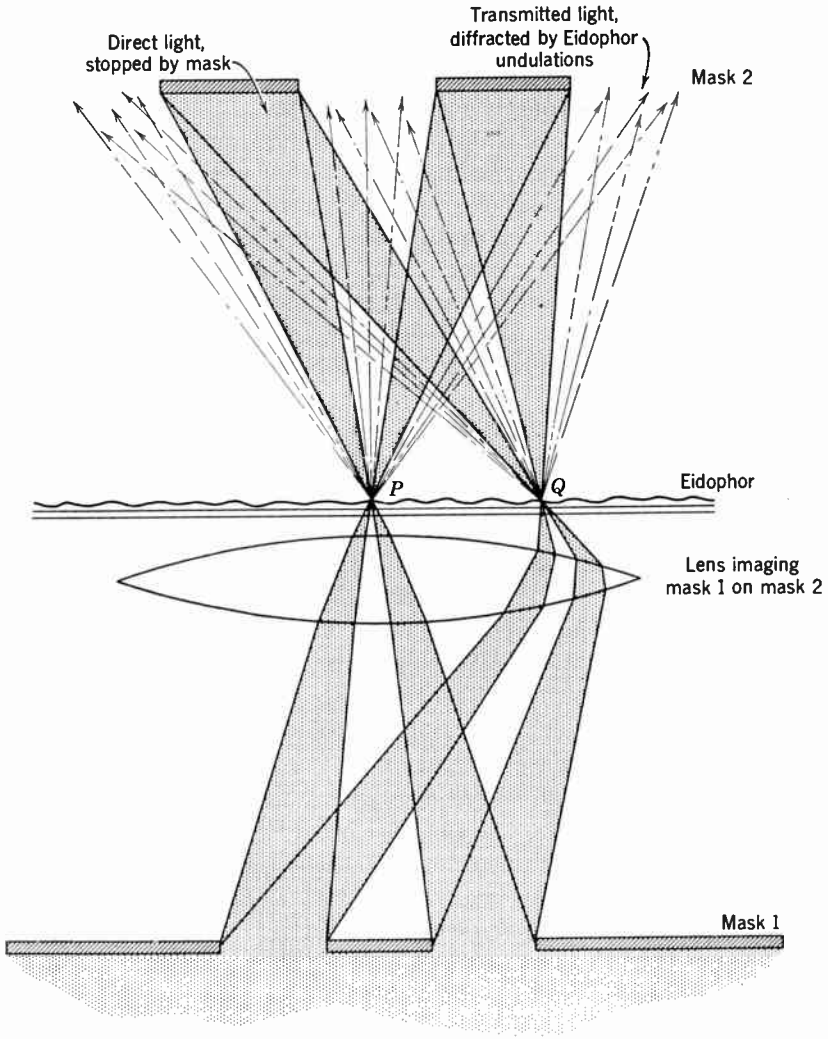


Fig. 8.22. Mechanism of Light Transmission by Eidophor Light Valve.

frequency of the signal, is applied to horizontal beam deflecting plates. The horizontal deflection is thus of the form

$$d = v_h t + A \sin \omega t$$

where v_h is the normal scanning velocity along a horizontal line and A is proportional to the video signal. With this type of scanning the

charge depositions and consequent distortions of the surface vary in amplitude with the video signal but have a uniform fundamental wavelength v_h/ω; the average charge density in an area of the order of a picture element or larger is uniform throughout.

The layer of liquid is maintained in an unaltered condition, and distortions at the edge of the scanning pattern are minimized by rotating the supporting glass disk, which is provided with a transparent metal deposit as conducting coating, at a very slow rate as compared to the frame frequency. After leaving the scanning area the liquid comes into contact with thermostated metal cooling plates and, before re-entering the scanning area, is flattened by a smoothing bar. The motion is so slow that it does not noticeably reduce the sharpness of the reproduced image.

It is evident that special demands are placed on the scanning beam by this system: The spot must be materially narrower than the fundamental wavelength of the Eidophor distortions in the line direction, i.e., at most one-fourth of a picture element width. It is defined by a slit, 0.013×0.1 square millimeters in size, which is placed in the crossover plane of the gun. The slit cheeks are of tungsten since they are exposed to very intense bombardment. The rectangular spot shape, with the long side perpendicular to the line, yields the maximum total intensity consistent with the high resolution required in the line direction. The cathode itself is novel in design. It is formed by the flattened end of a tungsten rod aligned on the gun axis. The rod is heated partly by radiation from a coaxial tungsten helix and partly by electron bombardment from the same helix. The helix is heated and rendered negative with respect to the cathode rod by pulses restricted to the horizontal return time, so that the electron-optical imaging system is not disturbed by the magnetic field of the heating currents.

A few remarks are in place regarding the general performance of the Eidophor system. With a beam voltage of 20 kilovolts and a beam current of about 20 microamperes the system shown in Fig. 8.21 was able to achieve the required spot size of about 0.02×0.1 square millimeters and to attain the necessary modulation of the Eidophor surface. The required deflecting signal amplitude, applied to a special pair of deflecting plates, was only of the order of 1 volt. The maximum light transmission of the light valve was 20 percent, not counting reflection losses: the application of nonreflecting films on all glass surfaces between the light valve masks is not only desirable from the point of view of efficiency, but also essential to achieve adequate contrast; light scat-

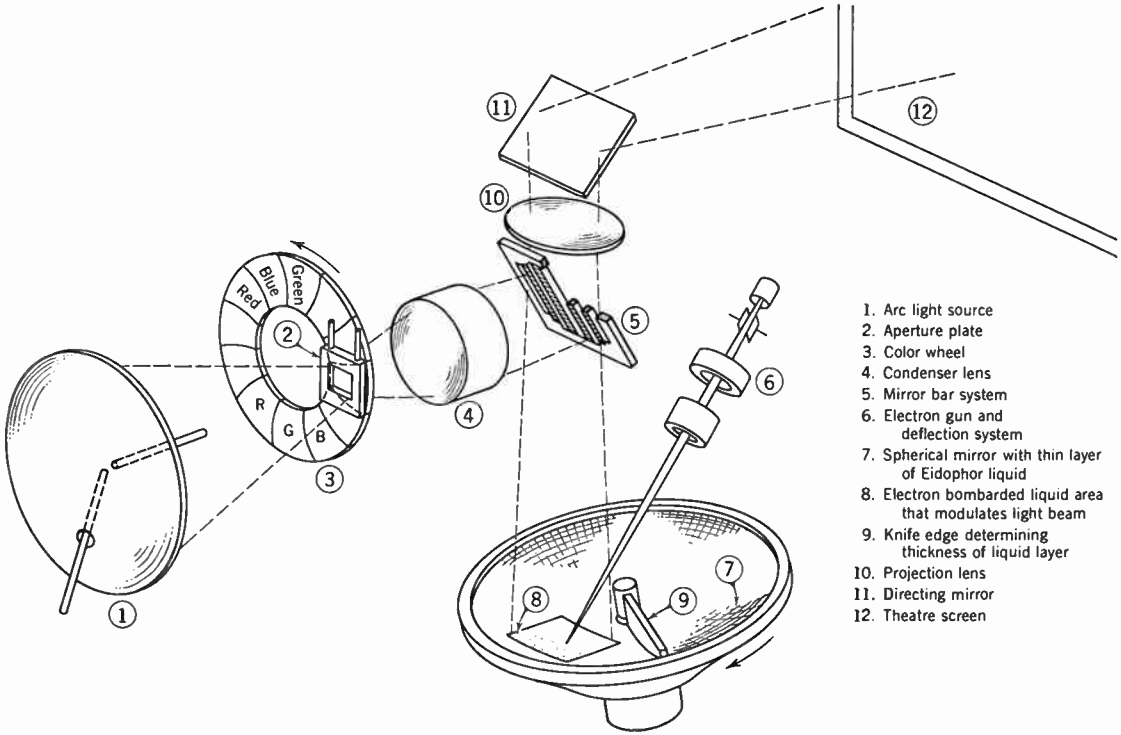


Fig. 8.23. Diagram of Folded Eidophor System. (Courtesy of Twentieth Century-Fox Film Corporation.)

tered within the system limits the darkness of the shadows which can be achieved in the picture.

More recently a great gain in compactness has been achieved by "folding" the optical system as indicated in Fig. 8.23. The reflecting bar mask, inclined by 45 degrees to the axis of the illuminating system, is placed at the center of curvature of a spherical mirror coated with the Eidophor liquid. In the absence of beam modulation the light inci-

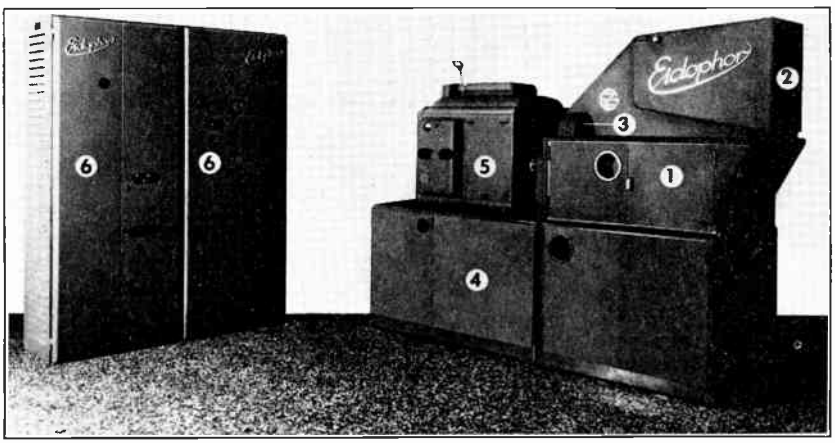


Fig. 8.24. Eidophor Large-Screen Color Television Projection Equipment. (1) Eidophor projector. (2) Projection light beam hood. (3) Color wheel. (4) Auxiliary services (vacuum pump, thermostat, and system for Eidophor cooling). (5) Projection lamp (Ventarc type). (6) Television receiver circuits. (Courtesy of Twentieth Century-Fox Film Corporation.)

dent on the spherical mirror is reflected onto the mirror bars and returned by them to the illuminating system. On the other hand, light which is deflected by Eidophor corrugations passes through the slots between the mirror bars and is utilized for forming the picture on the viewing screen. The color wheel shown in the diagram would, of course, be omitted in an Eidophor projector designed for black and white pictures. However, with the beam current raised to 70 microamperes (and a beam voltage of 17 kilovolts) even color pictures of acceptable brightness have been obtained on a screen 15×11.5 square feet in size. The external appearance of the complete projection equipment is shown in Fig. 8.24. The Eidophor system has been shown capable of producing excellent projected pictures, though obviously only at considerable cost in equipment and operating skill.

A quite different type of cathode-ray light valve tube which has found some application in radar technique, where the required storage

times are generally longer than television, is the dark-trace screen tube or Skiatron (Fig. 8.25). The light-valve action here rests on the fact that certain crystalline materials, such as potassium chloride (KCl), are darkened by electron bombardment. The scanning beam traces a negative image on the screen which, illuminated by an auxiliary source, is projected on the viewing screen. Obviously, a picture signal of reversed polarity is required to yield a positive reproduction of the image.

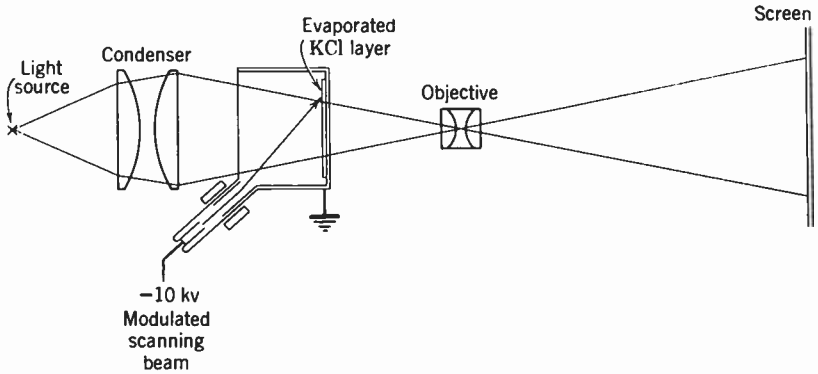


Fig. 8.25. The Skiatron (Dark-Trace Cathode-Ray Tube) as Television Projector.

The darkening of the crystalline screen by the incident electrons is interpreted as the production of color centers, i.e., the transfer of electrons from the negative chlorine ions of the potassium chloride lattice to chlorine ion vacancies. Electrons in such vacancies are capable of excitation by, and hence of the absorption of, visible light. Even in the absence of light, they are eventually thermally excited and return to neutral chlorine atoms to reform the original lattice ions and restore the transparency of the crystal. This process may be accelerated by raising the temperature of the crystalline layer or applying an electric field which tends to "sweep out" the electrons in the color centers. Thus Rosenthal * mentions that in a single crystal of potassium chloride a voltage of 600 volts applied across a 1-millimeter layer will clear the opacity produced by an electron beam in the course of $\frac{1}{30}$ second, i.e., a television frame period. Actually thin evaporated films of potassium chloride, giving a mauve coloration, or of 75 percent potassium chloride and 25 percent potassium bromide, giving a neutral gray, were employed. The persistence of the thinner screens proved to be, at room temperature, $\frac{1}{20}$ to $\frac{1}{15}$ second, without the application of an external

* See Rosenthal, reference 9.

field. High beam potentials (e.g., 10 kilovolts) proved advantageous for high contrast and absorption. Figure 8.26 shows the control characteristics of two potassium chloride layers employed in conjunction with a Wratten No. 59a filter, whose transmission is approximately complementary to that of the potassium chloride color centers. The maximum beam current corresponded here to 400 microamperes.

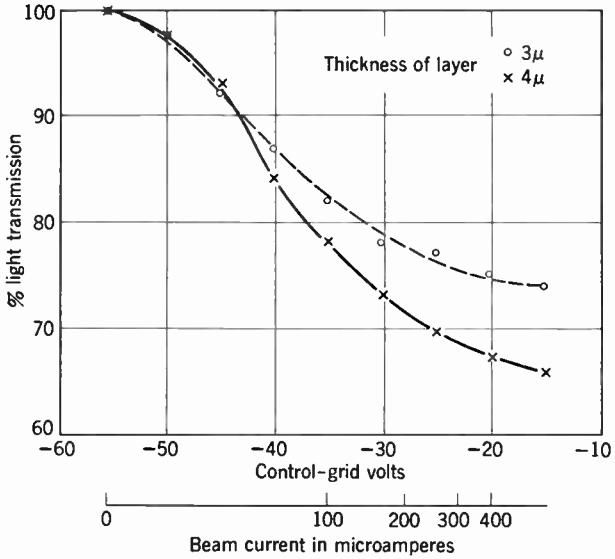


Fig. 8.26. Control Characteristics of KCl Targets in Combination with Wratten 59a Filter. (Rosenthal, reference 9; courtesy of *Proceedings of the Institute of Radio Engineers.*)

8.5 Phosphor-Screen Cathode-Ray Tubes. In the field of image reproduction the purely electronic viewing device has had outstanding success. By its use all mechanical moving parts are eliminated and the problem of producing and controlling light of the required intensity is greatly simplified. Consequently, the cathode-ray viewing tube is used almost exclusively today in practical television receivers. The history of the developments leading up to the modern cathode-ray tube is one with that of electronics itself, and as such would require many pages to present even in outline form. Therefore, it must be omitted from this necessarily brief account.

The first proposal to use a cathode-ray tube as a means of reproducing television images appears to have been made by Boris Rosing, a year or two before Campbell Swinton's suggestion of applying cathode

rays to the problem of pickup. Before the end of the first decade of the twentieth century, Rosing was able to demonstrate the synthesis of simple images by means of a Braun tube. The Braun tube is a modification of the Crookes discharge tube, arranged as shown in Fig. 8.27 to produce a narrow pencil of cathode rays which can be deflected by means of electrostatic plates or electromagnetic coils to any point on the fluorescent screen forming the front wall of the tube. The electrons forming the beam in this tube are produced by the ionization of gas in the vicinity of the cathode. An aperture defines the beam while another aperture, in conjunction with the deflection plates, serves to control the current reaching the screen. The scanning pattern is gen-

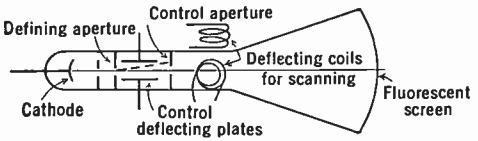


Fig. 8.27. Rosing's Electronic Viewing Tube.

erated by supplying sawtooth current waves of suitable periodicity to the two mutually perpendicular deflecting coils. This relatively primitive cathode-ray tube would not, of course, fulfill present-day requirements. The fine spot needed for 525 lines could not be obtained, control of the beam is inadequate, and high deflection frequencies are impossible with this type of gas discharge. However, it served to demonstrate the principle.

As the technique of picture transmission advanced and the standards of requirements were raised, greater demands were put upon the viewing tube. However, the improvement of the cathode-ray tube more than kept pace with these developments. Today it provides, in black and white, 525-line pictures suitable either for direct viewing or for projection on a large screen. Furthermore, it has been found capable of meeting even more stringent demands when and where needed.

The essential elements of the phosphor-screen cathode-ray tube as used for television reproduction are a means for producing a fine pencil of electrons, usually known as the electron gun; an arrangement for deflecting the electron beam in two mutually perpendicular directions; and a screen coated with luminescent material. These elements are assembled in a glass bulb which is evacuated to as low a pressure as is practicable. A typical example of a tube of this class is shown in Fig. 8.28. When the tube is operated the synthesized picture appears

on the large screen forming the front of the bulb, either to be viewed directly or else projected onto a suitable large viewing surface. This picture has the line structure characteristic of the scanning process used in television. The lines are produced by the motion of the bright spot which results from the cathode-ray beam striking the fluorescent materials of the screen. The motion is obtained by bending the electron stream close to the end of the gun. Either an electrostatic or magnetic field is suitable for producing the deflection. Two fields at right angles to the axis of the gun are employed in either case. These fields may be superimposed or may be sequential. The field strength of one is made

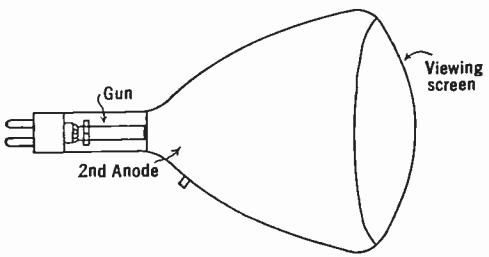


Fig. 8.28. Schematic Diagram of Direct-Viewing Kinescope.

to vary periodically at line frequency, whereas the other has a periodicity corresponding to frame (or field) frequency. The variation is such as to deflect the beam at a uniform rate across the screen and return it rapidly to its original position. The superposition of the two displacements results in a parallel line scanning pattern of the type described in Chapter 5.

The brightness of the reproducing spot is varied by controlling the current in the scanning beam. This control must be effected without altering either position or size of the spot, and for practical reasons should not require large amounts of video power.

The importance of the electron gun is evident. It is responsible for supplying to the screen all the power which is needed to produce the light. In order to obtain this power not only must the electrons be accelerated to a high velocity in the gun, but the beam current also must be quite large. Tubes employed in direct-viewing receivers require 6000 to 16,000 volts, home projection receivers demand 20,000 to 30,000 volts, and theatre projection tubes may be operated with as much as 80,000 volts or more. The beam currents in all types are of the order of several hundred microamperes, attaining peak values of

one or two milliamperes in high-power projection tubes. This power must be delivered into a very small spot if the resolution of the image reproduced on the screen is not to be limited by the cathode-ray tube. For a 525-line picture the current must be concentrated into a spot which is about a $\frac{1}{500}$ part of the vertical height of the picture in diameter—in other words, for a 12-inch direct-viewing tube, about 0.015 inch, whereas in a projection tube it may be as small as one-fifth of this value.

The gun consists of an electron source and an electron-optical arrangement for concentrating the cathode rays into a fine beam. The electron-optical system includes an element for controlling the current in the beam. The many forms of electron guns may be divided, broadly, into two classes: namely, those having purely electrostatic lenses, and those in which magnetic and electrostatic lenses are combined. Choice of the class of gun for a given cathode-ray tube depends upon the particular use intended and upon the operating conditions.

Most guns have a two-lens electron-optical system. The first lens is very close to the cathode and acts in such a way as to force the emitted electrons into a narrow bundle, usually called the "crossover." The control grid, which usually is part of the first lens, governs the flow of electrons into the crossover by regulating the field in the immediate vicinity of the cathode. The crossover, which is the narrowest region of the beam within the gun, is imaged on the fluorescent screen by the second lens. This second lens may be either magnetic or electrostatic, depending upon the class of gun. The theory and design of an efficient electron gun are far from simple, and their importance warrants a more complete discussion, which will be given in Chapter 12.

The luminescent screen on which the picture is reproduced is made by coating the inside face of the tube with a fluorescent material in powder form. The nature and synthesis of the fluorescent material itself have been discussed in Chapter 2; the special problems involved in applying these materials to the glass wall will be dealt with in Chapter 11. Needless to say, the layer of fluorescent material must be uniform, free from contamination, and of the correct thickness. The deposition of a thin, continuous aluminum film over the phosphor layer is of particular value in projection tubes, where it improves the efficiency of energy conversion into light as well as the contrast of the picture.

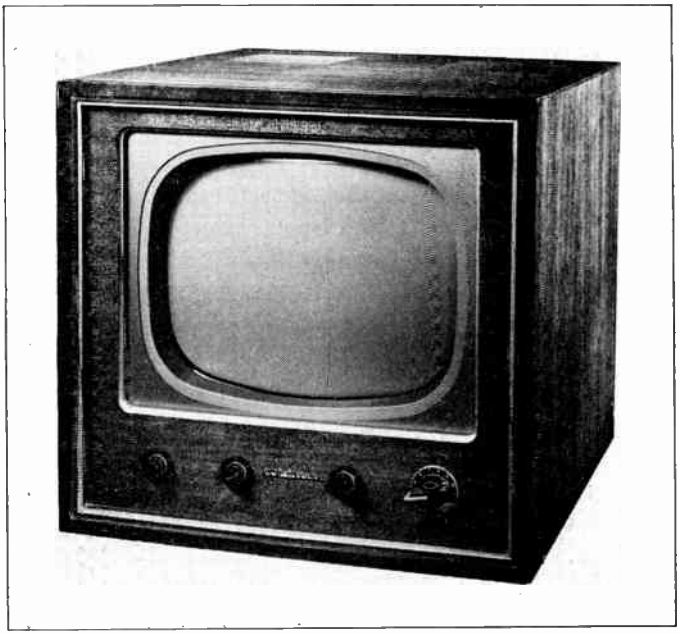


Fig. 8.29. Table Model Television Receiver (RCA Victor Division).

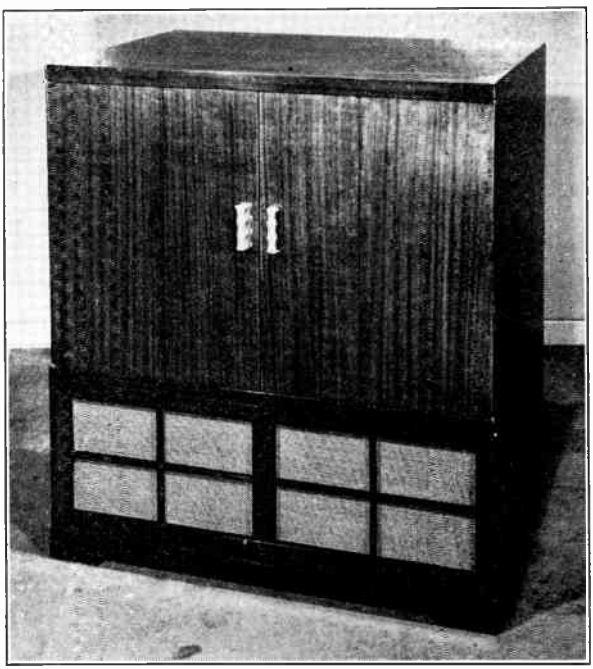


Fig. 8.30. Console Television Receiver (RCA Victor Division).

Present-day screens universally provide a black-and-white picture. Early television workers had to content themselves with the green and black reproduction characteristics of natural willemite. Fortunately

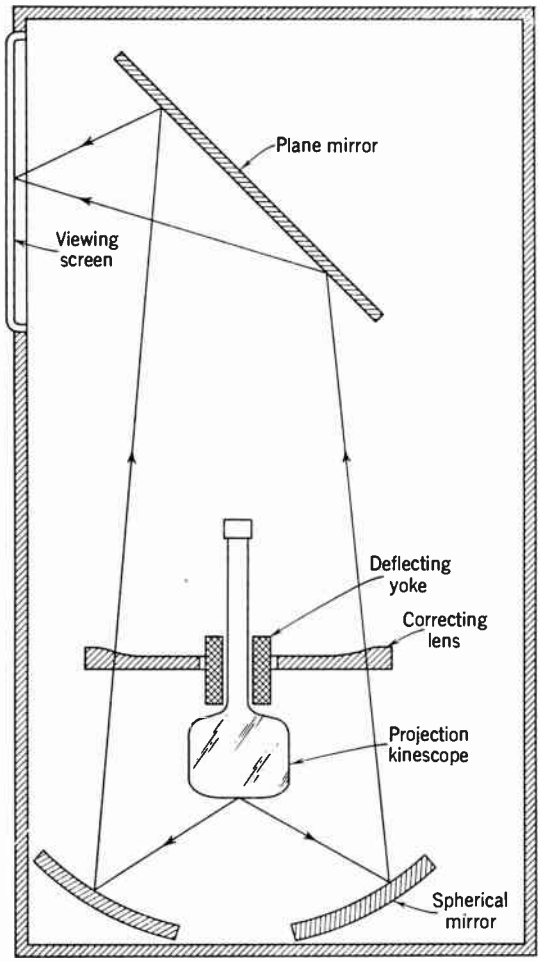


Fig. 8.31. Optical System of RCA Home Projection Receiver.

for the development of the art, this naturally occurring material has a fairly high efficiency in converting electrical power into light; otherwise the acceptance of the electronic system would have been greatly delayed.

Tubes employed in home receivers generally have screens ranging from 5 to 30 inches in diameter. On a 16-inch screen the picture is

approximately 11 by 14½ inches. In practically all direct-viewing receivers the tube is mounted horizontally; Fig. 8.29 shows a typical table model receiver. Console receivers, such as that shown in Fig. 8.30, commonly provide space in addition for broadcast and short-wave radio receivers, phonograph turntables, etc. The brightness of the

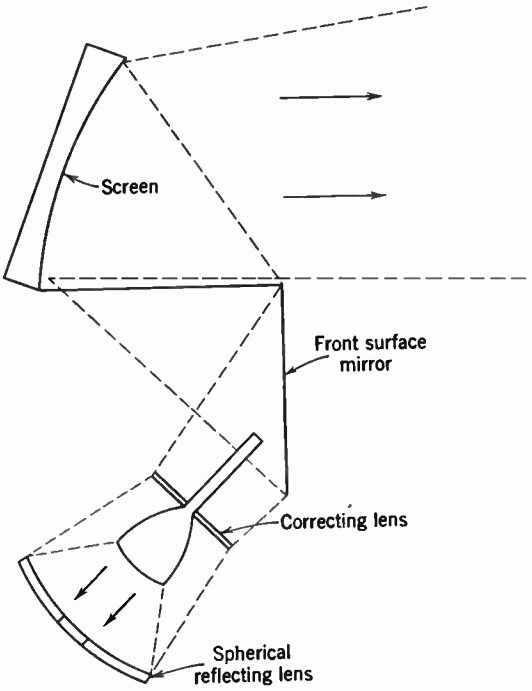


Fig. 8.32. Optical System of Philco Home Projection Receiver. (Bradley and Traub, reference 14; courtesy of *Electronics*.)

image reproduced on the screen of the viewing tube ranges from 10 to 25 foot-lamberts and can be comfortably viewed in a normally lighted room.

The physical size of the picture which can be obtained directly on the cathode-ray tube is limited by practical considerations of tube size. Even for a 16- or 20-inch tube the forces which the tube envelope must withstand are enormous. Before the engineering problems involved in the manufacture of very large tubes were overcome, projection tubes, with 4- to 5-inch tube faces, were widely employed, in conjunction with high-efficiency optical systems, to produce the largest home-receiver television pictures. Figures 8.31 and 8.32 show, schemat-

ically, two home projection receivers providing 15- by 20-inch pictures. Both employ a spherical mirror and an aspheric correcting plate, as well as an intermediate plane mirror, to project the image on the viewing screen. In the first example this screen is a beaded directional transmission screen; in the second, a metallic reflecting screen, which

Fig. 8.33. Home Television Projection Receiver (RCA Victor Division).

achieves the desired directional characteristics by curving and grooving. The external appearance of a home projection receiver is shown in Fig. 8.33.

A reflective projection system of similar character, which exceeds an $f/2$ projection lens by a factor of 10 in efficiency, has been employed with great success in theatre television. To comply with the brightness standards of motion-picture practice (7 to 14 foot-lamberts) for a picture 15 by 20 feet in dimension, it becomes necessary to scale up all the parameters of the projection system. The television projector shown in Fig. 8.34 employs a 25-inch mirror and a 7-inch projection kinescope operated at 80 kilovolts to achieve the desired result.

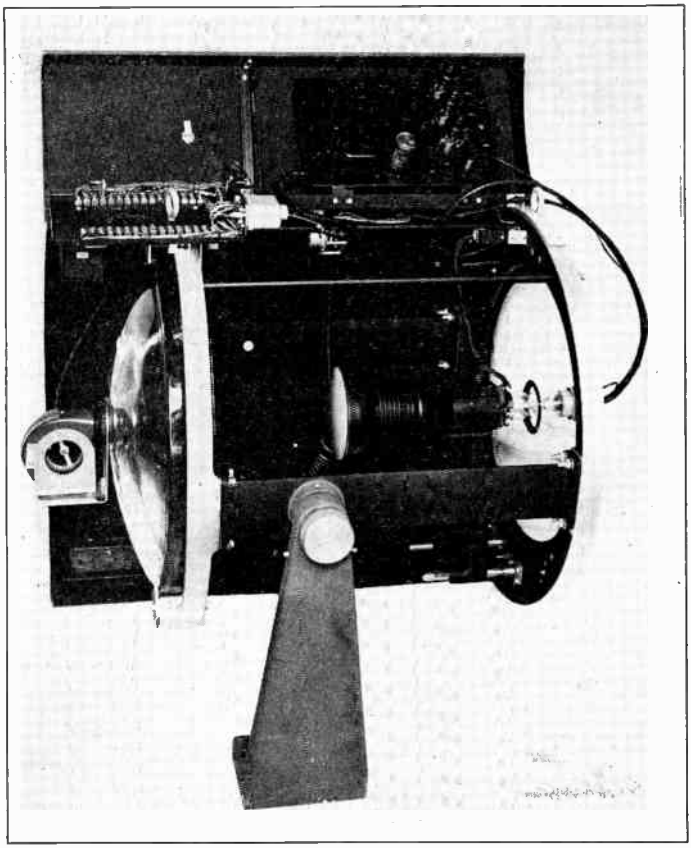


Fig. 8.34. Theatre Projection Receiver (RCA PT-100 Ultraspeed Projection System).

8.6 Viewing Tube with Electronic Storage. A system which combines features of the light valve tubes with image reproduction on a phosphor screen has been proposed by Schröter * (Fig. 8.35). The essential feature of the method is the creation, by the scanning beam, of a potential pattern on an insulating surface which controls the photoemission from the photocathode of an image tube; the television picture is reproduced, in this manner, on the phosphor screen of the image tube.

In greater detail, a beam of uniform intensity scans a highly insulating target after passing through a closely spaced modulated grid. The modulated grid itself consists of an array of very fine wires, approximately 0.0004 inch in diameter. The beam energy is chosen so

* See Schröter, reference 26.

that the secondary emission ratio of the insulator is greater than 1. As a result the scanned point of the insulating surface is brought, by the beam, to a potential slightly positive with respect to the momentary potential of the control grid. If the surface is initially more negative, it is charged positively by secondary emission until just as many electrons leave the insulator as arrive from the beam; on the other hand, if it is originally more positive, all the secondary electrons are returned to the insulating surface until the latter has been made so negative that the same balance between incident and emitted electrons is achieved.

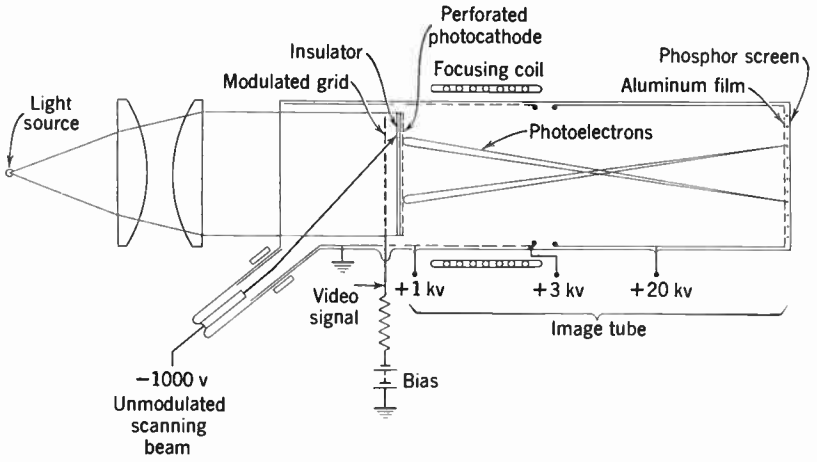


Fig. 8.35. Schröter Storage Viewing Tube.

The other side of the insulated target is coated with an electrically connected photoemissive grid which is illuminated by a constant light flux. The uncoated portions of the target act here as a coplanar grid, suppressing photoemission to the extent in which they are negative relative to the photocathode. Actually, a negative target potential of about 7 volts is found adequate to completely suppress the photoemission for the surface gradient, at the photocathode, of about 1000 volts per centimeter. The image-tube phosphor itself may be maintained, by an electron-permeable aluminum film, at a potential of the order of 20 kilovolts.

One of the obvious advantages of a tube of the type described is the complete absence of flicker. The potential pattern laid out by the beam in one frame scan remains unaltered, and hence also the image intensity on the viewing screen unchanged, until the next scan. This might permit a reduction in the frame period and a consequent econ-

omy is required bandwidth. A second advantage is that the ultimate limit in the light obtainable from a phosphor screen is increased in this method since, for equal total electron current incident on the screen, the *current density* at any point is very greatly reduced. It is too early to judge to what extent these advantages may be realized in practice.

REFERENCES

1. J. C. Wilson, *Television Engineering,* Pitman, London, 1937.
2. F. Schröter (editor), *Fernsehen,* J. Springer, Berlin, 1937.
3. D. G. Fink, *Principles of Television Engineering,* McGraw-Hill, New York, 1940.
4. V. K. Zworykin and R. E. Shelby, "Recent Television Developments," *Physical Society Reports,* Vol. 8, pp. 135–185, 1941.
5. George Newnes, Ltd., *Television Today,* London, 1936.
6. L. M. Myers, *Television Optics,* Pitman, London, 1936.
7. I. G. Maloff and D. W. Epstein, *Electron Optics in Television,* McGraw-Hill, New York, 1938.
8. H. W. Lee, "The Scophony Television Receiver," *Nature,* Vol. 142, pp. 59–62, July 9, 1938.
9. A. H. Rosenthal, "A System of Large-Screen Television Reception and on Certain Electrical Phenomena in Crystals," *Proc. I.R.E.,* Vol. 28, pp. 203–312, May, 1940.
10. V. K. Zworykin, "Description of an Experimental Television System and the Kinescope," *Proc. I.R.E.,* Vol. 21, pp. 1655–1673, December, 1933.
11. Max Knoll, "Electron Optics in Television," *Z. tech. Physik,* Vol. 17, pp. 604–617, 1936.
12. V. K. Zworykin and W. Painter, "Projection Kinescope," *Proc. I.R.E.,* Vol. 25, pp. 937–953, 1937.
13. D. W. Epstein and I. G. Maloff, "Projection Television," *J. Soc. Motion Picture Engrs.,* Vol. 44, pp. 443–455, June, 1945.
14. W. E. Bradley and E. Traub, "A New Television Projection System," *Electronics,* Vol. 20, pp. 84–89, September, 1947.
15. I. G. Maloff, "Optical Problems of Large-Screen Television," *J. Soc. Motion Picture Engrs.,* Vol. 51, pp. 30–36, 1948.
16. R. V. Little, Jr., "Developments in Large-Screen Television," *J. Soc. Motion Picture Engrs.,* Vol. 51, pp. 37–46, 1948.
17. H. Jaffe, "Electrooptic Effect in Crystals of the Primary Phosphate Type," *Phys. Rev.,* Vol. 73, p. 95, 1948.
18. B. H. Billings, "The Electro-Optic Effect in Uniaxial Crystals of the Type XH_2PO_4," *J. Opt. Soc. Am.,* Vol. 39, pp. 797–808, October, 1949.
19. J. S. Donal, Jr., "Cathode-Ray Control of Television Light Valves," *Proc. I.R.E.,* Vol. 31, pp. 195–208, May, 1943.
20. J. S. Donal, Jr., and D. B. Langmuir, "A Type of Light Valve for Television Reproduction," *Proc. I.R.E.,* Vol. 31, pp. 208–214, May, 1943.
21. F. Fischer and H. Thiemann, "Theoretical Considerations on a New Process for Large-Screen Television Projection," *Schweizer Archiv f. angew. Wiss. u. Technik,* Nrs. 1 and 2, pp. 1–23, 1941.

22. H. Thiemann, "Large-Screen Television Projection by the Eidophor Process," *Bull. assoc. suisse électr.,* Vol. 40, pp. 585–595, 1949.
23. "Eidophor Projector for Theatre TV," *Tele-Tech,* Vol. 11, pp. 57, 112–113, August, 1952.
24. H. W. Leverenz, "Luminescence and Tenebrescence as Applied in Radar," *RCA Rev.,* Vol. 7, pp. 199–239, June, 1946.
25. L. T. Sachtleben and G. L. Allee, "Ultraspeed Theater Television System," *J. Soc. Motion Picture and TV Engrs.,* Vol. 57, pp. 425–433, 1951.
26. F. Schröter, "Picture Storage Problems," *Bull. assoc. suisse électr.,* Vol. 40, pp. 564–566, 1949.

PART 3

Component Elements of an Electronic Television System

9 The Iconoscope

The iconoscope, as has been noted in Chapter 7, was the first pickup device to utilize the storage principle. Because of the effective utilization of this principle and in spite of its photoelectric inefficiency and other losses of signal which will be discussed below, it was the first television pickup device with sufficient sensitivity for a practical system. During the early years of commercial television, it was essentially the only camera tube available and carried the entire burden of program transmission from television studios, outdoor scenes, and motion-picture film. Because of the historical position of this tube and because it is a good example of the application of the storage principle, this chapter has been taken over from the earlier edition of this book. A section has been added, however, bringing the description of the iconoscope and its construction up to date.

Within the last few years the newer tubes, such as the orthicon and image orthicon, have made the iconoscope nearly obsolete for most live pickup, both of outdoor scenes and in the studio. At the time of writing, the iconoscope is still used for film reproduction. There is reason to believe that it will soon be obsolete even for this application and that the Vidicon will assume this role. The next chapter describes in detail the newer pickup tubes which have now almost completely taken over the function of producing a video signal from the scene before the camera.

In Chapter 7 a brief description was given of the iconoscope. It was described as consisting of a photosensitive mosaic and an electron gun assembled in a highly evacuated glass envelope. The gun is an electron-optical system serving to produce a fine pencil of cathode rays which is made to scan the sensitized side of the mosaic by means of a suitable magnetic, or electrostatic, deflecting system. The mosaic in the normal iconoscope is a very thin mica sheet covered on one side with a vast number of minute silver globules, photosensitized, and insulated from one another, and coated on the other with a metal film known as the signal plate. This metal film is coupled, on the one hand,

to the silver elements by capacity through the mica and, on the other, to the video amplifier through a signal lead sealed into the bulb.

The optical image is projected onto the silver-coated side of the mosaic. Each silver element, being photoemissive, accumulates charge by emitting photoelectrons. Thus, information contained in the optical image is stored on the mosaic in the form of a charge image. The scanning beam sweeping across the mosaic in a series of parallel lines releases the charge from each element in turn and brings it to equilibrium, ready to start charging again. The change in charge in each element induces a similar change in charge in the signal plate and, consequently, a current pulse in the signal lead. The train of electrical impulses so generated constitutes the picture signal.

9.1 Construction of the Iconoscope. The appearance of a typical iconoscope can be seen from Fig. 9.1. The shape and size of the blank containing the iconoscope elements are matters of expediency, and, though the design illustrated appears to have certain advantages, many alternative forms have been successfully used.

The main body of the tube must be large enough to house a mosaic of suitable size. It will become apparent as the discussion proceeds that the magnitude of the picture signal is a function of the mosaic area. Therefore, the mosaic should be as large as is permitted by practical considerations of the associated optical system. Both the scanning beam and the optical images must have unobstructed access to the entire surface of the mosaic. Optical considerations require that the axis of the lens system which projects the picture coincide with the central normal from the mosaic surface. Hence, directly over the face of the mosaic there is an optical window free from imperfections that might distort the picture. The gun is placed at an angle to the normal through the center of the mosaic instead of coinciding with it as it would there interfere with the optical system. In the tube illustrated, the gun is mounted in the tubular arm forming the handle of the dipper-shaped blank. The optical window is the curved shell capping the blank on the gun side.

Two side tubes are required in the processing of the iconoscope, one for the admission of cesium, the other for the purposes of exhaust. The second tube is often located in the press at the end of the gun arm, for reasons which will be given later. The cesium tube may enter the blank at any convenient point, although to insure a uniform distribution of the alkali over the photosensitive surface a location behind the mosaic is advisable.

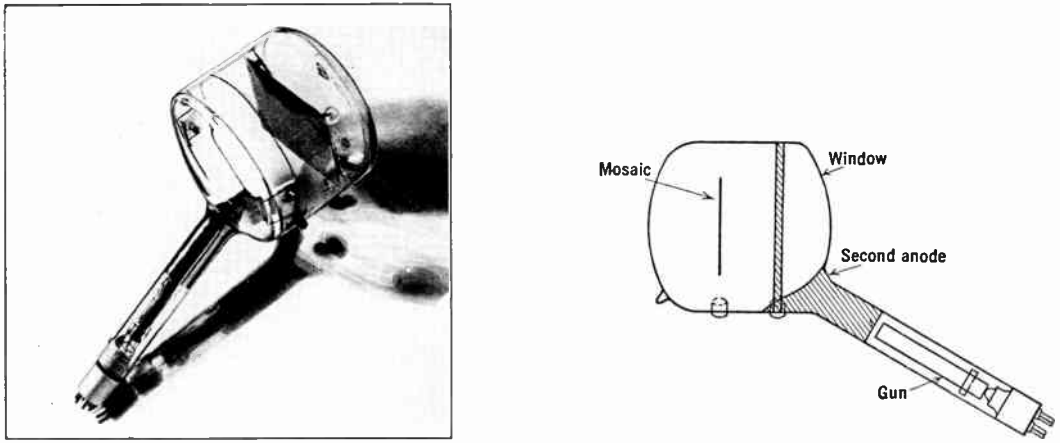


Fig. 9.1. Photograph and Diagram of an Iconoscope.

Several leads must be provided, of course, to carry the required potentials to the tube elements. Besides the wires in the gun press, two other leads are sealed into the walls of the tube, one to carry the picture signal, the other connecting the second anode. Additional leads may be necessary when special activation schedules, such as silver sensitization, are to be used.

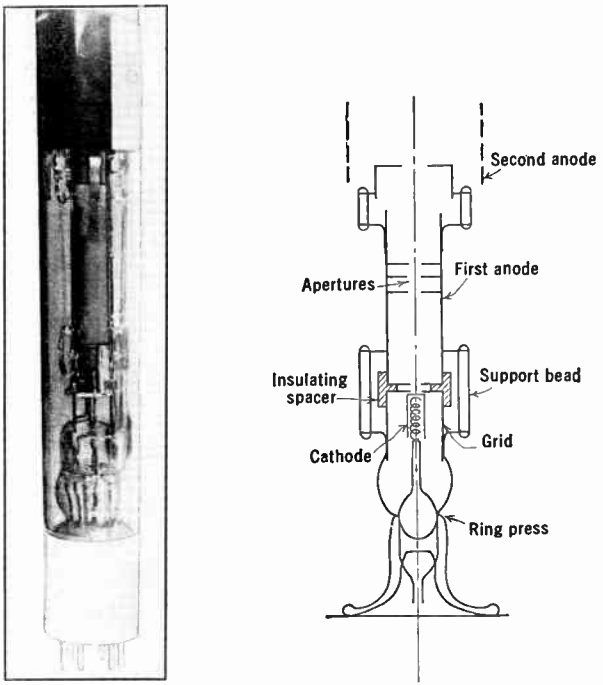


Fig. 9.2. Iconoscope Gun.

The mosaic should be fastened rigidly in the blank in such a way that it will not warp or become misaligned during the processing. The mounting is purely a structural problem which admits of a number of solutions. In the tube shown in Fig. 9.1 the mosaic is supported on the metal cross member that can be seen in the illustration.

The gun structure shown in Fig. 9.2 is a characteristic example of the type used in the iconoscope. It consists of a thermionic cathode, a control grid, and an insulating spacer, all mounted on a six-wire ring press together with the first and second anode cylinders. The general method of assembly and the approximate dimensions of this gun can be seen from the figure. A small, closed, nickel cylinder serves as

thermionic cathode and is indirectly heated by an internal tungsten heater. The cathode material is the usual barium and strontium oxide and coats the closed end of the cathode cylinder. The gun elements, all of which are cylindrical and mounted coaxially, can be assembled on an appropriate jig and spot-welded in place.

Essentially, the gun is a two-lens electron-optical system, one lens being located close to the cathode, the other between the first and second anode. The first lens is a multi-element system consisting of the cathode, control grid, and first anode. As is evident from its name, the control grid serves to govern the current in the scanning beam. An accelerating grid, when employed, minimizes the effect of changes in the first-anode voltage on the beam current so that the focus of the gun may be varied without altering the current to the mosaic. Between the first and second anode a second lens is formed which serves to image the narrow section of the beam near the cathode, known as the crossover, onto the mosaic. Three apertures are mounted in the first anode cylinder in such a way as to form the aperture stop for the second lens and to mask off the secondary electrons emitted by the first anode itself. The second anode is divided into two parts. The first is a cupped washer attached to the main gun assembly at the end of the first anode; the rest is in the form of a metallized coating on the inside of the gun arm and extending into the main body of the tube. Since, as will be described later, the second anode also serves to collect electrons emitted by the mosaic, it also surrounds the mosaic, usually in the form of a narrow ring. A more detailed account of the electron gun will be found in Chapter 12.

9.2 The Mosaic. The mosaic, as has been explained, consists of a dielectric sheet coated with a surface having high photosensitivity and low transverse conductance on one side and with a conducting film on the other. There are a great many ways of making such a mosaic, both as to the choice of dielectric and method of forming the sensitive layer.

The natural properties of mica make it a very suitable dielectric for the purpose. As a base for the mosaic of a normal iconoscope the mica should be cleaved into a thin sheet, the thickness being between 1 and 3 mils, depending upon the characteristics desired for the tube. In the subsequent theoretical considerations, a mosaic area of 100 square centimeters has been assumed. This requires a rectangle of approximately $3\frac{1}{2}$ by $4\frac{3}{4}$ inches. The sheet must be clean, preferably freshly cleaved, devoid of foreign inclusions, and free from steps due to fractured cleavage planes.

Since the cleavage operation is an important one, it is worth outlining an experimental technique for obtaining suitable sheets. A clear sheet of mica slightly larger than is required for the finished mosaic is selected. Separation of the sheets is started by gently tapping one corner. A strip of stiff paper is then inserted in the separation and worked around the edges of the mica, gradually moving toward the center. This process may sometimes be facilitated by immersing the mica in water (using a thin, celluloid spatula instead of

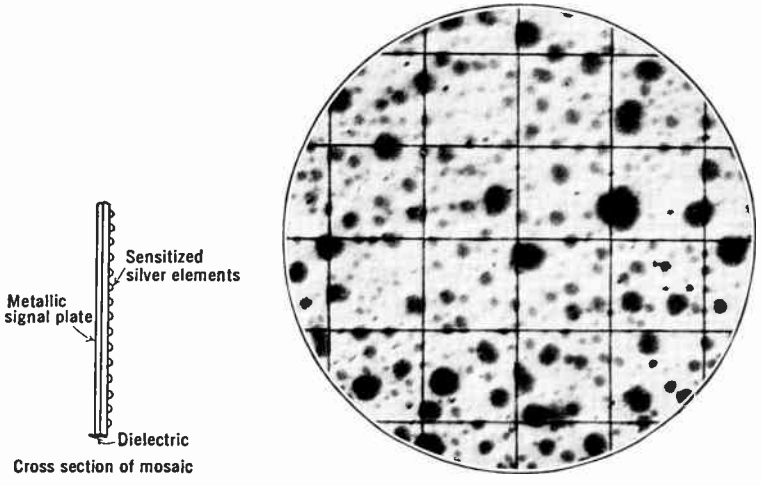


Fig. 9.3. Structure of Iconoscope Mosaic.

the paper strip). Freshly cleaved mica, of course, requires no cleaning, and is ready for the next step in the preparation of the mosaic, namely, the formation of the silver globules.

The silver elements may be formed in a number of different ways. For example, a silver film may be evaporated onto the mica in the form of a continuous film and then, by heating, broken up into minute droplets; a film only a few atoms thick, so thin that it is not conducting, may be deposited on the mosaic after the tube has been evacuated; or silver oxide in the form of a powder may be spread over the mica and reduced to silver globules. Since the last-mentioned method lends itself well to experimental technique, it will be described in greater detail.

Silver oxide which has been ground in a ball mill until it is reduced to a fine flour is used to form the elements. A thin layer of the oxide is spread over the sheet by allowing the powder to settle through air.

Without disturbing the coating the mica is heated rapidly to a temperature above the reduction point of the silver oxide. The rapid heating effects a reduction of the oxide without the individual droplets coalescing into a conducting film. It is often advantageous to lay down the mosaic surface in several thin layers formed in succession, rather than completing it in one operation.

After the silver surface has been formed the back of the mosaic is metallized to make the signal plate. This order of building the mosaic is usually preferable, particularly for experimental tubes, because it permits the use of transmission measurements to determine the coverage of the silver layer. The metallizing can be done by evaporation, sputtering, or with one of the many available chemical metallizing preparations.

In order to prevent the mosaic from buckling or warping, it is backed by a heavy sheet of mica to which it can be attached by means of a suitable strip metal frame. The backing sheet also facilitates making electrical contact to the signal plate.

Since the iconoscope is essentially a high-impedance device, the friction contact between the signal plate and a metal strip threaded through slits in the backing sheet will be ample.

Before taking up the exhaust and activation procedure, other types of mosaic insulation should be considered. Various dielectrics such as enamels, ceramics, metal oxide films, and thin powder layers may be substituted for cleavage sheets of mica.

An enamel to be used as dielectric must have a high specific resistance and be free from inhomogeneities. It is usually applied to a metal sheet, e.g., nickel, which serves as the signal plate. The coefficient of thermal expansion of the coating must match that of the underlying metal.

Metal oxide films can be formed by chemically oxidizing the metal of the signal plate. Aluminum oxide on an aluminum signal plate can easily be thus prepared. Other oxide dielectrics, such as barium borate, magnesium oxide, or silica, may be evaporated onto the signal plate.

Powder insulated mosaics have been found to be very satisfactory, even surpassing mica in some respects. Flowing, or spraying, a liquid suspension of the powder over the signal plate is fairly satisfactory, but other and more complicated techniques give much better results. Magnesium oxide, kaolin, barium titanate, and colloidal quartz are all suitable as dielectrics for powder mosaics.

9.3 Exhaust and Activation Schedule. The assembled iconoscope is sealed to the vacuum system by the exhaust tube in the gun press.

It should be noted that, though this location of the exhaust vent is not essential, it reduces the possibility of the mosaic's being contaminated by gases given off by the gun during activation. Cesium required for activation is admitted through the side tube already mentioned.

The exhaust requirements for the iconoscope are severe. Although the final pressure may be as high as 10^{-5} millimeter of mercury in an operative tube, the slightest trace of certain impurities is found to be absolutely fatal. The evacuation schedule for the most part follows standard practice. The tube is pumped to a low pressure at room temperature, and then heated to as high a temperature as the blank can safely withstand. During the initial exhaust and the first fifteen to thirty minutes of bake, a freezing-out trap normally is not used. This avoids collecting at the mouth of the trap large amounts of condensable vapors given off by the glass and metal parts. For the rest of the exhaust, the trap is used. The bake is continued until the pressure in the tube has dropped to 4 or 5×10^{-5} millimeter of mercury, after which the iconoscope is allowed to cool slowly to room temperature. In order to drive out gases occluded in the metal parts of the gun, these parts are brought to a dull red heat by high-frequency heating.

The activation of the mosaic follows closely that used for the hard cesium phototube, with this difference: in the phototube the only consideration determining the amount of cesium is the sensitivity, whereas in the mosaic both sensitivity and conductivity across the surface are dependent upon the amount of cesium. The process of sensitization consists of oxidizing the silver surface, admitting cesium, and baking to promote a reaction between the silver oxide and the alkali.

Oxidation of the silver elements on the mosaic can easily be effected by the procedure given in Chapter 1. Oxygen is admitted into the system to a pressure of about one millimeter, and a glow discharge is formed over the surface of the mosaic. The glow discharge can be produced with the aid of a high-voltage, high-frequency oscillator grounded on one side, the other side being connected to a movable electrode placed outside the walls of the tube. A very uniform oxidation can be obtained by moving the electrode judiciously over the bulb. The thickness of the oxide layer can be gauged by the interference colors produced by the film on the metal. The colors appear in the usual order for subtractive interference colors as the thickness is increased, namely: yellow, red, blue, blue-green, second yellow, second red, second green, etc. As in phototube practice, the first blue will produce a very satisfactory photosensitive mosaic.

Care should be taken to avoid a violent discharge during oxidation as this will cause silver from the globules to be sputtered over the mica, resulting in loss of both signal and resolution due to leakage between the elements. After the oxidation the excess oxygen is pumped out of the system. A short bake at 200°C will expedite the complete removal of the oxygen, but is not essential.

The thermionic cathode in the electron gun employs an emitting surface of barium and strontium oxide. The preparation and activation of this type of cathode were described in Chapter 1. In activating a cathode prepared in this way, a large amount of gas is released during the conversion of the carbonate to the oxide. As this gas may contaminate a sensitized mosaic, it is advisable to activate the cathode before proceeding with the cesiation. During the activation of the cathode the grid structure should be kept at a dull red heat to avoid the formation of an oxide-bearing film.

The final step in processing the iconoscope is the cesiation of the mosaic. A quantity of cesium, slightly less than that required for complete activation, is driven into the iconoscope. In order to promote a reaction between the silver oxide and the cesium, the tube is baked at 200°C for a few minutes. After it has again become cool the sensitivity is measured ballistically, by methods described in the next section. The final sensitivity should lie between 7 and 10 microamperes per lumen. By comparing the first sensitivity reading with this figure, the additional cesium necessary can be estimated. After this amount has been admitted the tube is again baked. Care should be taken to avoid an excess of cesium, as this will result not only in poor photosensitivity but also in loss of both resolution and storage due to leakage. If an excess has accidentally been admitted, its harmful effects can sometimes be reduced by a prolonged bake at 200°C. The presence of a tin oxide or lead oxide cesium getter in the tube will also reduce the harmful effect of too much cesium.

The simple sensitization schedule described is well suited to meet experimental laboratory needs. In the commercial production of iconoscopes a much more exacting procedure is employed. Various rather complicated techniques, such as sensitization with silver and argon or hydrogen glow discharges, have been developed which not only increase the sensitivity of the tube but also improve the color response. These methods lead to more than double the sensitivity resulting from the simple procedure outlined.

9.4 Performance Tests. The ultimate test of the merit of an iconoscope is, of course, its performance in the television system. There are,

however, a series of measurements which can be made during its construction and activation, before it is actually used, which will give an indication of quality and thus greatly facilitate producing a good tube.

The iconoscope must be tested for its behavior with regard to the following factors: (*a*) photosensitivity, (*b*) leakage over mosaic, (*c*) characteristics of gun, (*d*) perfection of mosaic, (*e*) resolution, (*f*) sensitivity, (*g*) color response, (*h*) spurious signal, (*i*) secondary emission from gun, (*j*) miscellaneous defects. Of these tests, the first three are made during the activation of the iconoscope, the remaining ones after its completion.

Test *a*, for photosensitivity, is extremely important and should be made at regular intervals during the final stages of the cesiation of the tube, so that it can serve as a guide to the treatment required.

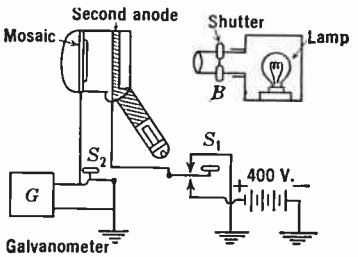


Fig. 9.4. Ballistic Sensitivity Test.

Since the mosaic is an insulator, transient measurements are necessary to determine the photoemissivity of its surface. The ballistic method described below is found convenient and is often used. The method consists of applying a potential between the signal plate and second anode, and then allowing a flash of light of known intensity and duration to fall on the mosaic. The charge released as a result of the light impulse is measured with a ballistic galvanometer. The set-up for this measurement is shown in Fig. 9.4. The light source projects a spot of light about 2 inches in diameter onto the mosaic. The duration of the light impulse is controlled by the shutter B; an ordinary camera shutter will serve the purpose admirably. The quantity of light found convenient is of the order of $\frac{1}{100}$ lumen-second, that is, for example, $\frac{1}{2}$ lumen for $\frac{1}{50}$ second.

Since the emission from an average mosaic lies between 7 and 15 microamperes per lumen, the charge released will be between 0.07 and 0.15 microcoulomb. Furthermore, since the capacity of the mosaic to its signal plate is of the order of 100 micro-microfarads per square centimeter, the rise in potential will be of the order of 60 volts. The voltage, V, applied to the second anode must be well in excess of this value to insure that the photoemission is at all times saturated. In practice, a potential of 200 to 400 volts is used.

When the test is started, the galvanometer G is short-circuited through switch S_2 and the second anode grounded by means of S_1. The

whole iconoscope is then illuminated with an intense light so that photoemission from the second anode and other metal parts will bring the mosaic to approximately ground potential. Next, the bulb is covered so that no light can reach the mosaic except through the shutter from the lamp. Switch S_1 is moved to the other contact, while S_2 still shunts the galvanometer to prevent the capacity surge from affecting the instrument. The galvanometer switch S_2 is then opened and the light flashed. From the charge indicated by the galvanometer deflection and the quantity of light the photosensitivity can be calculated.

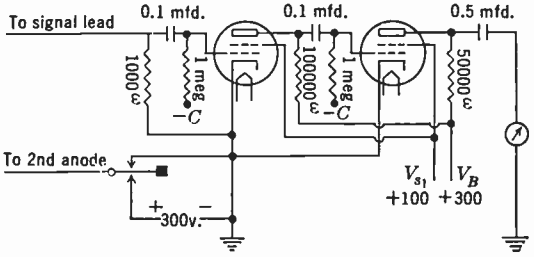


Fig. 9.5. Vacuum-Tube Meter for Sensitivity Test.

An alternative, very similar to the above method, is to illuminate the mosaic continuously, and to apply the second anode potential for a short known period of time. In general, this is found to be less satisfactory than the former procedure.

The ballistic galvanometer for these tests may be replaced by a vacuum-tube meter. A circuit suitable for the purpose is shown in Fig. 9.5.

Next in importance is the measurement of leakage over the mosaic. This can be made as follows. With the test equipment just described, and the potential applied to the second anode, sufficient light from the source is allowed to fall on the mosaic to cause the illuminated portion of the mosaic to reach second anode potential. The light required for this will be of the order of $\frac{1}{10}$ lumen-second. The mosaic is then left in darkness for 5 seconds. The deflection produced by the $\frac{1}{100}$ lumen-second impulse of light then applied to the mosaic will indicate the amount of charge which has leaked away from the charged area into uncharged regions. With 200 volts applied to the second anode the deflection thus observed should be small compared with that obtained in the normal sensitivity test. It should be noted that the time of exposure required to saturate the area of the mosaic illuminated will depend upon the sensitivity. If leakage tests are made during the

early stages of activation, when the sensitivity is low, the exposure time must be lengthened accordingly. However, the exposure time should be no longer than necessary as stray light and leakage may cause the unexposed portions of the mosaic to become saturated, thus vitiating the leakage measurement. Because the requirements of low leakage and high photosensitivity must both be fulfilled in a satisfactory iconoscope, the importance of these tests cannot be overemphasized.

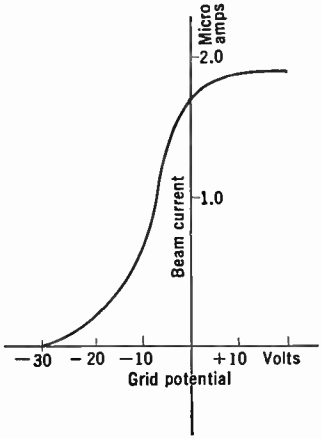


Fig. 9.6. Current Characteristic of a Typical Iconoscope Gun.

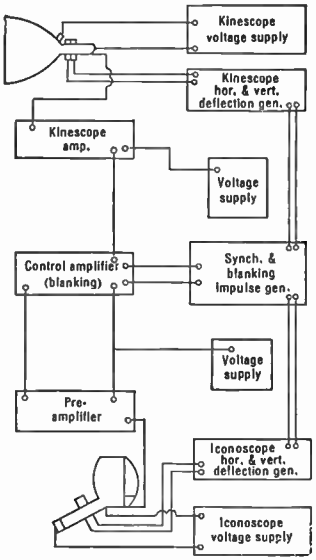


Fig. 9.7. Block Diagram of an Iconoscope Test Set.

Test *c*, of the gun performance, can be simply performed by applying the normal voltages to the gun and measuring the current to the second anode as a function of grid bias. In carrying out this test the beam should be deflected off the mosaic with a small permanent magnet to avoid burning a spot on the sensitized surface. A typical characteristic curve of the gun is shown in Fig. 9.6.

The remaining tests are performed after the iconoscope has been completed and sealed off from the vacuum system. These tests require equipment consisting of a video amplifier with an available gain of at least 8000, voltage supplies suitable for providing the potentials needed by the iconoscope, horizontal and vertical sawtooth generators, and a kinescope for viewing the reproduced picture. Descriptions of these

various units will be found in the succeeding chapters of this book. A simplified diagram of the complete test outfit is shown in Fig. 9.7; Fig. 9.8 is a photograph of a typical iconoscope test outfit.

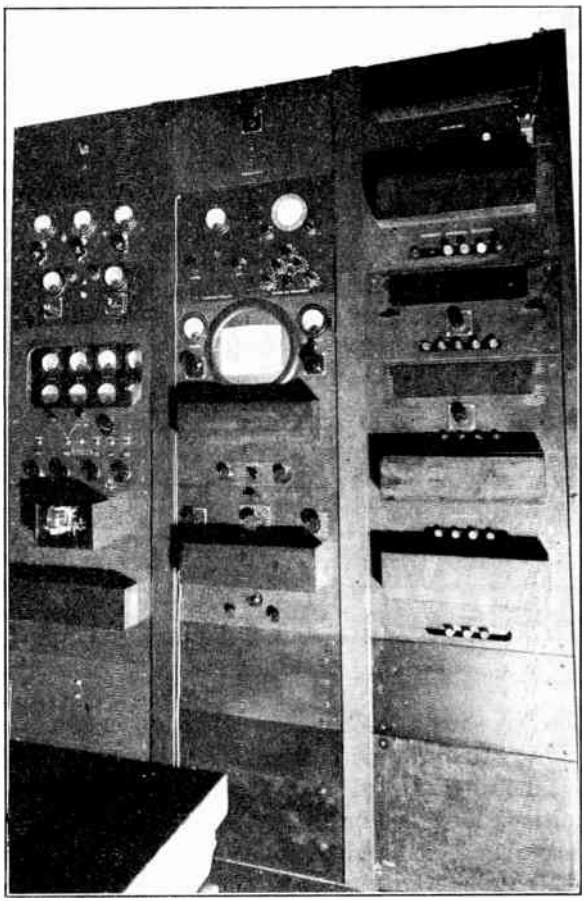


Fig. 9.8. View of a Typical Iconoscope Test Set.

The completed iconoscope is connected into the test set and the voltages adjusted to the values required for correct operation. With the tube in darkness the beam is allowed to scan the mosaic. The resulting picture on the kinescope should show the mosaic as a uniform clean surface. Small bright or dark spots are usually due to dust or dirt particles on the mosaic, or to pinholes in the signal plate. Large, irregular patches may be due to dirt, to nonuniform distribution of

cesium, or to lack of uniformity in oxidation. Sharp, straight-edged areas are due either to splits across cleavage planes of the mica or to scratches on the signal plate.

A uniform light is then projected over the surface of the mosaic to determine the uniformity of photoemission.

To test the definition of the tube, a resolution pattern, such as is reproduced in Fig. 9.9, is projected onto the mosaic. A good experi-

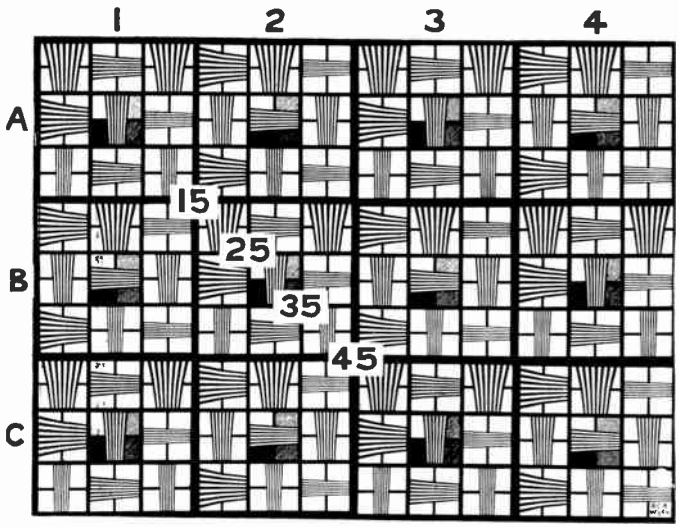


Fig. 9.9. Resolution Pattern.

mental iconoscope should resolve the pattern down to at least 600 to 800 lines. Since often the video amplifier response is limited to 3 or 4 megacycles, the resolution of the picture seen on the kinescope under normal conditions may be limited by the system. However, if the horizontal scanning is reduced to cover only a small fraction of the mosaic while keeping the kinescope deflection normal, an enlarged image of the pattern can be seen on the screen. The resolution of this enlarged section will be limited by the iconoscope only.

It is important that the spot produced by the gun be round, as any ellipticity due to misalignment will cause unequal resolution in various directions. The iconoscope deflection being contracted both vertically and horizontally, the resolution obtained along different axes will be a measure of the regularity of the spot.

For a routine estimate of the sensitivity, a picture of known quality, such as the resolution pattern just illustrated, is projected onto the

mosaic and the image is observed as the light in the projected picture is reduced. The amount of light required to produce a picture just visible over the amplifier noise is a measure of the sensitivity. This illumination can be compared with that required to produce a similar picture in an iconoscope of known performance.

A quantitative measurement of the sensitivity is much more difficult. It is usually made by measuring the signal output, when a small area of the mosaic—for example, a square covering an area equal to one-hundredth the total area of the mosaic—is illuminated with a known light intensity. A plot is then made of signal output against light intensity. Typical curves will be found in section 9.11.

The relative color response can be determined by projecting a spectrum of known light distribution onto the mosaic and measuring the signal output for the various wavelengths. A color chart of the type used for photographic measurements imaged on the mosaic affords a very convenient way of estimating the color response.

When the mosaic is scanned, even in darkness, a spurious signal in the form of shading will be observed on the viewing screen. This is variously known as "black spot," "tilt and bend," "shading," etc., and its cause will be discussed in a succeeding section. The magnitude of the signal increases rapidly with beam current. In a good tube the spurious signal should be small and fairly regular. Electrical compensation of this signal will be taken up in a later chapter.

Appreciable secondary emission from the electron gun is rather rare and is always due to some fault in the gun construction. The symptoms are a haze which resembles a partially transparent curtain over the picture. If the amplitude of the deflection is greatly reduced, this haze will be seen to take the form of a very blurred image of the mosaic. When this defect is present it can be overcome by operating the gun under conditions of inverse focus. This is done by applying a higher potential to the first anode than to the second anode. For example, a three-element gun which focuses at a ratio of 1 to 4 between first and second anode will be found to focus also with a ratio around 3 to 1. However, this mode of operation is not very satisfactory, from the standpoint either of the resolution obtainable or of the life of the tube.

The final test consists of examining the performance of the tube under as wide a variety of operating conditions as possible. This test should reveal whether the tube oscillates under high beam current, whether the scanning beam strikes the neck of the gun, and any other defects which might lessen its usefulness.

9.5 Theory of Operation—Characteristics of the Mosaic. In preceding chapters a very much simplified picture of the operation of the iconoscope was given. The actual details of operation are extremely complicated, and as interesting as they are complex. Quantitative measurements of the various factors involved in the operation are difficult, and in general require the introduction of measuring electrodes and probes, which alter to some degree the conditions existing in the normal tube. The extent of these changes cannot be estimated readily. However, it is possible to derive a working theory yielding quantitative results which are in fair agreement with the measured performance of the iconoscope.

The surface of the mosaic, as has been pointed out, consists of a vast array of minute photosensitive elements. The size of these elements depends upon the way in which the mosaic was formed and sensitized, and may range from droplets 0.0005 centimeter in diameter down to submicroscopic particles. In any case, these elements are very small compared to the diameter of the scanning spot, so that a large number of these particles are under the beam at any one time. Thus it is evident that, from the standpoint of the mosaic, the concept of a picture element has no real significance, but is only a convenient fiction for simplifying the description of the operation of the iconoscope. In what follows, the mosaic will be treated as a two-dimensional continuum capable of photoelectric and secondary emission but with zero transverse conductance.

The capacity of the sensitized surface to the signal plate depends upon the thickness of the insulating layer and upon the dielectric material. In general, it may be said to range between 50 and 300 micro-microfarads per square centimeter. The operation of the ordinary iconoscope is not very critical to the value used. For convenience, a value of 100 micro-microfarads per square centimeter will be assumed.

The secondary-emission ratio of the mosaic surface also may vary over a wide range. It has been measured to be as low as 2 for certain types of mosaics, to 8 or 9 for others. A fair average value to take as a basis for discussing the behavior of the tube is 4. The photoelectric sensitivity of a good silver-sensitized mosaic is in the neighborhood of 15 microamperes per lumen.

The electron scanning beam at the mosaic has a diameter of 0.01 to 0.02 centimeter and a current of 0.1 to 0.25 microampere. The current distribution in this beam is shown in Fig. 9.10.

For purposes of discussion, the spot will be considered to be a square 0.02 centimeter on a side and having a uniform current density 5×10^{-4} ampere per square centimeter. As it moves across the mosaic it has a linear velocity of 2.1×10^5 centimeters per second. This assumes a picture period of $\frac{1}{30}$ second and a 525-line picture. For simplicity it will be further assumed that the scanning is not interlaced. The effect of interlaced scanning will be discussed somewhat later. The dimensions of the mosaic commonly used are $3\frac{1}{2}$ by $4\frac{3}{4}$ inches, or approximately 9 by 12 centimeters—an area of slightly more than 100 square centimeters. On the basis of the average capacity per unit area given above, the total capacity of the sensitized surface of the mosaic to the signal plate is about 10,000 micro-microfarads. In this connection it may be mentioned that the capacity between signal plate and second anode is in the neighborhood of 6 micro-microfarads.

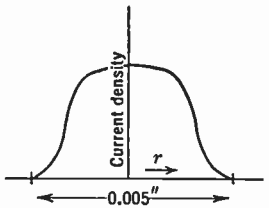


Fig. 9.10. Current Distribution in Scanning Spot.

9.6 Potential Distribution on the Mosaic. Contrary to what might be expected, the surface of the mosaic when scanned in darkness does not assume a uniform potential. Measurements made both on a cesiated silver element set into the surface of a mosaic and connected by a lead through the wall of the tube, and on tubes having a divided signal plate, show that directly under the scanning beam the mosaic assumes its most positive potential, and that it is at its most negative potential directly in front of the scanning beam. Between these two points there is a continuous transition of potential. At the point of bombardment the surface assumes a voltage between +2 and +3 volts with respect to the second anode. This potential is, to some extent, dependent upon the physical state of the surface of the mosaic and upon the velocity of the bombarding electrons, but it is to the first approximation independent of the current in the scanning beam. On the other hand, the lower limit of the potential is primarily a function of the beam current and secondary-emission characteristics. At very high currents, that is, between 0.5 and 1.0 microampere, the lower potential will be about $-1\frac{1}{2}$ volts. As the current is decreased, the lower limit gradually approaches the potential under the beam. Figure 9.11 is a map of the potential distribution over the mosaic for various beam currents.

The rather complicated potential distribution arises from the fact that two different types of potential equilibria are possible simultane-

ously at different points on the mosaic. These are, first, that of an insulated secondary-emissive area under electron bombardment and, second, the retarding potential necessary to prevent electrons scattered from the point under the bombarding beam from reaching the mosaic.

In order to account for the potential of the point directly under the bombarding beam, it is necessary to consider the secondary-emission

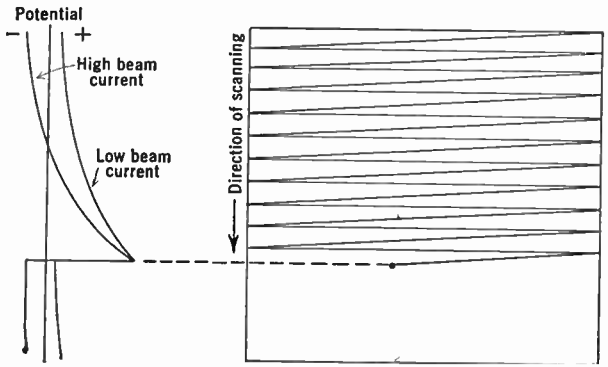


Fig. 9.11. Distribution of Potential on Mosaic.

phenomenon taking place at this area. Since the mosaic is surfaced with cesiated silver which has been activated according to the schedule used for the iconoscope, a saturated secondary-emission ratio of 4 can be assumed. When an insulated surface of this type is bombarded, more electrons will leave than arrive, causing it to become positive with respect to the electrodes which collect the secondary electrons. As the potential becomes increasingly positive it more and more pre-

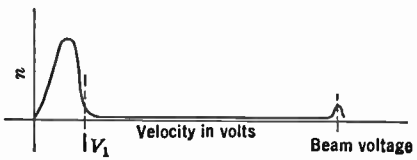


Fig. 9.12. Velocity Distribution of Secondary Electrons from a Mosaic.

vents secondary electrons having small initial velocities from reaching the collecting electrode or other parts of the mosaic. When the potential reaches such a value that the escaping current just equals the bombarding current, equilibrium is established and there is no further change in potential. To clarify this mechanism further, Fig. 9.12 is presented. The curve represents the velocity distribution of secondary

electrons from a surface of the type under discussion. The abscissa of the curve gives the initial velocity in volts; the ordinate, n, multiplied by the velocity range dV, gives the number of electrons having velocities lying between any value V and $V + dV$ for each primary electron. The integral of this curve,

$$\int_0^\infty n\,dV = R$$

gives the secondary-emission ratio R. Again, if the emitting surface has a potential V_1 the ratio of secondary to primary current will be:

$$R' = \int_{V_1}^\infty n\,dV$$

Equilibrium will be established when R' is equal to unity. The equilibrium potential for a surface of the type found in the mosaic lies, as has been stated, between 2 and 3 volts.

Owing to the positive potential of the point just considered, and to free electrons between the mosaic and second anode, which is the collecting electrode in this case, a certain fraction of the electrons is returned to points on the mosaic not under the scanning beam. This current, which reaches the mosaic in the form of a more or less uniform rain of redistribution electrons, causes the rest of the surface to become increasingly negative. When a region of surface has reached $-1\frac{1}{2}$ volts, the retarding potential is sufficient to prevent all or nearly all redistribution electrons from reaching this area. Thus, the limits of the potential variation over the surface are $+3$ volts and $-1\frac{1}{2}$ volts. If the beam current is small, the redistribution current will not be sufficiently great to cause any point on the mosaic to reach the lower equilibrium point in a frame period.

Thus, to summarize the cycle which occurs when the mosaic is scanned in darkness, it is found that an element of area under bombardment assumes a potential of $+3$ volts. As the scanning beam moves on, the element receives a small current in the form of redistributed electrons. This causes the area to become increasingly negative as the beam moves farther away. The decrease in potential continues until either the scanning beam returns to the element in question or until it reaches about $-1\frac{1}{2}$ volts.

9.7 The Mosaic under the Influence of a Light Image. The mechanism of the generation of the picture signal can be described on the basis of the phenomenon just discussed. Fundamentally, it is due to

the fact that the potential of an illuminated element of area, just before being reached by the scanning beam, is more positive than that of one which has not been illuminated. Since the mosaic is driven to the same positive equilibrium regardless of its initial potential, charge will be released at a different rate when the beam is traversing the lighted region from that when scanning an unlighted area. The change in current reaching the second anode constitutes the picture signal.

The difference in potential due to light is, of course, caused by the storing up of the charge resulting from the photoemission of the illuminated areas. Under actual operating conditions this photoemission is less than the redistribution current so that, relative to the second anode, every element of area becomes increasingly less positive with respect to the second anode. Those which are illuminated decrease in potential less rapidly than the remainder.

The photoemission from the mosaic is, in general, unsaturated, as will be evident when it is remembered that the potential of the mosaic is, for the most part, positive with respect to the second anode. This makes the calculation of the signal to be expected from a given amount of illumination very difficult because it means that the emission is a function of the potential of the area under consideration as well as of the incident light. A further complication results from the fact that it is entirely unnecessary for the photoelectrons to leave the mosaic in order that a picture signal be generated. In fact, under normal operating conditions, the fraction of photoelectrons from an illuminated area which permanently leave the mosaic is insignificant compared to that which is returned to the mosaic in the form of redistribution electrons. Experimentally, this can be easily shown by interrupting the image projected onto the mosaic at a frequency which is some multiple of the frame frequency. If care is taken to insure that the second anode and signal plate are not photoemissive, almost no signal will be observed corresponding to the interruption frequency. This indicates that no photoelectric current is traversing the circuit consisting of mosaic, second anode, amplifier coupling impedance, and signal plate.

In view of the complicated nature of the unsaturated photoemission responsible for the picture signal, it will not be amiss to discuss it in greater detail. Considering first the extreme case of a mosaic in the absence of a scanning beam, if a small area of the mosaic is illuminated with an intensity L_1, this area will emit electrons which in part go to the second anode and in part return to the mosaic. Eventually this area becomes sufficiently positive to prevent the escape of further

electrons. It is obvious that under these conditions the positive potential assumed by the area is dependent upon the initial velocities of the photoelectrons rather than upon the intensity of illumination. Therefore, if an intensity L_2—half that of L_1—had been used, the potential would have been approximately the same. If, however, two areas are illuminated at the same time, one with an intensity L_1, the other with L_2, these areas will not come to the same potential. This will be clear if the photoelectric saturation curve shown in Fig. 9.13 is examined and the behavior of the electrons considered. The photoemission curves are represented as straight lines, for convenience of discussion, rather than in their true form. The ordinate represents the actual number of electrons emitted by an elementary area of the mosaic at any potential, and not the net current, which is the algebraic sum of the current leaving and the redistribution current arriving from other portions of the mosaic. When the two areas emit simultaneously the return current to each illuminated element of area will be approximately equal,* represented by the line *ab*. In order that the net current to each elementary area be zero, the two illuminated regions must take on the potentials E_1 and E_2, respectively. Similarly, if an entire image is projected onto an unscanned mosaic, a corresponding potential image will be established as an equilibrium condition.

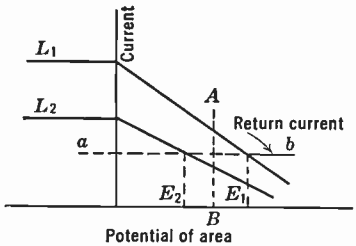


Fig. 9.13. Current-Density Characteristics of Differently Illuminated Mosaic Elements.

Under operating conditions, when a relatively weak beam is used and an intense image is projected onto the mosaic, the mechanism of the formation of the charge image will be essentially that outlined above, except that the current represented by *ab* will include not only the returning photoelectrons but also the redistribution electrons from the beam. In this case the charge image is fully established in less than a picture period, and further exposure produces only a slight increase in picture signal. (This increase is due to line sensitivity, which will be discussed in succeeding pages.) To avoid possible misunderstanding, it should be pointed out that this saturated mode of operation does not imply that the tube is operating as a nonstorage system, since, even if the picture is fully established in a time corresponding to,

* Strictly speaking, the magnitude of the return current is a complex function of tube geometry, mosaic activation, distribution of illumination, etc.

say, 10 lines, the storage factor is still some 5000 times. As the illumination is decreased, eventually a point is reached where the charge image is no longer fully established in a frame time. Under these circumstances the illuminated area is receiving redistributed photoelectrons and beam secondary electrons, and during the cycle becomes increasingly negative with respect to the second anode; however, its

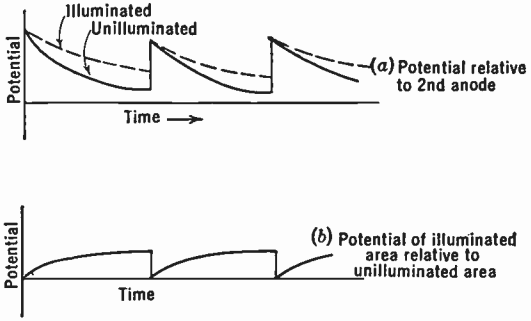


Fig. 9.14. Variation of Potential with Time for Unilluminated Elements.

potential is continuously rising with respect to an unilluminated area. Figure 9.14*a* shows the time variation of the potential of an illuminated area and an unilluminated area with respect to the second anode, whereas *b* shows the potential of the same illuminated area referred to that of the unilluminated region.

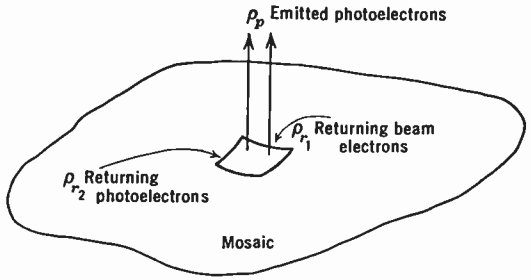


Fig. 9.15. Current Components for an Element Not under the Beam.

From a more quantitative point of view, if an element of area not under the beam ds_a (Fig. 9.15) is considered it is found to receive the following current components: ρ_p the current density due to photoelectrons leaving the area, ρ_{r1} redistribution electrons from the beam, and ρ_{r2} returned photoelectrons. The rate at which its potential is changing with respect to the second anode will be

$$\frac{dV_a}{dt} = \frac{(\rho_p - \rho_{r1} - \rho_{r2})}{C_0\, ds_a} ds_a$$

where C_0 is the capacity per unit area. The corresponding rate for an element ds_b, in darkness, will be

$$\frac{dV_b}{dt} = \frac{-(\rho_{r1} + \rho_{r2})}{C_0\, ds_b} ds_b$$

The three current-density components in the foregoing differential equations are functions of the potential with respect to the second anode and of the difference in potential between various portions of the mosaic. Actually, however, the variation of ρ_{r1} and ρ_{r2} with local variations in potential is relatively small. Assumptions could be made as to the functions involved, and the differential equations could be solved. However, since little is known of the nature of these functions, very little would be gained by doing this. It must suffice to say that, under conditions of low light levels, the average photoelectric current leaving a small, illuminated region, with the rest of the mosaic in darkness, is approximately 20 percent of the saturated photoemission.

To summarize the portion of the cycle discussed above: directly under the beam the mosaic assumes a potential V_2; examined farther and farther from the point under bombardment, the potential is found to decrease to some potential $V_1(x, y)$, which is not constant over the mosaic but is a function of the light distribution in the image and, therefore, a function of the coordinates, x, y, on the mosaic.

9.8 The Formation of the Video Signal. The final step in the cycle is the transition under the beam of the potential of an element from $V_1(x, y)$ to the positive equilibrium potential V_2. It is of fundamental importance, for, in this step, the picture signal is generated.

The beam in sweeping over the mosaic acts like a resistive commutator in series with a source of positive potentials of 2 or 3 volts. The resistance in this case is determined by the current-voltage relation of the secondary electrons from the bombarded area, and will, of course, be nonohmic. By means of small, cesiated silver electrodes inserted into the mosaic, it is possible to learn a little about the nature of this relation. Figure 9.16 shows typical curves of the current to or from such an element as a function of voltage. It is found that the curve may be approximated over a large part of its range by an ohmic impedance whose value is $z = z_0/i_b$, where i_b is the beam current and z_0 the coefficient of beam impedance having a value between 1 and 2

ohm-amperes. Although this relation is useful for certain purposes, where an averaging over a picture cycle is involved, it is not a sufficiently good approximation for the determination of the instantaneous condition of a bombarded element. It is probable that the emission from an element is actually very nearly saturated (that is, most of the electrons leaving it return to other points of the mosaic rather than to the element under bombardment) until it is very close to the equilibrium potential, V_2.

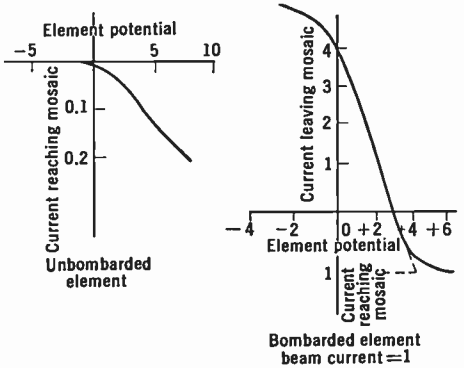


Fig. 9.16. Current-Voltage Characteristics for an Element of the Mosaic.

For purposes of determining the change during bombardment in potential ΔV of an element of area which is small compared with the size of the scanning spot, yet includes a large number of silver particles, let the saturated secondary-emission current density be ρ_s and the capacity

$$C = C_0\, ds$$

If it is assumed that the secondary emission is saturated for the duration of the time, T_e, the beam is on the element, the change in potential will be

$$\Delta V = \frac{(\rho_s - \rho_b)\, ds}{C_0\, ds} T_e \tag{9.1}$$

For a beam current of 0.20 microampere, ρ_b has a value 5×10^{-4} ampere per square centimeter, and, assuming a secondary emission ratio of 4, $\rho_s - \rho_b = 1.5 \times 10^{-3}$. The element time, T_e, can be found by dividing the linear scanning velocity, v, into the width of the spot, h, which gives 1.0×10^{-7} second for the time the element is being

bombarded. Introducing these values into Eq. 9.1, the maximum value for the change in voltage is around 1.5 volts.

Experimentally, it is found that, under the beam conditions described, the change in voltage $\Delta V = V_2 - V_1$ lies between 1 and 2 volts, which is consistent with the above.

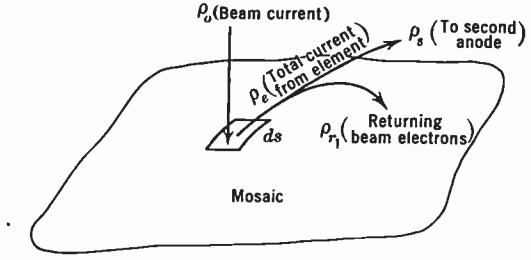


Fig. 9.17. Current Components for an Element under the Beam.

The net current leaving the elementary area (see Fig. 9.17), under the conditions just described, will be

$$di_e = -\left(C_0 \frac{dV}{dt}\right) ds$$

Since the beam current reaching the area is $\rho_b\, ds$, the total current from the element is

$$di_t = -\left(\rho_b + C_0 \frac{dV}{dt}\right) ds$$

This entire current will not reach the second anode because, on the average, as much current must reach the mosaic as leaves. Let $f(x, y, V)$ be the fraction of this current reaching the second anode; then the current to the second anode can be written as

$$di_s = f(x, y, V)\left(\rho_b + C_0 \frac{dV}{dt}\right) ds \tag{9.2}$$

The instantaneous current reaching the second anode is given by the integral of the above expression over the area of the scanning spot. Since the current i_s carries the picture signal, the solution of Eq. 9.2 would be highly desirable. This integration, however, cannot be performed since the function $f(x, y, V)$ is unknown.

It is possible to make certain approximations which will permit the calculation of an average value for the redistribution function. The

mosaic is an insulator and returns a current to the second anode equal to the beam current. Hence the average value of this function must be given by

$$\bar{f} = \frac{i_b}{i_t} = \frac{i_b}{i_b + i_e} \tag{9.3}$$

The average value of the element current, i_e, must be equal to the charge released by the beam divided by the time required for its release. On this basis, i_e is given by

$$i_e = \int_{\text{spot area}} \frac{(V_2 - V_1)C_0\, ds}{T_e}$$

$$= \frac{(V_2 - V_1)C_0 h^2}{h/v} = (V_2 - V_1)C_0 hv$$

Substituting this equation into Eq. 9.3, the average value for the redistribution function will be

$$\bar{f} = \frac{1}{1 + \dfrac{hvC_0(V_2 - V_1)}{i_b}} \tag{9.4}$$

For quantities representing ordinary operating conditions, the value of the function is found to be about 0.25. This may be interpreted to mean that approximately a quarter of the total charge which leaves the mosaic reaches the second anode, the rest being returned to the mosaic in the form of redistribution current. Since the variation in V_1 over the mosaic under conditions of low light levels is small, it is legitimate to treat f as independent of the charge image. Therefore, only about 25 percent of the stored charge is available for producing the picture signal.

The redistribution signal, which constitutes about three-fourths of the net stored charge, can very easily be shown experimentally. The iconoscope required has a mosaic whose sensitized surface is made in the ordinary way, but whose signal plate is divided into two halves, each of which is brought out on a separate lead. The video amplifier is connected to one half of the signal plate, whereas the other half is grounded. If a light image is projected onto the grounded side of the mosaic, a negative of this image will be seen on the screen of the viewing tube. This image is due, of course, to the redistribution electrons returning to the mosaic and carrying some of the signal back in the

form of a negative picture. With normal connections, this negative signal tends to oppose the positive picture, so that the net output of the tube is only about 25 percent of what might be expected if this redistribution phenomenon did not exist.

9.9 Line Sensitivity. Going back for a moment to the question of the photoemission of the mosaic, an examination of the potential distribution over the mosaic reveals that there is an abrupt potential rise at the line being scanned, and that the line just in front of it is subjected to a strong positive field which tends to saturate the photoemission. This means that the line in question is much more sensitive than any other part of the mosaic, and, furthermore, even if "saturation" light levels are used so that an equilibrium potential image has been established, further photoemission can take place at this line. This phenomenon, known as line sensitivity, is the reason for the statement that the picture signal keeps on increasing even when the light level is increased above that required to form an equilibrium potential image.

The phenomenon of line sensitivity can be demonstrated very strikingly by the following simple experiment. The image from a continuously run moving-picture film (i.e., that obtained by removing the intermittent and shutter from a moving-picture projector) is projected onto the mosaic of an iconoscope. The film is run at such a rate that its frame speed is equal to the field frequency and in a direction such that the image moves opposite to the vertical direction of scanning. Under these conditions the iconoscope transmits a clear image of two frames of the moving-picture film, although to the eye there appears to be only a blur of light on the mosaic.

Thus, it is evident that there are two sources of signal, one the stored charge on the entire mosaic surface, the second the sensitive line ahead of the scanning beam. At low light levels the surface sensitivity contributes the major portion of the signal, but at high illumination intensities, line sensitivity can contribute a large fraction.

9.10 Black Spot. Up to this point it has been assumed that the two terminal potentials, V_1 and V_2, had the same value at every point on an unilluminated mosaic. Actually, because both potentials are sensitive to the field conditions in their neighborhood, they are functions of the coordinates of the mosaic. Furthermore, the redistribution factor, f, is also dependent upon the location of the point under bombardment. As a consequence, there is a variation in the current reaching the second anode as the mosaic is scanned. This variation gives rise to spurious signal which, if not compensated, produces an irregular shading over

the picture. It is evident from the factors upon which this shading depends that it may be considered as divided into two parts. The first is a stored signal which depends upon $V_1(x, y)$; the second is an instantaneous effect depending upon the variation of $V_2(x, y)$ and $f(x, y)$ over the mosaic.

The instantaneous component of the spurious signal can be demonstrated experimentally if a metal plate is substituted in place of the usual mosaic. When such a plate is scanned in the ordinary way, being maintained at a potential close to that of the second anode, a spurious signal is generated which is very similar to that produced by the iconoscope. The mechanism of the generation of this signal, though not identical with that involved in the case of the mosaic, is quite similar. If the potential of the plate is made strongly positive or negative with respect to the second anode, so that the secondary-emission current reaching the second anode is either saturated or suppressed, the signal will disappear.

In the normal iconoscope the magnitude of the spurious signal depends upon: (1) the nonuniformity of scanning, (2) the shape of the glass envelope and extent of exposed glass, (3) the beam current, and (4) the activation of the mosaic. It can be reduced by attention to items 1, 2, and 4, but it cannot be completely eliminated. With respect to its dependence upon beam current, it is found to increase more rapidly than the picture signal as the beam current is increased. For this reason, it is common practice to operate the iconoscope with small beam currents (i.e., from 0.1 to 0.2 microampere) at the expense of some efficiency in signal generation.

In television transmission the spurious signal from the iconoscope is compensated by means of an electrical correcting network, which introduces a signal similar in shape to the shading signal but opposite in polarity into the video amplifier.

No mention has been made of space charge between the mosaic and second anode. Between these elements there necessarily exists an electron cloud, but under ordinary operating conditions its density does not become sufficient to create a potential minimum or a virtual cathode. Although the space cloud of electrons is incidental only, as far as the operation of the iconoscope is concerned, nevertheless it has an effect on the paths of the redistribution electrons, and can introduce transit time effects observable under certain unusual operating conditions. Very close to the element under bombardment, when it is near its equilibrium potential, a virtual cathode is undoubtedly formed.

9.11 Performance of the Iconoscope. From the theory of operation just presented, the signal output from a standard iconoscope falls below that expected from an ideal storage tube owing to lack of saturation of the photoemission and to redistribution losses. Assuming the signal to be reduced by a factor of 5 by the former and by a factor of 4 by the latter, the overall efficiency is expected to be in the neighborhood of 5 percent. In spite of this inefficiency, the very great advantage resulting from the application of the storage principle makes the iconoscope a very effective pickup tube.

On the basis of the mechanism predicated and the estimated efficiency, the output expected can be calculated. Let it be assumed that a small area of the mosaic is illuminated with an intensity of L lumens per square centimeter. Also the following additional data taken from the characteristics of the standard iconoscope are necessary:

Area of mosaic	$A = 100$ square centimeters
Photosensitivity	$p = 15$ microamperes per lumen
Overall efficiency	$k = 0.05$
Coupling resistor	$R = 10{,}000$ ohms

The increase in current reaching the second anode as the beam reaches the illuminated area will be given by the additional charge accumulated per unit area by the mosaic multiplied by the rate at which the beam traverses this area and the redistribution efficiency. In other words,

$$\begin{aligned} i_s &= (kpLT)h \cdot v \\ &= k \cdot p \cdot L \cdot T \cdot \frac{H}{n} \cdot W \cdot \frac{n}{T} \\ &= kpLA \end{aligned} \tag{9.5}$$

In this derivation, H and W are the height and width of the mosaic, n the number of lines, and T the picture period. The remaining symbols have already been defined.

The voltage appearing across the coupling resistor will be

$$V_s = Ri_s = kpLAR$$

Substituting the values given above, the signal voltage is found to be

$$V_s = 0.75L$$

where V_s is in volts and L in lumens per square centimeter, or, as is generally found more convenient, V_s is in millivolts and L in millilumens per square centimeter.

Actual response curves taken of good iconoscopes have been found to have an initial slope of more than 1 millivolt per millilumen per square centimeter. However, the agreement between the measured signal output curves and the response predicted is within the accuracy of the assumptions that were made.

The two sets of curves reproduced in Fig. 9.18 are typical of the response of the iconoscope. Those shown on the left represent the

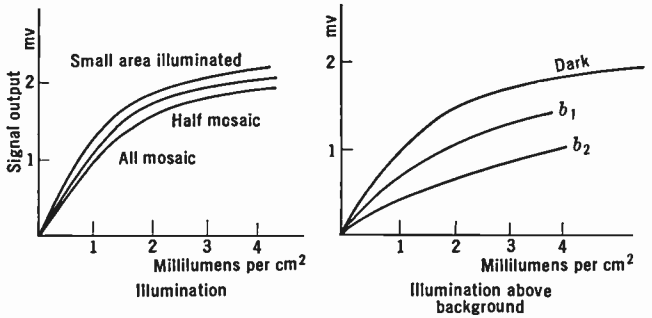


Fig. 9.18. Response of Iconoscope under Various Conditions of Illumination.

signal output as a function of the intensity of illumination for three different sizes of area illuminated. The family of curves on the right was obtained by measuring the signal voltage generated across a 10,000-ohm coupling resistance when a spot of light occupying 1 percent of the area of the mosaic was projected on the mosaic. In addition to the spot of light, a uniform background as indicated covered the entire mosaic. The highest curve a is the result of the measurement when the background illumination was zero. For this curve the initial slope corresponds to a sensitivity of 0.95 millivolt per millilumen. As the light intensity is increased, the slope of the curve becomes less. The shape of the response curve over a large part of its range is logarithmic, but departs from this at the two extremes of illumination. The curves b_1 and b_2 illustrate the effect of a background illumination. It is evident that, as the background illumination increases, the initial slope D is related to the background illumination L_B as follows:

$$D = \frac{D_0}{1 + mL_B}$$

where D_0 is the slope for zero background and m is an empirical constant. It should be noted that this relation is derivable from the logarithmic form of the response curve. However, the magnitudes of the constants are functions of the areas illuminated by the spot and the background.

The color response of the iconoscope depends upon the activation of the mosaic. By properly selecting the activation schedule, the response can be shifted to meet the operating requirements over a wide range. Figure 9.19 shows the color response of a typical normal iconoscope.

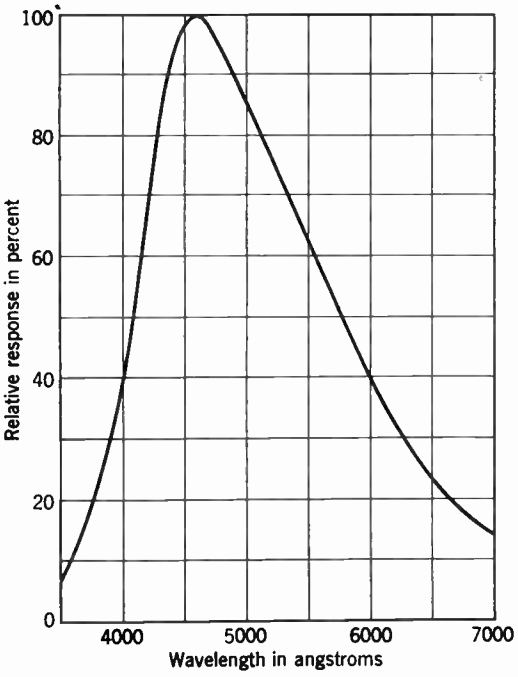


Fig. 9.19. Spectral Response of an Iconoscope.

Since the mosaic of the iconoscope is surfaced with silver particles which are very small compared with the beam size, the resolution which can be obtained is limited by the beam. The resolution of the tube is usually measured, as has been explained, by projecting onto the mosaic a test pattern consisting of groups of converging lines (see Fig. 9.9). With such a pattern it can be shown that the horizontal resolution of the normal iconoscope is in the neighborhood of 700 or

800 lines. Since the video amplifier response extends to about 4 megacycles, a picture of this definition cannot be transmitted over the system. However, the high limiting resolution insures a small aperture defect over the working region.

9.12 Limiting Sensitivity. The question of the statistical fluctuation and its relation to the minimum light that can be used to produce a picture is of fundamental importance. Broadly speaking, these fluctuations arise from the granular nature of electricity and are present in every electronic device and every circuit element having a resistive component. As has been shown, the iconoscope generates a certain signal voltage across the coupling impedance. At the same time, there appears across this impedance a random fluctuating voltage due to the thermal agitation of the electrons in the coupling resistance. The video amplifier amplifies not only the picture signal but also the voltage fluctuations of the coupling resistance, or, to adopt the terminology of amplifier practice, the "noise." If the voltage generated by the picture signal becomes of the same order as the noise, the picture becomes lost in noise and is no longer intelligible. This sets a definite lower limit to the light which can be used to produce a picture from a given tube.

Before making any calculations on the sensitivity of the iconoscope, it is necessary to assign quantitative values to the psychological effect of the picture-to-noise ratio. Tests have been made to determine the effect on the observer of various ratios of peak picture signal in an average picture to the root-mean-square noise. It has been found that, if the root-mean-square noise is equal to 30 percent of the picture signal, the picture is still recognizable, but the definition is decreased and the picture tiring to watch. In addition, it was observed that, if the ratio remains constant but the picture amplitude is decreased, the noise becomes less objectionable; however, the effective resolution is not increased.

If the ratio of signal-to-noise is 10 to 1, a good picture can be obtained, but the noise is still very noticeable. Such a picture is usable and has fair entertainment value.

When the noise is reduced to less than 3 percent of the picture signal it becomes practically unnoticeable, and the picture may be considered excellent.*

Most of the calculations of sensitivity which follow will be based on an allowable signal-to-noise ratio of 10. Although this amount of

* These tests were made with noise evenly distributed over the frequency band. If high-frequency noise predominates the percentage rms noise permissible is greater.

noise would be greater than should be tolerated if conditions made it possible to avoid it, such a picture would, nevertheless, have reasonably good entertainment value and could be broadcast where program continuity made its transmission necessary.

The noise which must be considered in sensitivity calculations arises primarily from two sources: first, the coupling resistance, or impedance, between the iconoscope and the video amplifier; and, second, the first tube of the amplifier.

The thermal noise generated by a resistance is found to have the following characteristics. It appears as a voltage generated across the ends of the resistance, and is entirely independent of any currents flowing through it. The fluctuations are completely random, so that their energy is distributed uniformly over the frequency spectrum. Furthermore, the mean square voltage is proportional to the temperature of the resistance and to its magnitude or, to be more general, the magnitude of the resistive component of any impedance. All this can be formulated in the following equation:

$$\overline{e_n^2} = 4kTRF$$

where k is the Boltzmann constant and T the absolute temperature, the product $4kT$ having a value of 1.6×10^{-20} joule at room temperature. R is the resistive component of the impedance, F the frequency band under consideration, and $\overline{e_n^2}$ the mean square noise voltage.

In addition to the noise generated by the coupling impedance, noise is generated by the amplifier itself. This noise is primarily due to the first tube of the amplifier, but the coupling impedance between the first and second stage may also contribute an appreciable amount. The noise produced by the amplifier depends, of course, upon the tubes used. A noise characteristic corresponding to a 500- or 1000-ohm resistor in the grid circuit of the first tube will represent a typical, well-built amplifier.

Heretofore a 10,000-ohm coupling resistance between the iconoscope and the picture amplifier has been assumed in discussing the performance of the tube. In actual television practice the coupling cannot be as simple as this, because the capacity of the iconosope and amplifier input circuit causes an attenuation of the high-frequency components of the picture signal. However, in order to estimate the sensitivity of the pickup tube, the simple coupling circuit consisting of a 10,000-ohm resistance will be retained and the attenuation of the upper frequencies neglected. Details of the more complicated circuits needed to maintain a uniform frequency response will be given in Chapter 13. There

it will be seen that the limiting sensitivity that can be obtained with the corrected circuit networks actually used does not differ greatly from that given under the present simplification.

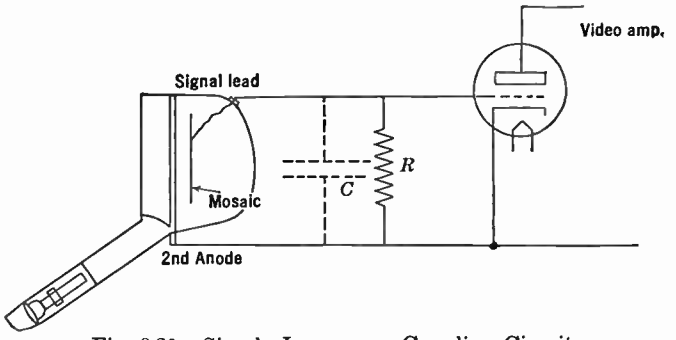


Fig. 9.20. Simple Iconoscope Coupling Circuit.

The mean square noise voltage generated by the 10,000-ohm resistor over a 4-megacycle frequency band is

$$\overline{V_N{}^2} = 1.6 \times 10^{-20} \times 10^4 \times 4 \times 10^6$$
$$= 6.4 \times 10^{-10}$$

The iconoscope, being a high-impedance current generator (actually its zero-frequency impedance is about 5 megohms when a 0.2-microampere beam is used), produces a signal voltage across the coupling resistor which is given by

$$V_s = pkALR \text{ volts}$$

or, if the values from page 333 are used, by

$$V_s = 0.75L \text{ volts}$$

The ratio S of signal to root-mean-square noise is therefore

$$S = \frac{V_s}{\sqrt{\overline{V_N{}^2}}} = 3 \times 10^4 L$$

Assuming that the signal-to-noise ratio cannot be less than 10, the lower limit of light which will give a usable picture signal is

$$L = 3 \times 10^{-4} \text{ lumen per square centimeter}$$

Similarly, to obtain a picture with a signal-to-noise ratio such that the noise is less than 3 percent of the picture signal, the illumination on the mosaic must be at least a millilumen per square centimeter.

The order of magnitude of the measured sensitivity agrees fairly well with these calculations. Tests on a good iconoscope have been made indicating that a satisfactory image can be obtained of scenes having a brightness of 15 candles per square foot, using a lens with an $f/4.5$ aperture. The illumination on the mosaic, taking account of reflection losses in the lens, will then be about 0.4 millilumen per square centimeter.

Usually the illumination for actual television transmission is made, if possible, at least three times this amount, to insure that the picture signal will be well above all spurious signals. This means that a surface brightness of 50 candles per square foot is desirable. A reflection coefficient of 25 percent being assumed, the incident illumination of the scene should be $50\pi/0.25 = 600$ foot-candles. Of course, a transmittable picture can be had with less than 20 percent of this illumination, but the figure given above can be regarded as desirable.

TABLE 9.1. SURFACE BRILLIANCE OF OUTDOOR SCENES

Scene	Location	Time (E.S.T.)	Date	Weather	Surface Brilliance (candles/ft^2)
Sixth Avenue	New York, N. Y.	9:30 A.M.	4–25–35	Clear	6½
Sixth Avenue	New York, N. Y.	1:15 P.M.	4–25–35	Overcast	40
Times Square	New York, N. Y.	1:30 P.M.	11– 6–34	Light rain	40
Parade	East Orange, N. J.	10:30 A.M.	11–29–34	Light rain	40 to 60
Street	Rockland, Me.	1:15 P.M.	7– 5–36	Overcast	100
Street	Warrenton, N. C.	3:15 P.M.	6–30–35	Clear	130
Street	Harrison, N. J.	3:30 P.M.	8–15–34	Hazy	130
Street	Harrison, N. J.	9:30 A.M.	8–15–34	Rain	16
River	New York, N. Y.	2:30 P.M.	10–24–35	Hazy	50
River	Pennsville, Del.	1:30 P.M.	6–29–35	Hazy	350
Bay	Cape Charles, Va.	10:00 A.M.	6–30–35	Clear	250
Beach	Atlantic City, N. J.	2:00 P.M.	8–18–34	Hazy	500
Football game	New York, N. Y.	1st quarter	11–17–34	Clear	55
		2nd quarter		Hazy	50
		3rd quarter		Hazy	27
		4th quarter			16
		End			2
Baseball game	New York, N. Y.	1:00 to 3:40	9– 8–35	Clear	70 to 100
Snow bank	Harrison, N. J.	10:00 A.M.	1–24–35	Bright sunshine	700
Open field	Bethel, N. C.	3:45 P.M.	7– 1–35	Severe thunderstorm	2

In order to give an idea of the brightness encountered in typical outdoor scenes, Table 9.1 is reproduced from a paper by Iams, Janes, and Hickok.* Since it can be assumed that the iconoscope has sufficient

* See Iams, Janes, and Hickok, reference 7.

sensitivity to transmit a usable picture at a brightness as low as 10 candles per square foot, the versatility of the tube will be evident.

If the walls of a normal iconoscope are photosensitive and illuminated with a small amount of light, the photoemission from the bulb so alters the conditions in the vicinity of the mosaic that the working sensitivity is about doubled and the shading improved. This expedient, which is of considerable practical importance, was first introduced by the Electrical and Musical Industries of England, and has since been widely adopted.*

9.13 Depth of Focus. The depth of focus of the objective used with an iconoscope is an important consideration. In fact, if it were not for this limitation, the iconoscope with very large-aperature objectives would be sufficiently sensitive to transmit a picture under all practical conditions, both indoors and out. However, the aperture can be increased only at the expense of depth of focus. In the photographic camera, depth of focus with short exposure times is obtained by means of high-aperture, short-focal-length lenses. This procedure permits decreasing the diameter of the lens without decreasing the intensity of illumination on the photographic plate. Since depth of focus in the object field is dependent only upon the diameter of the lens and not upon the numerical aperture, it can be increased without loss in sensitivity.

It has been shown above that in the iconoscope the response is proportional not only to the intensity of illumination but also to the mosaic area. Stated in another way, the response is proportional to the total light falling on the mosaic. This light is in turn proportional to the absolute free area of the lens rather than to its numerical aperture. Therefore, it is not possible to follow photographic practice and increase the depth of focus by using a miniature iconoscope and a short-focal-length lens. The conclusion is that, for a given sensitivity of the iconoscope and a given illumination of the scene transmitted, the signal obtained is inversely proportional to the square of the depth of focus in the object field.

The relation between depth of focus and the lens diameter can be shown as follows. Figure 9.21 shows that the object and image distances are x and y, respectively, and the focal length of the fens is f. These quantities are related by the equation

$$y = \frac{fx}{x - f}$$

* See Janes and Hickok, reference 10.

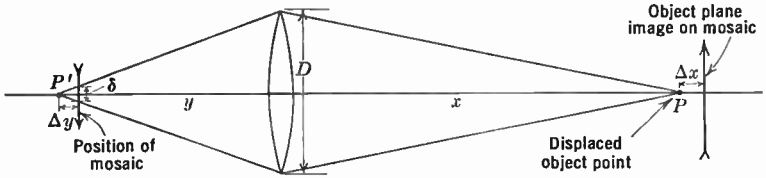


Fig. 9.21. Determination of Depth of Focus.

By differentiation, a change Δx in the object distance, x, is found to produce a change in the image distance

$$\Delta y = -\frac{y^2}{x^2}\Delta x$$

This will produce a circle of diffusion of diameter δ in an image at y, with

$$\delta = \frac{D}{y}\Delta y = D\frac{y}{x}\frac{\Delta x}{x}$$

where D is the diameter of the lens. For a given object, the effective resolution will be equal to the diameter of the circle of diffusion divided by the magnification. Since the magnification, m, is given by y/x, the quantity to be determined, that is, δ/m, can be expressed as follows:

$$\frac{\delta}{m} = D\frac{\Delta x}{x}$$

Therefore, the depth of focus in the object field $\Delta x/x$ is dependent only upon the diameter of the lens, D.

9.14 Pickups for Motion-Picture Film. So far, no mention has been made of the transmission from moving-picture films. Two methods of reproduction will be discussed. The first makes use of a continuous motion projector. Such a projector is arranged so that the light on the screen remains constant and the consecutive frames fade smoothly into one another. The operation of the iconoscope with such a projector is exactly the same as in studio or outdoor pickup. Furthermore, the frame frequency of the film need bear no relation to the scanning frequency of the tube.

A second method, somewhat more complicated but nevertheless found in practice to be quite satisfactory, is to project the picture onto the mosaic during the return time of the beam. An ordinary moving-picture projector, using a standard film which is run at 24 frames per second, is equipped with a shutter, usually in the form of a rotating,

apertured disk, which allows the light image to pass sixty times a second for a period of about $\frac{1}{800}$ second. With this arrangement it will be obvious that images from consecutive frames will pass through the shutter alternately two and three times. The scanning beam is synchronized with the shutter in such a way that the image is flashed on the mosaic only during the time between the end of one scanning pattern and the start of the next, while the beam current is biased to cutoff. The light image flashed on the mosaic causes a potential image to be formed in the manner described in the section discussing the theory of the iconoscope. This potential image gives rise to the picture signal when the unilluminated mosaic is then scanned. A reproduced picture thus formed differs in no way from the ordinary picture in appearance. Because of the short duration of the light image on the mosaic, the instantaneous illumination must be high, but this presents no particular difficulty since the light produced by an ordinary projector is ample.

9.15 The Type RCA1850-A Iconoscope. Although the essential features of the design, construction, and activation of the iconoscope are at present essentially the same as those of the iconoscope described in the first writing of this chapter in 1940, certain changes have been made which have considerably improved the operating characteristics of the tube. Figure 9.22 shows schematically the present form of the iconoscope. The dipper-shaped envelope shown in Fig. 9.1 is still used. However, it will be noted that there is a considerable change in the support of the mosaic and of the wall coating on the inside of the glass envelope.

The mosaic is formed on a carefully selected piece of mica 0.001 to 0.002 inch thick. The mosaic of silver is formed by allowing a fine dust of silver oxide to settle on the mica surface. Before the silver dust is deposited, this surface is, however, sand-blasted to give it a fine mat texture. The silver oxide dusting procedure is carried out in a specially designed double-chamber dust-settling unit which minimizes aggregations and nonuniform deposition of the fine silver oxide powder. The layer is not deposited all at once; instead, a thin coating is allowed to settle on the mosaic and it is then baked in air at 950°C (for approximately 30 seconds) to reduce the silver oxide to silver. More silver oxide is then deposited, and the bake is repeated. This is done three or four times until the desired amount of silver has formed on the mica. The correct amount of silver is measured by shining light through the mosaic and determining the reduction in light transmitted. A well-formed mosaic will transmit 60 to 70 percent.

After the silver mosaic has been formed, the signal plate is coated on the back of the mica. Graphite in the form of Aquadag is used for this purpose. The mosaic is mounted on a circular mica disk which just matches the inside diameter of the iconoscope envelope. It is held on this disk by four mica strips which are riveted to the disk. These mica strips are coated with a film of platinum on the beam side. The

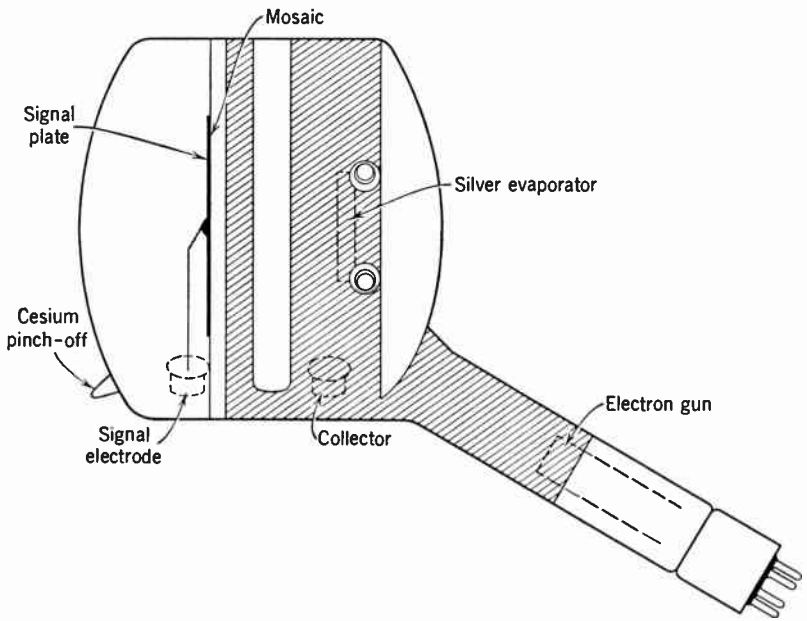


Fig. 9.22. Diagram of an 1850-A Iconoscope.

coating helps to reduce the spurious signal generated by the mosaic. A nichrome ribbon which passes through slots in the mica supporting disk and makes contact with the Aquadag signal plate of the mosaic serves as signal lead.

The mosaic supporting disk is held in place in the glass envelope by four glass rods which are sealed into the cylindrical walls of the tube. In order to avoid distortion and loss of contrast in the projected image, the front window of the envelope must be as optically perfect as possible. The glass is thin enough, however, so that it can be curved rather than flat. The rear closure of the envelope is also left clear to admit back lighting, a means of reducing spurious signals already mentioned in section 9.12 and one which will be discussed a little further in connection with the operation of the tube. A broad band is sand-

blasted around the inside cylindrical wall of the iconoscope to reduce internal reflections falling on the mosaic. This sand-blasted band is coated with evaporated aluminum which extends down into the gun neck. The aluminized coating serves as part of the high-voltage anode of the gun and also as the secondary electron collector of the beam electrons.

Two small silver evaporators are mounted on the cylindrical wall very close to the front window. The gun itself is similar to that used in earlier iconoscopes but has been modified considerably in detail. This gun will be described more extensively in Chapter 12.

After the tube has been evacuated and given an outgassing bake, it is activated in much the same way as the earlier iconoscopes. The silver is oxidized by a high-frequency glow discharge in oxygen at a few hundred microns pressure. The oxidation is carried to the first blue or blue-green. About 6 milligrams of cesium is then driven into the tube, and the tube is given a 10-minute bake at 180 to 200°C. It is then sealed off the system and placed in performance-testing equipment. A very small amount of silver is evaporated from the two side evaporators. Upon evaporation of this silver, the sensitivity is seen first to rise and then to fall. The deposition of silver is continued until the response drops to about 50 percent of its maximum value. The tube is then removed and given a short bake. If the activation schedule has been properly carried out, the sensitivity will rise and will finally become several times the maximum observed before the deposition of silver. The tube is tested not only for its sensitivity but also for image quality, shading, etc.

In operation, it is found that the performance of these tubes is markedly improved both with respect to the spurious signal and to some extent in effective sensitivity by the use of back lighting. This is a small amount of light which is allowed to enter the tube through the back window so as to fall on the aluminum layer on the side wall of the envelope surrounding the target. There is enough photoemission from this aluminum coating to improve very markedly the uniformity of the charge pattern on the mosaic and to keep the average potential of the mosaic from becoming positive with respect to the secondary electron collector band.

This description, while it covers the more important features of the present-day iconoscope, must of necessity omit a number of details both of construction and of activation which are involved in the actual manufacture of these tubes.

It will be noted that mica is still preferred as dielectric for the mosaic, in view of the great uniformity in thickness which is obtainable with it. However, the study of substitute materials, which are of particularly great value when good-quality mica is not readily available, has continued. Relatively good results have been obtained with barium titanate, whose very high dielectric constant permits the use of relatively thick layers to attain a prescribed capacity.

9.16 Summary. For a good many years, the iconoscope was the only pickup tube used for television broadcasting. Its sensitivity, while considerably below that of the pickup tubes used at present, was, nevertheless, sufficient for reproducing all studio scenes and a great many of the outdoor scenes encountered in practice. Where the illumination was as high as 1000 to 2000 foot-candles, the tube imposed essentially no limitation on the performance of the system. However, even under these conditions, great care had to be exercised in shading compensation if the picture was to come up to the standards required by broadcasting. As the illumination decreased below 1000 foot-candles, it was necessary to sacrifice picture quality, first by using larger-aperture, small-depth-of-focus lenses, and then by tolerating a large amount of noise in the picture. Below about 50 foot-candles, picture transmission became impractical.

At present, the iconoscope has been superseded by the image orthicon and Vidicon for most studio and outdoor pickup requirements. It is, however, used extensively for motion-picture film transmissions. The high definition and good storage properties make the tube very effective in this application. However, unless the iconoscope is operated with great skill, shading can be very objectionable, particularly in scenes where the light level and contrast are low.

REFERENCES

1. Fritz Schröter (editor), *Fernsehen*, Springer, Berlin, 1937.
2. Busch and Brüche (editors), *Beiträge zur Elektronenoptik*, Barth, Leipzig, 1937.
3. V. K. Zworykin, "The Iconoscope—A Modern Version of the Electric Eye," *Proc. I.R.E.*, Vol. 22, pp. 16–32, January, 1934.
4. V. I. Krasovsky, "On Light Storing Devices," *I.E.S.T.* (*Izvestia Elektro-Promishlenosti Slabovo Toka*), No. 2, pp. 24–34, February, 1936.
5. R. Urtel, "The Operation of Electron Beam Scanning Devices with Storage," *Z. Hochfreq. Elektroakust.*, Vol. 48, pp. 150–155, November, 1936.
6. V. K. Zworykin, G. A. Morton, and L. E. Flory, "Theory and Performance of the Iconoscope," *Proc. I.R.E.*, Vol. 25, pp. 1071–1092, August, 1937.

7. H. Iams, R. B. Janes, and W. H. Hickok, "The Brightness of Outdoor Scenes and Its Relation to Television Transmission," *Proc. I.R.E.*, Vol. 25, pp. 1034–1047, August, 1937.
8. W. Heimann and K. Wemheuer, "Operation of the Electron Beam Picture Scanner," *Elek. Nach. Tech.*, Vol. 15, pp. 1–9, June, 1938.
9. M. Knoll, "The Significance of the 'Redistribution Electron Effect' for the Operation of Picture Scanning Tubes," *Z. tech. Physik*, Vol. 19, pp. 307–313, October, 1938.
10. R. B. Janes and W. H. Hickok, "Recent Improvements in the Design and Characteristics of the Iconoscope," *Proc. I.R.E.*, Vol. 27, pp. 535–540, September, 1939.

10 Television Pickup Tubes

In the preceding chapter, the theory and performance of the iconoscope, the first type of pickup tube to find general employment, were discussed in some detail. The methods used in improving the sensitivity of pickup tubes, particularly in respect to sensitivity, will be the subject of this chapter.

Even though the iconoscope, under the exacting test of a number of years of television broadcasting service, proved itself to be a fairly satisfactory pickup device, there was a very real need of an increase in sensitivity in order to improve the television picture. Increased sensitivity would permit the use of lenses having greater depth of focus, would extend the possibilities of obtaining an optimum picture in outdoor transmission, and would improve studio working conditions.

The problem of increasing the sensitivity of the pickup tube may be approached in at least three different ways. First, the overall efficiency of the mosaic or target can be increased by improving its capability of storage or the recoverability of the stored charge. Second, the noise generated may be reduced while keeping the signal output the same. Finally, the quantity of charge acquired by the target per unit light flux may be increased. This list does not, of course, exhaust all possibilities but it will at least indicate some lines of investigation which have been pursued.

The remaining sections relating to pickup tubes will first consider some of the basic factors which set the limit of sensitivity of any pickup device. Then these approaches will be taken up, and some of the solutions which have been found for the problem will be described. It should be remembered that most of the developments described represent work that has been carried out in various research and development laboratories and has only been possible through the coordinated work of many scientists and engineers. Furthermore, the evolution of the pickup tube has not ended, and continued research will lead to more sensitive and better pickup devices.

10.1 Limiting Sensitivity of Pickup Devices. Before taking up methods of improving the sensitivity of television pickup tubes, an estimate must be made of the sensitivity that could be achieved by a perfect pickup device. The ideal device will be assumed to be a light-sensitive storage target combined with scanning beams for recovery of the stored information. It is further postulated that the information will be removed and amplified in such a way that the noise or random error in the information will not be increased over that which is contained in the initial stored information. In other words, the amplifier does not itself add more noise to the signal, although it does, of course, amplify the noise already existing in the information from the target.

Ordinarily, light is thought of as a continuous flux, and illumination is measured in lumens per square centimeters or watts per unit area. It is necessary here, since a statistical approach is to be adopted, to consider light in its more basic aspect, namely, as photons. Each photon has, as has been noted in earlier chapters, associated with it an energy hc/λ, where λ is the wavelength of the light, so that illumination may be expressed in terms of the number of photons per second incident on a unit area.

If a small radiation detector in the form of a square with sides of length h has an illumination L falling on it, this important question might be asked: Can a fundamental lower limit to a change of illumination ΔL be found below which it is impossible to detect the light change irrespective of the sensitivity of the device? Until the granularity of light is recognized, no such limit can be established. However, the essentially random arrival of light photons as postulated by quantum mechanics introduces an uncertainty which sets the lower limit of detectability. To determine the ultimate sensitivity which can be achieved with the detector, it is assumed that every incident photon produces a recoverable element of information and that the detector has a storage or retention time T over which the information is accumulated. The number of photons N arriving on the sensitive area during a period is, on the average,

$$N = \rho L h^2 T \tag{10.1}$$

where ρ is the number of photons per second per lumen. ($\rho = 1.4 \times 10^{16}$ photons per second per lumen for "white" light.)

However, since the photon arrival is random there is a statistical deviation in the number in any period which has a root-mean-square value $\delta = \sqrt{N}$. If the change in number of photons ΔN, due to the change ΔL in illumination, is so small that it is comparable with δ, the

change cannot be distinguished from a statistical fluctuation. In other words, if the detector records a change in any one period of

$$\Delta L = \sqrt{L/(\rho h^2 T)} \tag{10.2a}$$

the probability of its being a change in the average illumination (and not a random fluctuation) is of the order of 1/2. As ΔL becomes greater, the probability that the observed change is not accidental increases rapidly. This may be formulated by replacing Eq. 10.2*a* by the expression

$$\Delta L = k\sqrt{L/(\rho h^2 T)} \tag{10.2b}$$

where k is the "coefficient of certainty." If k is given the value 5 in the expression for the limiting sensitivity, the probability that ΔL represents a change in illumination rather than random fluctuation exceeds 99 percent.

Exactly the same type of reasoning applies in the determination of the limiting contrast which can be discerned with any type of image-forming pickup device. If the scene contains an object of brightness $B + \Delta B$ against a background of brightness B, the object therefore has a contrast $C = \Delta B/B$. The corresponding area imaged on the sensitive target of the pickup device will receive an illumination

$$L + \Delta L = (B + \Delta B)/(4A^2)$$

against a background illumination of

$$L = B/(4A^2)$$

where A is the f-number of the lens of the camera. The contrast of the image relative to the background is $C = \Delta L/L$.

Since the detectability of the difference in illumination depends upon the number of photons arriving in the area under consideration during the integration interval of the pickup device rather than the surface density of photons, the size of the area must be considered. It does not detract from the generality of the argument to assume that the object is a square area with sides of length h on the sensitive image target. The criterion of Eq. 10.2*b* can now be used to determine the illumination of the target at which an object of given size and contrast becomes detectable over the statistical fluctuation in the number of photons in a similar area of the background:

$$C = \frac{\Delta L}{L} = k\sqrt{\frac{1}{\rho h^2 LT}}$$
$$L_{\min} = \frac{k^2}{\rho C^2 h^2 T} \tag{10.3}$$

The illumination $L_{\min}$ on the sensitive target is the fundamental lower limit of illumination which will permit the detection or seeing of an object of area h^2 and contrast C. No form of optical transducer, be it a photographic film, a television pickup device, or the eye, can exceed this performance. Actually, all these devices fall considerably short of this sensitivity. It is possible now, with the aid of Eq. 10.3, to formulate a criterion of merit for any pickup device. Thus, a figure of merit M may be defined as

$$M = L_{\min}/L_E \tag{10.4}$$

where L_E is the experimental limiting illumination for the detection of a test object of contrast C and area h^2.

By employing the same type of reasoning used above, it is possible to derive an ideal figure of merit which will apply to a wide variety of pickup devices irrespective of their exact mechanism of operation. This derivation involves taking into account factors describing the efficiency of gathering and utilizing information from the light image and considering additional statistical fluctuations or noise which may be introduced by the device. In general, a photoelectric surface will not respond to every incident photon. A factor γ indicating the quantum efficiency of the sensitive surface must, therefore, be applied. This factor is the ratio of primary photoelectric events, e.g., emitted photoelectrons, in the case of a photoemitter, to the number of photons. A second factor ε will also be required to describe the efficiency of utilization of the photoelectric effects. This factor relates the number of elements of information with the number of photoelectric events required to produce them. If, for example, a photoemission pickup device has a photosensitivity of 40 microamperes per lumen and requires two electrons for a useful element of information, γ will have the value 0.02 and ε the value 0.5.

All real pickup devices introduce additional fluctuations so that the limit of contrast discrimination is not solely due to the statistical fluctuations from the granularity of the photoelectric phenomenon. These additional fluctuations may be amplifier noise, thermal effects, etc., in

the case of television devices; fogging in a photographic emulsion; and random nerve pulses in the eye. This may be taken into account by including a term n_D to express the mean-square deviation of these additional fluctuations. Then the root-mean-square deviation of the optical detector becomes $\delta = \sqrt{N + N_D}$ (where $N_D = n_D T$). Consequently, Eq. 10.2*b* becomes

$$\Delta L = k \sqrt{\frac{(\rho h^2 L + n_D)}{\rho^2 h^4 T}} \tag{10.5}$$

Including all the factors introduced in the preceding paragraph, an equation corresponding to Eq. 10.3 can be derived as follows:

$$C = \frac{\Delta L}{L} = k \sqrt{\frac{(\varepsilon\gamma\rho h^2 L + n_D)}{\varepsilon^2\gamma^2\rho^2 h^4 T L^2}}$$

$$C^2 = \frac{k^2}{\varepsilon\gamma\rho h^2 T L}\left(1 + \frac{n_D}{\varepsilon\gamma\rho h^2 L}\right) \tag{10.6}$$

with $M = \dfrac{L_{\min}}{L} = \dfrac{k^2}{L\rho C^2 h^2 T}$.

In pickup devices where n_D is very small, the term involving it can be neglected, and the expression for the limiting contrast becomes

$$C^2 = \frac{k^2}{\varepsilon\gamma\rho h^2 T L}$$

This leads to an ideal figure of merit of

$$M = \varepsilon\gamma \tag{10.7a}$$

If, on the other hand, the additional fluctuations are large enough to be dominant, so that

$$\frac{n_D}{\varepsilon\gamma\rho h^2 L} \gg 1$$

the expression for contrast becomes

$$C^2 = \frac{n_D k^2}{\varepsilon^2\gamma^2\rho^2 h^4 L^2 T}$$

leading to a figure of merit:

$$M = \frac{\varepsilon^2\gamma^2\rho h^2 L}{n_D} \tag{10.7b}$$

The iconoscope, which was described in the preceding chapter, falls in the category of devices to which Eq. 10.7*b* applies. The eye, on the other hand, together with certain types of pickup devices which will be discussed below, follows Eq. 10.7*a* over the range of useful vision.

In discussing possible ways of improving pickup devices, methods of improving the operation of the mosaic or target which increase γ and ε will be treated first. Next, ways of increasing the stored charge on the target, by the use of electron imaging, will be considered. This will be followed by a discussion of methods of reducing n_D by means of secondary emission multiplication. Finally, pickup devices not employing external photoemission, together with a number of special tubes, will be described.

METHODS OF INCREASING TARGET EFFICIENCY

The most immediately apparent way of increasing the effectiveness of a mosaic or target of a pickup device (i.e., increasing γ in Eqs. 10.7*a* and 7*b*) is by improving the photosensitivity of its surface. A great deal of work has been done along this line, with the result that in present-day iconoscopes the saturated photosensitivity is 15 to 20 microamperes per lumen as compared with one-third this value a decade and a half ago. Still further increase has been obtained by the employment of photocathodes which are separate from the storage target, as will be described in sections dealing with the image iconoscope and the image orthicon.

The second improvement that must be considered is the reduction of the losses which have been discussed in Chapter 9, describing the operation of the iconoscope. The two principal losses associated with the mosaic are the lack of saturation of photoemission of the mosaic and the lowering of efficiency due to electron redistribution. Two general methods have been explored for solving this second problem, namely, the use of a two-sided target and scanning with a low velocity beam.

10.2 The Two-Sided Target. The generation of the picture signal from the scanned mosaic of an iconoscope is possible only when the elements can be driven to an equilibrium potential such that the effective secondary emission ratio is unity. Therefore, the potential of the elements must necessarily be close to that of the electrode collecting electrons emitted from it, and the field collecting photoelectrons is consequently small. However, by constructing a target in such a way

that the elements extend through it, so that one side is scanned and emits secondary electrons while the other emits photoelectrons, the photoemission can be field-saturated.

The most straightforward form of two-sided target is one consisting of conducting elements which extend through an insulated metallic mesh. The target is mounted in its envelope so that its surface is normal to the axis of the electron gun. The ends of the elements on the side of the target away from the gun are photosensitized. A ring electrode is provided to collect photoelectrons. A second electrode is located on the beam side of the target to collect secondary electrons. Figure 10.1 illustrates the general arrangement of this type of tube.

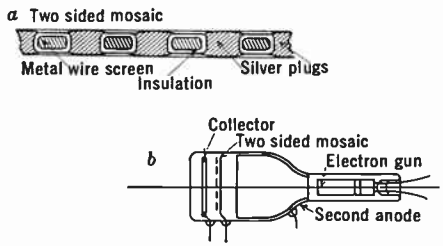


Fig. 10.1. Pickup Tube with Two-Sided Target.

The construction of this form of target is technically very difficult because of the extraordinarily high degree of perfection needed to generate a picture having the degree of excellence required by commercial television. Experimental two-sided targets have been made by the following procedure. A uniform mesh either in the form of a woven wire screen or an electrodeposited net is used as a base. In order to avoid a grainy appearance in the picture the mesh must be extremely fine, with, for example, 500 wires to the inch for a 3-inch target. The screen wires are coated with an insulating layer of vitreous enamel, lacquer, or ceramic. Great care must be taken to avoid cracks or pinholes as such imperfections will generate strong spurious signals.

The metal plugs which serve as elements can be made in a variety of ways. One method is to fill the interstices of the mesh with a paste of silver oxide suspended in a binder which can later be driven off by heating. It should be noted that the net area covered by the plugs should be as large as possible in order to insure efficient utilization of incident light.

This type of target permits the use of a sufficiently strong photoelectron-collecting field so that the emission is saturated. It does not, however, overcome the redistribution losses which occur with an icono-

scope mosaic. Also, the same spurious signal and shading consequent on redistribution will exist.

A so-called barrier grid will, in a large measure, overcome both these difficulties. The barrier grid is a conducting grid which surrounds each element on the beam side of the target. The general arrangement of a barrier-grid target is illustrated in Fig. 10.2*a*.

The target is mounted in the tube in the same way as the simple two-sided target, with the barrier shield maintained at about 45 volts negative with respect to the electrode, which is normally the secondary

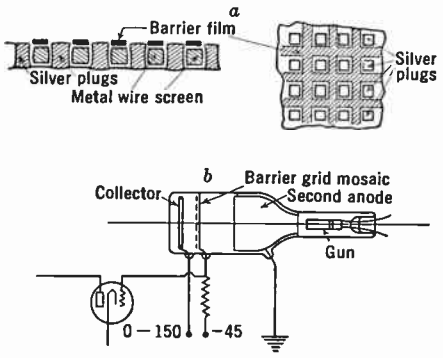


Fig. 10.2. Barrier-Grid Target and Pickup Tube.

electron collector. A circuit diagram for the tube is shown in Fig. 10.2*b*.

When the tube is in operation the equilibrium potential of the elements under the beam is slightly positive with respect to the barrier grid. Under these conditions electrons escaping from an element must either be collected by the barrier grid immediately surrounding it or by the secondary electron collector. Because of the field existing between the barrier and the collector there is no possibility of electrons going from one element to the next.

Experimental tubes of this type have been built which have six to ten times the sensitivity of a normal iconoscope, with no shading due to redistribution. The saturated signal, instead of being a few microamperes, as in the case of an iconoscope, may be several hundred times this amount.

The technical difficulties associated with the making of these targets with the high degree of perfection required are very great. Because of this, and the fact that simpler methods exist for attaining the same ends, as will be described in the next and succeeding sections, barrier grid pickup tubes are not used for general television work.

10.3 Low-Velocity Scanning; The Orthicon. Another method of obtaining photoelectric saturation and freedom from redistribution is through the use of a scanning beam which strikes the target with so low a velocity that its secondary emission ratio is below unity. The target for a tube operating in this way may, like that of the normal iconoscope, consist of an insulating layer coated on one side with a signal plate and on the other with photoemissive elements (real or virtual) which have very high transverse resistance between each other. When the target elements are in darkness and the beam falls on them they accumulate negative charge, and their potential falls until it reaches

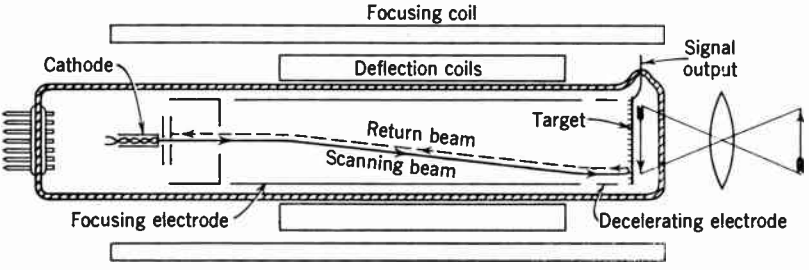


Fig. 10.3. Orthicon-Type Pickup Tube.

cathode potential. When the elements attain this potential, the beam can no longer reach them, and is turned back to be collected by more positive electrodes. Thus, the element equilibrium is established relative to the cathode potential of the gun. This form of tube is illustrated schematically in Fig. 10.3.

With this type of operation, since the secondary-emission ratio is less than unity because of the low bombarding velocity of the electrons, the element equilibrium is not affected by a positive field between the elements and other collector electrodes. Since the collector element surrounding the target is maintained positive with respect to the elements, the photoemission is saturated. Furthermore, secondary electrons emitted by the target cannot fall back on the target but are drawn to the collector. Thus, the principal causes of inefficiency and spurious signal that exist in the iconoscope are eliminated.

Although in principle the action of the low-velocity or orthicon pickup tube is quite simple, certain details of its operation require further explanation. In particular, this includes questions of the focusing and deflection of the beam, the conditions of the landing or turning back of the bombarding electrons, and certain aspects of the target.

As has been previously mentioned, the target may take the form of an iconoscope target. However, stability of the landing electrons requires that the beam strike the target nearly normal to its surface. This introduces serious optical problems if an opaque target is employed. Because of this, the target generally takes the form of a nearly transparent evaporated layer of silver, thin enough so that when it is oxidized and activated its transverse resistance is extremely high. A target thus formed permits the image to be focused on it from the rear, while electrons are emitted from its face. This type of target can be made to have an intrinsic photosensitivity as high as, or even higher than, that of the targets described in Chapter 9.

The success or failure of this form of pickup tube operation depends, for one thing, upon the possibility of maintaining in sharp focus an electron beam which reaches the target with essentially zero velocity. It has been found possible to use either electrostatic or magnetic focusing to form beams satisfying this condition. The electrostatic gun is similar to that used in a normal iconoscope. The electrons are first accelerated to a fairly high velocity, i.e., one or two hundred volts, and then decelerated in a short distance in front of the target. Adequate definition can be obtained with such a gun, but some difficulty is encountered in regard to beam stability. Unless the beam at all times approaches the target normal to its surface, a potential gradient is developed. This gradient, in turn, produces a lateral oscillation of the beam which gives rise to a moving shading in the reproduced image. Because of this, it is necessary to use a target which is curved in such a way that its center of curvature lies at the effective center of deflection of the beam when electrostatic focusing is employed.

A long, approximately uniform magnetic field is used for magnetic focusing. The gun could, in principle, be simply a very small point cathode. Actually, for purposes of better control, etc., it is somewhat more complicated than this, as will be explained in Chapter 12. Where magnetic focusing is used, the deflection becomes a difficult problem. The simplest solution consists of superimposing, upon the axial focusing field, fields at right angles in the x and y directions, which are made to vary in such a way as to give the required horizontal and vertical deflection. When the target is small this is a practical solution, and it is the method by which deflection is obtained in all low-velocity pickup tubes in current use.

However, if the target is large, as for certain earlier orthicons, the volume occupied by the varying magnetic fields is large, and an inordinately great amount of power is required to obtain the horizontal

deflection. In this case, magnetic deflection may be used for the vertical deflection, whereas horizontal deflection may be obtained by employing an electric field at right angles to the direction in which it is desired to deflect the beam. It will be recalled that an electron in a crossed electric and magnetic field follows a cycloidal path in a direction at right angles to the field, and that the time of completing the

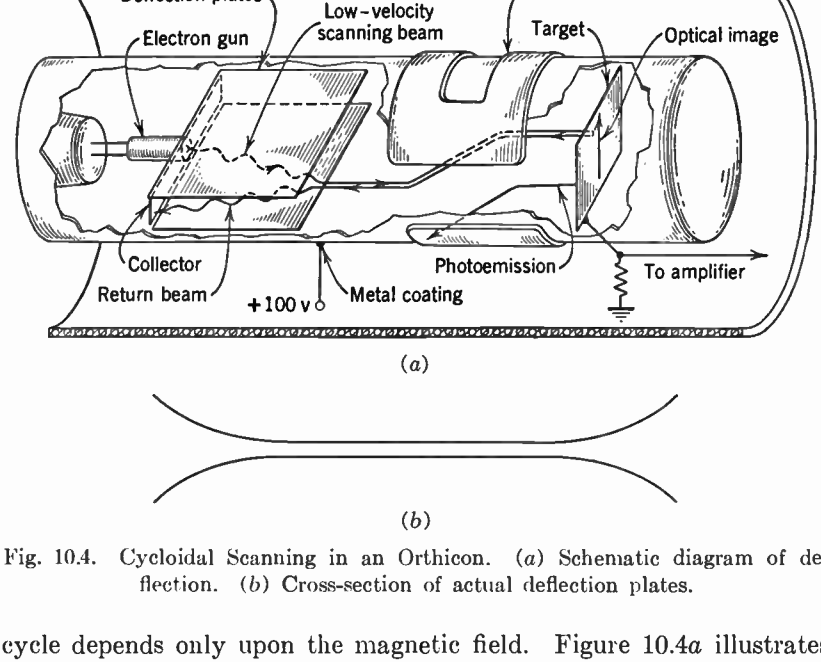


Fig. 10.4. Cycloidal Scanning in an Orthicon. (*a*) Schematic diagram of deflection. (*b*) Cross-section of actual deflection plates.

cycle depends only upon the magnetic field. Figure 10.4*a* illustrates schematically the arrangement of this deflecting system.

If the length of the deflecting plate, the velocity of the beam, and the value of the magnetic field are correctly chosen, the electrons will complete an integral number of cycles while under the deflecting plates. They will consequently emerge with zero lateral velocity, and no vertical (y) displacement, but with a horizontal displacement given by $x = \text{const } E/H$. Thus, by varying the deflecting plate voltage and, consequently, E, the beam is horizontally deflected in proportion to the deflecting voltage. Furthermore, by properly shaping the deflecting plates, the vertical displacement of the cycloid can be reduced (in principle it can be made zero), thus minimizing the space between the deflecting plates and reducing the voltage required to obtain the neces-

sary electric field. Figure 10.4*b* gives the form of the deflecting plate actually employed.

It has been pointed out that the electrons in the beam are turned back from the target when the tube is in darkness. Actually, because there is inevitably a certain amount of leakage, a small fraction of the beam, made up of electrons having the highest axial velocity, will reach the target. With light causing the elements to become more and more positive, a larger fraction of the beam is absorbed, the higher-velocity electrons being able to reach the elements first. Eventually, the elements may become sufficiently positive to be reached by all the beam electrons. In order to minimize the voltage change of an element required before this point is attained, the axial velocity spread of electrons in the beam should be as small as possible. This reduces the beam current required for complete charge erasure, which, in turn, makes possible a smaller beam diameter at the point of incidence. The question of beam velocity distribution will be discussed in greater detail in a later section. However, as the elements become more positive, the faster electrons, which reach the target with a not inconsiderable bombarding velocity, produce a greater number of secondary electrons. Eventually, a point is reached where the secondary-emission ratio is no longer below unity, and the elements are not returned to equilibrium. When this occurs, cutting off the light is no longer sufficient to return the elements to cathode potential. In addition, the electrode which collects secondary and photoelectrons from the target must be made negative. This instability with bright light is one of the objectionable features of the orthicon type of pickup tube. However, its high sensitivity and freedom from shading more than offset this relatively minor inconvenience.

INCREASING THE CHARGE STORED ON THE TARGET

As long as the target elements themselves are the source of the stored photoelectric charge, the only way of increasing this charge is by improving their photoemissivity. However, it is possible to multiply the charge when the photoelectrons come from a separate cathode and thus increase the charge stored on the target.

10.4 The Image Iconoscope. The method which has been found the most practical to obtain enhancement of the stored charge is through secondary-emission multiplication of the photoelectrons. For this, the optical image is projected onto a semi-transparent extended

photocathode. Electrons are emitted from this cathode with a density distribution identical with the intensity distribution of the light image. Thus, close to the cathode surface there is an electron image which is an exact reproduction of the optical image.

An electron-optical lens system accelerates the electrons away from the cathode and focuses them onto the target elements. These target elements are treated to have a secondary emission greater than unity. Hence, for each image electron which strikes the target several electrons leave. If the secondary emission ratio of the target is σ, the charge stored in the elements will be $\sigma - 1$ times as great as it would have been had the target elements only stored the emitted photoelectrons.

The problem of storing the image charge and removing the information with a scanning beam is essentially the same as for targets with photoemissive elements. A schematic diagram of an image iconoscope is given in Fig. 10.5.

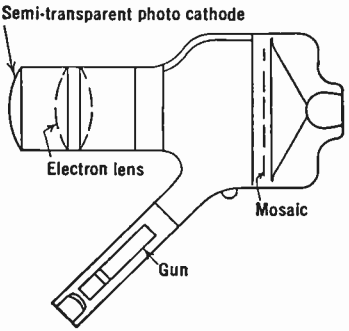


Fig. 10.5. Image Iconoscope.

The photocathode employed with image multiplier pickup tubes is almost always of the semi-transparent type, which is illuminated on one side while the electrons leave from the other. The precise sensitization of this cathode depends upon the spectral response most suitable for the application at hand.

Where the radiation employed is largely in the red and infrared, a cesium-cesium oxide-silver photocathode is best. It is formed by providing silver evaporators in the tube which are arranged so that a thin film of silver can be deposited on the inner face of the cathode. This silver film is oxidized completely by a glow discharge in oxygen at low pressure. (Note. This same discharge can also be used to oxidize the target surface as required.) Cesium is then added, and the tube is baked for a short period. The evaporation of a very small amount of additional silver and a final bake will complete the sensitization. The spectral response is given in curve a of Fig. 10.6.

At the other extreme of the visible-light high-sensitivity photocathodes is the cesium-antimony surface. This is formed by depositing a thin layer of antimony on the cathode face by evaporators located in positions similar to those used for the silver evaporations in the previously described activation. Cesium is next added, and the tube is

given a suitable thermal treatment. This type of photocathode has the highest quantum efficiency of any (visible light) photocathode known. Unfortunately, it has no red response and cannot be used where color

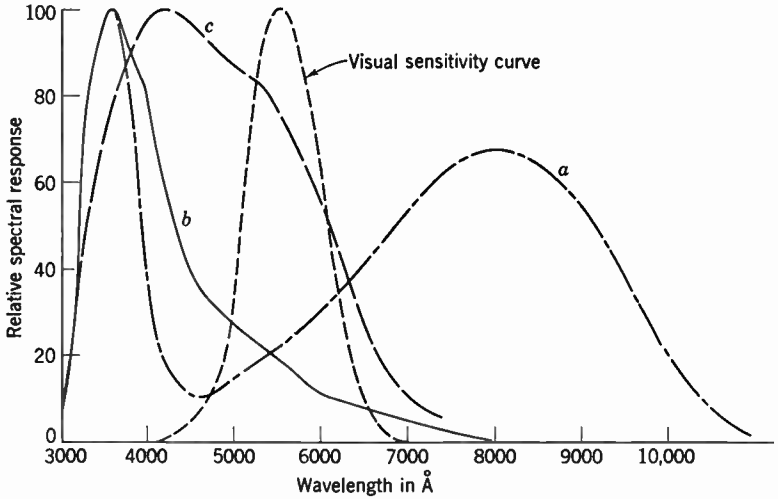


Fig. 10.6. Spectral Response of Representative Semitransparent Photocathodes.

tone fidelity is important. The spectral response is shown by curve *b* of Fig. 10.6.

Other cathodes, consisting of a bismuth-silver-cesium combination, can be formed which will have responses quite close to the visual response. Curve *c* in Fig. 10.6 is given to illustrate such a cathode. Photocathodes of cesium-cesium oxide-silver with the red response suppressed and antimony-bismuth-cesium may also serve in these tubes.

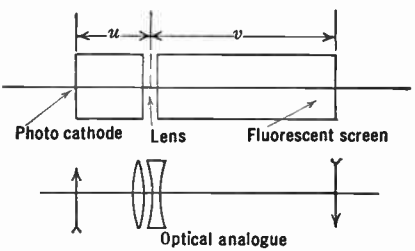


Fig. 10.7. Simple Electrostatic Image Tube.

Both electrostatic and magnetic electron optical systems have been used in image iconoscopes. The electrostatic systems are similar to those described for image tubes in Chapter 4.

The simplest electrostatic lens system is that formed by two coaxial cylinders, as illustrated in Fig. 10.7. This has been discussed in some

detail in Chapter 4. When a potential is applied between the cathode and anode cylinder, the latter being positive, electrons from any point on the cathode will be converged onto a point on the image side of the system. A target located at the locus of these points will receive an inverted electron image of the light pattern falling on the cathode.

In order to permit electrical focusing and to reduce aberrations, a somewhat more complicated lens structure is used in practice, just as is done in the image tube. A series of short coaxial cylinders, maintained at potentials which can be varied, but are fairly close to the cathode voltage, are located between the cathode and the main lens. These form adjustable correcting and focusing lenses. In order to reduce further the two main lens defects, namely, curvature of the image field and pincushion distortion, it is necessary to curve the photocathode. Because of the optical difficulties introduced by the curved cathode and the requirement of a high voltage to obtain good image quality, electrostatic lenses have not been widely used in image multiplier pickup tubes.

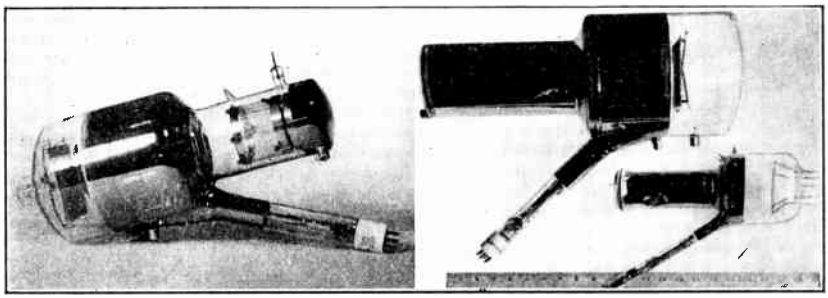
A uniform magnetic field is the simplest form of magnetic imaging system. Such a system produces an erect, distortion-free image with unit magnification whose principal aberration is due to the initial velocities of the photoelectrons. There is a continuous transition between this type of system and the short-coil lens. The short-coil lens is formed by a magnetic field which occupies only a small fraction of the space between the object and image planes, and produces an inverted image whose magnification is the ratio of the lens-image distance to the object-lens distance.

In practice the lens used is intermediate to the uniform field system and the short lens. Usually, the magnetic field is fairly uniform in the vicinity of the cathode and then decreases towards the target. The image from this electron optical system will be magnified and rotated by 30 to 40 degrees. The image plane is flat for a flat cathode, and the principal distortion is rotational, so that a straight line through the center of the object plane is imaged as an S-shaped line. This distortion can, however, be almost completely overcome by properly shaping the magnetic field to compensate the nonuniformity of the electrostatic accelerating field.

The electronic magnification of a magnetic image iconoscope may be of the order of 4. This magnification permits the use of conveniently small short-focus optics even when the image on the target is of the same size as in the iconoscope. In one recent version * of the image

* See Francken and Bruining, reference 6.

iconoscope the magnetic field coils are encased in a soft-iron shield and subdivided into three sections. Simultaneous variation of the three coil currents in a prescribed relationship permits the continuous variation of the picture magnification by a factor of 2.5, maintaining sharp focus and freedom from distortion throughout. The employment of a fine-mesh high-transparency anode screen spaced 0.4 inch from the photocathode both simplifies the focusing problem and prevents the concentration of ions on the target which may lead to its relatively rapid deterioration.



a *b*

Fig. 10.8. Electrostatic and Magnetic Image Iconoscopes.

The image multiplier principle has been used with a variety of types of targets. It was first employed with a more or less normal iconoscope mosaic. Figure 10.8 illustrates electrostatic (*a*) and magnetic (*b*) image iconoscopes. The higher photosensitivity and better efficiency of the cathodes of these tubes, together with a small secondary-emission gain at the target, give these tubes an improvement of four to six times over the normal iconoscope. Of course, the same problems of spurious signal exist in the image iconoscope.

Although this tube can give very satisfactory performance, it has not, until recently, been used in commercial television in this country, because of the development and adoption of the orthicon. It has, however, been used extensively abroad. Such tubes as the Super Emitron, the Eriscope, and the Multicon come under this tube class.

A pickup tube utilizing low velocity beam scanning lends itself very well to image multiplication. The image orthicon, which is now very widely used in commercial television, combines orthicon scanning with image multiplication. A discussion of this tube will be deferred until after the signal multiplication principle has been described in section 10.6.

10.5 Multi-Stage Image-Multiplier Pickup Tubes. Up to this point the discussion has been limited to image multiplier pickup devices which have only a single stage of image intensification. In principle, it is possible to focus the electron image from a photocathode onto a secondary emitting target, then to refocus the secondary electron image onto a second dynode surface, and so on until the image is intensified to the desired level, whereupon it is imaged onto a pickup tube target. In fact, operating pickup tubes involving more than one secondary emissive target electrode have been built. Practical problems of maintaining high definition in view of the high initial velocity of secondary electrons, however, constitute a serious stumbling block.

An alternative way to obtain additional intensification is to form the first electron image on a fluorescent screen which is closely coupled optically to a second photocathode. The electron image from this may be focused onto a target or onto a second fluorescent-screen-photocathode combination. This form of intensification will yield a current gain of ten or more per stage. Practical tubes with extremely high sensitivity have been built and used employing this principle. It might be noted in passing that this method has been used successfully in pickup tubes for the intensification of x-ray images.

In terms of the figure of merit for the tube, the effect of image multiplication is to increase the utilization factor ε. At the same time, since the signal from the target is correspondingly high, the noise factor is reduced relative to unity. In the limit, the figure of merit is not, however, identically $M = \gamma$ because the multiplication factor at each stage of image multiplication shows some random variation and introduces additional noise. It can be readily proved that the limiting form of M becomes approximately

$$M = \gamma \sqrt{\frac{\sigma - 1}{\sigma}}$$

where σ is the gain of each stage of a large number of stages. For example, if σ is 10, the sensitivity is 95 percent of the theoretical maximum for the given quantum efficiency γ of the initial photocathode.

THE SUPPRESSION OF NOISE

The third general method of improving pickup tubes is to reduce all nonfundamental noise.

In all the pickup tubes described in the preceding sections, the video signal was taken directly from a capacity-coupled signal plate asso-

ciated with the target. This signal is then amplified by a wide-band video amplifier. The constant noise factor appearing in the expression giving the figure of merit of the device is primarily due to the input noise (coupling resistor and first tube) of the video amplifier.

10.6 Signal Multiplication. A secondary-emission multiplier offers the possibility of obtaining the initial signal amplification almost without the introduction of noise. It has been pointed out earlier that there is a signal component of the electron current leaving the target which exactly equals the video signal in the signal lead. In the case of the normal iconoscope, this current is made up of secondary electrons from the target, whereas in orthicon-type tubes this signal is carried by the returned beam electrons. By a suitable electrode arrangement, it is possible to direct these electrons into a secondary-emission multiplier incorporated in the tube. In this way the video signal can be raised to a level well above amplifier noise.

The problem of directing the secondary electrons from an iconoscope mosaic is very difficult because the electric fields in the vicinity of the target must be small by the principle of its operation. Experimental tubes have, however, been built with secondary-emission signal multipliers. If the mosaic capacity is made small, very marked increase in low light sensitivity has been obtained, but all the tubes investigated have shown a considerable degree of shading due to nonuniform electron collection by the secondary-emission multipliers.

With an orthicon-type target, the problem is much simpler. Where magnetic focusing and deflection are employed the electrons from the target return along a path which is almost the same as that of the beam electrons. Usually there is a small amount of electrostatic lens action in such a tube so that the returning electrons have a small amount of scanning motion when they reach the gun. Also, the field can be adjusted so that the return beam is not sharply in focus at the gun. A secondary emitting disk at the final aperture of the electron gun can serve as the first dynode for the signal multiplier. Because of the scanning action of the returning electrons, the secondary-emission ratio of the surface of the disk must be very uniform if shading in the reproduced picture is to be avoided. There must be a small hole at the center of the disk to allow passage of the beam electrons, but this can be made very small so that it does not contribute a serious spurious signal. Secondary electrons from the disk are then drawn into the remaining dynodes which surround the gun. The arrangement of this type of multiplier is shown in Fig. 10.9.

The effect of the signal multiplier on the ultimate sensitivity of the pickup tube, or upon its figure of merit, can be estimated as follows. A moment's consideration will make it clear that when the target is in darkness, or from black areas in the image, a current equal to the beam current must enter the multiplier. This current will necessarily

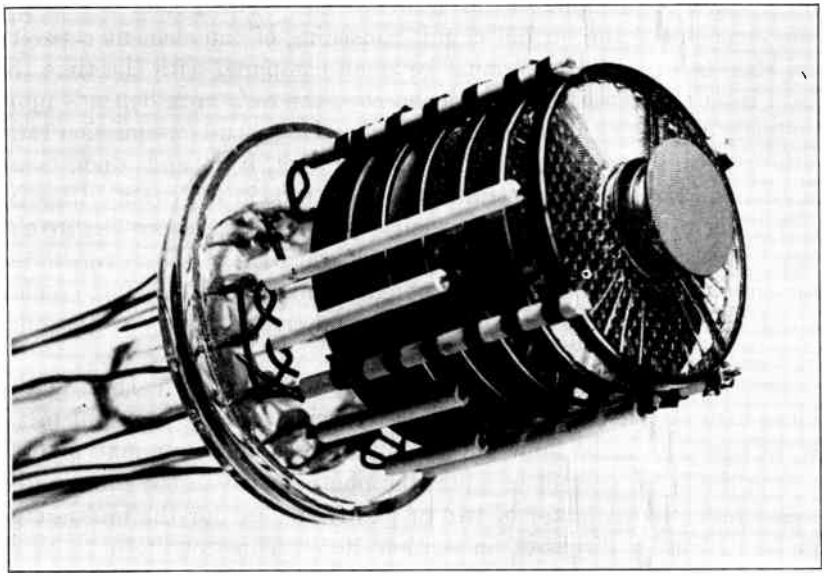


Fig. 10.9. Signal Multiplier for an Image Orthicon Pickup Tube.

generate noise at the output, which is equal to an uncertainty in the number of unit charges of

$$\overline{\Delta N}_{\text{rms}} = \sqrt{\bar{N}_B} \sqrt{\frac{\sigma}{\sigma - 1}} \quad \text{electrons}$$

where σ is the stage gain of the multiplier and N_B is the number of electrons entering the multiplier from the target area considered in a frame time. If the beam current is i_B and an area h^2 is considered in a target of area A, then

$$N_B = \frac{i_B}{1.6 \times 10^{-19}} \frac{h^2}{A} T$$

where T is, as before, storage or frame time. In like manner, if i_p is the photocurrent from an illuminated target, the number of photoelectrons from the area h^2 is

$$N_p = \frac{i_p}{1.6 \times 10^{-19}} \frac{h^2}{A} T'$$

It is evident that, if the beam is to be able to restore the target equilibrium, the minimum value that the beam can take is such that $N_B \geq N_p$; in other words, $i_B \geq i_p$. This is not the only consideration which sets a lower limit to the beam current. The second is that the time constant of the virtual circuit consisting of the element capacity and the "beam resistance" must be small compared with the time the beam is on an element. In the iconoscope the resistance depends upon the beam current and the variation of effective secondary-emission ratio with element potential. For an orthicon target, it depends upon beam current and the velocity distribution of electrons in the beam. Clearly, with either type of target, the capacity (per unit area or per picture element) must be small to achieve greatest sensitivity. However, if the capacity is too small, the element potential may saturate before enough charge can accumulate on it to give a good signal-to-noise ratio under favorable conditions of illumination.

Because of the many degrees of freedom in designing a signal multiplier pickup device, it is not possible to assign a single figure of merit to the device. In the limit the minimum detectable illumination (adjusting the beam current to equal the photocurrent in the illuminated areas) differs by a factor of two or less from that for the multi-stage image multiplier discussed in section 10.5. This condition is never actually realized in practice. More realistically, we should consider the figure of merit as given by Eq. 10.7*b* with ε determined by the target and the noise determined by the minimum practical beam current. The noise introduced is lower by two orders of magnitude than that introduced by a conventional video amplifier.

10.7 Image Orthicon. The image orthicon combines the principles of image multiplication and signal multiplication. In sensitivity, it is superior to any other pickup tube used for commercial television, being about one hundred times more sensitive than a normal iconoscope.

When this tube was initially introduced, it was primarily used only for pickups where the lighting might be unfavorable. As the tube was further developed, it became more widely used. Today it is by far the most widely used camera tube in this country for all types of pickup (except that from moving-picture film), including studio, sports, outdoor and indoor spot programs.

A schematic drawing of the construction of an image orthicon is given in Fig. 10.10. It will be noted that it combines the features of

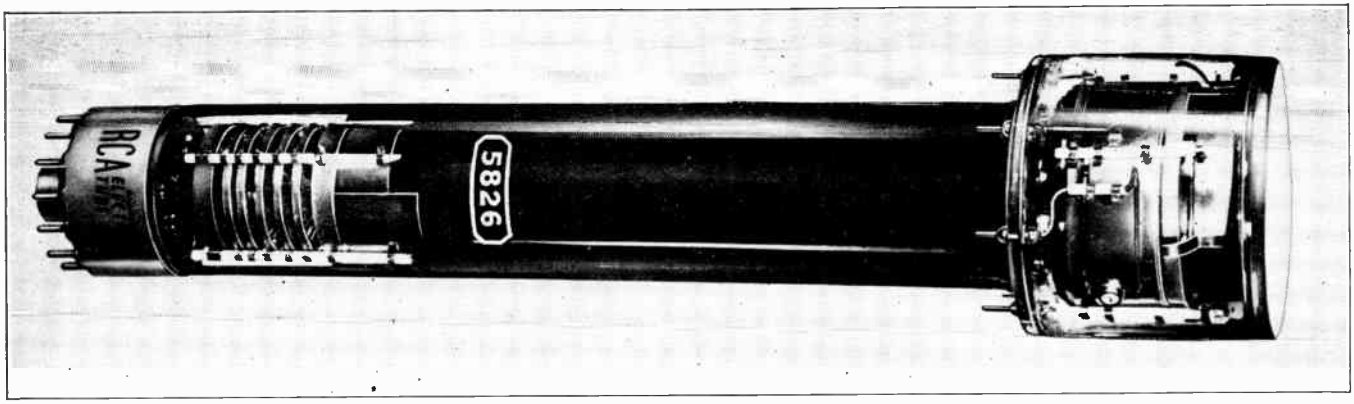


Fig. 10.10. The Image Orthicon.

image-multiplication, a two-sided target, orthicon operation, and signal multiplication.

In the figure the photocathode is on the extreme right. It takes the form of a sensitized semitransparent film on the inner end of the blank, so that when light is projected through the front end of the tube, electrons are released toward the target structure. The photoelectrons are focused by the magnetic field of the long focusing coil and form an electron image on the target. The secondary electrons produced are drawn away by the field produced by a fine mesh screen located very close to the target. Because of the close spacing between the target and screen, field enough to saturate the target secondary emission is obtained when the screen is only one volt positive with respect to the target. As the secondary electrons are drawn away the corresponding picture elements of the target become positive. However, they cannot become very much more positive than the screen, no matter how large the current density of the electron image. This means that the target remains stable under conditions of intense illumination. Thus, one of the serious objections to the orthicon type of operation is overcome.

The target, as has already been noted, is two-sided. It does not, however, employ the plug and mesh structure discussed in section 10.2. Instead, the target takes the form of an extremely thin film of high-conductivity glass. The thickness and conductivity are so chosen that the resistance through the glass is small compared with the transverse resistance over areas comparable in size with a picture element (i.e., spot size), and that charge equalization between the two sides takes place in a time of the order of a frame period. For this purpose, the glass has a specific resistivity between 3.10^{11} and 10^{12} ohm-centimeters at room temperature and the film thickness is of the order of a ten-thousandth of an inch.

The target and screen are assembled as a unit. A flat thin sheet of glass, about 1½ inches in diameter, is sealed to a metal ring on which is mounted a flat fine-mesh screen, with a mutual separation of about two thousandths of an inch. The screen is formed by a sputtering and electro-deposition technique from a master. It has 500 to 1000 meshes to the linear inch, with a transparency of 50 to 75 percent. The geometrical uniformity of this screen approaches that of a ruled grating, and it is very clean chemically. The technical problem of producing a target structure which has the required uniformity and will not break under the thermal processing required by the exhaust schedule is a major one, but one which has been successfully answered not only in the laboratory but also for manufacture.

The leaving secondary electrons cause the front side of the target to go positive. Since the target is isolated, the beam side assumes essentially the same potential as the front. When the low velocity beam sweeps over the surface, it brings each area on the beam side to cathode (zero) potential. The front of the target follows this potential to within about 0.01 volt, while the beam is on an element, and the charge, front and back, equalizes in the course of a few frame times.

Because the potential range of each element is only two volts (i.e., from cathode (0) to screen (+2) potential), it is essential that the beam strike the target normal to its surface. The beam in the image orthicon is magnetically focused and deflected. Upon leaving the cathode, the beam electrons are accelerated to 220 volts, and then slowed down to zero volts in the vicinity of the target. Unless great care is taken in the design of this electron optical system, the electrons will assume a helical motion owing to their having components of their velocity not parallel with the magnetic field lines. By careful design variations in axial kinetic energy resulting from this effect can be reduced to less than 0.1 electron volt.

Electrons which are returned from the target enter a five-stage secondary-emission multiplier. The multiplier is designed along the lines indicated in the preceding section. The first dynode is an apertured disk at the end of the gun. The remaining dynodes are disks with radial "pinwheel"-like slats in the form of annular rings around, and just back of, the first dynode. The first dynode is operated at 300 volts positive with respect to the target. About 300 volts are applied between the second and first dynode. The electrodes around and in front of the first and second dynode combination are all somewhat negative with respect to the first dynode. Therefore, nearly all the electrons from the first reach the second dynode. Secondary electrons from dynode number two are drawn through it and into the third dynode. A coarse screen between the second and third dynode, and maintained at third dynode potential, prevents the field of dynode two from suppressing the emission from dynode three. The remaining two stages are similar to the screen-pinwheel combination of dynode three. The gain of this multiplier is about 500, more than sufficient to give all the improvement in signal-to-noise that can be achieved with a signal multiplier.

10.8 Performance of the Image Orthicon. The sensitivity of the image orthicon is at least one hundred times greater than that of a normal iconoscope. A factor of 5 of this is derived from the image multiplier section, partly because of the better photocathode and partly

through secondary-emission multiplication of the image. The remaining factor of 20 results from using a signal multiplier. The useful gain from the multiplier can be estimated as follows. The equivalent noise current of a good video preamplifier with a 5-megacycle bandwidth is 2×10^{-9} ampere. The basic beam noise is

$$I_{\text{rms}} = (2e\Delta f I_b)^{1/2}$$

In normal operation the beam current cannot be much less than 0.01 microampere if reasonably good signal-to-noise ratio is to be maintained. The useful multiplier gain, and consequent improvement, is therefore about 20. At lower light levels, with smaller beam currents and a somewhat poorer signal-to-noise ratio, an even greater sensitivity advantage is to be had.

The sensitivity of a camera using an image orthicon approaches that of the eye down to scene brightness levels of a few hundredths of a footlambert (i.e., full moonlight). It exceeds that of a camera with equivalent optics, using Super XX film.

Representative curves of the signal as function of photocathode illumination are illustrated in Fig. 10.11. By beginning with low light levels it will be noted that the signal rises linearly with illumination until intermediate light levels are reached. At this level the target elements charge up to the collector screen potential. The curve is almost flat with further increase in target illumination. Beyond this there may be another small rise if the target-to-screen spacing is wide and then another leveling off.

The initial linear range corresponds to perfectly normal orthicon behavior.

The flat portion of the range corresponds to signal saturation for large areas. However, secondary electrons are returned to the target when the target elements charge to the screen potential. The redistribution of electrons is here sufficient to permit the formation of a charge image on the target. This is exactly the same type of redistribution photoelectric image which has been described in connection with the normal iconoscope. (See section 9.7.) Because of the small spacing between the screen and the target, this redistribution image is a small-area phenomenon. It might be pointed out that the redistribution phenomenon is for the most part very useful, since it enhances contrast under circumstances when the target would otherwise be saturated. However, the effect can show up as a spurious signal under certain circumstances, namely, if a very bright highlight is superimposed on a low-level background. Around the bright area there

will be a dark halo due to electrons redistributed into the region around the strongly illuminated area. Figure 10.12 compares an intermediate level highlight (*a*) and a bright highlight (*b*). The dark halo is clearly in evidence in the second reproduction. In addition, extremely bright portions of the picture may be accompanied by "ghosts" and white halos, which result from the return to the target of back-scattered or reflected primary electrons. These effects are always objectionable.

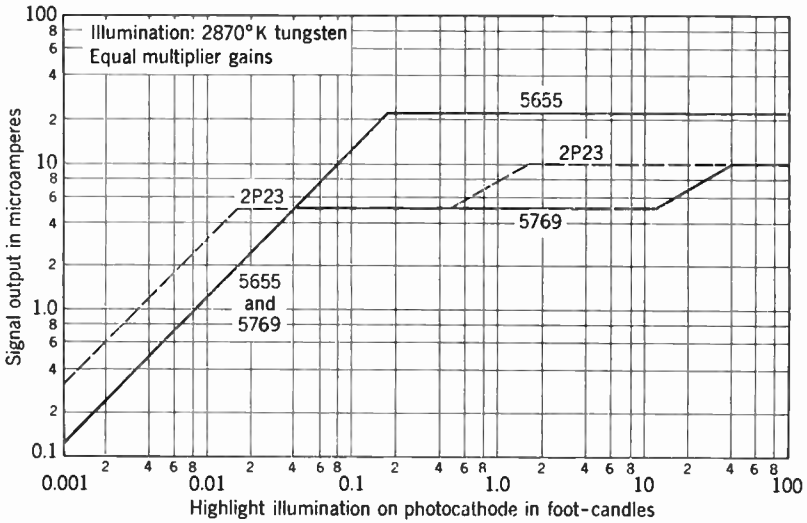


Fig. 10.11. Signal Response of the Image Orthicon. (Janes, Johnson, and Moore, reference 11.)

However, they are rarely in evidence; furthermore, ghosts can generally be brought to coincidence with the picture by careful design of the image section.

The spectral response of the photocathode, for various types of image orthicons, is shown in Fig. 10.13. The earliest commercial type, the 2P23, had a relatively wide-spaced target and employed a silver-cesium oxide-cesium photocathode; high-sensitivity tubes of this type exhibited a strong infrared response, and the response in the red was accentuated quite generally. In the (narrow-spaced) 5655 and the (wide-spaced) 5769 an antimony-cesium film replaced the cesium oxide–cesium layer and the peak of the response was shifted to the blue end of the spectrum. In the 5820 (and the 5826 and 1854, with narrower target spacing), finally, considerably higher absolute sensitivity coupled with a response similar to that of the human eye was

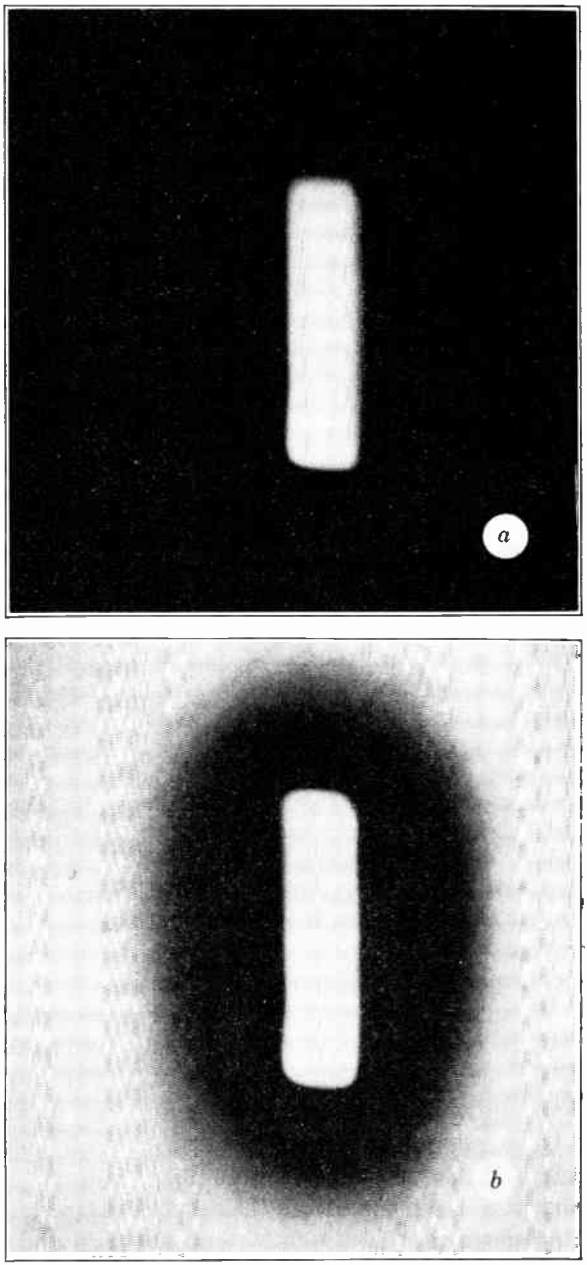


Fig. 10.12. Picture of Light Spot at Low (*a*) and High (*b*) Spot Brightness Transmitted by Image Orthicon.

achieved with an oxygen-treated silver-bismuth-cesium photocathode. The tubes with wide-spaced targets may be operated with smaller beam currents and hence yield a better signal-to-noise ratio at very low light levels. They are consequently preferred for outdoor pickup. On the other hand, more charge can be stored on the target with close target spacing, leading to higher signal-to-noise ratios at high light

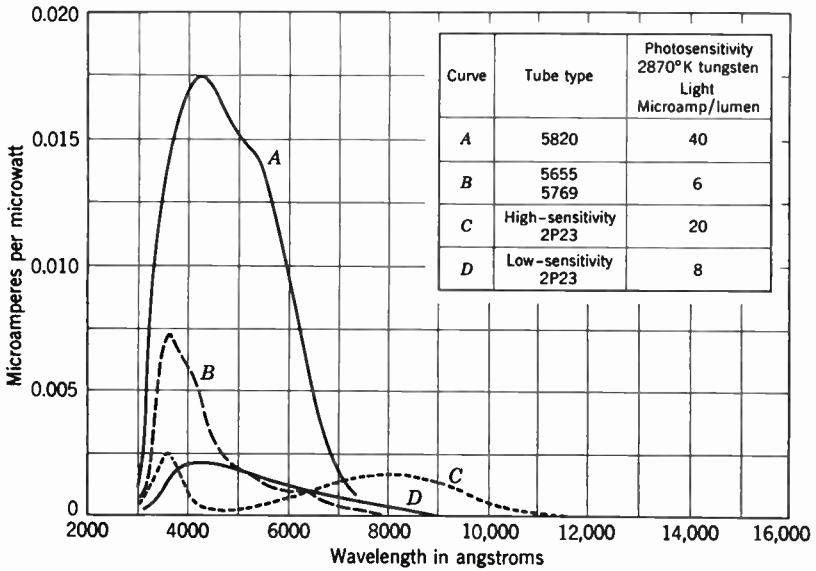


Fig. 10.13. Spectral Response of Various Commercial Image Orthicons. (Janes, Johnson, and Handel, reference 12.)

levels. Hence, tubes with close spacing are employed mainly for studio service, where adequate light levels can be provided without difficulty.

An examination of the properties of the image orthicon indicates that it can meet nearly all the requirements placed on a pickup tube by commercial television. The resolution is ample for all practical purposes. It shows no lag or trailing. The contrast range is very large, larger in fact than can be covered by other parts of the system, so that it does not limit overall performance in any way. Considering all of its features, the image orthicon constitutes a highly satisfactory all-purpose camera tube for modern black-and-white television.

10.9 The Isocon. An image orthicon with a signal multiplier cannot, in principle, reach the performance of an ideal pickup device at very low light levels. This is because the full beam current returns to

the multiplier at dark portions of the target. Thus, noise is a maximum for low light areas and a minimum for illuminated elements.

Maximum sensitivity can be realized only if either (*a*) there is sufficient image intensification so that an element subjected to a single photoelectric event causes enough charge to be stored in the target so that the signal from it overrides beam or amplifier noise, or (*b*) the target is arranged so that the returned electrons are a minimum for dark areas and a maximum for those illuminated. A multi-stage image intensifier may achieve the first. The image Isocon is one way of accomplishing the second.

The formation of the charge image on the target is the same in the Isocon as in the image orthicon. A low-velocity electron beam from the gun approaches or strikes the target as with an orthicon, but the method of handling the returned electrons is different.

With the target in darkness the electron-optical focusing system is so designed that the returning electrons focus onto a sharp spot. At the point of this focus a sharp edge is placed in such a way that it intercepts the returning electrons, preventing them from reaching the multiplier. As the target elements become more positive, the approach and return of electrons near the target are altered. This causes the electron beam to be somewhat out of focus at the barrier so that some electrons get past and into the secondary-emission multiplier. Over the working range, the more positive an element becomes the greater the defocusing and the larger the current reaching the multiplier.

Tubes of this type whose performance, when correctly adjusted, meets the requirements of television have been built both on a laboratory and on a developmental scale. They are, however, found to be very critical in adjustments.

At very low light levels there is a marked difference in the performance of the Isocon as compared with the image orthicon. The absence of noise in the darker picture areas gives a marked improvement in low-light detail reproduction. However, under conditions of operation where the Isocon is markedly superior to the image orthicon, the signal-to-noise ratio is fundamentally so low that the picture is undesirable for entertainment purposes. It may have application, however, when pictures are wanted under very difficult conditions.

10.10 Photoconductive Pickup Tubes. So far, all the pickup tubes described depend for their action upon photoemission. It is, however, also possible to make practical pickup tubes employing photoconductors to obtain an electrical signal from an optical image.

There are a number of ways in which photoconductors may be employed for the purpose. Two, in particular, will be discussed in this section. One of these uses a material which has a relatively low resistance and gets its storage through employing a photoconductor with a time constant comparable to frame time. The second uses a high-resistance long-time-constant photoconductor and stores charge between the photoconductor surface and a back conducting plate as does a photoemissive target.

The mechanism of signal generation of the first form can best be explained with the aid of Fig. 10.14. This figure is shown for operation with a beam whose secondary-emission ratio is greater than one, as for a normal iconoscope. Orthicon-type of operation can equally well be used. The current voltage diagram would be much the same as that shown.

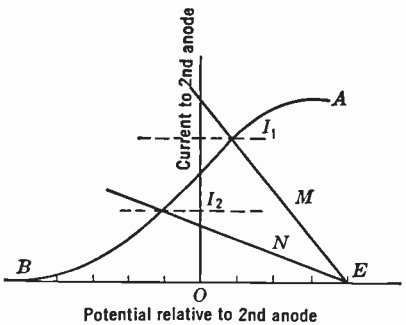


Fig. 10.14. Current Characteristics of a Low-Resistance Photoconductor Target.

The target is a conducting plate (which may or may not be transparent) coated on the beam side with a thin layer of photoconductive material. When a light image is formed on the photoconductor, areas which are illuminated become lower-resistant than unilluminated areas. The conducting back plate is made slightly negative with respect to the secondary electron collector, but not sufficiently so to saturate the secondary emission. Under these conditions, curve AB represents the current-voltage curve of the secondary-emission current to the target, the abscissa being with reference to the surface of the target. Line N is the current-voltage curve of an unilluminated area of the target, and point E the bias voltage of the target back conducting plate. The intersection of N and AB gives the current I_2 flowing to the electron collector from a dark area when the beam strikes it. The curve M is the current voltage curve of the illuminated area. The intersection now corresponds to a higher current I_1. The change of collector current with illumination gives rise to the video signal. The cumulative nature of the photoconductive effect in certain photoconductors gives the storage needed for efficient operation. This type of storage in the gain mechanisms of a photoconductor has been discussed briefly in Chapter 1.

The second type of photoconductive pickup tube (Fig. 7.21) more closely resembles the orthicon (or iconoscope) in its operation, except that charge is transported away from the target surface by photoconductivity instead of photoemission. Like the first-mentioned photoconductive pickup tube, the photoconductive material is formed as a thin layer on a conducting semitransparent plate. In this instance, material with a very high specific resistivity in darkness is employed. Normally, where the target is one or two microns thick the resistance back to front should be at least 10^7 or 10^8 ohms per square centimeter. For orthicon-type operation, the conducting plate is made slightly positive with respect to the gun cathode. In darkness the scanning beam drives the surface to cathode potential. As soon as the beam leaves an element of area, the dark current flowing from the conducting plate to the surface charges it slightly. However, owing to the high dark resistance, the amount of charge accumulated during a frame period is small. Where illumination falls upon an element the photocurrent flowing from the conducting plate to the surface is greater, so that the charge accumulated is greater. The amount of current taken from the beam (i.e., not returned to the collector or multiplier) is greater, therefore, for the illuminated area than for the dark regions. Since the charge accumulated is proportional to the photoconductive current, which is, in turn, proportional to the illumination,* the video signal has a one-to-one correspondence with the light image on the target.

Many materials may be used for photoconducting targets of the second type. Amorphous selenium, Sb_2S_3, Cd_2S_3, Cd_2Se_3, and composite layers of these materials are examples. The factors which govern the selection are photoconductive efficiency, spectral response, time lag, and stability, assuming that the dark resistance of the material is suitable. It might be noted that many of these materials require a rather elaborate activation treatment.

The desirable features of photoconductive pickup tubes are many. The primary quantum efficiency of some of the photoconductors is very high and can approach unity. It is possible to build into the material a gain mechanism which gives a multiplication of 10 or 100 or even 1000 times. This plays the same role as image multiplication. The targets take full advantage of the storage principle. Operated as an orthicon target, they lend themselves to the use of a secondary-emission multiplier. Finally, the construction of a photoconductive pickup tube is very simple and rugged.

* The relation $i_p = kL$ is not true for all photoconductive materials. Frequently $i_p = kL^{1/2}$. Other relationships are also possible.

In practical photoconductive pickup tubes it has been possible to achieve sensitivities approaching that of the image orthicon. Such tubes perform satisfactorily in a wide range of applications. The Vidicon (Fig. 10.15) is a photoconductive pickup tube which approaches the ultimate in simplicity. It consists only of a photoconductive target on the glass end of the tube and an electron gun at the other. It is only 1 inch in diameter and 6¼ inches long. In spite of its small size and simplicity, it has sufficient sensitivity to give an excellent picture in a normally lighted room. For a small compact unit, such as is needed for industrial television, the tube is almost ideal. A slight lag at low levels of illumination has been the major obstacle to its use as a general-purpose tube for broadcasting.

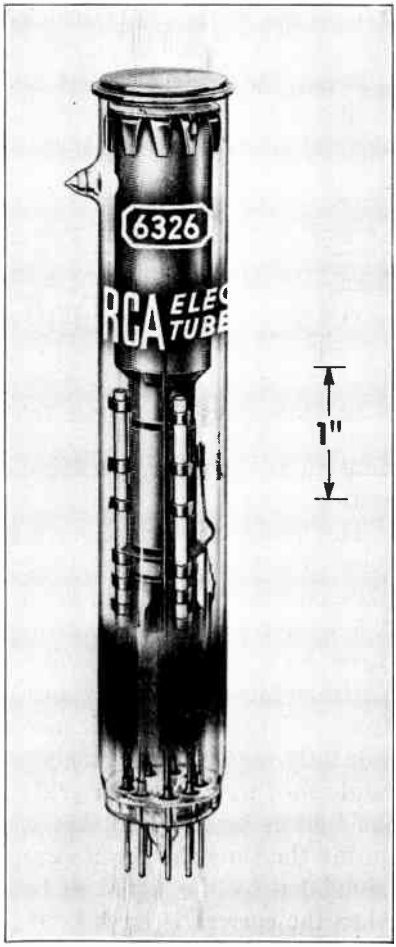


Fig. 10.15. Vidicon Pickup Tube.

10.11 The Storage Tube. Although the storage tube is not, strictly speaking, a television pickup tube, nevertheless, it employs many of the same principles and, therefore, is included in this chapter.

The storage tube consists of a means, usually in the form of a scanning beam, of recording upon a storage target the charge image of a video signal, the storage target itself, and a means for reading the stored information when it is required. Figure 10.16 illustrates the most elementary type of storage tube as an example of this class of device. It consists of a two-sided target which is scanned on either side by separate guns. The signal to be stored is applied to the secondary electron collector of one of the two guns. As the beam sweeps across the target the elements will assume a potential which is de-

pendent upon the instantaneous potential of the electron collector. When the scanning raster is complete, a potential pattern will exist over the target which is a record of the voltage variations applied to the collector. When scanned from the other side by the second gun, the charge pattern is transformed into a video signal which can be collected by a secondary-emission collector electrode.

It is quite evident from the discussions given in this and the preceding chapter that redistribution and other losses occurring on both sides of a target of this type would make a storage tube designed strictly along these lines very ineffective. In practice, a somewhat more sophisticated approach must be taken. A variety of storage tubes have been developed which overcome redistribution losses and other sources of inefficiency and have television capabilities. Space limitations, however, permit mention of only two of these in this section.

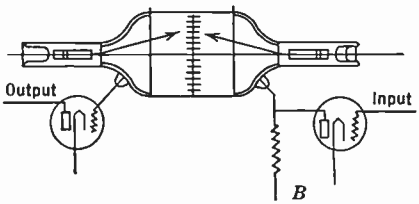


Fig. 10.16. Simplified Storage Tube.

The first is the Radechon, which is a one-gun barrier grid storage tube. The target of this tube consists of a metalized conducting signal plate covered with a thin layer of highly insulating aluminum oxide. A fine mesh screen is mounted very close to the insulating surface and serves as a barrier grid, which prevents redistribution of secondary electrons formed by the beam. As the beam sweeps across the target, areas under the beam emit more electrons than they receive, and consequently become increasingly positive until they reach an equilibrium established by the barrier grid and collector electrode. In normal use the tube is operated so that elements do not reach this equilibrium during the time the beam sweeps across them. The beam current is modulated by the signal to be recorded. Elements under the beam when the current is large become more positive than those under the beam when it is small. To read the recorded signal the back plate is made slightly positive with respect to the barrier grid. Capacity drives the front surface of the target positive but retains the charge image. The target is then scanned with a constant current beam which is intense enough to bring it to equilibrium. The difference in charge on the different areas causes a variation in the current returning from the target. This variable current is the read video signal.

The second type of storage tube which will be mentioned is the Metrechon (Fig. 10.17). This tube employs a two-sided target and two guns. The target consists of a thin film of slightly conducting glass (similar to that used for the image orthicon target) fused to a fine metal mesh. The target is mounted so that the metal mesh is exposed to the reading gun. A second fine metal mesh is mounted close to the target on the writing gun side. The writing is done with a high-velocity beam very similar to that used in the Radechon. The potential of the screen on the writing side which determines the equilibrium of the

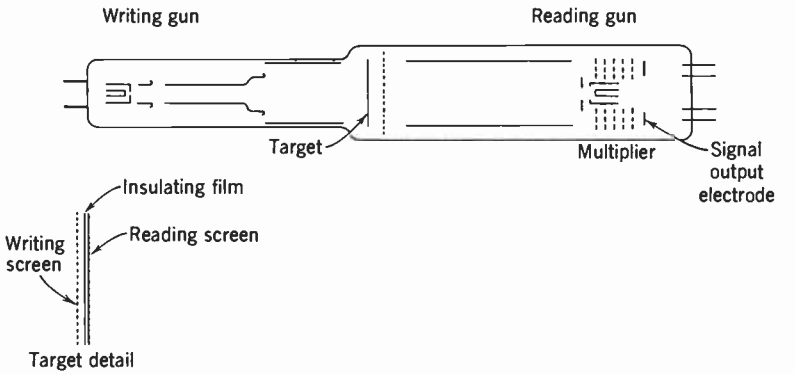


Fig. 10.17. Metrechon Storage Tube.

target elements under the writing beam is adjusted so that the beam drives the elements negative by 1 to 5 volts. The screen on which the target film is mounted is made very slightly positive with respect to the cathode of the reading gun. The reading beam, therefore, has the characteristics of the low-velocity beam of an orthicon. The target elements which were driven negative by the writing beam not only prevent electrons from reaching the elements themselves but also exert a grid action preventing some of the reading beam electrons from reaching the wire mesh. The fraction of beam electrons which actually reach this mesh depends upon the extent to which the elements in the immediate vicinity have been driven negative by the writing beam. Since the fraction of reading beam electrons absorbed by the mesh varies with the charge pattern on the target, the electrons returned from the target also vary with the charge image. These electrons are picked up by a secondary-emission multiplier just as are those from an image orthicon target, and the intensified current output from the multiplier constitutes the reading signal.

It will be observed that the operation of reading does not remove the charge image stored on the target. Once a pattern has been stored it can be read as many times as required without its being destroyed. When the writing beam sweeps over the target to record another pattern it removes the previously stored charge image.

In television storage tubes may be used in a variety of ways. Scanning pattern conversion, where the line number or frame frequency of the raster of one system differs from that of another, is one example. Again, if it is necessary to delay a sequence of pictures for any interval up to one or a few frame periods, storage tubes may be required. They may also find application for certain special types of picture combination or alteration.

10.12 The Monoscope. While the Monoscope is not a pickup tube, it shares with the camera tubes the property of being a picture signal

Fig. 10.18. The Monoscope.

generator tube. The Monoscope resembles the iconoscope in its makeup, but in place of a photosensitive target it has a target upon which is printed the image whose video signal it is desired to reproduce.

The metal surface of the target is of a material having a fairly high secondary-emission ratio. Aluminum with an oxide coating due to atmospheric oxidation is an example. The pattern to be reproduced is printed on this surface in carbon ink or other material having a different secondary-emission ratio from the background.

When this pattern is scanned, the change in current collected, or current to the back plate, due to the varying secondary-emission ratio constitutes a video signal. A typical Monoscope is illustrated in Fig. 10.18.

10.13 Conclusion. This discussion of the various forms of pickup tubes is not by any means exhaustive. Many very interesting types which have been, or are being investigated, were of necessity omitted.

The ones included were intended to give at least an idea of the major considerations in this phase of the television problem. It is hoped, also, that some of the fundamental limits to the sensitivity which can be achieved were made clear.

Color pickup tubes are still in an early developmental stage. Some of the pertinent problems are touched on in Chapter 17.

REFERENCES

1. V. K. Zworykin, G. A. Morton, and L. E. Flory, "Theory and Performance of the Iconoscope," *Proc. I.R.E.*, Vol. 25, pp. 1071–92, 1937.
2. H. Iams and A. Rose, "Television Pick-up Tubes with Cathode Ray Beams," *Proc. I.R.E.*, Vol. 25, pp. 1048–70, 1937.
3. H. Iams, G. A. Morton, and V. K. Zworykin, "The Image Iconoscope," *Proc. I.R.E.*, Vol. 27, pp. 541–547, 1939.
4. J. D. McGee and H. G. Lubzynski, "EMI Cathode-Ray Television Tubes," *J.I.E.E.*, Vol. 48, pp. 468–475, 1937.
5. M. Berthillier, "The Eriscope," *Le Vide*, Vol. 2, pp. 355–359, 1947.
6. J. C. Francken and H. Bruining, "New Development in the Image Iconoscope," *Philips Tech. Rev.*, Vol. 14, pp. 327–335, 1953.
7. H. Smith, "Multicon—A New TV Camera Tube," *Tele-Tech*, Vol. 12, pp. 57, 125, July 1953.
8. A. Rose and H. Iams. "The Orthicon, a Television Pickup Tube," *RCA Rev.*, Vol. 4, pp. 186–199. 1939.
9. J. D. McGee, "Distant Electric Vision," *Proc. I.R.E.*, Vol. 38, pp. 596–608, 1950.
10. A. Rose, P. K. Weimer, and H. B. Law, "The Image Orthicon—A Sensitive Television Pickup Tube." *Proc. I.R.E.*, Vol. 34, pp. 424–432, 1946.
11. R. B. Janes, R. E. Johnson, and R. S. Moore, "Development and Performance of Television Camera Tubes," *RCA Rev.*, Vol. 10, pp. 191–223, 1949.

12. R. B. Janes, R. E. Johnson, and R. R. Handel, "A New Image Orthicon," *RCA Rev.*, Vol. 10, pp. 586–592, 1949.
13. R. B. Janes and A. A. Rotow, "Light-Transfer Characteristics of Image Orthicons," *RCA Rev.*, Vol. 11, pp. 364–376, 1950.
14. A. Rose, "Television Pickup Tubes and the Problem of Vision," *Advances in Electronics* (L. Marton, Ed.), Vol. 1, pp. 131–166, Academic Press, N. Y., 1948.
15. P. K. Weimer, "The Image Isocon—An Experimental Television Pickup Tube Based on the Scattering of Low Velocity Electrons," *RCA Rev.*, Vol. 10, pp. 366–386, 1949.
16. M. Knoll and F. Schröter, "Translation of Electron Pictures and Drawings with Insulating and Semi-Conduction Layers," *Physik. Z.*, Vol. 38, pp. 330–333, 1937.
17. M. von Ardenne, "On Experiments with Photosensitive Semi-conducting Layers in Cathode Ray Tubes," *Hochfrequenztechn. u. Elektroakustik,* Vol. 50, pp. 145–149, 1938.
18. P. K. Weimer, S. V. Forgue, and R. R. Goodrich, "The Vidicon Photoconductive Camera Tube," *Electronics,* Vol. 23, pp. 70–73, May, 1950, or *RCA Rev.*, Vol. 12, pp. 306–313, 1951.
19. B. H. Vine, R. B. Janes, and F. S. Veith, "Performance of the Vidicon, a Small Developmental Television Camera Tube," *RCA Rev.*, Vol. 13, pp. 3–10, 1952.
20. W. Veith, "Signal Formation in Picture Converters. The Conductron," *Le Vide,* Vol. 5, pp. 887–895, 1950.
21. C. E. Burnett, "The Monoscope," *RCA Rev.*, Vol. 2, pp. 414–420, 1938.
22. M. Knoll and B. Kazan, *Storage Tubes and Their Basic Principles,* Wiley, New York, 1952.

11 The Kinescope

The kinescope, as was indicated in Chapter 8, is one of the two terminal tubes of the television system. The picture, as a consequence of a long chain of physical events, is reproduced in light and dark * on its viewing screen. The position of the viewing tube as the final link in the television chain places upon it the responsibility of converting a mass of electrical phenomena into a visible image allowing accurate cognizance of the scene being transmitted.

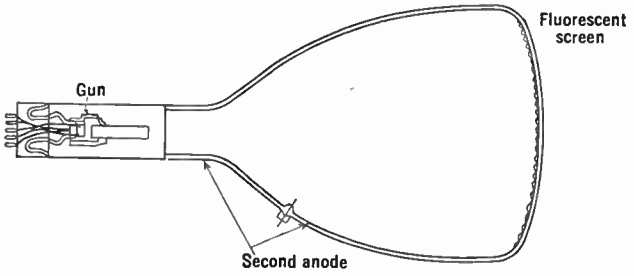


Fig. 11.1. Diagram of Kinescope.

Basically, the kinescope is quite simple, consisting of an electron gun and fluorescent screen assembled in a highly evacuated bulb. The electron gun, like that in the iconoscope, is an electron-optical system which concentrates the electrons from a thermionic cathode into a narrow pencil. It is located in a narrow neck which terminates the pear-shaped main body of the tube. Opposite the gun is the slightly curved broad end of the bulb, and on this surface is coated the fluorescent material or phosphor. The arrangement of the elements can be seen from Fig. 11.1.

The cathode-ray pencil from the gun strikes the screen in a small area called the "spot." Owing to the luminescent properties of the screen the spot is visible as a bright point of light. The current in

* The present chapter deals only with kinescopes for monochrome picture reproduction. Tricolor tubes will be treated in Chapters 17 and 19.

the electron beam and, therefore, the amount of light emitted at the spot can be varied by a potential applied to a control element in the gun.

Means are provided in the form of magnetic coils—or, in exceptional cases, electrostatic deflecting plates—for deflecting the spot to any position on the screen. The spot can, therefore, be made to sweep across the kinescope screen in the series of straight parallel lines required for the scanning pattern by means of suitably varying currents or potentials applied to these deflecting means. The entire scanning pattern is swept out at such a rate that, by virtue of the persistence of vision, the pattern is seen as static. That is, when examined closely the screen of the kinescope appears to be ruled with a grating of fine luminous lines, but at a distance of a few feet it appears uniformly luminous.

At the pickup end of the system an exploring spot sweeps out a similar pattern. Not only are the scanning patterns similar in number of lines and in aspect ratio, but also the spot at the receiver is so synchronized with that at the transmitter that the two spots are at every instant at the same relative positions on their respective patterns. Thus, if the brightness of the spot at the receiving end is controlled by the amplitude of the video signal, a reproduction of the image at the transmitter will be formed on the viewing screen. By supplying the video signal to the control grid of the kinescope gun this variation in spot brightness can be obtained and a luminous reproduction of the picture being televised will be formed on the fluorescent screen.

The generation of the video signal has already been discussed in Chapters 9 and 10. The amplification, transmission, and reception of the video signal, prior to applying it to the grid of the kinescope, will be taken up later. Also, the very important problem of synchronizing and deflecting the beam must be reserved for another chapter. The sole concern of the present chapter is, therefore, the requirements, construction, and performance of the tube on which the picture appears.

11.1 Requirements of the Kinescope. The general requirements of a high-definition picture have already been discussed without reference to any particular television system. When referring to a particular means for reproducing a picture, some of these requirements must be tempered because of practical considerations, and certain compromises must be made.

The size of the picture reproduced by a direct-viewing kinescope determines the size of the tube; this applies equally for tubes with round screens and with rectangular screens, though the bulk of the

latter is reduced appreciably by fitting the shape of the tube envelope to that of the picture. Early tests had indicated that, for general family use in the home, a picture between 1 and 2 feet on the diagonal would be desirable; it would seem that a picture larger than this could not easily be fitted into the average home decoration or environment. Experience has shown, in agreement with this, a strong customer demand for 21 and 24 inch tube sizes, with an apparent limit of acceptability at 27 inches; in any case, the bulk of tube production is concerned with screen diameters in excess of 16 inches or more. At the same time, smaller tubes continue to play a role in particularly low-priced and compact sets. On the other hand, very much larger pictures are required for public entertainment purposes. These are provided by the projection kinescope, which will be considered at the end of this chapter. Small projection kinescopes have, in addition, played a limited role in providing pictures of intermediate size, with diagonals 2½ to 5 feet, for the home.

Psychologically it would seem desirable to have the picture appear on a flat screen. However, where the end of the tube serves as the viewing screen, the very great reduction in thickness of glass that can be effected by slightly curving the tube face more than offsets the advantage of a flat screen.

One of the most important considerations in connection with picture quality is resolution. It was shown in Chapter 5 that a picture made up of 500 to 600 horizontal lines meets the basic requirements of high-definition television. Assuming video signals capable of giving this resolution, the factors necessary in the kinescope are a small scanning spot, accurate deflection, and fine-grain texture of fluorescent screen material. For a picture about 16 inches on the diagonal and with a 4 to 3 aspect ratio (i.e., 13 by 10 inches) the spot diameter must be less than 0.025 inch. The size of the particles of screen material should be such that at least four grains are under the beam at any one instant, and preferably many more. The deflection must be accurate to a fraction of the separation between lines.

The brightness of the screen is an important factor in the entertainment value of the picture, particularly if it is to be viewed over a long period of time. The brightness required is not a fixed quantity, but depends upon the average illumination of the surroundings and, to some extent, upon the content of the picture being viewed. If the receiver is in complete darkness, the light level required to avoid undue fatigue will be much lower than if it is in a moderately lighted room. This is so, of course, because the iris opening of the eye is determined

by the average illumination entering the pupil. Similarly, a black-and-white cartoon can be viewed without fatigue with less light than is required for a halftone picture. Generally, the brightness of a television screen must be sufficient so that an ordinary halftone reproduction can be viewed with comfort in a normally lighted room.

For orientation, Table 11.1 lists the brightness of a few familiar objects. From such data and various tests it has been concluded that

TABLE 11.1. BRIGHTNESS OF FAMILIAR OBJECTS

	Foot-Lamberts
Outdoor scene (bright day)	300 to 600
Moving-picture theater screen (high light)	2.7 to 5.2
Lighted page (minimum recommended)	10

the brightness of the screen must be 10 or more foot-lamberts; under special circumstances brightnesses of the order of 100 foot-lamberts have been found advantageous and have been achieved on the kinescope screen. The major factors governing the light which can be obtained from the screen are the fluorescent material itself, the current in the beam, and the accelerating voltage of the bombarding electrons. To a much smaller extent it is a function of spot size (due to saturation) and scanning frequency (due to phosphorescent decay time). Indirectly, the scanning frequency is of great importance since, with increase in light level, objectionable flicker effects persist up to higher field frequencies.

The color of the luminescence of the screen is, in a sense, related to the brightness in that the eye has a varying response over the visible spectrum. Green, or yellow-green, not only produces the maximum sensation for a given amount of radiation, but also is least fatiguing to watch. These reasons make it a desirable screen color. However, tests over a long period of time have revealed that these advantages are more than offset by the adverse psychological effect of any color other than white. Therefore, white or nearly white screens have come into universal use.

The scanning process, with reference to any small area of the screen, is essentially discontinuous and, accordingly, introduces the possibility of flicker. This flicker, when pronounced, is extremely objectionable. The magnitude of the sensation of flicker depends on vertical sweep frequency, duration of phosphorescence, and brightness of the screen. Irrespective of the second two factors, flicker can be made imperceptible if the vertical sweep frequency is sufficiently great. Increase

in the time of phosphorescent decay decreases the threshold frequency; increased brightness has the opposite effect, as has heretofore been pointed out. With the fluorescent materials available and the brightness ordinarily used, the flicker threshold occurs at a much higher frequency than that necessary for continuity of motion. In order to avoid the impression of flicker without necessitating a corresponding increase in bandwidth beyond that required by continuity, interlaced scanning is used.

For the mere conveyance of information, contrast is relatively unimportant. However, it is a factor of major importance in determining the quality of a picture. The meaning of the term contrast and its relation to noise, brightness, and visual threshold have been discussed in Chapter 5. Over most of the range of normal vision the visual response is approximately a logarithmic function of the brightness B of the object viewed. Therefore, to produce equal increments of sensation, the steps of brightness ΔB must be related by $\Delta B/B = \text{const}$. As has previously been pointed out, the relation ceases to be valid at low light levels and, over the useful range, has a threshold value of about 0.02. The darkest portion of the viewing screen of a kinescope reproducing a normal picture is not "black" but has a finite brightness B_0. The available visual contrast range for the kinescope will, therefore, be

$$\ln \frac{B_m}{B_0}$$

where B_m is the maximum brightness of the screen. The ratio B_m/B_0 is often called the brightness contrast range, as distinguished from the visual range. The contrast range is thus limited by the minimum illumination of the screen, as well as the maximum brightness that can be reproduced. In order to indicate the brightness contrast range that would be desirable for a television screen, Table 11.2 gives the contrast range encountered in familiar scenes. It will be apparent from

TABLE 11.2. CONTRAST RANGE

Clear sunlight and shadow	100 and up
Normal landscape	30 to 50
Interior with normal artificial illumination	20 to 50
Moving-picture screen	50 to 100
Good photograph	15 to 40

the wide variation of contrast range listed that the human eye is capable of adapting itself to an extreme variety of conditions. A

television picture with a range of 15 to 20 would be acceptable as a good picture, although considerable improvement can be obtained if the range is increased to 100. Increasing the contrast range above 100 results in further improvement, but the improvement is relatively small.

The six requirements listed, referring to size, resolution, brightness, color, flicker, and contrast, are essential and must be met by a viewing tube which is to reproduce a picture of high entertainment value. Besides these essentials, there are other practical considerations. The tube must be such that it can be manufactured economically. Its physical size (as distinguished from screen size) must be reasonable. The voltage and current needed for its operation must be such that they can be cheaply and safely supplied. Finally, the kinescope should be durable, both electrically and mechanically.

Formidable though this list of requirements may appear, present-day types of kinescopes meet the conditions named in a satisfactory manner.

11.2 Construction of the Kinescope Bulb—Round Glass Tubes. The design of the bulb itself is a major problem in kinescopes having large viewing surfaces. This is immediately apparent from an estimate of the force exerted by atmospheric pressure on the face of the tube. For example, on the screen alone of a 16-inch kinescope there is a force of 3000 pounds, or 1½ tons. The construction of a glass bulb which will stand up under these conditions and yet is not unduly heavy and costly because of excess material is far from simple. As will be seen below, the employment of a metal cone has been found advantageous for large-sized tubes from a number of points of view. Furthermore, shaping the tube face to fit the picture results, here, in an appreciable space saving. Accordingly, the use of the simple *round* glass bulb, which is discussed in the present section, though universal for tubes with screen diameters of 12 inches or less, is relatively limited for larger tube sizes.

The material of these bulbs may be either hard or soft glass. Hard glasses, such as Pyrex and Nonex, have the advantage of excellent thermal and mechanical properties. From the elastic properties of the glass it has been found that, for a 12-inch screen having a radius of curvature of 16 inches, except near the edges of the screen, a wall thickness of ¼ inch is ample, allowing a safety factor between 3 and 5.

At the juncture of the curved disk of the screen and the conical walls of the tube the stresses may become very high. Either the glass must be considerably thickened at this point, or else a graduated curve,

which can be calculated by approximate methods, must carry the curvature of the screen into the conical walls. The two types of bulbs are shown in Fig. 11.2. The last-mentioned bulb shape requires less glass, but the bulb diameter for a given screen size must be somewhat greater. Considerations of space economy have led to a preference for the first solution.

The conical portion of the tube does not require as great a wall thickness as the screen, $\frac{3}{16}$ to $\frac{1}{8}$ inch being ample to withstand the

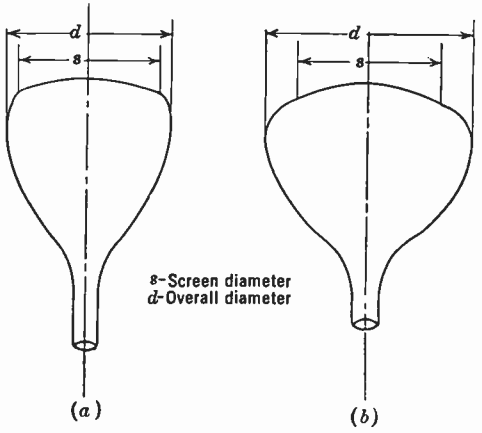


Fig. 11.2. Two Typical Round Glass Kinescope Bulbs.

stresses. In order to increase the strength for a given amount of material, and for reasons of contrast, the generatrices of the cone are slightly curved rather than straight, as can be seen from Fig. 11.2. The gun neck which is joined to the narrow end of the cone, as far as the atmospheric pressure is concerned, could be made of very thin glass. However, since it is subjected to other mechanical stresses, and to simplify the sealing of the cone to the neck, it is made the same thickness as the rest of the cone.

Before a bulb design is accepted it is put through rigorous pressure tests. The bulb is rough pumped and then placed in a pressure tank. The external pressure is then increased until the bulb implodes. Unless the bulb will withstand three to five times atmospheric pressure, the design is rejected.

The bulb must be arranged to meet the electrical requirements. In addition to the leads of the gun stem, a side contact is provided in

the conical walls of the bulb through which the high voltage is supplied to the ultor.*

The ultor consists of a conductive coating which extends into the gun neck, past the end of the gun. This coating also serves to collect the secondary electrons from the screen and therefore covers the inside walls of the tube, being carried almost to the screen. A metal film like that constituting the second anode of the iconoscope cannot be used in the kinescope because the light reflected back on the screen from the brighter portions of the picture would materially decrease the contrast range. For this reason the coating is made of a matte black material. A number of substances are suitable for this purpose, e.g., Aquadag, Aquagraph, Dixonac, lead sulphide, and carbon black in sodium silicate.

There are many methods of applying the blackening. In experimental tubes, especially those having a settled screen, the following procedure leads to a uniform coating without danger of contaminating the screen with the blackening material. A quantity of the blackening suspension somewhat greater than the volume of the tube to be coated is prepared. If commercial Aquadag is used it should be diluted with about three parts of distilled water. The tube in question is inverted and the gun neck closed with a two-hole stopper. The blackening suspension is admitted through one of the holes, while a glass tube through the second, extending to about ½ inch from the screen, serves as an air vent. A convenient way of controlling the amount of blackening liquid admitted is to supply the material from a reservoir which can be raised or lowered, and is connected to the bulb being treated by means of a rubber tube attached to an outlet at the bottom of the reservoir. Figure 11.3*A* shows the arrangement of the equipment. The reservoir is raised until the level of the liquid in the bulb just reaches the edge of the screen, then lowered to drain the bulb. Next the coating is dried with a gentle air blast until it is entirely free from moisture. In order to drive off the organic materials used as dispersing agent in the Aquadag, the final step is to bake the tube in air at 400°C. The finished tube will then have a uniform dull black conducting coating.

This process yields excellent results with experimental tubes, but it is too lengthy and expensive to meet the requirements of large-scale production. Manufacturing methods include spraying the inside of the bulb with special sprayers and masks. With suitable equipment,

* The term "ultor" designates the electrode to which the highest beam accelerating voltage before deflection is applied.

results equal to, or better than, those with the technique described above can be obtained. Alternative commercial processes for applying the black coating are the brushing method and the so-called rolling method. For the latter method the bulb is mounted in such a way that the axis of the tube makes an angle with the horizontal which is slightly greater than the half vertex angle of the conical portion of the bulb. The blackening suspension is flowed into the bulb until it just reaches the screen. By rotating the bulb the coating is made to flow uniformly over the inside walls of the tube, yet leaving the screen free

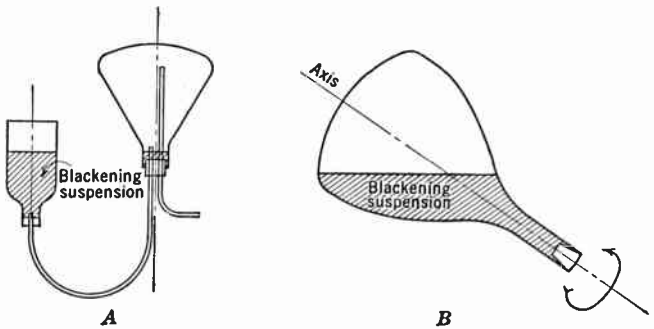


Fig. 11.3. Methods of Applying Blackening to Form Second Anode.

from contamination. Figure 11.3*B* illustrates the position of the bulb and the coating liquid.

Before the subject of the second-anode coating is left, mention should be made of the method of making contact between it and the lead wire sealed into the wall of the bulb. Usually, the lead consists of a short length of wire projecting into the tube. If the black coating is merely deposited over the wire, the contact will be very fragile and will be almost certain to open up during subsequent operation. However, if the wire and the glass, for an area of about a half an inch about the lead, are coated with silver paste which can be reduced to silver by heating, and the blackening layer is deposited over this silver, a firm, rugged connection between the external lead and the black coating can be effected.

11.3 Metal Tube Bulbs. For large screen sizes the weight and the bulk of early types of all-glass kinescope bulbs appeared unduly large and the maintenance of the narrow manufacturing tolerances required for mass production difficult. The metal kinescope was free from these drawbacks. Though improved design and manufacturing technique have overcome the earlier disadvantages of all-glass bulbs, the intro-

duction of metal kinescopes played a large role in the popularization of direct-view sets with picture diameters of 16 inches or more.

The metal tube bulb consists of three parts: the glass face plate, the metal cone, and the glass neck (Fig. 11.4). Owing to its mode of support, the face plate can be relatively thin; for a 16-inch tube it consists of $\frac{3}{16}$-inch window glass which is sagged or pressed so as to

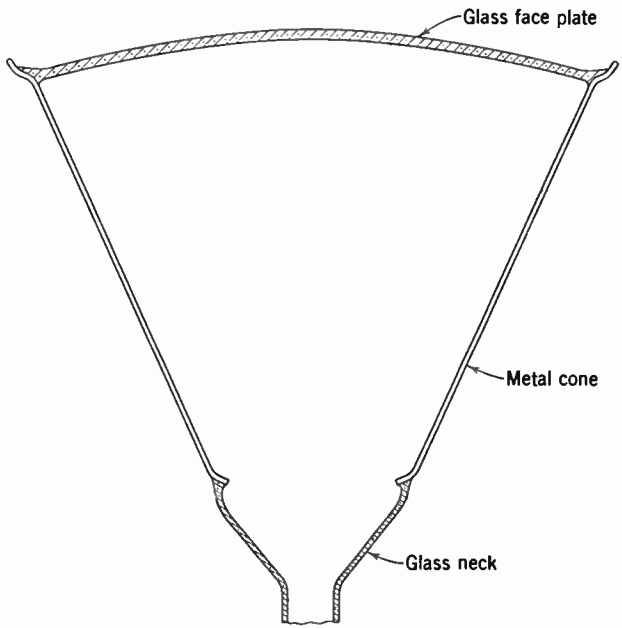


Fig. 11.4. Kinescope Bulb with Metal Cone. (Rose and Turnbull, reference 6.)

give a uniform radius of curvature of 27 inches. This small and uniform curvature permits maximum utilization of the face plate area for the picture.

The junction of the face plate and the metal cone is shown in detail in Fig. 11.5. The sealing surface, in the form of a truncated rim cone, is shaped to afford maximum resistance to tangential tension forces applied by the face plate. In the sealing process the cone and the face plate, placed on the sealing machine, are first brought uniformly to the annealing temperature of the glass (about 500°C). Then the sealing heat (of the order of 1200°C) is applied to the rim, causing the glass in contact with the metal to melt and form the seal, dissolving the adhering metal oxides. The shape of the face plate is maintained,

at the same time, by controlling the air pressure in the bulb. After sealing, the bulb is transferred to an oven near the annealing temperature of the glass and permitted to temperature-equalize. After this point it may be cooled rapidly in air. The coefficient of expansion of the metal (approximately $110 \cdot 10^{-7}$ degree^{-1}) being greater than that of the glass ($91 \cdot 10^{-7}$ degree^{-1}), the glass is placed under compression in both radial and tangential directions, giving it great strength.

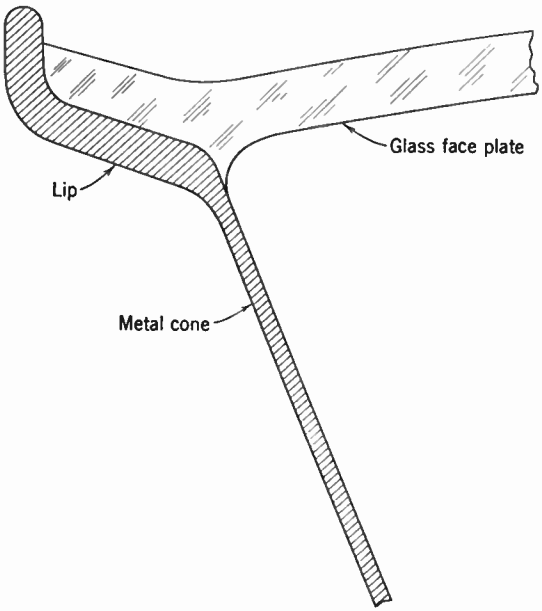


Fig. 11.5. Junction of Face Plate and Metal Cone in Metal Kinescope. (Steier, Kelar, Lattimer, and Faulkner, reference 5.)

The metal cone is made of a high-chromium iron alloy (AISI Type 430 or CRS). This has the advantage of good sealing properties—the chrome-iron oxides dissolve readily in the glass, forming a firm bond—and of the absence of phase changes within the temperature cycle. Such phase changes, which exist for alloys with smaller chromium content, may impose excessive sudden stresses on the glass in the cooling cycle and consequently break the seal. The cone is first spun and then machined.

The tube neck is made of lead glass (Corning No. 0120), possessing the high electrical resistance required for insulation between the deflection yoke (at ground potential) and the anode coating. Further-

more, the portion adjoining the metal cone is flared, to improve the external resistance between yokes and cone. A coating of insulating paint on the flared portion further prevents leakage.

Since the expansion coefficients of the metal and the lead glass are matched, it is possible to employ automatic machinery to form the butt weld between the neck and cone. Similarly, the coefficients of expansion of the neck and stem are matched to facilitate sealing.

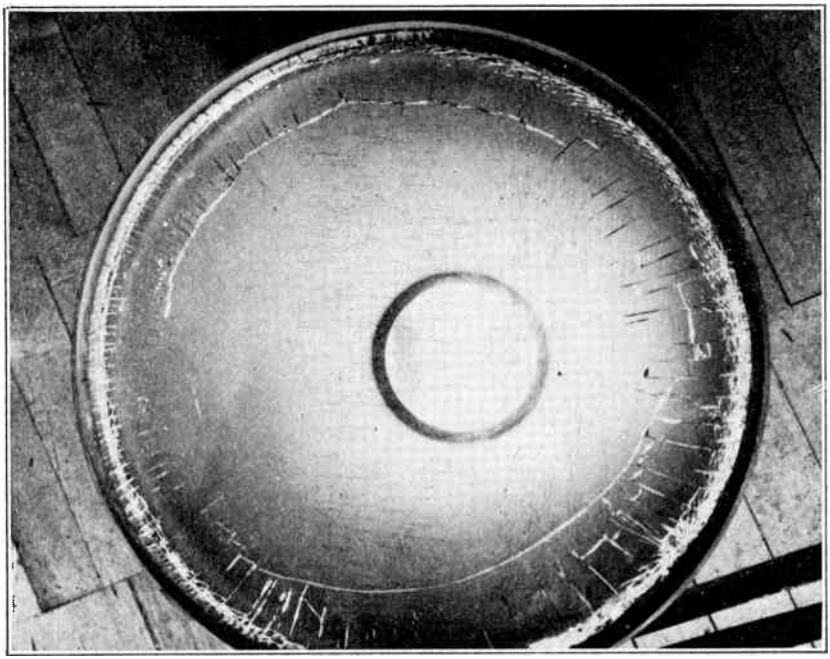


Fig. 11.6. Face Plate after Pressure Test Failure. (Steier, Kelar, Lattimer, and Faulkner, reference 5.)

The bulbs, finally, are subjected to an excess pressure test. If the pressure is increased until failure results, radial cracks develop as shown in Fig. 11.6. This is also the characteristic failure pattern for other types of shock. No violent implosion results; the cracked face plate is held in place by the vacuum until pressure has been equalized. As an example of the thermal stability of the structure, in one of the tests the bulbs are transferred from boiling water to liquid air, and vice versa, without resulting damage.

11.4 Rectangular Tubes. Kinescopes with rectangular faces are manufactured both with all-glass bulbs and with metal shells. As

already mentioned, these tubes have the advantage of having approximately 20 percent smaller height for a given picture width. This

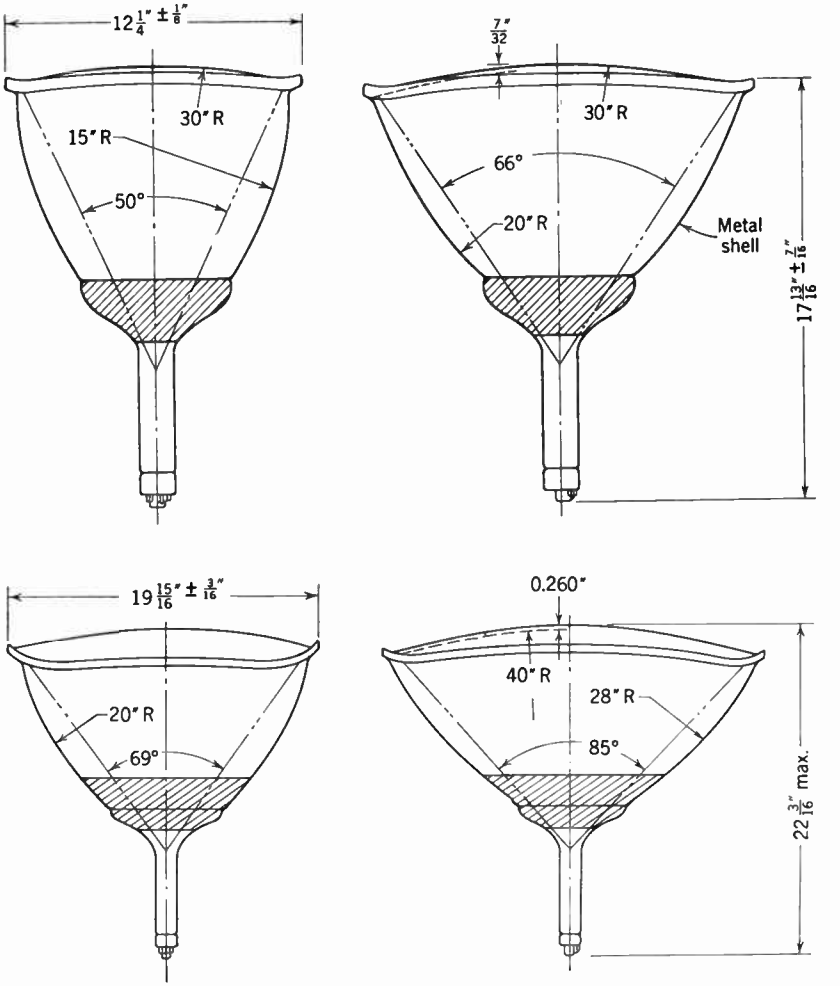


Fig. 11.7. Outline of Rectangular Metal Kinescopes (17CP4 and 27MP4—Vertical and Horizontal Projection).

represents a material gain in reducing set volume and accommodating the chassis. At the same time, the smaller degree of symmetry of these tubes introduces special problems in design and manufacture.

Figure 11.7 shows as an example outline drawings of the rectangular-screen metal kinescope 17CP4, giving a 14⅝ by 11 inch picture, and

of the much larger 27MP4, giving a $23\frac{7}{16}$ by $18\frac{1}{8}$ inch picture. Mechanical considerations dictate the slight curvature of the metal cone.

11.5 The Electron Gun. Like the electron gun in the camera tube, the kinescope gun is an electron-optical system for producing a narrow

Fig. 11.8. Typical Kinescope Gun.

pencil of electrons. The requirements of the kinescope gun are very different from those of the camera tube. It must be capable of producing several hundred microamperes of beam current instead of a mere fraction of a microampere; it must have control characteristics

such that it can be driven from the output of the video amplifier, and it usually is operated at a much higher overall voltage. On the other hand, a higher spot size can be tolerated.

The electron power which the gun is capable of delivering is one of the two factors determining the brightness of the reproduced picture, the other being the conversion efficiency of the fluorescent screen material. Because of its importance and because of the rather complicated theory back of the modern gun, a detailed discussion of it is reserved for a later chapter dealing with the gun problem alone.

The gun structure is assembled on a stem which is sealed into the end of the gun neck in such a way that the axis of the gun coincides with the axis of the tube. At the center of the stem a short length of tube serves as the exhaust exit. Although the processing of the kinescope does not require the admission of an activating material, nevertheless a getter must be provided to insure long life of the tube. The getter is generally attached to the gun structure. The channel containing the gettering material is open toward the tube walls and forms a closed loop with its U-shaped supporting wire; the gettering material is deposited on the tube wall when the loop is heated by induced high-frequency currents.

There are, of course, many different types of guns in current use. The great majority is based on the two-lens principle. A typical kinescope gun, designed for use with a magnetic second lens, is shown in Fig. 11.8.

11.6 The Fluorescent Screen. The fundamental importance of the fluorescent screen of the kinescope is obvious. Its properties determine the color of the image, its brightness for a given beam current, and also, to a certain extent, contrast, flicker, etc. The nature of the fluorescent materials available for coating the viewing screen was discussed in Chapter 2. Of the materials described, those most frequently used are mixtures of blue- and yellow-emitting phosphors, producing a white luminescence. The blue component is silver-activated zinc sulphide or titanium-activated calcium-magnesium silicate; whereas the yellow component is silver-activated zinc-cadmium sulphide or manganese-activated zinc-beryllium silicate.

The preparation * of the viewing screen does not end with the production of the fluorescent material. The material must be applied to the screen, an operation as important in determining the success or

* See Leverenz, reference 2.

failure of the tube as the actual manufacture of the screen material itself.

Quite generally in kinescopes the screen is viewed from the side opposite that from which the scanning electron beam strikes. The screen consists of a thin layer of finely powdered fluorescent material on the glass end of the tube. The material is normally attached to the glass with an inert binder, although it may also be ground and deposited in such a way that it adheres without the aid of any foreign substance.

In either case, the thickness of the layer of fluorescent material is rather critical. If the layer is too thin, the electron beam will not be stopped by the fluorescent material but will pass through it, giving up energy to the glass where it serves no useful purpose. On the other hand, a layer which is materially thicker than the penetration depth of the electrons will absorb some of the light unnecessarily as it passes through the underlying material which is not reached by the beam.

Not only the screen thickness but also the size of the particles making up the screen is important in determining its efficiency. From the standpoint of the resolution of detail, the size should be as small as possible. However, each individual grain is covered with a "dead layer," i.e., a shell of inert material which does not fluoresce. The thickness of the "dead layer" is approximately independent of the particle size, and the ratio of inert material to active phosphor increases as the particle size is diminished. Therefore, a compromise must be sought in the dimensions of the grains. An upper limit to the grain size is obviously set by the effective screen thickness, which, as has already been pointed out, is of the same order as the depth of penetration of the electrons in the phosphor. It will be shown below that, for the bombarding voltages employed in direct-view kinescopes, the depth of penetration is of the order of a few microns. This is only a small fraction of the spot diameter in a kinescope and hence constitutes not only an upper limit to, but also a most appropriate value for, the grain size.

Figure 11.9 shows the variation of light output with thickness for a settled silicate screen with an average particle size of 1 micron at 10,000 volts. The optimum thickness is seen to be that which results when 0.7 milligram of material per square centimeter is used. Since the density of the material is 4.2 grams per cubic centimeter, the effective thickness of this screen will be 1.6 microns. It should be noted that the effective thickness is very different from the actual thickness because the material is not close-packed.

The penetration of electrons through willemite can be shown, by means of relations developed by H. Bethe, to be given by

$$p = 2.5 \cdot 10^{-12} V^2 \text{ centimeter}$$

where p is the depth of penetration and V the accelerating voltage of the electrons striking the screen. Whereas the voltage applied in the measurements shown in Fig. 11.9 was 10,000 volts, the actual velocity of the electrons striking the screen corresponded probably to less than

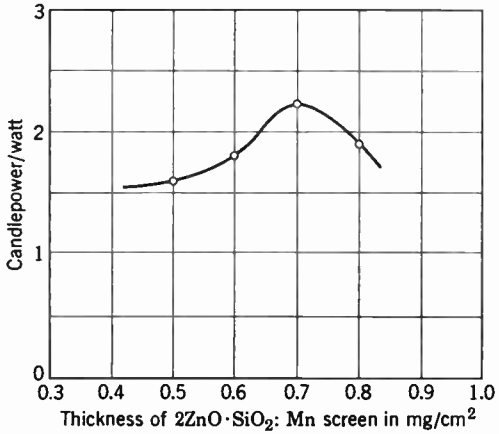


Fig. 11.9. Light Output as Function of Screen Thickness.

9000 volts, owing to the difference in potential between the ultor and the screen, which was pointed out in Chapter 2. Therefore, it can be concluded that the optimum thickness is only slightly less than the depth of penetration of the electrons. This is also found to be true for the sulphide phosphors, whose penetrations are given by

Zinc sulphide	$p = 2.83 \cdot 10^{-12} V^2$ cm
Cadmium sulphide	$p = 3.24 \cdot 10^{-12} V^2$ cm

They require screen thickness 10 to 30 percent greater than willemite for optimum thickness.

As was pointed out in Chapter 2, the break-point potential of a fluorescent screen is, at least in part, dependent on the screen thickness. In the case of the silicate phosphor, a screen of 0.7 milligram per square centimeter comes under the classification of intermediate thickness screens and has no sharply defined maximum bombarding voltage. Instead the difference between bombarding voltage and ultor potential

increases gradually after 4000 to 6000 volts have been exceeded. In projection kinescopes, operating with second-anode voltages ranging from 20 to 100 kilovolts, this difference is nullified by the application of a conducting aluminum film.

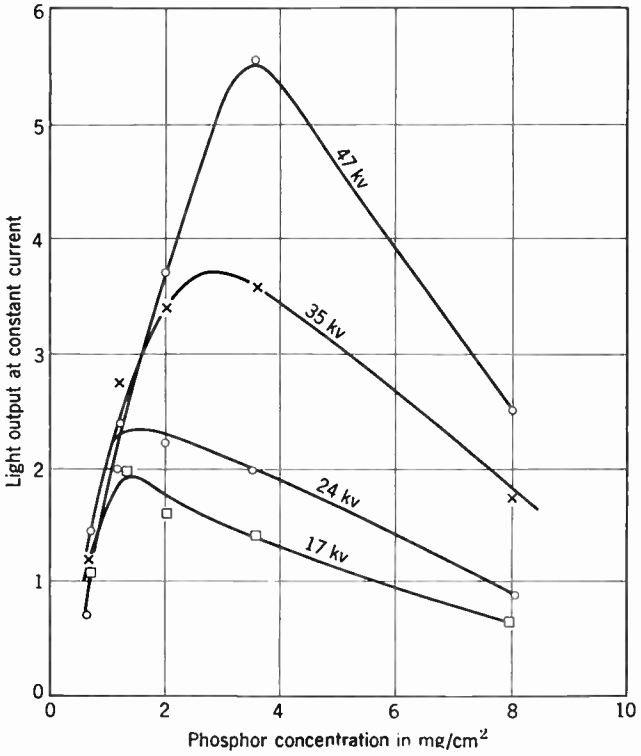


Fig. 11.10. Variation in Light Output with Screen Thickness for an Aluminized Zinc Sulphide Screen. (F. H. Nicoll, in Fonda and Seitz, *Preparation and Characteristics of Solid Luminescent Materials*, Wiley, 1948.)

The presence of the aluminum film also changes the optical properties of the screen, reflecting the backward emission of the phosphor. Figure 11.10 shows the variation in light output with screen thickness for such an aluminized film for a number of different operating voltages. The zinc sulphide screen in question had an average particle size of 3 microns. The decline from the maximum as the thickness is increased arises from two factors: the absorption of the light by the phosphor particles and the imperfect reflection of the aluminum screen. It can be demonstrated, however, that the first factor plays the major role.

11.7 Screening Procedure. The size of the freshly synthesized silicate phosphor particles may be as large as 0.1 millimeter when ordinary preparation technique is used. Before the material can be applied to a screen the grain size must be reduced by a factor up to a hundred. Since the efficiency of the material is dependent upon complicated and unstable structural properties, it is very sensitive to mechanical treatment. The grinding necessary to reduce the particle size should, therefore, be done in such a way that the minimum force required to fracture grains comes to bear on the material. This condition can best be met by grinding the phosphor in a ball mill, turned at a rate which produces a sliding rather than a tumbling action. The Moh hardness of the silicate phosphor is 5.5; consequently, quartz balls with a hardness of 7 are satisfactory for the silicate phosphor. The material is usually suspended in an inert liquid such as alcohol, acetone, or water. A section through a typical quartz ball mill is shown in Fig. 11.11.

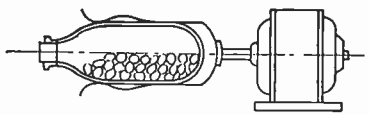


Fig. 11.11. Ball Mill.

The sulphide phosphors are even more delicate than the silicates, and, in general, it is not practical to reduce their size by grinding. However, by the proper preparation techniques these materials can be crystallized in grains of suitable size.

The material can be applied to the screen in a variety of ways. Those most frequently used are dusting, spraying, and settling through a liquid medium.

The last-named technique permits an accurate control of screen thickness and is, therefore, favored for both experimental and mass production of screens. It has the added advantage of wasting no material. A suspension of the ground fluorescent material in very pure distilled water is made, using an amount of phosphor just sufficient to cover the area of the screen to the desired thickness. The volume of liquid used in making the suspension is relatively unimportant but should be sufficient to cover the screen to a depth of 5 to 10 centimeters. Since the dielectric constants of the liquid and the particles are different (i.e., for water, $\varepsilon = 80$; for zinc orthosilicate, $\varepsilon = 15$), the suspended particles tend to become negatively charged. Such charges will cause the material to settle nonuniformly, producing a screen covered with fine lines which resemble the force lines on a surface dusted with iron filings in the presence of a nonuniform magnetic field. To avoid

this difficulty the particles must be kept charge-free, which can be done by adding a small amount of mild electrolyte to the distilled water. Ammonium carbonate has been found to be very satisfactory for this purpose. It should be used in a concentration of about 12.5 grams of the ammonium salt for each gram of the phosphor.

A very convenient procedure is to grind the phosphor in a concentrated aqueous solution of ammonium carbonate. The resulting mixture can be directly diluted to the desired concentration required for the settling suspension. Not only does the slightly alkaline ammonium carbonate solution serve to eliminate the charges on the particles, but also it may be made to reduce the dead layer by a slight decomposition of the material. Further, by a similar action, it increases the adhesion of the phosphor to the screen on which it is settled.

A suitable quantity of the phosphor in suspension having been prepared, the liquid is poured into the bulb and the screen material allowed to settle, after which the remaining liquid is decanted and the screen dried. The introduction of a "cushion layer" of potassium silicate solution, over which the phosphor suspension is flowed, aids in attaining screen uniformity. The settling may require 3 to 8 hours, and during this time the bulb must be kept quite free from vibration. Also, the liquid must be kept at a constant low temperature to avoid undesirable convection currents which will cause the screen to be "loaded" at the center or the edges. If the ambient temperature is lower than that of the liquid, the convection currents will be such that the center of the screen will be too thick, whereas a higher ambient temperature produces the opposite effect. It should be noted that small, controlled convection currents can be used to obtain a uniform screen on a curved surface such as is usually encountered in a normal kinescope.

When the settling is complete, no scattering should be observable from a light beam passed through the liquid parallel to the screen. The liquid can then be decanted slowly over a period of half or three-quarters of an hour. During this pouring process, vibration must be avoided. After the pouring has removed all the liquid, the screen is dried with a gentle current of clean warm air, eventually after removal of remnants of silicate from the tube neck by washing with dilute hydrofluoric acid.

In order to carry out the settling and pouring operation described in the laboratory it has been found convenient to use the apparatus illustrated in Fig. 11.12. The bulb is clamped to a small platform which is hinged to a massive cast-iron base. The base is suspended

by means of damped helical springs, so that it will be free from vibration. A smooth-running motor is arranged so that it will slowly tilt the hinged platform and the kinescope clamped to it, in this way performing the pouring operation. In mass production the bulbs are instead inserted, face down, in pockets on a slowly moving belt and the settling solutions introduced. At the end of the settling period the belt passes over a large-diameter pulley, so that the bulbs are slowly tilted and inverted, discharging the settling liquid.

Whereas liquid settling is preferred for the highest-quality screens, spraying of the screen material lends itself well to high-speed production methods. In this case the phosphor is often suspended in a volatile organic liquid, such as acetone, to which has been added a small amount of binder. The binder may be in the form of nitrocellulose or cellulose acetate, or some other material which can easily be removed by baking after the screen has been applied. In general, however, a binder is not necessary, as the screen material will adhere if the particle size is small. To avoid the charging of the phosphor grains it is often necessary to add a small amount of organic acid (e.g., formic acid or acetic acid) or base (e.g., trimethylamine) to the suspending liquid. The particle size suitable for spraying is 1 to 6 microns.

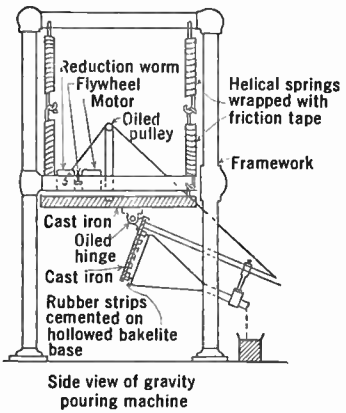


Fig. 11.12. Settling and Pouring Apparatus.

Larger particles, 10 to 30 microns in diameter, are usually deposited on the screen by dusting or settling through air. A binder in the form of sodium or potassium silicate is often required to cause the material to cling to the screen. This technique lends itself well to production methods and is extensively used by tube manufacturers. The powdered material is put into a dust gun which is inserted in the inverted kinescope. The gun produces a uniform cloud of phosphor which settles on the glass end of the tube. Two forms of the dust guns are shown in Fig. 11.13. Because it lends itself to the deposition of large particles, dusting is widely applied in the production of sulphide phosphor screens, which are easily damaged by grinding.

One basic advantage of settling procedures as compared to, for example, spraying resides in the fact that the larger particles are

deposited first, directly on the face plate, resulting in reduced optical contact between phosphor particles and face plate. Wherever optical contact is established, a large portion of the light emission is trapped by total reflection within the face plate, leading to reduced efficiency as well as reduced contrast. The same variation in the speed of settling with particle size may be utilized to control the color and efficiency of mixed phosphor screens. Thus, if the particles of the blue zinc sulphide

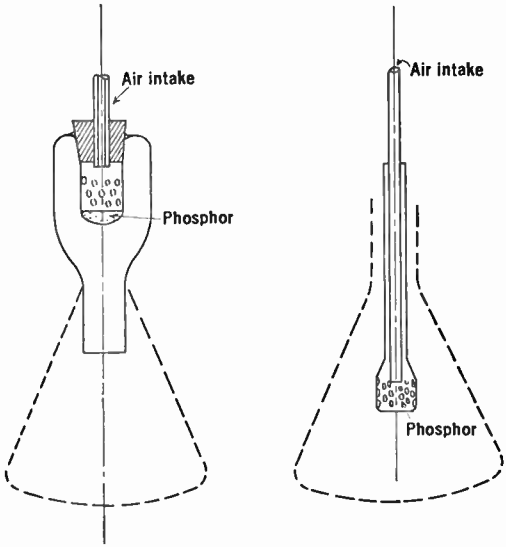


Fig. 11.13. Dust Guns for Applying Screen Material.

component of a white television screen are made larger than those of the yellow zinc–cadmium sulphide component, they will settle to the bottom. Since the blue radiation is absorbed by the zinc–cadmium sulphide crystals, whereas the yellow radiation is transmitted freely by the zinc sulphide, this leads to a screen of maximum efficiency. In projection tubes, employing yellow zinc silicate and blue zinc sulphide phosphors, the same arrangement minimizes color changes arising from saturation effects, as will be seen later. Conversely, in a cascade screen the yellow-emitting material is deposited first, and the blue emission is utilized specifically to excite the yellow phosphor.

11.8 Metal Backing of Kinescope Screens. An important improvement in the performance of kinescopes operating with relatively high voltages may be achieved by the deposition of a continuous thin metal film over the fluorescent screen. Such a metal film:

(1) serves to maintain the fluorescent screen at full ultor voltage, eliminating the "sticking effect,"

(2) protects the screen from bombardment with negative ions, which cause a loss in conversion efficiency ("ion spot"), and

(3) increases the efficiency of light utilization and improves contrast by reflecting the light emitted by the screen in a backward direction.

To possess these properties it is necessary that the film, though thin enough to transmit the electron beam with relatively small energy loss,

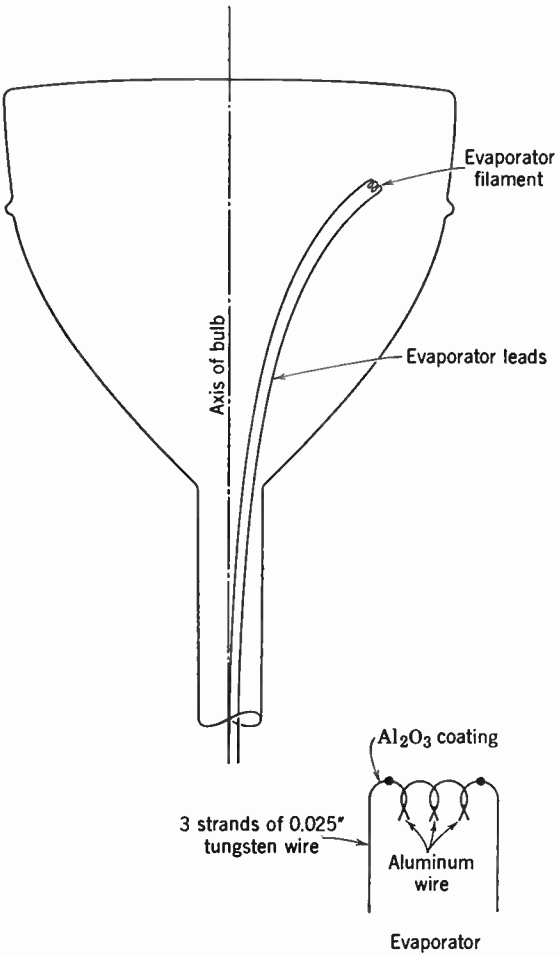


Fig. 11.14. Arrangement for Evaporating Aluminum Film on Screen. (H. P. Steier.)

reflect light fully and have sufficient mechanical strength to be unaffected by the electron bombardment; finally, it must have no chemical action on the fluorescent material.

Of all the metals tested, aluminum satisfies these properties most completely. To produce the flat metal sheet required, a thin collodion "blanket" is first formed over the screen, by letting a drop of collodion solution spread on water; the water is removed by siphoning or spinning after the film has solidified. Thereupon the tungsten evaporator filament with its aluminum wire "riders" is introduced into the bulb through a vacuum-tight bearing, in such a fashion that it assumes an off-axis position (Fig. 11.14). After evacuation the aluminum is evaporated by heating the filament while the bulb is rotated slowly about its axis. In this manner a relatively uniform film is obtained on the screen. It may be estimated that approximately one-fourth the total amount of aluminum evaporated contributes to the film over the screen. The supporting collodion blanket disintegrates during the later baking of the tube.

The optimum thickness of the aluminum film increases with the operating voltage of the tube. At high voltages it should be made such that the energy loss of the beam electrons within it is negligible while yet negative ions of equal energy are fully absorbed. At low voltages it may be necessary to purchase mechanical strength and optical reflecting power by increasing the thickness to a point where the energy loss within the film is appreciable. Even a loss of 50 percent may be balanced approximately by the light gain occasioned by optical reflection. In practice, the thickness of the aluminum films ranges from 500 to 5000 Å. Somewhat greater thicknesses are required with electrostatic focus (and magnetic deflection) than with magnetic focus, since, with electrostatic focus, the ion spot is more highly concentrated.

The energy loss within the aluminum film may be ascribed to two factors: the average velocity loss of the electrons, as given by the Thomson-Whiddington law,

$$v_0^4 - v^4 = b'\rho x, \quad v_0^4 = b'\rho p \tag{11.1}$$

and the absorption of electrons within the film given by Lenard's absorption law,

$$\frac{I(r + dr)}{I(x)} = \epsilon^{-\alpha\, dx}, \quad \alpha = \frac{b''\rho}{v^4} \tag{11.2}$$

Here v_0 is the velocity of the electrons entering the film, v, their most probable velocity at a depth x, ρ the density of the film, p the pene-

tration of the film for electrons of velocity v_0, $I(x)$ the electron current at depth x, and b' and b'' are two universal constants. Strictly, Eq. 11.2 is applicable only after the electrons have penetrated the film to a sufficient depth that they have acquired a random angular orientation. If this fact is neglected, the two equations, together, yield a simple expression for the fraction of beam energy transmitted after passing through a film thickness x:

$$\frac{W}{W_0} = \frac{mv^2}{mv_0{}^2}\,\epsilon^{-\int_0^x \alpha\,dx} = \sqrt{\frac{p-x}{p}}\,\epsilon^{-\int_0^x \frac{b''}{b'(p-x)}\,dx} = \left(1-\frac{x}{p}\right)^{\frac{1}{2}+\frac{b''}{b'}} \tag{11.3}$$

If the values of b' and b'' obtained by Terrill * from measurements on aluminum films ($b'\rho = 1.4\cdot 10^{43}$, $b''\rho = 2.7\cdot 10^{43}$) are substituted Eq. 11.3 becomes

$$\frac{W}{W_0} = \left(1-\frac{x}{p}\right)^{2.4} \tag{11.4}$$

The extent to which Eq. 11.4 can be regarded as reliable may be inferred from the comparison, in Fig. 11.15, of the relative energy loss as function of layer thickness as computed from Eq. 11.4 and as derived from measurements by K. Gentner † on the energy dissipation of 25.6-kilovolt electrons in aluminum. It is seen that Eq. 11.4 gives throughout too high values for the energy absorption—particularly for very thin films, for which the absorption of electrons is negligible since they have been scattered only slightly. The discrepancy is actually reduced, however, by the fact that the penetration depth employed in conjunction with curve *I*,

$$p_I = 8.7\cdot 10^{-13}V^2 \text{ centimeters} \tag{11.5}$$

is materially less than the penetration depth from the Thomson-Whiddington law (Eq. 11.1) as usually defined. Equation 11.5 is, at the same time, in perfect agreement with Terrill's data.

Curve I of Fig. 11.15 and Eq. 11.5 permit the calculation of curves indicating the transmitted energy of a film as function of the energy of the incident electrons (Fig. 11.16). Since, according to the measurements of Walkenhorst,‡ aluminum films over 500 Å thick have a reflection coefficient in the visible (5000 Å) which exceeds 90 percent

* See Terrill, reference 12.

† See reference 11; the distribution of the energy loss by back scattering, which totals about 8 percent, has been estimated.

‡ See W. Walkenhorst, reference 13.

(Fig. 11.17), it is seen that it should be possible to achieve a gain, rather than a loss, in brightness by the application of aluminum films for voltages in excess of about 4000 volts. Experimental measurements

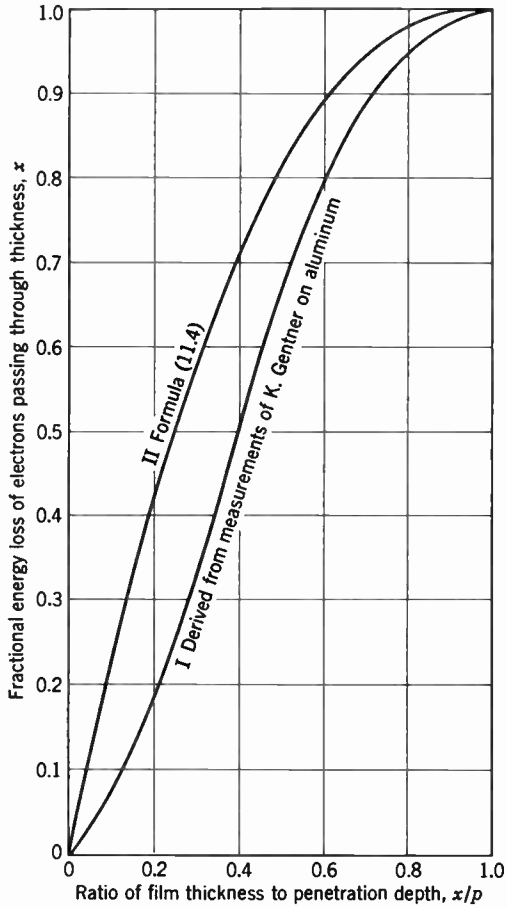


Fig. 11.15. Energy Loss of Electrons in Passing through Thin Films.

on kinescopes of similar construction with and without an aluminum film confirm this expectation (Fig. 11.18). If here the effective reflection coefficient is set, in view of the imperfection of the film and screen absorption, at 50 percent, the film thickness may be estimated (from the crossover point) to be 370 Å. The gain becomes greatest, however, at very high operating voltages where, without the metal film, the screen tends to lag far behind ultor potential. It is hence

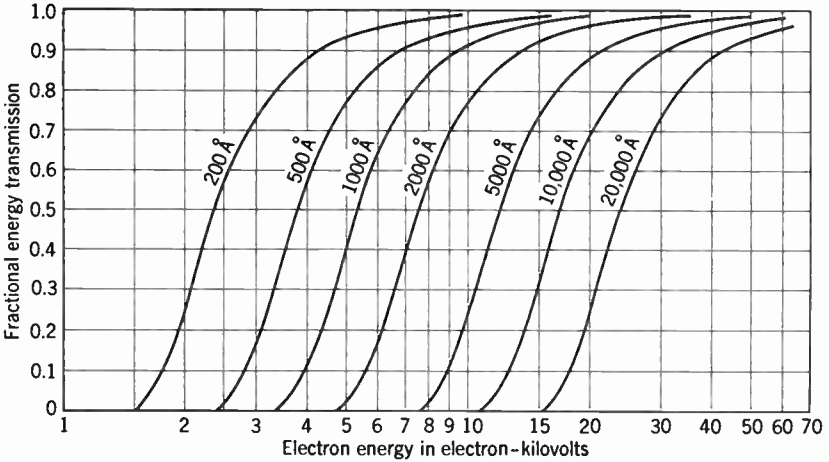


Fig. 11.16. Energy Transmitted by Aluminum Films of Different Thicknesses as Function of Voltage. (Based on Curve I of Fig. 11.15.)

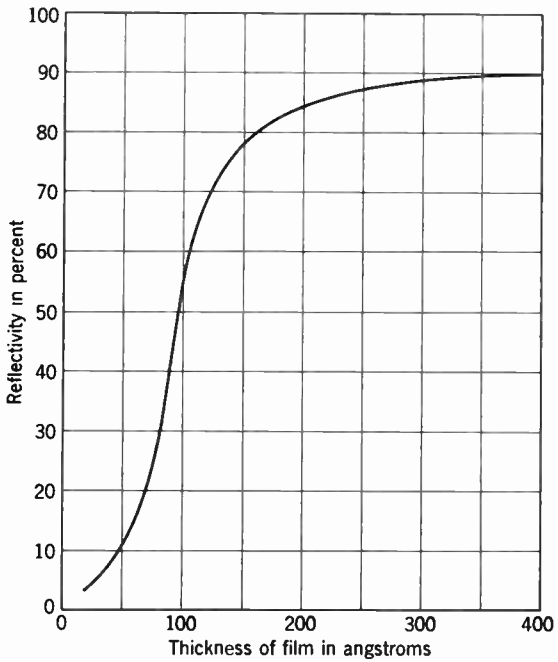


Fig. 11.17. Reflectivity of Aluminum Films for 5000 Å. Radiation as function of thickness. (Corrected for oxide layer—W. Walkenhorst, reference 13.)

universally employed in projection tubes. Also, in tricolor kinescopes, where slight charging effects can lead to color dilution, aluminizing is advisable. Its advantage is not so clear for direct-view tubes, since here the metal film not only increases the effective light output of the phosphor but, at the same time, enhances the fraction of incident ambient light which is reflected; under high ambient-light conditions it thus may reduce, rather than enhance, the contrast of the picture.

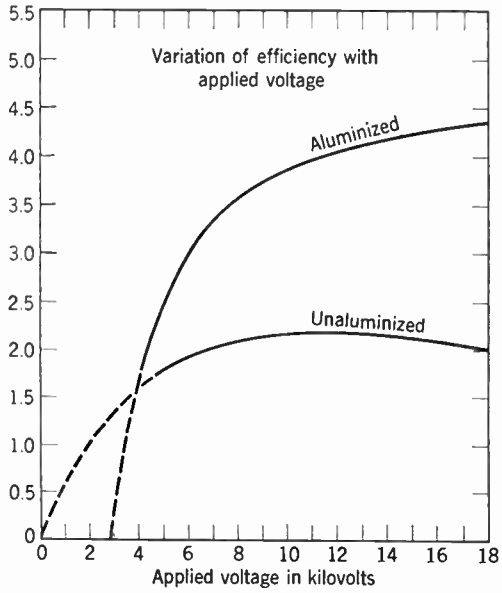


Fig. 11.18. Efficiency as Function of Voltage for Similar Aluminized and Unaluminized 12-Inch Kinescope Screens. (Epstein and Pensak, reference 9.)

It may be noted that the experimentally measured light gain resulting from the use of an aluminum film is not as great as might be expected from the curves of transmitted energy in Fig. 11.16 and the reflection characteristics shown in Fig. 11.17. This is not surprising since the aluminum films are unquestionably not truly flat.

The effectiveness of the film in suppressing ion spot depends not only on its thickness and the operating voltage of the tube, but also on the nature of the second lens (whether electrostatic or magnetic) and on the residual pressure in the tube. For example, a tube with magnetic focus operating at 11,000 volts or less will normally achieve ion-spot elimination for a life period of 500 hours with an aluminum film 1000 to 1500 Å thick, which may enhance the light output 20 to

25 percent. On the other hand, for an electrostatic focus tube with similar gas pressure operating at 10,000 or 11,000 volts effective ion-spot elimination demands a film thickness which may actually reduce the light output by 15 to 20 percent.

The employment of an aluminum film also has an appreciable effect on the color of the picture. Selective settling and reduced effective voltages and currents combine to enhance the relative brightness of the yellow component. For example, in a 10-inch tube operating at 9000 volts, the ratio of the blue component to the yellow component must be increased from 50:50 to 58:42 when an aluminum film is applied.

11.9 Processing of the Kinescope. The experimental exhaust and activation of a kinescope is relatively simple. An ordinary vacuum system capable of producing a pressure of 10^{-6} millimeter of mercury or less, and equipped with an oven, is satisfactory. The evacuation schedule is the same as that for the iconoscope, the tube being baked on the pump at a temperature of 350 to 450°C, depending upon the kind of glass used in the bulb. After the tube has been baked the gun parts are thoroughly outgassed by heating them to red heat for a period of 10 minutes to half an hour. The activation of the gun, which then follows, proceeds exactly as described in Chapter 1. Just before the tube is sealed off the system, the getter is exploded by heating the wire loop incorporating the getter channel with induced high-frequency currents.

A tube prepared in this way should give good service and have a long life. The principal precaution to be taken is to avoid residual gas. The beam current in a kinescope is relatively high, and therefore even a small amount of gas will result in the generation of positive ions which will greatly decrease the life of the thermionic cathode. For this reason the walls of the tube and the metal parts must be very carefully outgassed.

In the absence of an ion trap in the electron gun or an aluminum film deposited over the fluorescent screen, residual gas manifests itself by producing a dark, inactive spot in the center of the fluorescent screen. This spot is caused by the bombardment of the screen with negative ions which destroy the phosphor. The character of the ions producing this effect has been studied by Knoll and others by a rather ingenious method. Before the screen material is destroyed, the ions are capable of exciting fluorescence. If the tube is placed in a strong magnetic field with a known flux density, the electron beam can be

deflected off the screen whereas the paths of the ions are deviated by an amount proportional to their e/M ratio. Thus the tube itself serves as a mass spectrograph. By reversing the potentials on the tube positive ions emitted by the gun can be analyzed in similar fashion. The beam is generally broken up by the deflecting field so as to produce a number of spots, revealing the presence of more than one gas. By measurement of the positions of these spots the gases have been identified. The principal gases found in order of their occurrence by volume are oxygen, hydrogen, carbon monoxide, water vapor, nitrogen, and carbon dioxide. Other gases are present but in lesser quantity.

The production processing of the kinescope presents a rather different problem. Here time and economy of operation are of major importance. Particularly to be avoided are all hand-performed individual operations which require skilled workers; in operations such as gun mounting, which do require hand labor, the latter is simplified and accelerated by carefully designing mounting jigs. The sealing of the gun stem to the tube neck is done by automatic machinery, which requires only manual loading. For exhaust and processing, large kinescopes are mounted individually on mobile carts, incorporating a complete exhaust system. On these carts they pass slowly through the baking oven, where voltages for the several tube electrodes are provided by sliding contacts. In this manner exhaust, bake, outgassing of metal parts, and gun activation are carried out automatically in proper sequence. As a final step the copper tubulation is pinched off and the tube transferred to a conveyor belt, which transports it to the basing reel. Here the base is applied and, after the basing cement is cured with infrared heat, the leads are soldered to the base pins. Getters are flashed and the cathodes aged on the conveyor belt transferring the tubes from basing to testing.

11.10 Tests and Performance. Every responsible manufacturer must test the viewing tubes carefully to keep a constant check of the production methods. These tests can be divided into two classes. The first class consists of routine checks on the performance of every tube produced. The second consists of a more critical examination of a selection of tubes from, e.g., each day's production prior to tube shipment.

The routine tests must determine whether the screen on each tube is satisfactory, whether the spot meets the requirements of size and shape, and whether the control characteristic is correct. The tests also

should be of such nature that they can be performed simply and quickly. One form of test outfit which meets these requirements has been designed by C. E. Burnett.* This equipment, in addition to supplying the anode voltages and vertical and horizontal deflection to the tube being tested, is arranged to deliver a signal to the control

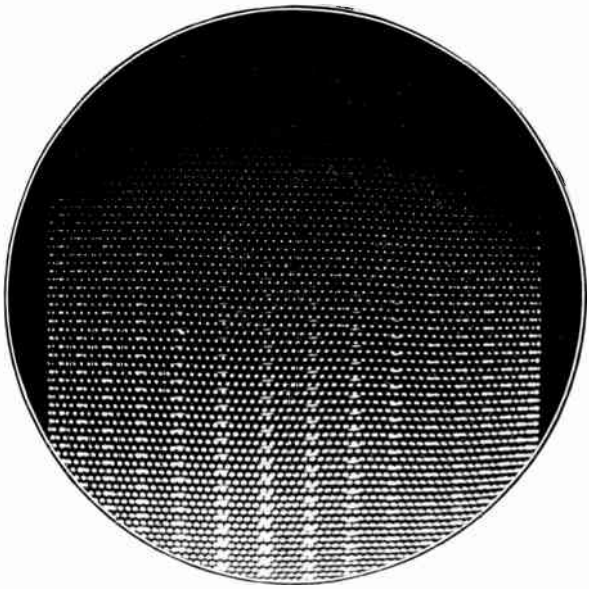


Fig. 11.19. Kinescope Test Pattern.

grid, which breaks the pattern up into dots of about picture-element size which vary in brightness over the screen. A pattern of this type is shown in Fig. 11.19. By comparing this pattern with a standard pattern, the performance of the tube can readily be judged. Visual inspection will reveal whether or not the spot dimensions will give the desired resolution. A comparison of the brightness of the pattern with the standard indicates the merit of the screen. The control characteristics can be estimated from the variations of pattern brightness over the screen. With this type of equipment the tests can thus be performed simply by inserting the kinescope into its socket and examining the pattern.

* See Burnett, reference 14.

The second class of examination is much more rigorous and must supply quantitative information about the performance of the kinescope not only under normal operating conditions but also under conditions of overload.

Thus a number of tubes are subjected to pressures of 45 pounds per square inch to evaluate their strength; others are life-tested and still others stored for a prescribed period and retested.

Electron-gun performance tests include a microscopic examination of the spot in various positions on the screen and micrometer measurements of spot size. The line produced by deflecting the beam is similarly examined. The extent to which the spot "blooms" or increases in size when the grid is made positive is also determined.

Furthermore, the cathode should be studied for flaws in the emitting coating. This can be done by adjusting the potentials on the gun in such a way as to produce a magnified image of the cathode on the fluorescent screen. In other words, the gun is used as an electron microscope.

In addition, measurements are made of the control characteristics of the gun and of the current to each of the gun elements. Overvoltage tests of the gun yield information as to the quality of the insulation and the extent of cold discharge.

The fluorescent screen should be examined to determine quantitatively the efficiency of the phosphor, and curves taken of light output against current and voltage. Also the spectral output should be measured.

If the several tests have been passed satisfactorily, the tubes in the production unit in question are transferred to the packing room for shipment or storage. At this stage metal tubes are also provided with coats of conducting protective paint on the metal cone and an insulating coating on the neck bulge, to minimize surface leakage.

11.11 Contrast. The behavior of the kinescope with respect to contrast is an extremely important factor in determining its overall performance. Therefore, the causes which contribute to loss of contrast must be studied in some detail. It will be seen that such an analysis * points out effective methods of improving picture contrast.

If a conventional kinescope of an earlier vintage is observed when the spot is stationary at the center of the viewing screen it will be noticed that, in addition to the single bright point of light due to the impact of the electron beam, considerable light is scattered from other

* See Law, reference 15.

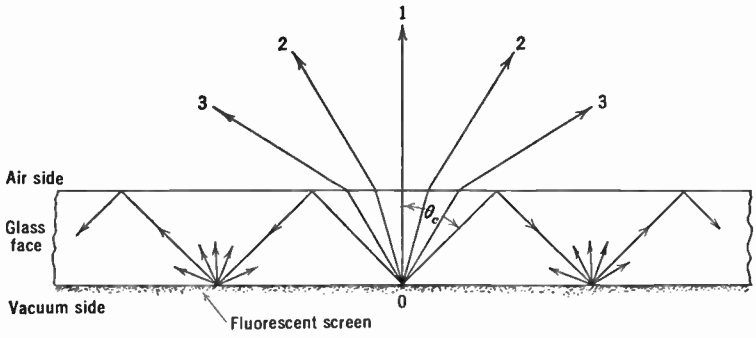
parts of the screen. Close to the beam, and extending over an area of several square centimeters, is a rather faint luminosity due to normal reflections. The spot is surrounded in addition by two or more concentric rings of light, the smallest having a radius of about a centimeter, due to total internal reflections in the glass face of the tube. Finally, there is a diffused glow over the entire screen which is more or less uniform but may have a slight maximum in the vicinity of the spot. This general scattering is due primarily to the effect of screen curvature, reflections from the bulb walls, and scattered electrons.

A consideration of these factors will make it immediately apparent that, for a given tube, the value of contrast range which can be obtained depends upon the type of pattern reproduced. In other words, much greater contrast is possible when the pattern consists of a small bright area on a dark background than when it consists of a dark area on a white background. Therefore, a single figure will not suffice to describe the performance of the tube with respect to contrast. Two measurements which, when taken together, give a fairly accurate estimate of the merit of the tube are those giving the detail contrast range and the overall contrast range.

The detail contrast range is measured by observing the brightness ratio of a small dark area on a bright background. For this measurement the beam is made to scan the normal picture area and a signal is applied to the control grid such that the beam has its maximum operating intensity except over an area which is small compared to the picture area, where the beam is biased off.

If the same area is scanned, but a signal is applied to the grid which biases the beam to cutoff over half the picture, overall contrast range can be obtained by measuring the brightness at the centers of the light and dark areas.

The major factor in reducing the contrast range of a picture is the scattering due to internal reflections in the glass. Figure 11.20 illustrates the nature of this type of scattering. The screen material is shown as embedded in the glass wall. Rays leaving the screen at the point of electron bombardment will be refracted at the interface between the glass and air. Any ray which leaves the screen material at an angle to the normal greater than θ_c will not emerge from the glass but instead will be totally reflected back to the screen. When this totally reflected light again strikes the screen material some of it will be scattered, producing the first visible ring, while the remainder will be totally reflected to form additional rings.



Cross section of kinescope face showing how total internal reflections produce halation

(*a*)

(*b*)

Fig. 11.20. Internal Total Reflection.

The critical angle required to produce total reflection can be calculated, from Snell's law and a knowledge of the index of reflection μ_1 of the glass, as follows:

$$\mu_1 \sin \theta_1 = \mu_2 \sin \theta_2$$

$$\mu_2 = 1, \quad \mu_1 = 1.54$$

$$\theta_2 = 90°, \quad \theta_1 = \theta_c = \text{arc sin} \frac{1}{1.54} = 41°$$

From the geometry of Fig. 11.20*a* it will be seen that the diameter of the first ring will be given by the relation

$$D = 4t \tan \theta_c$$

where t is the thickness of the glass and can be assumed to be ¼ inch. This gives a diameter of 0.85 inch.

It is evident that only a fraction of the light emitted at the spot escapes from the screen in the form of useful light. Since the distribution of light emitted by the fluorescent screen approximates Lambert's law,

$$L = B \cos \theta$$

where B is the light flux per unit area emitted by the screen and θ is the angle between the direction of observation and the normal, the fraction of light which can escape can be calculated with the aid of the following integrals, whose construction can be seen from Fig. 11.21a:

$$R = \frac{2\pi \int_0^{\theta_c} B \cos \theta \sin \theta \, d\theta}{2\pi \int_0^{\pi/2} B \cos \theta \sin \theta \, d\theta} = \sin^2 \theta_c = 42 \text{ percent}$$

Thus it will be seen that under the condition postulated about 60 percent of the light emitted in a forward direction is trapped by internal reflection.

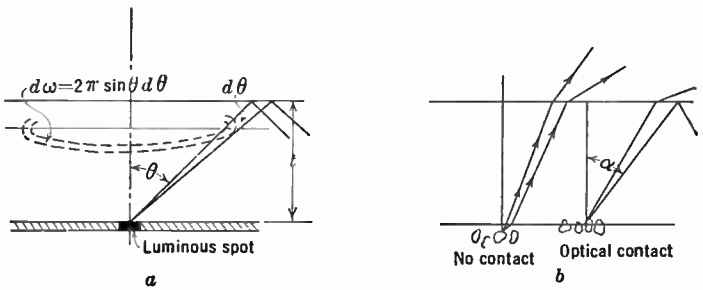


Fig. 11.21. Reflection and Refraction in the Tube Face.

The assumption was made, however, that all the phosphor grains were embedded in the glass. Actually, the optical contact between the fluorescent material and the glass is only about 20 to 30 percent, depending upon the way in which the screen is prepared. As can be seen from Fig. 11.21b, no internal total reflections exist in the case of luminescent grains which are not in optical contact with the glass. Therefore, assuming 30 percent contact, the actual loss by internal reflection is 60 percent times 30 per cent, or 18 percent of the emitted light. The internally reflected light is gradually scattered out of the

screen as it is reflected back and forth between inner and outer glass surfaces. At least half of the scattered light must be in the forward direction, that is, 9 percent, and contributes to loss of contrast. On this basis, and owing to this cause alone, a contrast ratio not greater

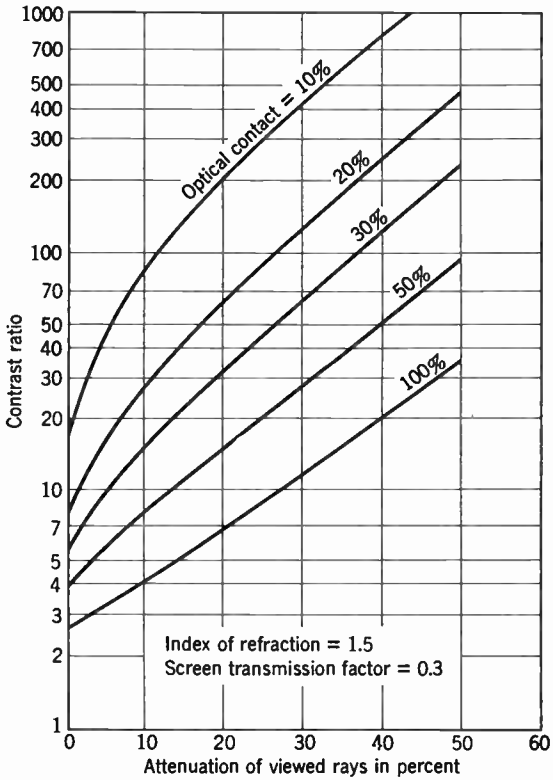


Fig. 11.22. Contrast Ratio as Function of Attenuation in Face Plate. (Law, reference 15; courtesy of *Proceedings of the Institute of Radio Engineers.*)

than 82 percent/9 percent = 9 can be expected. A more accurate calculation, including multiple scattering within the screen, absorption of the glass, and edge effect, reduces this figure to about 6. Loss of contrast resulting from normal reflections in the tube face, at less than the critical angle, plays only a minor role.

Since most of the loss of contrast is due to the scattering at the first ring of light, it has a much greater effect on detail contrast ratio than

on overall contrast ratio, a fact which is confirmed by measurements. The methods of overcoming this loss of detail contrast were clearly recognized by Law in his analysis of the problem. They are:

1. Reducing the optical contact between phosphor grains and glass.
2. Depositing the phosphor on a separate flat sheet of negligible thickness.
3. Making the tube face so thick that the first ring falls outside the picture area.
4. Incorporating a neutral absorbing material in the tube face, which attenuates the multiple-reflected rays much more strongly than those which are transmitted directly.

The degree of optical contact is influenced by the method of deposition of the phosphor (e.g., 30 percent for spraying, 20 percent for

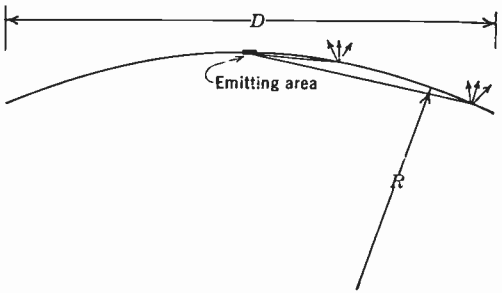


Fig. 11.23. Effect of Face Curvature on Contrast.

settling, and 15 percent for dusting), but cannot be made negligible. Modifying the thickness of the substrate as prescribed in solutions 2 and 3 is open to practical objections. Hence, the employment of absorbing glass (e.g., Filterglass) for the face plate represents the most practical method of minimizing the effect of reflections within the face plate and is widely utilized in modern television tubes. Law has calculated the contrast ratio for various degrees of optical contact as function of the attenuation of the directly viewed image (Fig. 11.22), for 30 percent light transmission by the screen and an index 1.5 of the face plate. It is seen that, at the cost of reducing the picture brightness by 20 percent, the detail contrast ratio may be increased by a factor of 6 to 8.

The remaining contrast-reducing factors influence primarily large-area contrast. One of these is the effect of screen curvature, permitting

light scattered at the point of incidence of the beam to reach other points of the screen, as shown in Fig. 11.23. The light spread over the screen can be readily calculated. If D is the diameter of the screen and R its radius of curvature, the maximum angle over which light can reach the screen, measured from a tangent to the shell at the point of impact of the beam (assumed to be the center of the screen), will be α, where

$$\sin 2\alpha = D/(2R)$$

By proceeding as before, the ratio of scattered light to forward light will be

$$S = 1 - \int_0^{\frac{\pi}{2}-\alpha} 2 \sin\theta \cos\theta \, d\theta$$

$$= \sin^2 \alpha = \frac{1}{2}\left\{1 - \left[1 - \left(\frac{D}{2R}\right)^2\right]^{\frac{1}{2}}\right\} \cong \frac{1}{16}\left(\frac{D}{R}\right)^2$$

Since at least half of the light scattered by the screen is in the backward direction, the ratio of brightness of the dark area to the bright area (assuming Lambert's law) is not greater than

$$S = \frac{1}{32}\left(\frac{D}{R}\right)^2$$

For a typical 12-inch glass tube ($D = 12$ inches, $R = 16$ inches), S is found to be 0.02, indicating a contrast ratio of 50. On the other hand, for a metal tube such as the 16GP4, with its much flatter face plate ($D = 16$ inches, $R = 40$ inches), the contrast ratio, as determined from this cause alone, is increased to 200.

Reflections from the bulb walls may have a similar effect on large-area contrast. Bulb reflections may be either specular or diffuse. If the bulb is so shaped that specular reflection from one part of the screen reaches other screen areas, the loss of contrast can be quite great. The bulb shown in Fig. 11.24*a* gives this type of reflection. If the conical part of the bulb is curved as shown in Fig. 11.24*b*, specular reflection cannot reach the screen. In the bulb in Fig. 11.24*a*, which does not have this curvature, the texture of the blackening material used on the walls is very important and even a surface which is only moderately glossy will greatly reduce contrast. Thus, it has been found empirically that the large-area contrast in metal tubes similar

to the 16AP4 may be as much as doubled by appropriate treatment of the inner wall of the metal cone.

Finally, scattered electrons may be a source of loss of contrast. They may arise from the gun, in the form of secondary emission from gun aperture edges and cold emission from gun electrodes, or, in the presence of an appreciable potential gradient along the screen, from secondary emission at the point of incidence of the beam on the screen. Modern gun design and construction effectively suppress the first source; the second source is significant only in projection kinescopes operating at high voltages and current densities, since only here large potential gradients may be established along the screen.

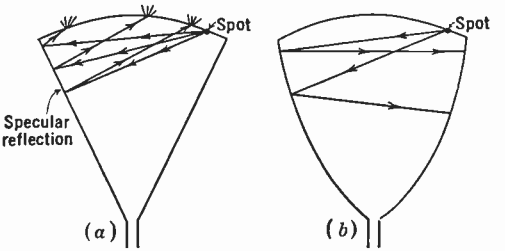


Fig. 11.24. Effect of Wall Shape on Bulb Reflections.

It is readily seen that all three of the sources of loss of overall contrast which have just been enumerated can be rendered negligible by the deposition of a metallic film over the fluorescent screen. This film suppresses at the same time all light emitted in a backward direction and places the fluorescent screen at a uniform potential. As indicated in Fig. 11.25, the light which, with the uncoated screen, gives rise to bulb-wall reflections and illumination of remote areas of the screen, is now redirected to enhance the picture brightness. Reflections in the face plate remain the one important contrast-limiting factor.

The great increase in overall contrast achieved by aluminizing the screen has been amply verified by measurement. For example, in a 5TP4 projection tube the overall contrast ratio was found to be increased (by G. M. Daly) from 130 to 500. Comparable increases were measured for larger, direct-view kinescopes.

However, all the above measurements were carried out in the absence of ambient illumination. This condition is fulfilled, in practice, for projection kinescopes, but not for direct-view kinescopes. The effect of ambient illumination in raising the light level of the dark

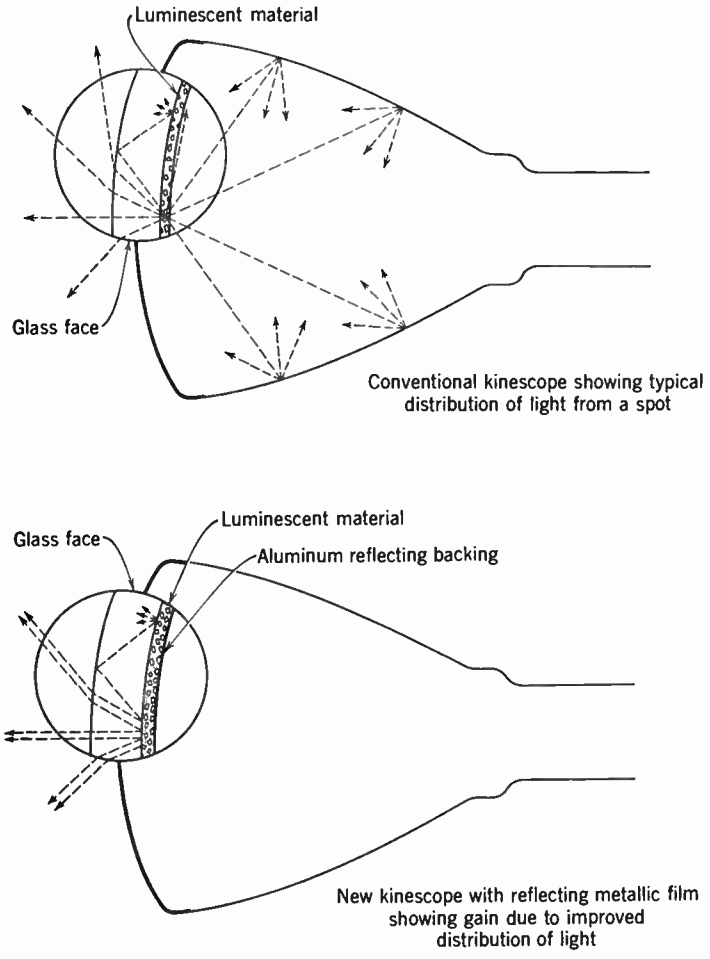


Fig. 11.25. Sources of Loss of Contrast in Kinescopes with Unaluminized and with Aluminized Screens. (Epstein and Pensak, reference 9.)

portions of the picture is enhanced by the reflecting backing along with the picture brightness. Thus with

B_{min}, B_{max}	dark area and highlight brightness of picture in absence of ambient illumination
E_0	ambient illumination
t	transmission of face plate
K	reflection coefficient of screen
K'	reflection coefficient of glass surface of face plate

the reciprocal contrast ratio becomes

$$S = \frac{B_{\min} + E_0(t^2K + K')}{B_{\max} + E_0(t^2K + K')}$$

Since K is generally of the order of 0.5, whereas K' is more nearly 0.04, reducing the transmission factor t of the face plate has a very

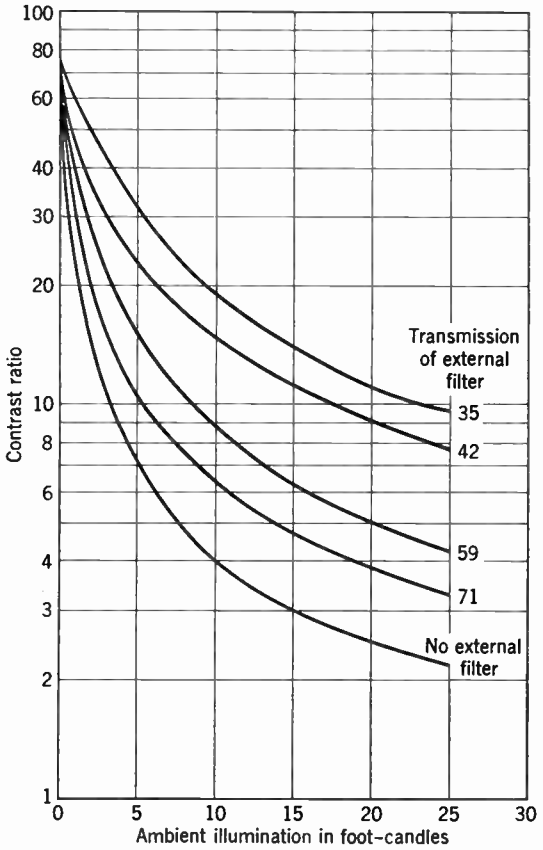


Fig. 11.26. Contrast Ratio as Function of Ambient Illumination for a 10-Inch Kinescope with 15.5 Foot-Lamberts Highlight Brightness. (L. S. Yaggy.)

material effect in diminishing the contrast-reducing action of ambient illumination. This is indicated by the measured curves of L. S. Yaggy, in Fig. 11.26, for a 10-inch kinescope with a highlight brightness of 15.5 foot-lamberts. In these measurements external filters of different absorption replaced the absorption within the face plate.

The effect of aluminizing on contrast in the presence of ambient illumination is shown, for a similar tube with clear face plate and at a higher highlight level, in Fig. 11.27. It is seen that, for the same

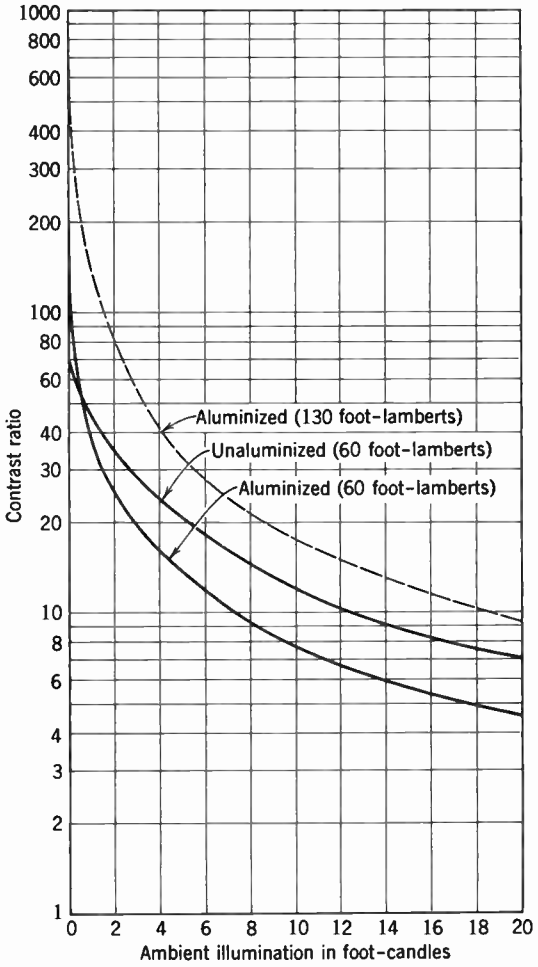


Fig. 11.27. Contrast Ratio as Function of Ambient Illumination for a 10-Inch Kinescope with Unaluminized and with Aluminized Screen on Clear Face Plate. (G. M. Daly.)

picture brightness, the contrast is less for the aluminized tube over most of the range of ambient illumination covered. The loss in contrast may be restored, however, if the greater light emission, of which the aluminized tube is capable, is fully utilized.

It may be noted that specular reflection at the front surface of the face plate may be objectionable quite apart from its effect on overall contrast; room reflections superpose extraneous detail on the picture which can be highly disturbing. This effect is removed in practice by lightly frosting the front surface of the face plate. With appropriate subsequent hydrofluoric acid treatment this frosting has only a slight effect in blurring the contours of the transmitted picture on the screen but greatly reduces external reflections.*

Summarizing, the employment of absorbing face plates (Filterglass) permits the attainment of detail contrast comparable with large-area contrast. Large-area contrast, with proper shaping or coating of the bulb and the relatively flat face plates currently employed, may be of the order of 100 in the absence of ambient illumination and can be made very much larger than this by aluminizing the screen. In the presence of ambient illumination contrast is greatly reduced and the beneficial effect of the aluminum film diminished. Absorption in the face plate, on the other hand, reduces the deleterious effect of ambient illumination on overall contrast. Specular reflections at the front surface, or even both surfaces, of the face plate, furthermore, can be rendered largely innocuous by lightly frosting the glass surface.

11.12 Direct-View Kinescopes. Modern direct-view kinescopes employ predominantly magnetic deflection and either magnetic or electrostatic focus. Ion-trap fields not only eliminate the deterioration of the screen by negative ions but, in addition, aid in the precise alignment of the electron beam. Large deflection angles, made possible by careful yoke design and high-efficiency deflection circuits, shorten the tubes.

The smaller kinescopes, face diameters 7 to 12 inches, are generally of all-glass construction and have circular faces. The 10BP4-A, widely employed as monitor kinescope, may serve as an example (Fig. 11.28). The tube, 17⅝ inches in length, employs a total deflection angle of 52 degrees. A double-field ion-trap magnet provides for the suppression of the negative ions. The spectral characteristic of the sulphide P4 phosphor employed in this as in the other direct-view kinescopes is shown in Fig. 11.29. If the full screen diameter is utilized, a rounded-end picture of 9 by 6¾ inches is obtained; as an alternative, the picture may be confined to a rectangle with rounded edges, 8 by 6 inches.

The following are typical operating characteristics of the 10BP4-A monitor kinescope:

* See e.g., Szegho, Amdursky, and Reed, reference 16.

Ultor (and grid No. 3) voltage	9000 volts
Grid No. 2 voltage	250 volts
Grid No. 1—cutoff voltage	−45 volts
Focusing coil current	115 ma
Ion-trap magnet current	155 ma

Fig. 11.28. The 10BP4-A Monitor Kinescope.

As indicated by the grid-drive characteristic of the 10BP4-A in Fig. 11.30, very high highlight levels can be achieved with this kinescope.

As has already been brought out, larger-sized kinescopes are manufactured with both glass and metal envelopes, and with round and rectangular face plates. Whereas the 21-inch rectangular glass kine-

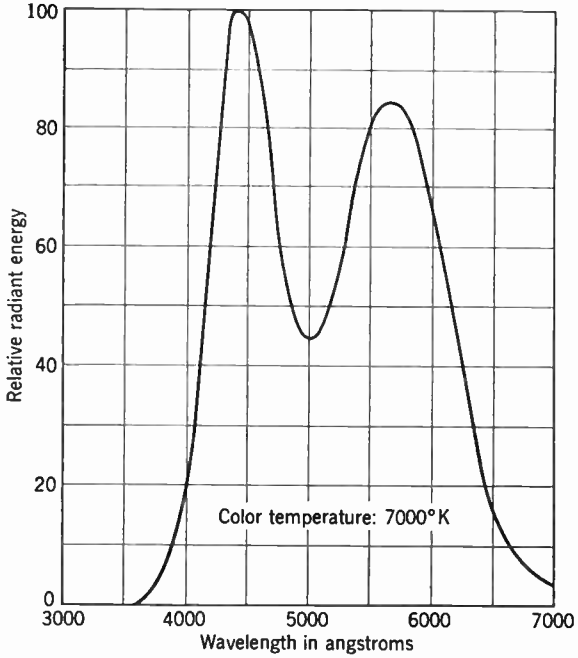


Fig. 11.29. Spectral Emission Characteristic of P4 Sulphide-Type Phosphor.

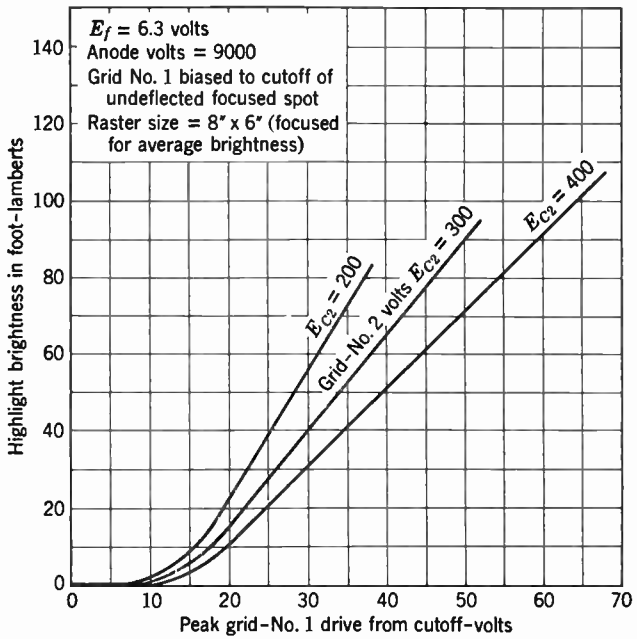


Fig. 11.30. Grid-Drive Characteristics of the 10BP4-A Kinescope.

Fig. 11.31. The 16GP4, a Round Metal Kinescope.

Fig. 11.32. The 17CP4, a Metal Kinescope with Rectangular Face Plate.

scopes play currently (1953) a leading role, the round metal 16GP4, the rectangular metal 17CP4 and 27MP4, and the rectangular glass 21EP4 (Figs. 11.31 to 11.33) may serve as examples of different designs. In the 16GP4, the 17CP4, and the 21EP4 the total deflection angle has been raised to 70 degrees, making possible a reduc-

Fig. 11.33. The 21EP4, a Glass Kinescope with Rectangular Face Plate.

tion in the overall length of the tubes to 17 11/16, 19, and 23 inches, respectively. In part, this reduction is achieved by reducing the distance between cathode and focusing lens: as the image distance is made smaller (by increasing the deflection angle for equal screen area), the object distance can be reduced in proportion without impairing the spot sharpness. Thus in the 16GP4 the gun structure of the longer 16AP4 is simply pushed forward into the focusing lens field and at the same time tilted to permit the use of a single-field ion-trap magnet near the cathode (Fig. 11.34). The electromagnetic focus shown here

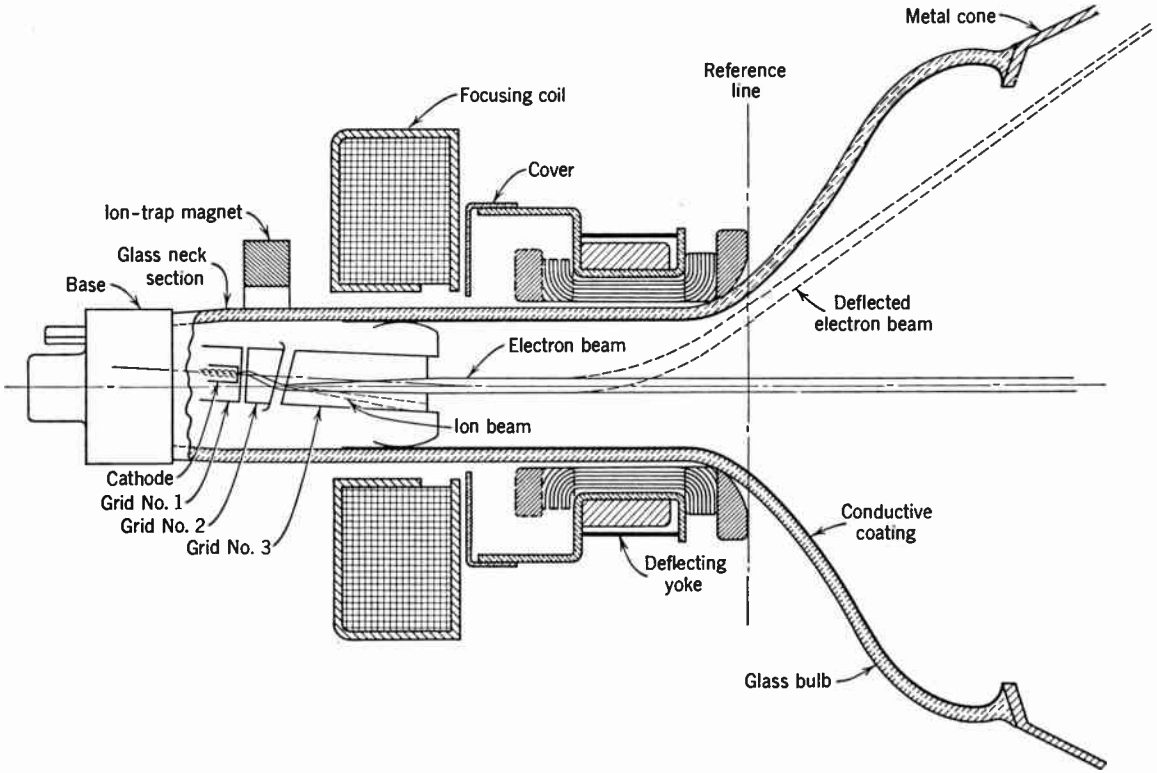


Fig. 11.34. Gun and Deflecting Coil Assembly for the 16GP4. (Swedlund and Steier, reference 7; courtesy of *Tele-Tech*.)

may be replaced by permanent magnet focus (Fig. 12.22) or, with minor changes in the gun, by electrostatic focus (Fig. 12.23).

A still further advance in the direction of greater total deflection angle and reduced length for given picture size is represented by the rectangular metal tube 27MP4, with 90-degree diagonal deflection

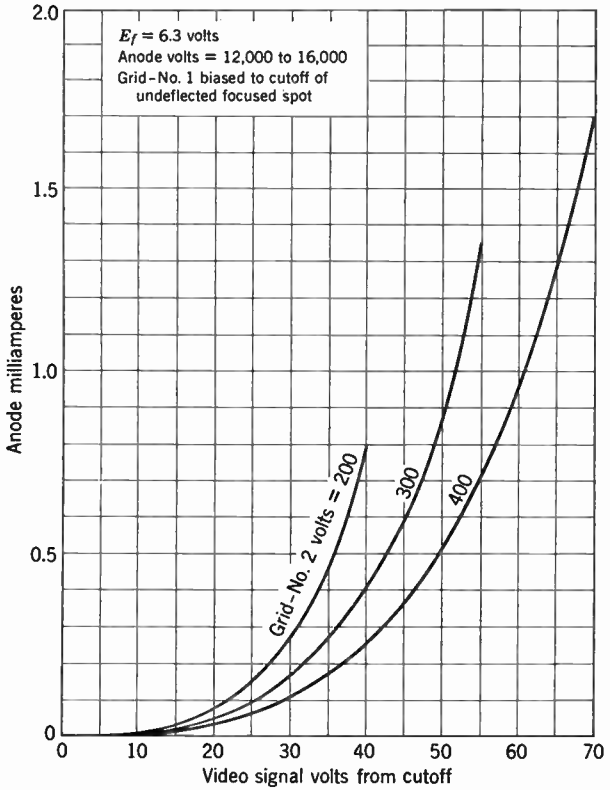


Fig. 11.35. Beam Current as Function of Grid Voltage (17CP4).

angle, a 23 7/16 by 18 1/8 inch picture and a tube length of 22 3/16 inches. The newer kinescopes also exemplify the trend toward flatter screens; in the 16GP4 the radius of the face plate is increased to 40 inches. This radius increase is made possible by a simultaneous slight increase in face-plate thickness to 7/32 inch.

Below are typical operating conditions for the 17CP4, which provides a 14 5/8 by 11 inch picture with rounded corners and slightly curved edges, and the 27MP4 giving a similarly shaped picture 23 7/16 by 18 1/8 inches in size:

	17CP4	27MP4
Ultor voltage	14,000	16,000 volts
Grid No. 2 voltage	300	400 volts
Grid No. 1 cutoff voltage	−33 to −77	−49 to −97 volts
Focusing coil current	104	110 ma
Field strength of single-field ion-trap magnet	50	50 gausses

The grid drive characteristics of the 17CP4, giving screen brightness and beam current as function of grid voltage, are given in Figs. 11.35 and 11.36. A comparison of the two sets of curves indicates the tendency of the sulphide phosphor to saturate at very high light levels.

It should be noted that, in direct-view kinescopes, the x-rays emitted at the screen quite generally do not present a health hazard as long

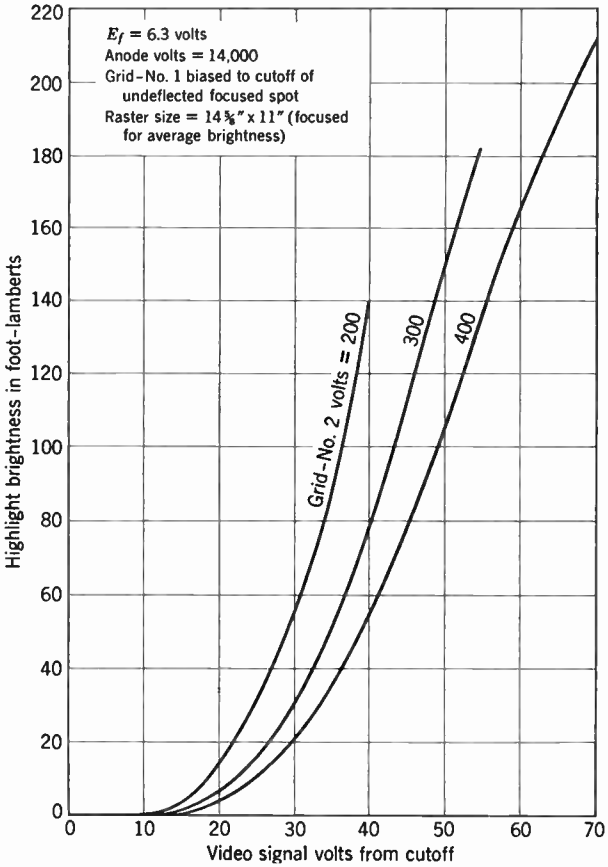


Fig. 11.36. Screen Brightness as Function of Grid Voltage (17CP4).

as the operating voltage is held below 16 kilovolts, since the face plate and the walls of the tube are sufficiently effective absorbers of the radiation in question. Normal precautions to prevent users and service personnel from coming in contact with high-voltage elements must, however, be taken; with metal kinescopes this precaution includes the protection of the metal cone and lip by an insulating sleeve and mounting ring of polyethylene or some similar insulating plastic. Finally, protection against flying fragments in case of an accidental implosion may be secured by mounting a flat pane of glass or plastic over the face plate. As has been pointed out, this risk is greatly reduced with metal tubes.

11.13 Projection Kinescopes. Projection kinescopes are operated at much higher voltages than direct-view kinescopes since their light output, for a relatively small fluorescent screen surface, must be sufficient to illuminate a much larger viewing screen without any diminution of picture sharpness. Their face plates are invariably curved, with a radius of curvature approximately half as great as that of the mirror of the reflective projection system utilized. In addition, they are commonly polished optically, since small flaws may, with the large-aperture optical systems employed, have a material influence on picture quality and contrast. The screen is necessarily aluminized. The aluminum film, which may be 5000 Å or more in thickness, not only maintains the screen uniformly at full anode potential but, in addition, renders an ion trap unnecessary.

Magnetic deflection and electrostatic focus are favored in projection kinescopes. The first renders unnecessary the introduction of high-voltage deflecting elements and decreases spot-defocusing problems. Electrostatic focusing, on the other hand, possesses the advantage of eliminating the optical obstruction of a focusing coil in a reflective projection system and simplifies the electrical supplies; the focusing voltage is derived from a voltage-regulated d-c power supply provided with a voltage control circuit. It may be noted that the 3NP4, which is designed to be used with a folded Schmidt projection system where a focusing coil does not interfere with the imaging, employs magnetic focus.

The phosphor of the projection kinescope is generally a mixture of yellow zinc-beryllium silicate and blue zinc sulphide. The layer is deposited in such a manner that the blue phosphor settles at the bottom, with the result that, owing to the spreading of the electron beam within the phosphor layer, the current density is smaller in the blue phosphor than in the yellow phosphor. Since the sulphide phosphor has a

greater tendency to saturate at high current densities than the silicate phosphor, color changes with increase in beam current are minimized by this procedure. As an alternative, an all-silicate phosphor mixture, in which zinc sulphide is replaced with calcium-magnesium silicate,

Fig. 11.37. The Philips 3NP4 Projection Tube. (Courtesy North American Philips Corporation.)

may be employed. The all-silicate phosphor completely avoids color change difficulties, but only at the expense of an appreciable loss in light output.

A further problem which does not arise in ordinary direct-view kinescopes is the discoloration of the glass under high-voltage electron bombardment, presumably arising from the separating out of neutral metal atoms within the glass. The effect varies greatly with the type of glass and is quite severe, e.g., for Pyrex. It requires the selection of special glasses (such as No. 707 tungsten sealing glass) for the face plate. With the proper choice of glass, the brown discoloration can be

largely, though not completely, removed by periodic baking with infrared lamps.

Commercial projection tubes range from 3 to 7 inches in face diameter. Three- and 5-inch tubes, such as the 3NP4 and 5TP4, are designed for home projection receivers, whereas the larger 7NP4 was developed specifically for theatre television projectors.

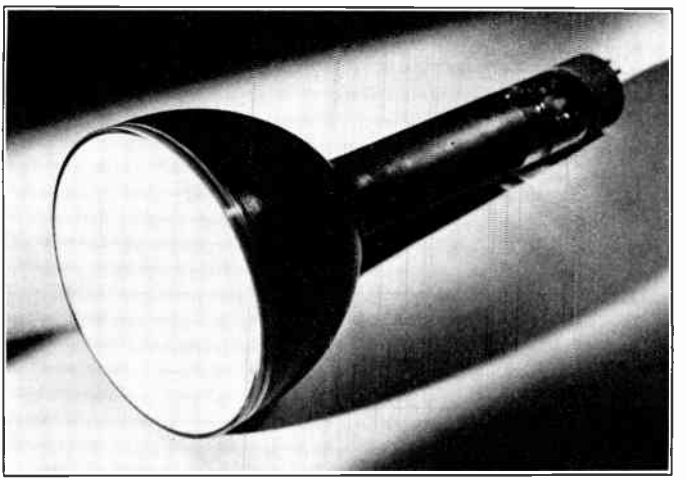


Fig. 11.38. The RCA 5TP4 Projection Kinescope.

The Philips 3NP4 (Fig. 11.37) and the RCA 5TP4 both operate at approximately 25 kilovolts. The former forms a picture 1.4 by 1.86 inches, the latter 3 by 4 inches. Since the highlight brightness of the two tubes is about the same (3000 and 4000 foot-lamberts, respectively), the light output of the 5TP4 is at least four times as great as that of the 3NP4. The overall tube lengths are only slightly different, 10.5 and 11.75 inches, respectively. The smaller picture size on the 3NP4 requires a smaller spot size (approximately 0.003 inch), and this, in turn, a relatively greater length of the electron gun.

Typical operating conditions of the 5TP4 (Fig. 11.38) are given below:

Anode No. 2 (Ultor) voltage	27,000 volts
Anode No. 1 voltage (for focusing 200-microampere beam)	4,320 to 5,400 volts
Grid No. 2 voltage	200 volts
Grid No. 1 cutoff	−42 to −98 volts
Anode No. 2 current	200 microamperes

Grid drive characteristics for the 5TP4 are shown in Fig. 11.39.

In the 3NP4 and the 5TP4, as well as the large projection tubes, a grounded external conducting coating is provided on the tube neck,

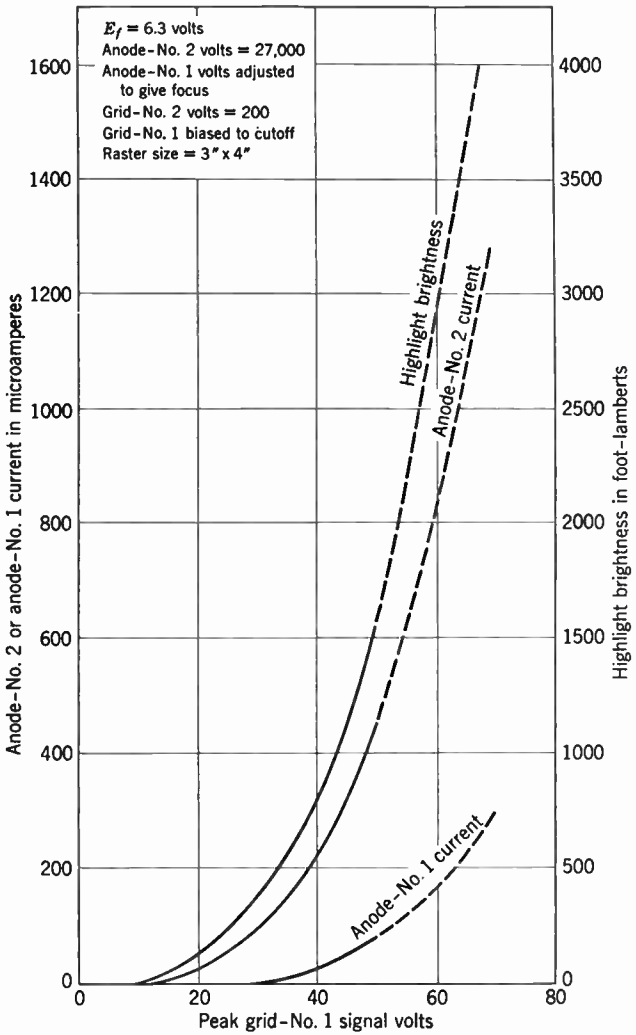


Fig. 11.39. Grid-Drive Characteristics of the 5TP4 Projection Kinescope.

protecting the deflection yoke, which is at ground potential, against high-voltage discharges.

Whereas the availability of direct-view kinescopes of large size has tended to minimize the always limited interest in projection kinescopes

for the home, these tubes fulfill a unique function in theatre television. Obviously, the demands on the light output of these tubes, which must illuminate a viewing screen 20 by 15 feet in size, are very great. They can be met only by simultaneously increasing the operating voltage and (in view of the tendency of phosphors to saturate under intense bombardment as well as thermal overload of the tube face) the screen area. At the same time, it is desirable to keep the size within reasonable limits so that the bulk and cost of the projection system may not be excessive.

Fig. 11.40. The 7NP4 Theatre Projection Kinescope.

Picture brightnesses comparable with the standards set by the Society of Motion Picture Engineers for first-class pictures (7 to 14 foot-lamberts) are obtained with the 7NP4 (Fig. 11.40) with a maximum operating voltage of 80 kilovolts; at 75 kilovolts highlight brightnesses of 30,000 foot-lamberts are obtained on the tube screen.

The second-anode voltage contact is visible near the screen; all other contacts are brought out through the plastic-filled base. The high-voltage first-anode pin is here isolated by leaving four adjoining pins on either side disconnected. The tube wall corrugations serve to increase leakage resistance; they and the adjoining tube neck flare are coated with an insulating moisture-repellent lacquer. The neck itself is coated externally with a conducting coating for the protection of the deflection coils. To obtain adequate insulation between this coating and the internal anode coating an inner cone-neck section is provided.

The disposition of the tube within the projection optics is shown in Fig. 11.41; the enlarged image is projected on a screen at a distance of 60 feet. A forced air flow of 40 cubic feet per minute prevents the tube face from overheating; this is necessary, since the average energy

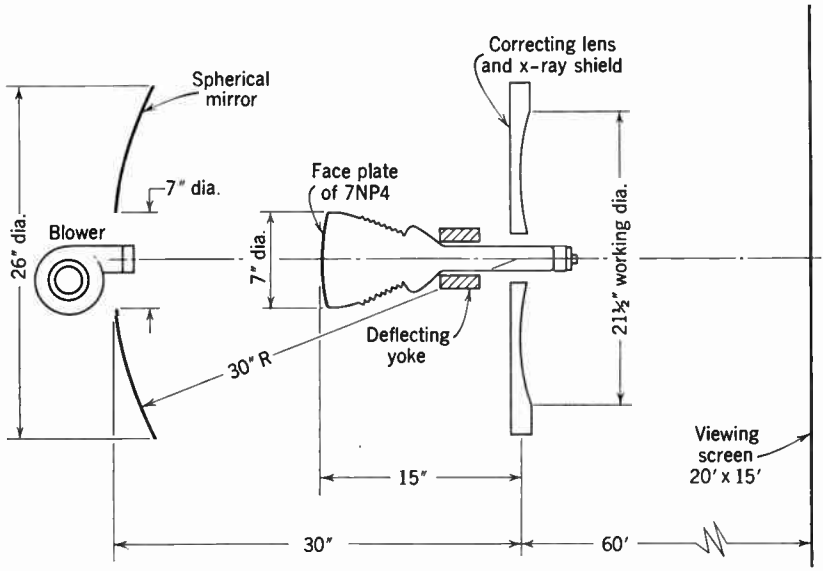


Fig. 11.41. Disposition of the 7NP4 within the Reflective Projection System.

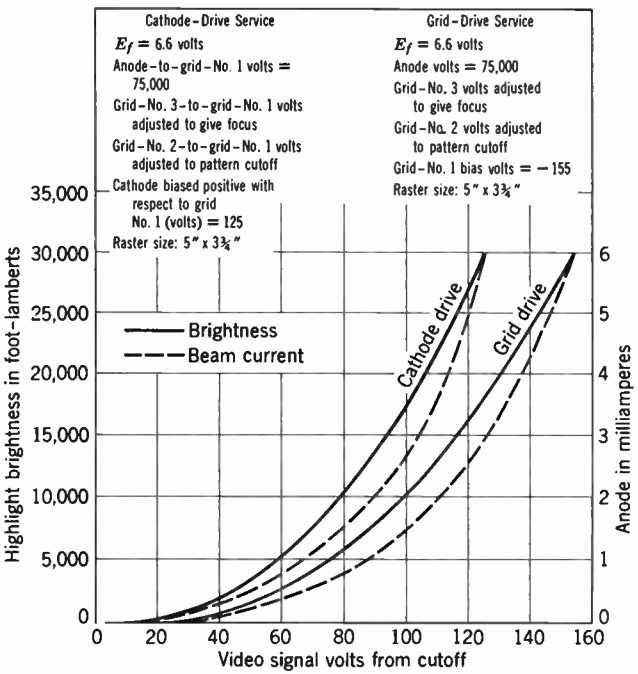


Fig. 11.42. Response to Video Signal of the 7NP4.

dissipated at the screen may be 150 watts. The air flow should be perpendicular to the tube face to prevent uneven heating and consequent cracking of the tube face.

The "quality rectangle" on the screen, which is normally made to coincide with the boundaries of the picture, is 5 by 3¾ inches in dimension. The variation of the picture brightness and beam current with the applied video voltage for this picture size are shown in Fig.

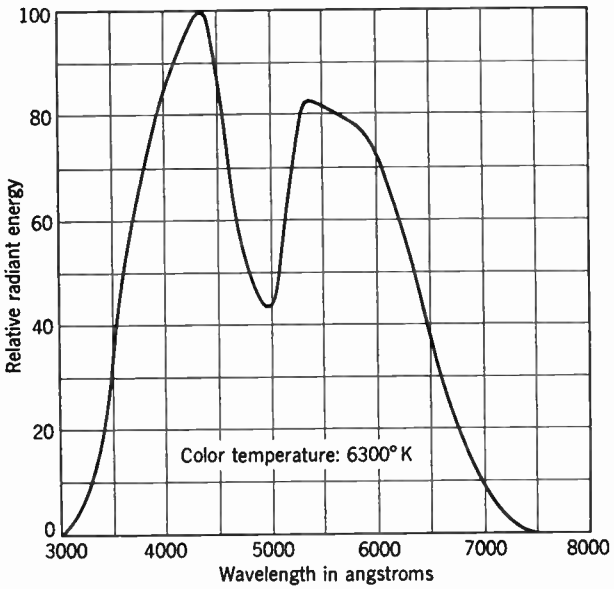


Fig. 11.43. Spectral Emission Characteristic of Mixed Sulphide-Silicate Screen of the 7NP4.

11.42. It is possible to apply the video voltage either to the grid, keeping the cathode at constant voltage, or the cathode, with the grid-voltage fixed. Cathode drive is to be preferred, since it reduces the signal strength required to pass from black to white. As the cathode is made more negative, the cathode-grid No. 2 voltage increases, so that the beam current increases more rapidly than with grid drive, where this voltage remains constant. Typical operating conditions of the 7NP4 with cathode drive are shown below:

Anode-to-grid No. 1 voltage	75.000 volts
Grid No. 3 to grid No. 1 voltage	16,000 to 18,000 volts
Grid No. 2 to grid No. 1 voltage	400 to 600 volts
Cathode-to-grid No. 1 voltage	
Black level	0 volt
White level	125 volts

Maximum grid No. 3 current	15 microamperes
Maximum grid No. 2 current	−15 to 15 microamperes

The spectral characteristic of the screen of the 7NP4, which, like that of the 5TP4, is a mixed sulphide-silicate screen, is shown in Fig. 11.43.

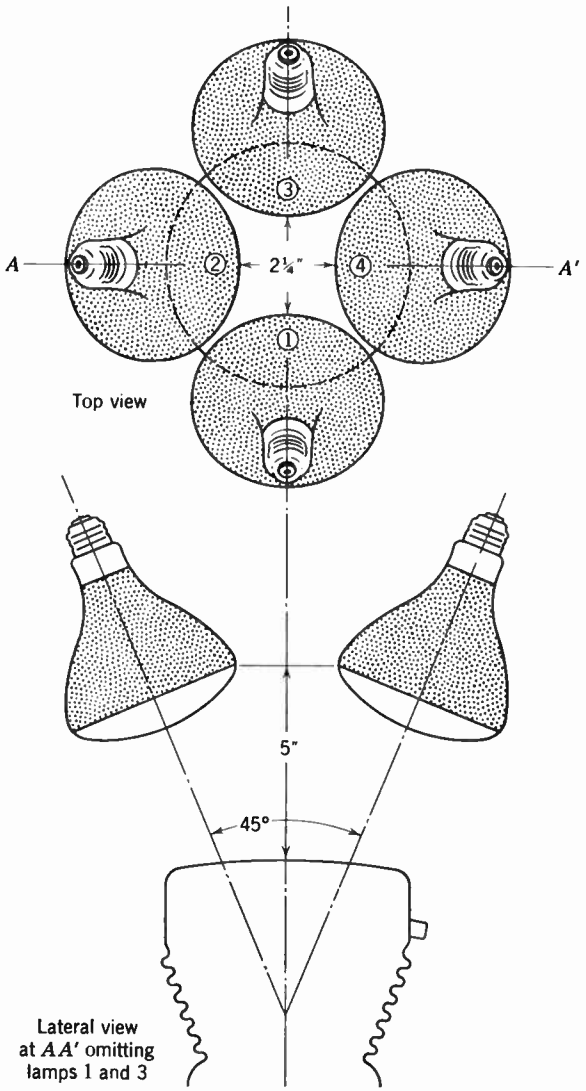


Fig. 11.44. Arrangement for Baking the 7NP4 to Remove Discoloration.

As with any high-voltage kinescope, the scanned area tends to discolor under bombardment. To reduce the discoloration, it is recommended that the tube face be exposed to the radiation of infrared lamps after each 150 hours of operation, as shown in Fig. 11.44. To equalize the radiation received by the screen from the four 250-watt reflector-type lamps, the tube is rotated through 90 degrees every 5 minutes, giving a total exposure of 20 minutes. Under this irradiation the tube face attains a temperature of approximately 200°C. After this treatment it is permitted to cool slowly.

REFERENCES

1. I. G. Maloff and D. W. Epstein, *Electron Optics in Television*, McGraw-Hill, New York, 1938.
2. H. W. Leverenz, *An Introduction to the Luminescence of Solids*, Wiley, New York, 1950.
3. V. K. Zworykin, "Description of an Experimental Television System and the Kinescope," *Proc. I.R.E.*, Vol. 21, pp. 1655–1673, 1933.
4. E. W. Engstrom, "A Study of Television Image Characteristics," *Proc. I.R.E.*, Vol. 21, pp. 1631–1651, 1933.
5. H. P. Steier, J. Kelar, C. T. Lattimer, and R. D. Faulkner, "Development of a Large Metal Kinescope for Television," *RCA Rev.*, Vol. 10, pp. 43–58, 1949.
6. A. S. Rose and J. C. Turnbull, "The Evaluation of Chromium-Iron Alloy for Metal Kinescope Cones," *RCA Rev.*, Vol. 10, pp. 593–599, 1949.
7. L. E. Swedlund and H. P. Steier, "Short 16-inch Metal-Cone Kinescope Development," *Tele-Tech*, Vol. 9, pp. 40–43, August, 1950.
8. H. P. Steier and R. D. Faulkner, "High-Speed Production of Metal Kinescopes," *Electronics*, Vol. 22, pp. 81–83, May, 1949.
9. D. W. Epstein and L. Pensak, "Improved Cathode-Ray Tubes with Metal-Backed Luminescent Screens," *RCA Rev.*, Vol. 7, pp. 5–10, 1946.
10. H. A. Bethe, "Theory of the Passage of Fast Corpuscular Rays through Matter," *Ann. Physik*, Vol. 5, pp. 325–400, 1930.
11. K. Gentner, "On the Energy Absorption of Fast Cathode Rays," *Ann. Physik*, Vol. 31, pp. 407–424, 1938.
12. H. M. Terrill, "Absorption of Cathode Rays in Aluminum," *Phys. Rev.*, Vol. 24, pp. 616–621, 1924.
13. W. Walkenhorst, "Optical Properties of Aluminum Films Evaporated in Vacuum," *Z. techn. Phys.*, Vol. 22, pp. 14–21, 1941.
14. C. E. Burnett, "A Circuit for Studying Kinescope Resolution," *Proc. I.R.E.*, Vol. 25, pp. 992–1011, 1937.
15. R. R. Law, "Contrast in Kinescopes," *Proc. I.R.E.*, Vol. 27, pp. 511–524, 1939.
16. C. Z. Szegho, M. E. Amdursky, and W. O. Reed, "Low-Reflection Picture Tubes," *Electronics*, Vol. 24, pp. 97–99, January, 1951.
17. J. de Gier, "Projection Television Receiver; II. The Cathode Ray Tube," *Philips Tech. Rev.*, Vol. 10, pp. 97–104, 1948.

18. R. V. Little, Jr., "Developments in Large-Screen Television," *J. Soc. Motion Picture Engrs.*, Vol. 51, pp. 37–46, 1948.
19. L. E. Swedlund and C. W. Thierfelder, "Projection Kinescope 7NP4 for Theatre Television," *J. Soc. Motion Picture Engrs.*, Vol. 56, pp. 332–342, 1951.
20. H. Moss, "Cathode Ray Tube Progress in the Past Decade," in *Advances in Electronics* II (L. Marton, Ed.), Academic Press, New York, 1950.

12 The Electron Gun

The concentration of electrons into a narrow beam is one of the very early phenomena observed in connection with cathode rays. Even before the nature of cathode rays was understood, it was known that in a gas-discharge tube the rays could be focused into a narrow pencil by properly shaping the disk from which the discharge originates. For example, this property was applied in the early gas x-ray tubes by Roentgen and others.

These early experiments also showed that such a beam could be directed onto a fluorescent screen, and deflected by means of an electrostatic or magnetic field. The first use of these effects was solely for the determination of the properties of cathode rays and of the electron. However, the recognition that the deflection of the focused spot could be used to measure the fields soon led to the development of the cathode-ray oscilloscope. Experiments by Rosing, which have already been described, used the gas-discharge oscilloscope as a means of reproducing television images. The last-mentioned field of application of the focused cathode-ray beam, namely, television, has gained an importance which is rapidly increasing. The requirements placed by this application on the means for producing the electron pencil, which is usually termed the electron gun, are very severe. In consequence, the design of the electron gun, for either a pickup tube or a viewing tube, has become a major problem.

The earliest cathode-ray tubes were gas-discharge tubes in which the low-pressure gases served not only to concentrate the beam but also to generate the electrons. The difficulty of controlling the current in the beam, and the fact that a beam focused in this way cannot be deflected at high speeds, made these tubes inadequate.

The first-mentioned difficulty was overcome by the introduction of a thermionic cathode to serve as the electron source. This made it possible to govern the current in the beam with a control grid. Such tubes were found applicable to the reproduction of low-definition television images. However, there still remained the difficulty incidental

to deflection which made them unsuited for high-definition pictures. The difficulty arises from the nature of gas focusing, and its dependence upon the presence of a positive-ion core at the center of the beam, which will not follow a rapid lateral motion of the beam. So fundamental did this obstacle prove to be that it became necessary to abandon the gas tubes in favor of hard cathode-ray tubes.

Without gas to aid the focusing of the beam, the design of the electron gun is much more difficult. The earliest hard cathode-ray tubes made use of an electrode system which produced a concentrating electrostatic field in the immediate neighborhood of the cathode; this field was intended to cause the electrons from the source to converge upon a point on the viewing screen. An example of this type of concentration system is to be found in the Wehnelt cylinder illustrated in Fig. 12.1. Such relatively simple electron-optical arrangements were satisfactory as long as the picture requirements did not demand a very small spot or very high current densities. However, when these two conditions became a necessity, a more complicated gun was required.

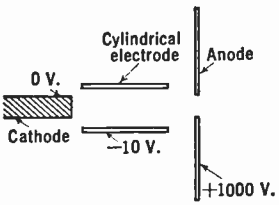


Fig. 12.1. Electron Gun with Wehnelt Cylinder.

12.1 Requirements of the Electron Gun. Before continuing with the discussion of the gun itself, the characteristics required of it should be reviewed. These characteristics differ, depending on its employment in high-velocity beam pickup tubes, low-velocity beam pickup tubes, direct-viewing tubes, or projection tubes.

The iconoscope, representing the high-velocity beam pickup tubes, has a gun which operates at relatively low power, but must be capable of forming an extremely small, well-defined spot. For this purpose the beam electrons are rarely required to have kinetic energies in excess of a thousand electron volts. Beam currents of 0.1 to 0.5 microampere are generally used, although certain special mosaics may require much higher currents. Although it is necessary that the beam current be controllable, the control characteristics are relatively unimportant. The spot must have a diameter not much greater than 0.01 centimeter for mosaics of the ordinary size; the meaning of the term "spot diameter" will be given later. It is important that the spot shape be regular and the spot well defined.

In the low-velocity beam pickup tubes—the orthicon, image orthicon, and Vidicon—the required beam current is of the same order as in the iconoscope, and the spot size less in proportion to the smaller target

areas of these tubes. In the Vidicon the spot diameter should not exceed approximately 0.002 centimeter. Apart from this, the gun requirements differ greatly from those of the iconoscope: in the low-velocity tubes the beam approaches the target with kinetic energy ranging from zero to a few electron volts; furthermore, the beam incidence must be perpendicular for all points of the target.

In the direct-viewing tubes both the currents and the voltages are considerably larger than in the iconoscope; the voltages range from 15 to 20 kilovolts, the currents from 100 to 500 microamperes. The spot size, on the other hand, may be much larger; with a 16-inch screen the spot diameter may be almost 1 millimeter.* More generally, the optimum spot size will be proportional to the picture dimensions.

The same rule as to spot size applies also for projection tubes; on the other hand, the power concentrated in the spot must here be much larger. This is achieved primarily by raising the operating voltage, which may range from 20 kilovolts for small home projection receivers to over 100 kilovolts for large-screen theatre projectors. Although the employment of such high voltages considerably increases the difficulties of insulation and suppression of cold emission, it facilitates the problem of concentrating large currents in small spots. The spot size must in general be smaller than for direct-viewing tubes since projection tube screens are, on the average, considerably smaller than direct-viewing tube screens. For a 5-inch projection tube the spot diameter should be about 0.03 centimeter.

Guns used in both types of viewing tubes must be designed in such a way that the current in the scanning beam can be readily controlled by the output from the video amplifier. Moderate voltage swing of the control element, low power requirements, and linearity of control are all desirable features.

The gun in the flying-spot tube must satisfy conditions similar to those which apply to the projection tube as far as power concentration in the spot is concerned. However, whereas the control characteristics of this gun are relatively indifferent, the spot can profitably be made smaller than in the projection tube.

12.2 The Two-Lens Electron Gun. With the exception of the guns in low-velocity pickup tubes, nearly all guns employed for television purposes are based on a two-lens principle. Such guns consist of a thermionic electron source, a first lens, almost invariably electrostatic, and a second lens, which may be either electrostatic or magnetic, or a

* Assuming 50 percent overlap, as described in Chapter 5, p. 204.

combination of the two. In practical viewing tube guns the action of the first lens is commonly supplemented by a prefocusing field which reduces the beam divergence in the region preceding the second lens.

The first lens is of the cathode-lens type described in Chapter 4, and serves to accelerate the electrons away from the cathode and to converge them into a bundle of narrow cross-section. This narrow portion of the ray bundle is not an image of the emitter, but corresponds more nearly to the exit pupil of the first lens system, and is usually referred to as the "crossover." Associated with the first lens is the control system. Various means of control have been successful, such as electrodes producing fields which restrict the emitting area of the cathode, elements causing the beam to be deflected or defocused at an aperture, and controls based on ordinary space-charge principles. Most control systems employed in practice owe their effectiveness to the restriction of the emitting area of the cathode in conjunction with the effects of space charge.

The second lens images the crossover formed by the first lens onto the target or fluorescent screen. With deflection, or defocusing, control the second lens may image the restricting aperture rather than the crossover when these do not coincide. In general, however, imaging the crossover itself leads to a maximum current for a given spot size. Where an electrostatic system is used for the second lens it may be arranged to have a voltage ratio greater than unity so that the beam electrons are accelerated by the lens. If the second lens is magnetic, on the other hand, the electrons receive their full acceleration at the first lens. The same applies for the electrostatic unipotential second lens commonly employed in iconoscope guns, where the suppression of secondary electrons ejected from the restricting aperture is highly desirable. Each of these lens systems has advantages and disadvantages which will become apparent as the discussion proceeds.

It may be noted parenthetically that, with very restricted cathode areas, a smaller image may be obtained on the screen by imaging the cathode on it in preference to the crossover. However, this condition is scarcely ever realized in television tubes.

12.3 Limiting Performance of an Electron Gun. The primary consideration in electron-gun design is to obtain a spot of small diameter and high current density. It is, therefore, important to investigate the factors upon which these two properties depend. The analysis of an actual gun is very complicated and, in general, cannot be rigorously

carried out. However, by postulating an idealized gun,* it is possible to determine the dependence of the current and spot size upon at least some of the parameters of the system.

The ideal gun will be assumed to consist of a cathode and a first and second lens system. The control grid, usually part of the first lens system, will be ignored entirely, no account will be taken of space charge, and, furthermore, both lens systems will be assumed to be free from aberrations.

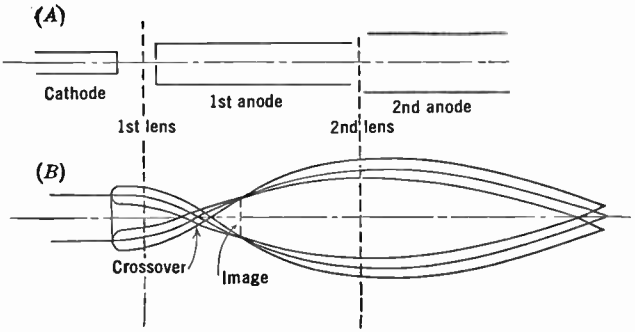


Fig. 12.2. Electrode Arrangement and Electron Ray Paths in Ideal Gun.

Figure 12.2*A* shows the general arrangement of the ideal gun; a schematic representation of the electron paths through the gun is shown in Fig. 12.2*B*. Electrons are emitted from every point of the cathode in all directions, as dictated by the distribution laws of thermionic emission. These electrons are accelerated toward, and through, the lens. The principal electron rays from all points of the cathode converge at the second focal point of the first lens. Other electrons, with radial initial velocities, are, of course, displaced a small radial distance from the axis at this point. Beyond this point the electron ray bundle from each point of the emitter may converge, forming a real image of the cathode, or the electrons of each bundle may diverge and the image of the cathode be virtual. In either case the diameter of the entire electron beam is a minimum at the point where the principal rays cross the axis. It has been noted that this point, which in the electron gun is designated as the crossover, is analogous to the exit pupil in conventional optics. The second lens, when adjusted to give a minimum spot, instead of using the image of the

* See also Langmuir, reference 14; and Law, reference 15.

cathode formed by the first lens as its object, images the crossover on the screen.

The electrons are emitted from the thermionic cathode with initial velocities which have a Maxwellian distribution. If the radial velocities are expressed in electron volts, the number of electrons having radial initial velocity components lying between $(2e\phi_r/m)^{1/2}$ and $(2e(\phi_r + d\phi_r)/m)^{1/2}$ will be

$$B_{\phi_r}\, d\phi_r = A\epsilon^{-e\phi_r/(kT)}\, d\phi_r \tag{12.1}$$

where A is a constant of emission, k is Boltzmann's constant, and T is the temperature of the cathode in degrees absolute. On this basis, the current I_V, consisting of electrons having radial velocities less than, or equal to, $(2eV/m)^{1/2}$ is given by

$$I_V = e\int_0^V B_{\phi_r}\, d\phi_r \tag{12.2}$$

For purposes of this analysis, it is convenient to assume that the first lens may be represented by a combination of a uniform field and

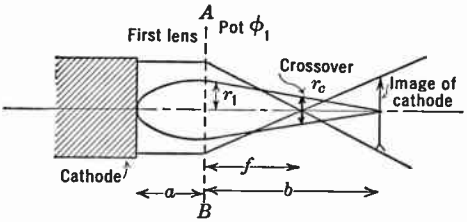


Fig. 12.3. First Lens of Ideal Gun.

a thin unipotential lens as indicated in Fig. 12.3. The electrons are accelerated by the uniform field up to the lens, which is indicated by the dotted line AB. The lens, which has the focal length f, bends the electrons toward the axis. A potential ϕ_1 will be assumed at the lens, and, furthermore, the region beyond the lens up to, and including, its second focal point, will be regarded as field-free.

An electron emitted from the cathode at a point on the axis with an initial radial velocity given by ϕ_r will follow a parabolic path until it reaches the lens AB. At the lens its radial distance from the axis becomes

$$r_1 = 2\sqrt{\phi_r/\phi_1}\, a \tag{12.3}$$

where a is the distance between the lens and the cathode. The lens will deflect the electron toward the image point corresponding to the emitting point. If the second focal length is f, the distance b between the image and the lens is given by the relation

$$\frac{1}{2a} + \frac{1}{b} = \frac{1}{f} \tag{12.4}$$

the effective object distance being twice the distance between the cathode and lens plane.

The radial distance r_c between the electron and the axis at the crossover (i.e., the second focal point) can be found as follows:

$$r_c = r_1 \frac{b - f}{b} = 2\sqrt{\frac{\phi_r}{\phi_1}}\, a \frac{b - f}{b} = \sqrt{\frac{\phi_r}{\phi_1}}\, f \tag{12.5}$$

It has been assumed that the electrons had their origin on the axis. However, the same result would have been obtained, for an aberration-free lens, for a point off the axis. For a real lens the circular cross-section of an off-axis pencil at the crossover is distorted into an ellipse, whose eccentricity increases with the angle α between the principal ray and the axis.

The radial initial velocities and the radius of the beam cross-section at the crossover are obviously related by Eq. 12.5. With the aid of Eqs. 12.1 and 12.2 the fraction of the total emitted current of electrons with radial velocity components no greater than $(2e\phi_r/m)^{1/2}$ can be determined and is given by

$$\frac{I_c}{I_T} = \frac{\displaystyle\int_0^{\phi_r} \epsilon^{-e\phi_r/(kT)}\, d\phi_r}{\displaystyle\int_0^{\infty} \epsilon^{-e\phi_r/(kT)}\, d\phi_r}$$

Evaluation of the integrals leads to:

$$\frac{I_c}{I_T} = 1 - \epsilon^{-e\phi_r/(kT)} \tag{12.6}$$

Substitution from Eq. 12.5 in Eq. 12.6 yields the following value for the current within the radius r_c at the crossover:

$$I_c = I_T\left(1 - \epsilon^{-\frac{r_c^2}{f^2}\frac{e\phi_1}{kT}}\right) \tag{12.7}$$

Finally, by differentiation with respect to r_c and division by $2\pi r_c$, the current density as function of r_c can be found. It is given by the following relation:

$$\rho(r_c) = \frac{I_T}{\pi f^2}\frac{e\phi_1}{kT}\epsilon^{-\frac{r_c^2}{f^2}\frac{e\phi_1}{kT}} \tag{12.8}$$

If, in this expression, we introduce ρ_0, the current density at the cathode, and α, the maximum angle of inclination of the principal rays at the crossover, in place of f, it may be written in the form

$$\rho(r_c) = \rho_0\alpha^2\frac{e\phi_1}{kT}\epsilon^{-\frac{\pi\rho_0}{I_T}\cdot\alpha^2 r_c^2\frac{e\phi_1}{kT}} \tag{12.8a}$$

It may be shown that this relation may be derived from thermodynamical considerations and hence possesses much more general validity than

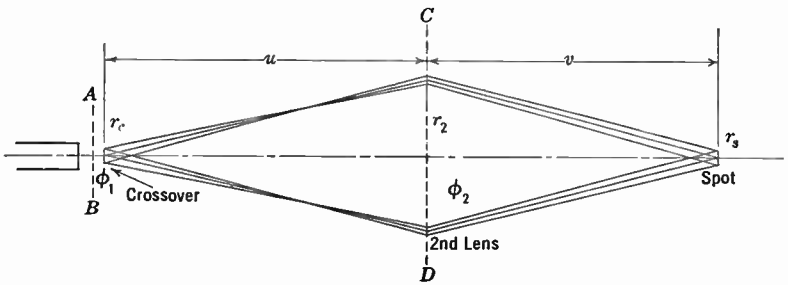


Fig. 12.4. Second Lens of Ideal Gun.

might be surmised from its derivation for a highly specialized model. *It represents the optimum current density attainable in an axially symmetrical beam with a beam convergence α and kinetic energy $e\phi_1$, which may be derived from a cathode at temperature T with an emission current density ρ_0 and a total emission current I_T, subject to the restriction $\alpha \ll 1$ and $e\phi_1 \gg kT$.*

Equation 12.8 readily leads to an expression for the distribution of current in the spot formed by the ideal gun. Since the second lens is adjusted to image the crossover onto the viewing screen, or target, any circle of radius r_c will be imaged into a circle of radius r_s at the spot. The second lens is shown schematically in Fig. 12.4 as a thin lens represented by the line CD at distances u and v from the crossover and screen, respectively. The potential in the image field of the second lens will be taken as ϕ_2. Therefore, the radii of the image and object will be related by

$$\frac{r_s}{r_c} = \frac{v}{u}\sqrt{\frac{\phi_1}{\phi_2}} \tag{12.9}$$

The fraction of the total emitted current within a radius r_s of the spot, and the current density distribution, can be found with the aid of Eq. 12.9 from Eq. 12.7. These relations are

$$\frac{I_s}{I_T} = 1 - \epsilon^{-\frac{r_s^2 e\phi_2}{f^2(v/u)^2 kT}} \tag{12.10}$$

and

$$(r_s) = \frac{I_T}{\pi f^2 (v/u)^2}\frac{e\phi_2}{kT}\epsilon^{-\frac{r_s^2 e\phi_2}{f^2(v/u)^2 kT}} \tag{12.11}$$

Equation 12.11 may again be written in the form of Eq. 12.8*a*, with ϕ_2 replacing ϕ_1.

Next, the diameter of the beam can be determined as it passes through the second lens. For the aberration-free lens which has been

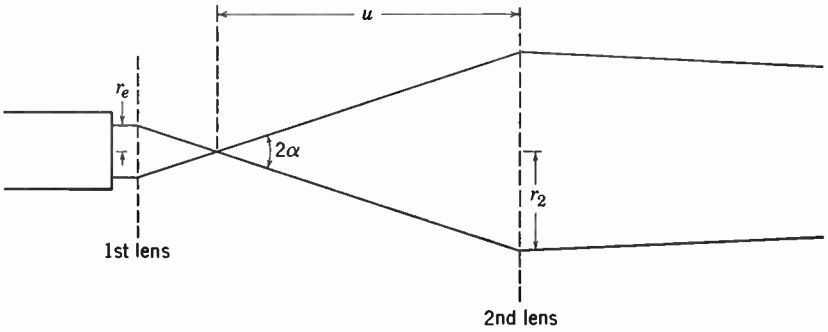


Fig. 12.5. Marginal Rays through Second Lens.

assumed, this is of no consequence, but for an actual lens in which spherical aberration plays an important role, the spread of the beam must be considered. Figure 12.5 again shows the idealized gun, with principal rays originating at the edge of the cathode, a distance r_e off the axis. These, after passing through the first lens, are deflected so as to meet the axis at the crossover. At this point they make an angle α with the axis:

$$\alpha \cong r_e/f \tag{12.12}$$

Since the space between the two lenses, which are both treated as thin, is at a constant potential ϕ_1, the electron rays are here straight. The

principal rays from the edge of the cathode, which form the boundary of the electron beam in the second lens, hence traverse the latter at a distance from the axis

$$r_2 = u\alpha = r_e u/f \tag{12.13}$$

Certain important conclusions can be drawn from Eqs. 12.7, 12.8, 12.10, 12.11, and 12.13 concerning the operating characteristics of the gun. Because of the simplifying assumptions made as to the nature of the first and second lens, caution is necessary in applying these deductions to an actual gun.

Equation 12.11, which describes the current distribution in the spot, indicates that the spot has no sharp boundary but, instead, the current density decreases exponentially with the square of the distance from the center. Therefore, the term "spot size" must be arbitrarily defined. One very practical definition, which has operational significance, is that which limits the radius of the spot to a value for which the current density is reduced to a certain fraction of the density at the center, or, symbolically:

$$K = \frac{\rho(\beta)}{\rho(0)} = \epsilon^{-\frac{\beta^2}{f^2(v/u)^2}\frac{e\phi_2}{kT}} \tag{12.14}$$

where β is the radius of the spot and K the limiting fraction of current density. The actual value of K is unimportant for purposes of the present discussion. In practice it depends on the use to which the gun is to be put.

Equation 12.14 can be rewritten as

$$\beta = f\left(\frac{v}{u}\right)\sqrt{-\frac{kT}{e\phi_2}\ln K} \tag{12.15}$$

It is evident from Eq. 12.15 that the "spot size" is independent of the first lens voltage ϕ_1, since the focal length f does not depend on it and, for given f, the ratio v/u is a function only of the position of the second lens. *The spot size is, however, inversely proportional to the square root of the total voltage* ϕ_2.

Equation 12.10, together with Eq. 12.14, can be used to determine the current in the spot, as follows:

$$I_s = I_T(1 - \epsilon^{-\frac{\beta^2 e\phi_2}{f^2(v/u)^2 kT}}) = I_T(1 - K) \tag{12.16}$$

In other words, practically all the current emitted by the cathode is concentrated into the spot. This result is to be expected since, as yet, no account has been taken of possible limiting apertures.

The radius of the beam at the crossover (defined in the same terms as the spot size) varies inversely with the square root of the first-lens voltage. As has been shown, in the absence of limiting apertures the size of the spot is independent of the crossover diameter. However, this does not mean that the first-lens voltage is unimportant. In the first place it should be high enough to saturate the emission from the cathode when the maximum beam current is required. Furthermore, it must be large enough to minimize the defocusing action of the space charge in the neighborhood of the cathode. On the other hand, a high first-lens voltage tends to make the control of the electrons more difficult.

In the idealized gun postulated, the first-lens voltage has no effect on the diameter of the beam through the second lens. If a potential gradient had been assumed between the first and second lens the diameter of the ray bundle at the second lens would have been found to vary with the first-lens voltage.

The radius of the ray bundle through the lens is also directly proportional to the radius of the emitting area. When the current control for the beam is obtained by restricting the emitting area of the cathode, the size of the beam at the second lens varies with beam current. Under these conditions, the presence of spherical aberration in this lens will cause the spot size, also, to vary with beam current.

Because of the aberration inherent in all electron lenses, the spot size in a real gun will be considerably greater than that of the ideal gun. In order to minimize the effect of these aberrations, it is common practice to distribute limiting apertures throughout the system.

The effect of restricting the radius of the emitter leads to no first-order change in the spot size. However, for a given specific emissivity of the cathode material, the beam current is directly proportional to the area of the cathode, i.e., to the square of the radius. The diameter of the beam through both the first and the second lenses is approximately proportional to the diameter of the cathode. Because of this, for reasons connected with the aberrations of these lenses, in particular of the first lens, a very large cathode area cannot be used without increasing the spot size. By curving the surface of the emitter some of these defects can be minimized, thus permitting a larger cathode area and, consequently, a higher beam current.

An aperture at the crossover has many consequences. The spot size under these conditions is no longer independent of the first-lens potential but is given by the expression

$$r_s = \beta = r_a \frac{v}{u} \sqrt{\frac{\phi_1}{\phi_2}} \tag{12.17}$$

where r_a is the radius of the aperture, ϕ_1 is the first-anode voltage, and ϕ_2 is the second-anode voltage. The current passing through the aperture can be expressed by the integral

$$I_b = \int_0^{r_a} 2\pi r \rho_c(r)\, dr = I_T \int_0^{r_a} \frac{r}{f^2} \frac{2e\phi_1}{kT} \epsilon^{-\frac{r^2}{f^2}\frac{e\phi_1}{kT}}\, dr$$

$$= I_T\left(1 - \epsilon^{-\frac{r_a^2}{f^2}\frac{e\phi_1}{kT}}\right) \tag{12.18}$$

These two relations indicate that the spot size can be decreased by decreasing the radius of the crossover aperture or, for fixed crossover aperture, by reducing the first-anode voltage. However, this reduction can be accomplished only at the expense of beam current. It should be noticed that the two methods of reducing the spot size cause an equal decrease in beam current if the total emission current I_T is constant. Both have little effect on the diameter of the beam at the second lens. In practical guns this conclusion is materially modified by space charge and aberration effects.

When the second lens is provided with an aperture to limit the beam diameter, its effect on the aberration-free system is practically the same as limiting the area of the emitter. The difference in action is due to the fact that the diameter of the beam at this lens is not exclusively determined by the marginal principal rays, as was heretofore assumed, but is also a function of the crossover diameter. In a practical gun, an aperture placed at, or near, this point to reduce the effect of the spherical aberration of the lens limits the spread of the beam resulting from the aberrations in the first lens, as well as that determined by the first-order paths of the rays. It is thus very much more effective for the purpose than limiting the cathode area (although the cathode is usually restricted as well to avoid excessive power dissipation at the aperture limiting the second lens).

12.4 The Cathode. The design of the thermionic cathode and the choice of emitting material for the practical gun are of fundamental importance. In the preceding section it was pointed out that it is generally necessary to restrict the emitting area, and, consequently, the

current which can be obtained in a spot of given size is, except for space-charge considerations, directly proportional to the emissivity of the material. Furthermore, it was shown that the radial initial velocities play an important role in determining the distribution of current at the crossover, and consequently at the spot. Another consequence of high initial velocities, which was not mentioned in the discussion of the ideal gun, is that they increase the spot size as a result of the chromatic aberration of the first lens. In order to minimize these undesirable effects, the cathode should be operated at as low a temperature as possible, as is evident from the form of the thermionic emission equations.

The velocity distribution of the electrons from the various cathodes can be determined from the Maxwell-Boltzmann distribution law. According to Eq. 12.1, the relation giving the number of electrons emitted with radial velocity components between $(2e\phi_r/m)^{1/2}$ and $(2e(\phi_r + d\phi_r)/m)^{1/2}$ is

$$B\, d\phi_r = A\epsilon^{-e\phi_r/(kT)}\, d\phi_r$$

Similarly, the axial distribution is given by

$$D\, d\phi_a = E\epsilon^{-e\phi_a/(kT)}\, d\phi_a$$

These two initial velocity distributions are important in determining the crossover radius and chromatic aberration of the first lens.

Cesium-activated cathodes would be highly desirable from the standpoint of high emissivity and low operating temperatures. However, such cathodes are, in general, much too delicate for this purpose and are deactivated almost immediately by ion bombardment.

Both thoriated tungsten and cathodes coated with barium and strontium oxide are suitable for use in electron guns. Because of their high specific emission and lower operating temperature, the oxide-coated cathodes are preferable and are used almost universally in sealed-off tubes. Thoriated tungsten cathodes are often advantageous in demountable systems because this type of emitter can be readily reactivated. Where more rugged cathodes are required, pure tungsten or tantalum may be employed.

The conventional oxide-coated cathode assembly consists of a heating element, a cathode cup, and the emitter. Indirect heating of the cathode has the advantage of minimizing the interaction between the magnetic and electrostatic fields produced by the heater current and the beam electrons. Typical forms of oxide-coated emitters are illustrated in Fig. 12.6. The heater is usually a tungsten filament in the

form of a noninductively wound helix covered with a refractory material to insulate it, and fitted inside a nickel cup. The nickel cup serves as a base for the cathode material. The cathode area of the nickel cup is coated with a mixture of barium and strontium car-

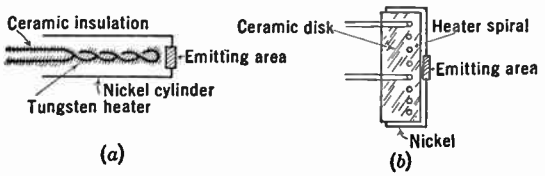


Fig. 12.6. Typical Indirectly Heated Cathodes.

bonate. This material can be decomposed by heating and activated as described in Chapter 1.

A typical kinescope cathode assembly is shown in Fig. 12.7. The heater is a tungsten twist designed to operate at 0.6 ampere and 6.3

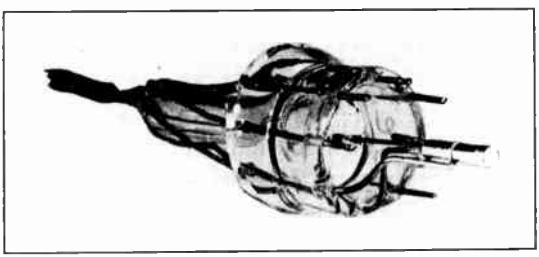


Fig. 12.7. Photograph of Cathode Assembly.

volts. This double spiral is insulated from itself and from the cathode cup with a refractory clay. A cylinder closed at one end with a disk forms the cathode cup. The heater is inserted in the cathode cup, which it heats partly by radiation and partly by conduction. The

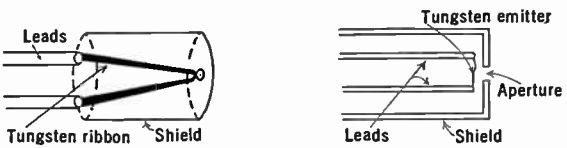


Fig. 12.8. Two Forms of Demountable Cathode.

terminal disk coated with emitting material forms the cathode surface. Curving the cathode is sometimes advantageous to minimize the aberrations in the first lens.

The assembly of a thoriated tungsten cathode depends on the kind of demountable vacuum tube for which it is intended. Usually, the cathode is heated directly and is in the form of a thoriated tungsten ribbon or wire. In order to restrict the emitting area, the cathode is covered with an apertured shield. Two typical cathode assemblies of this kind are shown in Fig. 12.8.

12.5 First Lens and Control Grid. The simplest first-lens system is composed of three elements: the cathode, the control grid, and the first anode or, more generally, the second grid. The arrangement is illustrated in Fig. 12.9. The control grid consists of a short cylinder closed by an apertured disk close to the end of the cylinder. In the example

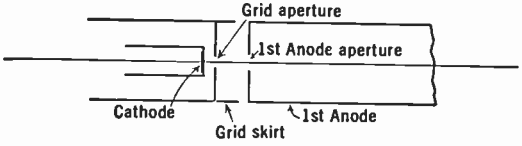


Fig. 12.9. Simple Three-Element First Lens.

the disk is placed so as to allow a portion of the cylinder known as the grid skirt to extend beyond it in order to achieve a desired control characteristic. The first anode or second grid is also a cylinder, of the same diameter as the grid, and is closed at the grid end by an apertured disk.

The first-lens region in the electron gun is, from an analytic point of view, by far the most difficult part of the whole electron-optical system. The electrostatic field in this region is so complicated that it is completely unamenable to mathematical manipulation. Furthermore, the control grid, which forms a part of the first lens, varies in potential, altering the optical properties of the system. Finally, the electron densities in this region are sufficiently great so that the space charge appreciably alters the field. In spite of these difficulties, this region has been sufficiently analyzed so that the behavior of the electrons over this portion of their trajectories is fairly well understood.

Figure 12.10 shows the equipotential surfaces in the neighborhood of the cathode in a typical three-element first-lens system for two values of grid potential, as measured by Maloff and Epstein.*

These field maps show that the potential of the control grid acts in three different ways. At, and near, the center of the cathode it controls the current in the beam by forming a space-charge barrier in the

* See Maloff and Epstein, reference 11.

manner of a conventional control grid. Furthermore, the control grid acts to restrict the area from which emission can take place. Thus, even if the emission is saturated at the center of the cathode, the current in the beam can be increased by decreasing the negative bias applied to the grid, because this will permit electrons to leave from a greater area of cathode. The final consequence of varying the control grid voltage is to alter the position of the cardinal points of the lens system. As the grid is made more negative, the second focal point and, consequently, the crossover move closer to the cathode.

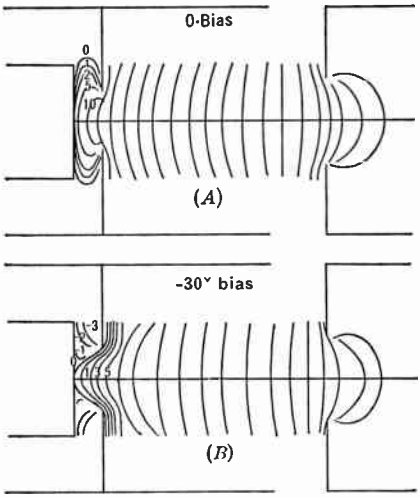


Fig. 12.10. Equipotentials in Three-Element First Lens. (After Maloff and Epstein.)

The cardinal points of the first lens have been determined on the basis of potential maps. The general arrangement of these points is shown in Fig. 12.11. Their position depends, of course, upon the velocity with which the electrons are emitted. Calculations show that, in the particular configuration with the bias at -30 volts, the second principal point and focal point are rather insensitive to the initial velocities, whereas the first focal point and principal point move from the surface of the

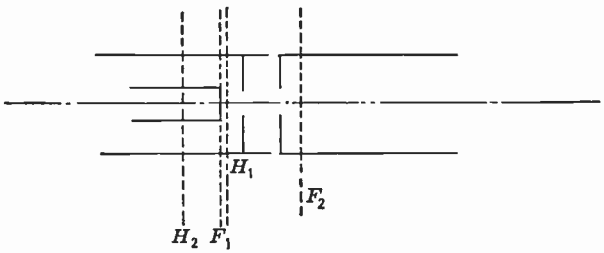


Fig. 12.11. Cardinal Points of First Lens. (After Maloff and Epstein.)

cathode to a distance of a few thousandths of a centimeter in front of the cathode as the energy of emission goes from 0 to 0.5 electron volt. It is also found that at zero volt bias the shift of the cardinal

points, corresponding to chromatic aberration in the first lens, is much less.

The equipotential plot in Fig. 12.10, determined by means of an electrolytic plotting tank, represents the field configuration in the absence of space charge. This approximation is not sufficiently good for purposes of determining the actual performance. In order to estimate * the effect of space charge, the configuration and charge density of the beam in the absence of space charge are calculated, and from these the new potential is determined. To attain a better approximation this procedure may be repeated. Figure 12.12 shows the envelope and density distribution of the beam; Fig. 12.13 illustrates the potential distribution corrected for the effect of space charge.

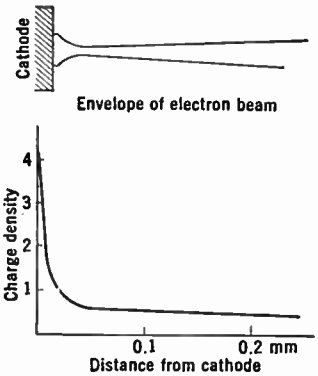


Fig. 12.12. Envelope and Electron Density of Beam in First Lens. (After Maloff and Epstein.)

Three important conditions must be satisfied by the ideal first-lens system: (1) the configuration must be such as to permit the flow of high current without too large a crossover (i.e., the product of second-grid voltage and crossover area must be small); (2) the control grid characteristics must be such as to permit ease of

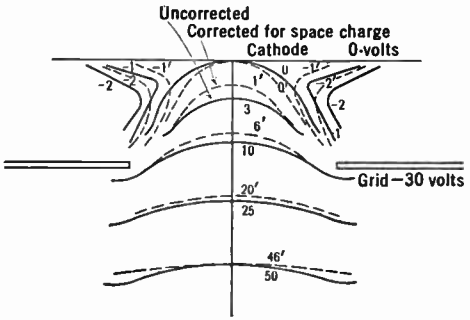


Fig. 12.13. The Effect of Space Charge in the First-Lens Region. (Maloff and Epstein, reference 11.)

control; and finally, (3) the effect of control voltage on spot size must be a minimum. No first-lens design has been worked out which satis-

* See Maloff and Epstein, reference 11.

fies completely all these conditions. There are several systems, however, which are very good from a practical standpoint.

The system which was described above can, by suitable selection of the size of the grid aperture and the length of the grid skirt, be made to have excellent control characteristics and to deliver considerable current into a small spot. However, the spot size depends on the grid bias, which is very undesirable.

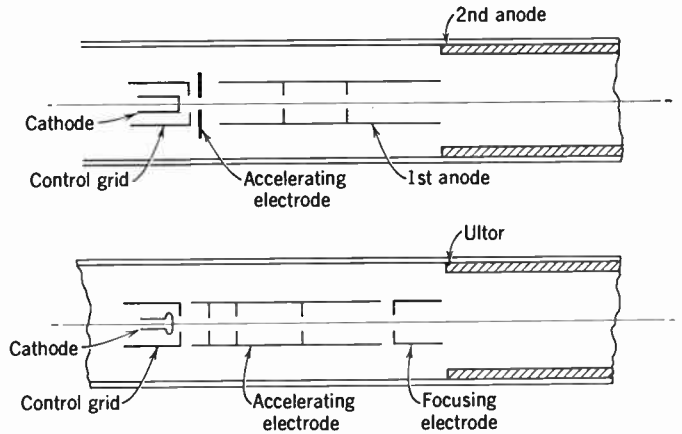


Fig. 12.14. Lens Systems Employing Additional Electrodes.

An accelerating grid, or other element, in the first lens is often used to make the control characteristic of the gun independent of second-lens focus and, incidentally, to make the condition of optimum spot focus independent of the grid bias. This is shown in the first example in Fig. 12.14. In modern kinescope guns (e.g., Figs. 12.19, 12.20, 12.22, and 12.23) the accelerating grid, denoted by G_2, generally takes the form of a cylindrical cup. The relatively strong converging lens formed at its junction with the first anode or focusing electrode (G_3), whose voltage may exceed that of the accelerating grid by a factor of 100, then collimates the beam diverging from the crossover. This collimation minimizes current loss at succeeding gun apertures and reduces the beam cross section at the second lens.

In the second example shown in Fig. 12.14 the first anode is replaced by an accelerating electrode which may be maintained at second-anode or ultor voltage; here the focusing electrode is the central element of a unipotential second lens, an arrangement favored in the iconoscope gun. Here, where a small, clean spot is essential, it is also customary to employ a physical crossover aperture. This aperture

may be either circular or rectangular, with the narrow dimension in the direction of horizontal scanning.

Finally, in kinescopes whose fluorescent screen is not protected by a metal film, the first-lens structure is commonly modified by the presence of an ion trap, designed to keep negative ions formed in the cathode region from bombarding the screen and reducing its efficiency. As shown, for example, in Fig. 12.20, the ion trap consists of the transverse magnetic field of a small permanent magnet, whose deflecting action on the electron beam is compensated by the tilt of the electrostatic lens between the accelerating grid G_2 and the first anode G_3. The heavy ions strike the wall of the gun since they are practically unaffected by the magnetic field, but are deflected by the electrostatic field of the tilted lens just like the electrons.

12.6 The Second Lens. The second-lens system images the crossover into the spot on the viewing screen or the pickup tube target. In general, this lens is relatively simple, and the first-order-image properties can be readily calculated by the methods given in Chapter 4.

In an electrostatic gun of the acceleration type, the second lens is based essentially on the lens formed between two coaxial cylinders at different potentials. Various departures from the simple cylinder lens have been studied, and some are in use in practical tubes. These modifications are dictated by the following considerations: (*a*) reduction of spherical aberration; (*b*) ability to withstand high voltages (particularly in viewing tubes); (*c*) ease in construction and alignment; and (*d*) circuit considerations.

For electrostatically focused viewing tubes and pickup tubes, the diameter of the glass neck of the tube envelope determines the maximum diameter that the electrodes of the second lens can have. Under these circumstances, it can be shown that the minimum spherical aberration is obtained when the lens is in the form of a pair of cylinders having this maximum diameter. Any shaping of the lens elements can be effected only by reducing the diameter of the lens, and this reduction in diameter more than offsets the improvement in spherical aberration resulting from the shaping. The same thing is true of making one cylinder different in diameter from the other.

The optical properties of the equidiameter cylinder lens have been discussed in Chapter 4 and are represented graphically in Fig. 4.17. The four curves shown there suffice to determine completely the first-order imaging properties of the lens, so that for any preassigned values of crossover to lens and lens to screen distance the voltage ratio required to form a sharp Gaussian image can be obtained.

In determining the position of the second lens, there are a number of considerations. They include practical features of tube design (e.g., the position of the deflection yoke, the maximum permissible tube length, etc.), the total current required, and the amount of space charge that can be tolerated in the first lens. The analysis given below applies to the simple two-lens gun. The effect of collimation by the lens between the accelerating grid cup and the first anode ("prefocus") will be noted briefly at the end.

As the position of the second lens is changed, the current density at the center of the spot remains nearly constant because the distance between the second lens and screen is large compared to that between lens and crossover, and a constant fraction of the second lens, as determined by the limiting aperture, is filled by the electron stream. This follows directly from Eq. 12.11, transformed to the form of Eq. 12.8*a*. However, both the spot size and the total current in the spot increase with decreasing distance between the second lens and crossover. The current and spot size for a given first-lens system is, therefore, one of the factors determining second-lens position.

Also, as the lens is moved towards the crossover, it is necessary to increase the strength of the second lens. Since usually the overall voltage is fixed by the operating requirements, and the diameter of the lens electrodes cannot be reduced because of spherical aberration considerations, this increase in lens strength means reducing the first-lens voltage. Because of space-charge effects, etc., a low first-lens voltage is not desirable. Therefore, even when a crossover aperture is employed to limit the size of the spot, the second lens should not be too close to the crossover. In practice a voltage ratio ϕ_1/ϕ_2 in the neighborhood of 1/4 to 1/5 is found to yield satisfactory results.

The spherical aberration of this type of lens can be computed from the potential distribution. The separation Δr_s from the axis at the Gaussian image plane of a ray which originates from an axial object point and passes through the aperture of the lens at a distance r_a from the axis is given by

$$\Delta r_s = S r_a^3 \tag{12.19}$$

where S is an aberration constant of the system * and r_a the radius of the beam cross-section at the second lens. If r_s^0 is the radius of the Gaussian image of the crossover at the screen, the actual radius of the

* For large magnification M the constant S is identical with the quantity MC/f^2 on p. 165, f denoting the focal length of the lens.

spot will be $r_s^0 + \Delta r_s$. To present the effect of the second-lens aberration on the spot size in a form independent of the lens magnification, it is convenient to divide the increase in spot size by the magnification, yielding the equivalent increase in crossover radius:

$$\Delta r_c = (S/M) r_a^3 \tag{12.20}$$

The coefficient S/M $(= C/f^2)$ is shown as function of $[(\phi_2 - \phi_1)/(\phi_2 + \phi_1)]^2$ for an equidiameter two-cylinder lens in Fig. 12.15; the

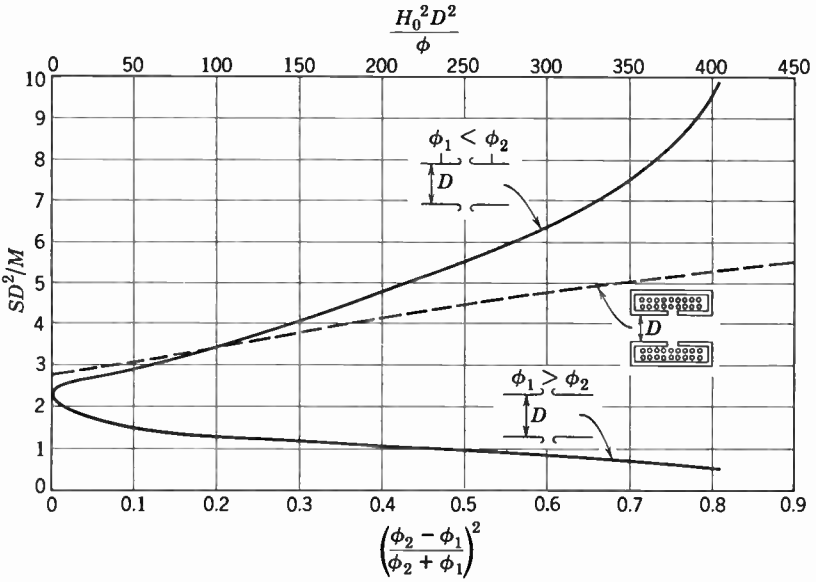


Fig. 12.15. Spherical Aberration of a Two-Cylinder Electrostatic Lens and a Magnetic Pole-Piece Gap Lens (for High Magnification M).

same quantity for a simple pole-piece gap lens, plotted as function of H_0^2/ϕ (H_0 = maximum field strength, ϕ beam potential), is given for comparison. Essentially the same data have been presented in different form in Fig. 4.30.

To illustrate the application of Figs. 4.17 and 12.15 in gun design, consider the specific case of a second lens formed by two cylinders 0.8 inch in diameter with a voltage ratio of 5. From Fig. 4.17 the values of the first and second focal lengths and the positions of the focal points relative to the plane of symmetry of the lens are

$$f_1 = -1.32 \text{ inches}, \qquad z_1 = -2.25 \text{ inches}$$
$$f_2 = 2.95 \text{ inches}, \qquad z_2 = 1.92 \text{ inches}$$

For a throw of 14.5 inches from the center of the gun to the fluorescent screen the lens equation places the object (i.e., the crossover) a distance of 2.6 inches from the center. The magnification of the system then becomes 4.3.

If the tube requirements dictate that the diameter of the final spot be 0.08 inch and the minimum practical diameter of the crossover obtainable with the first-lens system selected is 0.01 inch, the maximum fraction of the second lens which can be utilized can be determined as follows. In the equation $r_s = M(r_c + \Delta r_c)$ substitution of $r_s = 0.04$ inch, $r_c = 0.005$ inch, and $M = 4.3$ leads to $\Delta r_c = 0.0043$ inch. Since, from Fig. 12.15, the value of S/M is 5 diameters^{-2} for $\phi_2/\phi_1 = 5$ or $\{(\phi_2 - \phi_1)/(\phi_2 + \phi_1)\}^2 = 0.44$, Eq. 12.20 yields for the permissible beam diameter in the first principal plane $2r_a = 0.16$ inch. The proper diameter of a limiting aperture placed a distance b from the center of symmetry would be

$$D_{ap} = \frac{2.6 + b}{2.6 + (z_1 - f_1)} \cdot 2r_a = 0.096(2.6 + b)$$

Neglecting the curvature of the rays as they approach the plane of symmetry, the beam diameter in this plane becomes 0.25 inch, or 0.3 of the entire lens diameter. In practice the aperture would of course be placed well back from the symmetry plane so as not to affect the operation of the lens. Furthermore, the cathode area would be restricted to prevent excessive collection of electrons by the first anode and the aperture diaphragm. Finally, the spot size would be smaller than that indicated by the foregoing calculation, since the second lens would be focused to yield the minimum spherical aberration figure rather than the true Gaussian image of the crossover. The resulting spot-size reduction may amount to 30 percent.

With prefocus, i.e., collimation by a lens formed between the accelerating grid and first anode, the required strength of the second lens is reduced and the dependence of spot size and beam current on second-lens position greatly diminished. In view of the reduced beam diameter at the second lens, the aberration of this lens becomes less important and the aperture diaphragm in the focusing electrode serves primarily as a shield for secondary electrons and ions. Thus, though the paraxial spot size is increased somewhat, this effect is outbalanced by the reduced effect of second-lens aberration and the completeness of current concentration in the beam.

Defining the beam by a second-lens aperture is preferable to simply restricting the cathode area. In the latter case the aberrations of the

both to supply a sharply defined object for the second lens and to limit the angular divergence of the beam. The second lens itself is an equipotential lens with the central, or focusing, electrode at a lower potential. This arrangement has two advantages: first, any secondary emission from the limiting aperture is completely suppressed by the low potential of the focusing electrode, and, second, the beam can be focused without altering the beam current. The portion of the second-lens aperture which is utilized by the beam is so small that the relatively high spherical aberration of the second lens is entirely immaterial.

Because of its small beam current, the performance of the iconoscope gun with reference to spot size cannot be conveniently measured on a fluorescent screen. In order to make this measurement, the gun is mounted in a regular iconoscope bulb which contains, instead of a mosaic, a specially prepared screen. This screen consists of a sheet of mica, or glass, coated with a metallic film through which have been ruled groups of fine vertical lines. Each group has a different line separation, for example, 0.002, 0.003, 0.004, and 0.005 inch. When this screen is scanned and the signal output observed on the screen of a kinescope, the line pattern will be reproduced if it is resolved by the spot. By decreasing the horizontal amplitude of scanning in the test tube, the pattern of the lines can be magnified on the kinescope screen, thus eliminating the possibility of the resolution of the reproduced pattern being limited by the viewing tube or amplifier system, rather than the gun under test.

The performance of the gun can be estimated by observing which groups of lines are resolved under these conditions. A correct estimate of the resolution by this method requires some experience. It is quite common for a beginner to assign too small a value to the spot diameter since lines which are only partially resolved because of the nonuniform current distribution in the spot can easily be mistaken for truly resolved lines.

12.8 The Kinescope Gun. Whereas the greater current demand and the need for good control characteristics increase the difficulty of the design problem of the kinescope gun, they are partially offset by the larger permissible spot size and the higher overall voltage.

A simple gun similar to that shown in Fig. 12.16 may function satisfactorily as a kinescope gun if the limiting apertures are enlarged and the first-lens spacings are modified to permit the required higher beam current. However, reduction in the dependence of spot focus on grid voltage (beam current), suppression of negative-ion current to the screen and other factors to be noted below favor more complex designs.

The size of the spot required is determined by the size of the viewing screen, the number of scanning lines, and the overlap desired. For example, a 5-inch projection kinescope will produce a picture about 3 by 4 inches in size. Therefore, the separation of lines in a 525-line pattern will be 0.006 inch, or 0.15 millimeter. In order that two lines do not run together, the brightness at a distance of 0.075 millimeter from the center of a line must be less than a certain predetermined fraction of that at the center of the line. For the present illustration, let it be assumed that this ratio is 0.01. From these data the "spot diameter" can be determined.

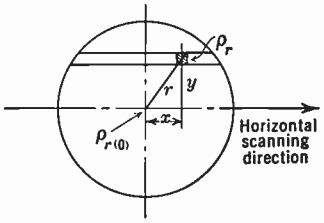


Fig. 12.18. Determination of Current Distribution in Scanning Line.

As has already been pointed out, the term "spot diameter" is rather meaningless unless arbitrarily defined. In section 12.3 the diameter was defined as that within which the ratio of current density at any point to current density at the center was greater than a certain fraction K. It is apparent that the diameter of the spot so defined will not necessarily equal the line width, because the line is scanned with a circular spot. It will be shown, however, that line width and spot diameter agree if the current distribution in the spot is given by Eq. 12.11 so that the density may be written in the form

$$\rho = A\epsilon^{-Br^2}$$

In order to determine the density variation across the line it is necessary to integrate the spot density over elementary strips parallel to the scanning line as indicated in Fig. 12.18. Since, here, $r^2 = x^2 + y^2$, the density distribution in the line becomes

$$\rho_L(y) = 2\int_0^\infty A\epsilon^{-By^2}\epsilon^{-Bx^2}\,dx = A\sqrt{\pi/B}\,\epsilon^{-By^2}$$

The ratio of current density at a distance y from the center of the line to that at the center, $\rho_L(0)$, will, of course, be

$$\frac{\rho_L(y)}{\rho_L(0)} = \frac{A\sqrt{\pi/B}}{A\sqrt{\pi/B}}\epsilon^{-By^2} = \epsilon^{-By^2}$$

If the "line width" 2δ is defined by the relation

$$\frac{\rho_L(\delta)}{\rho_L(0)} = K$$

a comparison with Eq. 12.14 shows that it is equal to the "spot diameter" 2β. Therefore, the spot diameter should be 0.15 millimeter for a 5-inch screen if no overlap is desired, or 0.3 millimeter for a 50 percent overlap.

The current required in the spot depends upon the type of fluorescent material used on the screen, the bombarding voltage, the screen size, and the brightness required. With white-emitting zinc sulphide as the screen material, and 12 kilovolts on the ultor of the gun, a beam current of the order of 1 milliampere will produce a high-light brightness of 140 foot-lamberts on a 16-inch screen (with a scanning pattern of 10 by 13¼ inches). This is more than adequate for the direct viewing of the tube in a normally lighted room. If the kinescope is to be used for projection the screen brightness must be much higher, both because of the greater size of the projected image as compared with the screen image and because of the light loss of the optical projection system. For example, if the high-light brightness of a 15 by 20 inch picture obtained from a 5-inch tube is to be 40 foot-lamberts, and the efficiency of the optical system is 25 percent, the brightness of the tube screen must be 4000 foot-lamberts. This is attained in projection tubes such as the 5TP4 with a beam current of 1.3 milliamperes at an operating voltage of 27 kilovolts.

The actual construction of the 5TP4 gun is shown in Fig. 12.19. Under normal conditions of operation (corresponding to an average beam current of several hundred microamperes) the spot diameter of this tube is, in accord with the requirements derived above, 0.15 millimeter or 0.006 inch. Cathode and control grid have the conventional form, with the grid aperture separated by a few thousandths of an inch from the cathode surface. The grid is followed by a screen grid which is operated at approximately 200 volts, and then a focusing electrode at 4900 volts. The screen grid serves to make the beam current independent of the focus of the second lens, and also, by the adjustment of its voltage, to control the cutoff voltage of the control grid. Normally, this is made equal to -70 volts.

The lens formed between the screen grid and the focusing electrode has a relatively short focal length, its first focal point being a small

distance behind the crossover. This lens forms a virtual image of the crossover, lying behind the real crossover. As a result, the principal rays through the crossover are effectively collimated, so that the amount of current intercepted by the limiting aperture of the second lens is greatly reduced.

The second lens is formed between the focusing electrode and the wall coating on the neck of the tube which serves as second anode. Since the field strength is quite high at the end of the focusing-electrode cylinder, the mouth of this electrode is carefully rounded and freed

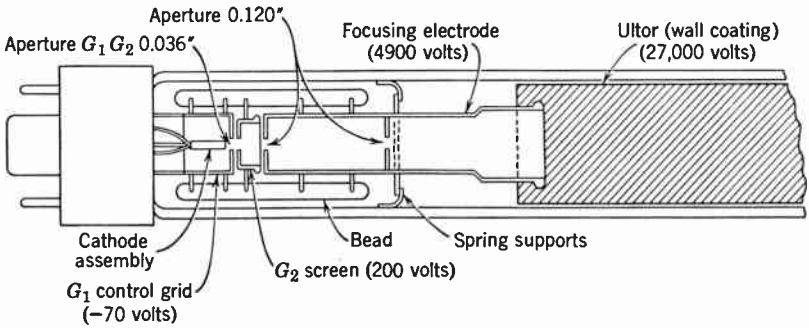


Fig. 12.19. Schematic Diagram of High-Voltage Gun of RCA 5TP4 Projection Kinescope.

from points and roughness to avoid cold emission. For similar reasons the end of the screen grid facing the focusing electrode is also carefully smoothed and rounded.

In the 5TP4 kinescope, as in projection tubes generally, the presence of a thin aluminum film deposited on the fluorescent screen eliminates the destructive action on the screen of negative ions emitted by the gun. Guns of direct-viewing tubes whose screens are not similarly protected require "ion traps" which complicate, to some extent, their construction. The gun for a small direct-viewing kinescope with a 7-inch screen, the 7DP4, is shown schematically in Fig. 12.20. A tilted lens, formed between the screen grid and a cylinder at ultor potential (6000 volts), deflects the negative ions away from the axis of the tube; its similar action on the electron beam is compensated by the transverse magnetic field provided by the pole pieces of an external small magnet which has inappreciable effect on the paths of the much heavier ions.

In greater detail, the ion-trap fields in direct-view kinescopes may take a variety of forms (Fig. 12.21). In one of these two separate magnetic fields are employed (Fig. 12.21*a*). The first, coinciding ap-

proximately with the tilted lens, overcompensates the deflecting action of the tilted lens on the electrons, requiring the second magnetic field, of reverse polarity, to center the beam once more on the gun axis. With a single magnetic field and an axially aligned gun this condition can only be attained with effective coincidence of the electric and magnetic deflecting fields, which is not readily achieved. However, a second magnetic ion-trap field becomes superfluous with the "tilted gun" (Fig. 12.21*b*), in which the axis of the gun electrodes is tilted

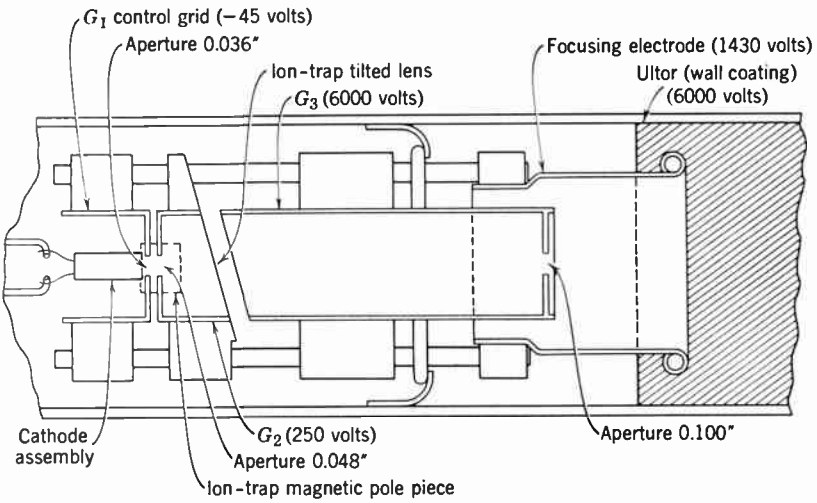


Fig. 12.20. Schematic Diagram of Gun of RCA 7DP4 Kinescope.

with respect to the tube axis. As an alternative, the tilted lens may be omitted and the anode tube be bent as shown in Fig. 12.21*c*; in such a "bent gun" the ion-trap magnetic field is placed at the bend in the anode tube. Still another ion-trap system, employed in England,* derives its ion-deflecting properties from the decentering of the grid and screen-grid apertures (Fig. 12.21*d*). These apertures are, furthermore, so displaced relative to each other that a superposed magnetic field, nullifying the deflecting action of the asymmetric electric fields between the apertures on the electrons, causes the electron beam to be displaced onto the tube axis.

The ion-trap magnetic field required in the single-field traps is materially weaker than that required in the double-field trap. Among the single-field traps the tilted gun has the further advantage of re-

* See Moss, reference 4.

quiring neither the manufacture of asymmetric elements (as the bent gun) nor very precise adjustment (as the decentered aperture gun).

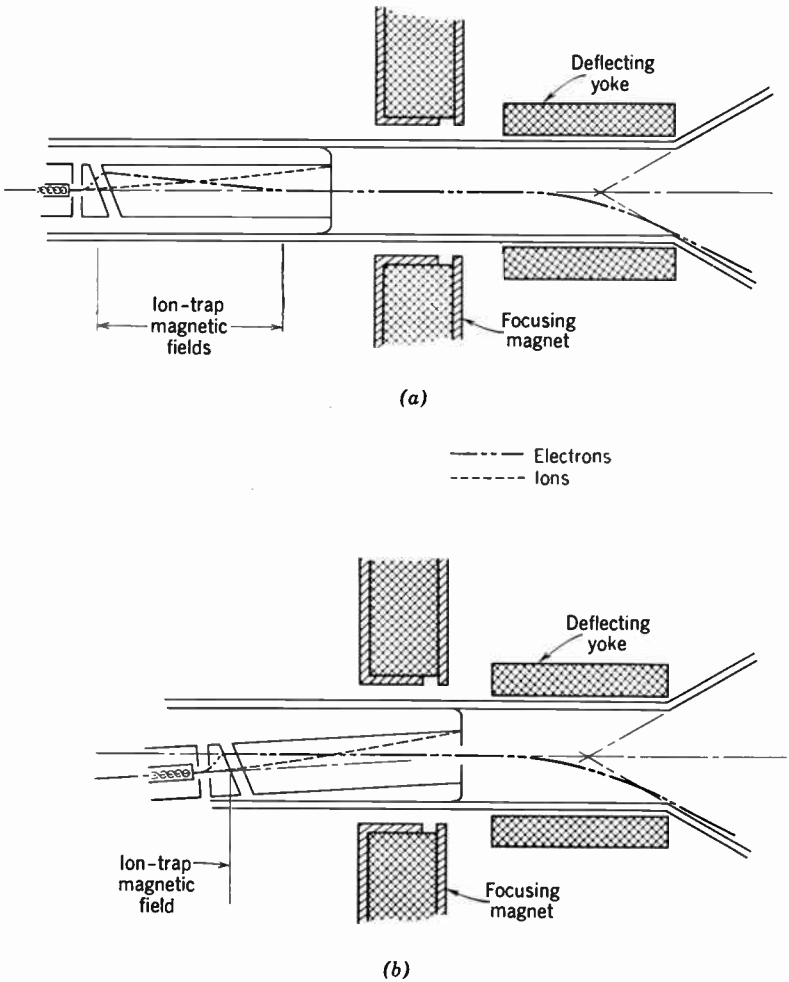


Fig. 12.21. Ion-Trap Types. (*a*) Double magnetic field trap. (*b*) Tilted gun.

The second lens of the 7DP4 (Fig. 12.20) is seen to be an electrostatic unipotential lens with its center element at a relatively low potential. The limiting aperture for the lens is placed at the end of the cylinder at ultor potential. With this arrangement the current collection by the focusing electrode is very small, so that a relatively high-resistance

bleeder may be employed to supply the focusing voltage without risking defocusing as the beam current is varied. In this manner the required capacity of the high-voltage supply is reduced. This consid-

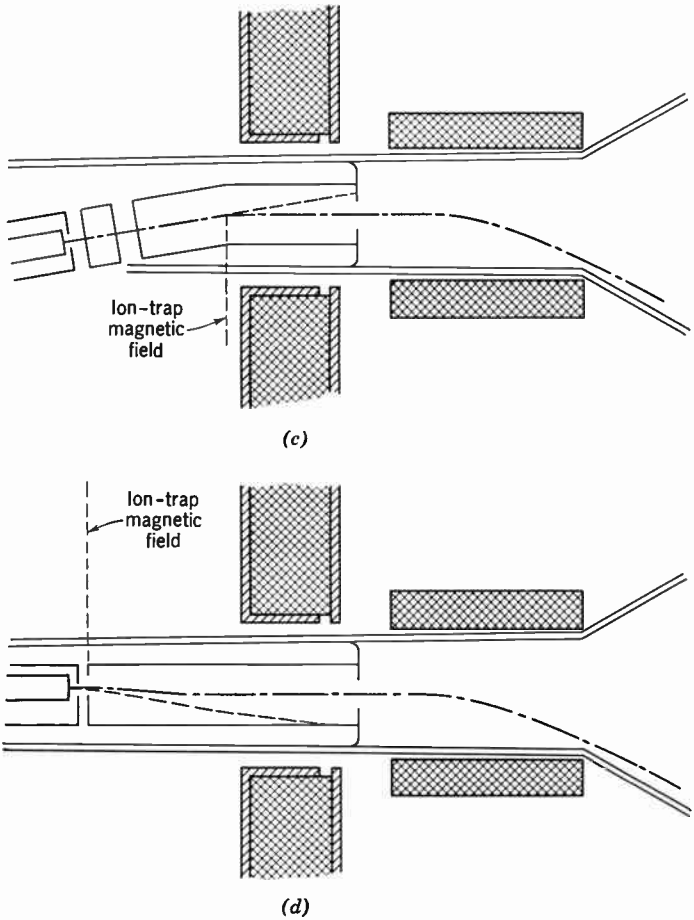


Fig. 12.21. Ion-Trap Types. (*c*) Bent gun. (*d*) Decentered aperture gun.

eration must be weighed against the disadvantage of higher spherical aberration inherent in this type of lens.

In the larger direct-view kinescopes, in which a greater beam current and, hence, a fuller utilization of the second-lens aperture are required, a magnetic second lens is commonly employed. Figure 12.22 shows schematically the electron gun of the 16GP4 kinescope. The gun as-

sembly is here mounted at a small angle with respect to the tube axis so as to simplify the structure of the magnetic field required in the ion trap. Apart from this, the first-lens region (including the ion trap) resembles that of the 7DP4 kinescope. On the other hand, the electrons, after having passed through the tilted lens forming part of the

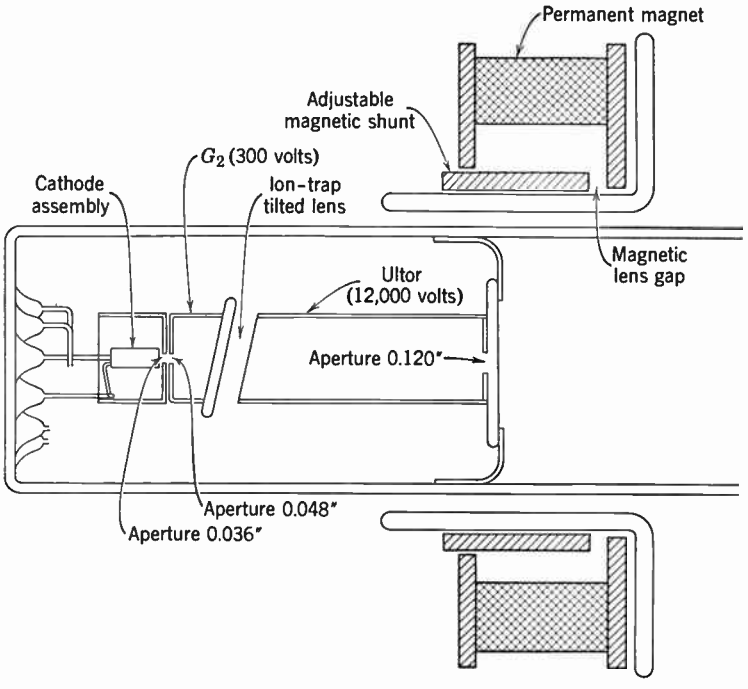


Fig. 12.22. Schematic Diagram of Gun of RCA-16GP4 Kinescope with Permanent-Magnet Second Lens.

ion trap and having entered the ultor cylinder, undergo no further accelerations or decelerations.

The magnetic lens which focuses the beam may be either electromagnetic, permanent-magnet-excited, or a combination of the two. In the example shown in Fig. 12.22 the magnetomotive force is provided entirely by a set of permanent magnets mounted between two soft-iron disks. Focusing is accomplished by mechanically displacing a soft-iron cylinder which forms a magnetic shunt between the two disks; by increasing or decreasing the gap between the cylinder and one of the disks the magnetic field along the axis of the tube is increased or decreased, respectively.

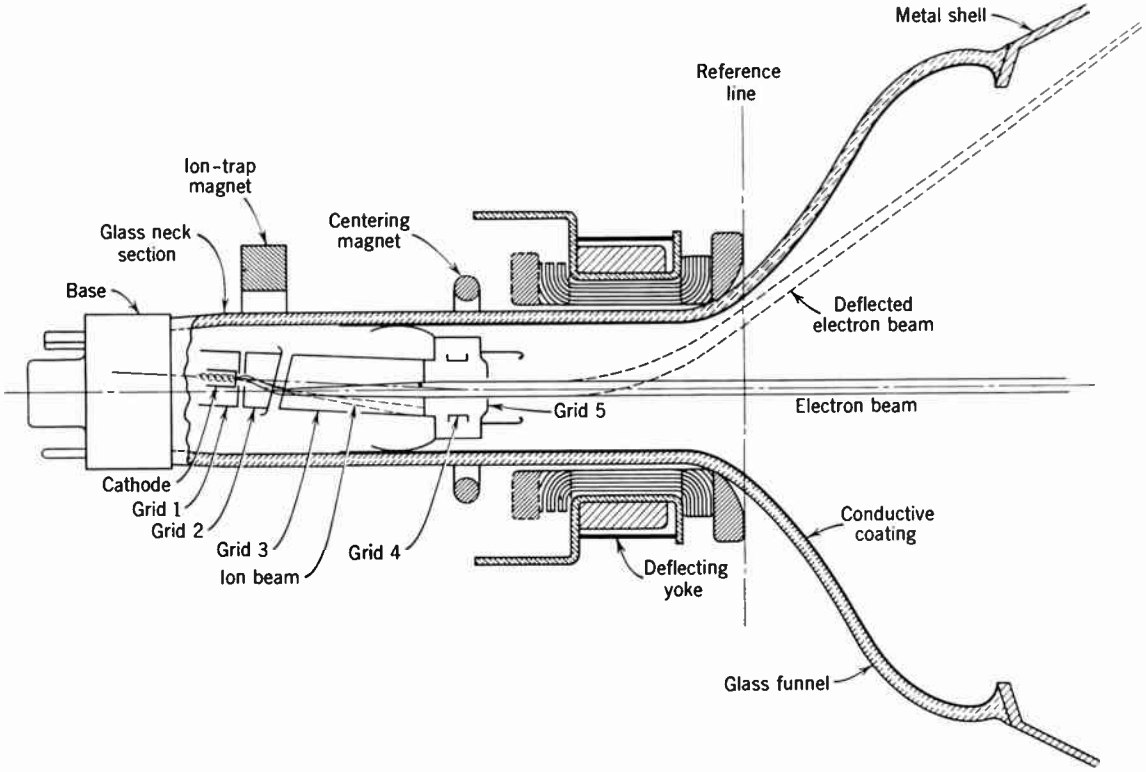


Fig. 12.23. Cross-Section of Electrostatic Focus Gun. (Swedlund and Saunders, reference 25; courtesy of *Electronics*.)

A saving in scarce metals, particularly cobalt and/or copper, may be effected by replacing the magnetic second lens with an equipotential electrostatic lens, as indicated in Fig. 12.23.* For production economy and interchangeability, the two types of guns (Fig. 12.24) are designed to fit into identical tube envelopes and to operate with the same de-

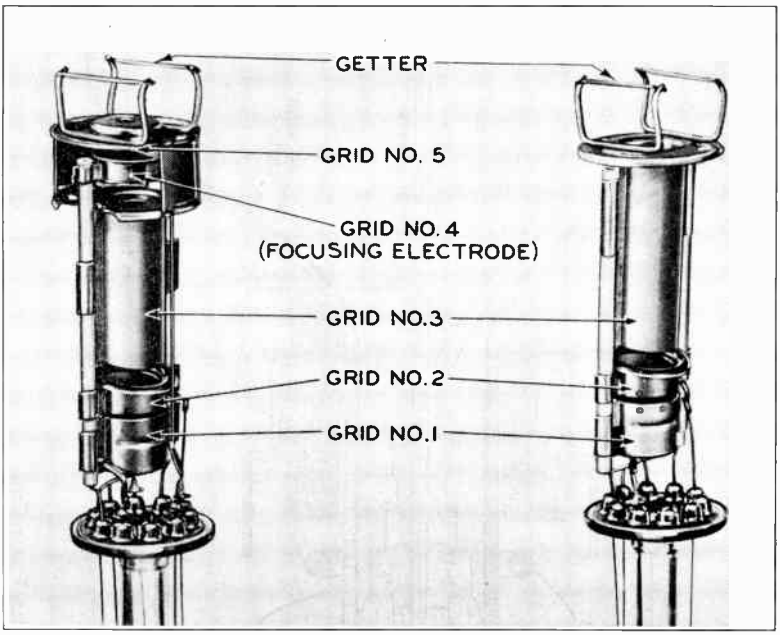


Fig. 12.24. Structure of Electrostatic Focus Gun Compared with Magnetic Focus Gun. (Swedlund and Saunders, reference 25.)

flection yokes. The added short focusing cylinder in the electrostatic gun (grid No. 5) operates at a voltage of 5 kilovolts or less with respect to the cathode. The distance between cathode and second lens is reduced somewhat, reducing the beam size in the deflection field and, hence, deflection defocusing, at the cost of somewhat larger spot size at the center of the picture. Since, in tubes with magnetic focus, it has become customary to center the scanning pattern on the tube screen by slightly tilting the second-lens assembly, a small centering magnet ring is provided with the electrostatic tube to accomplish the same purpose.

* See Swedlund and Saunders, reference 25.

An incidental advantage of the electrostatic gun is that its focus is not influenced by high-voltage changes, provided that the focusing voltage is derived from the high-voltage supply. As a still further step in the direction of making spot focus independent of the overall tube voltage, kinescope guns have been designed with a unipotential second lens, whose center element is connected to the cathode.* In practice

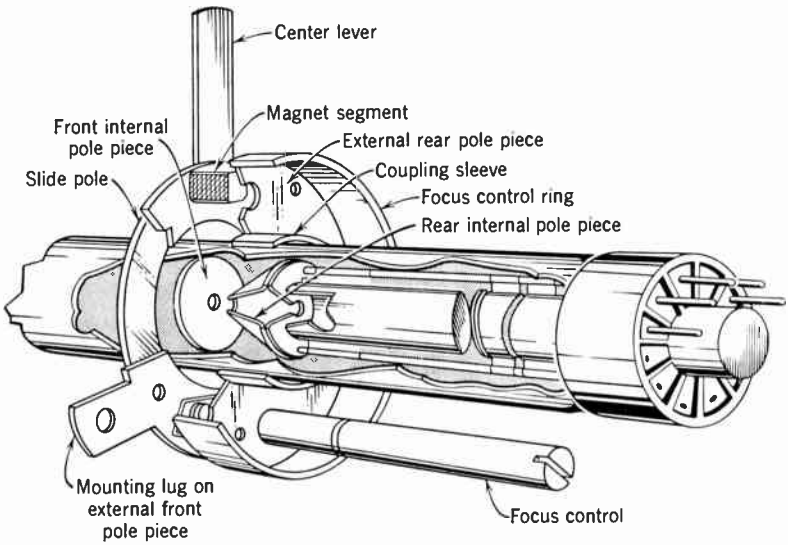


Fig. 12.25. Kinescope Gun with Internal Pole Pieces. (Fogelberg, Morse, Reiches, and Ingle, reference 27; courtesy of *Electronics*.)

provision must be made for applying a small focusing voltage between the cathode and the middle lens electrode, or extremely stringent tolerances must be maintained in the manufacture of the gun.

An alternative approach to the problem of conserving magnetic materials or ampere turns of focusing power consists of placing the magnetic pole pieces of the second lens inside the tube neck. By this means the volume over which the focusing field extends, and hence the total magnetic flux required for focusing, are greatly reduced. At the same time, the stricter localization of the focusing fields lessens difficulties from interaction between the deflection field and the focusing field.

*See Bentley, Hoagland, and Grossbohlin, reference 26.

One form of an internal pole piece gun is shown in Fig. 12.25.* The internal pole pieces are here a hollow soft-iron cone and a flat apertured disk mounted on the end of the gun structure. Two flat external soft-iron rings and a series of permanent magnet bars between them (or a sleeve enclosing the field coil) complete the magnetic circuit. Lens strength is adjusted by the displacement of one or both of the external pole pieces along the tube axis, centering, by the transverse displacement of a sliding pole piece ring on the front external pole.

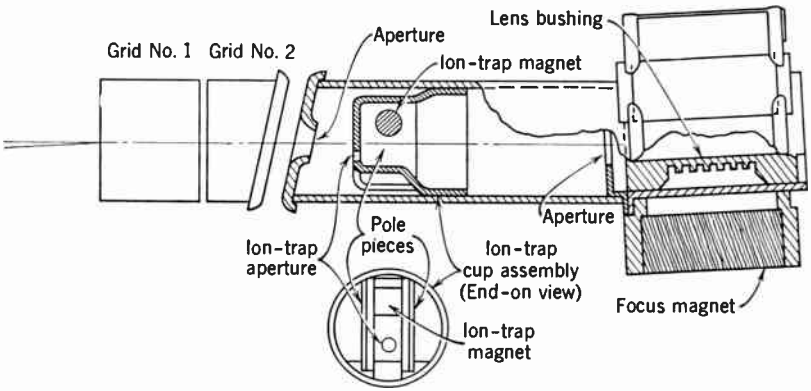


Fig. 12.26. Internal Pole Piece Gun of the General Electric Company.

The General Electric internal pole piece gun shown in Fig. 12.26 represents a more radical departure from conventional construction. Here, both the pole pieces and the permanent magnets are placed within the tube neck. The pole pieces, which are placed within the ultor tube, are machined out of a single iron cylinder, high-reluctance gaps being formed by reaming out the cylinder to such a thickness that the residual wall is fully saturated. The ion-trap magnet is also placed within a gun cylinder. Both the focusing magnets and the ion-trap magnet are magnetized in place by passing high-current pulses through external coils placed on the tube neck. In practice, they are first overmagnetized and then demagnetized in short steps with reversed-current pulses until spot focus and spot alignment are satisfactory.

12.9 Guns for Low-Velocity-Beam Tubes. The gun required in low-velocity-beam pickup tubes, such as the orthicon, image orthicon, and Vidicon, must satisfy conditions which differ materially from those imposed on kinescope and iconoscope guns. In the low-velocity guns, the

* See Fogelberg, Morse, Reiches, and Ingle, reference 27.

electrons are accelerated away from the cathode, move down the tube at velocities corresponding to an accelerating potential of 50 to 200 volts, and are then slowed down to virtually zero velocity at the target. The spot formed by the intersection of the beam with the target must be small—in the Vidicon, in which the scanning pattern is only half an inch wide, less than 0.001 inch in diameter. In addition it is desirable—particularly in the image orthicon—that the electrons approach the target (for given target potential) with as nearly uniform a velocity component normal to the target as possible. Hence, the gun should not introduce a larger range in the direction of motion of the electrons than necessary. The only factor which makes it possible to satisfy all the above conditions simultaneously is that the required beam current is small—of the order of 0.1 microampere in the orthicon and Vidicon, of 0.02 microampere in the image orthicon.

In all three tubes the beam is focused on the target by a longitudinal magnetic field of sufficient strength to form a series of intermediate images of the spot between guns and target. This magnetic field extends into the gun region, although it is, in general, attenuated here.

The simplest electron gun would here consist of a small emitting source which is imaged by the uniform magnetic field on the target. Uniform electric fields would serve both to accelerate the electrons away from the cathode and to retard them as they approach the target. If the target potential were equal, under these circumstances, to that of the emitting surface of the cathode (due account being taken of contact potential) and aberrations introduced by the magnetic focusing field were neglected, a perfect image of unity magnification would be formed of the source at the target and the current density in the spot would be exactly equal to the emission density of the cathode. As the target is made more negative with respect to the cathode, the current density in the spot would become

$$\rho = \int_{\phi_T}^{\infty} \rho_0 \frac{e}{kT} \exp\left(-\frac{e\phi_0}{kT}\right) d\phi_0 = \rho_0 \epsilon^{-e\phi_T/(kT)} \tag{12.21}$$

Here ρ_0 is the emission density and ϕ_T the potential difference between cathode and target.

The performance described can be approached quite closely even with practically realizable, nonuniform accelerating and retarding fields. Even so, the point-cathode gun does not represent a practical solution for the low-velocity pickup tube beam source: (1) It makes no provision for controlling the beam current except by adjusting the

temperature of the cathode; this makes it difficult to adjust to optimum beam current for the immediate operational requirement and does not permit cutting off the beam during the return periods of the scanning beam. (2) The current obtainable in a spot of the prescribed size is often not large enough to discharge the target under lighting conditions giving optimum signal-to-noise ratio. The second factor is particularly significant in the orthicon and Vidicon.

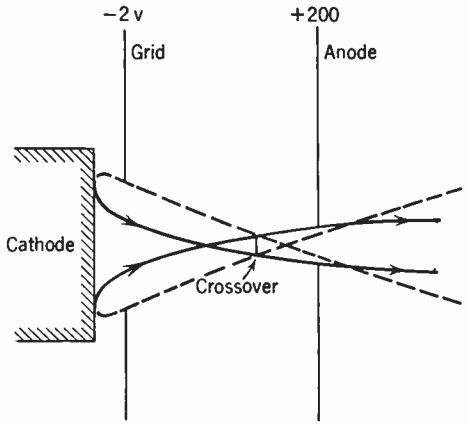


Fig. 12.27. Low-Velocity Gun.

A gun which overcomes both the above-mentioned objections is shown in Fig. 12.27. This gun consists of a cathode with a fairly large emitting area, a control grid close to the cathode, and a small defining aperture spaced a distance from the cathode equal to four or five times the cathode-grid spacing. The magnetic field extends from the cathode to the target and focuses the beam electrons.

As operated, the grid-cathode structure forms a crossover which lies on the cathode side of the small defining aperture. In a typical electrode assembly of this kind, the crossover will be about 0.005 inch in diameter. Its position depends on the grid potential. For smallest spot, the magnetic field is adjusted to focus the small defining aperture as sharply as possible on the target.

The current obtained from a gun of this type is greater, for a given spot size, than where a small cathode is imaged on the target. However, the range of velocities is also greater. Quite generally, it is not possible to obtain a greater current density of electrons in any small velocity range than that corresponding to the maximum current density in an equal velocity range in the low-velocity gun first described.

Figure 12.28 illustrates the current reaching the target as a function of target potential for the two types of guns. The practical voltage range for the directly imaged cathode is less than 0.2 volt whereas for a large-cathode crossover gun the range may be 1 to 1.5 volts.

The velocity range in the large-cathode crossover gun depends on the bias voltage applied to the control grid. As the grid is made more negative, the crossover moves back toward the cathode and electrons

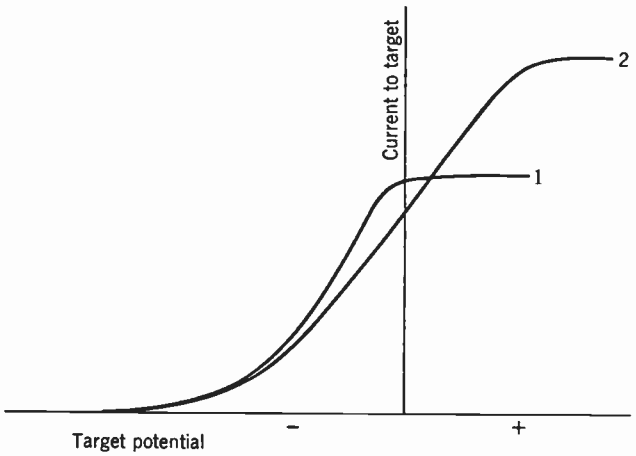


Fig. 12.28. Current to Target from Low-Velocity Gun. (1) Point cathode gun. (2) Focused gun.

from a smaller area of the cathode enter the beam. Since the components of the beam having low axial velocities consist of electrons coming from the outer portions of the cathode, to which the lens action of the grid-cathode system has imparted transverse velocities at the expense of longitudinal velocities, the more negative grid potential tends to remove these components and decrease the velocity range.

In the orthicon and Vidicon, in which the signal is removed by means of a signal plate on the back of the target, a large beam current with considerable velocity range is not undesirable as long as a sharp spot can be maintained. In the image orthicon, on the other hand, where the magnitude of the return beam determines the noise superposed on the signal, the beam current should be made no larger than required to discharge the brightest areas of the target. Furthermore, this required current will be least if the velocity range of the electrons is very small since, then, a relatively large proportion of the beam electrons can

land, for any given target potential, on the target and contribute to the discharging process.

An image orthicon gun of the type represented disgrammatically in Fig. 12.27 is shown, in section, in Fig. 12.29. The magnetic field

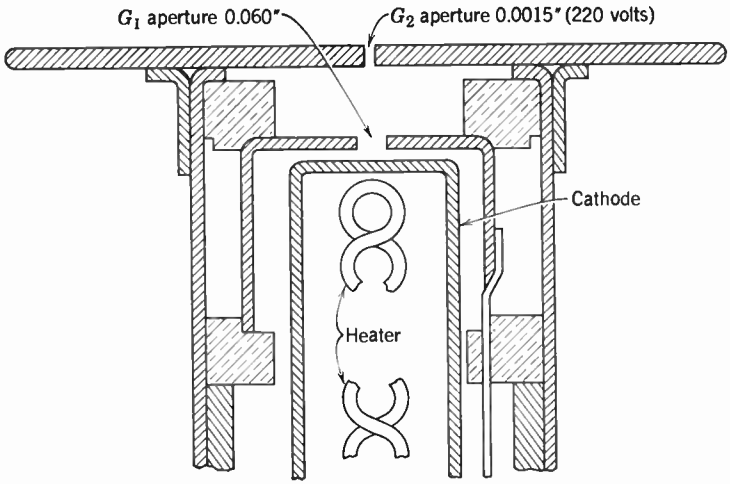


Fig. 12.29. Schematic Diagram of Gun of RCA 2P23 Image Orthicon.

strength at the gun is actually only a third or half of that in the main body of the tube. As compared with the point-cathode gun it possesses the practical advantages, not mentioned before, of greater ease of acti-

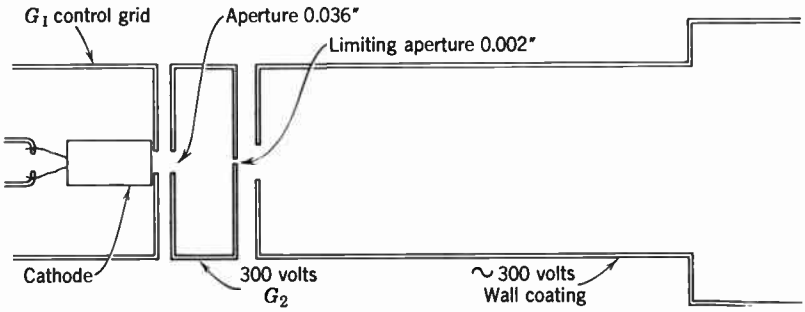


Fig. 12.30. High-Current Gun for the Vidicon (Schematic).

vation and reliability of the cathode and of shielding the photosensitive elements in the tube from the light of the cathode. With a magnetic focusing field of the order of 100 gausses and a beam energy of 200

electron volts, a real image of the defining aperture is formed at points approximately an inch apart, so that the image formed at the target is of relatively high order.

In the Vidicon currents of 1 to 2 microamperes, extending the usefulness of the tube to higher light levels, may be obtained with the gun shown in Fig. 12.30. With this arrangement the current-transmitting region in front of the center of the cathode is imaged in the plane of the defining aperture. A grid bias of the order of 100 volts is required for cutoff, a voltage swing of 20 to 30 volts to vary the beam current between zero and its maximum value.

REFERENCES

1. E. Brüche and O. Scherzer, *Elektronenoptik*, Springer, Berlin, 1934.
2. I. G. Maloff and D. W. Epstein, *Electron Optics in Television*, McGraw-Hill, New York, 1938.
3. V. K. Zworykin, G. A. Morton, E. G. Ramberg, J. Hillier, and A. W. Vance, *Electron Optics and the Electron Microscope*, Wiley, New York, 1945.
4. H. Moss, "Cathode Ray Tube Progress in the Past Decade, with Special Reference to Manufacture and Design," *Advances in Electronics* (Ed. L. Marton), Vol. II, Academic Press, New York, 1950.
5. K. R. Spangenberg, *Vacuum Tubes*, McGraw-Hill, New York, 1948.
6. F. Schröter, *Fernsehen*, Springer, Berlin, 1937.
7. A. Wehnelt, "Increase in the Sensitiveness of the Braun Tube by Use of Cathode Rays of Lower Velocity," *Physik. Z.*, Vol. 6, p. 732, 1905.
8. E. Brüche and O. Scherzer, "Braun Tube as an Electron-Optical Problem," *Z. tech. Physik*, Vol. 14, pp. 464–466, 1933.
9. M. Knoll, "Electric Electron Lens for Cathode Ray Tubes," *Arch. Elektrotech.*, Vol. 28, pp. 1–8, 1934.
10. M. Knoll and J. Schloemilch, "Electron-Optical Current Distribution in Electron Tubes with Control Electrodes," *Arch. Elektrotech.*, Vol. 28, pp. 507–516, 1934.
11. I. G. Maloff and D. W. Epstein, "Theory of the Electron Gun," *Proc. I.R.E.*, Vol. 22, pp. 1386–1411, 1934.
12. M. Knoll, "Electron Optics in Television Technique," *Z. tech. Physik*, Vol. 17, pp. 604–617, 1936.
13. D. W. Epstein, "Electron Optical System of Two Cylinders as Applied to Cathode Ray Tubes," *Proc. I.R.E.*, Vol. 24, pp. 1095–1139, 1936.
14. D. B. Langmuir, "Limitations of Cathode Ray Tubes," *Proc. I.R.E.*, Vol. 25, pp. 977–991, 1937.
15. R. R. Law, "High-Current Electron Gun for Projection Kinescopes," *Proc. I.R.E.*, Vol. 25, pp. 954–976, 1937.
16. V. K. Zworykin and W. H. Painter, "Development of the Projection Kinescope," *Proc. I.R.E.*, Vol. 25, pp. 937–953, 1937.
17. E. Gundert, "Electron Lens Aberrations Shown by Thread Beams," *Physik. Z.*, Vol. 38, pp. 462–467, 1937.

18. K. Spangenberg and L. M. Field, "Some Simplified Methods of Determining the Optical Characteristics of Electron Lenses," *Proc. I.R.E.*, Vol. 30, pp. 138–144, 1942.
19. K. Spangenberg and L. M. Field, "The Measured Characteristics of Some Electrostatic Electron Lenses," *Electrical Communications*, Vol. 21, pp. 194–204, 1943.
20. J. R. Pierce, "Limiting Current Densities in Electron Beams," *J. Appl. Phys.*, Vol. 10, pp. 715–724, 1939.
21. B. J. Thompson and L. B. Headrick, "Space Charge Limitations on the Focus of Electron Beams," *Proc. I.R.E.*, Vol. 28, pp. 318–324, 1940.
22. H. Moss, "The Electron Gun of the Cathode Ray Tube," *J. Brit. Inst. Radio Engrs.*, Vol. 5, pp. 10–25, 1945; Vol. 6, pp. 99–129, 1946.
23. L. M. Field, "High-Current Guns," *Revs. Mod. Phys.*, Vol. 18, pp. 353–361, 1946.
24. G. A. Morton, "Electron Guns for Television Application," *Revs. Mod. Phys.*, Vol. 18, pp. 362–378, 1946.
25. L. E. Swedlund and R. Saunders, Jr., "Material-Saving Picture Tube," *Electronics*, Vol. 24, pp. 118–120, April, 1951.
26. A. Y. Bentley, K. A. Hoagland, and H. W. Grossbohlin, "Self-focusing Picture Tube," *Electronics*, Vol. 25, pp. 107–109, June, 1952.
27. C. V. Fogelberg, E. W. Morse, S. L. Reiches, and D. P. Ingle, "Using C-R Tubes with Internal Pole Pieces," *Electronics*, pp. 102–105, October, 1952.

13 Video Amplifiers

13.1 The Amplifier. In the foregoing chapters the two basic terminal tubes of the television system have been discussed in some detail. It is evident from this discussion that, even where the camera tube merely operates a monitoring kinescope, it is necessary to amplify its signal output by a factor of many thousands before the signal voltage will be sufficient to swing the control grid of the viewing tube. If the camera tube is an element in a complete television broadcast system,

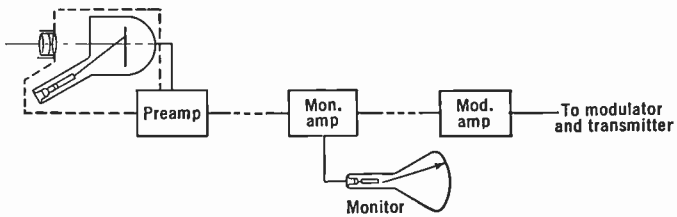


Fig. 13.1. Video Amplifier Chain at Transmitter.

even further amplification is necessary. Not only must the picture signal from the camera tube be amplified up to the modulation level of the transmitter, but the signal also must be amplified at the receiver to the level required to operate the kinescope.

The amplification of the transmitter may, for convenience, be divided into three steps, as was mentioned in Chapter 6. A camera amplifier with a gain of about 1000 is located at the television camera. The final stage of this amplifier has a fairly low output impedance and feeds a low-capacity flexible cable. The input of the monitoring amplifier is fed from this cable, and the signal is amplified up to monitoring level, and also to a level sufficient to supply the signal to a coaxial cable connecting the monitor to the transmitter. The output of a second cable supplies the modulation amplifier, which amplifies the signal to such an extent that it can be used to modulate the transmitter. This chain is illustrated in Fig. 13.1.

At the television receiver the incoming signal must again be amplified. Here the problem is somewhat different in that a large part of the amplification takes place in the radiofrequency and intermediate stages of a superheterodyne receiver. This is shown in Fig. 13.2. However, the output of the second detector must usually be amplified by a video amplifier similar to that at the point of transmission before the signal is sufficient to operate the kinescope control grid.

Since the requirements of these video amplifiers are the same except for the power which they must handle, the principles which will be considered in this chapter apply for the most part equally to all of them.

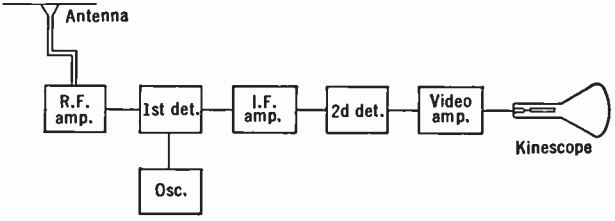


Fig. 13.2. Amplifiers in Superheterodyne Receiver.

13.2 Requirements of a Video Amplifier. The video amplifier is a thermionic vacuum-tube amplifier, and its basic principles are similar to those of the ordinary audio amplifier familiar in radio practice. Because the general amplifier problem, as represented by the audio amplifier, has been treated so ably and completely in numerous textbooks, it is unnecessary to discuss here the entire theory of the amplifier. Instead, only problems resulting from differences in the requirements of audio and video amplifiers will be treated.

The two principal demands which distinguish the video amplifier from the audio amplifier are:

1. A constant frequency response over a band several megacycles wide.
2. A constant time delay over the usable frequency range.

The differences between the two types of amplifiers arise from these requirements of the video amplifier. Some attendant problems which must be considered are those of maintaining response at both high and low frequencies and at the same time keeping a linear variation of phase delay with frequency, of shielding without introducing high capacity, of decoupling the power supply to avoid low-frequency oscil-

lation or motor boating, and of keeping a satisfactory signal-to-noise ratio.

The frequency band required to transmit a picture was discussed in some detail in Chapter 5. Here it was concluded that to transmit a 525-line picture with an aspect ratio of 4 to 3 at 30 frames per second required a bandwidth of approximately 4.5 megacycles. It should be noted that most audio amplifiers have a constant response over less than 10,000 cycles or, in other words, have a bandwidth of less than one three-hundredth of that required of a video amplifier.

Since it is a property of the human ear that it cannot (within rather wide limits) detect phase distortion, the problem of maintaining constant time delay does not enter into the design of an audio amplifier. However, it is apparent from the Fourier analysis of the picture signal given in Chapter 5 that the phase of the various frequency components making up the picture is very important. This is also obvious if the mechanism of the formation of the picture by scanning is kept in mind. The scanning beam in a 12-inch kinescope moves across the screen at a speed of 1.5×10^5 inches per second. Thus, if there is a time delay of only one microsecond for a particular frequency, the position of light maxima and minima resulting from this frequency will be shifted by 0.15 inch. If there is an equal time delay for every frequency it will merely result in a shift of the picture as a whole and does not introduce any distortion. Therefore, the time delay over the frequency band used must be zero or a constant. Time delay is related to the phase angle as follows:

$$\Delta t = \phi/(2\pi f)$$

where ϕ is the phase angle. Hence, a constant time delay requires a phase angle which increases or decreases in proportion to the frequency. The phase angle of the video amplifier must, therefore, be proportional to frequency over the 4.5-megacycle band required.

13.3 Amplifier Types. A rather wide variety of types of amplifying systems for sound equipment is available, differing from one another in the means used to couple the successive vacuum tubes. The principal types are (1) inductance coupled, (2) transformer coupled, (3) direct coupled, (4) resistance coupled, and (5) cathode coupled. This classification is not truly justified, as will become apparent as the discussion proceeds; however, it simplifies a preliminary survey of the field. The five amplifier types are shown schematically in Fig. 13.3. There are, of course, many modifications of each.

When each is examined in turn as to its desirability as a video amplifier, it becomes evident that the resistance-coupled amplifier most nearly meets the requirements.

The inductance-coupled amplifier, though economical in power consumption, is impractical because the great range of frequencies involved makes it impossible to use any of the known types of coupling chokes.

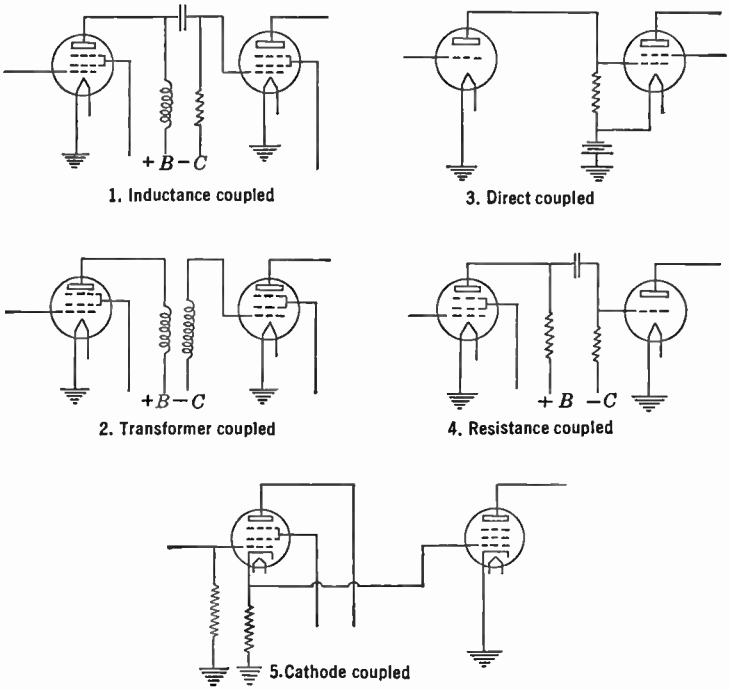


Fig. 13.3. Principal Types of Class-A Amplifier Couplings.

Air-core chokes cannot be used because the very large number of turns necessary to obtain sufficient impedance at low frequencies makes the distributed capacities high, which spoils both phase and amplitude response at high frequencies. Iron-core chokes, on the other hand, exhibit a lack of response at high frequencies, and therefore are not suitable. Magnetite cores, or those using a powdered iron alloy in an insulating binder, as well as ferrite cores consisting of suitably processed complex metal oxides, come closer to yielding chokes which have the desired properties. Even these, however, have undesirable phase shifts and variation in impedance over a band as wide as that required by television transmission.

Transformer-coupled amplifiers suffer from the same type of defects as inductance-coupled amplifiers and are, therefore, also impractical.

Direct-coupled amplifiers are those in which each tube as a whole is maintained at a higher potential than the preceding tube, so that the plate of one tube is at the d-c potential of the grid potential of the next. This type of amplifier can be made to have excellent phase and amplitude characteristics. However, its adjustment is extremely critical, and, furthermore, it is very sensitive to voltage fluctuations of the power supply. Even though this type of amplifier has been successful in experimental installations, its operation is considered too unreliable to make it really serviceable for a practical television system.

The resistance-coupled amplifier is by no means ideal; nevertheless, by resorting to certain corrective measures, it is possible to obtain a satisfactory response. This type of amplifier is the one chiefly used as video amplifier and, therefore, will be considered in detail in the succeeding sections.

The final class, the cathode-coupled amplifiers, cannot be used for voltage amplification since their gain never exceeds unity. However, they have their place in television as a means of coupling a low impedance load, such as a cable, to a voltage amplifier.

13.4 Resistance-Coupled Amplifier. The principle of the operation of the resistance-coupled amplifier can be seen from the circuit diagram in Fig. 13.3. If a positive impulse is impressed upon the grid of the first tube through its coupling condenser, the plate current through this tube increases. The rise in current increases the IR drop through the external plate resistance of the tube, causing its plate end to become more negative. The negative impulse thus produced is transmitted through the second coupling condenser onto the grid of the second tube, where it causes a decrease in the plate current of that tube.

In order to predict the behavior of a resistance-coupled amplifier as a video amplifier, a quantitative analysis is necessary. A single stage of the amplifier, together with its equivalent circuit, is shown in Fig. 13.4. The symbols in this figure have the following significance. The input voltage is e_1, and the voltage amplification μ, so that in the equivalent circuit the first tube can be represented as a voltage generator whose output is $e_1\mu$. R, r_p, and R_g are the external plate resistance, the internal plate resistance, and the grid resistance, respectively. The condenser C is the coupling condenser, C_p is the capacity of the plate of the first tube to ground, and C_g is the total input capacity of the succeeding tube.

It is evident from Fig. 13.4 that the interstage coupling in the resistance-coupled amplifier is an impedance network rather than a pure resistance. Similarly, for the other amplifiers shown in Fig. 13.3, the coupling is a network and the designation indicates merely the predominant character of the network over part or all of the working range of the amplifier. The coupling network of the resistance-coupled amplifier exhibits the characteristics of a resistance only over that portion of the total frequency band for which the impedance of the condenser C is small compared with R_g and the impedance of the tube capacity is large in comparison with R.

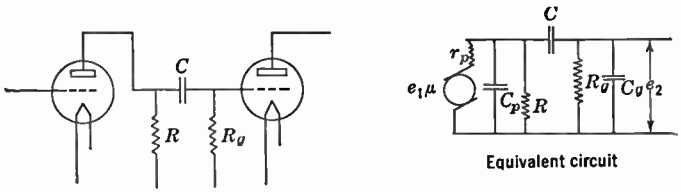


Fig. 13.4. Simple Resistance-Coupled Stage and Its Equivalent Circuit.

The response of this network to a voltage applied to the grid of the first tube can be determined from the circuit constants. If the voltage e_1 is applied at the first grid, the voltage e_2 appearing at the grid of the second tube will be *

$$e_2 = \frac{Z_1 Z_2 \mu e_1}{Z_1 Z_2 + r_p Z_1 + r_p Z_2 - j(r_p X_c + X_c Z_1)} \tag{13.1}$$

where

$$Z_1 = \frac{R_g - j\omega C_g R_g^2}{1 + \omega^2 C_g^2 R_g^2}$$

$$Z_2 = \frac{R - j\omega C_p R^2}{1 + \omega^2 C_p^2 R^2}$$

and

$$X_c = 1/(\omega C)$$

It is apparent that the output voltage e_2 is not independent of the frequency $f = \omega/(2\pi)$, and that therefore the response of the amplifier may not be flat over the frequency band required by the video signal.

* To conform with electrical engineering practice, j is used instead of i to express $\sqrt{-1}$ in this and in subsequent chapters.

Since for an amplifier of this type $C \gg C_p$ or C_g, and R_g is very large, the expression given in Eq. 13.1 reduces to

$$e_2 = \frac{\mu e_1 R}{R + r_p}$$

over the middle range of frequencies (i.e., frequencies such that $1/(\omega C_g)$ and $1/(\omega C_p)$ are large compared with R, and $1/(\omega C)$ is small in comparison with R_g). At the two extremes e_2 falls off for a given input voltage, and, furthermore, there is a varying phase relation between the input and output voltages.

Instead of evaluating Eq. 13.1 directly, which leads to a rather cumbersome algebraic expression, more information as to the effect of the various circuit constants on the frequency response is obtained if approximations are made which permit a simple evaluation of the response at the highest and lowest frequencies within the transmitted band.

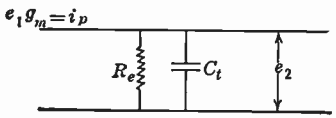


Fig. 13.5. High-Frequency Equivalent Circuit of a Resistance-Coupled Stage.

It is well known that an amplifier tube may be represented by a constant current generator shunted by the internal impedance of the tube. Representing the transconductance by g_m, the current output of this generator is given by the relation

$$i_p = e_1 g_m$$

Furthermore, since the television amplifier must operate at low frequencies, the coupling condenser C must be large and will, therefore, present negligible impedance at the upper frequencies. At these frequencies the equivalent circuit can be represented with fair accuracy by Fig. 13.5. In this figure $C_t = C_p + C_g$ is the total shunt capacity, and R_e is the effective shunt resistance of R, r_p, and R_g in parallel, namely:

$$\frac{Rr_pR_g}{Rr_p + R_gr_p + R_gR}$$

The output voltage e_2 under these conditions is

$$\begin{aligned} e_2 &= e_1 g_m Z \\ &= \frac{R_e - j\omega C_t R_e^2}{1 + \omega^2 C_t^2 R_e^2} e_1 g_m \end{aligned} \tag{13.2}$$

The magnitude of the gain as a function of frequency is, therefore,

$$G = \frac{R_e g_m}{\sqrt{1 + 4\pi^2 f^2 C_t^2 R_e^2}} \tag{13.3}$$

whereas the phase angle is

$$\phi = -\arctan (2\pi f C_t R_e) \tag{13.4}$$

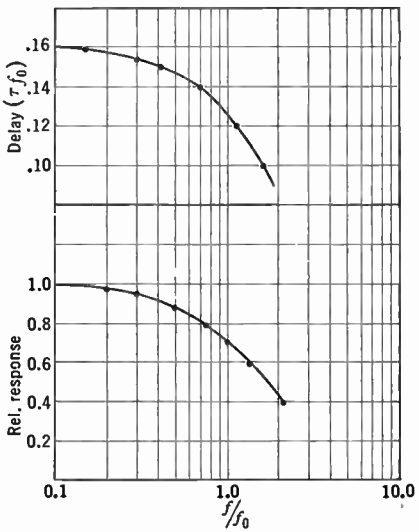


Fig. 13.6. Response of the Coupling Shown in Fig. 13.5.

For convenience, a certain maximum or critical frequency f_0 will be defined by

$$f_0 = \frac{1}{2\pi C_t R_e} \tag{13.5}$$

Later, when considering the correction of the amplifier at high frequencies, the reason for this will become apparent. The magnitude and phase angle of the response in terms of this critical frequency are

$$G = \frac{R_e g_m}{\sqrt{1 + (f/f_0)^2}} \tag{13.6}$$

$$\phi = -\arctan (f/f_0) \tag{13.6a}$$

These are shown plotted as a function of frequency in Fig. 13.6.

At very low frequencies the impedance of the coupling condenser C cannot be considered negligible, whereas the shunt load impedance due

to the capacity of the tube and wiring becomes so high that it can be neglected. The equivalent circuit for this case will, therefore, be that given in Fig. 13.7. The output voltage of this circuit will be

$$e_2 = e_1 g_m R_a \frac{R_g}{(R_g + R_a) - jX_c}$$

where

$$R_a = \frac{Rr_p}{R + r_p}$$

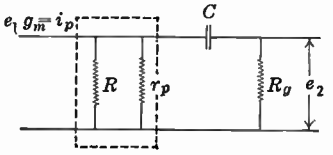


Fig. 13.7. Low-Frequency Equivalent Circuit of a Resistance-Coupled Stage.

Since R_a, in this type of amplifier, is small compared with R_g,

$$e_2 = e_1 g_m R_g R_a \left(\frac{R_g + jX_c}{R_g^2 + X_c^2} \right) \tag{13.7}$$

The magnitude of the gain is, therefore,

$$G = g_m R_g R_a \frac{1}{\sqrt{R_g^2 + X_c^2}} \tag{13.8}$$

and the phase angle is

$$\phi = \arctan \frac{X_c}{R_g} \tag{13.9}$$

where $X_c = 1/(2\pi fC)$. These two functions are shown in Fig. 13.8.

It is clear from the foregoing that it may be necessary to correct both the low-frequency and the high-frequency response of this type of

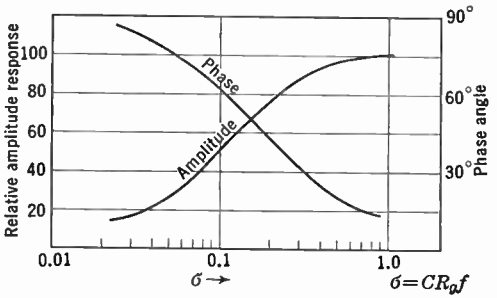


Fig. 13.8. Response of Coupling Shown in Fig. 13.7.

amplifier if it is to be used for a very wide frequency band. The extent to which this correction is required depends on the characteristics of available amplifier tubes.

13.5 Vacuum Tubes Suitable for the Video Amplifier. It is evident from Eqs. 13.5 and 13.6 that the most important considerations in the choice of suitable tubes for the video amplifier are the transconductance and the capacity of the tube. An approximate criterion on which to base judgment of the merit of a tube for this purpose is the ratio of transconductance to capacity. However, this criterion must be used with caution because the wiring capacity is added to the tube capacity, which decreases the dependence of gain on tube capacity alone.

Table 13.1 lists a few of the available types of tubes in order to estimate their performance in a resistance-coupled amplifier. Values

TABLE 13.1. AMPLIFIER TUBE DATA

Tube Type	Capacity, micromicrofarads			Transconductance, micromhos	Plate Resistance, ohms
	Grid-Ground	Plate-Ground	Grid-Plate		
6C5	3.0	13	2.0	2,000	10,000
6C6	5.0	6.5	...	1,200	1,000,000
954	3.4	3.0	...	1,400	1,500,000
955	1.0	0.6	1.4	2,000	12,500
6AK5	4.0	2.8	...	5,100	690,000
6AG5	6.5	1.8	...	5,000	800,000
6AB7	8.0	5	...	5,000	700,000
6AC7	11	5	...	9,000	1,000,000
6AG7	13	7.5	...	11,000	130,000
6AU6	5.5	5.0	...	5,200	1,000,000
6BQ7(½)	2.85	1.35	1.15	6,000	5,800
6CB6	6.3	1.9	...	6,200	600,000
6CL6	11	5.5	...	11,000	150,000

are given for the grid-to-cathode, plate-to-ground, and grid-to-plate capacities, and also for transconductance and plate resistance. The first tube listed is a triode. The total input capacity is, therefore, the grid-to-ground capacity plus the grid-to-plate capacity times the voltage gain of the tube. If the gain of this stage is large, the input capacity is high owing to the reflected capacity. This fact excludes in general the use of triodes in video amplifiers. An exception to this statement occurs in the first stage of a video amplifier, where, from the standpoint of signal-to-noise ratio, it may be desirable to use a triode. This exception will be discussed in greater detail in a later section.

The second tube, a pentode, is much more satisfactory. The total shunt capacity due to the tube electrodes is 11.5 micromicrofarads.

This leads to a figure of merit of 1200/11.5, or, approximately, 100. The next tube, an acorn pentode, played an important role in earlier camera amplifiers. Its figure of merit is about 220. There follows an acorn triode, the RCA 955. Here again the high reflected capacity limits its applicability in this type of amplifier. However, it can fulfill certain useful functions because of its very low static capacity. The remaining tubes listed were developed, largely, for television purposes. They are designed to meet the requirements of wide-band amplification. The figures of merit of these tubes, as determined from the data in Table 13.1, are

6AK5	750
6AG5	600
6AB7	380
6AC7	560
6AG7	540
6AU6	500
6BQ7	135
6CB6	760
6CL6	660

The calculated gain of a single stage will serve to indicate the performance of various members of this group of tubes. For this purpose Eq. 13.6 is used, and a frequency is assumed such that f/f_0 is small compared to 1. The gain under these conditions will be merely $R_e g_m$, where

$$R_e = \frac{1}{2\pi C_t f_0}, \quad \text{so that} \quad G = \frac{g_m}{2\pi C_t f_0}$$

Furthermore, a wiring capacity of 5 micromicrofarads has been added to the tube capacity to obtain the total capacity C_t. The results are tabulated in the following table, assuming f_0 to be 4.5 megacycles. The

TABLE 13.2. CALCULATED GAIN OF AMPLIFIER STAGE

Type	C_t	R_e	Gain
6C6	16.5 $\mu\mu$f	2100	2.6
6AK5	11.8 $\mu\mu$f	3000	15.3
6AC7	21 $\mu\mu$f	1700	15.1
6AU6	15.5 $\mu\mu$f	2300	11.9

superiority of the special television tubes is very apparent; they have, accordingly, replaced all other types for the low-level stages of video amplifiers.

13.6 High-Frequency Correction. The gain per stage of a simple resistance-coupled amplifier is given by Eq. 13.6. This equation shows that the gain has dropped to 70 percent of its intermediate-frequency value by the time the applied frequency has reached the characteristic frequency f_0. An amplifier with this characteristic would be serviceable only over a frequency range well below f_0. In order to accommodate the wide band required by the video signal, f_0 must, therefore, be very high. However, f_0 can be increased only by decreasing the coupling resistor R, since the shunt capacity is fixed. This, of course,

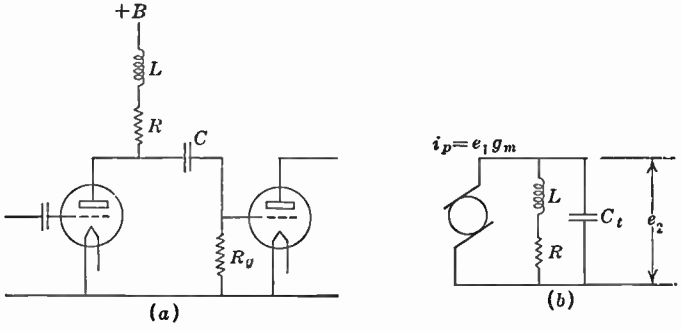


Fig. 13.9. Response Correction by the Addition of an Inductance to a Resistance-Coupled Stage.

decreases the intermediate-frequency gain. It is obviously desirable to compensate the stage in order to improve its response at high frequencies.

The inclusion of a small inductance in series with a load resistance R is one of the simplest methods of obtaining partial correction of the response. This circuit, together with its high-frequency equivalent, is given in Figs. 13.9*a* and 13.9*b*. The voltage amplification for this stage is

$$G = \frac{e_2}{e_1} = g_m |Z| \tag{13.10}$$

where Z is the impedance of the coupling net at the frequency f. By ordinary circuit calculation methods this impedance is found to be

$$Z = \frac{\dfrac{R}{2\pi f C_t} + j\left(\dfrac{L}{C_t} - 4\pi^2 f^2 L^2 - R^2\right)}{2\pi f C_t \left[R^2 + \left(2\pi f L - \dfrac{1}{2\pi f C_t}\right)^2\right]} \tag{13.10a}$$

The absolute value of the gain is, therefore,

$$G = g_m \frac{\sqrt{\frac{R^2}{4\pi^2 f^2 C_t^2} + \left(\frac{L}{C_t} - 4\pi^2 f^2 L^2 - R^2\right)^2}}{2\pi f C_t \left[R^2 + \left(2\pi f L - \frac{1}{2\pi f C_t}\right)^2\right]} \tag{13.11}$$

whereas the phase angle is given by

$$\phi = \arctan \frac{L/C_t - 4\pi^2 f^2 L^2 - R^2}{R/(2\pi f C_t)} \tag{13.11a}$$

At intermediate frequencies such that $1/(2\pi f C_t) \gg R$ and $R \gg 2\pi f L$, the gain becomes

$$G_i = g_m R$$

As expressed by Eq. 13.11, the effect of the various circuit constants on the gain is not immediately apparent. A simplification can be effected by the substitution

$$f_0 = \frac{1}{2\pi C_t R}, \qquad K = \frac{L}{C_t R^2}$$

In terms of f_0, K, and G_i, Eqs. 13.11 and 13.11a yield for the gain as function of frequency

$$\frac{G}{G_i} = \sqrt{\frac{K^2\left(\frac{f}{f_0}\right)^2 + 1}{K^2\left(\frac{f}{f_0}\right)^4 - (2K - 1)\left(\frac{f}{f_0}\right)^2 + 1}} \tag{13.12}$$

and for the phase angle

$$\phi = \arctan \frac{f}{f_0}\left[(K - 1) - K^2 \frac{f^2}{f_0^2}\right] \tag{13.12a}$$

The value of K giving a gain ratio of unity at $f = f_0$ can be seen from Eq. 13.12 to be $K = 0.5$. This value is obtained by using the following circuit constants, as derived from the definitions of f_0 and K:

$$R = \frac{1}{2\pi f_0 C_t}, \qquad L = \frac{1}{8\pi^2 f_0^2 C_t}$$

The response under these conditions is not flat but rises to a value somewhat greater than unity at frequencies slightly below f_0 and then falls rather rapidly. The time delay is not constant, which is a more serious fault than the slight rise in the response.

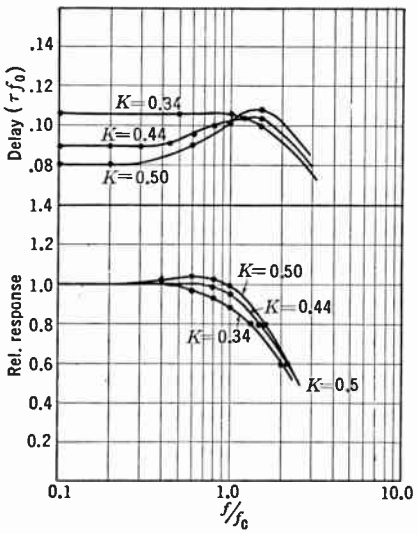


Fig. 13.10. Families of Response Curves for Coupling Circuit Shown in Fig. 13.9.

More favorable response characteristics can be obtained with other circuit constants. The most nearly flat response is to be had if $K = 0.41$; $K = 0.32$ gives the most nearly constant time delay. In practice, a compromise value of K is generally used. Figure 13.10 gives a family of response curves corresponding to various K values.

13.7 The General Coupling Network. The resistance-coupled amplifier stage with a correcting element added to it is a special case of the general two-terminal coupling network. Although to date only a comparatively few members of this class of networks have been thoroughly examined, nevertheless valuable results have already been obtained and interesting generalizations have been reached.

An amplifier stage with a generalized two-terminal circuit network is shown in Figs. 13.11*a* and 13.11*b*. One limitation imposed on this network is that the tube capacities, C_t, must appear across the input of the circuit as shown in Fig. 13.11*c*. H. A. Wheeler * has shown, without rigorous proof, that, if Z is the most general two-terminal filter possible, the response giving a flat characteristic over a bandwidth from a very low frequency to the characteristic frequency f_0 is limited to $G = g_m/(\pi f_0 C_t)$. Actually, this response cannot be fully realized in practice.

The response of the general circuit (Fig. 13.11*a*) is, of course,

$$G = g_m Z(f)$$

and its constancy is determined by the constancy of the impedance $Z(f)$. The general conditions of constancy for a given network can be

* See Wheeler, reference 11.

easily determined.* The square of the impedance is expressed as a ratio of power series of the frequency, as shown:

$$|Z(f)|^2 = |Z(0)|^2 \frac{1 + a_1 f^2 + a_2 f^4 + \cdots}{1 + b_1 f^2 + b_2 f^2 + b_3 f^6 + \cdots} \tag{13.13}$$

If now the coefficients of terms of like powers in the numerator and denominator are equal, the response is flat. Similarly, the derivative of the phase response can be expressed as a ratio of two polynomials,

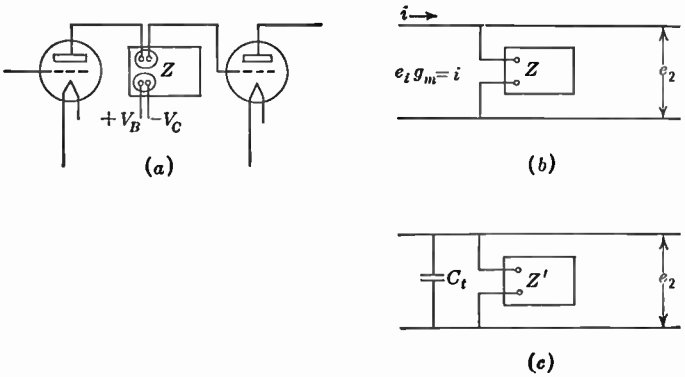


Fig. 13.11. Two-Terminal Network as Amplifier Coupling.

and a correspondence of coefficients of like terms is the criterion for a constant time delay.†

Equation 13.12 describes the amplitude response of the circuit discussed in section 13.6 by a fraction of the type shown in Eq. 13.13. The fact that the denominator contains $(f/f_0)^4$, whereas the numerator does not, shows that this type of coupling net cannot be made truly flat. However, for the range of frequencies over which $(f/f_0)^4$ may be neglected in comparison to $(f/f_0)^2$, the amplifier response will be constant if

$$K^2 = 1 - 2K$$

or

$$K = 0.414$$

* G. V. Braude showed that, if the impedance is developed as a Taylor's series and the coefficients are equated to zero, the response is flat. Tellegen and Verbeck originated the simpler method described, and showed it to be equivalent to Braude's method. *J. Tech. Physics, U.S.S.R.*, Vol. 4, Nos. 9–10, 1934.

† See Herold, reference 10.

If this value of K is introduced into Eq. 13.12, the amplitude response at $f/f_0 = 1$ is found to be 93 percent.

The phase angle for this circuit was given by Eq. 13.12*a*, which is

$$\phi = -\arctan\left\{\left(\frac{f}{f_0}\right)\left[1 - K + K^2\left(\frac{f}{f_0}\right)^2\right]\right\}$$

Differentiating Eq. 13.12*a* and equating coefficients, the condition for constant delay is fulfilled when

$$K^3 + 3K - 1 = 0$$

$$K \cong 0.32$$

It has been estimated that, if the amplitude response at the maximum frequency in the video signal does not fall below 90 percent of the

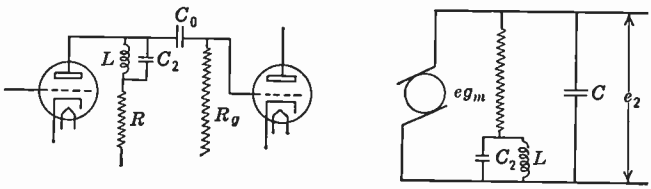


Fig. 13.12. Coupling Network, Including Inductance and Capacity for Response Correction.

intermediate-frequency gain, it may be considered satisfactory. If the value $K = 0.37$ is used, the amplitude response at $f/f_0 = 1$ will be about 90 percent of that at intermediate frequencies, and the departure from constancy of time delay over this region will be $0.004/f_0$. This is to be compared with a drop of the amplitude response to 90 percent at $f/f_0 = 0.48$ accompanied by $0.005/f_0$ delay error for the circuit without the inductance (Fig. 13.4).

A condenser C_2 placed across the inductance, as shown in Fig. 13.12, increases the frequency range over which the gain remains above 90 percent. By the methods described above, it can be shown * that, for $K = 0.38$ and $C_2 = 0.30C$, the range of frequency is $1.25f_0$ and the time delay variation $0.008/f_0$.

A four-terminal network, as shown in Fig. 13.13, may also be used as interstage coupling. An advantage is gained in this case in that the tube capacities are not lumped but are at the two ends of the net.

* See Herold, reference 10.

Wheeler has shown by very general methods that a maximum amplitude response of

$$G = \frac{g_m}{\pi f_0 \sqrt{C_p C_g}} \tag{13.14}$$

is obtainable with a flat characteristic when the impedances at the two ends of the net are properly matched to the respective capacities.

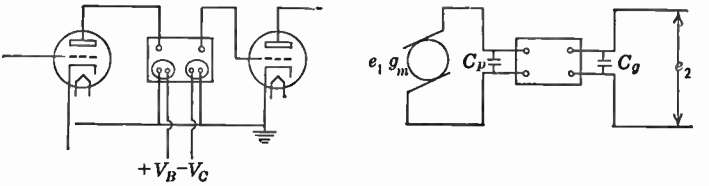


Fig. 13.13. Four-Terminal Network as Amplifier Coupling.

A π network with an added inductance * in series with the load resistor, as shown in Fig. 13.14, represents a typical four-terminal coupling network. The circuit constants giving flat response are $L_1 = 0.12CR^2$, $L_2 = 0.52CR^2$, $C_p = 0.34C$, and $C_g = 0.66C$. With these constants, the response does not drop below 90 percent up to $f/f_0 = 1.8$, and the delay error is limited to $0.009f_0$ in the same range. Circuits similar to that shown in Fig. 13.14, eventually with a resistance

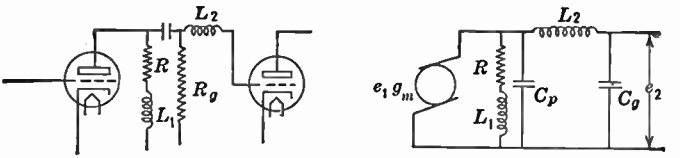


Fig. 13.14. Simple Embodiment of the Four-Terminal Network as a Coupling Means.

large compared to the load resistance shunting the series inductance, find extensive application in video amplifiers.

The problem of interstage coupling from the standpoint of two-terminal and four-terminal filter theory has been investigated by Wheeler.† Considered from this point of view, the two-terminal coupling becomes a low-pass filter whose input impedance determines its effect on the response. A filter of constant-k sections, concluded

* See Braude, reference on p. 499, and Herold, reference 10.

† See Wheeler, reference 11.

with an m-derived half section and giving an image impedance approximately matching the terminal resistor R (which in this case is

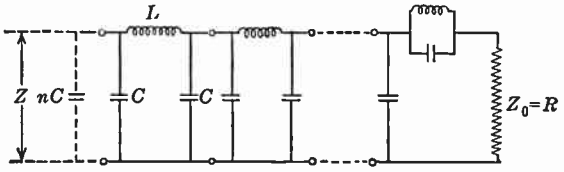


Fig. 13.15. Constant-k Network.

the external plate resistor), is shown in Fig. 13.15. Such a filter, if made up of a large number of sections, has an impedance given by

$$|Z| = \frac{R}{\sqrt{1 - (1 - n^2)\left(\frac{f}{f_0}\right)^2}} \tag{13.15}$$

and a phase angle

$$\phi = \arctan \frac{n\frac{f}{f_0}}{\sqrt{1 - \left(\frac{f}{f_0}\right)^2}} \tag{13.15a}$$

where f_0 is $1/(2\pi RC)$. The total tube capacity in this case is $(1 + n)C = C_t$. The impedance for various values of n is shown in Fig. 13.16 together with the phase angle. The amplitude response becomes perfectly flat when n is equal to unity, and under these conditions the gain is

$$G = g_m R = \frac{2g_m}{2\pi f_0 C_t}$$

Under conditions of maximum constant gain, the time delay is not constant. In order to maintain as nearly uniform delay as possible, n must be somewhat greater than 1, and, as is apparent from the form of the expression for Z, the response is not constant under these conditions.

It has also been shown that the three two-terminal coupling circuits which have been described above are elementary forms of this type of filter. With optimum circuit constants the circuit shown in Fig. 13.4 has an approximately uniform gain of 20 percent of that of the ideal

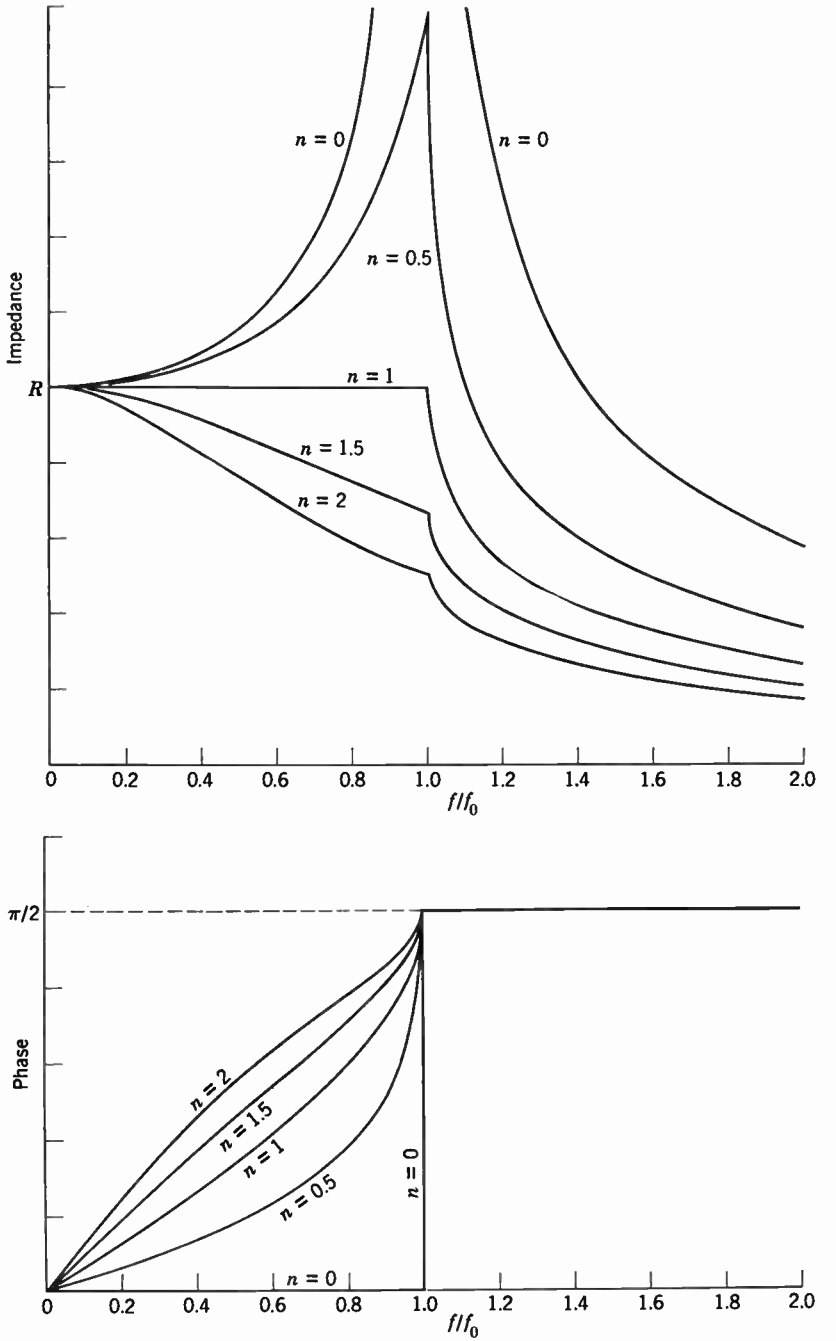


Fig. 13.16. Characteristics of Network Shown in Fig. 13.15.

two-terminal network. The circuit of Fig. 13.9 is a single constant-k half section and gives 55 percent of the ideal response. Figure 13.12 is an m-derived section and is capable of yielding 70 percent. It is interesting to note that a circuit which includes a constant-k half section and an m-derived half section will give a gain which is 95 percent of the theoretical maximum.

Another approach toward obtaining large bandwidth with tubes of given characteristics employs "distributed amplification," discussed briefly in Chapter 15, p. 654. Here the input and output circuits form parallel low-pass filter chains with equal time delay per section, joined by one amplifier tube per section; the filter capacities are provided by the tube capacities. In this manner gain can be obtained at arbitrarily large bandwidths, simply by the addition of tubes and filter sections, as long as the attenuation in a filter section is less than the signal enhancement by an amplifier stage. However, the requirements of the video amplifier are generally sufficiently modest to make this refinement unnecessary.

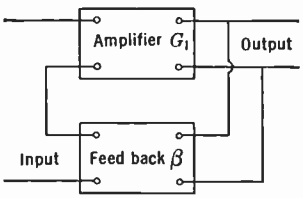


Fig. 13.17. Block Diagram of Feedback Amplifier.

Although high theoretical efficiencies should be obtainable with these more complex circuits, and are in fact realized, they are limited by factors such as the distributed capacity of coils and unwanted resistance losses. Four-terminal networks of the type shown in Fig. 13.14, and eventually of even greater complexity, find wide employment in video amplifiers.

Negative feedback provides a more general method of improving amplifier response. This type of feedback consists of returning the output voltage to the grid through a suitable network, in such a way that it opposes the input voltage. Figure 13.17 illustrates such a circuit schematically. If it is assumed that the gain of the amplifier is G_1 and the attenuation of the return network is β (either G_1 or β, or both, may be complex), the following relation exists between output voltage e_2 and input voltage e_1:

$$e_2 = G_1(e_1 - \beta e_2)$$

The net gain of the amplifier with feedback is, therefore,

$$\frac{e_2}{e_1} = G = \frac{G_1}{1 + G_1\beta} \tag{13.16}$$

If, now, the voltage feedback is large compared with the input voltage, or, in other words, if the term $G_1\beta$ is large compared with 1, the net gain becomes

$$G \cong 1/\beta$$

Thus the effective gain becomes independent of the properties of the amplifier and is determined solely by the characteristics of the return network. For example, if an amplifier has a gain of 2000 at intermediate frequencies, and 1000 at the maximum frequency, and the loop has an attenuation of 0.01 independent of frequency, the net gain would be

$$G = \frac{2000}{1 + 20} = 95 \text{ at intermediate frequencies}$$

and

$$G = \frac{1000}{1 + 10} = 91 \text{ at maximum frequency}$$

The effective gain is therefore nearly constant in spite of the 2-to-1 change in amplifier response.

Actually, because of the difficulty in stabilizing such an amplifier over the wide frequency band required for the video signal, it is used in practice only in relatively simple forms. A frequently encountered simple feedback circuit is shown in Fig. 13.18. The feedback here cannot be made sufficient to render the net gain independent of the amplifier characteristics; consequently, the return network in the cathode circuits is not made flat but is given properties compensating for the errors in the amplifier. The return net shown consists of a resistor shunted with a small condenser, making the negative feedback at high frequencies less than at low, compensating for the falling gain of the amplifier.

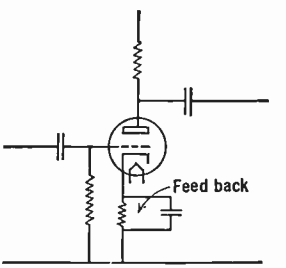


Fig. 13.18. Simple Feedback Circuit for Correcting Frequency Response.

13.8 Low-Frequency Response. The television amplifier must not only respond to high frequencies, but must also be capable of amplifying signals down to below frame frequency of the picture being transmitted. The frequency range should theoretically extend to zero frequency if pictures of moving objects are to be perfectly transmitted. However, if an amplifier is good to a little below repetition frequency and very low frequencies corresponding to the average light are re-

inserted by methods which will be described later, the error introduced is negligible.

According to Eqs. 13.8 and 13.9 the expressions * for the response of a resistance-coupled stage at low frequencies are

$$G = g_m R_g R \frac{1}{\sqrt{R_g^2 + X_C^2}} = \frac{R g_m}{\sqrt{1 + (X_C^2/R_g)}}$$

$$\phi = \arctan \frac{X_C}{R_g}$$

$$X_C = \frac{1}{2\pi f C}, \qquad R \ll r_p \text{ and } R_g$$

where R and R_g are the plate and grid resistors and C is the coupling capacity.

If the choice of C and R_g were not limited, the gain could be made constant down to any desired frequency. However, the choice of R_g is restricted by the grid characteristics of the tube in question, and in

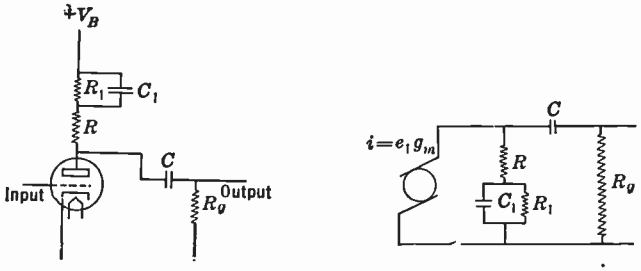


Fig. 13.19. Circuit for Improving Low-Frequency Response.

general cannot exceed about 500,000 ohms. The coupling condenser is limited by the fact that the physical size of a large capacitor introduces a large shunt capacity between grid and ground, and thus, as has been shown, has a very adverse effect upon the high-frequency response. A practical limit to the value of the coupling condenser may be taken as 0.1 microfarad.

Assuming a frame frequency of 30 cycles, the impedance of a 0.1-microfarad coupling condenser will be about 50,000 ohms. This means that the gain will be dropped one-half percent of its value for intermediate frequencies. The phase delay, however, will be quite

* These apply equally to the circuits discussed in sections 13.6 and 13.7 with the exception of those employing negative feedback.

large, being about $5 \cdot 10^{-4}$ second. In other words, a sine wave of this frequency impressed upon the grid will be delayed nearly 3 percent of the vertical distance on the screen for each stage. This delay can, to some extent, be corrected by the filter shown in Fig. 13.19.

The correction factor to be applied to the gain as a result of the circuit illustrated is

$$K = \frac{\{[RR_1^2 + X_{c1}^2(R + R_1)]^2 + (X_{c1}R_1^2)^2\}^{\frac{1}{2}}}{RR_1^2 + X_{c1}^2R} \tag{13.17}$$

where $X_{c1} = 1/(2\pi fC_1)$ and the term

$$\phi' = -\arctan \frac{X_{c1}R_1^2}{RR_1^2 + X_{c1}^2(R + R_1)} \tag{13.18}$$

is added to the phase angle.

As before, R is determined by the high-frequency response desired. This expedient does not yield complete correction, but, by properly

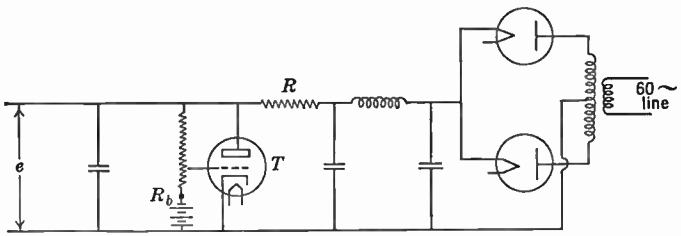


Fig. 13.20. Regulated Voltage Supply.

choosing R_1 and C_1, the response at frame frequency can be made sufficiently good, from the standpoint of both phase and gain, for practical picture work.

Since this type of amplifier is made to respond to low frequencies, great care must be taken to decouple the common high-voltage supply, so as to avoid low-frequency relaxation oscillations or motor boating. It is often found necessary to place a condenser of 1000 or more microfarads across the plate supply of such an amplifier.

Another method used, either alone or in conjunction with a large condenser, to lower the impedance of the voltage supply, is a vacuum-tube voltage regulator. One form of this regulator is shown in Fig. 13.20. The regulator tube T is arranged so that, as its plate current increases, the IR drop across the resistance R decreases the output

voltage e. The grid potential with respect to the cathode is determined by the potential across R_b of the voltage divider. If now, for any reason, e tends to increase, the grid of tube T becomes more positive, causing the plate current to increase, which produces a drop along R opposing the increase of e. Greater sensitivity can be obtained if the single tube T is replaced by several stages of amplification. This type of constant-voltage plate supply is very useful for television work.

13.9 Overall Amplifier Response. The response of only a single stage of amplification has been considered so far. In practice, the com-

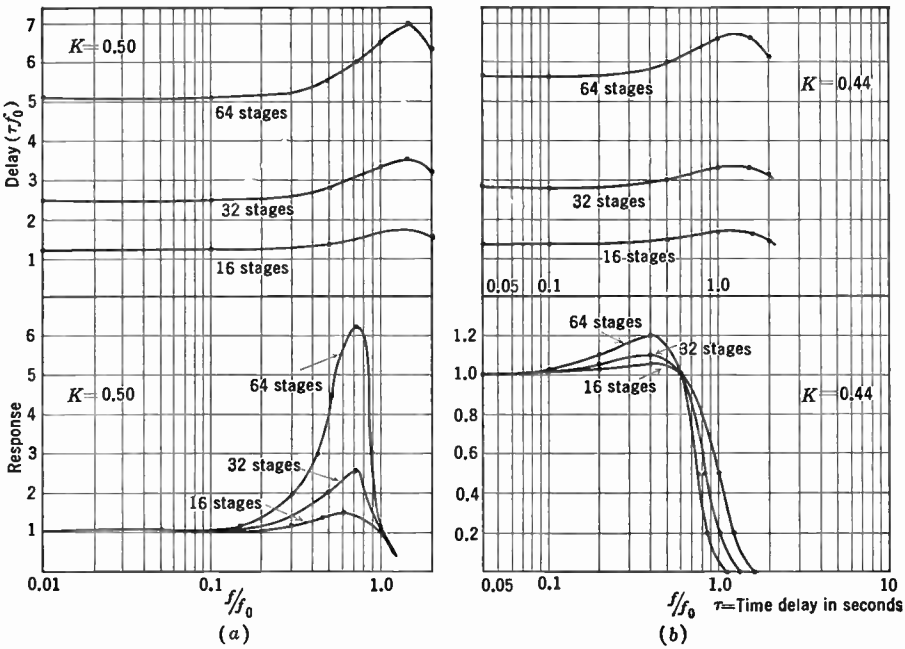


Fig. 13.21. Response of Multi-Stage Amplifiers.

plete amplifier consists of a great many cascaded stages, and the overall response is the important consideration. If the amplifier consists of identical stages and there is no accidental regeneration resulting from improper shielding, the overall gain will be the gain of a single stage at each frequency raised to a power equal to the number of stages, whereas the time delay will be the product of the delay for each stage and the number of stages. The amplitude and delay curves * for 16,

* See Bedford and Fredendall, reference 12.

32, and 64 stages of amplification coupled as shown in Fig. 13.9, with $K = 0.5$ and $K = 0.44$, are given in Fig. 13.21.

It is often advantageous to employ stages which are not identical but differ in such a way that they tend to mutually compensate their individual defects. Thus, a filter-coupled stage which has a flat response but a positive time delay error over part of the useful frequency band can be used with another constant gain stage which has a negative delay error over the same portion of the band. If the characteristics of the individual stages differ, the overall gain is the product of the gains at each frequency, and the total delay error is the algebraic sum of the individual errors.

In amplifiers having a great many stages, as at the pickup, monitoring, and modulating units of a transmitter, it was formerly the practice to use the complex coupling circuits required for high gain only in the initial stages, that is, up to the point where the signal level was well above the level of the noise generated by the amplifying tubes and coupling circuits. The remaining stages were simple wide-band stages of low gain, requiring little frequency compensation, simplifying the problem of maintaining constant gain and delay characteristics for the amplifier as a whole. Improvements in design and circuit components make it practical now to employ high-gain stages throughout, reducing the total number.

13.10 Frequency Response Characteristics and Picture Quality. The analysis of the various amplifier circuits discussed in this chapter yields response characteristics in terms of amplitude and phase angle as function of frequency. It is difficult even after considerable experience to estimate directly from such response curves the quality of the picture that will be reproduced. Consequently, it is desirable to express the response characteristic in a form which can be more conveniently interpreted in terms of picture quality.

The response of an amplifier to a voltage step is a very convenient indication of its behavior as a picture amplifier. The application of the gain and phase curves to determine the corresponding transient response involves the use of a Fourier integral. Furthermore, here, as elsewhere, the treatment is greatly simplified by considering high- and low-frequency response separately.

The high-frequency transient response requires a knowledge of only the upper portion of the gain and phase curves, and assumes that from the intermediate frequencies down to zero frequency the gain and delay

are constant. On the other hand, the low-frequency analysis assumes that the upper end of the response spectrum is perfect.

Let it be assumed that the impulse shown in Fig. 13.22*a* is applied to the input of the amplifier. For convenience, the height of the voltage step is taken as unity. Therefore, the input signal can be expressed as

$$f_1(t) = 0, \qquad -\infty < t < 0$$

$$f_1(t) = 1, \qquad 0 < t < \infty$$

The Fourier representation of this function is

$$V_0 = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \frac{\sin 2\pi ft}{f}\,df$$

This equation can be interpreted to mean that the frequency spectrum of the voltage impulse applied to the input of the amplifier consists of

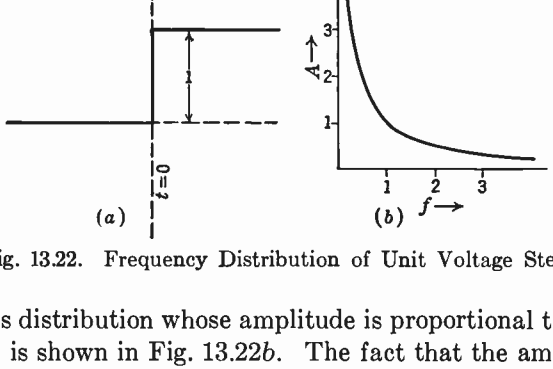


Fig. 13.22. Frequency Distribution of Unit Voltage Step.

a continuous distribution whose amplitude is proportional to $1/f$. This distribution is shown in Fig. 13.22*b*. The fact that the amplitudes become infinite as f approaches zero causes no difficulty when a-c amplifiers are considered. As far as the low-frequency portions of the response characteristics are concerned, the gain of the amplifier becomes zero before zero frequency is reached. For the high-frequency analysis this portion of the spectrum is unimportant.

After passing through the amplifier the frequencies represented in the spectrum will be amplified by amounts which can be determined from the gain curves and will also be delayed by known phase angles. Consequently the output frequency spectrum will not be identical with the input spectrum. It will, however, represent some form of voltage impulse. The Fourier integral giving the form of this impulse is

$$V = \frac{A(0)}{2} + \frac{1}{\pi}\int_0^\infty \frac{A(f)\cos\phi(f)}{f}\sin 2\pi ft\,df + \frac{1}{\pi}\int_0^\infty \frac{A(f)\sin\phi(f)}{f}\cos 2\pi ft\,df \quad (13.19)$$

where $A(f)$ is the gain of the amplifier as a function of frequency and $\phi(f)$ is the phase shift.

In view of the oscillatory character of the integrands, a numerical evaluation of the integrals in Eq. 13.19 is extremely laborious. As pointed out by A. V. Bedford and G. L. Fredendall,* the desired information can be obtained much more simply by considering the response of the amplifier to a square wave with a fundamental period longer than the time during which the response of the amplifier to a step voltage deviates appreciably from the final value. Empirically, the highest frequency f_p for which the gain and phase delay are sensibly constant may be employed as square-wave frequency for any real amplifier. The Fourier series for the square wave has the form

$$V_0 = \frac{1}{2} + \frac{2}{\pi}\left\{\sin 2\pi f_p t + \frac{1}{3}\sin 6\pi f_p t + \frac{1}{5}\sin 10\pi f_p t + \cdots\right\} \quad (13.20)$$

so that the response of the amplifier to the square wave may be written, employing the same notation as for Eq. 13.19:

$$V = \frac{A(0)}{2} + \frac{2}{\pi}\{A(f_p)\sin[2\pi f_p t + \phi(f_p)] + \tfrac{1}{3}A(3f_p)\sin[6\pi f_p t + \phi(3f_p)] + \tfrac{1}{5}A(5f_p)\sin[10\pi f_p t + \phi(5f_p)] + \cdots\} \quad (13.21)$$

$$[A(0) = A(f_p)]$$

Since f_p is chosen so as to lie reasonably close to the upper boundary of the transmission band of the amplifier, a relatively small number of terms will generally suffice to give a faithful representation of the response to the square wave and, hence, to the voltage step represented by the rise of the square wave.

Figure 13.23 shows the response of 16, 32, and 64 amplifier stages of the type shown in Fig. 13.9 to a voltage step, calculated by Bedford and Fredendall by utilizing a square wave with a fundamental frequency $0.44/(2\pi\sqrt{LC_t})$. Figure 13.23*a* shows the transient response curves for the three amplifiers having the frequency characteristics

* See Bedford and Fredendall, reference 12.

given in Fig. 13.21*a*, where $K = 0.5$; Fig. 13.23*b* shows similar curves for the response given in Fig. 13.21*b* for $K = 0.44$. It is immediately apparent that, if the rest of the system is assumed to have an ideal response, the amplifier producing the curves in Fig. 13.23*a* cannot be used because of the extreme overshoot which would result in a dark line following every change from black to white in the picture, or a

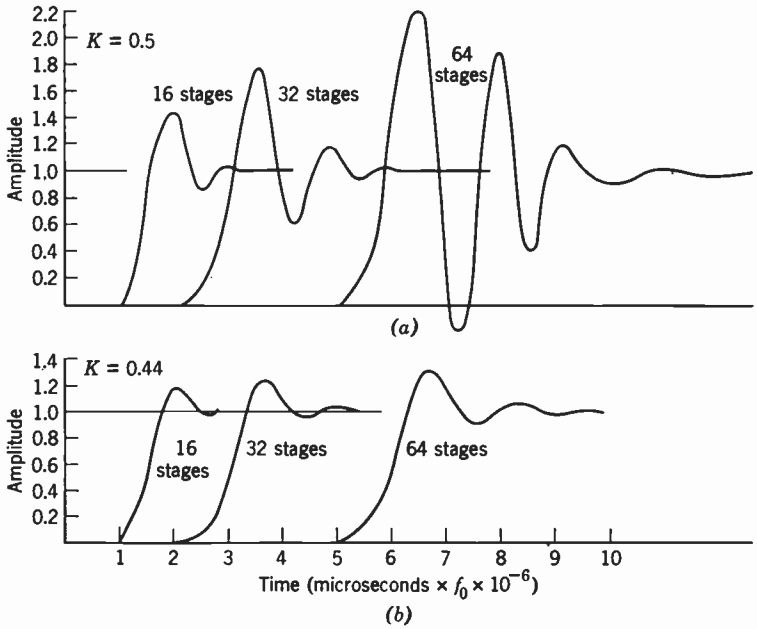


Fig. 13.23. Transient Response to Unit Step of Amplifiers with Characteristics Shown in Fig. 13.21.

white line following a change in the opposite direction. The transient response shown in Fig. 13.23*b* will, however, be usable since the overshoot is only about 10 percent, which is not objectionable. The "region of blurring" is greater for the second amplifier.

As will be seen in the next section, the analysis of the transient response to a unit voltage step may be extended to include the low-frequency portion of the amplifier spectrum. However, from the point of view of picture reproduction the response of the amplifier to a square wave of field frequency (representing a stationary horizontal bar) or of slightly less than field frequency (a horizontal bar moving in the vertical scanning direction) yields the needed information more di-

rectly. If A_i and D_i are the gain and time delay, respectively, at intermediate frequencies, the response may conveniently be written in the form

$$V(t) = A_i V_0(t - D_i) - \sum_{n=0}^{\infty} \frac{2}{(2n+1)\pi} B_{2n+1} \sin\left(2\pi[2n+1]f_p t - \theta_{2n+1}\right) \quad (13.22)$$

with

$$B_n = \{[A(nf_p)\cos\phi(nf_p) - A_i\cos(2\pi nf_p D_i)]^2 + [A(nf_p)\sin\phi(nf_p) + A_i\sin(2\pi nf_p D_i)]^2\}^{\frac{1}{2}}$$

and

$$\theta_n = \arctan\frac{-A(nf_p)\sin\phi(nf_p) - A_i\sin(2\pi nf_p D_i)}{A(nf_p)\cos\phi(nf_p) - A_i\cos(2\pi nf_p D_i)}$$

The first term represents the undistorted response of the amplifier, the sum, the correction required by the fact that the actual response is

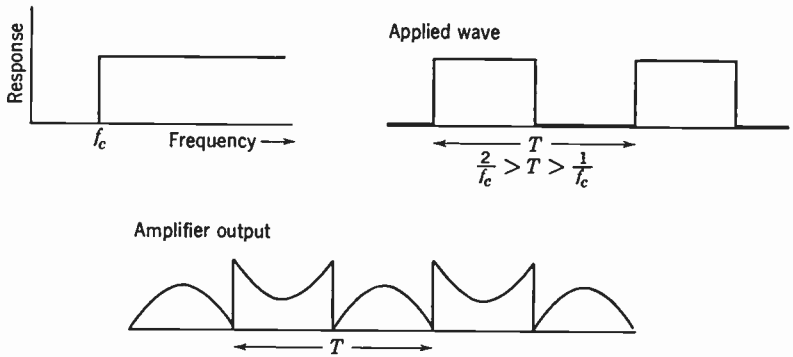


Fig. 13.24. Effect of Sharp Low-Frequency Cutoff on the Reproduction of a Square Wave with Fundamental Frequency below Cutoff Frequency.

nonuniform at low frequencies. The higher terms of the sum vanish since, by assumption, the gain and delay are constant for intermediate and high frequencies. For infinitely sharp cutoff at the frequency f_0, $f_p < f_0 < 2f_p$, the correction consists of a superposed sine wave $-\frac{2A_i}{\pi}$ $\sin(2\pi f_p(t - D_i))$. This is shown in Fig. 13.24. In any true amplifier there is a gradual variation of both gain and phase delay with frequency, and the spurious shading effects are much less drastic. Thus Fig. 13.25 shows the response of a simple uncorrected network with a 500,000-

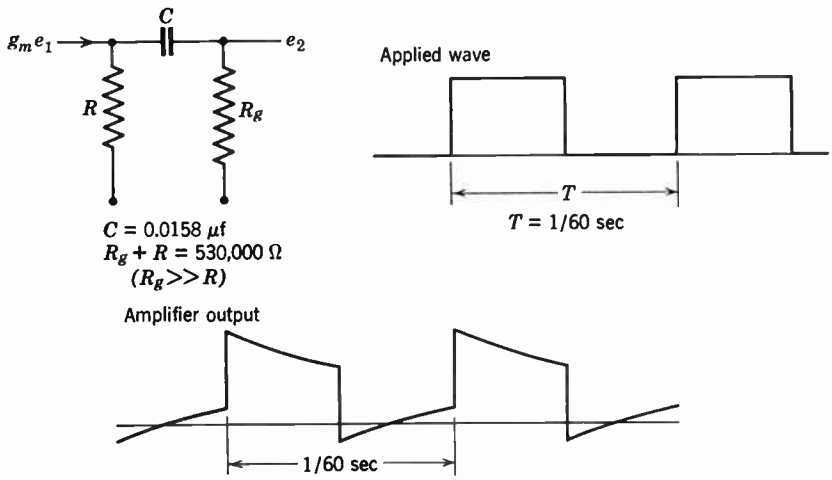


Fig. 13.25. Response of Uncorrected Coupling Network to Square Wave at Field Frequency.

ohm grid resistance and a 0.016-microfarad blocking capacitance to a 60-cycle square wave. Equation 13.22 yields in this special case

$$A_i = \frac{g_m R R_g}{R_g + R}, \qquad D_i = 0$$

$$B_n = \frac{A_i}{\sqrt{4\pi^2 n^2 f_p^2 C^2 (R_g + R)^2 + 1}}$$

$$= \frac{A_i}{\sqrt{10n^2 + 1}} : \qquad \theta = \arctan 2\pi n f_p C(R_g + R)$$
$$= \arctan 3.16n$$

so that

$$V(t) = A_i\{V_0(t) + 0.192 \cos (120\pi t + 17°32') + 0.022 \cos (360\pi t + 6°01') + 0.008 \cos (600\pi t + 3°37') + 0.004 \cos (840\pi t + 2°35') + \cdots\}$$

The very low coupling capacitance has been chosen to demonstrate the effect more clearly. With, e.g., 0.1 microfarad the first term would have been only $0.05 \cos(120\pi t + 2°52')$. If amplitude and phase response are corrected in the manner described in section 13.8, the spurious shading is further reduced. Since gain and phase characteristics of real amplifiers vary smoothly with frequency, an amplifier

which gives undistorted response down to frame frequency will have reasonably good response also for somewhat lower frequencies and thus reproduce moving, as well as stationary, horizontal bars without undue shading.

13.11 Amplifier Response by Laplace Transformation. In the preceding section a convenient method has been outlined for inferring the response of an amplifier to a unit voltage step from its reaction to a square wave of suitable fundamental frequency. In this manner a close approximation to the step response can be calculated simply and rapidly for an amplifier of arbitrarily many stages, provided that its gain and delay are known as functions of frequency. The method of the Laplace transformation is usually somewhat less simple to apply, particularly for an amplifier of many stages; on the other hand, it yields an exact expression of the step response, in analytical form. The Laplace transformation makes it possible to reduce the solution of the differential equations governing the currents and voltages in a coupling network resulting from a step input to a series of algebraic operations similar to those required to calculate the stationary response of the network to a sinusoidal input. Furthermore, since an arbitrary transient can be constructed by the superposition of step functions, the method of the Laplace transformation is applicable to arbitrary transients.

If we consider a function $f(t)$ which is zero for all values of $t < 0$, its Laplace transform is defined as

$$F(s) = \mathfrak{L}\{f(t)\} = \int_0^\infty \epsilon^{-st} f(t)\, dt \tag{13.23}$$

Furthermore, integrations by parts show that:

$$\begin{aligned}
\mathfrak{L}\left(\frac{df}{dt}\right) &= -f(0) + sF(s) \\
\mathfrak{L}\left(\frac{d^2f}{dt^2}\right) &= -\frac{df}{dt}(0) - sf(0) + s^2F(s) \\
\mathfrak{L}\left(\frac{d^3f}{dt^3}\right) &= -\frac{d^2f}{dt^2}(0) - s\frac{df}{dt}(0) - s^2f(0) + s^3F(s), \text{ etc.}
\end{aligned} \tag{13.24}$$

Since, also, obviously

$$\mathfrak{L}\{f_1(t) + af_2(t)\} = \mathfrak{L}\{f_1(t)\} + a\mathfrak{L}\{f_2(t)\} = F_1(s) + aF_2(s) \tag{13.25}$$

it is seen that the Laplace transformation converts any linear differential equation with the independent variable t into a simple algebraic

equation with the variable s. If this algebraic equation is solved for $F(s)$, the function of t corresponding to $F(s)$ is obtained by applying the inverse Laplace transformation:

$$f(t) = \mathfrak{L}^{-1}\{F(s)\} = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} \epsilon^{st}F(s)\,ds \tag{13.26}$$

where γ is any positive number, so that the integration is carried out along some ordinate axis in the positive (right-hand) half of the complex plane for s.* In the circuit problems to which we wish to apply the Laplace transformation $F(s)$ quite generally has the form of a ratio of two polynomials in s, the denominator being of higher order

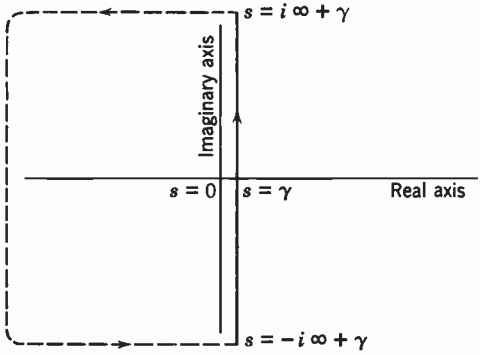


Fig. 13.26. Integration Loop for Evaluating Inverse Laplace Transformation.

in s than the numerator. If the denominator is factored appropriately, it is thus permissible to write

$$F(s) = \frac{G(s)}{\prod_i (s - a_i)^{n_i}} \tag{13.27}$$

where the a_i are the roots of the denominator whose real parts are less than or equal to zero, the n_i are integers, and $G(s)$ is a function which

* The reader can readily verify the validity of the inverse transformation (13.26) by substituting the expression for $F(s)$ from Eq. 13.23; γ is any real number for which

$$\lim_{T\to\infty}\int_0^T |\,\epsilon^{-\gamma t}f(t)\,|\,dt < \infty$$

the existence of a number γ being the condition of transformability for the function $f(t)$. In the examples given here this condition is fulfilled for any positive γ. For further details regarding the Laplace transformation, see, e.g., J. A. Stratton,

is regular throughout the negative complex half plane ($\mathrm{Re}(s) \leq 0$). Since for this form of $F(s)$, nothing is added to the value of the integral in Eq. 13.26 by adding a loop about the negative complex half plane as indicated in Fig. 13.26, this integral (γ being chosen arbitrarily small) may be regarded simply as a contour integral about the complex half plane. It is thus possible to apply Cauchy's integral theorem, to the effect that the value of an integral about a closed loop in the complex plane is equal to $2\pi i$ times the sum of the residues of the poles of the integrand in the interior of the loop. Thus Eq. 13.26 becomes

$$f(t) = \sum_i \frac{1}{(n_i-1)!} \left[\frac{d^{(n_i-1)}}{ds^{(n_i-1)}} \{\epsilon^{st}(s - a_i)^{n_i} F(s)\} \right]_{s=a_i} \tag{13.28}$$

In particular, if all the roots of the denominator are different ($n_i = 1$)

$$f(t) = \sum_i \frac{G(a_i)\epsilon^{a_i t}}{\prod_{j \neq i} (a_i - a_j)} \tag{13.29}$$

The application of the procedure just outlined is best illustrated by determining the response to a unit step for an actual circuit, such as that shown in Fig. 13.9*b*. The Laplace transform for the step function

$$e_1 = 0, \quad t < 0; \qquad e_1 = 1, \quad t > 0$$

is readily evaluated and found to be

$$\mathfrak{L}(e_1) = 1/s$$

Furthermore, if i_1 is the current through L and R, $i_2 = \dfrac{dq_2}{dt}$, that through C_t, and e_2 is the input voltage for the next stage, the following relations are satisfied:

$$i_1 + i_2 = g_m e_1$$

$$L\frac{di_1}{dt} + Ri_1 = \frac{q_2}{C_t} = e_2$$

If the second equation is differentiated once and i_2 is substituted from the first equation,

$$L\frac{d^2 i_1}{dt^2} + R\frac{di_1}{dt} + \frac{i_1}{C_t} = \frac{g_m e_1}{C_t}$$

reference 23, pp. 309–321; Churchill, reference 24; and, for its specific application to television amplifiers, McLachlan, reference 25.

At $t = 0$, $\dfrac{di_1}{dt} = i_1 = 0$; hence, if $F_i(s)$ denotes the Laplace transform of i_1, an application of the Laplace transformation to the above differential equation leads to

$$\left(Ls^2 + Rs + \frac{1}{C_t}\right) F_i(s) = \frac{g_m}{sC_t}$$

so that

$$F_i(s) = \frac{g_m}{s(LC_ts^2 + RC_ts + 1)}$$

and the Laplace transform of e_2 becomes

$$F_2(s) = LsF_i(s) + RF_i(s) = \frac{g_m(Ls + R)}{s(LC_ts^2 + RC_ts + 1)}$$

The roots of the denominator of $F_2(s)$ are seen to be 0, A, and B, where

$$A = -\frac{R}{2L} + j\sqrt{\frac{1}{LC_t} - \frac{R^2}{4L^2}}; \qquad B = -\frac{R}{2L} - j\sqrt{\frac{1}{LC_t} - \frac{R^2}{4L^2}}$$

If Eq. 13.29 is applied, the response of the network thus becomes

$$e_2 = \frac{g_m}{LC_t}\left\{\frac{R}{AB} + \frac{LA + R}{A(A - B)}\epsilon^{At} + \frac{LB + R}{B(B - A)}\epsilon^{Bt}\right\}$$

Substitution of the values for A and B leads, after a certain amount of algebraic manipulation, to

$$e_2 = g_mR\left[1 - \frac{\epsilon^{-\frac{R}{2L}t}}{RC_t\sqrt{\dfrac{1}{LC_t} - \dfrac{R^2}{4L^2}}}\cos\left\{\sqrt{\frac{1}{LC_t} - \frac{R^2}{4L^2}}\,t + \arcsin\left(1 - \frac{R^2C_t}{2L}\right)\right\}\right] \quad (13.30)$$

or, in the notation introduced in section 13.7,

$$e_2 = g_mR\left[1 - \frac{2K}{\sqrt{4K - 1}}\epsilon^{-(\pi f_0t)/K}\left\{\cos\frac{\pi f_0t}{K}\sqrt{4K - 1} - \arcsin\left(\frac{1 - 2K}{2K}\right)\right\}\right]$$

This variation is shown in Fig. 13.27 for particular values of the constants. For moderate values of t it is practically identical with the result obtained by the square-wave analysis described in the preceding section. It may be noted that, as follows from and examination of Eqs. 13.28 and 13.29, quite generally the transient term depends on time only through a sum of decaying exponentials which may eventually be multiplied by sinusoidal functions and powers of t.

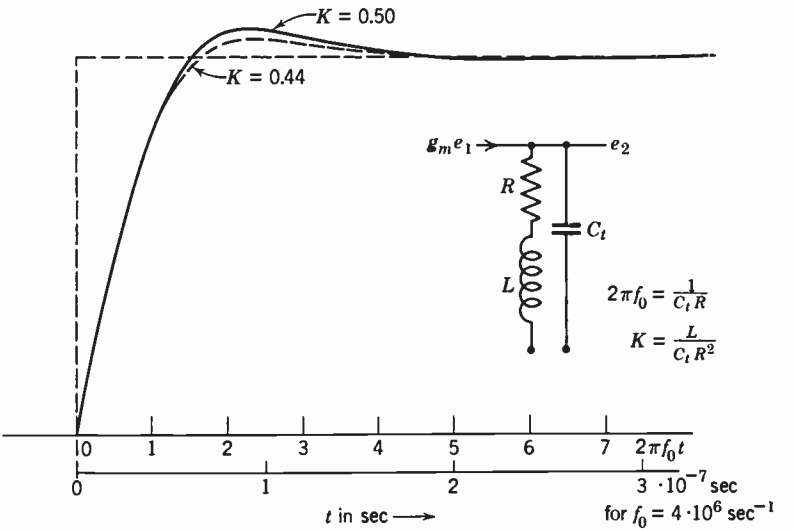


Fig. 13.27. Response of Coupling Circuit Shown in Fig. 13.9*b* to Unit Voltage Step.

It is perfectly possible to treat the complete network, including blocking condenser, tube capacitances, and inductance compensation at high frequencies simultaneously in deriving the response to a voltage step. However, this requires the solution of a third-order algebraic equation and leads to unduly complicated expressions for the constants appearing in the formula for the response. The fact that the effects of the blocking condenser impedance appear only a relatively long time after the impression of the voltage step—in fact, a time so long that the effects of the tube capacities and the compensating inductance on the response have decayed completely—permits once more a separate treatment. The "long-time" response of the network is again given by the equivalent network in Fig. 13.25. In the same manner as before, the Laplace transform of the response is readily found to be

$$F_2(s) = g_m \frac{R_g R}{R_g + R} \frac{1}{s + \dfrac{1}{(R_g + R)C}}$$

and the corresponding value of the step response (by Eq. 13.29) is

$$e_2(t) = g_m \frac{R_g R}{R_g + R} \epsilon^{-\frac{t}{(R_g+R)C}} \tag{13.31}$$

This response, which might, of course, have been derived in more elementary fashion, is plotted, for comparison with the square-wave response in Fig. 13.25, in Fig. 13.28. The complete response curve for the

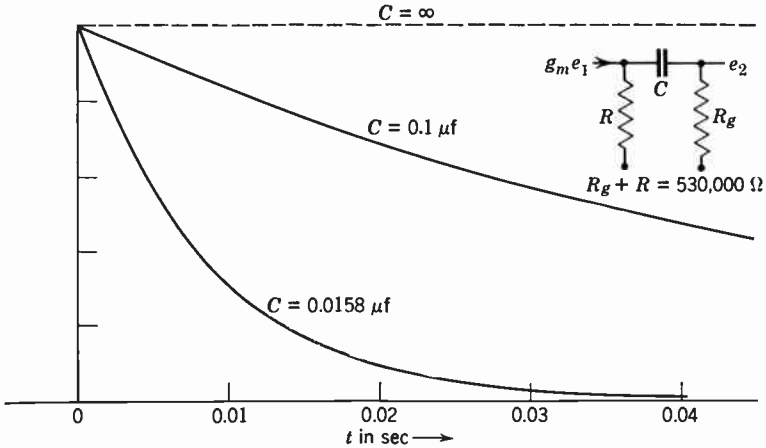


Fig. 13.28. Response of Uncorrected Coupling Network Shown in Fig. 13.25 to Unit Voltage Step.

compound network would be obtained by replacing the initial portion of the curve in Fig. 13.28 by the curve in Fig. 13.27. It should be noted that the "equivalent resistance" R in Eq. 13.30 is identical with the quantity $R_g R/(R_g + R)$ in Eq. 13.31.

In conclusion, we may outline the method employed for applying the Laplace transformation to an amplifier consisting of a number of stages. Let g_n be the mutual conductance of the nth tube, $f_{n+1}(t)$, with the Laplace transform $F_{n+1}(s)$, be the response of the nth coupling network for a unit step voltage input to the amplifier. Since the input to the second stage may be regarded as a superposition of successive voltage steps of height $\left(\dfrac{de_2}{dt}\right) dt$, the output of the second stage may be written

$$e_3(t) = g_2 \int_0^t f_3(t - \tau) \frac{de_2}{dt}(\tau)\, d\tau \tag{13.32}$$

Furthermore, the relation

$$\mathfrak{L}\left\{\int_0^t f_1(t - \tau) f_2(\tau)\, d\tau\right\} = F_1(s) \cdot F_2(s) \tag{13.33}$$

follows from the definition of the Laplace transform (Eq. 13.23) in elementary fashion. Hence Eq. 13.32 yields

$$\mathfrak{L}(e_3) = g_1 g_2 s F_2(s) F_3(s)$$

and, generally,

$$\mathfrak{L}(e_{n+1}) = s^{n-1} \prod_{i=1}^{n} g_n F_{n+1}(s) \tag{13.34}$$

The time variation of e_{n+1} may then be obtained by an application of Eq. 13.28. It is readily seen, however, that the expressions obtained, in particular, for a large number of identical stages become decidedly involved, so that the procedure outlined in section 13.10 is preferable here.

13.12 Additional Frequency Correction. All pickup devices at present in use in practical television systems give a response which is more or less dependent on frequency. This situation arises partly from the electrical characteristics of the device, and in part from the aperture defect. In order to obtain undistorted image reproduction it is necessary to correct the amplifier circuits in such a way that the overall response is flat.

Such compensation is necessary with the iconoscope. The electrical characteristics of this tube are those of a constant current generator shunted by a capacitance or, more exactly, a voltage source in series with a resistance of several megohms with a condenser of about 10 micromicrofarads across the output. The capacity is, of course, in parallel with the resistance coupling the tube to the video amplifier. Since the coupling resistance must be high in order to maintain a high signal-to-noise ratio there will be an attenuation of the signal at high frequency. Smaller coupling resistances are permissible with pickup tubes of relatively high level signal output, such as the image orthicon.

Scanning with a finite aperture always results in a decrease in signal at the higher frequencies. If the scanning aperture is very small the frequency distortion produced may be negligible over the working range of the video amplifier. However, in most practical pickup tubes aper-

ture defect cannot be neglected. It has already been shown in Chapter 5 that the effect of the aperture on the signal is equivalent to that of a filter in series with an ideal signal generator. This equivalent filter has an attenuation which increases with frequency, but produces no change in phase delay.

To compensate for the impedance characteristics and aperture defect, a circuit network with a frequency attenuation characteristic corresponding to the inverse of that of the pickup tube device must be provided in the amplifier. This network is not necessarily separate from those used to compensate the frequency response of the amplifier itself, and, in fact, the correction is often made by overcompensating one or more of the amplifier stages with the circuits described in the preceding sections. Occasionally, however, it is advisable to use a separate filter for this correction, and in this case a high-pass filter (either dissipative or nondissipative) can be designed on the basis of circuit theory to fit the attenuation requirements.

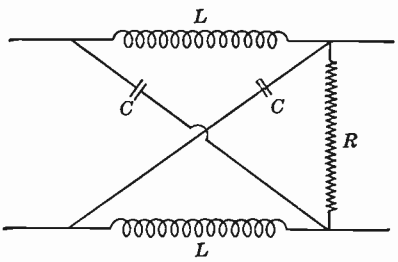


Fig. 13.29. Phase-Correcting Network.

Since the attenuation due to aperture defect occurs without phase changes, and most filters used to compensate for this attenuation introduce a degree of nonuniformity in the delay characteristics, a phase-correcting network may be necessary. In general, these nets are derived from the basic lattice network shown in Fig. 13.29. If the constants of such a network are so chosen that

$$R^2 = L/C$$

the attenuation will be constant while the phase, which is given by

$$\theta = \arctan \frac{2R}{2\pi fL - \dfrac{1}{2\pi fC}} \tag{13.35}$$

is a function of frequency. The phase angle as a function of f/f_0, where $f_0 = 1/(2\pi\sqrt{LC})$, is given in Fig. 13.30. Other types of phase variation can be obtained by substituting various combinations of inductance and capacitance in the bridge arms.

With the aid of either the amplifier compensating circuits or, where necessary, a high-pass filter and a phase-compensating net, it is possible to meet the requirements of overall uniformity in amplitude and delay as a function of frequency.

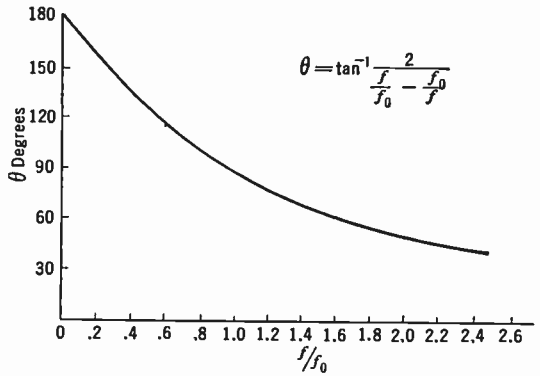


Fig. 13.30. Phase Response of Network Shown in Fig. 13.29.

13.13 Nonlinear Amplification. All vacuum-tube amplifiers have a nonlinear voltage response. The nonlinearity can be made very small if the voltage swing on each stage is kept within the so-called linear portion of the tube characteristic, but it can never be eliminated. As has been seen in the foregoing chapters, both the iconoscope and the kinescope are likewise nonlinear in their response. As a result, the overall amplitude characteristic of the complete system, also, is invariably nonlinear. The effect of this nonlinearity is to distort the contrast values in the reproduced picture, as has already been pointed out in Chapter 5. The response of a typical iconoscope-kinescope system to a series of bright bands having equal increments of light intensity is shown in Fig. 13.31.

Fortunately, the eye is accustomed to seeing under such a wide range of contrast conditions that a contrast distortion even greater than that shown in Fig. 13.31 would be entirely unnoticeable. In this respect the eye differs greatly from the ear, which is very sensitive to nonlinear distortion. Nonlinearity compensation does become desirable, however:

1. With a linear pickup device, such as the flying spot scanner; here the compression in the low lights effected by the power-law transfer characteristic of the kinescope tends to destroy detail in the shadows.

2. With kinescope recording and retransmission; here the recording and the subsequent iconoscope film pickup process add their transfer characteristic, which, uncorrected, resembles Fig. 13.31, to that of the system for direct television transmission. This has, in particular, the effect of washing out the highlights, such as the faces of the actors.

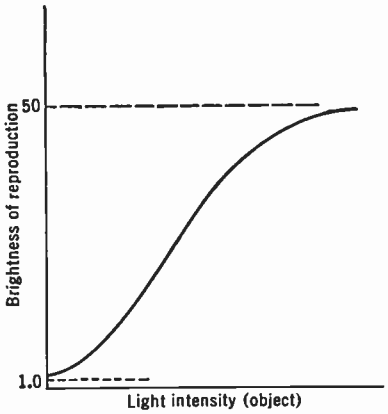


Fig. 13.31. Nonlinearity of a Typical Iconoscope-Kinescope System.

3. In color television, where nonlinearity distorts the rendition of color values as well as brightness values.

In all instances the correction can be effected by the insertion of a nonlinear amplifier stage. Its complexity will, in general, depend on the accuracy with which a linear overall response must be realized. The required accuracy is greatest for color television. Networks suitable for simulating any desired transfer characteristics will accordingly be discussed in greater detail in Chapter 19.

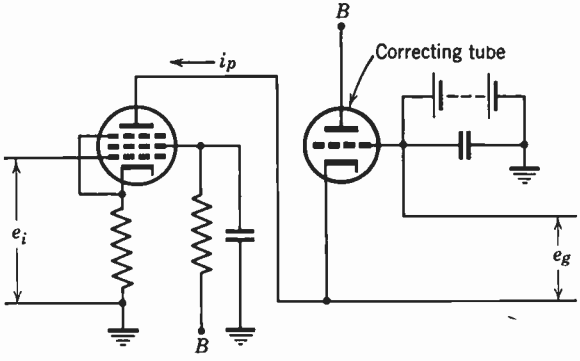


Fig. 13.32. A "Rooter" Amplifier.

If, with a linear pickup device, only a rough compensation for the power law variation of kinescope screen brightness with grid voltage is required, this can be accomplished with a tube whose plate-current versus grid-voltage characteristic exhibits a similar power-law variation. An appropriate circuit is shown in Fig. 13.32. The cathode of

the correcting tube, whose plate and grid are grounded with appropriate fixed bias, is connected to the plate of a high-plate-impedance pentode amplifying the video signal. The desired signal, precompensated for the kinescope characteristic, then appears between the cathode and the grid of the correcting tube. This may be seen readily.

If n is the exponent in the power law representing both the kinescope characteristic and the grid-plate characteristic of the correcting tube,

$$i_p = ke_g^n$$

However, the plate current i_p of the correcting tube is also the plate current of the pentode, with mutual conductance g_m. Hence,

$$g_m e_i = ke_g^n \qquad \text{and} \qquad e_g = k'\sqrt[n]{e_i}$$

n normally has values between 2 and 3. Since the compensating amplifier effectively extracts a root of the input signal, it is commonly known as a "rooter amplifier." *

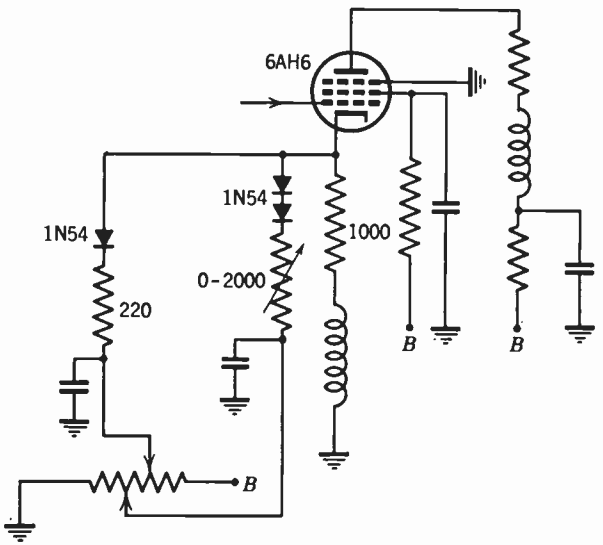


Fig. 13.33. Nonlinear Stage in Orthogam Amplifier (Schematic).

An unavoidable drawback of the rooter amplifier is that, in expanding the contrast range in the shadows, it also accentuates the noise in the darker portions of the picture.

* See, e.g., Oliver, reference 27.

The expansion of the contrast range in the highlights, required in an iconoscope-kinescope chain, may be accomplished with an "Orthogam amplifier" developed by Goodale and Townsend.* Figure 13.33 shows, schematically, the nonlinear compensating stage. Two diodes with resistors in series shunt the cathode circuit. They are inactive until the signal voltage across the cathode circuit equals their bias setting. Above this point (or, rather, these two points) their series resistance shunts the cathode circuit. At these higher signal levels the degenerative feedback is correspondingly reduced, and the gain of the stage is increased.

13.14 Noise Considerations. The importance of noise in relation to the picture signal has already been mentioned in the discussion of the operation of various pickup tubes. It has been shown that the limit to the sensitivity of the pickup tube is determined by the noise generated either by the tube itself or by the amplifier. Observations were cited in Chapter 9 which indicated that for a completely satisfactory picture the ratio of the voltage generated by the highlights and the root-mean-square noise voltage must be at least 30; in other words, a signal-to-noise ratio of 30 or better is required. It was also pointed out that a picture having a signal-to-noise ratio of 10 could be considered usable.

The basic noise in a television system is generated by the pickup tube and, eventually, in the first few amplifier stages. Additional noise may be introduced at other points in the system as, for example, at the input of the radio receiver used to pick up the television signal, but this added noise is not fundamental and may be eliminated by improving the receiving antenna, increasing the transmitter power, etc.

Where an iconoscope or image orthicon is used as the pickup device the noise generated by the tube itself is determined primarily by the current in the scanning beam. By using a low beam current the noise generated by the iconoscope can be reduced to such a value that it can be neglected in comparison with that produced by the amplifier. With a normal iconoscope there is an optimum beam current for which the magnitude of the picture signal is a maximum. If the tube is operated below this optimum, signal output is reduced, but there is an even greater decrease in spurious signal, so that the overall picture is improved if the noise level of the amplifier permits operating in this range. The lower limit beyond which the beam current cannot be re-

* See reference 28.

duced in the normal iconoscope is determined by the current required to discharge the mosaic completely. If this minimum beam produces a noise which is much in excess of the amplifier noise (a condition never realized with the present thermionic amplifiers), a further decrease in beam current without loss in efficiency can be effected by reducing the capacity per unit area of the mosaic, as was explained in Chapter 10.

In the image orthicon the multiplier gain is generally adjusted so that the noise output of the tube overshadows the amplifier noise. Here, too, the lower limit of the beam current is set by the requirement that the target be discharged completely by the beam.

There are two fundamental sources of noise in a vacuum-tube amplifier after such avoidable sources as poor connections or insulation, bad batteries, microphonics, and external pickup have been eliminated. These are the noise generated by the thermal agitation in the resistors and the emission noises in the vacuum tubes.

13.15 Thermal Noise of Resistors. It has been shown experimentally by Johnson,* after whom the effect has been named, and theoretically by Nyquist * that a resistor, regardless of the material of which it is made and independent of the current flowing through it, generates an emf across its terminals owing to the fact that its conduction electrons are in thermal equilibrium with the atoms. It can be shown that this noise is made up of purely random fluctuations, and that, like shot noise discussed in Chapter 1, the noise power is distributed uniformly over the entire frequency band. Furthermore, the mean square voltage is proportional to the resistance of the resistor or, in the case of a complex impedance, to its resistive component. Finally, the power is found to be proportional to the absolute temperature of the resistive elements.

The relation between the voltage generated and the above-mentioned factors can be formulated as follows:

$$\overline{E^2} = 4kRT(f_2 - f_1) \tag{13.36}$$

where k is Boltzmann's constant; T, the absolute temperature; $\overline{E^2}$, the mean square voltage; f_2 and f_1, the upper and lower limit, respectively, of the frequency band over which measurements are being made; and R, the resistance.

At room temperature this expression becomes

$$\overline{E^2} = 1.6\ 10^{-20}R(f_2 - f_1)\ \text{volts}^2 \tag{13.37}$$

* See Johnson, reference 15; and Nyquist, reference 16.

13.16 Tube Noise. Three types of noise can occur in vacuum tubes. They are flicker effect, shot noise, and space-charge limited noise. In addition, there is noise due to ions, secondary emission, and leakage, but in a well-designed tube used for television purposes these last are negligible.

Flicker effect is caused by changes in emission over small parts of the cathode, apparently due to chemical and crystal-structure changes of the emitter. The fluctuations resulting from this effect are important over a rather narrow frequency band at the lower end of the spectrum and become negligible over 1000 cycles. In general, it is not a serious obstacle in television work.

Shot noise has been discussed in detail in Chapter 1. In a temperature-limited vacuum tube this type of noise predominates over all other tube noises. It can be quantitatively expressed in terms of the emission current I by the following equation:

$$\overline{i^2} = 2e(f_2 - f_1)I \tag{13.38}$$

When a tube is operated under space-charge limited conditions, there is a reaction between the fluctuations in emitted electron current and the height of the potential barrier at the virtual cathode which reduces the effect of shot noise in the anode current. This does not mean, however, that the current output from the tube will be free from statistical fluctuations. Theoretical investigation * of this effect has led to a quantitative explanation under certain operating conditions. An approximate rule applicable to screen-grid tubes states that the mean square fluctuations in the plate current will be equal to the temperature-limited shot noise of a current equal to that reaching the screen.† In general, triodes exhibit a lower noise level than the equivalent multigrid tubes.

Tubes are frequently rated for their noise characteristics in terms of an equivalent noise resistance, which is determined as follows. The mean square plate noise current is divided by the square of the transconductance. This yields the mean square voltage which, impressed on the grid, would produce the same noise output in an identical "noiseless tube." The equivalent noise resistance is the resistance whose thermal noise voltage at room temperature is equal to this voltage. This gives a rating which is independent of the frequency band in-

* See Thompson, North, and Harris, reference 19.

† The rule can be applied only to tubes in which the screen-grid current is large; it is not applicable to such tubes as the RCA 6L6.

volved and the gain of the tube. For example, if upon measuring the plate current fluctuations of a 57-type tube, whose transconductance is 1200 micromhos, over a 10-kilocycle band, the noise is found to be 3.4×10^{-18} ampere^2, the rating in terms of equivalent ohms across the grid will be

$$\overline{V^2} = \frac{\overline{i^2}}{g_m{}^2} = 2.4 \times 10^{-12} \quad \text{volt}^2$$

$$R_t = \frac{\overline{V^2}}{1.6 \times 10^{-20} \times F} = \frac{2.4 \times 10^{-12}}{1.6 \times 10^{-20} \times 10^4} = 15{,}000 \text{ ohms}$$

For screen-grid tubes falling within the scope of the rule given above the equivalent noise resistance will be given by

$$R_t = \frac{32 \times 10^{-20} i_s \, \Delta f}{1.6 \times 10^{-20} \, \Delta f \, g_m{}^2 10^{-12}} \tag{13.39}$$

where i_s is the screen current in amperes and g_m is the transconductance in micromhos.

The noise rating * of a few characteristic tubes are listed in Table 13.3.

TABLE 13.3. NOISE RATING OF AMPLIFIER TUBES

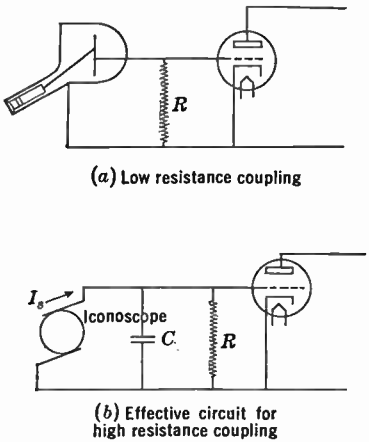
Tube Type	Noise Rating, R_t	Mode of Operation
6AC7	220	triode
6AC7	720	pentode
6AK5	385	triode
6AK5	1880	pentode
6AU6	2660	pentode
6AG5	1650	pentode
6J6(½)	470	triode
6L6G	600	triode
6L6G	1050	pentode

13.17 Signal and Noise in the Video System. The performance of an amplifier can be predicted from a knowledge of the fluctuation characteristics of the tube and coupling circuits. If the gain of the first stage of the amplifier in question is more than three or four, the only noise which produces any appreciable effect upon the signal-to-noise ratio will be that of the coupling resistance and that generated

* For a more complete listing see, e.g., F. Langford Smith, *Radiotron Designers Handbook*, 4th Edition, 1952. See also Kauzmann, reference 21.

by the first tube. Where the gain of the stage is less than three, the noise of the first interstage coupling circuit and the second tube must be taken into account.

The problem considered in this section relates specifically to the iconoscope and its associated video amplifier. Although the reasoning employed is applicable to any pickup system, the employment of a secondary-emission multiplier ahead of the amplifier, as in the image orthicon, minimizes the influence of amplifier noise, reducing the importance of the following considerations.



(*a*) Low resistance coupling

(*b*) Effective circuit for high resistance coupling

Fig. 13.34. Iconoscope Coupling Circuits.

There are a number of ways in which the iconoscope may be coupled to the amplifier. The simplest form of coupling consists of a resistance between the signal lead and ground. The shunt capacity of the iconoscope and grid of the first tube is in parallel with this resistor. Unless the resistance of the resistor is quite low, this capacity has an important effect on both the frequency response and the signal-to-noise ratio. In Chapter 9 the signal-to-noise ratio of an iconoscope coupled with a 10,000-ohm resistance was estimated, leaving out of consideration the shunt capacity. However, since the frequencies involved are as high as 4 or 4.5 megacycles, the effect of this capacity cannot be neglected, unless the coupling resistance is no greater than 2000 ohms.

Where the coupling resistance is greater than 2000 ohms, the simple coupling shown in Fig. 13.34*a* should be represented by the effective circuit shown in Fig. 13.34*b*. The impedance of this circuit is

$$Z = \frac{R - 2\pi jfCR^2}{1 + 4\pi^2 f^2 C^2 R^2} \tag{13.40}$$

Hence, the magnitude of the signal impressed on the grid of the first tube is

$$V_i = |Z| i_s = \frac{Ri_s}{\sqrt{1 + 4\pi^2 f^2 C^2 R^2}}$$

where i_s is the signal current from the iconoscope. In order to correct for the attenuation of the signal with increasing frequency, the amplitude response of the video amplifier must have the following form:

$$G = B\frac{\sqrt{1 + 4\pi^2 f^2 C^2 R^2}}{R}$$

The signal voltage output is independent of frequency and is given by

$$V_s = Bi_s \tag{13.41}$$

The resistive component which determines the noise generated by the coupling impedance is

$$\frac{R}{1 + 4\pi^2 f^2 C^2 R^2}$$

This, together with the equivalent noise resistance of the first tube, R_t, determines the noise input to the amplifier. The input noise in the frequency range f to $f + df$ is

$$\overline{dV^2} = 4kT\left(\frac{R}{1 + 4\pi^2 f^2 C^2 R^2} + R_t\right) df$$

In passing through the amplifier, the noise voltage is amplified in the normal manner. Therefore, the mean square noise voltage appearing at the output of the amplifier over a frequency band extending from 0 to f is

$$\begin{aligned}\overline{V_N{}^2} &= \int 4kT\left(\frac{R}{1 + 4\pi^2 f^2 C^2 R^2} + R_t\right)\frac{B^2(1 + 4\pi^2 f^2 C^2 R^2)}{R^2} df \\ &= \frac{4kTB^2}{R^2}\left[(R + R_t)f + \frac{4\pi^2 C^2 R^2 R_t f^3}{3}\right] \end{aligned} \tag{13.42}$$

The signal-to-noise ratio is obtained by dividing Eq. 13.41 by the rms noise voltage from Eq. 13.42:

$$S = \frac{V_s}{\sqrt{\overline{V_N{}^2}}} = \frac{Ri_s}{\sqrt{4kT}\sqrt{(R + R_t)f + \dfrac{4\pi^2}{3} C^2 R^2 R_t f^3}} \tag{13.43}$$

If the coupling resistance R is small compared with the impedance of the shunt capacity, Eq. 13.43 reduces to

$$S = \frac{Ri_s}{\sqrt{4kT(R + R_t)f}} \tag{13.43a}$$

which is identical with the expression for the signal-to-noise ratio found in Chapter 9. On the other hand, when the resistor R is very large, the equation becomes

$$S = \frac{1}{4\pi}\sqrt{3/(kT)}\,\frac{i_s}{C\sqrt{R_t f^3}} \tag{13.43b}$$

It is interesting to note that the coupling resistance does not enter into the expression for the signal-to-noise ratio when this resistance is large.

For comparison, the two cases are evaluated for the following representative conditions:

Bandwidth	$f = 4.5$ megacycles
Shunt capacity	$C = 15$ micromicrofarads
Equivalent noise resistance	$R_t = 1000$ ohms

Case I. (*Eq. 13.43a*). The resistance coupling of the pickup tube and amplifier must be less than the capacitative reactance. To meet this requirement R is assumed to be 2000 ohms. The signal-to-noise ratio is

$$S = \frac{2000 i_s}{\sqrt{4kT \cdot 4.5 \cdot 10^6 (1000 + 2000)}}$$
$$= 1.4 \cdot 10^8 i_s$$

Case II. Where the coupling resistance is of the order of 200,000 ohms, the expression giving the signal-to-noise ratio is given by Eq. 13.43*b*. Under the conditions given above this has the value

$$S = \frac{1}{4\pi}\sqrt{3/(kT)}\,\frac{i_s}{15 \cdot 10^{-12}\sqrt{1000 \cdot 91 \cdot 10^{18}}}$$
$$= 4.8 \cdot 10^8 i_s$$

The coupling employing a high resistance and an amplifier corrected for the attenuation at high frequencies can be seen from these examples to give more than three times the signal-to-noise ratio. The increased complexity of amplifier design required for complete correction has not proved a serious obstacle to its adoption.

13.18 Blanking and Signal Insertion. It has already been pointed out that the complete video signal must contain the information necessary to synchronize the scanning pattern at the receiver in addition to the picture signal.

The signal obtained directly from the pickup tube does not contain all the required information. To the pickup tube signal must be added

not only the synchronizing impulse whose formation is discussed in the next chapter, but also information as to the average illumination falling on the scene, since the tube is, in general, an a-c device. Furthermore, a disturbance is generated during the return time of the scanning beam, which must be removed from the signal output of the tube.

The disturbing signal can be removed by rendering the amplifier inoperative during the return time of the beam. This process is commonly known as blanking the amplifier. Blanking may be accomplished by driving the grid of one stage of the amplifier below cutoff for the length of time it is desired that the amplifier be inoperative.

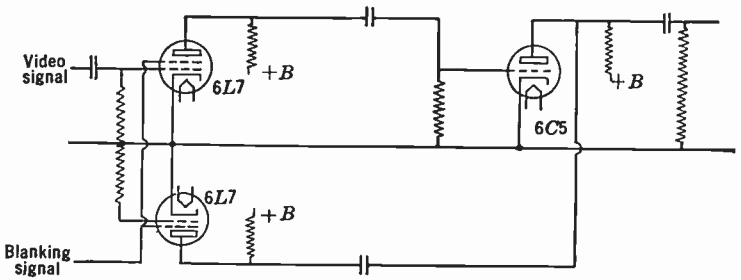


Fig. 13.35. Balanced Blanking Circuit.

As a consequence of biasing the blanking stage to cutoff, a pulse is generated. Since this pulse is unidirectional and positive, excess portions can easily be removed by arranging the bias of the succeeding stage in such a way that the tube is saturated during the blanking interval, or by the use of any one of the many conventional limiters. A second way of blanking the amplifier is shown in Fig. 13.35. This makes use of an additional compensating tube which balances out the effect of the blanking signal on the succeeding stages.

The addition of an impulse is an alternative to rendering the amplifier inoperative during the return time for blanking purposes. Thus an impulse of large amplitude whose duration is slightly in excess of return time may be added in the direction of a black signal to the video signal. The resulting composite signal is then put through a limiter, which removes the top part of the impulse and, with it, the unwanted disturbance.

As was pointed out in Chapter 9, the iconoscope does not yield information as to the average illumination of the scene. In other words, the signal generated by a gray line on a black background will be identical with that of a white line on a gray background. The signal representing either of these images before and after blanking is shown

in Figs. 13.36*a* and *b*. To supply the missing information a phototube can be placed in the iconoscope camera, so positioned that the light it receives is proportional to the average illumination in the picture. The output of this phototube, therefore, represents wanted information. There are a number of ways of adding this information to the video signal. For example, the current from the phototube may be made to control an impulse generator in such a way that the amplitude of these

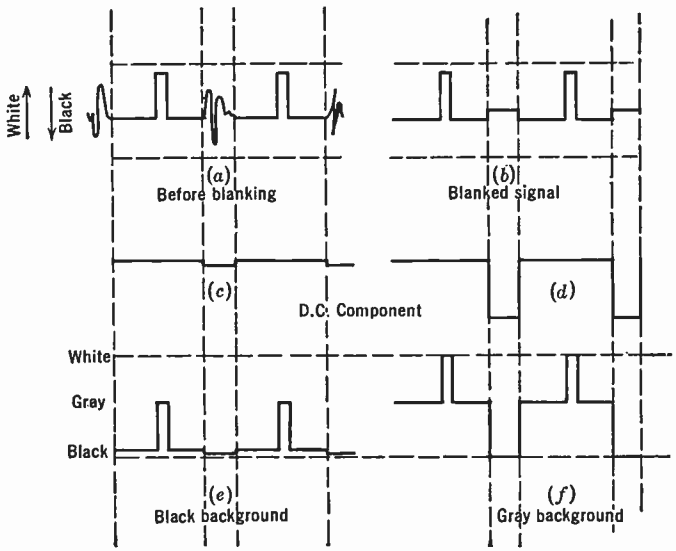


Fig. 13.36. Formation of Video Signal with Pedestal Indicating Mean Light Level.

impulses is proportional to the average light. This generator is so adjusted that the pulse it produces coincides both in time and in duration with the blanking operation. The impulses from a gray bar on black and a white bar on gray are shown in Figs. 13.36*c* and *d*. These impulses are added directly onto the video signal, resulting in a final signal which contains the desired information shown in Figs. 13.36*e* and *f*. The average illumination may also be introduced by utilizing the amplified output of the phototube to control a blanking limiter.

The addition of an impulse such as that indicating the average illumination, or the synchronizing impulses, may be carried out merely by introducing the signal to be added at the bottom of the grid resistor of the stage where the addition is to take place. By inserting the signal in this way, additional shunt capacity is not introduced into the amplifier. A circuit of this type is shown in Fig. 13.37.

Figure 13.38 shows a "clamp circuit" employed for fixing the black level in an image orthicon camera. A rectangular pulse coinciding with the blanking time is applied to an amplifier tube with equal plate and

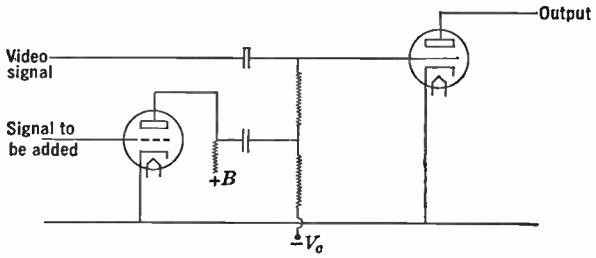


Fig. 13.37. Simple Injector Circuit.

cathode resistances. The pulse voltages appearing across these resistances, which are equal in amplitude and opposite in polarity, are applied to the two sides of a double diode in such a direction as to

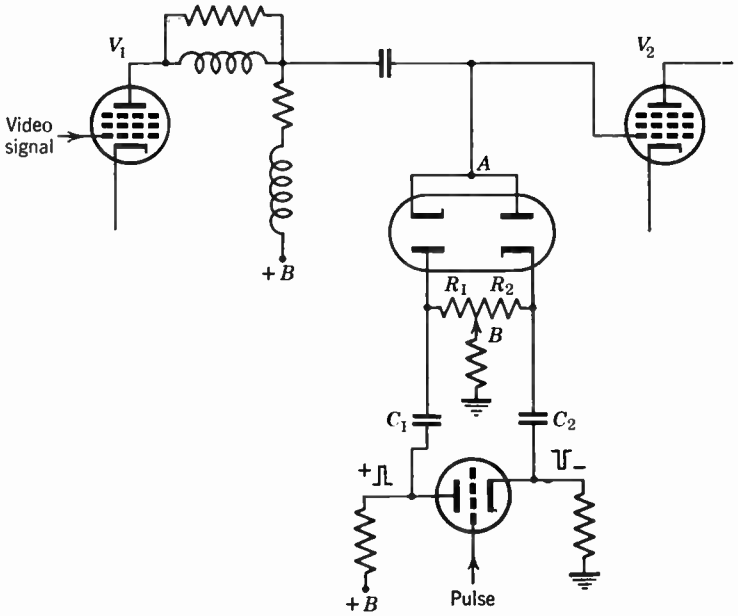


Fig. 13.38. Double-Diode Clamping Circuit.

render them conducting. Under these circumstances the two diode sections form a Wheatstone bridge with the resistances R_1 and R_2, and the junction point A is held at the potential of the center point B irre-

spective of the video signal transmitted from tube V_1. Between pulses the charge on the condenser C_1 and C_2 biases the diodes to cutoff, and the video signal is transmitted from tube V_1 to the grid of tube V_2 without interference.

13.19 Video Cable. Often it is necessary to interconnect two remotely placed amplifiers. This requires a cable that can carry an extremely wide frequency band. Such a cable may consist of a small conductor in a relatively large metal shield. The conducting core is separated from the shield by means of small, low-loss insulators spaced at intervals of a few inches. In this way, the capacity per unit length is made much smaller than it would be if the space between the shield

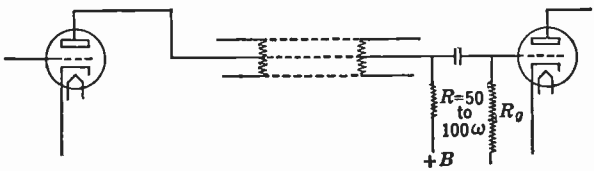


Fig. 13.39. Plate-Coupled Driving Circuit for Cable.

and the conductor were filled with an insulating material. Furthermore, such a cable has a very low leakage conductance for a given length.

The capacity of this type of cable is between 15 and 25 micromicrofarads per foot. Where short lengths are used, as for example between the preamplifier and the monitoring amplifier, the capacitative impedance is high enough so that, if the load at the output end has a fairly low impedance, the frequency and phase distortion produced by the cable is quite small and can be compensated by the peaking means described in earlier sections.

In order to couple the output of an amplifier efficiently to the cable, the output tube must be capable of driving a low impedance. Because of the size of the coupling condenser that would be necessary with the conventional arrangement of load resistance and capacitor, the circuit shown in Fig. 13.39 is commonly employed. The connection indicated requires that the core of the cable be at the plate potential of the driver tube.

The cathode-coupled circuit shown in Fig. 13.40 is very well adapted to driving a low impedance. A tube coupled in this way cannot have more than unity voltage gain. However, because of the low impedance into which the output tube must work, the gain of this stage will, in general, be far below unity, irrespective of the circuit.

Where a long cable is necessary the problem of its termination is much more difficult. The termination must correct for the attenuation at high frequency and for the form of the phase delay. It must also match the impedance of the cable in order to avoid reflections. In

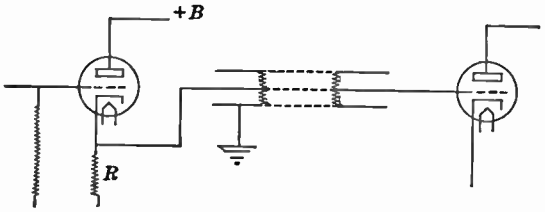


Fig. 13.40. Cathode-Coupled Driving Circuit for Cable.

other words, the termination must equal the surge impedance so that the cable will act as though it were infinite in length.

The surge impedance of a cable is given by

$$Z_s = \frac{r + j\omega L}{g + j\omega C}$$

where r and g are the resistance and leakage conductors per unit length, L and C the inductance and capacity also per unit length. This becomes invariant with frequency *if*:

$$\frac{r}{g} = \frac{L}{C}$$

However, it is not practicable to make a cable having this characteristic as the leakage conductance would have to be so high that the loss would be extremely great. Therefore, the input impedance of the terminus must vary with frequency so as to match the surge impedance, and then the attenuation of this unit must vary with frequency in order to produce constant overall response. This usually requires an attenuation which decreases with increasing frequency. The surge impedance of a television cable such as described is about 80 ohms and increases to a very high value at low frequencies.

13.20 The Complete Amplifier. In order to knit together the various elements discussed in this chapter, Fig. 13.41 is given, showing a circuit diagram of a typical camera amplifier used in television work. The complete amplifier is shown as a block diagram in Fig. 13.42, together with the signal injection points.

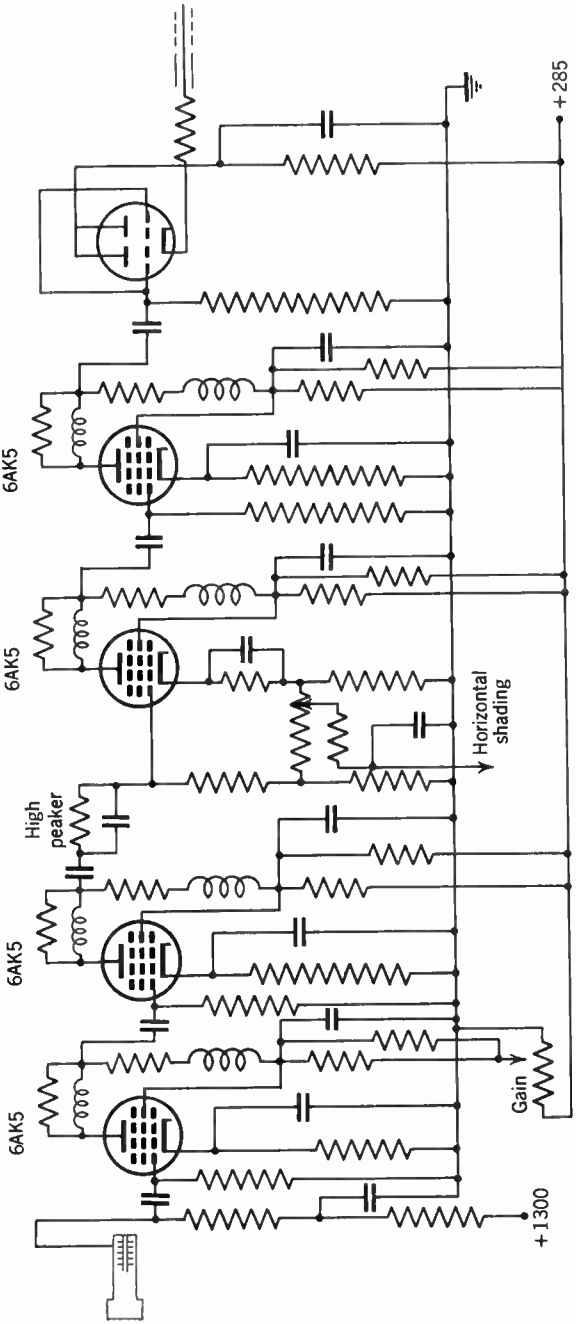


Fig. 13.41. Typical Image Orthicon Camera Amplifier.

This chapter does no more than outline the basic principles involved in video amplifier design. It is not intended that this summary should be used in the actual design and construction of an amplifier, as a

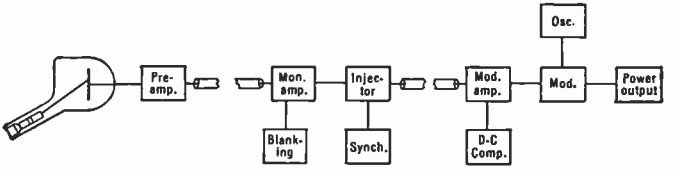


Fig. 13.42. Block Diagram of Complete Amplifier, Including Injector and Blanking Units.

treatment of the details of this problem obviously requires a textbook in itself. Those interested are referred to the many original articles on this subject, some of which are listed in the bibliography immediately following.

REFERENCES

1. D. G. Fink, *Television Engineering,* 2nd Ed., McGraw-Hill, New York, 1952.
2. F. E. Terman, *Radio Engineering,* 3rd Ed., McGraw-Hill, New York, 1947.
3. K. McIlwain and J. G. Brainerd, *High Frequency Alternating Currents,* Wiley, New York, 1931.
4. T. E. Shea, *Transmission Networks and Wave Filters,* Van Nostrand, New York, 1929.
5. T. S. Gray, *Applied Electronics,* The Technology Press, M.I.T., and Wiley, 1954.
6. G. D. Robinson, "Theoretical Notes on Certain Features of Television Receiving Circuits," *Proc. I.R.E.,* Vol. 21, pp. 833–843, June, 1933.
7. S. W. Seeley and C. N. Kimball, "Analysis and Design of Video Amplifiers," *RCA Rev.,* Vol. 2, pp. 171–183, October, 1937, and Vol. 3, pp. 290–308, January, 1939.
8. A. Preisman, "Some Notes on Video Amplifier Design," *RCA Rev.,* Vol. 2, pp. 421–432, April, 1938.
9. F. A. Everest, "Wide-Band Television Amplifiers," *Electronics,* Vol. 11, pp. 16–19, January, 1938.
10. E. W. Herold, "High-Frequency Correction in Resistance-Coupled Amplifiers," *Communications,* Vol. 18, pp. 11–14, August, 1938.
11. H. A. Wheeler, "Wide-Band Amplifiers for Television," *Proc. I.R.E.,* Vol. 27, pp. 429–438, July, 1939.
12. A. V. Bedford and G. L. Fredendall, "Transient Response of Multistage Video-Frequency Amplifiers," *Proc. I.R.E.,* Vol. 27, pp. 277–284, April, 1939.
13. H. A. Wheeler, "The Interpretation of Amplitude and Phase Distortion in Terms of Paired Echoes," *Proc. I.R.E.,* Vol. 27, pp. 359–384, June, 1939.
14. O. J. Zobel, "Distortion Correction in Electrical Circuits with Constant Resistance Recurrent Networks," *Bell System Tech. J.,* Vol. 7, pp. 438–534, July,

1928. See also "Electric Wave Filters," *Bell System Tech. J.*, Vol. 10, pp. 284–341, April, 1931.

15. J. B. Johnson, "Thermal Agitation of Electricity in Conductors," *Phys. Rev.*, Vol. 32, pp. 97–109, July, 1928.
16. H. Nyquist, "Thermal Agitation of Electric Charge in Conductors," *Phys. Rev.*, Vol. 32, pp. 110–113, July, 1928.
17. F. B. Llewellyn, "A Study of Noise in Vacuum Tubes and Attached Circuits," *Proc. I.R.E.*, Vol. 18, pp. 243–265, February, 1930.
18. G. L. Pearson, "Fluctuation Noise in Vacuum Tubes," *Physics*, Vol. 5, pp. 233–243, September, 1934.
19. B. J. Thompson, D. O. North, and W. A. Harris, "Fluctuations in Space-Charge Limited Currents at Moderately High Frequencies," *RCA Rev.*, Vol. 4, pp. 269–285, 441–472, 1940; Vol. 5, pp. 244–260, 371–388, 505–528, 1940–41; and Vol. 6, pp. 114–124, 1941.
20. W. S. Percival and W. L. Horwood, "Background Noise Produced by Valves and Circuits," *Wireless Engineer*, Vol. 15, pp. 128–137, March, 1938; pp. 202–207, April, 1938.
21. A. P. Kauzmann, "New Television Amplifier Receiving Tubes," *RCA Rev.*, Vol. 3, pp. 271–289, January, 1939.
22. R. D. Kell, A. V. Bedford, and M. A. Trainer, "An Experimental Television System," *Proc. I.R.E.*, Vol. 22, pp. 1246–1265, November, 1934.
23. J. A. Stratton, *Electromagnetic Theory*, McGraw-Hill, New York, 1941.
24. R. V. Churchill, *Modern Operational Mathematics in Engineering*, McGraw-Hill, New York, 1944.
25. N. W. McLachlan, "Reproduction of Transients by a Television Amplifier," *Phil. Mag.*, Vol. 22, pp. 481–491, 1936.
26. G. E. Valley and H. Wallman, *Vacuum-Tube Amplifiers*, McGraw-Hill, New York, 1948.
27. B. M. Oliver, "A Rooter for Video Signals," *Proc. I.R.E.*, Vol. 38, pp. 1301–1305, 1950.
28. E. D. Goodale and C. L. Townsend, "The Orthogam Amplifier," *RCA Rev.*, Vol. 11, pp. 399–410, 1950.

14 Scanning and Synchronization

The complete deflection system for scanning consists of two pairs of deflecting coils or plates, two sawtooth generators, and two discharge impulse oscillators. One set of these elements operates at horizontal or line frequency, the other at vertical or field frequency. Both transmitter and receiver are, of course, equipped with such systems, which must be exactly synchronized. Synchronization is accomplished by transmitting, along with the picture, synchronizing signals which, by suitable circuits, can be separated into vertical and horizontal synchronizing impulses. These impulses are used to actuate or control the respective deflection generators at the receiver. In practice the synchronizing signal controls not only the receiver scanning, but also the deflection circuits of the camera tube; in other words, the deflecting system at the camera tube does not generate the synchronizing impulse, but rather both camera and receiver deflection are controlled by an independent synchronization generator.

In the United States a ratio of frame frequency to scanning line frequency of 525 and a frame frequency of 30 pictures per second are prescribed by the Federal Communications Commission. Since interlaced scanning is used, the frequency of the vertical deflection is 60 cycles per second, two vertical strokes constituting a single complete picture period. The frequency of 30 frames per second was chosen originally because most of the electrical power used in the United States is generated at 60 cycles and, by making the frame frequency a simple submultiple of this, much trouble from "hum," that is, a-c ripple, can be avoided. For similar reasons the British Broadcasting Company has adopted a repetition rate of 25 frames per second and a vertical deflection frequency of 50 cycles per second. As has already been mentioned, the higher field frequency of 60 cycles has proved advantageous in permitting higher picture brightnesses without perceptible flicker.

14.1 Requirements of Scanning. The theory of scanning in relation to the analysis and synthesis of the picture itself was taken up in

some detail in Chapter 5. From this consideration it is quite evident that the success of the transmission of the picture depends upon an exact correspondence in the position of the scanning beam in the camera tube and the kinescope. In order to maintain this correspondence of position it is necessary to employ rather special synchronizing and scanning equipment. This equipment and its operation form the subject of the present chapter.

The treatment of the problem divides itself naturally into three parts: first, the means of producing the motion of the spot in the receiving and transmitting tubes; second, the generator required to supply the power for producing the motion of the spot; and, finally, methods of synchronizing the transmitter and receiver. However, before taking up these topics, some of the principles of scanning will be reviewed in order better to formulate the problem as a whole.

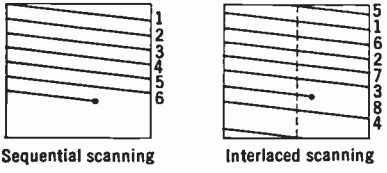


Fig. 14.1. Sequential and Interlaced Scanning.

In any television system using a storage camera tube and a kinescope at its two terminals, scanning consists of causing the electron beam to sweep over the image area (either target or fluorescent screen) in a series of straight parallel lines. The pattern thus generated is illustrated in Fig. 14.1. These lines may be laid down in sequence or, in accord with the more common practice, may be interlaced, i.e., laid down in the order 1, 3, 5, etc., 2, 4, 6, etc.

A perfect scanning pattern will be defined as one in which the following conditions are fulfilled:

1. The beam moves at a constant velocity along each line.
2. The lines are straight.
3. The lines are equally spaced.
4. The lengths of the lines are equal.

If the beam at the camera tube sweeps out a perfect pattern, then, in order that the received picture be a geometrically exact reproduction of the image being transmitted, it is necessary that the scanning pattern on the viewing tube or kinescope also be perfect. Furthermore, the scanning beam at the transmitter and receiver must start and reach the end of the picture simultaneously. That these are sufficient conditions is obvious and requires no further explanation.

On the other hand, if the pattern at the camera tube is distorted, an exact reproduction of the image can be obtained at the kinescope only if its scanning pattern is distorted in an identical manner. Figure 14.2 illustrates a camera tube pattern whose shape is distorted owing to varying lengths of the scanning line, and shows the images obtained on a perfect and on a similarly distorted kinescope pattern. As far as geometrical fidelity of the image is concerned, a television system with both the transmitting and receiving patterns distorted is perfectly possible. However, since a rectangular picture area is considered desirable, the use of a distorted pattern does not permit the fullest use of

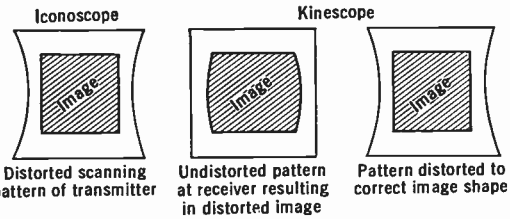


Fig. 14.2. Effect of Distorted Pattern on Reproduced Image.

the available picture area. Similarly, if the scanning pattern at the transmitting tube is not perfect on account of a varying velocity of the beam along the scanning line, a geometrical distortion of the received picture will occur at the kinescope unless its scanning beam likewise moves with a nonuniform velocity. Similar nonuniformity in both patterns, although causing no geometrical distortion in the received image, is undesirable because it leads to unequal horizontal resolution in different parts of the picture. This is a consequence of the fact that, the higher the velocity of the scanning beam, the lower the resolution for a given band of transmitted frequencies. In general, the most economical transmission of pictures requires as nearly perfect scanning patterns at the receiver and transmitter as are practicable.

Referring again to Fig. 14.1, it is apparent that the scanning pattern is generated by a displacement of the beam, which can be resolved into two mutually perpendicular displacements. If the motion of the scanning beam is separated into its vertical and horizontal components, and they are plotted as a function of time, sawtooth curves are obtained. The period of the sawtooth wave representing vertical displacement is the length of time required for the scanning beam to cover the entire picture area or, in other words, one field period; the period of the horizontal sawtooth is the length of time required by the scanning beam to

produce one line. The ratio of the vertical to the horizontal period is, of course, equal to the number of lines making up a single field. Sawtooth waves which may represent the horizontal and the vertical displacements are shown in Fig. 14.3. It will be seen that each wave consists of a gradual motion of the beam across the image area and an abrupt return to its initial position. Since the only useful portion of the cycle is the gradual displacement, it is desirable to make the return time as short as possible. In practice it is found that the vertical return time cannot be made less than about 7 percent of the total period, or the horizontal return time less than 15 percent. This return time is not entirely wasted as part of it is used for the transmission of the synchronizing impulse.

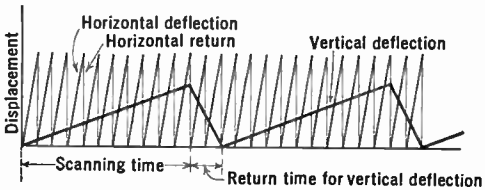


Fig. 14.3. Sawtooth Displacement Components Required for Scanning.

14.2 Deflection of Electron Beam. The process of electron beam deflection differs materially in the kinescope, iconoscope, and flying-spot tube, on the one hand, and the low-velocity camera tubes, such as the orthicon, image orthicon, and Vidicon, on the other. Finally, deflection in the image dissector differs from both groups although it resembles more closely that employed in the low-velocity tubes. In the present section only the first type of beam deflection, represented principally by the kinescope, will be discussed.

The electron beam from a properly centered electron gun in a kinescope or iconoscope will, if undisturbed, strike the surface of the fluorescent screen or mosaic near its center.* In order to produce scanning, the beam must be moved across the screen independently in two directions at right angles to each other.

With a conventional gun, such as is described in Chapter 12, one of the most satisfactory ways of displacing the spot is to deflect the beam through an angle close to the end of the gun, as is shown in Fig. 14.4. This deflection of the electron beam can be accomplished in either of two ways. An electrostatic field can be established in the region near

* The angle of the mosaic relative to the axis of the gun positions the undeflected spot below the center to permit symmetric vertical angular displacement in scanning.

the end of the gun and parallel to the direction in which the beam is to be deflected, or a magnetic field may be set up at right angles to the direction of deflection. At present, magnetic deflection for both the vertical and the horizontal displacement is employed universally for iconoscopes and all kinescopes, with the exception of some small special-purpose tubes. Its primary advantage is lowered tube cost. In addition, as will be seen below, the deflection energy required increases more rapidly with voltage for electrostatic deflection than for magnetic deflection. Finally, at very high voltages, the introduction of the high-voltage deflection plates introduces difficult problems of tube design.

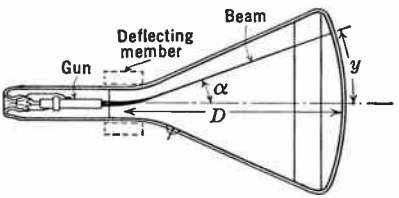


Fig. 14.4. Deflection of Beam in Kinescope.

It is, nevertheless, of interest to give a brief analysis of the action of an electrostatic deflection field on an electron beam. Electrostatic deflection may be effected by means of a pair of parallel plates, placed at the end of the gun in such a way that the electron beam passes between them. The arrangement is shown in Fig. 14.5. In this figure G_1 and G_2 are the first and second anodes of the gun and A–B, A'–B' are the two deflecting plates. A potential difference between the plates produces the electrostatic field which deflects the beam. Idealized field lines are shown between the plates. In order to calculate the deflection produced by these plates, three simplifying assumptions will be made: first, that the length d of the plates is small compared with the distance between the center point of the plates and the fluorescent screen; second, that the plates are placed far enough from the first anode so that there is no interaction between the field of the second lens of the gun and the deflecting field; and, finally, that the fringe fields have a negligible effect on the deflection. The error introduced by these assumptions will be considered later.

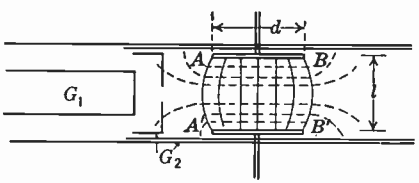


Fig. 14.5. Arrangement of Deflecting Plates.

The field E between the plates is given by

$$E = \frac{V_A - V_{A'}}{l}$$

where V_A and $V_{A'}$ are the potentials applied to the deflecting plates and l is the separation between the plates. This field acts on the electrons leaving the gun over a portion of their path equal to the distance d. The electron path over this length is parabolic, and the displacement Δ from the axis is

$$\Delta = \frac{1}{4}\left(\frac{V_A - V_{A'}}{V}\right)\frac{d^2}{l} \tag{14.1}$$

where $eV = e(V_A + V_A')/2$ is the kinetic energy of the electrons associated with their axial component of velocity. The ray is, therefore, deflected through an angle α, where

$$\alpha \cong \tan \alpha = \frac{2\Delta}{d} = \frac{1}{2}\left(\frac{V_A - V_{A'}}{V}\right)\frac{d}{l} \tag{14.2}$$

This deflection produces a displacement y of the spot at the screen which, if D is taken as the distance between the screen and the center of the deflecting plates, is given by

$$\begin{aligned} y &= \Delta + \alpha\left(D - \frac{d}{2}\right) = \frac{1}{2}\left(\frac{V_A - V_{A'}}{V}\right)\frac{d}{l}\left(\frac{d}{2} + D - \frac{d}{2}\right) \\ &= \frac{1}{2}\left(\frac{V_A - V_{A'}}{V}\right)\frac{dD}{l} \end{aligned} \tag{14.3}$$

This equation indicates that within the limits of the approximation the displacement on the screen is proportional to the potential difference between the deflecting plates. Therefore, a sawtooth voltage wave must be applied to the deflection plates to produce the displacement of the beam required for scanning.

It remains to examine the assumptions on which Eq. 14.3 is based. The first two, namely, a distance between plates and screen which is large in comparison with the length of the plates and noninterference between the focusing fields and deflecting fields, are quite readily fulfilled in most cases. The fringe fields around the plates, on the other hand, cannot be neglected so readily. These fringe fields have three major effects. First, they introduce a cylindrical lens into the electron-optical system of the gun, which may change its focusing properties so that the position of minimum spot size is made a function of the deflection, and, even more important, may distort the shape of the spot. Second, they make the displacement of the spot on the screen a func-

tion, not only of the potential applied to the deflecting plates, but also of any displacement of the beam in the plane of the plates, introducing "cross talk" between horizontal and vertical deflection. Third, the fringe field may result in a slight departure from linearity in the relation between deflecting potential and displacement. In order to overcome these defects, it is necessary to use specially shaped or positioned deflecting plates rather than plane, parallel plates. Flaring the plates, to match approximately the paths of the beam when deflected through the maximum angle, is also desirable from the point of view of maximum voltage and power sensitivity.

As has already been pointed out, it is necessary to deflect the beam in two mutually perpendicular directions. This can be done by having two pairs of deflecting plates, one ahead of the other, so that between the first the beam is deflected through an angle in one plane and then bent through an angle perpendicular to this between the second pair of plates (Fig. 14.6). The alternative arrangement of a single deflecting field produced by both pairs of plates (Fig. 14.7*a*) restricts the freedom in shaping the plates to such an extent that it becomes difficult to avoid excessive spot and pattern distortion. Good results have been obtained, however, by sawtooth interleaving of the deflection plates and applying the deflection voltages through the frequency sensitive network shown in Fig. 14.7*b*.* In this manner the lateral fringe fields are effectively eliminated.

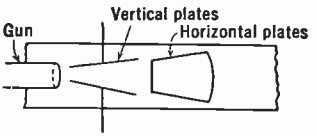


Fig. 14.6. Successive Electrostatic Deflection in a Vertical and Horizontal Direction.

With magnetic deflecting fields the superposition of the two deflecting fields has no adverse effect on spot and pattern and is, in fact, universally practiced. The deflecting elements consist of pairs of coils without or, more commonly, with cores of magnetic materials; the latter reduce the current required to achieve a given deflection.

The deflecting coils are mounted over the tube neck, at a point just beyond the end of the electron gun. They form, in general, an integral part of the receiver and need not be exchanged with the tube. The arrangement of two pairs of air-cored coils on the tube neck is shown in Fig. 14.8. The arrangement of the coils in a deflection yoke with magnetic core is quite similar; Fig. 14.9 shows a single pair of coils with a ring core. The folding of the ends of the coils as shown in the

* See Schlesinger, reference 21.

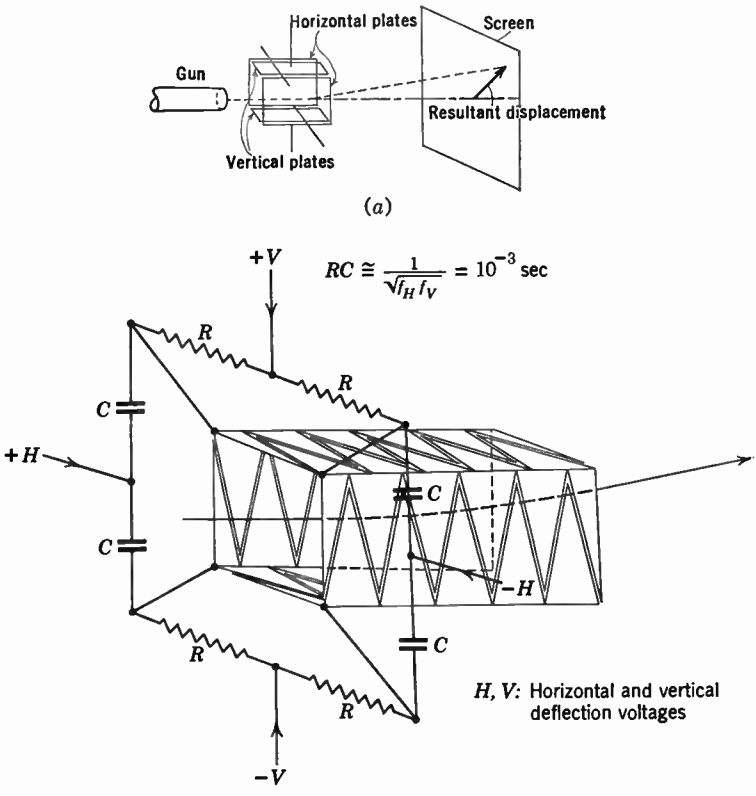


Fig. 14.7. Combined Vertical and Horizontal Electrostatic Deflection.

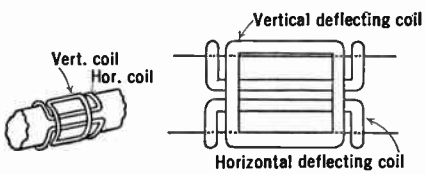


Fig. 14.8. Simple Form of Air-Core Deflecting Coils.

bottom diagram is advantageous both for achieving maximum deflection sensitivity (deflection for a given number of ampere turns in the coil) and, particularly, for minimizing defocusing of the beam for large deflections. Figure 14.10 shows the external appearance of a commercial deflection yoke as well as its several components.

The action of the magnetic deflecting field on the electron beam can be readily determined if, once more, some radical simplifying assumptions are made. The first of these is that the magnetic field strength H is uniform over a region of length d, and zero outside this region. This simplification entirely neglects fringe field effects. The second assumption is that the distance D from the center of the field to the screen is large compared with the length of field d and with the displacement y on the screen.

As shown in Fig. 14.11, the electrons move along circular paths within the field and emerge traveling in a direction which makes a small angle α with their original motion. In the field-free region beyond the deflecting field the paths are straight. The electron stream arrives at the screen at a point displaced an amount y from its undeflected position.

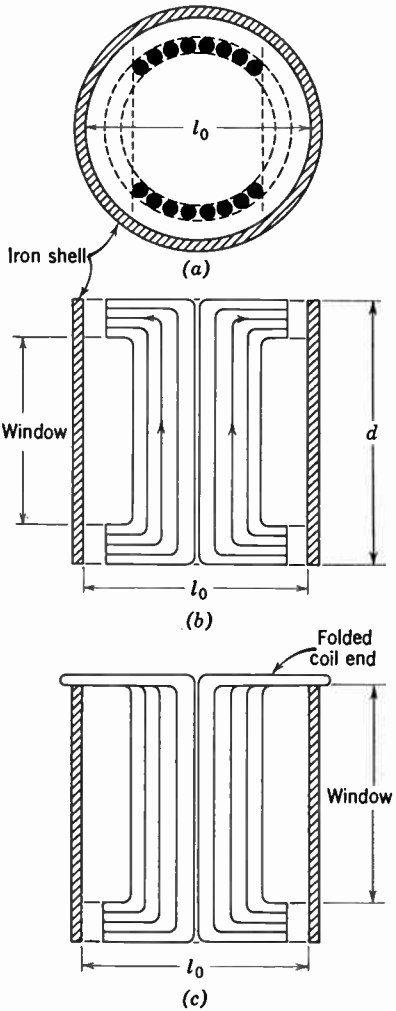


Fig. 14.9. Cross Sections of Single Pair of Deflecting Coil with Magnetic Ring Core. (*a*) Transverse section. (*b*) Longitudinal section. (*c*) Longitudinal section of coil with folded end. (Schade, reference 9.)

The force acting on an electron at right angles to its direction of motion is

$$F = eH\sqrt{2eV/m} \tag{14.4}$$

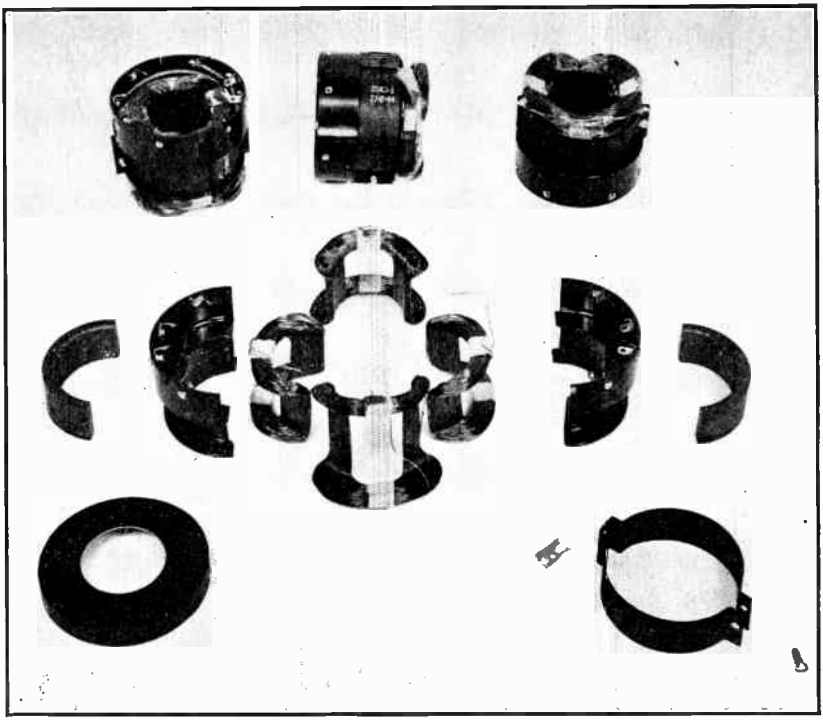


Fig. 14.10. External Appearance of Commercial Deflection Yoke. (Obert and Needs, reference 6.) (Courtesy of *Tele-Tech*.)

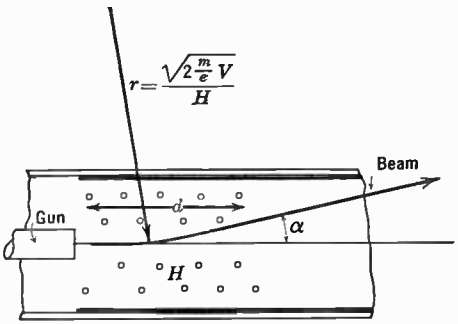


Fig. 14.11. Deflection of Beam in Magnetic Field.

Therefore, it will move in a circular path of radius:

$$R = \frac{1}{H}\sqrt{2mV/e} \tag{14.5}$$

The angle α through which it is deflected in traversing the distance d is, therefore,

$$\alpha \cong \sin \alpha = \frac{d}{R} = \frac{Hd}{\sqrt{2mV/e}} \tag{14.6}$$

and the displacement Δ at the end of the circular path

$$\Delta = R - \sqrt{R^2 - d^2} \cong \frac{d^2}{2R} = \frac{d^2H}{2\sqrt{2mV/e}} \tag{14.7}$$

The total displacement y is, therefore,

$$y = \Delta + \alpha\left(D - \frac{d}{2}\right) = \frac{D\,dH}{\sqrt{2mV/e}} \tag{14.8}$$

It will be noticed that the deflection for a given field H is inversely proportional to the square root of the anode voltage, whereas in the case of electrostatic deflection it is inversely proportional to the voltage itself. As has already been pointed out, this is an important factor in favor of magnetic deflection where high beam voltages are required, as, for example, in projection tubes.

Just as for electrostatic deflection, it is not possible to have a sharply defined region of uniform field. This can readily be seen from the form of the Laplace equations governing the magnetic field in free space. The fringing of the field tends to produce aberrations in the electron beam and also to distort the scanning pattern.

An exact analysis of the defects encountered in magnetic deflection is highly complex and can be carried out only with reference to a specified deflection yoke and gun. However, the origin and nature of some of the difficulties can be shown rather easily.

Figure 14.12 illustrates an electron beam under the influence of a nonuniform magnetic deflection field. The lack of uniformity, in both diagram *a* of the plane, including the axis of the beam and the pole pieces (of a type not normally employed), and diagram *b*, which shows a cross-section of the beam, has been greatly exaggerated for the sake of clarity. Since the cross-section of the beam has finite size, the field is not constant over its area, causing different parts of the beam to be

deflected different amounts.* Along the diameter from point 1 to point 3 in diagram *b* the field decreases; therefore, the electrons in the neighborhood of 1 are deflected more than those near 3, resulting in a compression of the spot along this diameter. Furthermore, since the

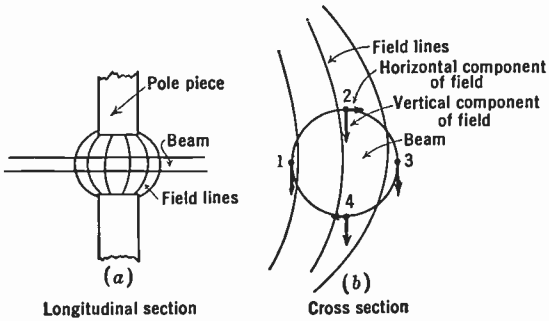


Fig. 14.12. Influence of Nonuniform Field on Electron Beam.

field is not uniform, the magnetic field lines are curved as shown. As a consequence of this curvature, the direction of the field at the top and bottom of the beam is not the same. Considering points 2 and 4, there are, as well as the field components producing the desired deflection, components parallel to the direction of deflection. This causes the electrons in the top portion of the beam to be bent upward, and those in the lower section downward. The net effect, therefore, is that the beam is compressed in the plane of deflection and expanded at right angles to it, resulting in an elliptical spot. If the beam is not in the median plane between the two poles, the effect is more complicated, but also leads to an elliptic spot whose major and minor axes may be rotated at arbitrary angles with respect to the direction of deflection.

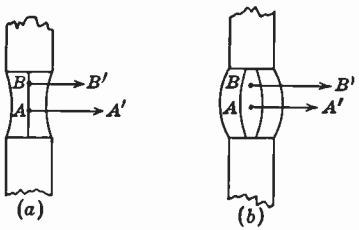


Fig. 14.13. Cause of Scanning Pattern Distortion by Nonuniform Field.

That the nonuniform field distorts the scanning pattern can be readily seen from Fig. 14.13, which again shows a section of the horizontal

* Strictly speaking, it is not the magnetic field in a particular beam cross-section, but rather the line integral of the transverse field component along the path of a particular beam electron which determines the deflection. Elementary considerations show, however, that, in practical cases, this line integral increases with the field in the midplane, though at a less than linear rate.

deflecting field normal to the beam. If the beam is undeflected in the vertical direction so that it lies in the median plane between the horizontal pole pieces corresponding to position A of the diagrams, it will be deflected in this plane by an amount represented by the vectors AA'. For convenience it may be assumed that this deflection is equal for the two types of distorted field illustrated. However, if the beam has been deflected in the vertical direction so that it passes through points B, the horizontal deflections will no longer be equal, for in the field represented in a it has moved into a region of lower field strength, whereas

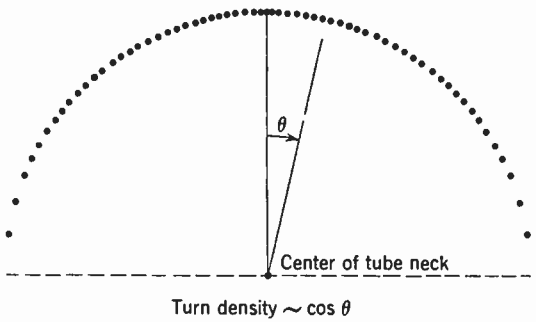


Fig. 14.14. Optimum Distribution of Turns in Air-Cored Coil.

in b it is in a region of higher field strength. Therefore, if the field has "barrel"-shaped distortion (b), the resulting scanning pattern will be pincushioned, whereas if the field lines are "hourglass"-shaped (a) the pattern will have barrel distortion.

These defects can be minimized by shaping the pole pieces properly and distributing the windings so that a nearly uniform field pattern is produced. With a long air-cored coil a very high degree of uniformity can be achieved if the individual rectangular turns are distributed around the periphery of the tube neck with a density proportional to the cosine of the azimuthal angle θ, as shown in Fig. 14.14.

The arrangement of the turns shown in Fig. 14.9, with a high-permeability ring core, also produces a deflecting field of high uniformity. As has already been mentioned, the folding of the turns at the ends of the coil reduces defocusing at the edges of the picture. With increasing deflection the magnetic field of the end sections of the coil turns is aligned increasingly with the beam and produces, hence, an additional focusing effect. The field component along the beam can be made relatively small, however, by making the average radius of the end sections large, as with a folded coil.

The current requirement for a coil of the type shown in Fig. 14.9 can readily be written down. Since the magnetic field is very small within the high-permeability material, the work done in carrying unit pole around the N turns of the pair of coils is simply that of carrying it along a diameter l_0 of the ring core. Accordingly, the field at the center of the deflection system is given by

$$H = 4\pi NI/(10l_0) \tag{14.9}$$

where I is the current measured in amperes and H the field measured in oersteds. The number of ampere turns can be related to the angle of deflection α and the beam voltage with the aid of Eq. 14.6:

$$NI = \frac{10l_0}{4\pi}\frac{\sin\alpha}{d}\sqrt{2mV/e} = 2.69\frac{l_0\sqrt{V}}{d}\sin\alpha \tag{14.10}$$

It is here assumed that V is expressed in volts. Finally, the fact that $LI^2/2$ represents the energy stored in the magnetic field in joules and that $Al_0H^2/(8\pi)$, where A is the mean window area of the coil turns, measures the same quantity in ergs, leads to the following expression for the self-inductance of the coil in henries:

$$L = \frac{4\pi N^2 A}{l_0}10^{-9} \quad \text{henries} \tag{14.11}$$

Normally, a linear variation in the deflecting current will lead to a pattern with pincushion distortion. Thus, for a perfectly uniform field $\sin\alpha$ is, according to Eq. 14.10, proportional to I, and, for a flat screen, $\tan\alpha$ should be proportional to I. The last condition is very nearly fulfilled in many modern kinescopes. Accordingly, if $T - R$ is the rise time of the deflecting current and α_0 is the maximum angle of deflection, the variation in deflecting angle should be given by

$$\tan\alpha = \left(\frac{2t}{T-R} - 1\right)\tan\alpha_0 \tag{14.12}$$

Thus, for a uniform deflecting field, the variation of current with time should be given by

$$I = k\sin\alpha = \frac{k\tan\alpha_0\left(\dfrac{2t}{T-R} - 1\right)}{\sqrt{1 + \tan^2\alpha_0\left(\dfrac{2t}{T-R} - 1\right)^2}} \tag{14.13}$$

The value of k follows, e.g., from Eq. 14.10. Figure 14.15 indicates the deviation from sawtooth linearity required for straightening out the pattern. The correction will be less than shown if, either, the face plate has an appreciable curvature or the deflecting field is so modified that the angle of deflection, rather than its sine, is proportional to the deflecting field.

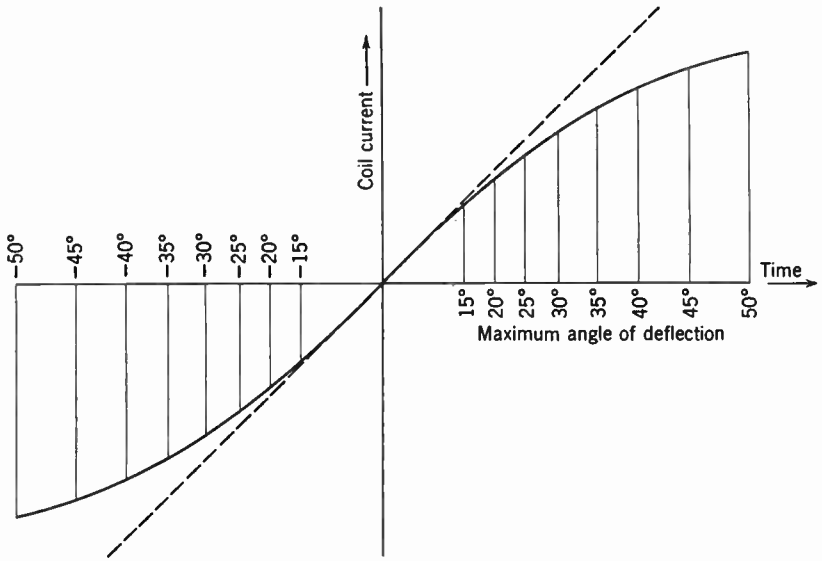


Fig. 14.15. Deviation from Linearity of Deflection Current Required for Producing Scanning Pattern on Flat Screen with Uniform Magnetic Deflecting Field.

The deflecting field quite generally exerts a focusing action on the beam, which adds to its convergence to an increasing degree with increasing deflection. Even if it does not, i.e., if the deflecting field is throughout at right angles to the beam, the image distance increases for a flat face plate with the angle of deflection; hence, a beam sharply in focus at the center of the picture will be slightly out of focus at the edges. This effect may be corrected by impressing currents supplied by the deflection generator on a small auxiliary focusing coil. These currents oppose the constant focusing field, reducing its magnitude just sufficiently to bring the beam into focus at the edge of the picture.

A word should be added regarding the material of the magnetic core. Since the horizontal deflecting current consists of a fundamental oscillation of 15,750 cycles per second, together with harmonic components of relatively high order, it is essential that the core material should

retain its high permeability even at high frequencies. Compact iron cores are unsatisfactory since here eddy currents dissipate energy and shield the interior of the core from the magnetic field. Cores made up of iron wire or thin iron laminations are usable, but have been found inferior to cores of electrolytic or sponge iron powder molded at pressures of the order of 30 tons per square inch. Figure 14.16 shows

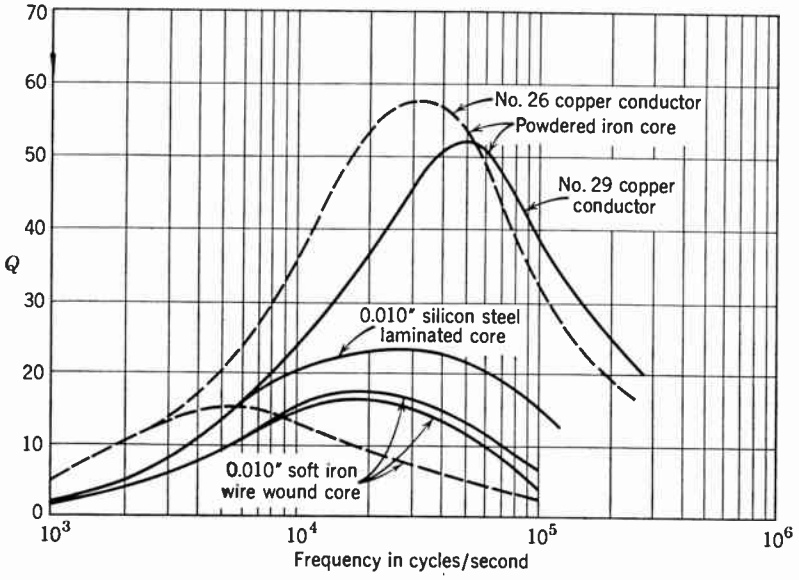


Fig. 14.16. Q of Kinescope Horizontal Deflection Coils with Various Cores and Wire Sizes. (Friend, reference 5.)

the variation with frequency of the Q, measuring the ratio of the peak energy of the magnetic field to the energy dissipation per cycle, of horizontal coils with cores of these materials. "Ferrite" cores, prepared by molding and heat-treating mixtures of iron oxide and other metal oxide powders, have characteristics even superior to powdered iron cores.

14.3 The Deflection Generator. If suitably shaped coils and cores are assumed, or correctly designed electrostatic plates, a perfect scanning pattern can be obtained only if they are supplied with current or voltage having a very specific time variation.

The wave form of the current through the coils producing the vertical deflecting field of the magnetic system must be sawtoothed and have a periodicity equal to the field frequency of the pattern. That supplied to the horizontal coils must also be sawtoothed, but its periodicity is

much higher, corresponding to line frequency. The current in each case builds up linearly with time for most of the cycle, and then returns abruptly to its initial value. It is obviously impossible for the return to be instantaneous, because this would necessitate an infinite voltage across the coils owing to their finite inductance. Therefore, a certain amount of time must be allowed to re-establish initial conditions, which interval is known as the return time. As has already been pointed out, this return time in practice occupies about 7 percent of the vertical deflection period and about 15 percent of the horizontal. For electrostatic deflection it is the voltage wave rather than the current which must have the sawtooth form. However, since, as already mentioned, magnetic deflection is employed almost exclusively in television practice, the following treatment will be restricted to magnetic deflection.

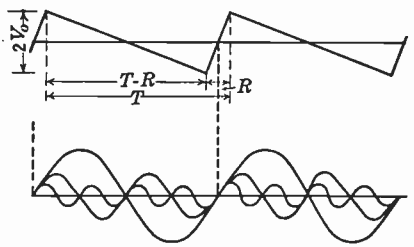


Fig. 14.17. Sawtooth Wave and Its Fourier Components.

The generator which supplies the power to these deflecting elements consists in general of an oscillator, a sawtooth generator, and a deflection amplifier or output stage. The oscillator is controlled by a series of synchronizing impulses which serve to coordinate the timing of the patterns at the transmitter and receiver. In the vertical deflection generator, in particular, the sawtooth generator and oscillator frequently have circuit elements in common even though they represent quite distinct functions. The output stage converts the sawtooth voltage wave provided by the sawtooth generator into a sawtooth current wave of sufficient amplitude to deflect the beam.

In order to formulate better the circuit requirements of the deflecting system, particularly where the current or voltage wave must be amplified, an analysis of the harmonic content of the sawtooth wave is required. The wave form, which is shown in Fig. 14.17, can be represented by a Fourier series as follows:

$$V(t) = b_1 \sin \frac{2\pi t}{T} + b_2 \sin \frac{4\pi t}{T} + \cdots b_n \sin \frac{2\pi n t}{T} \qquad (14.14)$$

$$b_n = \frac{2}{T} \int_0^T V(t) \sin \frac{2\pi n t}{T} \, dt$$

where

$$V(t) = \begin{cases} \dfrac{2V_0 \cdot t}{R} & -\dfrac{R}{2} < t < \dfrac{R}{2} \\ \dfrac{V_0(T - 2t)}{T - R} & \dfrac{R}{2} < t < T - \dfrac{R}{2} \end{cases}$$

The indicated integration leads to the following values for the coefficients of the series:

$$b_n = V_0 \frac{2T^2}{(\pi n)^2 R(T - R)} \sin \frac{\pi n R}{T} \tag{14.15}$$

The amplitudes of the first ten harmonic components for a 15 percent return time ($R = 0.15T$) are given in Table 14.1. It will be seen from

TABLE 14.1. AMPLITUDE OF HARMONICS IN SAWTOOTH WAVE

n	b_n/V_0	n	b_n/V_0
1	0.72154	6	0.01364
2	0.32146	7	−0.00508
3	0.17442	8	−0.01460
4	0.09448	9	−0.01748
5	0.04496	10	−0.01590

this table, and from the form of Eq. 14.15, that the coefficients go through zero at frequencies for which $n = T/R$. Usually, in designing the amplifiers and circuits required to handle the sawtooth voltage or current waves, it is not necessary to consider frequencies beyond the point where the coefficients go through zero, unless the return time is extremely long. The series has no constant term for a centered pattern.

If seven harmonics are required for the horizontal sawtooth wave, the amplifier and circuits for horizontal deflection must respond to frequencies of 15,750 to 110,250 cycles per second; those for vertical deflection, demanding 14 harmonics, must respond to frequencies of 60 to 840 cycles.

14.4 Deflection Circuit and Output Tube. Although the current wave shapes in the vertical and horizontal deflection coils are identical in shape, they differ by a factor of 262.5 in frequency. Since the peak magnetic field which must be provided by the two sets of the deflection coils is approximately equal, it follows that, for equal energy dissipation per cycle in the deflection circuits (equal Q), over 200 times as

much energy must be supplied to the horizontal deflection coils as to the vertical deflection coils.

The energy which must be stored in the magnetic field for a deflection angle α may be calculated from Eqs. 14.10 and 14.11 as well as Eq. 14.7 for the maximum deviation Δ of the beam from the axis within the deflecting field (the tube neck). It is:

$$W = 9.1 \cdot 10^{-8} KDV \sin \alpha (1 - \cos \alpha)$$
$$= 5.7 \cdot 10^{-7} V \sin \alpha (1 - \cos \alpha) \text{ joules} \qquad (14.16)$$

if the coil diameter D (= tube neck diameter) and the ratio $K = l_0/(2\Delta)$ are set equal to 3.8 centimeters and 1.66, respectively. For a deflection angle $\alpha = 30$ degrees and an operating voltage of 10,000 volts the stored energy becomes $3.8 \cdot 10^{-4}$ joule. If this energy is completely dissipated in every cycle, the vertical deflection circuits would have to provide 23 milliwatts of power, whereas the horizontal deflection would have to supply the sizable amount of 6.0 watts. Consequently, whereas power economy is of minor import in the vertical deflection circuit design, it is vital in the horizontal deflection circuit.

A second, rather obvious, difference is that the horizontal coils must be made of a much smaller number of turns and lower inductance and distributed capacity than is permissible in the vertical coils. This is necessary because their natural frequency must lie well above line frequency to permit adequately rapid retrace. In addition, high-inductance coils may demand excessive insulation between the horizontal coils and the core and vertical coils to prevent breakdown during flyback. If, for these reasons, the impedance of the horizontal coils is made relatively small, a proper impedance match to the output tube demands the use of a stepdown transformer in the horizontal circuit. In the vertical circuit, also, a transformer is commonly employed since it improves the impedance match and, eventually, facilitates the application of d-c centering currents to the vertical yoke. In particular, if the available B-supply voltage is high, efficient operation demands a very high impedance output, which is most readily secured by the use of a stepdown transformer.

The transformer in the vertical deflection circuit may be either a transformer with isolated windings, as in the circuit shown in Fig. 14.18, or an autotransformer, as shown in Fig. 14.19. Whereas the isotransformer has the advantage of permitting the application of d-c centering currents, as mentioned above, the autotransformer yields closer coupling and is more compact. Since the centering can be ade-

quately met, in general, by tilting the focusing coil slightly, the last considerations have favored the use of the autotransformer. In either case, the transformer transfers the impedance of the deflection coils,

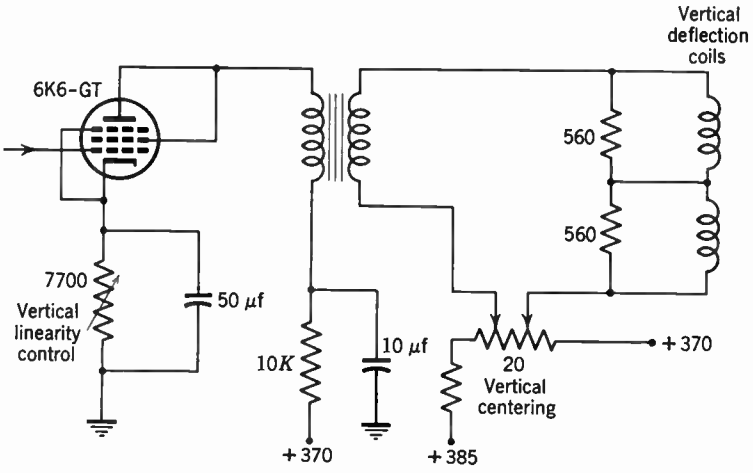


Fig. 14.18. Vertical Deflection Circuit with Isotransformer.

multiplied by the square of the turns ratio (which may be as high as 18:1), into the plate circuit of the output tube.

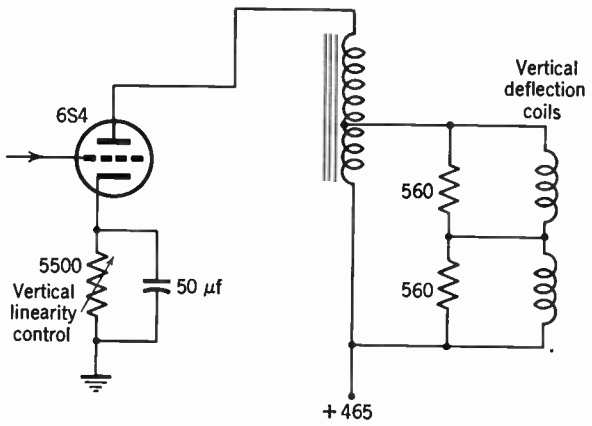


Fig. 14.19. Vertical Deflection Circuit with Autotransformer.

The shape of the sawtooth current delivered by the output tube depends on the signal applied to its control grid and the transfer characteristic of the tube; this, and hence the linearity of scan, may be

modified by changing the d-c bias of the tube by means of the variable cathode resistor.

The current variation during flyback requires special attention. If it is excessively steep, the insulation of the stepdown transformer may prove inadequate, and sparkover to the core may take place. On the other hand, a lower limit to its steepness is set by the requirement that retrace should be completed within a certain fraction of the vertical scanning period. The total flyback time, during which the scanning beam is blanked, is 5 to 8 percent of the return time. However, the retrace must not begin before the vertical synchronizing pulse is received and must be fully completed well before the end of the blanking period, so that the subsequent scan is linear from the start. In view of the required tolerances at both ends, only about 3 percent of the field period is actually available for retrace. If the load presented by the deflection coils were purely inductive, the voltage across the coils would be least for linear retrace; in magnitude the back-voltage pulse across the coils would be just about thirty times the forward voltage during scan.

Actually, however, the resistance of the coils is quite comparable to their inductive reactance; typical values for the inductance and resistance of a vertical-deflection coil are 41 millihenries and 48 ohms, respectively. Hence, for a simple sawtooth current in the coils, the voltage across them consists of the sum of the inductive pulse voltage $L\frac{di}{dt}$ and the resistive sawtooth voltage Ri (Fig. 14.20*a*). It is seen that the maximum excursion of the voltage wave can be materially reduced by increasing the initial steepness of the retrace current variation and letting it taper off (Fig. 14.20*b*).

The question arises finally how the grid-voltage wave applied to the output tube must vary to yield the desired current variation in the coils. The output current of the tube, which is given by the product of its mutual conductance g_m and the grid voltage E_g, must be equal to the sum of the currents through the coils and those through the (reflected) shunt resistances shown in Figs. 14.18 and 14.19 in parallel with the plate resistance of the tube. The plate resistance is significant, in particular, if the output tube is a triode as in Fig. 14.19. If the current through the coils is given by the current wave in Fig. 14.20, the added current through the parallel resistances is given, except for a scale factor, by the resultant voltage wave in the same figure. It is thus seen that the signal which must be applied to the grid of the out-

put tube contains, in addition to the sawtooth, a negative voltage pulse coinciding with retrace time.

In the above consideration of the vertical deflection circuit no attention has been paid to efficiency, for the reasons stated at the beginning of this section. With the horizontal deflection circuit this is no longer

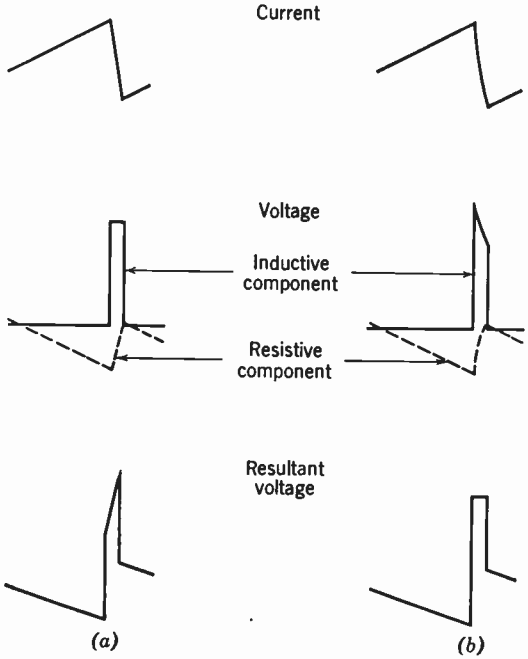


Fig. 14.20. Current and Voltage Variation of Vertical Deflection for (*a*) Linear Retrace, and (*b*) Retrace Minimizing Back Voltage.

permissible. It is convenient to start here, as does Schade,* with the idealized deflection circuit shown in Fig. 14.21. It consists of the effective inductance L of the deflection coils with their effective series resistance r, shunted by a small capacitance C, made up of winding and tube capacitances, the voltage supply E, and the switch S. As the switch is closed, the current through the coil increases nearly linearly up to a peak value which is much less than E/r. As this peak value is reached, the switch is opened. The resonant circuit then executes slightly over one-half period of its natural vibration, whereupon the current has reached a negative peak value which, in view of the dis-

* See Schade, reference 10.

sipation of the network, is slightly less than the initial positive peak value. At this point the switch is closed. The current in the coil decreases linearly (since $E \gg ir$), charging up the battery, and then increases up to the original positive peak value, whereupon the switch is opened again and the process is repeated. It is evident that, with this system, the power supplied by the battery is simply that which is dissipated in the resistance of the circuit. A simple calculation shows

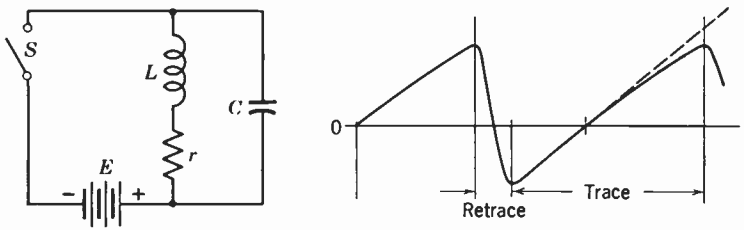


Fig. 14.21. Idealized Horizontal Deflection Circuit and Corresponding Coil Current Variation. (Schade, reference 10, Fig. 1.)

that the ratio of the energy dissipation per cycle to the maximum energy stored in the coil is, on this basis,

$$\frac{\pi}{Q} + \frac{2}{3}\frac{r}{L}(T - R)$$

where Q is the Q-value of the deflection coil at its resonant frequency (set equal to $1/(2R)$), T the horizontal scanning period, and R the retrace time. This formula applies for relatively high values of Q. By contrast, in the vertical deflection circuit the value of Q is so low that the energy stored in the coil is fully discharged in the course of the retrace time; throughout the trace the current flow in the coil is unidirectional and controlled by the output tube.

If the current variation in the coil is to resemble that shown in Fig. 14.21, it is essential that the switch permits the flow of current in both directions. In view of its high operating speed and the need for control by the discharge tube it is necessarily electronic. Since vacuum tubes permit current flow in one direction only, the switch must contain at least two tubes. Figure 14.22 shows a basic deflection circuit in which a diode shares the switching action with the grid-controlled output tube. The battery B compensates the voltage drop across the output tube; the relatively small drop across the diode may, for convenience, be neglected.

The sequence of operations in this circuit is the following. With the arrival of the negative pulse on the output tube, both output tube and diode are cut off and remain so for retrace time. The coil current swings from its maximum positive value to its maximum negative value, which is slightly less. $L\frac{di}{dt}$ is negative, so that the voltage across the coil has the same polarity as the battery E, and a large back voltage is impressed across the diode.

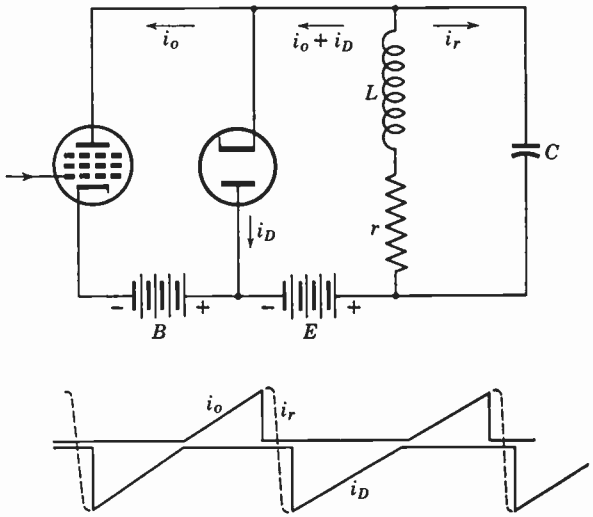


Fig. 14.22. Deflection Circuit with Electronic Switch Consisting of Output Tube and Damper Diode.

At the end of retrace the free oscillation continues until $\frac{di}{dt}$ has attained a sufficiently positive value that $L\frac{di}{dt} + ri = -E$. At this point, which occurs when the half period of free oscillation has been exceeded by about 5 percent, the diode begins to conduct. This maintains the voltage drop across the coil at a constant value. Except for the resistive component of the coil, the current in the coil would decrease at a uniform rate, charging the battery E. Actually, the rate of current decrease becomes less with time.

As the current through the coil approaches the value zero, the sawtooth voltage applied to the output tube grid causes the output tube to conduct, sending a current through the coil in an opposite direction.

The diode remains blocked off as long as the current supplied by the output tube is such that:

$$L\frac{di}{dt} + ri + E \leq 0$$

Thus, the current in the coil continues to build up to its peak value at a slightly decreasing rate.

In the circuit in Fig. 14.22 the battery E is gradually discharged, since the charging currents transfer less charge than the discharging

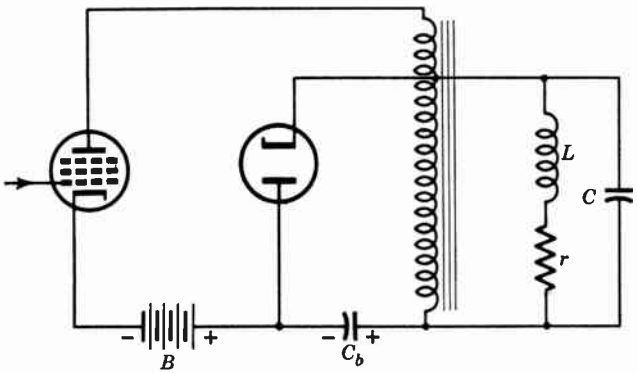


Fig. 14.23. "Bootstrap" Circuit with Boost Capacitance.

currents (in the second half of the trace). The circuit can, however, readily be modified so that integrated charge and discharge currents become equal. Then the battery may be replaced by a capacitance, as shown in Fig. 14.23. Here an efficient autotransformer has been inserted into the circuit. Since its inductance is very high compared to that of the coil, the circuit during the charging period, when the output tube is blocked off, is practically unchanged. On the other hand, during the discharge period, when the diode is blocked off, the current provided by the condenser is less than that flowing through the coil by the turns ratio of the autotransformer. In practice the turns ratio is chosen much larger than needed for compensation of the resistive component of the coil. The "boosting" capacitance C_b normally supplies power also to the vertical deflection output stage and the high-voltage supply for the kinescope, as will be seen later. If the turns ratio is made too large, the charging period is automatically decreased and the horizontal deflection currents (as well as the vertical deflection currents and the kinescope voltage) are increased.

The advantage of the two-way switch circuits operating with a low-loss coil become particularly evident when the energy dissipation in the output tube is considered. For optimum efficiency the B voltage on the deflection tube will be chosen so that the tube operates, during the second half of trace, as close to the knee of its characteristic as possible, e.g., at 40 volts plate voltage. For a completely loss-free circuit (Fig. 14.24*a*) current will flow for just half the period of trace, and the average current will be one-fourth of peak current. For a circuit with complete energy dissipation during retrace (Fig. 14.24*b*),

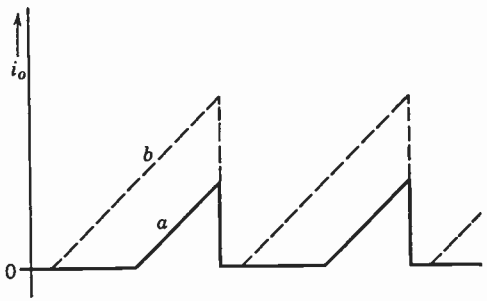


Fig. 14.24. Output-Tube Current Variation for Completely Loss-Free Deflection Circuit (*a*) and Completely Damped Circuit (*b*).

such as is customary for vertical deflection, the peak current is twice as great and the average current half the peak current. Thus the power dissipation at the output tube anode is four times as great in the second case as in the first. For example, if the coil, with an inductance of 8 millihenries, is coupled directly to the output tube, it will require a peak current, in the loss-free circuit, of about 0.275 ampere, so that here the anode dissipation would be 2.75 watts. For the second circuit the dissipation would be 11 watts. Whereas the anode dissipation would be the only energy to be provided by the B supply in the loss-free circuit,* the completely damped circuit would have to be provided with an additional 80-volt battery to compensate the yoke voltage drop and replace the boost capacitance in the loss-free circuit. The corresponding energy loss of 22 watts would arise in the resistive components of the deflection coil.

* The voltage drop across the diode, which is proportional to the two-thirds power of the current passing through it, is in effect added on to the drop across the resistive component of the yoke during the charging period. The corresponding energy dissipation is generally smaller than that in the coil and much smaller than that in the deflection tube.

The principal shortcoming of the circuit in Fig. 14.23 is its nonlinearity. Figure 14.25 indicates ways in which this can be overcome. Without any compensation the resistive component of the coils together with

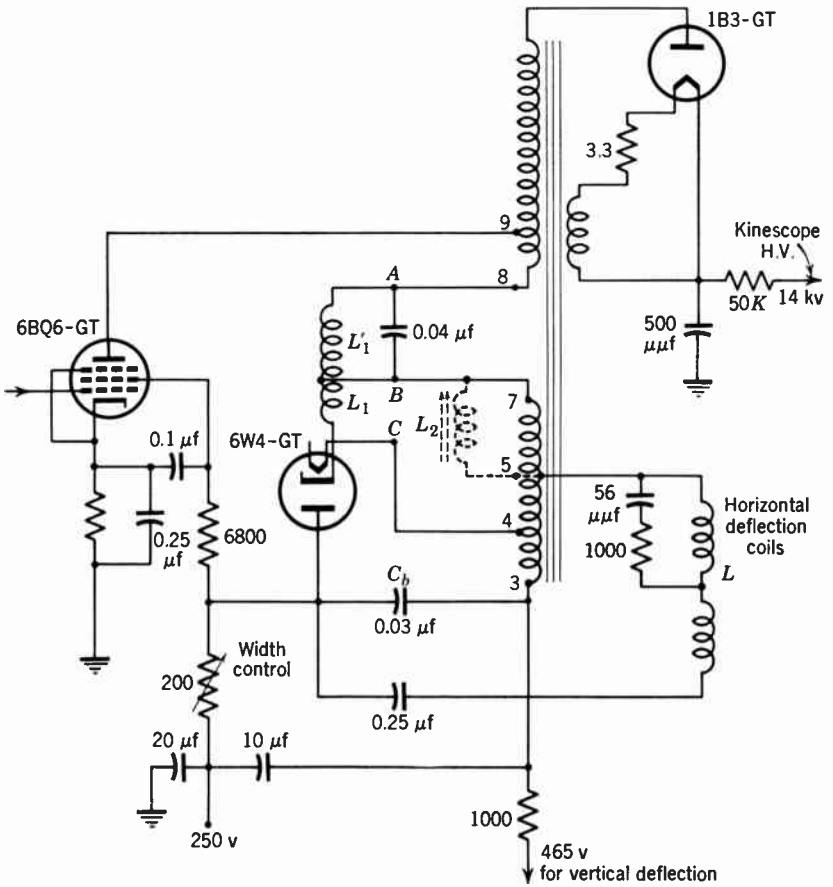


Fig. 14.25. A Practical Horizontal Deflection Circuit, Incorporating a Saturating Inductance for Trace Linearization and High-Voltage Supply for Kinescope.

the voltage drop in the damper diode cause the rate of change of the coil current to decrease in the course of the trace. The method employed to counteract the resistance in the circuit consists in the insertion of a saturating choke, whose inductance decreases as the current increases. The change in voltage drop across it, $L_1 \dfrac{di}{dt}$, with change in current is opposite in sign to the voltage drop Ri across the effective

resistance and may be made equal to it. The choke section BC controls linearization in the charging cycle, section AB in the discharge cycle. Section BC is made smaller than section AB, since the former is in the secondary circuit, the latter in the primary circuit of the transformer. To a first approximation the current in the primary circuit is smaller, the resistive voltage drop larger than in the secondary circuit by a factor equal to the turns ratio. Thus the turns ratio of the linearity control should be of the same order as for the transformer.

Figure 14.25 also indicates some other features of a practical horizontal deflection circuit. Thus, it contains as width control a variable resistance in the B-supply lead. Since the vertical deflection power is derived from the positive side of the boost capacitance, increasing this resistor reduces both deflection amplitudes without altering the aspect ratio; the latter may be adjusted with the aid of the height control in the vertical deflection generator, which is commonly inserted in the coupling between the vertical sawtooth generator and the deflection tube. The choke L_2 represents another form of width control; it places a variable inductance in parallel with the yoke, which diverts part of the charging and discharge current.

During retrace time the full pulse voltage across the yoke is impressed across the diode. For a 14-millihenry yoke, a horizontal deflection angle of 53 degrees ($\alpha = 26.5$ degrees, maximum deflection angle at the corners 66 degrees), and an operating voltage of 14 kilovolts Eq. 14.16 yields for the peak current a value of 0.232 ampere. For a free oscillation period of 70 kilocycles, corresponding approximately to a 12 percent retrace time, this leads to a maximum voltage drop $2\pi viL = 1430$ volts. Empirically, the maximum voltage drop is found to be as large as 2200 volts. Since the boost voltage is about 225 volts and the B voltage is 250 volts, connecting the heater to the cathode would demand an insulation of the heater transformer adequate to withstand a peak voltage of 2700 volts. On the other hand, grounding the heater would impress a similar voltage, which is well above the tube rating, between heater and cathode.

In practice the heater is connected about two-thirds down on the secondary of the transformer. In this fashion the d-c voltage between heater and cathode is nullified, and the pulse voltage reduced to about 1500 volts. At the same time the voltage to ground, and the primary of the power transformer, is limited to about 1200 volts. Finally, the effective capacitance of the heater winding in the yoke circuit is reduced to about one-tenth its actual capacitance in view of the turns ratio between the transformer coil sections 4–3 and 5–3.

It is customary to utilize the high pulse voltage generated across the horizontal deflection transformer during retrace to charge the high-voltage supply for the kinescope. Since the beam energy utilized for the generation of the picture in the kinescope within a line period is, on the average, only a small fraction of the peak energy stored in the deflection coils, this has merely the effect of decreasing the effective Q, or increasing the damping, of the deflection circuit. For example, if the mean beam current is 100 microamperes and the accelerating potential 14 kilovolts, the energy supplied to the fluorescent screen in a single trace is $7.5 \cdot 10^{-5}$ joule, whereas the energy (for a horizontal angle of 53 degrees) stored in the horizontal deflection coils is $3.8 \cdot 10^{-4}$ joule; the assumed beam current corresponds to high average brightness.

Figure 14.25 shows the manner in which the kinescope voltage is obtained from the deflection transformer. Since the maximum voltage which may be applied to the 6BQ6-GT tube is 4000 volts and the primary winding of the transformer must be adjusted so that this value is not exceeded on flyback, it is evident that the transformer must be provided with an additional high-voltage winding with a turns ratio of the order of 3.5 for an output voltage of 14 kilovolts. The opposite peaks of the voltage pulses charge, through a diode, a capacitance, replacing the charge supplied to the kinescope in the preceding line scan. The capacitance is made large enough that the decrease in anode voltage in the course of the line scan is negligible. For example, for an average beam current of 100 microamperes and a capacitance of 500 micromicrofarads the total change in anode voltage is

$$\Delta V = \frac{10^{-4} \cdot 0.85}{5 \cdot 10^{-10} \cdot 15{,}750} = 10 \quad \text{volts}$$

Whereas this decrease in voltage *during* line scan is negligible, the change in anode voltage with picture brightness can be quite appreciable. The withdrawal of energy during retrace results, as already mentioned, in an increased damping of the free oscillation of the deflection circuit. As the voltage peak occurs ¼ period after the free oscillation is initiated, it is reduced in amplitude as the kinescope beam current is increased. For example, for the circuit in Fig. 14.25 an increase in beam current from 0 to 140 microamperes results in a reduction of anode voltage from 14 to 12.9 kilovolts.

This decrease in the anode voltage is accompanied by a reduction in the horizontal deflection currents and, since the vertical deflection depends on the value of the boost voltage, also by a reduction in the

vertical deflection current. An effort is made, in the design of the deflection circuits, to adjust the values of the circuit components so that the decrease in deflection currents just compensates the drop in anode voltage, leaving pattern size and aspect ratio unaltered. Nevertheless, since perfect compensation cannot be achieved in practice, it is advisable to keep the change in anode voltage with picture brightness within a range of about 10 percent. This is all the more necessary with magnetically focused electron guns, since the drop in anode voltage results in a defocusing of the beam. With electrostatic focusing it is possible to derive the focusing voltage from a potentiometer supplied by the anode voltage and to avoid, in this manner, any defocusing effects.

In the present treatment of the horizontal deflection circuit emphasis has been placed on one particular example which illustrates several of the basic design principles. Circuits proposed and employed in practice vary widely, in the coupling of the coils to the output tube, the damping circuit, methods of linearization, and width control. For some of these variations the reader is referred to the references given at the end of this chapter.

14.5 Generation of the Sawtooth Signal. In both horizontal and vertical deflection circuits a sawtooth signal must be applied to the grid of the output tube. Its amplitude determines the amplitude of deflection, its shape the linearity during all or part of the trace. Finally, the sharpness of cutoff affects materially the efficiency of the horizontal deflection circuit and the initial linearity of trace in the vertical deflection circuit.

In essence, the sawtooth is generated quite generally by the charge of a capacitance through a resistance, terminated periodically by discharge through a discharge tube. The insertion of a "peaking" resistance in series with the capacitance serves to add a negative pulse voltage to the sawtooth transferred to the grid of the output tube. The discharge itself is triggered by a pulse. Whereas this pulse could, in principle, be the synchronizing pulse transmitted with the picture signal, separated from the latter and amplified, it is commonly supplied by an oscillator, which in turn is controlled by the synchronizing signal. This procedure decreases the effect on the scanning pattern of noise, which may cause premature triggering of the deflection or even suppress one of the synchronizing pulses. In addition, it may yield pulses of the required shape and amplitude with a smaller number of tubes. Not infrequently, the function of the discharge tube and the oscillator may be combined in one tube.

Automatic frequency control in the horizontal deflection circuits serves further to improve noise immunity of the scanning pattern. Whereas automatic frequency control circuits differ in detail, they utilize the phase difference between the synchronizing signal and the

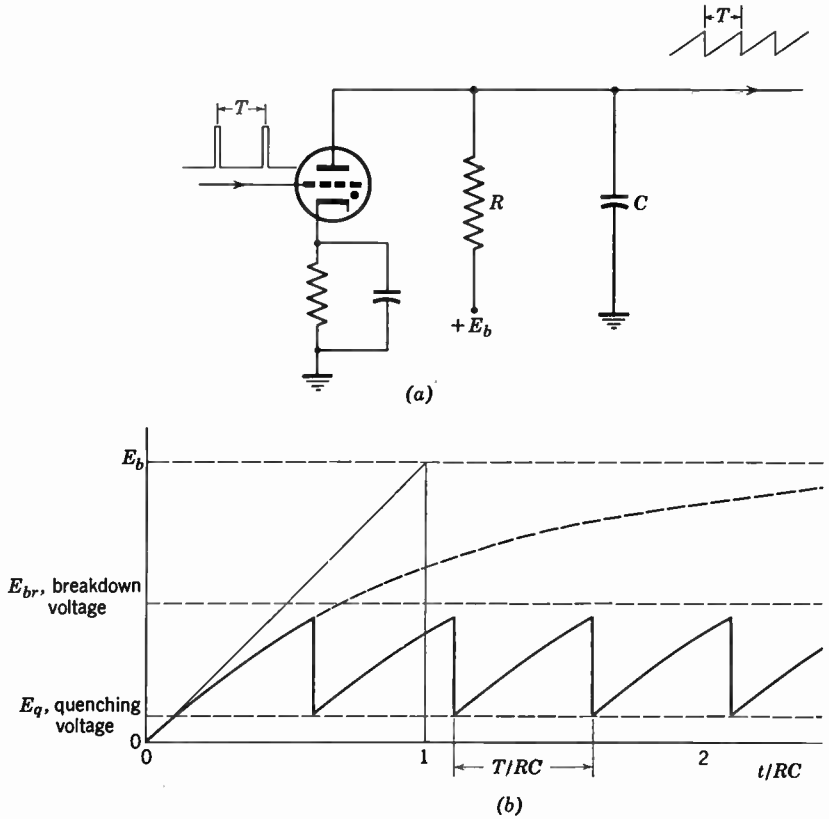


Fig. 14.26. Thyratron Sawtooth Generator. (*a*) Circuit. (*b*) Output voltage.

local oscillation, averaged over many cycles, to correct the frequency of the latter. Since the vertical synchronizing pulse is itself generated by a process of integration and is thus inherently more immune to noise, the vertical deflection circuits do not demand this refinement. Automatic frequency control circuits will be discussed in the next section.

One of the simplest sawtooth generators, which incorporates the function of the control oscillator, is the gas-tube circuit shown in Fig. 14.26*a*. Whereas, at present, it has practically disappeared from Amer-

ican television practice, it still enjoys considerable popularity as a time base for oscilloscopes. A rare-gas thyratron serves as discharge tube. The capacitance C, which may be assumed to be discharged at the beginning of the cycle, is charged through the resistance R by the B supply (Fig. 14.26*b*) until it reaches the breakdown voltage of the gas tube; the degree of curvature of the charging curve:

$$E_c = E_b(1 - \epsilon^{-t/(RC)}) = \frac{E_b}{RC}\, t\left(1 - \frac{t}{2RC} + \frac{t^2}{6(RC)^2} - \cdots\right) \quad (14.17)$$

$$E_c \ll E_b$$

depends on the ratio E_{br}/E_b of the breakdown voltage to the voltage of the B supply, whereas the period of the charging cycle is given, to a first approximation, by the product of this ratio with the time constant RC.

The natural period of the circuit, $\cong RCE_{br}/E_b$, is made slightly larger than the interval between synchronizing pulses, so that the latter, applied in positive polarity to the thyratron grid, cause the discharge to take place slightly in advance of the free breakdown period. Once breakdown has been initiated the gas tube continues to conduct, irrespective of the grid voltage, until the plate voltage has reached a very low value, of the order of the ionization potential of the gas; for argon this is, e.g., 15.7 volts. At this point the discharge is quenched, and the charging process is repeated.

The gas tube has fallen into disfavor in television practice because of the difficulty of maintaining the return time constant. In addition, the time required for quenching the discharge, which involves the drift of the gas ions to the walls and electrodes of the tube, tends to be too long for the generation of a satisfactory sawtooth in the horizontal deflection generator.

The employment of hard-vacuum tubes, which avoids these difficulties, requires greater circuit complexity. If discharge pulses of adequate amplitude are available, it is true, the gas tube in Fig. 14.26 may be replaced by a hard-vacuum triode with little other change (Fig. 14.27). The blocking condenser C_b and the grid resistance R_g are chosen so that their time constant is large compared with the period of the sawtooth. As the positive pulse reaches the grid, it drives it positive and causes it to draw grid current until its potential is brought down to zero relative to the cathode. At the same time, the tube presents a low-impedance path for discharging the condenser C for the exact duration of the pulse.

Relaxation oscillators serve, in general, to supply the discharge pulses. In these oscillators, of which the gas-tube circuit in Fig. 14.26 is an example, the period of oscillation is determined by the discharge of a capacitance or inductance through a resistance. The wave shape of their output tends to be irregular compared with that of a sine wave oscillator and may provide the desired pulses or sawtooth directly.

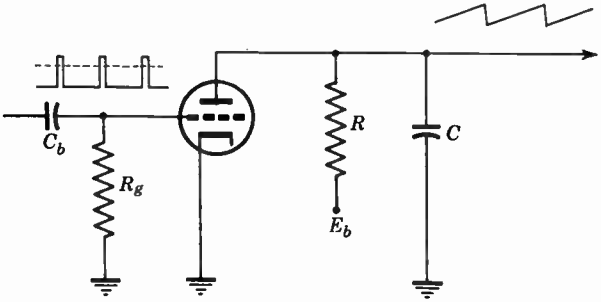


Fig. 14.27. Hard-Vacuum Tube Sawtooth Generator.

A relaxation oscillator which finds wide employment in television practice, particularly as a frequency divider, is the multivibrator (Fig. 14.28). It consists of two vacuum tubes, the grid of one being tied to the plate of the other. Assume that, to begin with, the grid of V_1 is

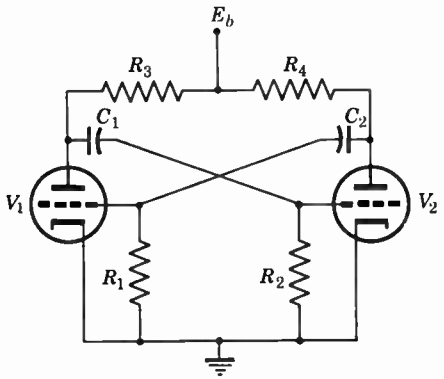


Fig. 14.28. The Multivibrator.

just below cutoff and increasing in potential. As V_1 begins to draw current, its plate becomes less positive and, correspondingly, a negative voltage is applied to the grid of V_2. This suppresses current in V_2 and, owing to the capacitative coupling between the plate of V_2 and the grid of V_1, causes V_1 to draw still more current, its grid potential being

stabilized at zero by grid collection. Tube V_1 continues to draw current in this fashion and V_2 to be cut off until C_1 has discharged sufficiently through R_2 and V_1 that the grid voltage of V_2 has risen once more to the cutoff point. At this point the current in V_2 rises rapidly and that in V_1 is cut off, and the cycle is repeated with the roles of V_2 and V_1 interchanged.

It is immediately evident that a small positive pulse applied to both grids shortly before the cutoff point of one of them has been reached will precipitate the conducting condition in the corresponding tube. The multivibrator is thus readily controlled by synchronizing pulses. Its period is determined by the two time constants C_1R_2 and C_2R_1 and

$C_2R_1 >> C_1R_2$

Fig. 14.29. Voltage Variation across Coupling Condenser C_1 of Multivibrator.

the ratio of the plate potentials in the conducting and the nonconducting phase. Since the time of charge of C_1 through R_2 and R_3 is determined by the time constant C_2R_1, the time of discharge through V_1 and R_2, by C_1R_2, the voltage variation across the condenser C_1 is a sawtooth whose rise time and decay time are proportional to these two time constants (Fig. 14.29).

Figure 14.30 shows a slight modification of the standard multivibrator circuit * which yields a quite linear sawtooth with short return time as one of the plate voltages and a sequence of narrow pulses at the other plate. The capacitance C_2 charges through the relatively large resistance R_4 during the trace, at which time V_1 conducts and V_2 is cut off. V_2 begins to conduct when capacitance C_1 has been discharged through V_1, R_k, and R_2 sufficiently so that the grid of V_2 reaches cutoff. As V_2 begins to conduct, the current through V_1 is reduced by the increased voltage drop across R_k; this increases the plate voltage of V_1 and hence the grid voltage of V_2 further until the grid of V_2 draws current, charging C_1 rapidly. At the same time C_2 discharges rapidly through V_2 and the small resistance R_k. As C_1 is charged and the grid current is reduced, the bias provided by R_k decreases, and V_1 begins to conduct. The corresponding drop in its plate voltage is communicated to the grid of V_2, which is consequently cut off; thus the

* See Potter, reference 13.

charging cycle of C_2 recommences. It is seen that the amplitude of the sweep is determined by R_4, its period by the time constant $C_1(R_2 + R_k)$. Good linearity of sweep is achieved as long as $C_2R_4 \gg C_1(R_2 + R_k)$. Since, as has been seen, a certain curvature

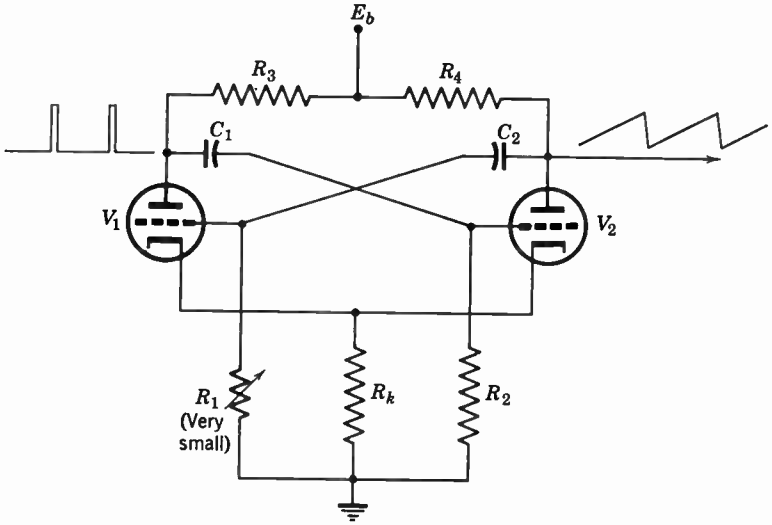


Fig. 14.30. Multivibrator Modified to Supply Sawtooth Output. (Potter Oscillator.)

of the sawtooth is desirable for most horizontal deflection circuits, this condition is not necessarily fulfilled in practice. Thus, Fig. 14.31*a* shows an output wave shape suitable for application to the deflection

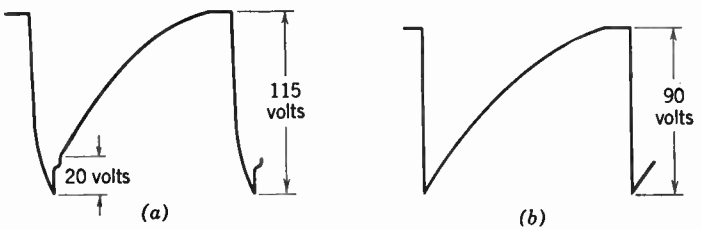


Fig. 14.31. Voltage Wave Shapes Delivered to Output Tube Grid by (*a*) Potter Oscillator and (*b*) Blocking Oscillator.

tube grid of the circuit in Fig. 14.25 which may be obtained from the oscillator shown in Fig. 14.30.

Perhaps the most generally useful sawtooth generator is the blocking oscillator, which provides an exceedingly rapid retrace pulse, as exem-

plified by the wave shape shown in Fig. 14.31*b*. A blocking oscillator circuit suitable for vertical deflection is shown in Fig. 14.32. The characteristic feature of all oscillators of this type is the tight regenerative transformer coupling between grid and plate. During trace the tube is cut off and the capacitance C is charged through the resistance R which is made variable to provide a height control. At the same time the negatively charged capacitance C_g discharges through the grid resistance R_g, which is adjusted for matching the natural period of the

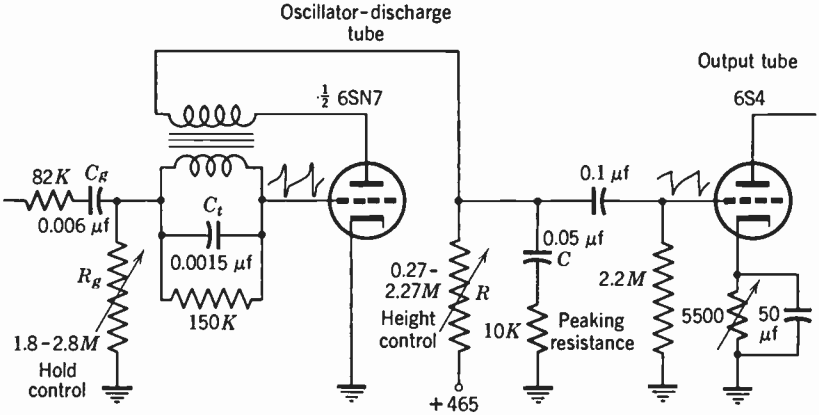


Fig. 14.32. Blocking Oscillator for Vertical Deflection.

oscillator to the field period ("hold control"). As the tube begins to conduct, the plate voltage drops, which causes a more rapid rise in grid voltage and tube conductance. Finally, grid current is drawn and the capacitance C_g charged negatively, opposing the further rise of the grid voltage. As the anode voltage drops, furthermore, the mutual conductance of the tube and the rate of change of the plate current decrease, so that the voltage across the transformer decreases and changes sign. Thus the tube is cut off, and C, discharged through the tube during its conducting period, begins to be charged once more. The purpose of the capacitance C_t across the transformer is to increase the retrace time, which depends on the natural period of the transformer; the shunt resistance is added to damp the transformer oscillations sufficiently so that the grid voltage is not driven above cutoff by them.

Characteristics similar to those of the blocking oscillator may also be obtained by employing feedback from the deflection transformer to the grid of the oscillator tube, as indicated in Fig. 14.33. The delay pro-

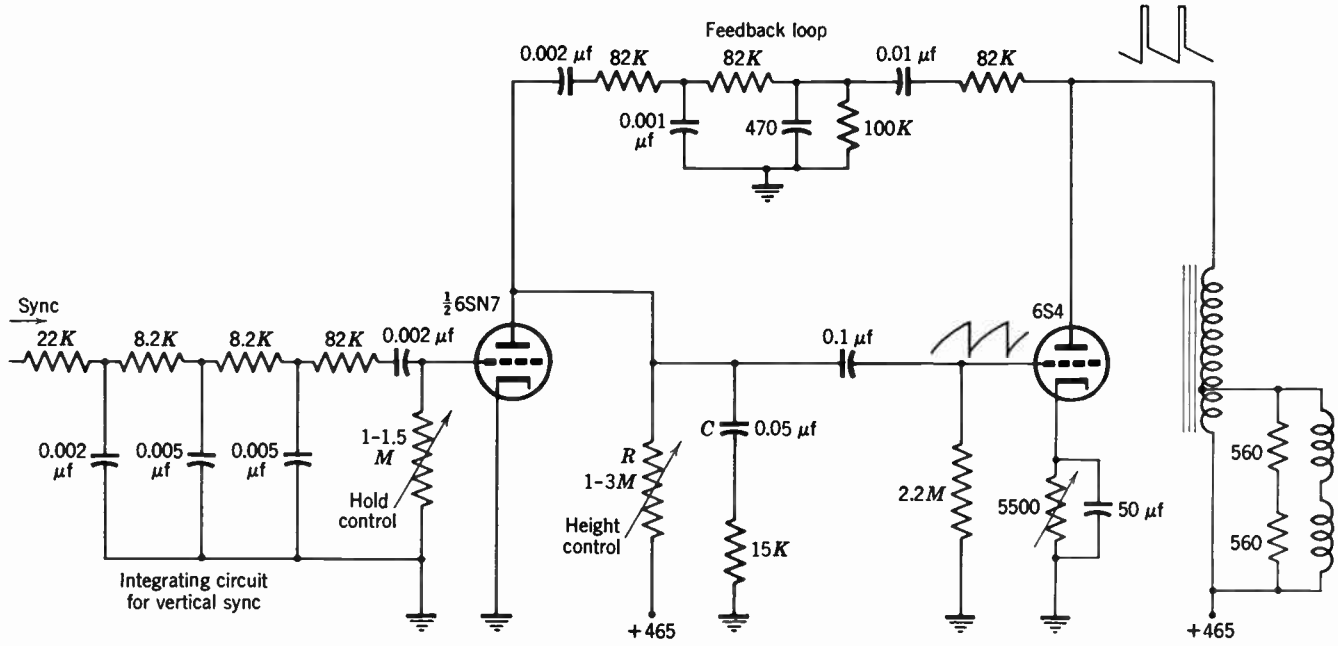


Fig. 14.33. Feedback Oscillator for Vertical Deflection.

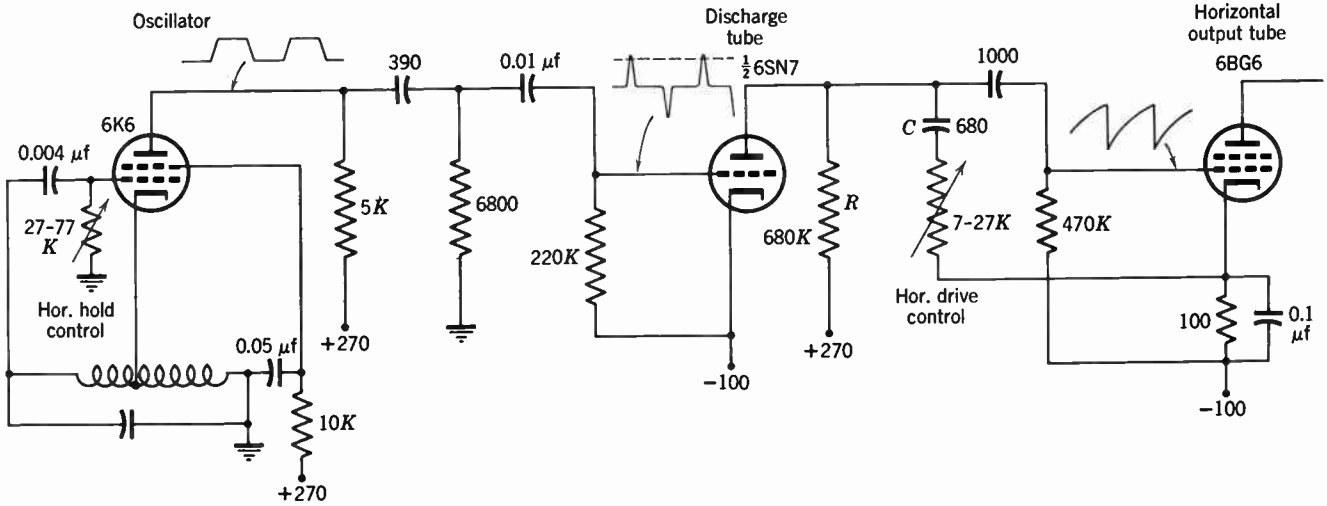


Fig. 14.34. Generation of Sawtooth by Sine-Wave Oscillator.

vided by the feedback network controls the voltage variation of the grid of the output tube during return time.

Sine wave oscillators are employed only in connection with certain types of automatic frequency control. Figure 14.34 indicates how the sawtooth voltage may be derived from them. The oscillator tube is a beam tetrode, whose screen-grid circuit constitutes a Hartley oscillator. The amplitude of the control grid voltage, e.g., 130 volts peak-to-peak, is sufficiently high so that the plate current is a square wave. This square wave is differentiated by a resistance-capacity circuit, which applies the resulting alternating positive and negative peaks to the discharge tube grid. The negative peaks are without effect, since the positive peaks, through grid current, fix the grid bias so that the maximum excursion of the positive peak corresponds very nearly to zero grid-cathode potential. The capacitance C, which is charged during the trace, is discharged through the tube during the positive peaks.

14.6 Automatic Frequency Control.* In all automatic frequency control systems a d-c, or slowly varying, voltage is generated by a phase comparison of the deflection voltages (or deflection control oscillations) and the synchronization pulses, and this slowly varying voltage is utilized to shift the frequency of the control oscillator in a direction tending to correct the phase difference.

The earliest system, commonly designated as the sawtooth system, is represented in Fig. 14.35.† The synchronizing pulses are applied to a double diode V_2 through a center-tapped transformer in such polarity that conduction takes place at the peak of the pulses. The capacitances C_1 and C_7 in the diode leads are charged by the diode current so that conduction is restricted to these instants. The symmetry of the circuit is such that, during these periods of conduction, K_1 and P_2 tend to assume the instantaneous potential of the transformer center tap, point A. Thus, if a sawtooth, derived from the pulses on the deflection transformer, is applied to point A, condensers C_2 and C_3 will be charged positively or negatively, depending on whether a positive portion (point 3) or a negative portion (point 2) of the sawtooth coincides with the synchronizing signal. The potential developed across the capacitances C_2 and C_3 determines the grid voltage, and hence the plate potential of V_3, which in turn provides grid bias to the blocking oscillator V_4. The period of the latter is reduced as the grid bias is made more positive, lengthened, as it is made more negative, so that the circuit tends

* See Clark, reference 16.

† See Wendt and Fredendall, reference 15.

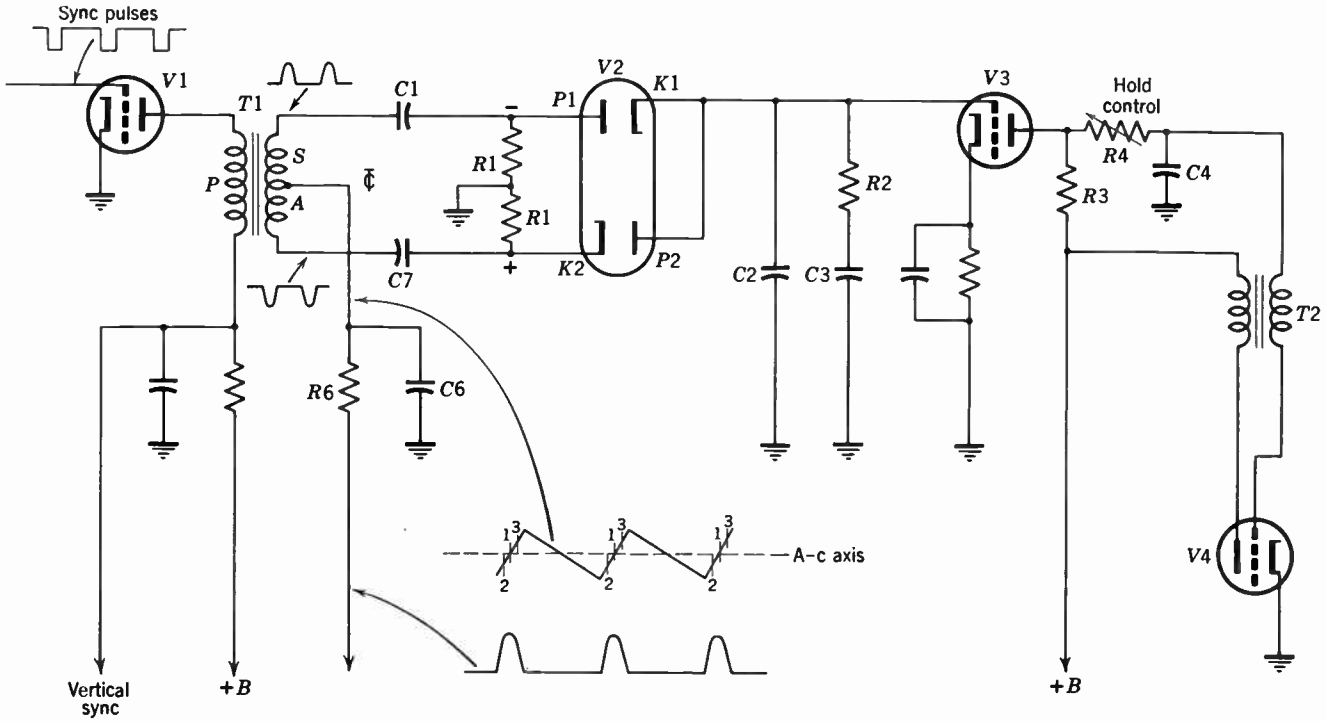


Fig. 14.35. Sawtooth Automatic Frequency Control System. (Clark, reference 16.) (Courtesy of *Proceedings of the Institute of Radio Engineers.*)

to shift the sawtooth to position 1, where the pulse coincides with the center of the retrace period.

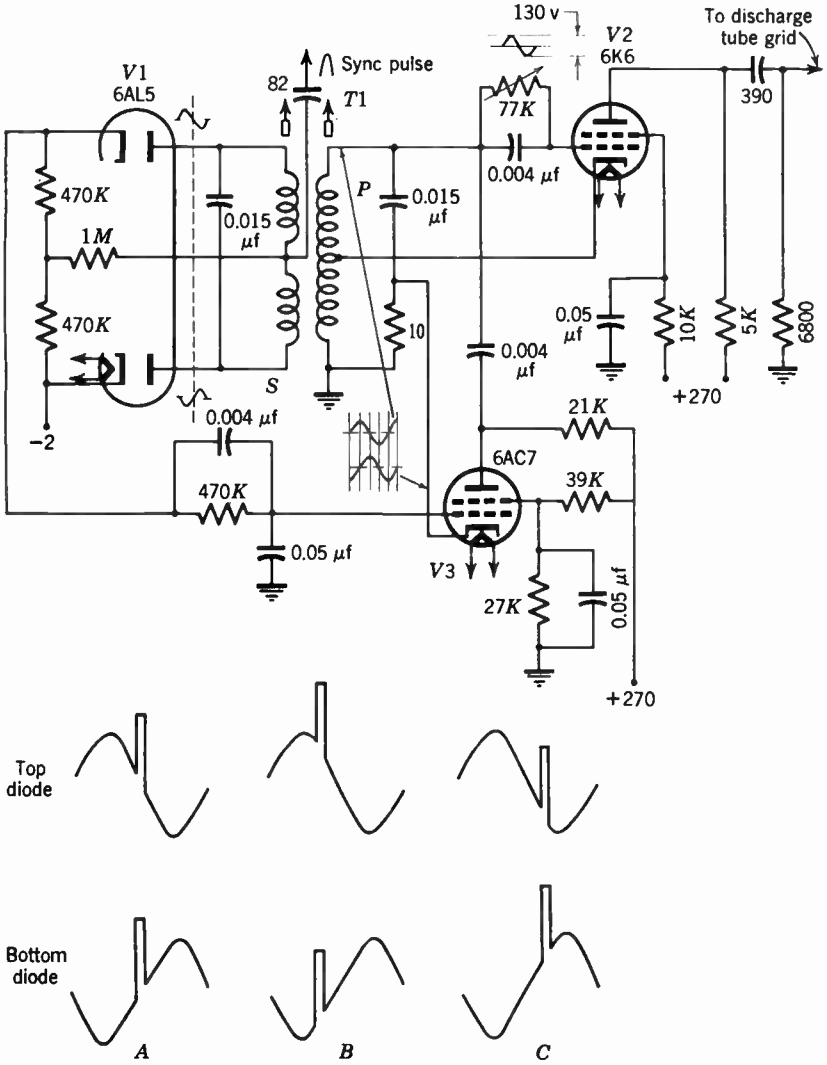


Fig. 14.36. Sine-Wave Automatic Frequency Control System. (Clark, reference 16.) (Courtesy of *Proceedings of the Institute of Radio Engineers.*)

In the sine wave system, represented in Fig. 14.36, the output of a stable sine wave deflection control oscillator is applied through a closely coupled transformer to the plates of a double diode. The synchronizing

signals are applied to the center tap of the secondary of the same transformer. Thus, at the plates, identical synchronizing pulses are superposed on the sine waves of opposite polarities. As a result, the voltages developed across the load resistors of the double diode integrate in the course of a cycle to zero if the pulses coincide with the zero point of the sine wave (A), but average to a net positive or negative value (B and C) if the zero point of the sine wave leads or lags relative to the pulse. The output of the comparator V_1 is integrated by a resistance-capacity network and applied to the grid of the reactance tube V_3, which, depending on its conductance, modifies the effective value of the capacitance in the resonant circuit of the Hartley oscillator V_2 and, hence, its frequency. It should be noted that the secondary of the transformer should be tuned to a frequency below the frequency of oscillation to obtain the proper phase relationship between the ultimate sawtooth and the synchronizing pulses.

Whereas the last system, in particular, is highly satisfactory from the point of view of noise immunity and speed of attaining synchronism, both the preceding systems demand a relatively large complement of tubes and circuit components. The pulse-time system, shown in Fig. 14.37, is far more economical. It employs a single double triode to deliver the sawtooth signal and maintain it in proper phase relative to the synchronizing signals.

The system depends on applying the synchronizing pulse superposed on the peak of the sawtooth to the oscillator control grid and employing the integrated cathode current to supply bias for a blocking oscillator, namely, the second half of the double triode. The magnitude of the integrated cathode current is given essentially by the area of the compound wave lying above cutoff; the peak is maintained at a constant voltage by grid current. This area is reduced as a larger portion of the synchronizing pulse slides off on the steep front of the sawtooth, increased as it shifts more toward the flat plateau at the peak. The control voltage characteristic obviously depends materially on the exact shape of the sawtooth fed back to the control section grid; thus the variable shunt capacitance serves to round the peak of the sawtooth and to extend the range over which the oscillator will snap into synchronism with the sync signals, simultaneously reducing the sensitivity. A fine adjustment for the magnitude of the bias voltage supplied to the oscillator is provided by a potentiometer in the plate circuit of the control section.

The blocking oscillator is somewhat unconventional in design; it employs an autotransformer with B voltage applied to mid-tap. A

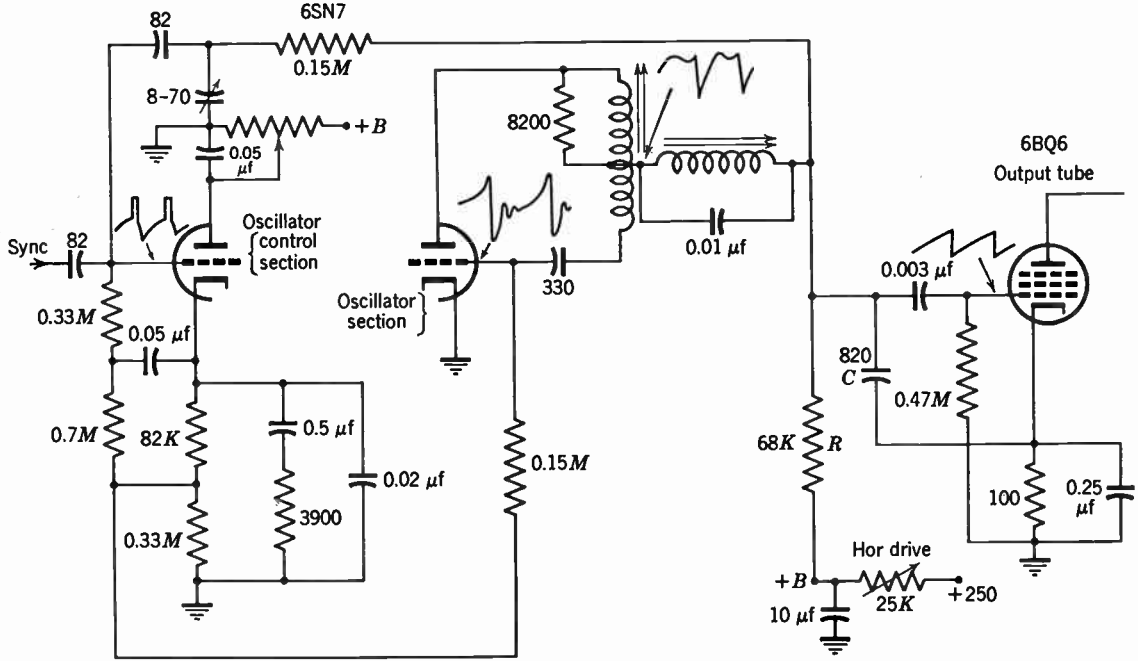


Fig. 14.37. Pulse-Time (Width-Control) System of Automatic Frequency Control.

third coil with shunt capacitance acts as filter for the damped oscillations taking place in the plate circuit during tube cutoff, and at the same time, by superposing its natural oscillation, near line frequency, on the grid voltage, sharpens the grid voltage variation near cutoff. In this manner the frequency of the blocking oscillator is stabilized. The condenser C and charging resistance R play the same role as in other sawtooth-generating circuits; discharge takes place through the horizontal oscillator tube.

14.7 Synchronization. The impulses which determine the phase of the control oscillators of the horizontal and vertical deflection generators accompany the picture signal, being a part of the complete video signal. These synchronizing impulses must meet the following requirements:

1. They must not interfere with the picture.
2. They must be easily separable from the picture signal.
3. They must be such that vertical and horizontal control pulses can be distinguished electrically.
4. They must be capable of controlling the oscillator in the presence of ordinary interference.

The first condition is met by placing the synchronizing pulses within the return time of the scanning beam, as was described in Chapter 6.

As was pointed out in the chapter mentioned, the synchronizing signals cannot be distinguished from the picture signal by their frequency content and must, therefore, be selected from the video signal by virtue of some other attribute. They are characterized by their time of occurrence and their amplitude. The time cannot readily be used as a means of selection, because it is the pulse which times the system itself. Therefore, of necessity, amplitude selection is used.

Separation of vertical and horizontal control pulses can be effected by means of differences of shape, duration, or amplitude of the two types of pulses. Choice of the most suitable means involves circuit considerations which will be discussed in a succeeding section.

The effect of interference upon synchronization depends upon the nature and amplitude of the interfering signals. If the amplitude of the synchronizing signal is sufficient to insure its positive separation from the picture signal, random noise will not interfere with synchronization unless its level is well above that which can be tolerated in the picture itself. However, sharp interfering pulses may cause the deflection generator to start prematurely if the synchronizing signals are applied to it directly. Much more rarely, a disturbance could oppose

the synchronizing pulse and delay the start of the control oscillator. Automatic frequency control (AFC), by averaging out the effect of noise pulses over a long period of time, largely eliminates their effect on synchronization. Without AFC, a disturbing pulse will only affect synchronization if its amplitude is comparable with the control signal itself, and then only if it occurs just before the end of a line or frame, at which time the control oscillator is easily started (see Fig. 14.38). The amplitude of the synchronizing impulses is made large enough so that, in general, interference which is not too obtrusive on the kine-

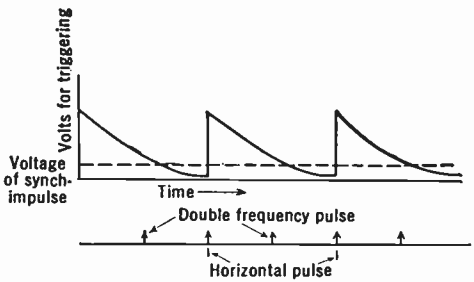
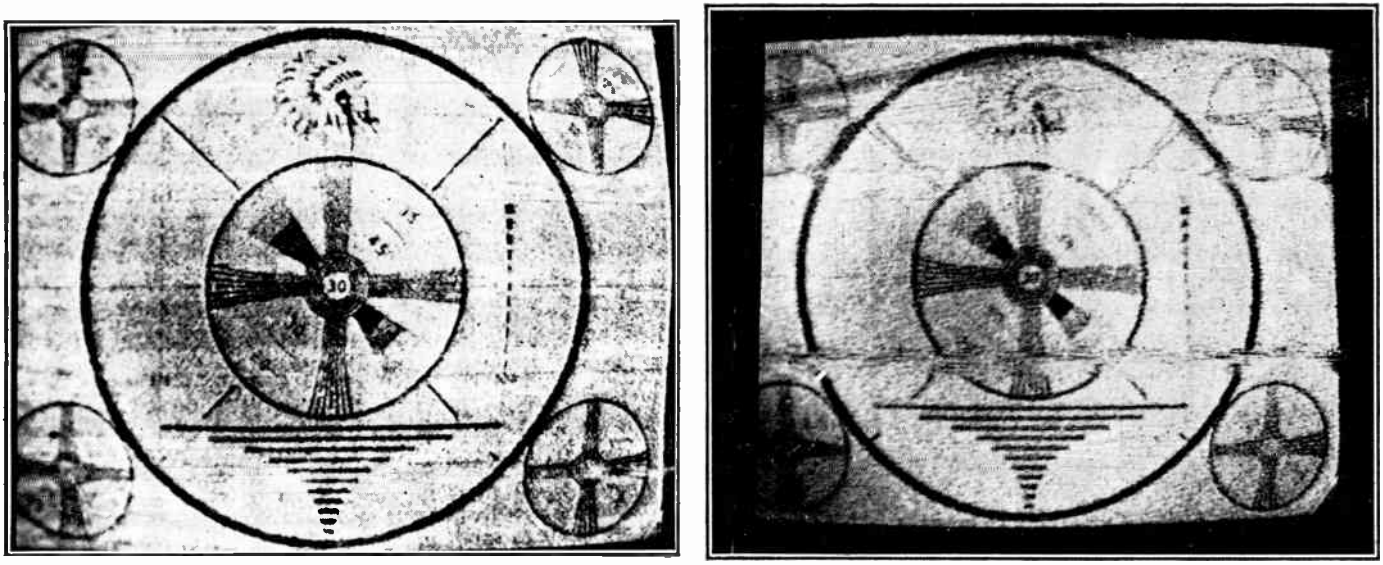


Fig. 14.38. Sensitivity of Control Oscillator as a Function of Time (in the Absence of AFC).

scope screen will not destroy synchronization. On this basis the amplitude of the synchronizing impulse is made 33 percent greater than the maximum picture signal. Experience has shown that, even so, AFC materially increases the acceptability of the picture under low-signal or high-noise level conditions. This is exemplified by the comparison in Fig. 14.39.

As far as immunity to interference and ease of separation are concerned, the polarity of the synchronizing impulse may be in the direction of either black or white. In other words, the signal may be "blacker than black" or "whiter than white." Since the former drives the grid of the viewing tube below cutoff during return time, some circuit simplification is effected by its use.

14.8 Separation of Video Signal and Synchronizing Impulses. The control pulses are selected from the complete video signal by a limiter tube. In a limiter the nonlinearity of the tube characteristics is utilized to remove unwanted portions of the signal. Thus, in Fig. 14.40, information lying above the level AB, constituting the synchronizing signals, is to be separated from the picture information carried by the lower portions of the curve.



(*a*) (*b*)

Fig. 14.39. Effect of Automatic Frequency Control on Picture in the Presence of High-Level Interference (Interference from Electric Razor). (*a*) Picture obtained without AFC. (*b*) Picture obtained with AFC. (Wendt and Fredendall, reference 15.) (Courtesy of *Proceedings of the Institute of Radio Engineers.*)

The operation of the many different limiter circuits employed for separating the synchronizing signals from the video signal differs but little. The diode circuit shown in Fig. 14.41 may be regarded as basic. The chosen polarity of the diode corresponds to positive synchronizing signal. During the positive synchronizing peaks the diode conducts, charging the condenser C through the small resistance R_c. Between synchronizing pulses C discharges through the large resistance R_g; the

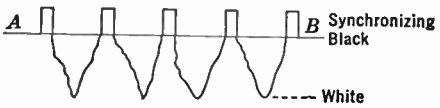


Fig. 14.40. Separation of Synchronizing Pulses from Picture Signals by Limiter or Clipper.

time constant CR_g is generally made about 6000 microseconds or approximately 100 line periods, so that the voltage change on the condenser is very slight. The clipping level below the synchronizing peaks is determined by the condition that the transfer of charge during the pulse time τ is just equal to that in the interval between pulses, $T - \tau$. Thus, if E_p represents the voltage of the synchronizing pulses, E_v, the

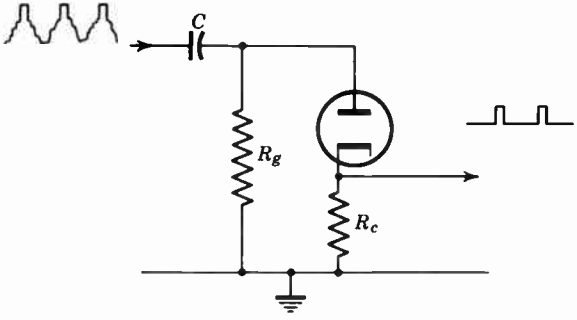


Fig. 14.41. The Diode Clipper.

average signal voltage between pulses, and E_c, the signal voltage for which the drop across the diode is zero (the clipping level),

$$\tau \frac{E_p - E_c}{R_c + R_t} = (T - \tau) \frac{E_c - E_v}{R_g} \tag{14.18}$$

Here R_t represents the diode resistance. This equation can be solved to yield

$$\frac{E_c - E_v}{E_p - E_v} = \frac{\tau R_g}{(T - \tau)(R_c + R_t) + \tau R_g} \tag{14.19}$$

Television standards demand that the ratio at the left be 0.75 for the least value of E_v, corresponding to an all-white picture. Thus this equation leads to a maximum value of the load resistance R_c; e.g., for $\tau/T = 0.07$,

$$\frac{R_c + R_t}{R_g} \leq 0.025$$

The value of $E_p - E_c$, reduced by the factor $R_c/(R_c + R_t)$, yields directly the amplitude of the synchronizing signal output.

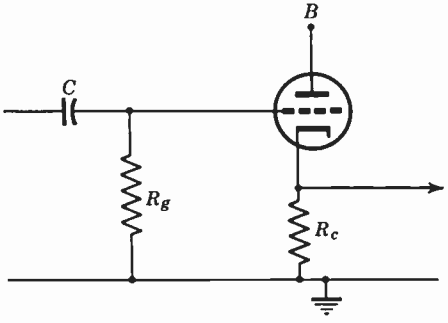


Fig. 14.42. The Cathode-Follower Clipper.

If the diode is replaced by a sharp cutoff triode, a cathode-follower clipper is obtained (Fig. 14.42). The amplitude of synchronization in the input signal must here be appreciably greater than the grid cutoff voltage, which is given approximately by the plate voltage E_b divided by the voltage amplification μ of the tube, since the clipping level is not given by E_c in Eq. 14.18, but by $E_c - E_b/\mu$. Correspondingly, for equal signal output, the resistance R_c must be made much smaller than for a diode.

In the more conventional circuit in Fig. 14.43*a* the triode amplifies as well as separates the synchronizing pulses, which, here, invert their polarity. Otherwise, the operation is similar to that of the two preceding circuits. As in them, interelectrode capacitance may cause video signals to contaminate the output to some extent. This difficulty may be avoided by replacing the triode by a pentode and deriving the synchronizing signals from the screen-plate circuit (Fig. 14.43*b*). A third modification of the "grid-leak bias clipper" (Fig. 14.43*c*) is suitable for signals in which the synchronization has negative polarity: The grid is grounded and the signal applied to the cathode.

The cathode-bias clipper shown in Fig. 14.44 is somewhat different in operation. The cathode resistor is made very large and shunted by a condenser so that the time constant of the cathode circuit is about 6000 microseconds. The bias developed across the condenser cuts off

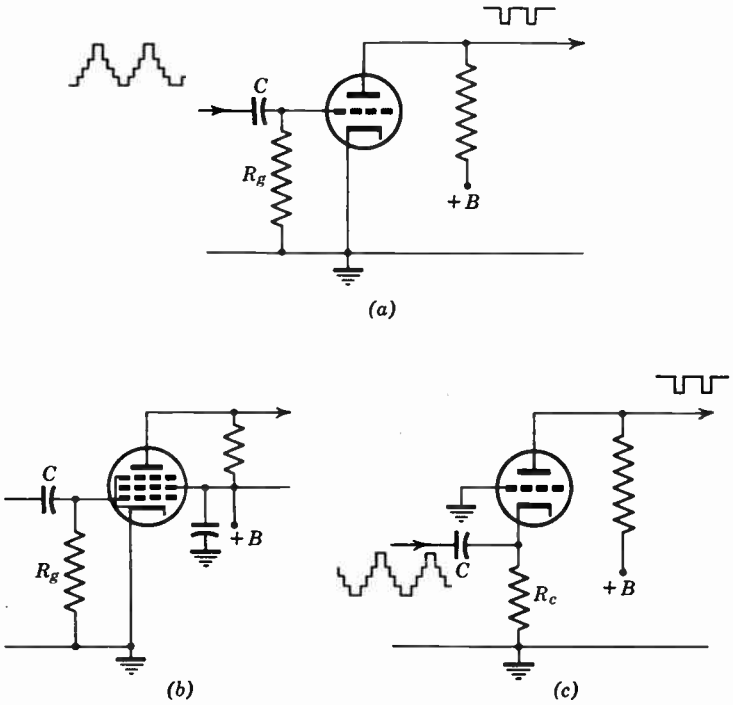


Fig. 14.43. Grid-Bias Clippers. (*a*) Triode circuit. (*b*) Pentode circuit. (*c*) Cathode-input circuit.

plate current except during the synchronizing pulses. An analysis such as that given for the diode clipper shows that the clipping level is given by

$$E_p - E_c = \frac{T - \tau}{\tau} \frac{E}{R_c g_m} \tag{14.20}$$

where E is the voltage across the condenser C. Since the quantity on the right must be less than the cutoff potential of the tube E_g, R_c must be chosen more than thirteen times as great as the reciprocal mutual conductance of the tube multiplied by the ratio E/E_g. The amplitude of the pulses in the output is given by

$$\frac{R_L r_p}{R_L + r_p} \frac{T - \tau}{\tau} \frac{E}{R_c} = \frac{R_L r_p}{R_L + r_p} g_m(E_p - E_c) \qquad (14.21)$$

For an a-c input signal the value of E will be large for a bright picture, small for a dark picture, so that, as with the diode clipper, the amplitude of the output will depend on the picture content. These synchronizing separators are thus useful, primarily, when the d-c level of the signal is set beforehand. With the grid-leak bias clipper the input signal must be larger but yields an output which is practically

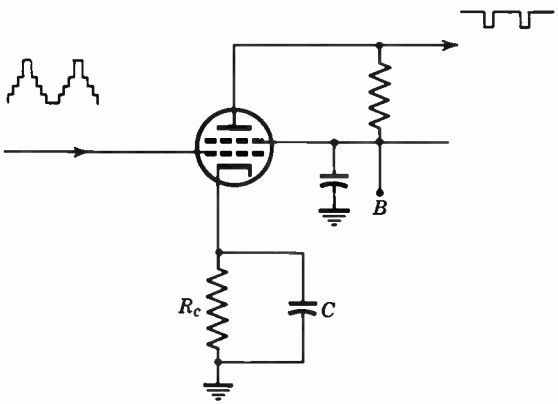


Fig. 14.44. Cathode-Bias Clipper.

independent of picture content. Thus, here, it is necessary that the synchronizing signal lie above the pedestal by more than the grid cutoff voltage E_g, whereas the output amplitude becomes, for sufficiently large R_g and r_p, $g_m R_L E_g$.

14.9 Selection of Vertical and Horizontal Synchronizing Impulses. After the vertical and horizontal impulses have been isolated from the complete signal, there remains the problem of separating the two impulses. This selection is effected by making the two impulses somewhat different in nature. They may differ in amplitude, shape, or duration. Thus the vertical pulse might be made much higher than the horizontal pulses—at the cost of greatly increased transmitter load during its transmission—and be separated from the horizontal pulses by a clipper circuit. As an alternative, horizontal and vertical pulses could be distinguished by having a steep leading or trailing edge, so that differentiation would lead to a strong positive and negative pulse, respectively. The method adopted in practice differs from either. It employs pulses of greatly differing duration. The vertical pulse, which

lasts three full horizontal lines, is serrated, so that differentiation leads to an unbroken succession of horizontal synchronizing pulses, whereas integration results in a single prominent vertical peak for every vertical pulse (Fig. 14.45), accompanied by smaller peaks produced by the individual horizontal pulses.

The complete synchronizing signals, as prescribed by the Federal Communications Commission, are shown once more in Fig. 14.46. They are complicated to some extent by the requirements of interlaced scanning. The serrations of the vertical pulse itself have twice hori-

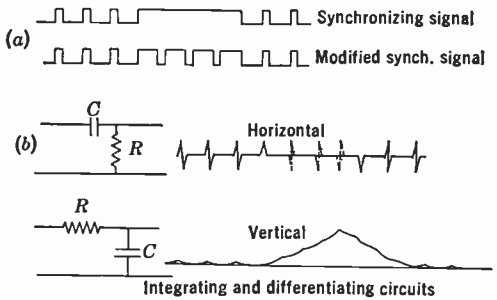
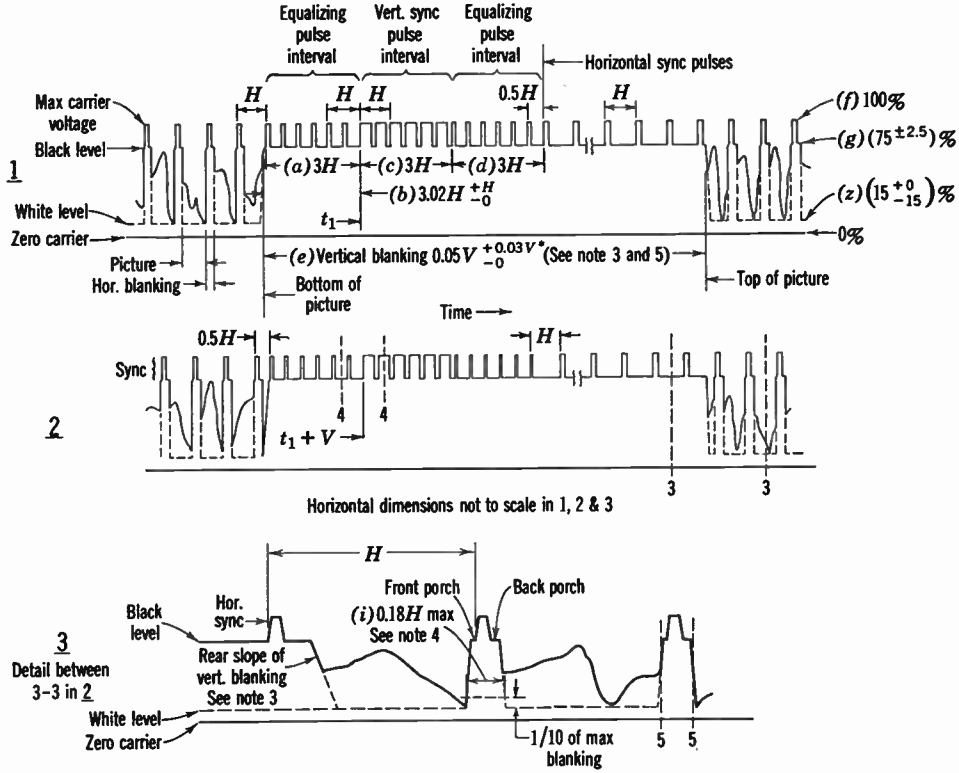
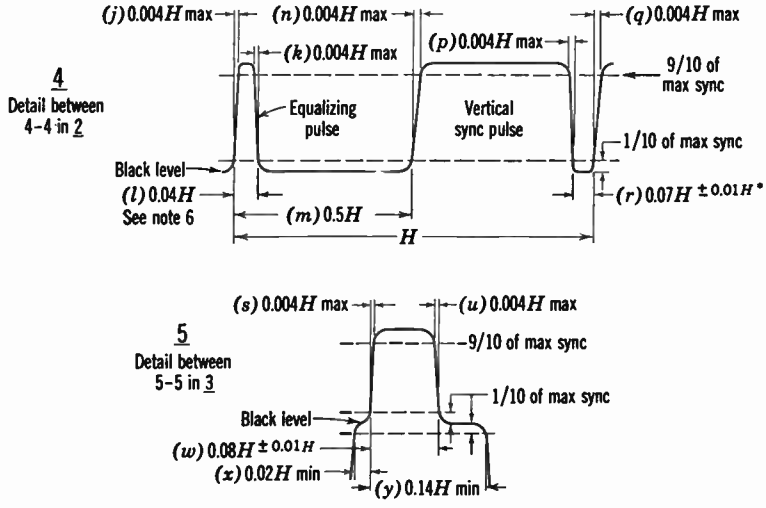


Fig. 14.45. The Vertical Synchronizing Pulse and Pulse Selection by Differentiating and Integrating Circuits (Schematic).

zontal frequency and, furthermore, the vertical pulse is preceded and followed by "equalizing pulses" with half the width and twice the frequency of the horizontal pulses for three line periods. The purpose of this modification is to render the shape of the integrated vertical pulse identical for successive fields, in which the relative position of the vertical and horizontal pulses is displaced by half a line period. The added pulses and serrations, inserted at the middle of the horizontal trace, do not trigger the horizontal oscillator (even without AFC) since they fall in an insensitive period of the oscillator. Equalizing pulses follow as well as precede the vertical pulse since the trailing part of the pulse has a certain effect on the rate of discharge of the vertical blocking oscillator condenser and, hence, on the length of the vertical period.

Figure 14.45 illustrates an integrating circuit consisting of a single condenser and resistance. In practice, an integrating network of several meshes is employed, as illustrated in Fig. 14.47. In this manner the peak amplitude produced by the individual horizontal pulses can be further attenuated relative to that corresponding to the vertical pulse; the shape of the integrated pulse, produced by the network in





Note:

1. H = Time from start of one line to start of next line.
2. V = Time from start of one field to start of next field.
3. Leading and trailing edges of vertical blanking should be complete in less than $0.1H$.
4. Leading and trailing slopes of horizontal blanking must be steep enough to preserve minimum and maximum values of $(x + y)$ and l under all conditions of picture content.

*5. Dimensions marked with an asterisk indicate that tolerances given are permitted only for long time variations, and not for successive cycles.

6. Equalizing pulse area shall be between 0.45 and 0.5 of the area of a horizontal sync pulse.
7. Refer to text for further explanations and tolerances.

Fig. 14.46 F.C.C. Standard Television Signal Wave Form. (Institute of Radio Engineers, reference 18.)

question, is also shown in Fig. 14.47. The design aims to let the most steeply ascending portion of the pulse coincide with the instant of triggering, since in this manner the greatest stability can be attained. Normally, the integrating circuit is preceded by a limiting amplifier which trims off noise pulses exceeding the synchronizing pulses in amplitude and thus improves the noise immunity of the horizontal deflection.

Several methods have been proposed for increasing the precision of vertical synchronization. One of these consists of adding the differ-

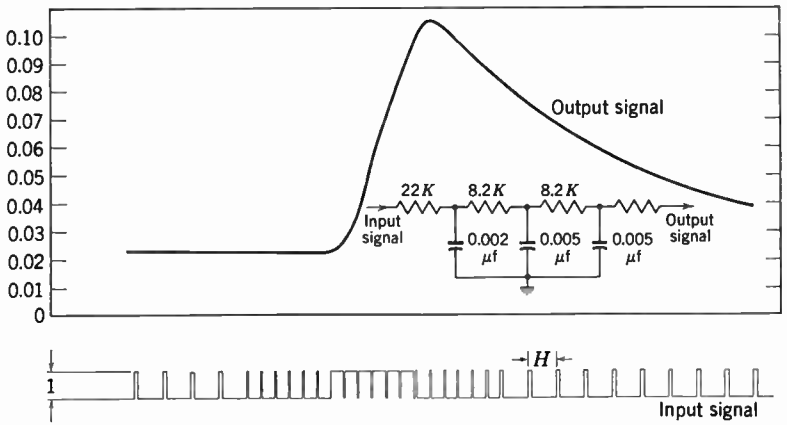


Fig. 14.47. Multimesh Integrating Network for Forming Vertical Synchronizing Pulse and Pulse Shape Obtained with Standard Signal Wave Form.

entiated horizontal signal to the integrated vertical signal and triggering the vertical signal by the pulse corresponding to one of the serrations.

14.10 Formation of the Complete Synchronizing Signal. The type of synchronizing signal having been selected, there still remains the problem of forming the impulses and adding them to the picture signal.

The impulses are formed by a chain of suitable pulse oscillators and shaping circuits. These impulses are added together and to the picture signal by means of keying or mixing tubes. There are many chains which are satisfactory, and the choice depends upon the particular installation. Only the broader aspects can be taken up here.

The highest pulse frequency is twice line frequency. Therefore, there is a primary oscillator delivering pulses at a rate of 31,500 per second. Its output is applied to a series of frequency division circuits which deliver pulses at line frequency $(31{,}500 \div 2)$ and field frequency $(31{,}500 \div 525 = 31{,}500 \div (7 \cdot 5 \cdot 5 \cdot 3))$ as well as a number of

intermediate frequencies. Pulse-forming, shaping, and keying circuits combine these into the complete synchronizing signal, as indicated schematically in Fig. 14.48.

The preferred frequency division circuit consists of a step counter in combination with a blocking oscillator with fixed cathode bias (Fig. 14.49) or a multivibrator. The step counter consists of a pair of diodes, V_1 and V_2, in combination with an input capacitance C_1 and a larger storage capacitance C_2. As a positive pulse is applied to the input of the counter, C_1 and C_2 are charged positively through V_1, in inverse ratio to their capacitance. At the end of the pulse, as the voltage decreases, C_1 discharges through V_2. The charge of the storage capacitance C_2, however, remains unaltered since the polarities of V_1 and V_2 block discharge. The next positive pulse increases the stored charge by a further increment which is smaller than the first since the charging voltage is reduced by the voltage drop already existing across C_2. The potential across C_2 thus increases stepwise until it reaches the firing voltage of the grid of the blocking oscillator with which it is connected. The number of steps required to bring the grid to its firing potential, or the frequency division factor, is determined by the fixed bias applied to the blocking oscillator cathode. A buffer amplifier transmits the positive pulse from the blocking oscillator grid, with polarity reversed, to the next frequency division stage.

The pulse forming and shaping system employs largely multivibrators, limiting amplifiers, addition networks, and keying stages. A keying stage (Fig. 14.50) normally utilizes a pentode, the signal to be keyed being applied to the control grid and the keying signal to the screen grid.

The master oscillator at twice line frequency may be either stabilized by a crystal maintained in a carefully temperature-controlled enclosure or be locked in with the alternations of the local 60-cycle power system. In the latter instance the 60-cycle pulses derived from the master oscillator by frequency division are compared in phase with the sinusoidal oscillation of the power line by a circuit similar to that employed in automatic frequency control (Fig. 14.36), and the output of the comparator is utilized to modify the frequency of the master oscillator.

14.11 Special Problems of Scanning. *Keystoning.* The problem of providing a scanning pattern which deviates materially from a rectangle arises in the iconoscope and may be encountered in other special-purpose tubes. In the iconoscope the mosaic makes an angle of 60

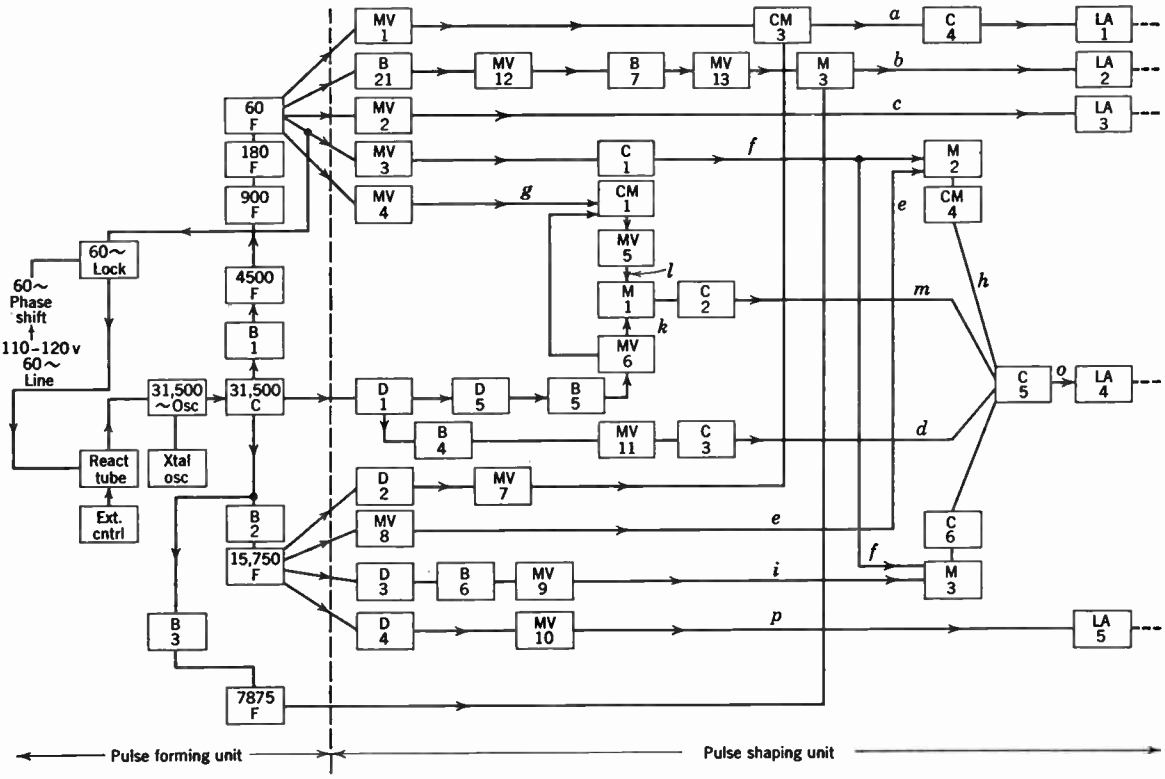


Fig. 14.48. Block Diagram Indicating the Formation of the Complete Synchronizing Signal.

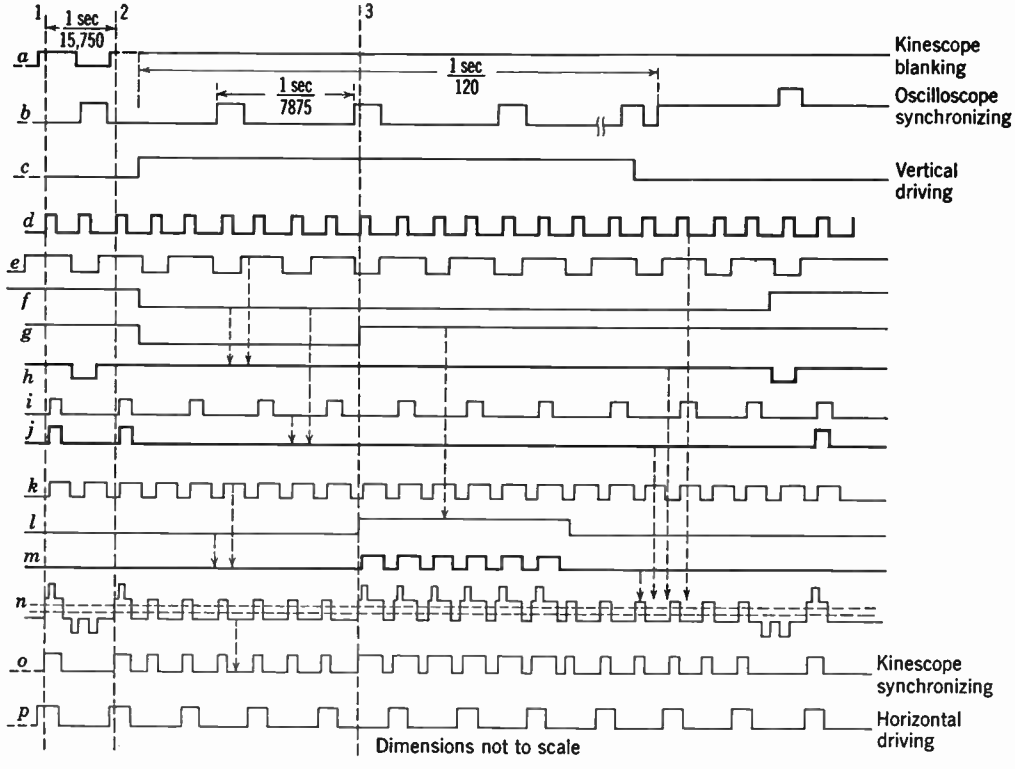


Fig. 14.48 (*continued*).

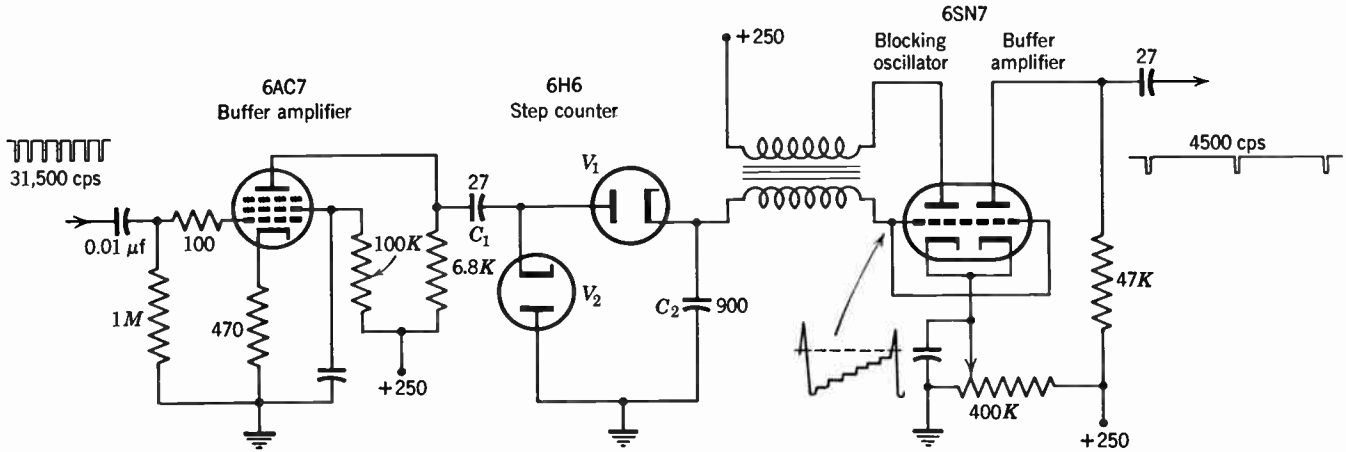


Fig. 14.49. Frequency Division Stage Consisting of Step Counter and One-Shot Blocking Oscillator.

degrees with the axis of the gun. Therefore, if the scanning pattern is formed by vertical and horizontal deflections having constant angular amplitude, the pattern will not be rectangular. Since the bottom of the mosaic is closer to the gun than the top, the top lines will be longer than the bottom lines, and the area swept out on the screen

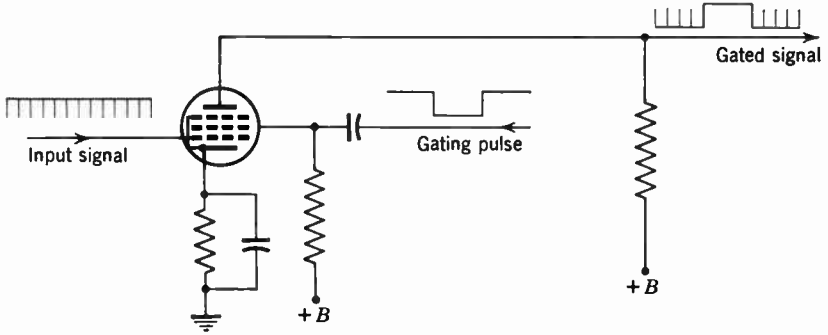


Fig. 14.50. Keying Circuit.

will be trapezoidal, or keystone, in shape. This will be clear from Fig. 14.51*a*.

To correct for this, it is necessary to modulate the horizontal deflection with the vertical sawtooth wave. In this way a decrease in angular amplitude at the top and an increase toward the bottom can

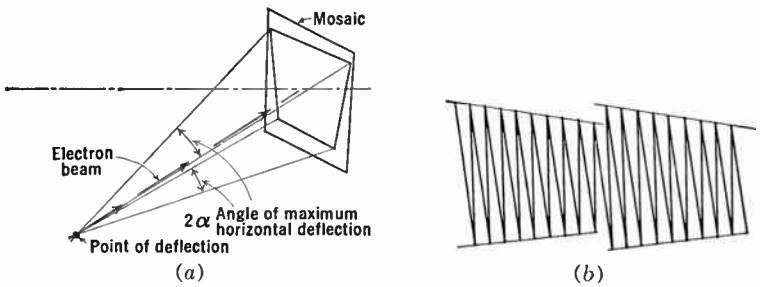


Fig. 14.51. Origin and Correction of Keystone Distortion.

be made to overcome the trapezoidal distortion in the pattern on the slanting screen. The appearance of the modulated sawtooth wave is shown in Fig. 14.51*b*.

This modulation can be carried out in a number of ways. Essentially, the gain of the deflection amplifier must be linearly increased

periodically at vertical frequency, and the vertical sawtooth wave component, present in addition to the modulated horizontal sawtooth, must be removed. Figure 14.52 shows a simple circuit which accomplishes this.

The grid bias of a variable-mu horizontal deflection amplifier tube is varied periodically by the vertical sawtooth. This produces a modulated horizontal sawtooth, plus a vertical component arising from the variation in bias. The second component is removed by applying an inverse vertical sawtooth of amplitude equal to that present in the output of the modulator. The condition for exact cancellation of the vertical sawtooth component is achieved by proper adjustment of the potentiometer contact on the vertical sawtooth input after the degree of modulation has been set with the aid of the cathode-bias adjustment.

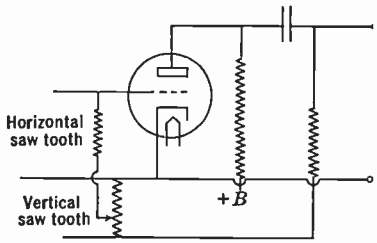


Fig. 14.52. Simple Keystone Correcting Circuit.

Interference and Scanning. The three main types of interference to be guarded against when designing a scanning system are: (1) effect of noise and static upon the synchronizing mechanism; (2) cross-talk between vertical and horizontal deflection; and (3) a-c ripple or hum from the power supply feeding through to the deflection. There are, of course, many other less important causes of interference.

The effect of interference on synchronizing has already been pointed out. With automatic frequency control of the horizontal deflection and appropriate clipping of the synchronizing signals ahead of the vertical integration circuits it is not serious unless of such magnitude as to spoil the picture itself.

Cross-talk is the term applied to the influence which may be exerted by the vertical deflection currents on the horizontal deflection, and vice versa. It is the result of inadequate electrical isolation between the two deflecting systems, or a poorly designed deflecting yoke. The former may be the result of insufficient shielding of the deflection amplifiers or leads or of coupling through a common power supply.

If the vertical sawtooth affects the horizontal deflection, it causes a distortion of the pattern shape. In case of a simple addition of the sawtooth, the pattern will be sheared; in case of modulation, it will be keystoned. Usually only some of the harmonics of the vertical

deflection wave interfere, resulting in wavy sides of the scanned area. Harmonics of the horizontal sawtooth appearing in the vertical deflection, on the other hand, cause the scanned lines to be wavy instead of straight. This defect is particularly serious in camera tubes such as the iconoscope since here, in addition to pattern distortion, they may give rise to appreciable spurious signals.

The effect of hum on the received picture is most serious when the hum frequency is not an integer multiple of the vertical deflection frequency. For example, Fig. 14.53 indicates the effect of 60-cycle hum

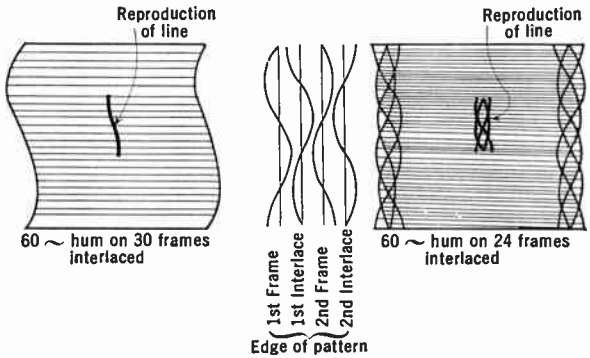


Fig. 14.53. Effect of Hum on Scanning Patterns.

influencing the horizontal deflection on a 30-cycle-per-second and a 24-cycle-per-second interlaced scanning pattern, respectively. Whereas in the first hum modulates the outlines of objects in the picture without blurring them, its effect on successive fields differs in the second so that the picture obtained by superposition is unsharp. The same applies to the effect of hum on the vertical deflection, where it causes a variation in line separation. A slight deviation between the hum frequency and the vertical-deflection frequency causes an apparent swaying of the picture, which proves very objectionable. This swaying may be prevented by locking the master oscillator at the transmitter with the local power system, provided that the power systems at the transmitter and the receiver are interlocked.

Wherever the power systems in the service area of a transmitter are not synchronized, it is necessary to rely on careful filtering of the power supply, shielding, and isolation of the heaters in vacuum tubes with indirectly heated cathodes. It has been found that, with proper precautions of this type, receivers can be made to function to com-

plete satisfaction with a frame frequency of 30 cycles per second in areas serviced by 50-cycle power systems.

REFERENCES

1. I. G. Maloff and D. W. Epstein, *Electron Optics in Television,* McGraw-Hill, New York, 1938.
2. S. Deutsch, *Theory and Design of Television Receivers,* McGraw-Hill, New York, 1951.
3. W. T. Cocking, *Television Receiving Equipment,* 3rd. Ed., Iliffe & Sons, London, 1950.
4. O. S. Puckle, *Time Bases,* 2nd Ed., Chapman & Hall, London, 1951.
5. A. W. Friend, "Television Deflection Circuits," *RCA Rev.,* Vol. 8, pp. 98–138, 1947.
6. M. J. Obert and W. A. Needs, "Ferrite Core for Wide Deflection-Angle Kinescopes," *Tele-Tech,* Vol. 9, pp. 42–44, October, 1950.
7. B. B. Bycer, "Design Consideration for Scanning Yokes," *Tele-Tech,* Vol. 9, pp. 32–35, August, 1950.
8. C. V. Bocciarelli, "Improved Deflection and Focus," *Electronics,* Vol. 23, pp. 94–95, August, 1950.
9. O. H. Schade, "Magnetic Deflection Circuit for Cathode Ray Tubes," *RCA Rev.,* Vol. 8, pp. 506–538, 1947.
10. O. H. Schade, "Characteristics of High-Efficiency Deflection and High-Voltage Systems for Kinescopes," *RCA Rev.,* Vol. 11, pp. 5–37, 1950.
11. A. W. Keen, "Television Time Base Linearization," *Proc. I.R.E.,* Vol. 37, p. 61, 1949.
12. O. S. Puckle, "Time Bases," *J. Inst. Elec. Engrs.,* Vol. 89, pt. 3, pp. 100–119, 1942.
13. J. L. Potter, "Sweep Circuit," *Proc. I.R.E.,* Vol. 26, pp. 713–719, 1938.
14. J. Haantjes and F. Kerkhof, "Projection Television Receiver: IV. The Circuit for Deflecting and Electron Beam. V. The Synchronization," *Philips Tech. Rev.,* Vol. 10, pp. 307–317 and 364–370, 1949.
15. K. R. Wendt and G. L. Fredendall, "Automatic Frequency and Phase Control of Synchronization in Television Receivers," *Proc. I.R.E.,* Vol. 31, pp. 7–15, 1943.
16. E. L. Clark, "Automatic Frequency Phase Control of Television Sweep Circuits," *Proc. I.R.E.,* Vol. 37, pp. 497–500, 1949.
17. K. Schlesinger, "Locked Oscillator for Television Applications," *Electronics,* Vol. 22, pp. 112–117, January, 1949.
18. Institute of Radio Engineers, *Standards of Television—Methods of Testing Television Receivers,* New York, 1948.
19. W. Heiser, "Sync Separator Analysis," *Electronics,* Vol. 23, pp. 108–111, July, 1950.
20. A. W. Keen, "Frame Synchronizing Signal Separators," *Electronic Eng.,* Vol. 21, pp. 3–9, January, 1949.
21. K. Schlesinger, "Internal Electrostatic Yokes," *Electronics,* Vol. 25, pp. 105–109, July, 1952.

15 The Television Transmitter

The difference between the type of transmitter required to radiate high-definition television signals and that used for sound broadcasting is so great as to be almost fundamental. The reason for this difference is the tremendous frequency band needed for picture transmission. Not only does the wide frequency band present a technical obstacle in itself, but it also makes necessary the use of a carrier in the very-high or ultra-high-frequency range, as was pointed out in Chapter 6. The Federal Communications Commission has assigned five 6-megacycle bands in the range from 54 to 88 megacycles and seven additional bands between 174 and 216 megacycles for television broadcasting. In Europe and elsewhere, similar bands have been given over to video transmission. Additional allocations are being made in the ultra-high-frequency band, from 470 to 890 megacycles. This chapter deals primarily with transmitters for the very-high-frequency range. The special problems of ultra-high-frequency transmitters are treated in brief.

Like the transmitter used for sound broadcasting, the television transmitter consists of a carrier generator, a modulator, power amplifiers, a transmission line carrying the modulated carrier to the antenna, and finally the antenna itself. In principle, these elements are similar to those used in sound transmission; however, because of the very high frequencies involved, problems which in sound transmission are only of secondary importance assume the proportions of major considerations. At the frequencies used for picture transmission, the series impedance due to the inductance of every lead and the shunt capacity between every electrode, or lead, and ground, must be taken into account. In fact, every lead must be considered as a transmission line whenever its length is comparable with that of the wave traveling over it.

15.1 Character of the Transmitted Signal. The modulation of the carrier by the video signal to form the r-f signal has been discussed in Chapter 6. The modulating video signal may be of either polarity,

giving rise to either positive or negative modulation. With positive modulation the amplitude of the radio frequency is a maximum for white portions of the picture whereas with negative modulation the converse is true, that is, the black regions of the picture increase the r-f amplitude.

As has already been pointed out, practical considerations dictate the use of synchronizing impulses in the direction of a black signal. Therefore, if positive modulation is used, the synchronizing signal causes the r-f signal to drop to a very low value (usually zero amplitude), whereas negative transmission results in a maximum signal for the synchronizing impulses.

Certain advantages may be urged for both positive and negative transmission. On careful examination, negative transmission, adopted as part of American television standards, appears preferable, however. If the effect of interference is neglected, negative transmission is advantageous on two scores. On the one hand, the synchronizing peaks provide a ready reference for automatic gain control; on the other, the average transmitter power requirements are reduced, since the carrier signal attains at most 75 percent of the maximum amplitude in the course of picture transmission.

With interference the choice is not quite so clear-cut. It is convenient to consider the effects of interference on synchronization and its direct effect on the picture separately. If interference is sufficiently serious to destroy synchronization completely, it will probably render the picture valueless even if synchronization is artificially maintained. This statement applies to either system. However, with directly triggered circuits, interference, even if not sufficiently severe to destroy the picture, may occasionally cause the impulse generator of the deflecting circuits to operate too soon. This occurs when a disturbing pulse of the right polarity arrives just ahead of the normal synchronizing impulse, at which time the deflection generator is extremely sensitive to any incoming signal. Negative picture transmission is somewhat more sensitive to this type of interference than positive modulation because to interrupt synchronization in positive transmission the interfering signal must be of such amplitude and phase that it just cancels the incoming signal. If the phase of the interference is not exactly opposite to that of the carrier, it will not cause the signal to become zero, and therefore will not interrupt the deflection. Likewise, if the interference is in phase opposition but its amplitude is greater than that of the video signal, it will result in overmodulation rather than loss of signal. Negative transmission, on the other hand, is sus-

ceptible to any impulse occurring near the end of a line or frame which causes the signal to rise to a value equal to, or greater than, that required for synchronization. In practice, this advantage of positive modulation is only apparent. With automatic frequency control in the horizontal deflection circuits and the insertion of an amplitude-limiting stage ahead of the integrating networks for the vertical synchronizing pulse, the synchronization of the receiver can always be rendered adequately immune to interference.

In the picture itself, interference becomes much less objectionable with negative modulation than with positive modulation. Momentary signal increases appear as dark spots with negative modulation, as white flashes with positive modulation. Bright flashes on the screen are much more objectionable than dark spots of similar size. The fact that the white spots may be increased in size by "blooming" makes them even more prominent.

Negative picture transmission has been adopted in the United States and in most other countries of the world where television standards have been fixed; the notable exceptions are Great Britain, France, and Morocco-Tunisia, where positive transmission is employed.

The signal, whether positively or negatively modulated, can be radiated from the transmitter in the form of either a-c or d-c transmission. A-c transmission, as the name implies, means that the video signal supplied to the modulator has been amplified by an amplifier passing only its a-c components. Because of this, the signal at the modulator represents the voltage variation about an axis or reference voltage which is the time average of the signal. Thus, if the picture consists of a narrow white line on a black background, the carrier envelope for negative modulation will vary from approximately half value for the background to zero carrier at the white band, whereas a black line on a white background leads to a carrier which has approximately half value for the white background and maximum value at the black line. It will be apparent that the instantaneous magnitude of the r-f signal has, therefore, no real meaning in terms of picture brightness. However, the complete signal carries the information which determines the correct light level of the picture, for the pedestal which accompanies the synchronizing signal always corresponds to black. Relatively simple circuits at the receiver can, by making use of the pedestal, establish the correct viewing-tube bias, as will be explained in the next chapter.

The video signal, on the other hand, may be amplified up to modulation level by means of a d-c amplifier or its equivalent. If this is

done, the magnitude of the r-f signal has absolute significance in terms of picture brightness. For negative modulation, a black area, irrespective of size, results in maximum carrier amplitude (exclusive of the synchronizing signal), whereas a white area leads to zero or minimum carrier. The carrier envelopes for a bright line on a black background and a dark line on a bright background are shown for a-c and d-c transmission in diagrams *a* and *b*, respectively, of Fig. 15.1.

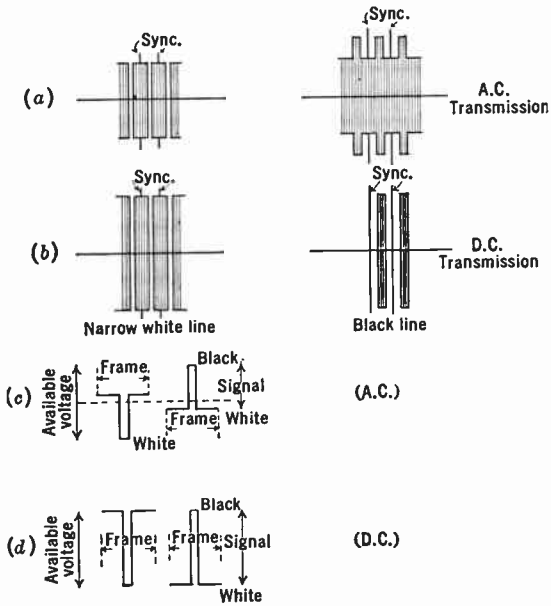


Fig. 15.1. A-C and D-C Transmission of Video Signals.

D-c transmission is generally favored since considerable economy in the use of any given transmitting equipment can be effected with it. This saving will be made clear by reference to Figs. 15.1*c* and *d*. A given transmitter has available a certain maximum voltage range for the carrier envelope. This range, for convenience, is referred to the voltage applied to the modulator and is shown in Fig. 15.1*c* by the double-headed arrow whose ends represent maximum and minimum carrier. The reference axis about which the video signal voltage swings in a-c transmission must be at or near the midpoint of the modulator voltage range in order to transmit the two types of picture signal shown in the diagram. It will be seen that the voltage change from black to white is only about one-half the total available voltage

swing. On the other hand, if a d-c video signal is applied to the same modulator, a "black" signal will correspond to a maximum positive voltage, whereas white corresponds to a maximum negative voltage; hence the full voltage swing of the modulator is available for the signal voltage. In other words, the useful voltage range with a-c transmission is only half as great as it is with d-c transmission. Actually, the power obtainable by d-c transmission is not four times as great as by a-c transmission, which might be expected from the above discussion. In the first place, the synchronizing signal and pedestal which occupy nearly 20 percent of the transmission time reduce the power saving (*Note:* For d-c transmission, synchronizing must always produce maximum carrier, whereas for certain types of pictures with a-c transmission it may be only slightly more than half maximum carrier.) Furthermore, the transmitter is limited not only by maximum voltage range but also by average power dissipated. The actual gain in power by the use of d-c transmission is between 1.5 and 2.5 times.

A carrier modulated in the ordinary way can be resolved into the carrier and two symmetrically placed sidebands. The spectrum will then appear as shown in Fig. 15.2*a*, which corresponds to a carrier modulated with a video signal containing all frequencies up to 2.875 megacycles. The position of the audio band relative to the video signal is also indicated in this figure. This band is not related to the video carrier, but has its own carrier, placed at a fixed distance from the video carrier.

In Chapter 6, it was shown that a single sideband, together with the carrier, contains *all* the information necessary to synthesize the picture, and that considerable advantage is obtained if only a single sideband is used. This being so, obviously it would be highly desirable if only a carrier and a single sideband, as illustrated in Fig. 15.2*b*, were generated by the transmitter. In this manner, the space in the frequency band required to transmit a given picture can be greatly reduced.

Unfortunately, at the present state of the art it is not possible to make an efficient modulator and power amplifier which will supply only the wanted carrier and single sideband to the antenna. As will become more apparent as the discussion proceeds, this is due to the fact that the tuned circuits which must be used have phase characteristics which invariably lead to inefficient amplification unless the carrier coincides with their resonant frequency.

A partial improvement can be effected, however, by filtering out one sideband from the signal by means of filters between the transmitter and the antenna. The transmitter in this case generates both sidebands, and the power contained in one sideband is dissipated in the filter. Whereas this implies a certain waste of transmitter power, the economy in utilization of the available frequency band is retained.

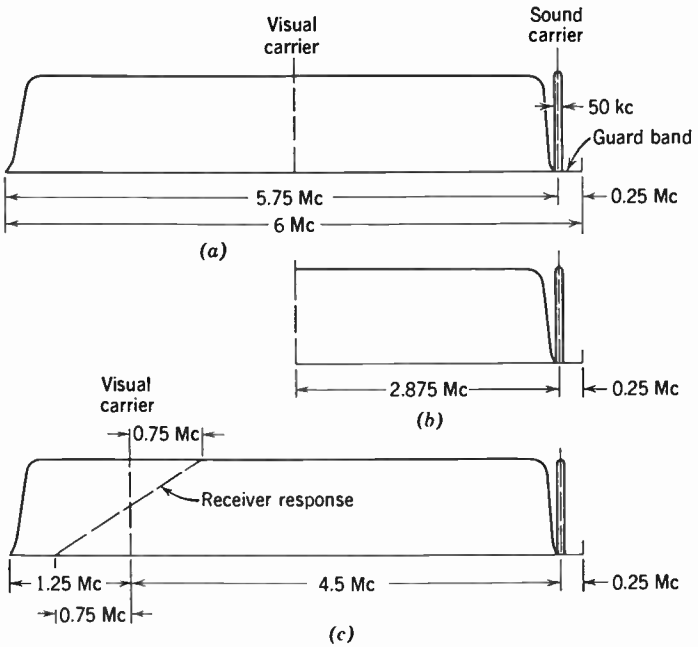


Fig. 15.2. Frequency Distribution for Transmitters Radiating Single- and Double-Sideband Signals.

Even this compromise cannot be carried out in full because of filter limitations. If a filter could be designed which had a suitable cutoff almost at the carrier, and no phase distortion over the passband, so that the resulting signal was of the form shown in Fig. 15.2*b*, faithful picture reproduction could be obtained, provided that similar requirements could be met by the response of the receiver. However, practical filter considerations have made it advisable to transmit approximately 1¼ megacycles of the unwanted sideband. The final shaping of the passband takes place in the receiver. Though the saving is not as great as if this expedient were not necessary, nevertheless, there is a gain of nearly 50 percent in frequency utilization.

If modulation of the carrier takes place at a lower level, so that several stages of amplification are needed to raise the modulated carrier power to the value required for transmission, the desired attenuation of one sideband may be achieved by off-center tuning of these amplifier stages. Even then the single-sideband filter is found necessary in practice to confine the radiated signal to the allocated frequency band. The relative advantages and disadvantages of high-level and low-level modulation are taken up in the next section.

To summarize: The signal which is considered most favorable at present is one which is negatively modulated, with "blacker-than-black" synchronization, and contains the d-c component. One sideband should be partially suppressed so as to permit a wider video band to be transmitted within a given radio channel. Finally, the carrier frequency lies preferably in the very-high-frequency range, namely, between 50 and 300 megacycles. The ultra-high-frequency channels, between 500 and 900 megacycles, are also suitable for television transmissions but present certain additional problems.

15.2 Very-High-Frequency Video Transmitters—General Plan. Practical video transmitters differ primarily with respect to the level at which the carrier is modulated by the video signal. Figure 15.3 shows block diagrams of three transmitters with 5-kilowatt output, in which modulation takes place at high level, low level, and intermediate level, respectively. All these transmitters employ crystal oscillators as primary frequency sources, succeeded by a chain of frequency multipliers; the last frequency tripler in this chain is employed only if the carrier frequency lies in the 174 to 216 megacycle range. In the high-level and medium-level transmitters, this is followed by one or two sharply tuned amplifier stages, which, in the first instance, raise the carrier power to about 400 watts. The output of these amplifier stages or, in the low-level modulated transmitter, that of the final frequency multiplier is applied to the modulated stage.

In addition, the transmitters contain a video amplifier chain which amplifies the video signal from about 2 volts to 625, 150, or 25 volts peak-to-peak across a 500-ohm load, respectively. The video amplifiers vary in design, both with respect to d-c re-insertion and interstage coupling. Grid modulation is employed in all three transmitters, producing modulated carrier outputs ranging from 5 kilowatts to 1 watt.

The three transmitters differ most in the section beyond the modulated stage. In the high-level modulated transmitter, the output of the modulated stage is adequate for transmission. It is, hence, ap-

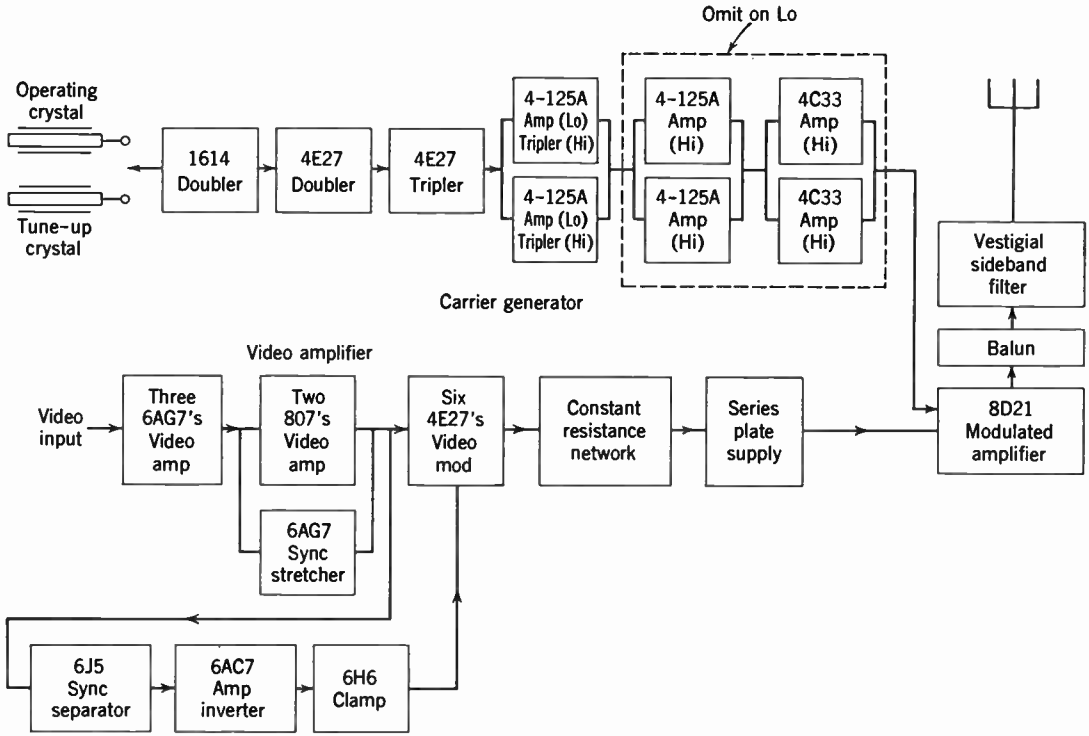


Fig. 15.3*a*. Block Diagrams of High-Level, Low-Level, and Intermediate-Level Modulated 5-Kw Television Transmitters. RCA TT-5A high-level modulated transmitter.

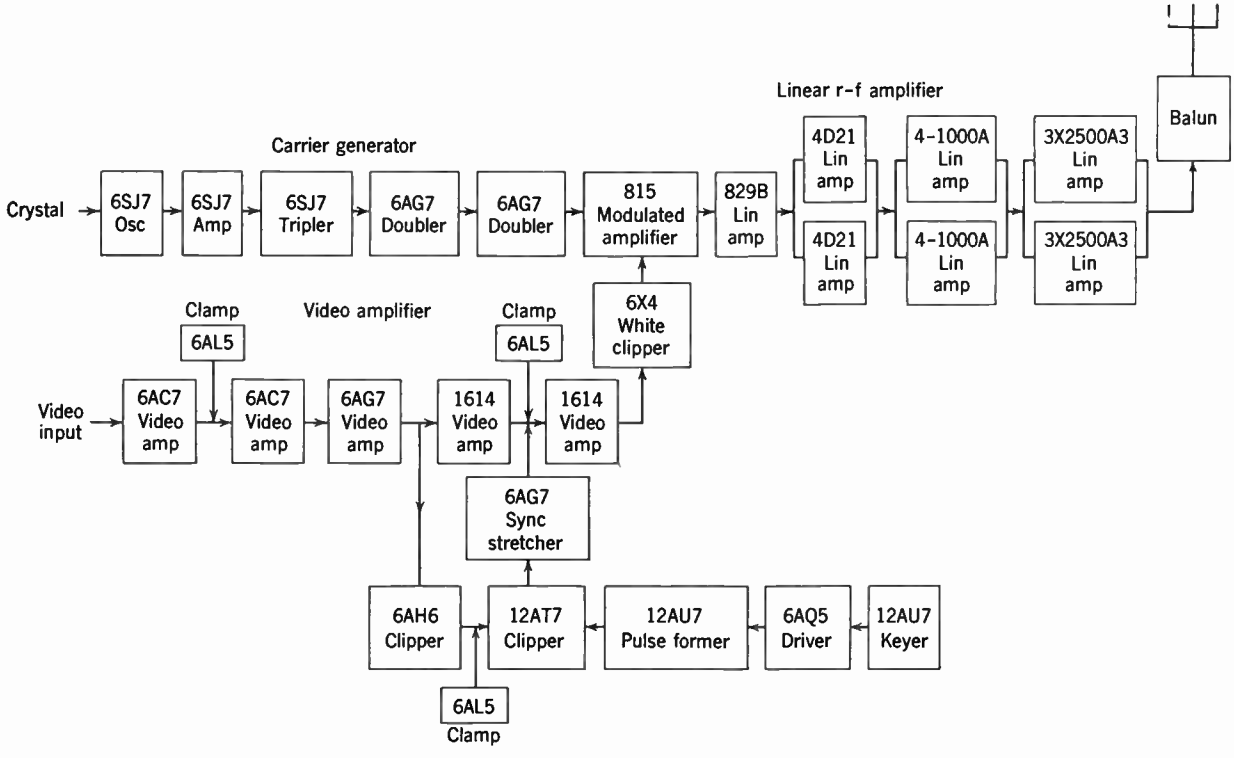


Fig. 15.3*b*. Block Diagrams of High-Level, Low-Level, and Intermediate-Level Modulated 5-Kw Television Transmitters. GE TT-10A low-level modulated transmitter.

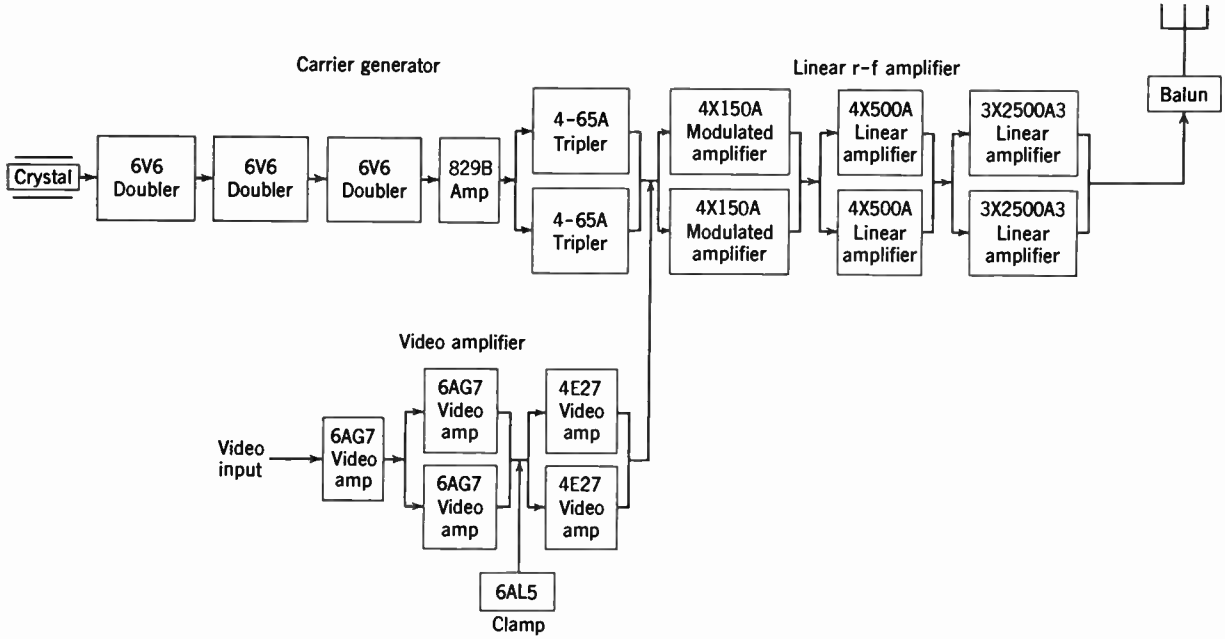


Fig. 15.3c. Block Diagrams of High-Level, Low-Level, and Intermediate-Level Modulated 5-Kw Television Transmitters. DuMont, Type 8000, intermediate-level modulated transmitter.

plied to a "balun," i.e., a circuit which converts a balanced output into a single-ended signal, and subsequently to a vestigial-sideband filter which removes the undesired portion of the modulated r-f signal. The resultant signal complies with the rules laid down for the transmitted signal by the Federal Communications Commission and the R.T.M.A.

In the low-level and medium-level modulated transmitters, a number of linear class B amplifying stages follow the modulated stage so as to bring the modulated carrier up to the power level required for transmission. These stages are tuned approximately 1.6 megacycles above the carrier frequency and are coupled so as to provide a signal cutoff on either side of the carrier which is similar to that shown in Fig. 15.2*c*. A vestigial-sideband filter (not shown) is required in practice to prevent interference with adjoining channels. It may be noted that the cost of the additional class B stages appears more than offset by the simpler video amplifier made possible with low-level modulation.

The primary advantage of high-level modulation rests in the greater stability of the transmitter characteristics and the greater ease of transmitter adjustment. Here, the picture quality is affected only by the video amplifier and the modulated stage. Since the video amplifier characteristics are highly stable, the possibility of spontaneous changes in the transmission characteristics is minimized.

Tuning of the high-level modulated transmitter can be accomplished simply with the aid of meters. A separate tuning crystal, yielding a carrier 1.6 megacycles above the main carrier, is provided for this purpose. With this crystal connected, the stages of the carrier generator chain, including the modulated stage, are tuned successively for maximum output. Then the tuning crystal is replaced by the operating crystal, and the carrier generator chain is retuned, leaving the modulated stage unaltered. Thus, the modulated stage introduces a certain amount of single-sideband attenuation, reducing the power dissipation required in the vestigial-sideband filter. Checking of the modulated stage with a sweep generator and oscilloscope is recommended when the flatness requirements for the video band are high. This applies particularly to color transmissions.

With low-level modulation, the tuning of the class B amplifier stages beyond the modulated stage requires the use of oscilloscopes for checking the frequency characteristic of each stage. These are generally built into the transmitter, along with a sweep-frequency generator. The latter is coupled to the horizontal deflection of the oscilloscope and provides the testing signal for the individual stages. Since the transmission band is a small fraction of the carrier fre-

quency, drifts in these stages can seriously affect the transmitted signal. Furthermore, the class B stages must be accurately linear to prevent re-insertion of the attenuated sideband.

At the present time, even though high-level modulation predominates, all three types of transmitters are employed by television broadcasters.

15.3 The Carrier Generator. As has already been pointed out, the principal elements making up the carrier generator are a master oscillator, a chain of frequency multipliers, and, eventually, one or two stages of power amplification.

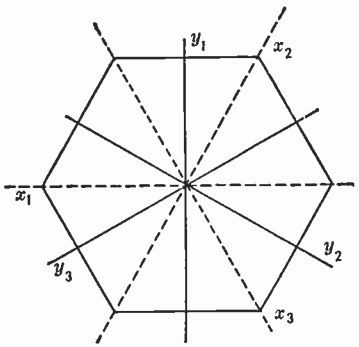


Fig. 15.4. Orientation of Electrical and Mechanical Axes in Quartz Crystal.

The piezoelectric crystal oscillator is universally employed as master oscillator in television transmitters. In principle, a high-Q tank circuit oscillator can also be used for this purpose.

Crystal oscillators are common in broadcast practice. In essence, a crystal oscillator is a sharply resonant mechanical vibrator with coupling between an electric circuit and the mechanical motion. The piezoelectric effect present in certain crystals provides this coupling. Quartz crystals have proved most satisfactory for this purpose.

A natural quartz crystal exhibits threefold symmetry about an axis known as the optic axis. This optic axis is electrically inactive. However, along the axes lying in a plane normal to the optic axis, quartz manifests a marked piezoelectric effect. Figure 15.4 shows a crystal projected in a plane perpendicular to the optic or z-axis. In this figure, the three axes Y_1, Y_2, Y_3, normal to the three main faces, are known as the mechanical axes, whereas the axes X_1, X_2, X_3, at right angles to the mechanical axes, are the electric axes. If the crystal is strained along one of the three mechanical axes, charges will be induced on the surfaces of the crystal normal to the electric axis, which is at right angles to the direction of strain. Similarly, an electric field applied parallel to an electric axis will produce a deformation along the mechanical axis perpendicular to the field.

If a rectangular segment is cut out of a quartz crystal in such a way that its face includes the Y and Z axes, as shown in Fig. 15.5a, and an alternating field is applied normal to the face, mechanical

vibration will occur in the Y direction. As the frequency is varied, a rate of vibration can be found which corresponds to the mechanical resonance frequency of the Y direction of the rectangle. At this frequency, a field of small amplitude can sustain a very considerable mechanical vibration. Furthermore, if the electrical oscillations are stopped, the crystal will continue to vibrate at this frequency for a large number of cycles owing to its elastic properties, inducing alternating charges on its two faces. In other words, the electrical be-

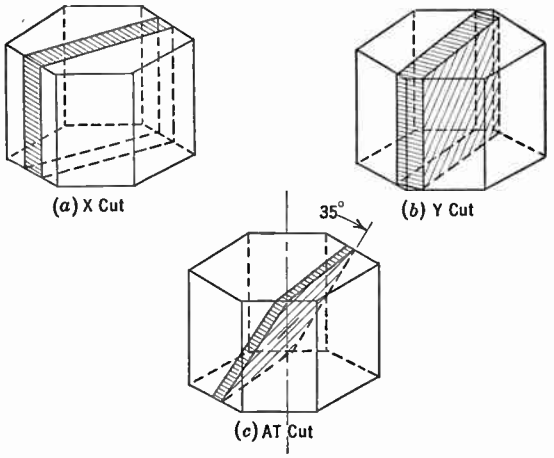


Fig. 15.5. Cuts Yielding Oscillator Crystals.

havior of the crystal plate is exactly the same as that of a very slightly damped (high-Q) parallel resonant circuit.

Under the resonant conditions just described, longitudinal mechanical standing waves whose motion is in the Y direction are set up. Similar standing waves can be established normal to the YZ face which will correspond to a much higher frequency.

Useful crystals can also be cut to include other crystallographic directions. A slab cut so that its face includes the X-axis and makes an angle of about 35 degrees to the optic axis results in a crystal whose resonant frequency is practically independent of temperature. In addition, a crystal so cut will be almost entirely free from parasitic oscillations which result from coupling between the various crystallographic directions. Such a cut is known as an AT cut. Other orientations of the crystal plate are also employed in practice to obtain high stability. The V-cut crystal, for example, is another temperature-independent crystal which finds frequent application.

The low damping of these crystal resonators makes them valuable as a means of controlling the frequency of vacuum-tube oscillators. Such an oscillator, together with its electrical equivalent circuit, is shown in Fig. 15.6. It will be seen that this oscillator is similar to the conventional tuned-plate, tuned-grid circuit. Various modifications of this circuit have been employed to increase the effective impedance of the crystal unit at the operating frequency and to enhance the sharpness of resonance.

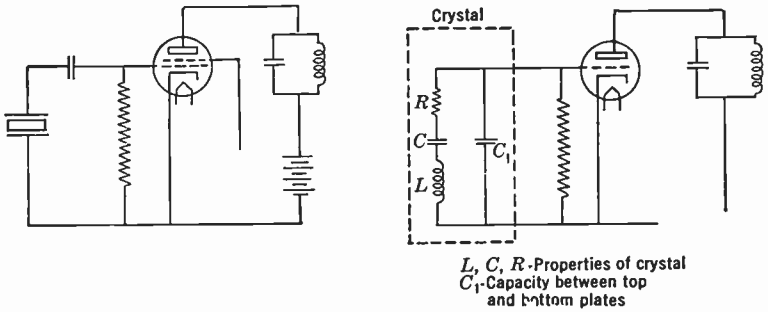


Fig. 15.6. Crystal Oscillator (Schematic).

The quartz resonator itself has also undergone considerable development since its introduction as frequency standard by W. G. Cady in 1921. Whereas, in the early days of radio communication, a frequency stability of one part in a thousand appeared quite adequate, modern television practice has raised these requirements to approximately one part in a million. Thus, with offset-carrier operation (see section 15.11) the video carriers of neighboring television stations operating on the same channel should differ by ± 10.5 kilocycles with a tolerance of the order of 1000 cycles in order to minimize co-channel interference. Since the carrier frequency approaches 1000 megacycles in the ultra-high-frequency band, the tolerance $\Delta f/f$ for the carrier frequency, as well as the crystal frequency from which it is derived, becomes 0.000001.

Similarly, stringent frequency stability requirements are imposed on the master oscillators of the picture and sound channels in compatible color television systems by the requirement that interference between the sound and the color subcarrier should have minimum visibility (see section 18.9). The maximum tolerance for the picture-sound carrier separation permitted by intercarrier sound reception for monochrome transmissions is greater by less than an order of magni-

tude (see section 16.13). On the other hand, since the required accuracy applies here to a frequency difference between two emissions originating in the same station, it is automatically satisfied if the two carriers are derived from the same crystal oscillator and becomes trivial if, with two crystals, one serves in essence to supply the difference frequency.*

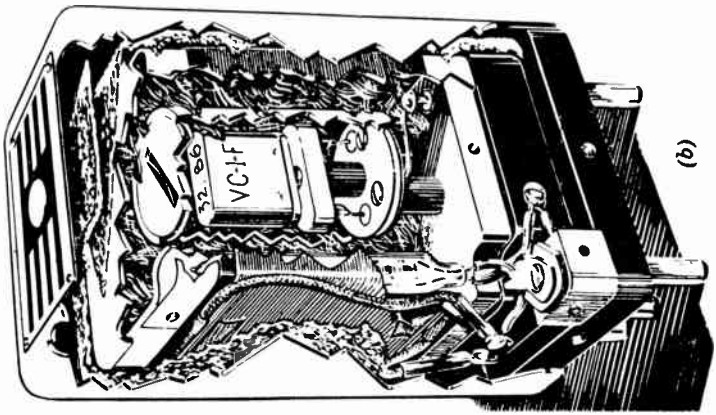
The crystal unit most widely employed as television transmitter master oscillator is an *AT* or *BT* cut crystal plate vibrating in the thickness shear mode; the *BT* cut differs from the *AT* cut only in the angle of inclination (Fig. 15.5*c*), which is here −49 degrees in place of 35 degrees. The thickness of the *AT* cut is $1.67 \cdot 10^6/f$ mm, that of the *BT* cut, $2.56 \cdot 10^6/f$ mm, f being the fundamental resonant frequency. The crystal plate is held clamped on two opposite edges between two gold-plated quartz electrodes in a hermetically sealed envelope (Figs. 15.7, *a* and *b*). The assembly is mounted in a thermostatically controlled crystal oven which maintains the temperature at 75°C.

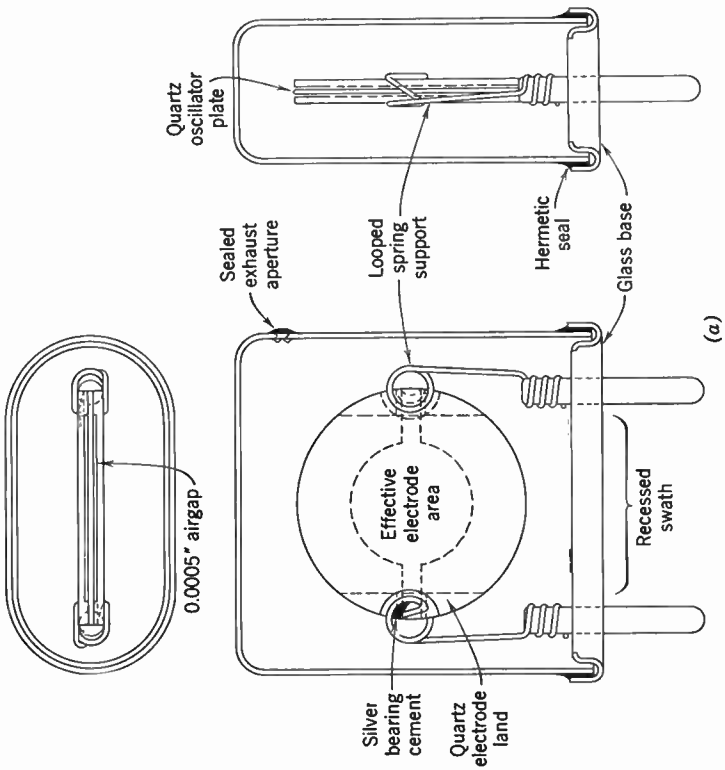
In practice, crystals with resonant frequencies in the range of 4.5 to 11 megacycles are employed in conjunction with frequency-multiplying circuits which raise the frequency of the output by a factor ranging from 8 to 36 for the very-high-frequency channels and even higher factors for the ultra-high-frequency channels. Higher oscillation frequencies, of the order of 30 megacycles, are obtained with reasonable thicknesses by employing the crystal in the third overtone thickness shear mode vibration. A maximum frequency deviation for a period of 30 days which is no greater than 5 parts in 10,000,000 is readily attained in this manner. Figure 15.7*c* shows an oscillator circuit with buffer amplifier suitable for use with such a crystal unit.†

The individual frequency-multiplying circuit is, normally, a harmonic generator. It consists essentially of a class C amplifier whose plate tank circuit is resonated to the desired harmonic of the input. The curves shown in Fig. 15.8*b* are given to clarify its operation. The bottom curve represents the voltage input supplied to the grid of the generator. The plate current and load voltage wave forms are also indicated. Figure 15.8*a* is the schematic circuit diagram of such a generator. Because the current output of the tube operated class C is high in harmonic content, such a device is a fairly efficient power

* G. A. Olive and J. G. Reddeck, "Two Methods of Maintaining the Difference between Sound and Picture Carriers at UHF," unpublished report, Radio Corporation of America, 1952.

† See Herbst and Washburn, reference 14.





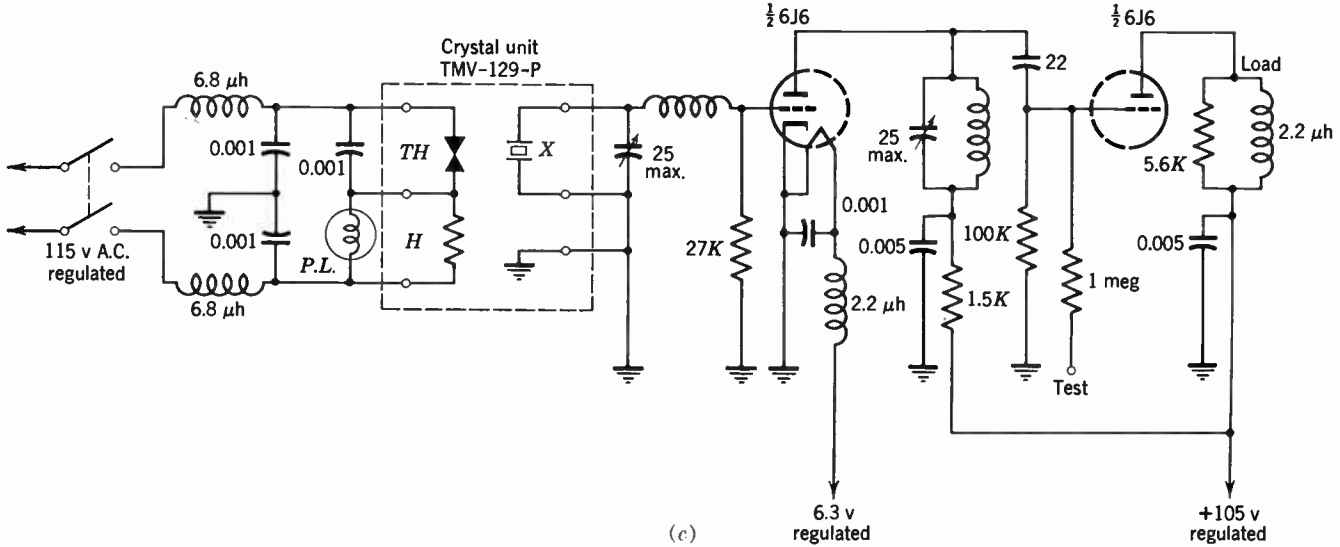


Fig. 15.7. Crystal Unit for Master Oscillator (Herbst and Washburn, reference 14). (*a*) Mounting of crystal. (*b*) Unit in thermostated insulated crystal oven. (*c*) Typical circuit of crystal oscillator and buffer amplifier.

generator. Actually, the power delivered by this type of generator is roughly inversely proportional to the order of the harmonic used; thus, operating on the second harmonic, it will deliver about 60 percent of the power that the same tube would deliver as an ordinary class C amplifier, whereas on the fourth harmonic, the power would be 30 percent. For this reason, the frequency gain per stage is commonly limited to 2 or 3.

A quartz crystal oscillator can deliver as much as 50 watts of power output. However, at very high frequencies, it is advantageous from the standpoint of stability to make the power required from the oscil-

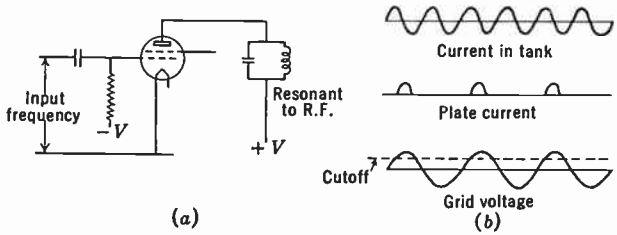


Fig. 15.8. Schematic Diagram of Harmonic Generator and Its Operation.

lator a minimum. This avoids overheating the crystal as a result of internal friction. The power required to operate the harmonic generator is small and, therefore, at most, one stage of amplification is required between it and the crystal oscillator. Examples of complete crystal-controlled carrier generators are shown in Fig. 15.3.

The low-loss "transmission line" oscillator finds more limited employment as master source for the carrier power generator. This oscillator is a vacuum-tube oscillator controlled by a resonant circuit. The resonant circuit is not a conventional coil and condenser tuned circuit, but, instead, a quarter-wavelength coaxial transmission line or some other type of resonant cavity. Such a circuit is practical only at very high frequencies, where a quarter-wave line has moderate dimensions. At present, it finds employment mainly in microwave relay systems, particularly in conjunction with klystron oscillators; at a frequency of 1000 megacycles a quarter wavelength is only 3 inches.

In the transmission line resonant circuit, the two cylinders are connected together at one end by means of a heavy metal disk, which also serves to hold the cylinders rigidly coaxial. The surface conductivity of the tank circuit must be very high because the "skin effect" at ultra-high frequencies is so great that all the current is localized within a fraction of a mil of the surface. It is particularly important that the

contact at the bridging disk be very good; therefore, often not only is solder or brazing used to hold the disk in place, but also the joint, and preferably the whole tank circuit, is plated with copper or silver.

The stability of this type of tank is a function of the Q of the circuit, where Q has its ordinary meaning of $2\pi fL/R$. Ordinarily, the frequency drift due to normal changes in driving power, load, etc., will not be greater than that which causes a 30 percent decrease in the impedance of the tank. On this basis, it can be shown that

$$\frac{\Delta f}{f} = \frac{1}{Q}$$

where f is the resonant frequency. The Q of the circuit is related to the geometry, physical size, and material of the line. It will be shown in a succeeding section that a diameter ratio of 3.6 between inner and outer conductor yields the greatest efficiency and consequently the highest Q for practical construction. Inasmuch as the factor Q is approximately proportional to the circumference of the conductors (because surface rather than volume determines the conductivity), is is desirable to make the diameter of the outer cylinder large. Resonant circuits for the very-high-frequency range require diameters of 1 or 1½ feet. When properly constructed, a concentric line tank circuit can be made to have values of Q between 5,000 and 10,000. The frequency stability may approach a few hundred cycles in 50 million.

The oscillator may be a conventional tuned-plate, tuned-grid type. For various practical reasons, the coaxial line is usually placed in the grid circuit, whereas the tuned-plate load consists simply of a relatively low-Q two-wire transmission line. Figure 15.9 shows schematically a typical push-pull oscillator using a high-Q tank. The driving unit in the example illustrated is loosely coupled to the tank circuit by a short loop near the inner conductor close to the point where it is attached to the outer cylinder. This loop is, in reality, a transformer and links the magnetic flux produced by the current in the inner cylinder. The location close to the bridging disk is, of course, chosen so as to be near the current maximum of the standing wave which exists on the line. The plate tank is coupled loosely to the load.

The high-frequency oscillation supplied by the master oscillator, whether it is crystal-controlled or controlled by a resonant cavity, must be greatly amplified before it has the power level required by the output stages. This amplification may be provided by frequency multipliers or separate amplifiers. Both operate commonly in a class C

mode. A class C push-pull amplifier is shown in Fig. 15.10, together with the form of the voltage wave applied to the grid and plate of the tubes and of the plate current. The high efficiency of this form of

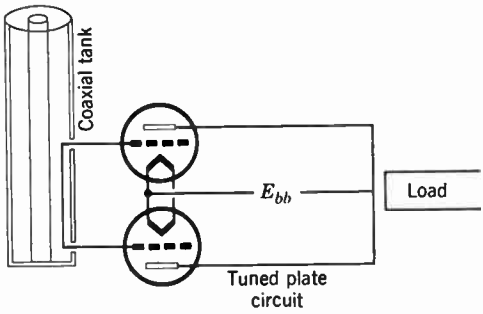


Fig. 15.9. Push-Pull Coaxial Tank Oscillator.

amplifier is a consequence of the fact that the tubes are biased to cutoff except when the plate voltage across the tubes is a minimum, making the ratio of power dissipated in the tube to power delivered to the

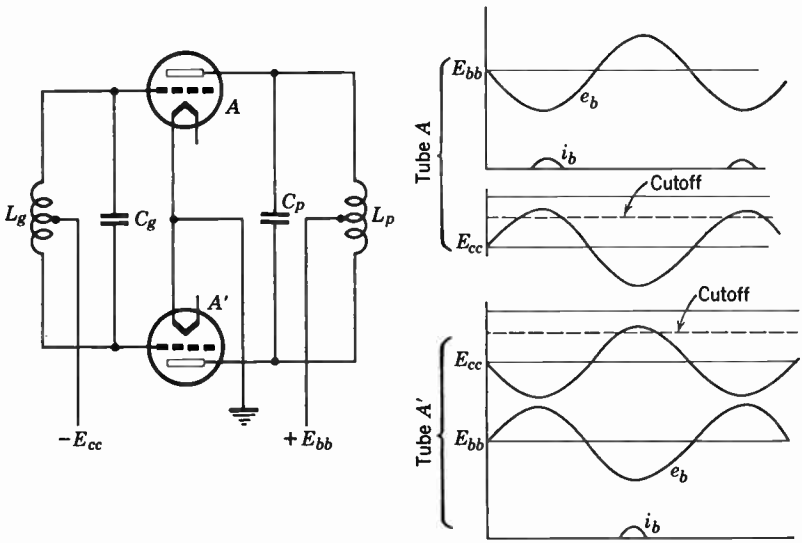


Fig. 15.10. Circuit and Operation of Class-C Push-Pull Amplifier.

load a minimum. Efficiencies of 85 percent or more can be obtained with such amplifiers, where the plate load impedance is a sharply resonant circuit with little damping, i.e., a high-Q circuit. This is to be

compared with the 70 percent efficiency which may be obtained with a class B amplifier, where each tube is biased so that it delivers current for half the radio-frequency cycle, and the very much lower efficiency of the conventional class A circuit, typified by the single-ended amplifier discussed in Chapter 13.

15.4 Modulation. In addition to the carrier generator, the transmitter contains a number of stages of video power amplification which raise the video signal to a sufficiently high level to permit nearly complete modulation of the carrier output. According to the requirements set up by the Federal Communications Commission, the modulated carrier amplitude for minimum video signal or "white" must be no greater than 15 percent of the maximum amplitude. The design of the modulated stage presents some of the most difficult problems facing the television engineer.

Modulation, as has already been pointed out, consists of varying the amplitude of the carrier in such a way that the envelope of the carrier wave corresponds to the modulating signal. The principal methods of modulation used at present for sound transmission are based on an amplifier stage whose gain is proportional to the modulating signal. Fundamentally, they may be divided into amplifiers whose gain is varied by a control of the plate voltage, and those whose gain is controlled by the mean value of the grid potential. These two classes are themselves subdivided into specific circuits of varying degrees of complexity, designed to fulfill the requirements of particular types of transmitters.

The basic circuit of a plate-modulated stage is illustrated in Fig. 15.11. The variable gain amplifier is shown as a push-pull circuit, consisting of two power tubes A and A'. The tuned-grid circuit is excited from the carrier generator, as in an ordinary radio-frequency amplifier. The plate voltage is supplied through the coupling impedance Z_c, and depends upon the voltage drop through the impedance. The current through Z_c is not only the current through the amplifier tubes, but also that to the modulator tube B. When the grid of B is negative so that the plate current is small, the voltage drop through the impedance is small, and the voltage applied to the amplifier stage a maximum. Under these conditions, the radio-frequency power delivered to the load is high. On the other hand, when the modulator tube is conducting so that the voltage drop through the coupling impedance is large, the gain of the amplifier is low. If the modulated amplifier is operated class C, a linear relationship between the modulating voltage and the radio-frequency output amplitude is obtained.

In sound transmission, the efficiency of this type of modulator can be made fairly high provided that a class B push-pull amplifier is employed as modulator stage. This is possible only where the modulation varies symmetrically about a mean carrier level. In television broadcasting, this a-c transmission is not economical, as has already been pointed out. Instead, the d-c component of the video signal is included. As a consequence, the variable gain stage must be controlled by a single-ended class A amplifier. This greatly decreases the efficiency of the system.

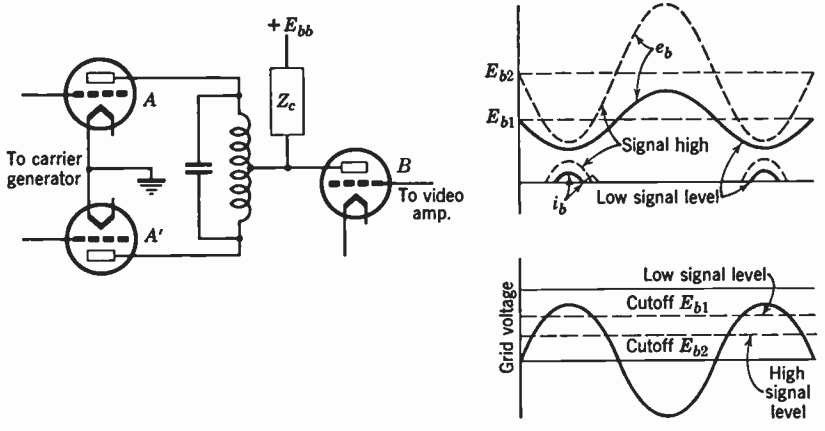


Fig. 15.11. Schematic Diagram of Plate Modulator.

For high-level modulation, plate modulation is out of question since it requires an excessive amount of modulating power. To achieve the almost complete modulation demanded by the standards set for the transmitted signal, the maximum voltage drop produced by the video current across the coupling impedance must equal approximately the d-c plate potential of the modulated stage, which may be of the order of 8 kilovolts. Furthermore, the coupling impedance, to yield uniform response over the entire video band, must in effect be a resistance, small compared to the reactance of the parallel tube capacitance at the maximum frequency of the video band. The peak video power V^2/R_c required for plate modulation of the output stage of a video transmitter may hence be shown to be of the order of 100 kilowatts.

In grid modulation, now universally employed in television transmitters, this video power requirement is much less severe. A schematic diagram of a grid modulator is shown in Fig. 15.12. Like the preceding modulated amplifier, it is essentially a variable-gain radio-frequency

power amplifier. The gain in this instance is controlled by the grid bias of the stage. Constant amplitude radio frequency is supplied to the power tubes, A and A', in phase opposition through the tuned circuit which couples them to the carrier generator. The bias for the amplifier tubes is introduced at the midpoint of the grid tank, and is obtained from a voltage supply through the coupling impedance Z_c. This impedance also carries the plate current of the modulator tube B.

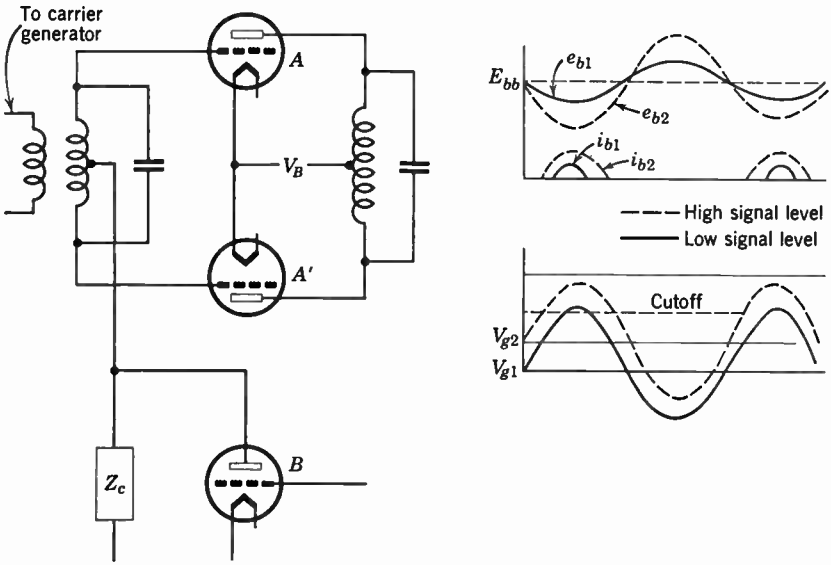


Fig. 15.12. Schematic Circuit Diagram of Grid Modulator.

The mean grid voltage applied to the amplifier has its maximum positive value when the modulator tube is at cutoff, and becomes more negative as the plate current in the modulator tube increases. To obtain linear modulation, the bias voltage supplied to the coupling impedance is generally so adjusted that when the modulator is at cutoff, each amplifier passes current for about 180 degrees of the radio-frequency cycle. As the grids are made more negative, the angle over which the tubes conduct decreases, and, at the same time, the peak value of their plate current becomes lower. Thus, the power delivered to the load decreases. As compared with the plate modulator, the plate voltage swing of the grid modulator required for 100 percent modulation is less by the voltage amplification of the modulated stage. The video power output may accordingly be less by a factor equal to

the square of this voltage amplification. This factor may be as much as 100.

It is also of some interest to consider the efficiency of the modulated stage and the power which may be derived from it, since they may serve as a guide for the tube types to be employed in a transmitter with a given power output. The conclusions apply qualitatively also to additional power amplifying stages which may be inserted between the modulated stage and the antenna, although these stages are normally operated class B rather than class C.

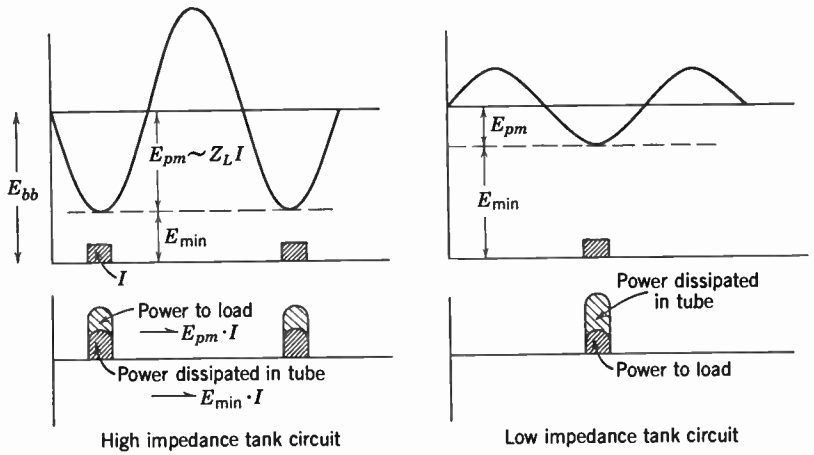


Fig. 15.13. Power Relations in Class-C Circuits.

The efficiency of the stage depends on the maximum plate current of the amplifier tubes, the minimum plate voltage for which it can be obtained, and the impedance of the tuned load. The significance of and the reason for this relationship are shown in Fig. 15.13. The first two factors are functions of tube design and are fixed by the tubes available. The third factor depends upon the properties of tuned circuits, and its maximum value is limited by the frequency bandwidth and the capacity of the amplifier tubes.

The power W delivered to the load from a class C amplifier is given by

$$W = \frac{(E_{bb} - E_{\min})I}{2} \tag{15.1}$$

where E_{bb} is the applied voltage across the load and tube, $E_{\min}$ is the minimum voltage across the tube at the r-f negative peaks, and I is

the peak value of the fundamental a-c component of the plate current. It will be seen that the power is dependent upon the radio-frequency voltage amplitude $E_{bb} - E_{\min}$, as well as upon the tube current and voltage which are determined by the amplifier characteristics. The r-f voltage is in turn proportional to the load impedance Z since

$$E_{bb} - E_{\min} = IZ \tag{15.2}$$

At resonance, when the current and voltage are in phase, the resonant circuit presents a purely resistive load, and the power is given by

$$W = I^2R/2 \tag{15.3}$$

Obviously, the power increases with increasing R.

Were only a single frequency involved, the resistance of the load at resonance could be made very large by employing a high-Q circuit. However, the impedance of the circuit must remain high over the band of frequencies covered by the sidebands. If the power for the maximum modulation frequency f_m, namely, at $f_c \pm f_m$, is not to be less than 50 percent of that at the carrier frequency f_c, the Q of the circuit cannot be greater than that given by the relation

$$\frac{f_m}{f_c} = \frac{1}{2Q} \tag{15.4}$$

At carrier frequency, the impedance of the load is

$$R = 2\pi f_c LQ \tag{15.5}$$

and, furthermore,

$$2\pi f_c L = 1/(2\pi f_c C) \tag{15.6}$$

where L is the inductance of the load and C the total capacity. Substituting from Eq. 15.6 and Eq. 15.4 in Eq. 15.5 gives the maximum impedance of the load at resonance as

$$R = 1/(4\pi f_m C)$$

Consequently, the power output available is

$$W = I^2/(8\pi f_m C)$$

Therefore, the power delivered depends not only upon the current and voltage characteristics of the tubes employed in the modulated stage and any power amplifying stages beyond it, but also upon their capacity, since this capacity is the minimum that can appear across

the load. For tubes in general use at present, this capacity places a serious limitation on the performance that can be expected from them. This is particularly true of tubes designed for high-power output.

From the foregoing, it is apparent that any modulation system, in which there is modulation of the carrier current in a tank circuit, is subject to this limitation.

Systems employing absorption modulation, impedance transfer networks, etc., arranged so that the generator maintains a constant voltage across the tank circuit, irrespective of the power delivered to the antenna, might be expected to overcome this difficulty. An interesting

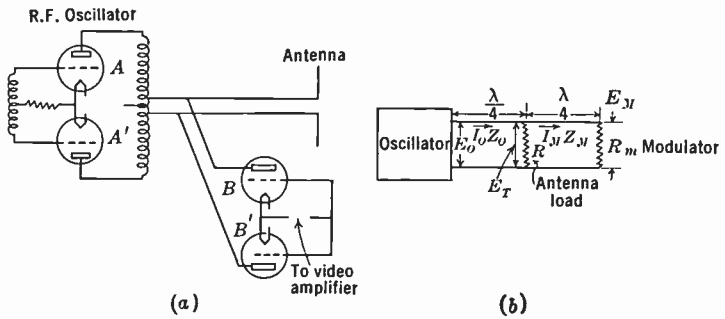


Fig. 15.14. Parker Modulation System.

example of this type of modulator is described by W. N. Parker.* The system in question makes use of the impedance inversion property of a quarter-wave line, and modulates the radio frequency in the antenna transmission line rather than attempting to modulate the power in the tank circuit of the output stage. A schematic diagram of the modulator is shown in Figs. 15.14*a* and *b*. The circuit consists of a conventional ultra-high-frequency oscillator coupled to a quarter-wave line at whose far end are two branches, one being a quarter-wave line to the antenna. In the equivalent circuit diagram, Fig. 15.14*b*, the antenna is represented by the load resistance R whereas the modulating tubes are shown as the variable conductance element R_m. It is a property of a quarter-wave line that, if its surge impedance is Z_s and the terminating impedance is R, its input impedance, as will be seen in section 15.8, is given by

$$Z_i = Z_s^2/R$$

Therefore, if the modulating tubes are biased to cutoff so that R_m becomes infinite, the impedance of the line L_m at the load resistance

* See Parker, reference 17.

approaches zero. Furthermore, since the line L_0 is also a quarter-wavelength long, its impedance at the generator becomes very high, and no power is delivered to the load. Conversely, if the grids of the modulating tubes are made positive, the value of R_m is small, and the input impedance to the line L_m is high. Under these conditions, the line L_0 is essentially terminated with the antenna load resistance R and its impedance at the generator becomes Z_0^2/R; hence power is delivered to the load.

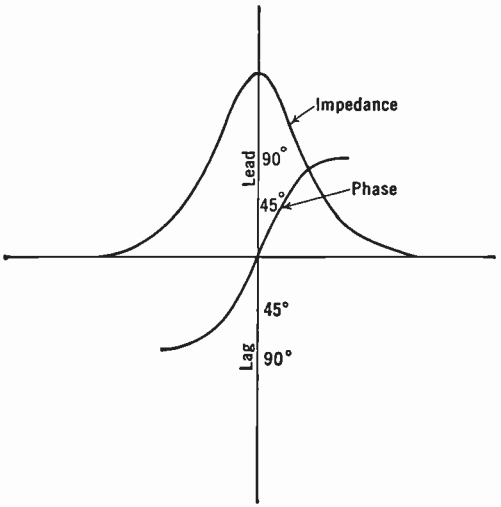


Fig. 15.15. Impedance Characteristics of Tuned Circuit.

Although this modulator does not require a rapid variation of the energy in the plate tank circuit of the power amplifiers, and, therefore, permits a more efficient utilization of these tubes, it is not independent of frequency because of the two quarter-wave lines involved. At carrier frequency, these lines can be made equal to exactly $\lambda/4$. However, this condition will not be fulfilled at the sideband frequencies, so that for these frequencies the impedance presented by the line at its input end is not purely resistive. The capacity between plate and ground of the modulator tubes and the capacity of the line itself have an effect on the response of the line which is quite analogous to that of the parallel capacity of a tuned circuit. This limits the amount of power that can be modulated with given tubes when a wide frequency band is involved. Furthermore, unless the effective resistance of the modulator tubes is very low when their grids are made positive, the impedance of the line L_m at the antenna load R will not be high enough

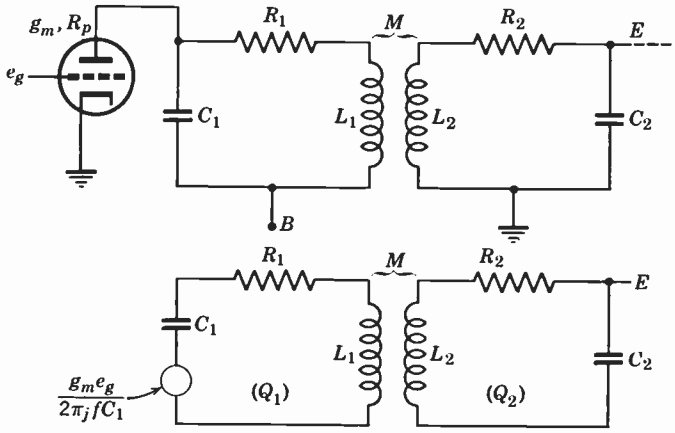


Fig. 15.16*a*. The Amplitude Response of Coupled Circuits. Circuit diagrams.

to prevent a large part of the power which should go to the antenna from being dissipated in the modulator tubes. It will be seen, therefore, that the tube requirements for this modulator are similar to those of the grid and plate modulators, namely, a low output capacity and a low minimum plate resistance. Up to now, absorption modulation has not been found advantageous in very-high-frequency television transmitters. The absorption principle offers, however, a logical approach for the modulation of high-power oscillators, such as magnetrons, for which conventional methods fail.*

The possibility of modulating at low power level for single sideband operation and amplifying the modulated signal up to the desired power has already been mentioned. On the surface, this appears to be a very attractive solution to the modulation problem because it reduces the bandwidth which the power amplifiers must handle. A difficulty is found, however, in the fact that a resonant circuit must be used as the plate load if efficiency is to be obtained from an r-f amplifier. The impedance over the resonance peak goes from capacitative reactance, through pure resistance, to inductive reactance, as shown in Fig. 15.15. Since for single sideband transmission the carrier lies at one or the other extreme of the band to be passed, the plate load for carrier frequency will be complex, and the load line an ellipse. Therefore, the carrier amplification will be inefficient. On the other hand, if the

* See, e.g., Donal and Busch, reference 18.

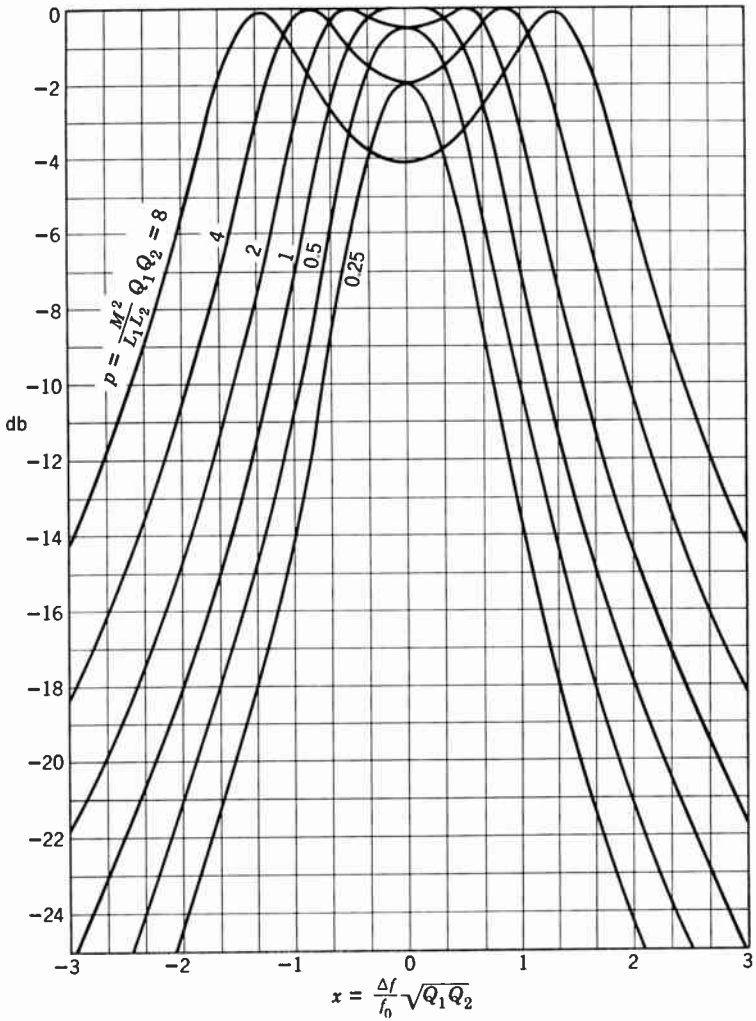


Fig. 15.16*b*. The Amplitude Response of Coupled Circuits. Amplitude response curves.

resonant circuit is made broad enough so that the carrier can be placed at the center of the peak, making the load at the carrier frequency purely resistive, then, from the same considerations as above, the plate load impedance must necessarily be low and the amplification inefficient.

An increase in gain for given bandwidth, or in bandwidth for given gain, can be achieved by using relatively strongly coupled tuned circuits for interstage coupling.* If the resonance frequency of the two tuned circuits is identical, the coupling causes, first, a broadening and flattening of the resonance peak and, for still stronger coupling, a splitting into two symmetrically placed peaks. This is illustrated, for equal Q of the two circuits, in Fig. 15.16. A flat overall response may be obtained either by the employment of critical coupling,

$$p = M\left(\frac{Q_1Q_2}{L_1L_2}\right)^{1/2} = 1$$

or an overcoupled double-tuned stage in sequence with a single-tuned stage. Double-tuned stages are relatively difficult to adjust. Hence, they may be employed in transmitter circuits but do not find favor in receiver design.

Outphase modulation, first proposed by Chireix † and applied by him to high-power radio broadcast transmitters, represents another promising approach to a low-level modulation television transmitter. The principle of the system ‡ is illustrated in Fig. 15.17. The carrier is phase-modulated by the video signal with plus and minus polarity. The two phase-modulated signals are then amplified to a high level by identical r-f amplifier chains and the outputs added at the antenna to provide an amplitude-modulated carrier signal. Since, with this system, amplifier nonlinearity has little effect on the output signal, it lends itself to the use of unconventional power tubes, such as resnatrons, klystrons, traveling-wave tubes, etc., which are efficient in the ultra-high-frequency range but are not readily amplitude-modulated.

In more detail, suppose that $e'(t)$ represents the video signal $e(t)$, eventually subjected to some prescribed nonlinear distortion, and f_c the carrier frequency. Then the desired phase-modulated signals $\sin(2\pi f_c t \pm ke')$ may be produced, e.g., by modulating the quadrature components of the carrier and adding the resultants according to the scheme:

$$\sin(2\pi f_c t \pm ke') = \cos(ke')\sin(2\pi f_c t) \pm \sin(ke')\cos(2\pi f_c t) \quad (15.7)$$

If $\cos(ke')$ is replaced by unity, $\sin(ke')$, by $ke(t)$, a simple balanced modulator is seen to be sufficient; the amplitude modulation introduced

* See, e.g., Aiken, reference 15.

† See Chireix, reference 19.

‡ See Evans, Jr., reference 20.

in this case may be eliminated by clipping and filtering in the amplifiers. Otherwise, specially designed phase modulator tubes may be required.

At the output, the addition of the two amplified phase-modulated signals leads to a radiated amplitude-modulated signal, $\sin(2\pi f_c t + ke') + \sin(2\pi f_c t - ke') = 2\cos(ke')\sin(2\pi f_c t)$.

This presupposes, however, no reaction of one output circuit on the other. In practice, the coupling through the antenna load can be large

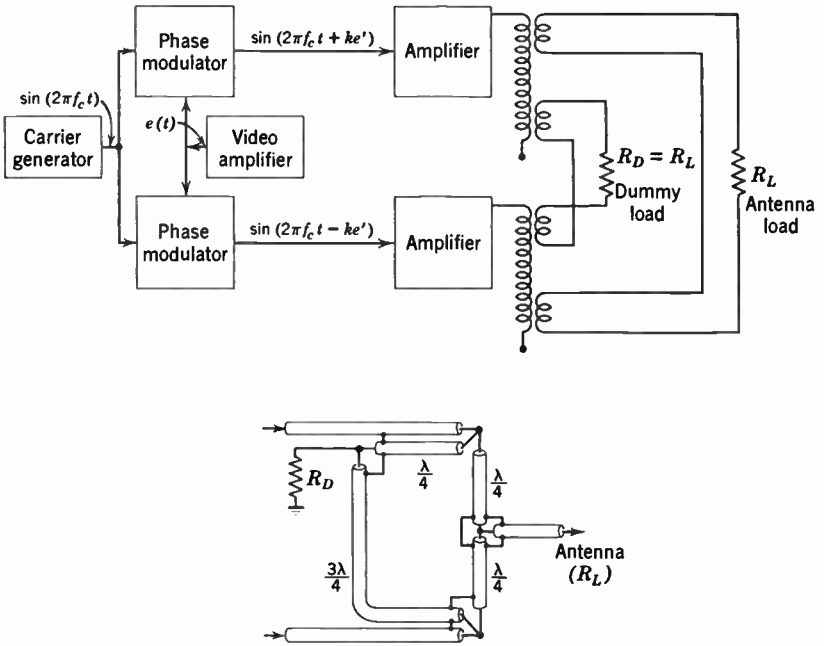


Fig. 15.17. Principle of an Outphase Modulation System. (Evans, reference 20.)

enough to alter profoundly the operation of the final power stage. To prevent this, a dummy load equal to the antenna load may be coupled to the output stages as shown in Fig. 15.17. With this arrangement, the interaction becomes zero and the load of each stage constant. The amplifier operates continuously at peak power; the power is distributed between the antenna load and the dummy load in accord with the modulation.

The variation of the power distribution with the video signal $e(t)$ depends on the relation between $e(t)$ and $e'(t)$. It is shown here for three specific examples:

System	Modulation Factors in Eq. 15.7: $\cos(ke') \rightarrow$	Modulation Factors in Eq. 15.7: $\sin(ke') \rightarrow$	Antenna Power	Dummy Load Power
Linear output	$1 - \frac{e}{e_{max}}$	$\sqrt{\frac{2e}{e_{max}} - \left(\frac{e}{e_{max}}\right)^2}$	$\left(1 - \frac{e}{e_{max}}\right)^2$	$\frac{2e}{e_{max}} - \left(\frac{e}{e_{max}}\right)^2$
Direct phase modulation	$\cos(ke)$	$\sin(ke)$	$\cos^2(ke)$	$\sin^2(ke)$
Balanced modulator	1	ke	$\frac{1}{1 + k^2e^2}$	$\frac{k^2e^2}{1 + k^2e^2}$

Although the linear-output system is more difficult to realize than the other two, it is to be preferred from the point of view of a high signal-to-noise ratio at low levels of the video signal; in the other two systems, the video signal is compressed at low signal levels.

In practice, the greater difficulty of adjustment of low-level modulated transmitters and the greater sensitivity of the transmitted picture to changes in the operating conditions of the power amplifier stages have tended to give preference to high-level modulation.

15.5 The Modulating Amplifier. The general principles involved in the design of a video amplifier have already been covered in Chapter 13. Methods for establishing the proper d-c level with reference to synchronizing signal pedestal by means of clamping circuits are discussed in Chapter 16. Since d-c transmission is employed, these also find application in the video stages of the transmitter.

Special problems arise in the transmitter, however, as a result of the very high video power which must be generated. Thus, load resistances must be generously dimensioned to provide the required power dissipation without changing their characteristics and have correspondingly large shunt capacities. These are added to the tube capacities and, hence, lower the attainable gain per stage.

The harmful effect of the shunt capacity of the load resistance can be largely overcome by replacing this resistance by a so-called constant resistance network, as shown in Figs. 15.18*a* and *b*. The network shown in Fig. 15.18*a* can be shown to have a constant resistance R for all frequencies provided that

$$\sqrt{L/C} = R \tag{15.8}$$

Thus, here, the energy dissipation is divided between two resistances equal to the original load resistance and the shunt capacity (in the RL

loop) can be correspondingly reduced. Still further improvement is possible by replacing the resistance in the RC network itself by a second constant resistance network and repeating this process several times, as shown in Fig. 15.18*b*, with

$$\sqrt{\frac{L_1}{C_1}} = \sqrt{\frac{L_2}{C_2}} = \sqrt{\frac{L_3}{C_3}} = R \tag{15.9}$$

The values of L and C, as well as the dimensions of the resistances, are increased with increasing index, so that the transfer characteristics of

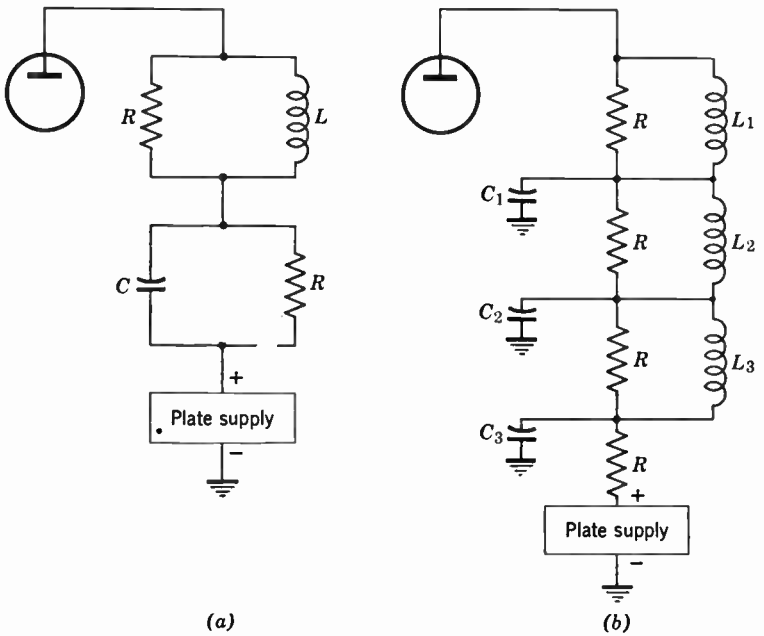


Fig. 15.18. Constant-Resistance Network.

the network for high frequencies are influenced primarily by the top meshes of the network. It will be seen that this network serves at the same time as an effective ripple filter for the plate supply of the modulating stage.

A further problem is introduced by the necessity of direct coupling between the output stage of the video amplifier and the modulated stage, required for the formation of a modulated carrier with fixed synchronizing signal amplitude. Since the d-c plate voltage of the

video output stage is of the order of 1000 volts and the cathode of the modulated stage is operated at ground, a bucking supply of some 1000 volts must be inserted between the modulator and the modulated stage. Such a supply, again, will have high capacity to ground and might hence tend to attenuate the high-frequency components of the modulating signal. Figure 15.19 shows how this effect may be avoided. In the circuit shown, the transfer impedance at low frequencies is simply

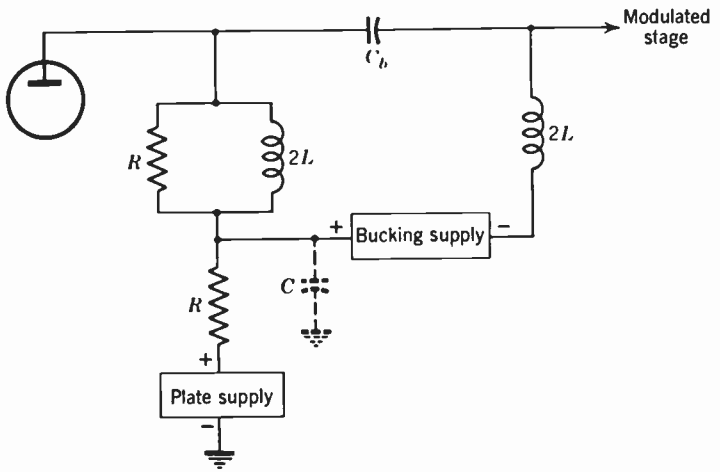


Fig. 15.19. Direct-Coupling Network for Compensating Capacity of Bucking Supply.

R, since here the low impedance of $2L$ shunts out the parallel resistance R and the capacitative impedance to ground of the bucking supply (indicated by C) is large. At high frequencies, on the other hand, both the blocking capacitance C_b and the bucking supply present negligible impedance, so that the network is essentially a constant resistance network if L is adjusted so that $\sqrt{L/C} = R$. Both principles, with slight modifications, are incorporated in the coupling network between the video output and the modulated stage in the RCA TT5A transmitter (Fig. 15.20). Other procedures have been successfully applied for attaining the same objective.

The practically complete elimination of low-frequency hum components which may be superposed on the video signal in the stages succeeding the insertion of the blanking and synchronizing signals may be effected by the employment of a so-called keyed clamping circuit, such as is shown in Fig. 15.21. The horizontal synchronizing

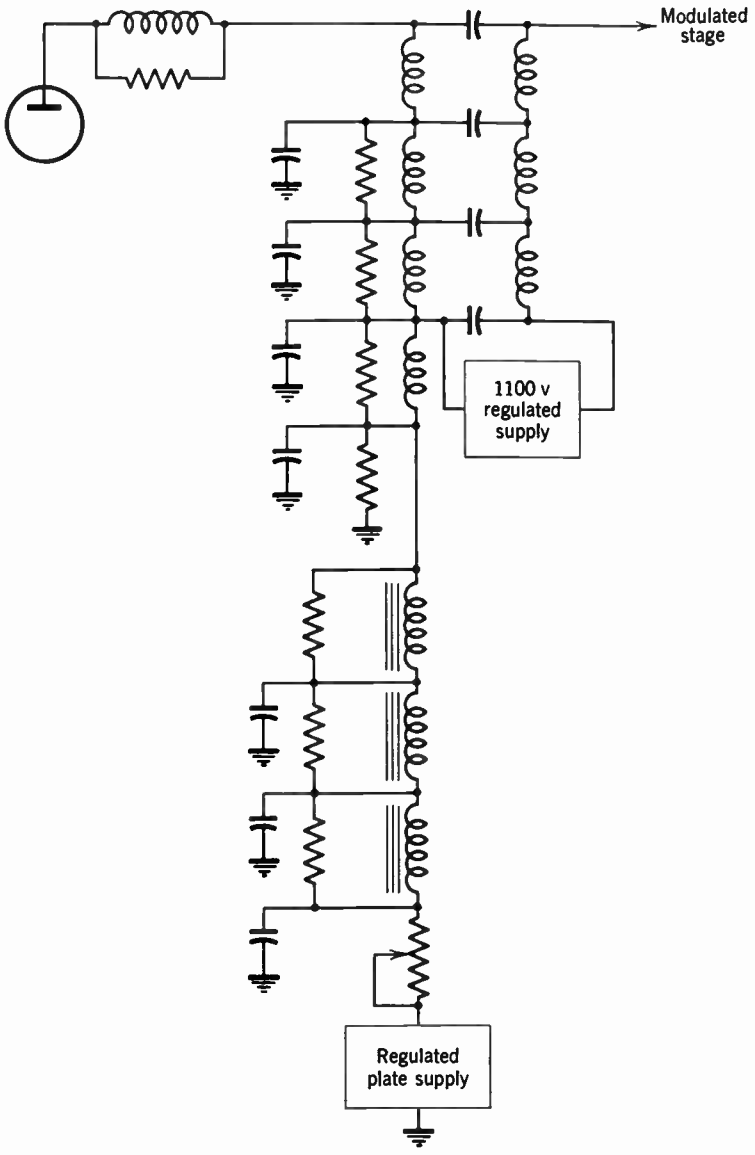


Fig. 15.20. Coupling Network for Video Amplifier and Modulated Stage in RCA TT5A Transmitter.

signal is separated, amplified, and clipped and applied through two condensers to the two halves of a double diode in opposite polarity. The remaining two electrodes of the double diode are connected to the grid of the final video stage. The diodes conduct, in view of the

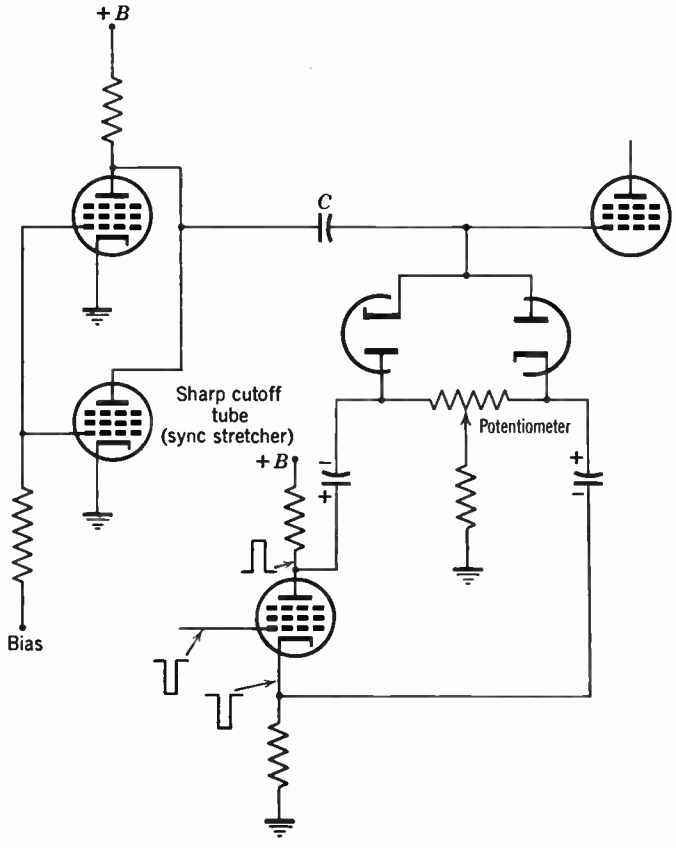


Fig. 15.21. Keyed Clamping Circuit and Synchronizing Pulse Stretcher.

bias built up on the two condensers, only when the keying pulse is applied, which coincides in time with the back porch of the pedestal of the video signal. During this brief period, the blocking condenser C is charged to a potential determined by the keying signal height and the potentiometer setting of the clamp circuit. Since the grid resistance of the clamped video amplifier stage is infinite between keying signals, the reference voltage is maintained perfectly for the succeeding line. Thus, as is shown in Fig. 15.22, a superposed hum wave affects

the video signal only by its change in amplitude in the course of a scanning line period.

It will be noted that, in the presence of hum, the clamping process may cause the synchronizing signals to be unduly reduced in amplitude after clipping. This effect may be compensated by providing a "sync stretcher." It is an amplifier stage placed in parallel to one of the regular video amplifiers, but with sharp cutoff so that it amplifies only

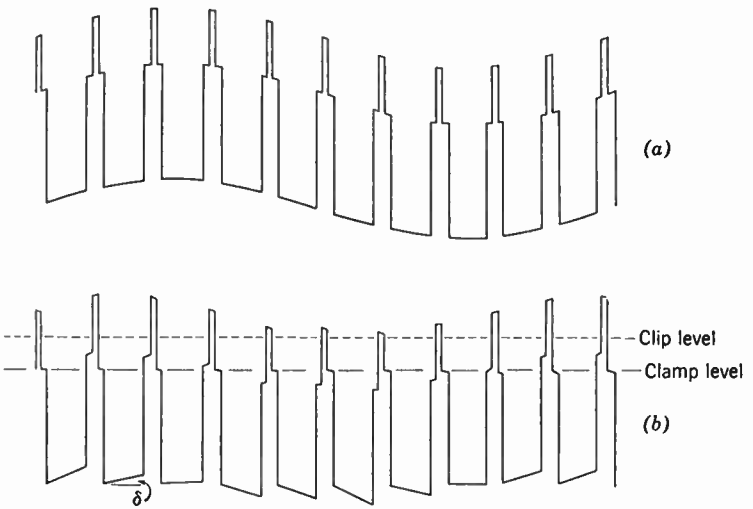


Fig. 15.22. Hum Elimination with Keyed Clamping Circuit. Signal with superposed hum (*a*) before and (*b*) after clamping. (From George E. Anner, *Elements of Television Systems*. Copyright, 1951, by Prentice-Hall, Inc., New York, p. 509. Reprinted by permission of the publisher.)

the synchronizing signals. The output of this amplifier is added to that of the parallel stage in a common load impedance.

15.6 Neutralization. The geometry of any grid-controlled vacuum tube is such that the various elements have a certain capacity to one another. Because of this capacity, when a voltage is applied to the anode, a voltage will be induced on the grid, resulting in a form of feedback. If an alternating voltage is applied to the anode, the importance of this feedback becomes rapidly greater as the frequency is increased. Thus, at radio frequency, it is necessary to take corrective measures to overcome the effects of this feedback, whereas at ultra-high frequency, this correction is a major problem. The correction process is known as neutralization. Perfect neutralization implies that voltage changes in the output circuit are without effect on the voltages appearing in the input circuit. This objective may be approached by

internal screening within the amplifier tube, as in pentodes and tetrodes as well as in the grounded-grid circuit to be considered later, or by the addition of neutralizing circuit elements.

The principal interelement capacities in a triode are indicated in Fig. 15.23. From this diagram, it is evident that, in the absence of the external circuits, the capacities C_{ag} and C_{gf} are, in effect, a voltage divider between grid and anode, and the induced voltage on the grid will be in phase with the anode voltage. The impedance of this voltage divider is inversely proportional to the frequency.

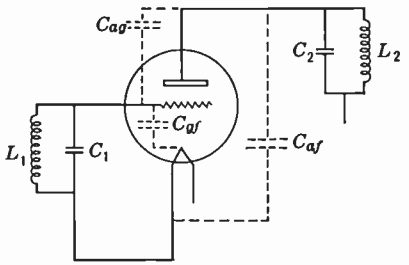


Fig. 15.23. Interelectrode Capacities of a Triode.

If both the plate load and the grid input circuit are resistive, the potential induced on the grid is in quadrature with the applied grid excitation. If the plate load is either capacitative or inductive (and the grid input resistive), the induced voltage will be in opposition to or in phase with the grid excitation, respectively. Thus, with tuned circuits as grid and plate loads, there will, in general, be some frequency near resonance for which the circuit as a whole may become unstable and oscillate. This is particularly true at very high frequencies where the grid-plate capacity impedance is small.

The high-frequency amplifier may be stabilized, or neutralized, by arranging the external circuit in such a way that a voltage is applied to the grid opposing that supplied by the interelectrode capacity. Examples of appropriate circuits are shown in Fig. 15.24.

The action of these circuits is more easily visualized if they are redrawn in bridge form as shown in Fig. 15.25. Circuits of the type shown in Fig. 15.25*a* are known as grid-neutralized, and correspond to the actual connections of Fig. 15.24*a* and *c*, whereas those of Fig. 15.25*b*, corresponding to *b* and *d* in Fig. 15.24, are plate neutralized.

For very-high-frequency work, experience has shown that bridges of the type illustrated, wherein the circuit elements used in symmetrical arms are physically different, can be balanced only with great difficulty. Therefore, it is general practice to use push-pull stages, making it possible to design a bridge which is physically symmetrical. A neutralized push-pull amplifier is illustrated in Fig. 15.26, together with the corresponding bridge.

In the very-high-frequency and ultra-high-frequency range, furthermore, it does not suffice to compensate for the interelectrode capacities

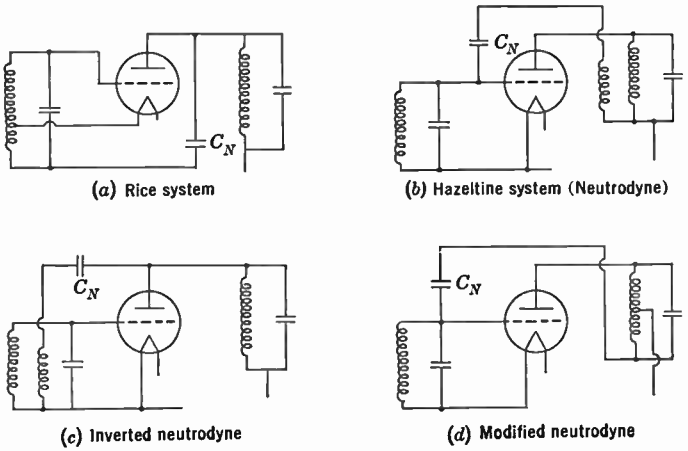


Fig. 15.24. Methods of Neutralizing Unwanted Capacities.

alone. In addition, it is necessary to take into account the inductance of the leads, which, as is evident from Fig. 15.27, makes it difficult to apply neutralizing voltages to the points required for full compensation.

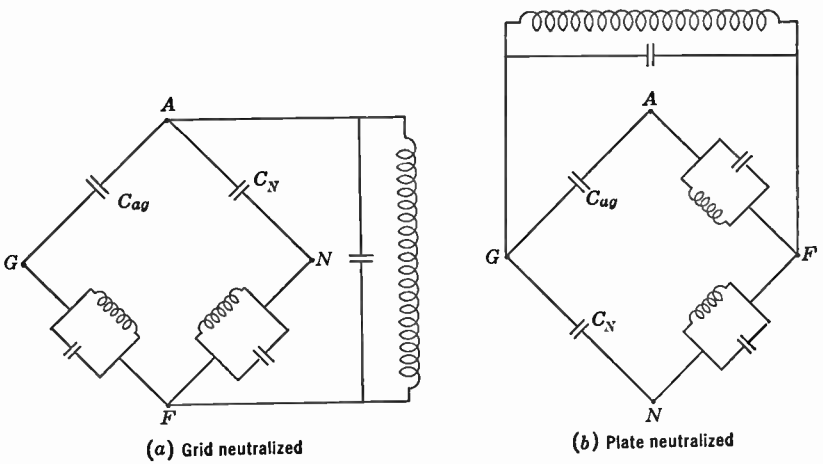


Fig. 15.25. Unsymmetrical Neutralizing Bridges.

For simplicity, only the inductance of the grid lead is shown in the figure.

It is obvious from this figure that, if the circuit elements of the bridge are such that $C_{ag} = C_N$, and a voltage is applied to the anode points, a voltage will appear at the true grid points G_A and G_B, and consequently between the filament points and the grids. In order to

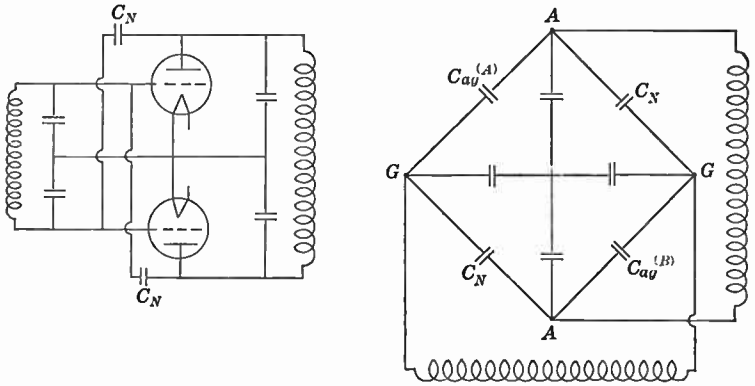


Fig. 15.26. Push-Pull Neutralizing Circuit and Equivalent Bridge.

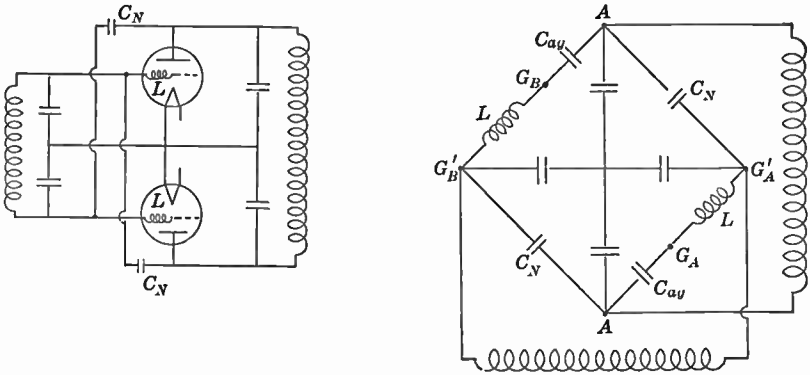


Fig. 15.27. Effective Grid Lead Inductances in Neutralizing Circuits.

illustrate the magnitude of the voltage thus induced, the following typical tube characteristics will serve for a qualitative example:

f_c = 50 megacycles, carrier
V_{af} = 3000 volts anode swing
C_{ag} = 20-micromicrofarad grid anode capacity
L_g = 0.15 microhenry inductance of grid lead

The voltage V_g appearing on the grid will be

$$V_g = V_{af} 4\pi^2 f_c^2 C_{ag} L_g = 900 \text{ volts}$$

if the input circuit impedance is neglected in comparison with the grid-anode capacity impedance. This voltage is to be compared with the 1000 to 1500 volts ordinarily used to drive such a tube.

The bridge can be neutralized in such a way as to overcome the effect of the grid lead inductance if a capacitance is put in series with the grid lead, such that its impedance just equals the impedance of the lead. This is shown diagrammatically in Fig. 15.28. The conditions for balance and no voltage at the true grid points are

$$C_N = C_{ag}$$
$$\frac{1}{2\pi f_c C_g} = 2\pi f_c L_g \tag{15.10}$$

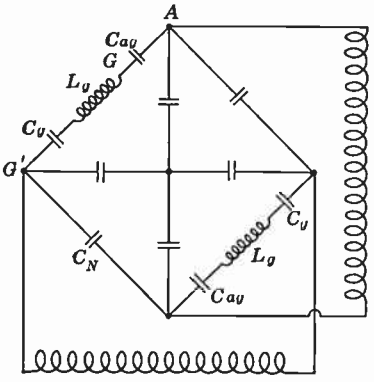


Fig. 15.28. Compensation for Grid Lead Inductances in Neutralizing Bridge.

It is immediately evident, however, that these conditions can be fulfilled only at one frequency. Where a wide communication band is used, as in television, the compensation will not be complete for the sideband frequencies. By continuing the illustrative example, and assuming a bandwidth Δf of 4 megacycles either side of the carrier, the added grid voltage at the sideband extremes is calculated as follows:

$$V_g = 4\pi^2 V C_{ag} L_g (f_c^2 - [f_c \pm \Delta f]^2) \tag{15.11}$$
$$\cong \pm 8\pi^2 V C_{ag} L_g \, \Delta f \, f_c \tag{15.11a}$$
$$\cong \pm 145 \text{ volts}$$

This voltage is in phase at frequencies above the carrier, and has phase opposition at frequencies below the carrier, therefore resulting in a slight dissymmetry of the sidebands.

This difficulty, as well as the effects of the lead inductances of the remaining electrodes, have been overcome by circuits of greater complexity. However, with increasing frequency the process of neutralization by external circuit means becomes increasingly difficult and critical. The problem can be greatly simplified, however, by modifications in tube construction and, eventually, by the employment of the grounded-grid circuit.

At moderate and intermediate power levels, the replacement of the triode by the tetrode or pentode, incorporating a screen grid to isolate the grid from the plate, appears to provide the answer. In particular, internally cross-connected dual tetrodes with built-in neutralizing capacitances provide push-pull amplifying stages which overcome difficulties from lead inductances.

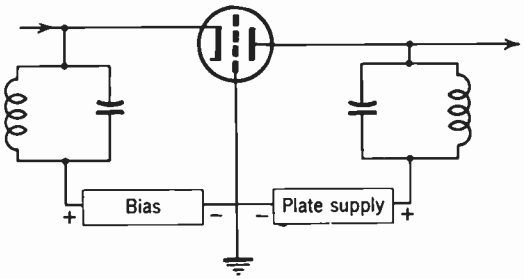


Fig. 15.29. Basic Grounded-Grid Amplifier Circuit.

At very high power levels, the construction of multigrid tubes becomes unduly costly. Here triodes in the grounded-grid connection provide the desired isolation of the input and output circuits. The basic grounded-grid circuit is shown in Fig. 15.29. It is seen that the grounded-grid plane effects perfect separation of the input and output. In practice, however, the lead inductance of the grid will prevent the

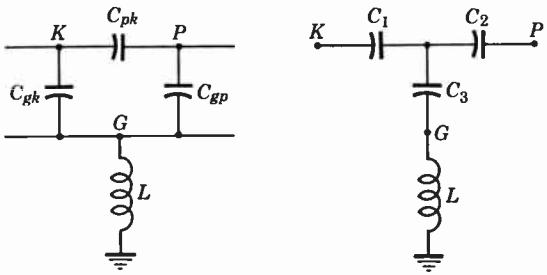


Fig. 15.30. Equivalent Circuits for Interelectrode Capacities.

grid itself from being grounded. It is then necessary to apply neutralization by inserting an appropriately dimensioned capacitance in the grid lead. The magnitude of this capacitance may be determined as follows (Fig. 15.30): The interelectrode capacitances of the tube form a π-network. If, in the equivalent T-network, the branchpoint is grounded, potentials applied at P (plate, output circuit) are seen to be

without effect on the voltage appearing at K (cathode, input circuit). The reactance of C_3:

$$C_3 = \frac{C_{gk}C_{pk} + C_{pk}C_{gp} + C_{gk}C_{gp}}{C_{pk}} \tag{15.12}$$

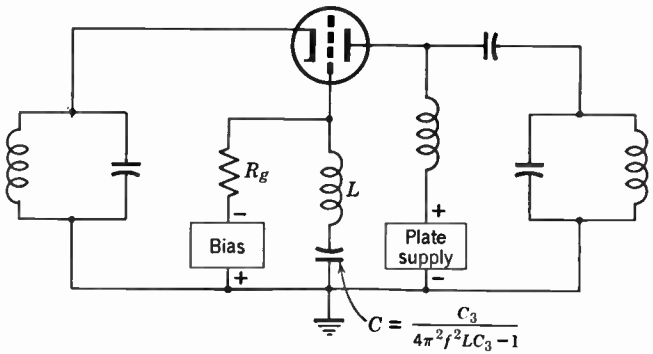


Fig. 15.31. Grounded-Grid Amplifier Compensated for Grid Lead Inductance.

is generally much smaller than that of the lead inductance L, so that, normally, capacity has to be added in series to the grid lead and the d-c grid bias applied by shunt feed (Fig. 15.31). It is clear that this

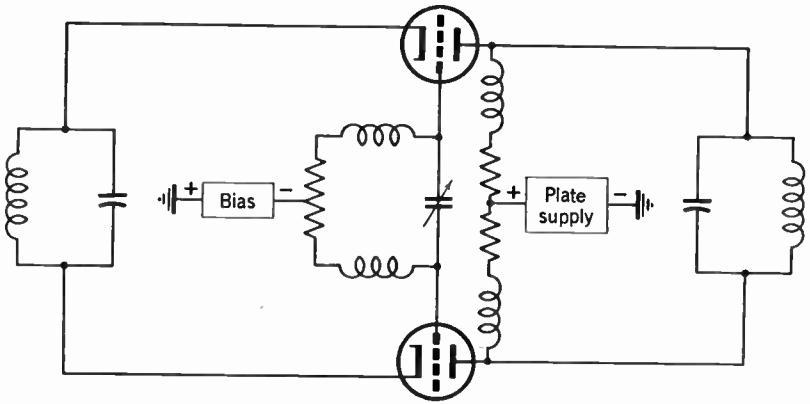


Fig. 15.32. Compensated Grounded-Grid Push-Pull Amplifier.

neutralization applies strictly only for the resonant frequency and is hence not suitable for broad-band circuits. More commonly grounded-grid amplifier stages for the modulated radio-frequency carrier are class B push-pull stages and take the form shown in Fig. 15.32.

Apart from the isolation of the input and output circuits, the grounded-grid circuit has a number of other characteristics which distinguish it from the conventional, grounded-cathode, triode circuit:

1. The power output obtainable from a given tube is increased in the grounded-grid connection, since part of the driving power is added to the output. Thus, if E_p is the (a-c) plate voltage relative to the cathode and E_g the grid voltage relative to the cathode, the power output is given by $(E_p - E_g)I_p$, where I_p is the plate current and the sign of E_g is opposite to that of E_p. For the grounded-cathode circuit, the output power is, of course, E_pI_p.

2. The output impedance for a given bandwidth can be increased since the shunt capacitance in the output circuit is reduced to one half or less. This, again, increases the power output for a given tube current.

3. The driving power required by the grounded-grid circuit, namely, $E_g(I_p + I_g)$, is higher than for the grounded-cathode circuit (E_gI_g). This is the chief drawback of the grounded-grid circuit. Its materially lower power gain per stage causes its use to be confined to high-power stages operating at very high frequencies, where the construction of pentodes and tetrodes has proved uneconomical or impractical.

15.7 Very-High-Frequency and Ultra-High-Frequency Power Tubes. The difficulties encountered in designing tubes to operate at very high frequencies increase very rapidly with the power output required. This, of course, is because of the contradictory demands on the construction. In order to dissipate large amounts of power and to withstand high voltages, the physical size must be large. On the other hand, the dimensions must be small enough so that the elements are only a small part of a wavelength, so that the interelectrode capacities are small, and so that the transit time of the electrons will be short. A compromise must be effected between these conflicting requirements.

Certain specific characteristics made evident from the discussion in the foregoing sections may be listed as follows:

1. The output d-c resistance and capacity must be as low as possible (i.e., I_p/E_p large).

2. The tube must be capable of handling the required voltages and of dissipating the power incident on the tube losses.

3. The input admittance must be low.

4. The geometrical configuration must be such that the desired potentials may be applied to the tube elements.

Each of these general requirements merits consideration, from the standpoint both of operation and of construction.

The first requirement follows directly from the discussion in section 15.4, in which the question of the output circuit of a broad-band modulator or amplifier is considered. Here it is shown that the load impedance is inversely proportional to plate capacity. Hence, since the carrier voltage for a given voltage on the tube during the time it is passing current depends upon the load impedance, it is inversely proportional to the capacity. Furthermore, the larger the current which can be driven through the tube for a given applied voltage, the greater the power that can be delivered to the load.

From the point of view of construction, the first item means that the area of the elements must be small and the spacing as large as possible in order to reduce the capacity. These conditions must be fulfilled without too great a reduction in available plate current. Therefore, emission from the cathode must be great and the elements so designed and disposed that a large electron current can reach the anode with a minimum applied voltage.

A tube operating under the conditions described for grid modulation (or as a broad-band amplifier) develops a very high voltage between anode and cathode during that part of the r-f cycle when the tube is not passing current. The peak of this voltage will, of course, be $2E_{bb} - E_{\min}$, where E_{bb} is the d-c voltage applied to plate circuit, and $E_{\min}$ is the minimum voltage appearing between plate and ground. As the efficiency of the tube is increased, this voltage becomes higher, since in the limit the efficiency as a function of voltage is $(E_{bb} - E_{\min})/E_{bb}$, assuming the operating angle to be very small. Therefore, an efficient power tube must be insulated to withstand high voltages both with respect to breakdown and dielectric loss. The elements must be formed so that there are no unshielded sharp points or edges from which cold discharge may take place or about which, when they are imbedded in a dielectric, an intense field may develop.

The power lost in the tube must, of course, be dissipated. This dissipation must take place at temperatures low enough not to melt or warp metal or glass parts. Also the temperature of the electrodes must be kept below that at which there is appreciable thermionic emission. Because of the small size of the electrodes, special provision (usually in the form of water cooling or forced air cooling) must be made to carry away the heat developed by: (1) the electron current in the tube; (2) dielectric losses; and (3) the cathode heater.

From the standpoint of operation, the requirement that the grid admittance be low needs almost no comment. The reactive component must be high since it limits the impedance of the external circuit which can be used for a given bandwidth. Obviously the grid conductance must be low since the power required to drive the grid is proportional to it; in fact, if the grid conductance becomes equal to the transconductance, the tube can no longer function as a current amplifier.

Grid capacity can be reduced only by making the tube elements small. However, this alone does not suffice. Even though the capacity of the grid may be very small when the tube is cold, the tube may have high grid capacity due to induction by the space current under operating conditions. This effect is exactly the same as that encountered in receiving tubes at very high frequencies.

The phase relation between grid voltage and current also shifts with increasing frequency, owing to electron transit time, causing the existence of a real component in the displacement current. However, because of the very much higher voltages used in connection with transmitting tubes, transit time loading of the grid circuit at very high frequencies is relatively unimportant. Another cause of grid loading is the fact that it is often necessary to drive the grid positive in order to make full use of the emission current from the cathode, and under these conditions the grid may draw large electron currents.

The final requirement, that of being able to apply the desired potentials to the elements, and only those potentials, is both interesting and difficult to meet. It is this requirement, or rather the inability to satisfy this requirement, that makes necessary the neutralizing systems described in the preceding section.

The ideal tube should be of such a size that the elements are only a negligible fraction of the wavelength used, and there should be no lumped capacitance or inductance in the tube or leads. Obviously, this ideal cannot be even approximately realized. However, capacities, and in particular that between anode and grid, can be made small; leads can be made short to avoid inductance; and the tube elements so disposed as to make them as nearly as possible part of the transmission lines comprising the external circuit.

Plate-to-grid capacity, which is perhaps the worst offender from the standpoint of inducing unwanted potentials on the tube elements with the conventional grounded-cathode operation, cannot be avoided without the aid of a screen grid. An example of a screen-grid tube with high-power output suitable for frequencies up to 300 megacycles is the

RCA 8D21, employed in the output stage of the TT-5A 5-kilowatt television transmitter.* Extremely compact construction is reconciled with the need for high-power dissipation by the circulation of distilled water in a closed circuit at a rate of 1.7 gallons per minute. Not only

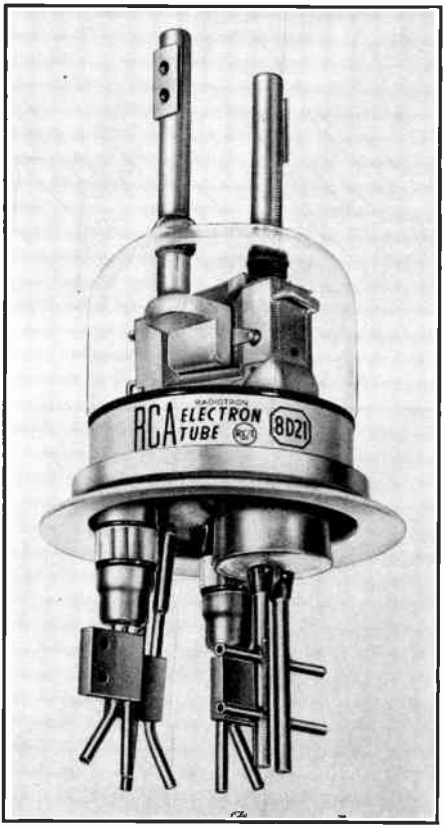
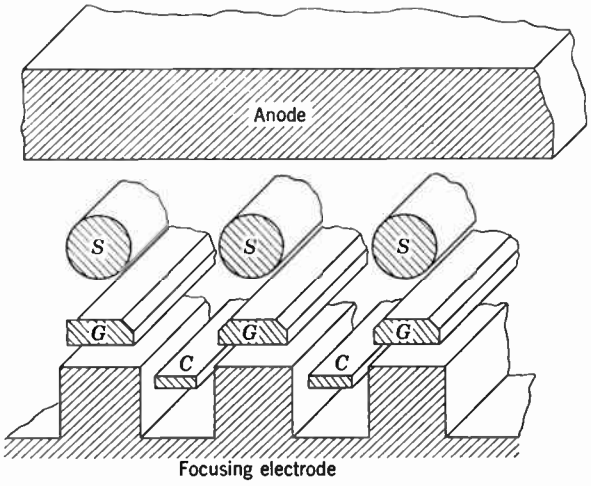


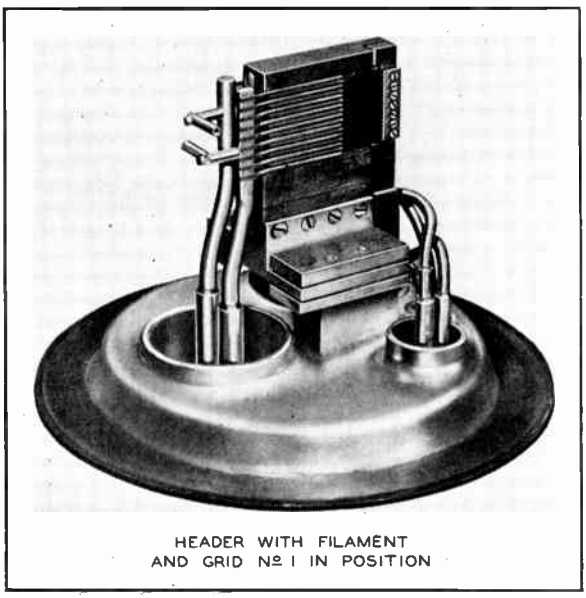
Fig. 15.33. The RCA 8D21 Double Tetrode.

the plates, but the filament mounting blocks, the screen grid, and the control grid supports are cooled in this fashion. Construction in the form of a dual tetrode makes possible neutralization of the residual grid-plate capacity within the tube, by neutralizing tabs attached to the control grids and projecting through the tube structure toward the opposite plate.

* See Smith and Hegbar, reference 23.



(*a*)



(*b*)

Fig. 15.34. Internal Structure of the 8D21 Double Tetrode. (*a*) General arrangement of beam-forming elements. (*b*) View with screen grids and anodes removed.

The overall appearance of the tube is shown in Fig. 15.33. It is mounted on a metal plate which separates the input and output stages. Both the plate leads and the control-grid leads form parallel transmission lines with provision for introducing the cooling water. The eight-stranded thoria-coated U-shaped tantalum ribbon filament is mounted at the center of the tube structure (Fig. 15.34). Its individual strands are aligned with the molybdenum control grid bars and the tubular screen grid. Grooves in a focusing block backing the filament cause the electron emission to be confined to flat beams which pass between the grid bars, minimizing grid dissipation. The water cooling of the plates is so effective that they are capable of dissipating 6000 watts with an exposed surface of only one square inch apiece. The plate-to-plate capacitance is only 6 to 5 micromicrofarads.

The water connections are made through semiflexible plastic tubing; adequate insulation of the several tube elements is maintained by inserting loops in the water connections. The screen grid supports are firmly clamped to the filament supports with mica insulation; in this manner, alignment is accurately maintained and, at the same time, an internal bypass condenser is provided between screen-grid and cathode.

Typical operating conditions for the 8D21 in the output stage of a very-high-frequency television transmitter, operating at up to 300-megacycle carrier frequency, are given below:

Filament voltage	3.2	volts
Filament current	125	amperes
Direct plate voltage	5000	volts
Direct plate current:		
Synchronizing signal	1.9	amperes
Pedestal	1.45	amperes
Grid No. 2 voltage	800	volts
Grid No. 1 voltage:		
Synchronizing signal	−220	volts
Pedestal	−400	volts
White	−820	volts
Peak Grid No. 1 to Grid No. 1 voltage	1300	volts
Grid No. 2 current pedestal	−0.025	ampere
Grid No. 1 current:		
Synchronizing signal	0.050	ampere
Pedestal	0.010	ampere
Driving power	300 to 500	watts
Output power:		
Synchronizing signal	5300	watts
Pedestal	3100	watts

A tetrode with even higher plate dissipation (10 kilowatts) is the RCA 6166 shown in Fig. 15.35. It is forced-air-cooled and has the typical coaxial structure favored for use with coaxial line input circuits as well as for multiple insertion in cavity circuits. Its power gain is approximately 9 and its peak signal output 12 kilowatts.

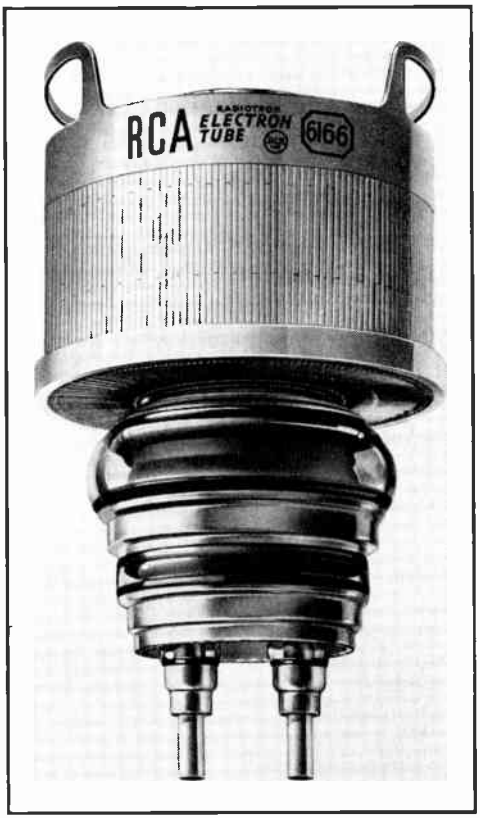


Fig. 15.35. Air-Cooled 10-Kw Tetrode (RCA 6166).

The same disk-seal construction finds general application in the power triodes employed in grounded-grid circuits, as, for instance, the 9C24's in the output stage of General Electric's TT-7A Transmitter or the 6C22's in the output stage of the Sutton-Coldfield Transmitter in England. Figure 15.36 shows a group of seven 5762 grounded-grid triodes employed as a single-power amplifier stage to raise the power output of the TT-5A transmitter from 5 to 20 or 25 kilowatts. Typical

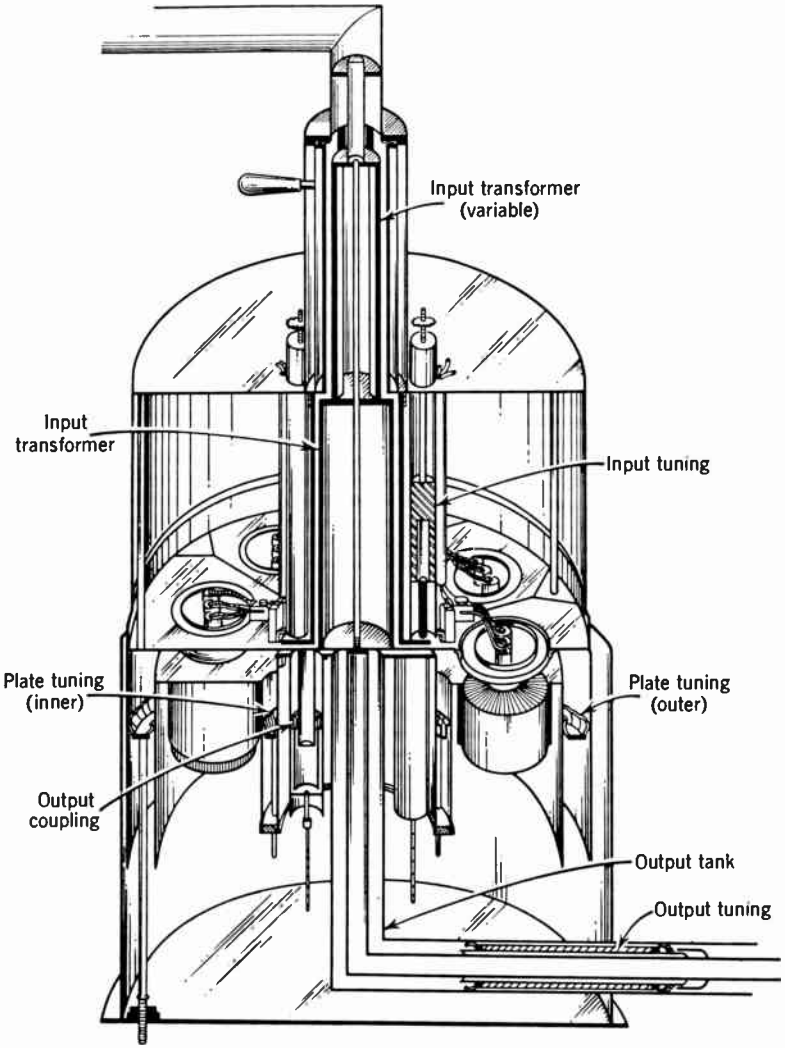


Fig. 15.36. Grounded-Grid Triode Output Stage of 20-Kw Transmitter for the High-Frequency VHF-Band with Seven RCA 5762's.

operating characteristics of the 5762 as grounded-grid class B power amplifier are given below:

Direct plate voltage	3200	volts
Direct grid voltage	−110	volts
Peak r-f grid voltage:		
Synchronizing level	435	volts
Pedestal level	310	volts
Direct plate current:		
Synchronizing level	1.8	amperes
Pedestal level	1.35	amperes
Direct grid current:		
Synchronizing level	0.400	ampere
Pedestal level	0.130	ampere
Driving power	770	watts
Power output:		
Synchronizing level	4000	watts
Pedestal level	2300	watts

It is seen that, as expected, the power gain is less than for a tetrode stage. The grid-to-plate and grid-to-filament capacitances of the 5762 are approximately equal, namely, 19 micromicrofarads. The plate-to-filament capacity, on the other hand, is exceedingly low, namely, 0.5 micromicrofarad, in view of the effective shielding by the grid.

The employment of a large number of moderate-sized tubes in parallel in resonant cavity circuits is commonly economically preferable to the use of a single high-power tube. This applies also in the ultra-high-frequency range. For example, the grid-modulated output stage of RCA's experimental transmitter in Bridgeport, Connecticut, operating with a carrier frequency of 530.25 megacycles, delivered 1 kilowatt peak power with eight 4X150 tetrodes in parallel.

The principle of distributed amplification * permits the grouping of a large number of tubes in parallel so as to add their power outputs without increasing the effective output and input capacities of the stage. To this end, the individual input and output circuits of the tubes are linked into artificial transmission lines, bilaterally terminated by their characteristic impedances. Figure 15.37 illustrates the simplest example, namely, that in which the transmission lines take the form of low-pass filters, with the grid and plate capacities as shunt elements. This arrangement would be suitable, e.g., for a video amplifier stage. The output voltage represents a growing wave passing from the left to the right, since the voltages are in phase at successive tubes. The

* See Ginston, Hewlett, Jasberg, and Noe, reference 25.

terminations prevent reflections. Finally, it is evident that the output impedance is independent of the number of tubes or meshes in the circuit.

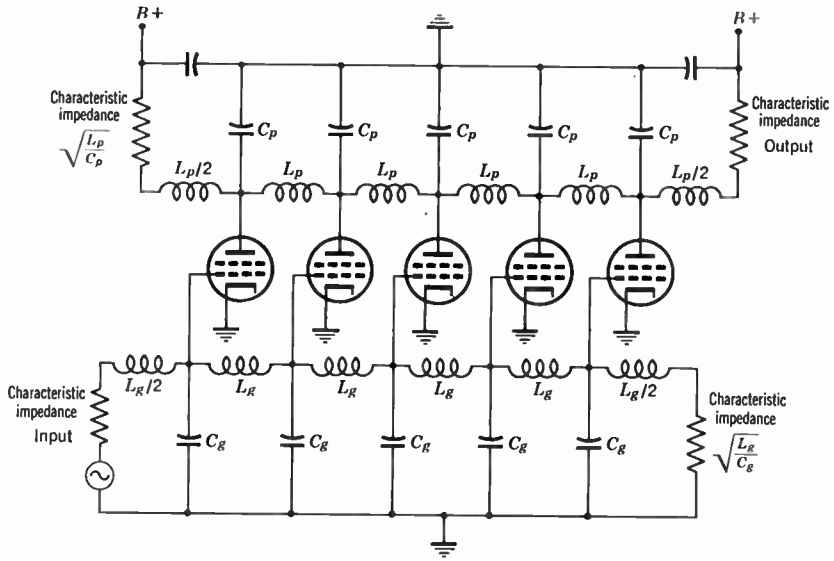


Fig. 15.37. The Principle of Distributed Amplification Applied to a Video Amplifier Stage.

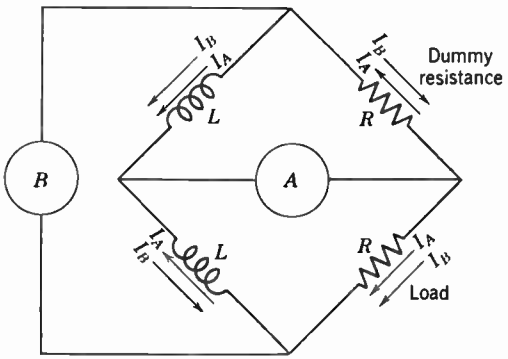


Fig. 15.38. Principle of Multiple Operation of Power Tubes.

The principle of another method for the multiple operation of tubes, which has actually been applied in the power stage of an experimental ultra-high-frequency television transmitter, is illustrated in Fig. 15.38.*

*See Brown, Morrison, Behrend, and Reddeck, reference 26.

The outputs of the two tubes A and B are applied to the two cross-arms of a balanced bridge containing two reactances and two resistances (the load resistance and a dummy resistance). In this manner, the operation of either tube is unaffected by that of the other. At the same time, the output currents are added in the load and cancel (for equal output currents) in the dummy resistance. If, now, the load resistance in two such bridge circuits, incorporating four tubes, are replaced by the cross-arms of a third similar bridge, the outputs of all four tubes are combined in the load resistance of the last bridge. Generally, the outputs of 2^n tubes, where n is any integer, can thus be added without increasing the effective output capacity. Although the drive of the several tubes should be simultaneous, this requirement is not very critical; even with considerable phase differences in the individual tube outputs, the power addition is affected only slightly. In practice, the bridge circuit is realized by a suitable combination of coaxial line elements.

It has already been pointed out that the difficulties of attaining high-power outputs with conventional tubes increase rapidly as the operating frequency is increased. These difficulties have been overcome in part by shaping the electrodes to form part of external resonant circuits, increased efficiency of water or air cooling, and improved shielding by the screen grid or grounded grid made possible by disk-seal construction. Difficulties from dielectric losses are reduced by avoiding insulators within the tube, that is, supporting all tube elements exclusively from external seals, and providing electrical shielding for external seals. A grounded-grid triode with cylindrical electrode structures suitable for the ultra-high-frequency range is shown schematically in Fig. 15.39.* With an oxide-coated cathode area of 7.5 square centimeters and an anode potential of the order of 1000 volts, a pair of these tubes delivers a kilowatt. Grid emission is prevented by selection of the material (tungsten) of the grid wires and by mounting them in a manner to carry heat off most effectively. The grid and grid sleeve effect an almost complete electrical separation of the input and output circuits.

A geometrically similar tetrode structure is realized in the Resnatron,† which was developed for radar jamming purposes and delivered as much as 50 kilowatts at 600 megacycles. Both the grid and the screen are maintained at r-f ground, effecting excellent separation of

* See Law, Burnside, Stone, and Whalley, reference 27, and Law, Whalley, Stone, reference 28.

† See Salisbury, reference 29.

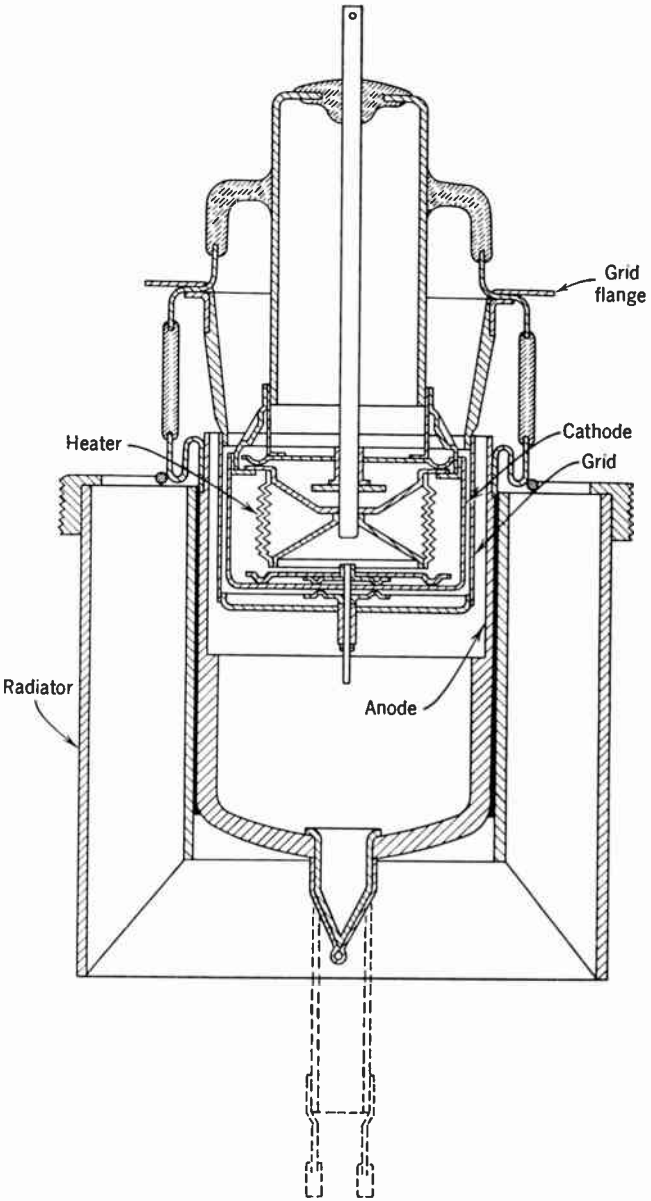


Fig. 15.39. Schematic Diagram of Experimental Ultra-High-Frequency Power Triode with Oxide Cathode. (Law, Burnside, Stone, and Whalley, reference 27.)

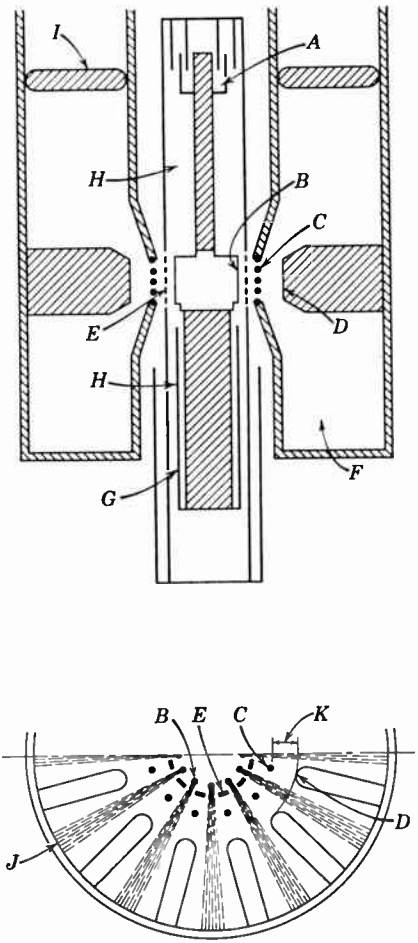


Fig. 15.40. Schematic Diagram of the Resnatron. (From Warnecke and Guenard, reference 34.) *A*, adjustable tuning capacity for the cathode-grid circuit. *B*, cathode. *C*, screen grid. *D*, reentrant portion of anode, reducing space charge and secondary emission effects. *E*, control grid. *F*, screen-grid anode cavity. *G*, by-pass condenser. *H*, cathode-grid cavity. *I*, tuning plug of anode cavity. *J*, anode. *K*, screen-grid anode interaction space. (Courtesy *Annales de radioélectricité*.)

the input and output circuits. With the screen and plate operated at a d-c potential of 17.5 kilovolts, transit time effects are minimized and a very high efficiency is realized. The structure and the beam formation in it are shown schematically in Fig. 15.40. Up to the present time, this structure has not been modified to meet television transmitter requirements.

An indication of the extreme measures which may be required in the adaptation of conventional tubes to the microwave range is given by the design of the BTL 1553 triode (WE 416 A),* employed as r-f amplifier in television relay installations operating at 4000 megacycles. Figure 15.41 shows this tube with its input and output resonant cavity circuits, which are separated by the grounded-grid plane. The grid-to-cathode spacing in this tube is only 0.0006 inch, the grid wire diameter 0.0003 inch. A compromise between output capacity, optimum coupling with the output cavity, and adequate field strength at the cathode surface to draw large currents without requiring a positive grid leads to a grid anode spacing corresponding to about a quarter period transit time or 0.012 inch at 250 volts operating volt-

* See Morton and Ryder, reference 30.

age. With 25 milliamperes plate current, the tube has a gain of 9.9 decibels and a bandwidth of 110 megacycles at 4000-megacycle carrier.

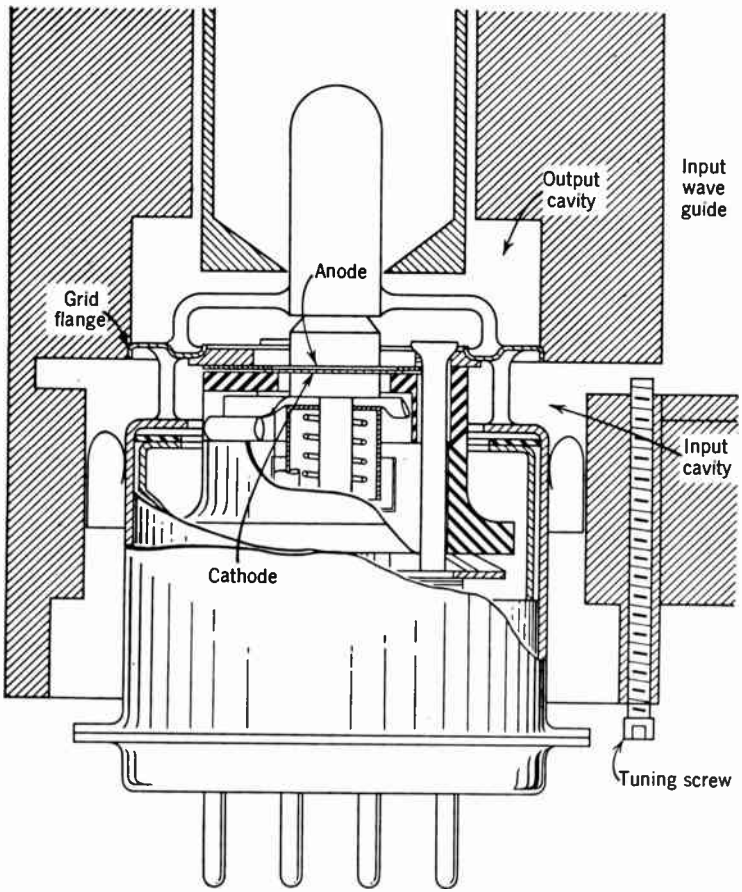


Fig. 15.41. Schematic Diagram of Microwave Triode BTL 1553.

In spite of the remarkable performance of this tube, it is evident that its design is not readily extrapolated for very high power handling capacities. To achieve this, particularly in the microwave range, it becomes advisable to consider tube constructions in which tube dimensions bear an entirely different relationship to the operating frequency. Examples are velocity modulation tubes or klystrons,* traveling-wave

* See R. H. and S. F. Varian, reference 31, and Webster, reference 32.

tubes, and traveling-wave magnetrons. Klystrons, in particular, have found application in television relay installations * and have been developed into amplifiers with large power-handling capacity.†

One form of klystron consists essentially of an electron gun forming a collimated high-velocity electron beam, two resonant cavities with pairs of grids or apertures permitting passage of the beam, and an equipotential "drift space" separating the "interaction spaces" in the two resonant cavities (Fig. 15.42). A collector with radiating fins is commonly added to dissipate the energy of the electron beam after it

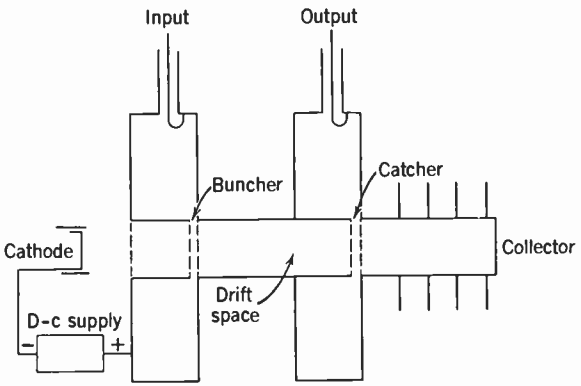


Fig. 15.42. Schematic Diagram of Klystron Amplifier.

has passed through the second interaction gap. The collimation of the beam is commonly assisted by a magnetic field parallel to the beam.

The klystron operates in the following way: Consider electrons passing through the interaction gap of the first cavity (the "buncher") during the half period of increasing accelerating field in the gap. The electrons which pass through the gap after the moment of zero field are then speeded up, whereas those that have passed through it earlier are slowed down to an increasing degree. Hence, there is a tendency for all these electrons to catch up to or fall back to those which have passed through the gap at zero field as they proceed through the drift space (Fig. 15.43). In the gap of the second resonant cavity (the "catcher"), the "electron bunches," which pass through with the frequency of the exciting signal induce fields which reinforce an existing oscillation. The klystron thus operates essentially as a class C ampli-

* Friis, reference 33.

† See Warnecke and Guenard, reference 34; Varian, reference 35; and Learned and Veronda, reference 36.

fier, with a maximum efficiency which may be computed to be 58 percent.

It is immediately obvious that the distance between the cathode and the interaction gaps bears no relation to the operating frequency of the device. Furthermore, the beam is generally accelerated before reaching the buncher so that its velocity is of the order of a tenth of

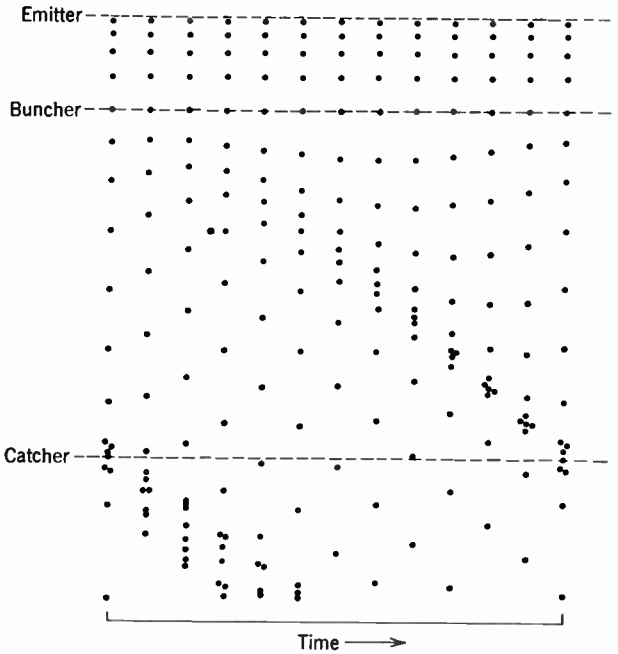


Fig. 15.43. Bunching Effect in Klystron. (Varian and Varian, reference 31.) (Courtesy of *The Journal of Applied Physics.*)

that of the velocity of light; it is then an easy matter to make the transit time through the interaction gap a fraction of a period of oscillation even for centimeter waves. The drift space between the two interraction gaps, finally, may readily be of the same order as a wavelength in dimension. In consequence, the klystron is well suited for the ultra-high-frequency range and for microwaves and is limited, rather, in the direction of the longer wavelengths.

The klystron is essentially a narrow-band device. Thus Friis finds a natural bandwidth of 5 megacycles for his klystrons operating at 4000 megacycles and achieves a desired flatness over 10 megacycles for a four-stage amplifier only as a result of stagger tuning and double

tuning. Coupling loops transfer the excitation from the output cavity of one stage to the input cavity of the next. Figure 15.44 shows one of the klystrons with magnetic collimator and external coaxial input and output cavities fitting onto disk seals on the tube. Metal screw plugs in the latter are provided for mechanical tuning. The operating

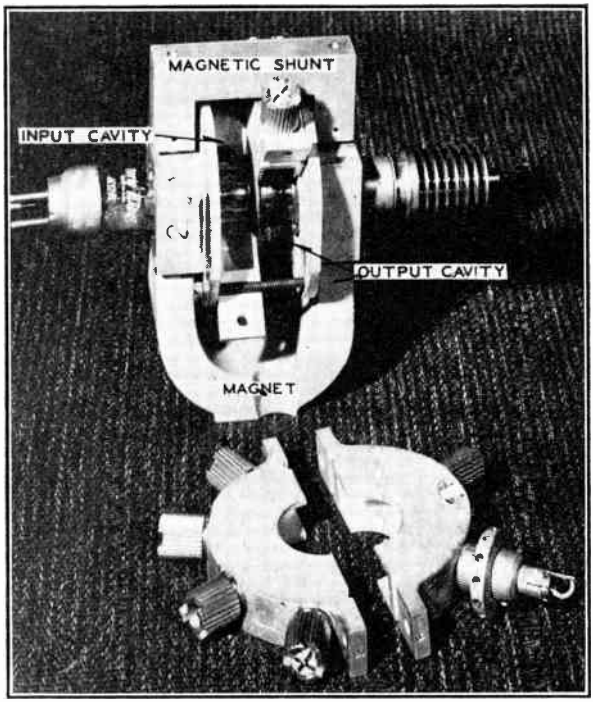


Fig. 15.44. Klystron Employed in Relay Transmitter. (Friis, reference 33.) (Courtesy Bell Laboratories.)

voltage of the klystrons was 1500 volts, and the output of the amplifier 0.7 watt.

Klystrons also find application as microwave generators. In this case, the output cavity is coupled to the input cavity or, more commonly, a single interaction gap and cavity are made to serve as both buncher and catcher by providing a retarding electrode to reflect the beam after a first passage through the gap; in this manner, the reflex klystron is obtained. The two-cavity klystron, furthermore, may be used effectively as a frequency multiplier. The very sharp pulse ob-

tained for maximum bunching is obviously very rich in harmonic content; in consequence, even the eleventh harmonic of the input frequency can be excited efficiently in the output cavity.

The gain as well as the power output of a klystron can be greatly increased by the addition of a third interaction gap and cavity. Figure 15.45 shows a tube of this type. Three-cavity magnetrons of the Sperry Gyroscope Company for operation near 3000 megacycles, such as the 2K35, are listed as having a gain of 33 decibels and a power out-

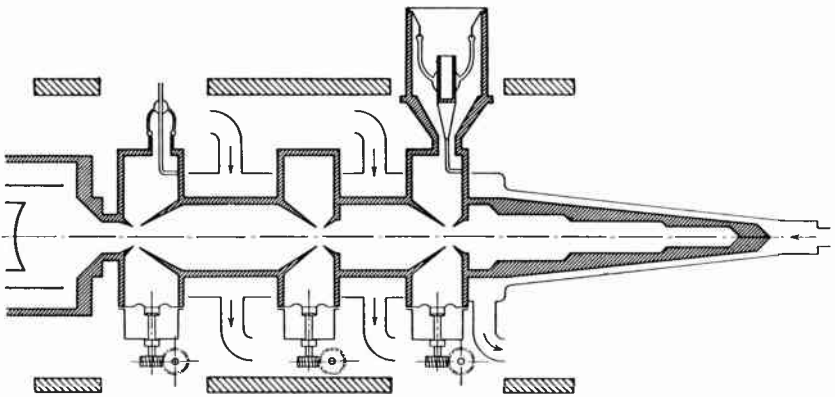


Fig. 15.45. Three-Cavity Klystron Amplifier. (From Warnecke and Guenard, reference 34.) (Courtesy of *Annales de radioélectricité*.)

put of 25 watts. Guenard, Epsztein, and Cahour * describe a tube specifically for television service in the ultra-high-frequency range (900 megacycles) with a gain of 20 decibels and a passband of 6 megacycles flat within 1 decibel which delivers 5 kilowatts. A uniform magnetic field, from which the cathode and the collector are screened to permit the narrowing and the spreading of the beam respectively, prevents dissipation of the beam on the cavity electrodes. Apertures replace the grids at the interaction gaps. The three cavities are tuned separately by deformation of one of the cavity walls, a common practice in klystron tuning. The intermediate cavity, which improves the bunching of the beam, floats with respect to the r-f potential. Tuning is, of course, so adjusted as to yield the desired broadband characteristic of the amplifier; as with other tubes, an increase in bandwidth can be achieved only by a loss in gain.

* See reference 37.

The other two tube types which have been mentioned, the traveling wave tube* and the traveling wave magnetron amplifier,† have not found, up to the present time, practical application in television transmitters. They have, however, distinctive properties which give them great promise in the microwave field and justify at least a brief description.

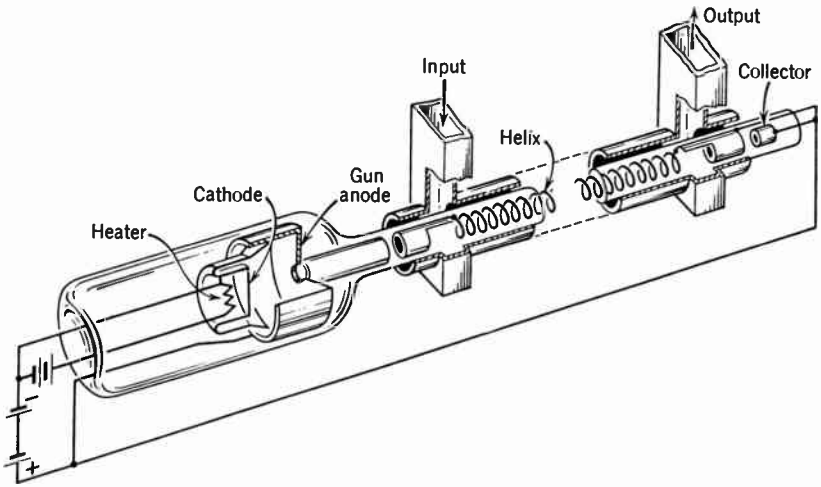


Fig. 15.46. Traveling-Wave Tube. (From J. R. Pierce, *Traveling-Wave Tubes*, Van Nostrand, New York, 1950.)

The principal distinctive feature of these tubes is the absence of resonant circuits in the input and output sections of the tube. The gain provided by the tube is the result of an exchange of energy between the signal wave traveling along a transmission line and an electron beam traveling at slightly higher velocity in close proximity to it. Since the velocity of travel of the signal along the line is practically independent of frequency, this also applies very largely to the energy exchange process. Accordingly, the passband of the traveling-wave tube circuit is not limited in the same manner as that of conventional amplifier circuits; L. M. Field has reported a bandwidth extending from 350 to 1050 megacycles.

A diagram of Pierce's traveling-wave tube is shown in Fig. 15.46. An electron beam of about 10 milliamperes and with a kinetic energy of some 1500 electron volts passes through a long helix and is even-

* See Kompfner, reference 38, and Pierce, references 39 and 40.

† See Warnecke and Guenard, reference 34, and Brossart and Doehler, reference 41.

tually absorbed in a collector. The input signal is applied to one end of the helix by a wave guide and the output signal derived from the other by a second wave guide. An externally applied uniform magnetic field of several hundred gausses confines the beam to the interior of the helix. Since the speed of propagation of the signal wave along

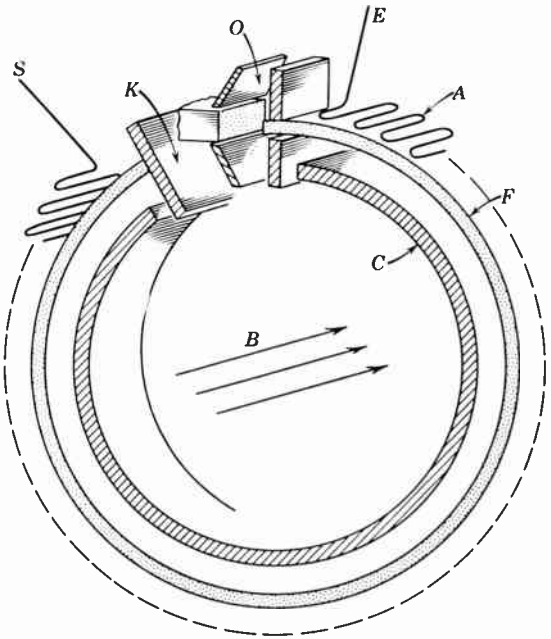


Fig. 15.47. Schematic Diagram of a Traveling-Wave Magnetron Amplifier. (From Brossart and Doehler, reference 41.) *E* and *S*, input and output antennas. *O*, beam-forming system. *A*, delay circuit (spiral). *F*, beam. *C*, central electrode. *K*, collector. *B*, magnetic field. (Courtesy of *Annales de radioélectricité*.)

the wire of the helix is approximately equal to the velocity of light, the pitch of the helix is matched to the velocity of the electrons; for 1500-volt electrons (velocity = one-thirteenth of the velocity of light) a ratio of pitch to diameter of 1:4 makes the speed of propagation along the helix and the velocity of the electrons approximately equal.

The gain of the traveling wave tube is, for otherwise constant design parameters, proportional to the length of the helix. Gains as high as 30 decibels have been achieved, and experimental tubes have been built to operate at frequencies as low as 200 megacycles and as high as 48,000 megacycles. However, it has not been possible to achieve high efficiencies and high power outputs. It is very difficult to obtain the

desired close coupling between the helix and the beam without loss of a considerable portion of the electrons by impingement, and consequent heating, of the helix. Furthermore, the requirement of a small difference in velocity between electrons and wave propagation prevents the utilization of more than a small fraction of the initial electron energy for amplification.

The last difficulty is overcome in the traveling-wave magnetron amplifier (Fig. 15.47) by causing the electrons to travel in a region of higher electric potential as they lose energy to the external circuit, here in general in the form of a flattened spiral. Essentially the same property leads to the high efficiency of the standard magnetron employed as oscillator. In the traveling-wave magnetron amplifier, the flattened spiral is given a positive d-c potential with respect to the ring electrode facing it. The electron beam passes between facing electrode and spiral along a trochoidal path, which, under certain circumstances, degenerates into a circle. As the electrons lose energy to the circuit, they drift closer to the spiral.

15.8 The Transmission Line. The modulated very-high-frequency or ultra-high-frequency output from the final power stages of the

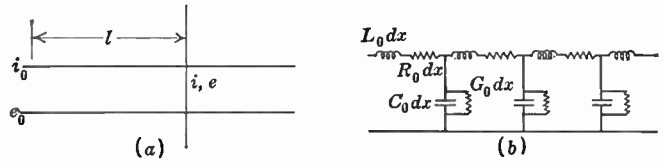


Fig. 15.48. Transmission Line and Its Effective Circuit Equivalent.

transmitter is radiated from an antenna, which, in general, is located at some distance from the generating units. Therefore, the problem arises of carrying the signal over a long length of transmission line without phase or amplitude distortion and without too great loss of power. This problem differs from that of the transfer of power at low frequencies in that the wavelength of the electrical oscillation is short compared with the length of the line under consideration. The voltage at any instant between the pair of conductors making up the line varies from point to point, going through a series of maxima and minima, and the instantaneous current likewise undergoes a series of variations. A typical transmission line is represented schematically in Fig. 15.48*a*. Between such a pair of conductors there is a certain capacity C_0 per unit length of the line, and the line itself has inductance per unit length to be designated by L_0. Furthermore, since the metal wires are not

perfect conductors, or the insulation separating the wires a perfect insulator, there will be resistance R_0 per unit length, and a conductance G_0 between them. The line may be represented by the equivalent circuit given in Fig. 15.48*b*.

For an elemental length of the line dx, the current and voltage will change by increments:

$$de = -i(R_0 + j\omega L_0)\, dx \qquad (15.13a)$$

$$di = -e(G_0 + j\omega C_0)\, dx \qquad (15.13b)$$

A complete solution of these equations describes the behavior of the line. However, such a solution is unnecessarily complicated for the present discussion. In any practical line used for short-wave power transmission, the leakage conductance can be neglected and, furthermore, $R_0/(\omega L_0)$ is very small compared with unity. Under these conditions, the voltage e and current i at a distance l from the input end of a semi-infinite line are given by the following relations:

$$e = e_1 \epsilon^{-\alpha l - j\beta l} \qquad (15.14a)$$

$$i = i_1 \epsilon^{-\alpha l - j\beta l} \qquad (15.14b)$$

where e_1 and i_1 are the input voltage and current and

$$\alpha = \frac{R_0}{2}\sqrt{C_0/L_0}, \qquad \beta = \omega L_0 \sqrt{C_0/L_0}$$

It might be noted here that, since the velocity of propagation along the line is

$$v = \frac{1}{\sqrt{L_0 C_0}} \qquad (15.15)$$

and the wavelength along the line is

$$\lambda = \frac{2\pi v}{\omega}$$

the phase factor can be expressed as

$$\beta = 2\pi/\lambda$$

The input impedance of the semi-infinite line can, with the aid of the above equations, be shown to be

$$Z_s = \sqrt{L_0/C_0}\left(1 - j\frac{R_0}{2\omega L_0}\right) \qquad (15.16)$$

This impedance is known as the characteristic or surge impedance * of the line, and is, of course, the impedance of a finite line terminated in an impedance equal to the surge impedance. For any practical line, the imaginary term $j(R_0/(2\omega L_0))$ will be less than 0.1 percent. It can, therefore, be omitted in most considerations. Under certain resonant conditions which are discussed later, it plays an important role, however, in determining the properties of the line.

When a finite line is terminated in any impedance Z_2 other than its surge impedance Z_s, it may be demonstrated to have the following characteristics:

$$Z_1 = Z_s \frac{Z_2 \cosh ml + Z_s \sinh ml}{Z_2 \sinh ml + Z_s \cosh ml} \tag{15.17a}$$

$$i_2 = i_1 \frac{Z_s}{Z_2 \sinh ml + Z_s \cosh ml} \tag{15.17b}$$

$$e_2 = e_1 \frac{Z_2}{Z_2 \cosh ml + Z_s \sinh ml} \tag{15.17c}$$

where the subscripts 1 and 2 refer to the input and output ends of the line, respectively, and

$$m = \left(\frac{R_0}{2\omega L_0} + j\right)\frac{2\pi}{\lambda}$$

The characteristics listed in Eqs. 15.17*a* to 15.17*c* are due to the fact that the abrupt change in impedance at the termination produces reflections back along the line. When the line length is a simple fraction of the wavelength, it has certain interesting and valuable properties. From a practical standpoint the most important of these special lines are those having lengths given by $l = \lambda/2$, $\lambda/4$, and $\lambda/8$.

* The characteristic impedance when the leakage conductance is considered is

$$Z_s = \sqrt{\frac{R_0 + j\omega L_0}{G_0 + j\omega C_0}}$$

Some authors distinguish between surge impedance and characteristic impedance, defining the terms as $\sqrt{L_0/C_0}$ and $\sqrt{\dfrac{R_0 + j\omega L_0}{G_0 + j\omega C_0}}$, respectively. This distinction is not made in the present discussion, since $\sqrt{L_0/C_0}$ can always be considered as an approximation of $\sqrt{\dfrac{R_0 + j\omega L_0}{G_0 + j\omega C_0}}$.

A half-wavelength line, when it is short-circuited, has an input impedance which can be expressed as

$$Z_1 = R_0\lambda/4 \tag{15.18}$$

and has properties of a series-resonant circuit. When a similar line is terminated with an infinite impedance, the input impedance becomes

$$Z_1 = \frac{4L_0}{R_0\lambda C_0} \tag{15.19}$$

and the line performs like an antiresonant circuit. Neglecting the small quadrature component of the surge impedance and writing $Z_s = \sqrt{L_0/C_0}$, a half-wave line behaves in such a way that the terminal impedance is transferred to the input impedance, i.e.,

$$Z_1 \cong Z_2 \tag{15.20}$$

This approximation requires that Z_2 be not too far different from Z_s.

A quarter-wavelength line has the following properties. The input impedance is

$$Z_1 = \frac{8L_0}{R_0\lambda C_0}, \quad \text{when } Z_2 = 0 \tag{15.21}$$

and the circuit is antiresonant. On the other hand, when $Z_2 = \infty$, the input impedance can be written as

$$Z_1 = R_0\lambda/8 \tag{15.22}$$

and the circuit is resonant. Under the same assumptions which led to impedance transfer for a half-wave line, the quarter-wave line becomes an impedance inverter, i.e,

$$Z_1 = Z_s^2/Z_2 \tag{15.23}$$

An open-circuited line whose length is equal to $\lambda/8$ has an impedance equal to the surge impedance, and a leading phase angle of 90 degrees. A short-circuited line of similar length has the same impedance, but a lagging phase angle. Similarly, very short lines under these conditions become more reactive in their behavior, approaching as a limit pure capacities and inductances having a magnitude $C_0 l$ and $L_0 l$, respectively.

These special lines are very important in the design of ultra-high-frequency tank circuits, transformers, and filters. Examples of the use of these transmission line equivalents of ordinary circuit elements will be found in the coaxial line tank circuit described above and in the

filter for removing the unwanted sideband in single sideband transmission, discussed later in this section.

The coaxial line is a very important special case of the transmission line. It consists of two coaxial cylindrical conductors, insulated from each other with low-loss spacers, so that for all practical purposes the dielectric medium between them is air. Because of its high efficiency, due to low radiation loss, this type of line is used almost exclusively in very-high-frequency power transfer. A coaxial line is illustrated in cross-section in Fig. 15.49. The surge impedance of such a line can be calculated directly, if the assumption is made that skin effect causes most of the current to flow on the outside of the conductors, so that internal inductance plays no role (a condition readily met in very-high-frequency cables). Use is made of the fact that the velocity of propagation along the cable is equal to the velocity of light and of Eq. 15.15:

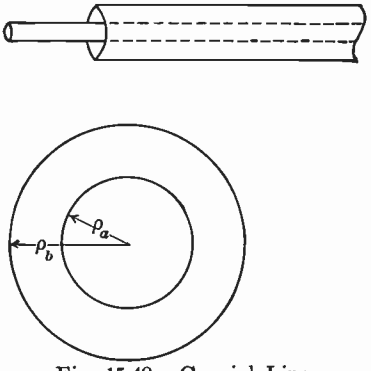


Fig. 15.49. Coaxial Line.

$$\sqrt{L_0 C_0} = 1/c, \quad \text{where } c = 3 \cdot 10^{10} \text{ cm per sec}$$

The surge impedance is therefore given by

$$Z_s = \sqrt{\frac{L_0}{C_0}} = \frac{1}{cC_0} = \frac{1}{c}\int_{\rho_a}^{\rho_b} \frac{2dr}{r} = \frac{2}{3 \cdot 10^{-10}} \ln \frac{\rho_b}{\rho_a} \text{ (esu)} \tag{15.24}$$

$$\text{or } 60 \ln \frac{\rho_b}{\rho_a} \quad \text{(ohms)}$$

Here ρ_a and ρ_b are the radii of the inner and outer conductor, respectively.

The resistance per unit length R_0 determines the power loss and also many of the electrical characteristics of the cable or transmission line. This resistance is not, however, simply the product of the d-c resistivity and the cross-sectional area of the conductor because, as has already been mentioned, the high-frequency current is concentrated within a small fraction of an inch of the surface. A fairly accurate

value can be obtained for the "skin-effect" resistance of a cylindrical conductor at high frequencies from the formula

$$R_0 = \frac{3.16 \cdot 10^{-5}}{\rho} \sqrt{\sigma f} \quad \text{ohms per cm} \tag{15.25}$$

where ρ is the radius (cm), σ the resistivity (ohm-cm), and f the frequency (sec^{-1}). In addition to the resistance of the conductor itself, radiation and dielectric loss may add to R_0 for certain types of lines, but these contribute a negligible amount in a well-designed coaxial line. It is interesting to note that the resistance varies inversely with the radius, rather than with the inverse square of the radius, as it does at low frequencies.

The optimum radius for the inner conductor, based on the most efficient power transfer, can be readily determined for a coaxial line of given radius of the outer conductor. Let I be the current flowing; then the power loss is given by

$$P_L = I^2(R_b + R_a)$$

where $R_b = k/\rho_b$ is the resistance of the outer conductor and $R_a = k/\rho_a$, that of the inner. The power transfer is

$$P = I^2 Z_s$$

The efficiency η can, therefore, be expressed as

$$\eta = \frac{P}{P_L} = \frac{Z_s}{R_b + R_a} = \frac{60 \ln (\rho_b/\rho_a)}{k\left(\frac{1}{\rho_b} + \frac{1}{\rho_a}\right)} \tag{15.26}$$

Differentiating with respect to ρ_b/ρ_a and equating to zero gives

$$\ln \frac{\rho_b}{\rho_a} = 1 + \frac{\rho_a}{\rho_b}$$

which leads to a value

$$\rho_b/\rho_a \cong 3.6$$

irrespective of the other constants of the line. This, of course, is the radius ratio mentioned in connection with coaxial tank circuits in an earlier section.

It should be noted that the neglect of the insulating wafers in the determination of the electrical characteristics of the line is justified only if the spacing of the wafers is small compared to the minimum

wavelength of the transmitted signal. This condition cannot be satisfied in the ultra-high-frequency range without the introduction of excessive dielectric loss. Specifically, if the insulator spacing is an integer multiple of a half wavelength, the extra capacity at the insulators adds so as to impress a very low shunt reactance across the terminating resistance. D. W. Peterson * has shown that the effect of the extra shunt capacity introduced by the insulators can be compensated by undercutting and grooving the inner conductor as indicated in Fig. 15.50. The grooving introduces series reactance which offsets the

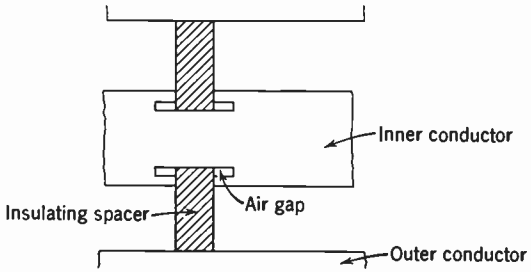


Fig. 15.50. Compensation of Electrical Effect of Spacers. (D. W. Peterson.)

increase in shunt reactance, leading to a constant value for the surge impedance $Z_s = (L/C)^{1/2}$ for any infinitesimal section of line.

The surge impedance of a line having a 3.6 radius ratio is, from Eq. 15.24, found to be 76.6 ohms. Other things being equal, it is, of course, advisable to use a line having approximately this impedance. However, as can be seen from Eq. 15.26, the efficiency curve has a very broad maximum. Lines having surge impedances ranging from 50 to 125 ohms are almost equally effective as a means of power transfer. In practice, other considerations will determine what surge impedance, within these limits, is best suited to the application in question. Problems of impedance matching are, in general, the determining factor.

Coaxial lines of the type described are used in practically all television installations to convey the power from the transmitter to the antenna. When the line is terminated with its own surge impedance, the input impedance will, from Eq. 15.24, be independent of the frequency of the signal. The attenuation of the output voltage is not entirely frequency-independent, however, because the resistance given by Eq. 15.25, which enters in Eq. 15.17*c*, is a function of frequency. In view of the square-root dependence of the resistance on frequency,

* D. W. Peterson, "High-Quality 0–1000 Mc Transmission Line," unpublished.

the variation over the bandwidth employed is unimportant in practice. The phase of the sideband frequency relative to the carrier is not fixed, but depends on the length of the line and the frequency. However, the variation is linear with frequency and hence does not introduce distortion in the video signal. These characteristics, together with its high efficiency, make a correctly terminated coaxial cable well suited to the transfer of television power.

When the transmission line terminates in an antenna, the impedance of the antenna must match the surge impedance of the line. Unless this match is fairly exact, reflections at the termination will introduce serious phase and amplitude distortion. It does not suffice to terminate the line in a load which equals the surge impedance at the carrier frequency alone, but rather the termination must match the line at all frequencies transmitted. The problem of matching the cable involves the load characteristic of the antenna; it will be considered briefly in the next section.

As has already been mentioned, circuit elements are often inserted in a transmission line to filter out unwanted components of the frequency spectrum. Specifically, a filter is required for single-sideband transmission when both sidebands are supplied by the output stage of the transmitter. Such a filter must be dissipative and must match the surge impedance of both the cable from the transmitter and that from the antenna if reflections are to be avoided. Figure 15.51 illustrates a filter element for sideband removal designed by G. H. Brown.* This filter is in reality a high-pass filter between the antenna and the transmitter, whose circuit equivalent is shown in Fig. 15.51*b*. The circuit constants are so chosen that L_2 and C_2 resonate at the frequency f_0, which is equal to or slightly lower than carrier frequency. L_3 and C_3 resonate at f_c, about 2 megacycles lower than f_0. At f_0, the network becomes in effect a π-section terminated by the antenna, whereas for the lower frequency f_c, it is a π-section supplying the dissipating resistance R. The circuit presents an impedance of approximately 77 ohms to the line over the entire frequency range.

In the coaxial line filter (Fig. 15.51*a*) employed in the television transmitter, half-wave (or three-half-wavelength) lines replace the resonant circuits, and open and closed one-eight-wavelength lines the capacitance and inductance. The dissipation of the lower sideband energy takes place in a water-cooled metallized ceramic resistor which forms part of the inner conductor of one of the coaxial lines.

* See Brown, reference 42.

The transmission characteristic of the filter in question is shown in Fig. 15.52. Three such filters in series are employed, in general, to achieve the desired attenuation. In addition ,"notching filters," constructed in similar fashion, are added to greatly increase the attenuation a quarter megacycle below the lower limit of the transmission

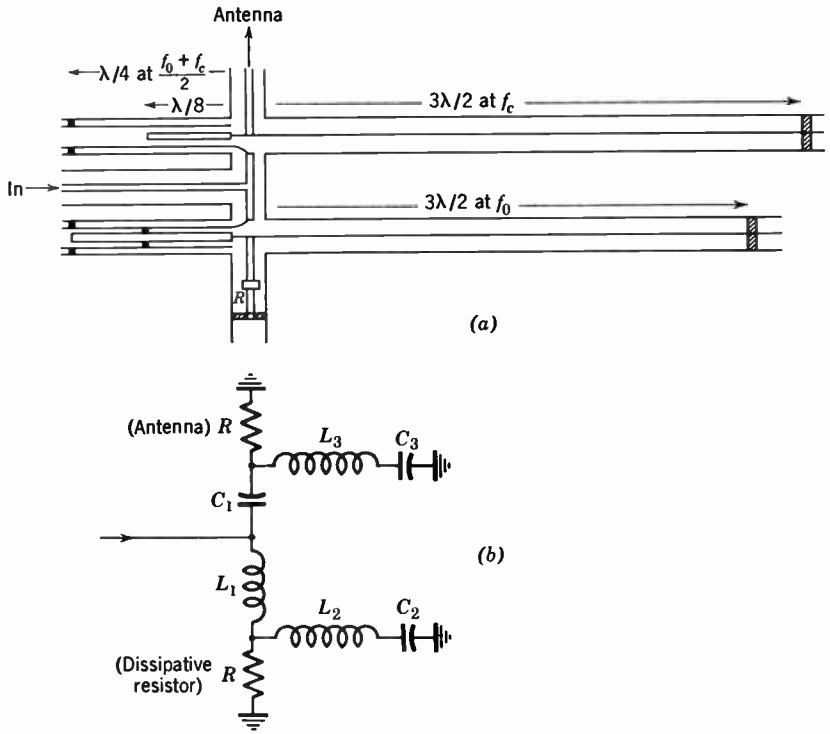


Fig. 15.51. Coaxial Line High-Pass Filter and Equivalent Circuit (Design for $f_0 = 45$ Mc).

channel, since the sound carrier of the adjoining channel is located at this point.

The recognition that the energy in the reject band which must be absorbed is small and can readily be dissipated if reflected to the output stage has favored the introduction of simpler, purely reactive, filter designs.* Since filter and coupling dimensions can be held down so that the reflection time is only of the order of 0.01 microsecond, no perceptible echoes are produced by the reflected signal.

* See Bradburd, Alter, and Racker, reference 43.

In principle (Fig. 15.53), the filter consists of two m-derived T-sections between two m-derived T half sections. The m-value for the center sections is chosen so that the attenuation approaches infinity at 1.5 megacycles below the visual carrier, i.e., at the sound carrier of the adjacent band; for the high very-high-frequency bands, this requires $m \cong 0.07$, for the low very-high-frequency bands, $m \cong 0.15$. The m-value of the half sections is set at $m = 0.30$ so as to produce an opti-

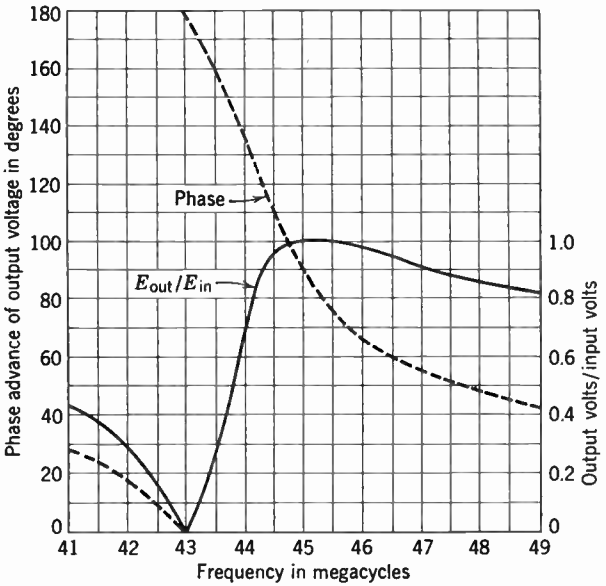


Fig. 15.52. Transmission Characteristic of Filter in Fig. 15.51.

mum impedance match to the 51.5-ohm generator and antenna lines throughout the unattenuated portion of the television channel.

The series-resonant circuits of the shunt arms are approximated by half-wave or full-wave transmission lines short-circuited at the far end; the approximation is improved by employing two sections of the central conductor with slightly differing diameters. The series capacitances are realized by gaps in the central conductor with dielectric spacers. Since the attenuation of the m-derived filters increases very rapidly beyond the cutoff frequency, relatively few sections are sufficient to satisfy the requirements laid down by the FCC and RTMA.

It may be noted that the coaxial line may be regarded as a special case of a wave guide, i.e., a system of conductors which confines an electromagnetic wave field to a restricted region of space. It has the

distinction of fulfilling this function for oscillations of all wavelengths or frequencies; the insulation between the inner and the outer conductor permits the maintenance of an electric field between them even

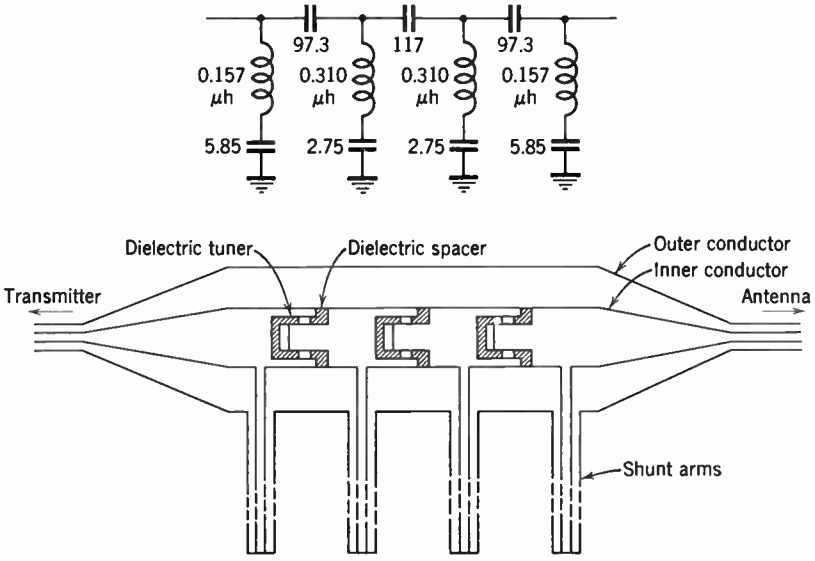


Fig. 15.53. Principle of Reactive Vestigial Sideband Filter (Values for Channel 7). (Bradburd, Alter, and Racker, reference 43.)

at arbitrarily low frequencies. Furthermore, for all wavelengths materially larger than the transverse dimension of the coaxial line, the

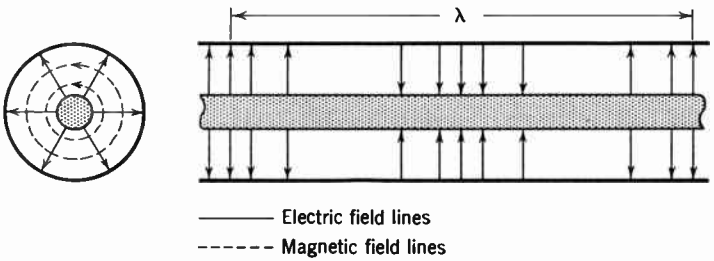


Fig. 15.54. Field Pattern of Coaxial Line.

character of the field, or the "mode," is unique. The electric field lines are radial, the magnetic field lines circular and at right angles to the electric field lines (Fig. 15.54).

Similar waves can be propagated along a single wire of *finite* conductivity. This was shown by Sommerfeld over fifty years ago.* Although the attenuation for such surface waves on wires is small, they have proved of little practical value since they extend over a large volume of space surrounding the wire and are scattered by any obstacles in their path. Goubau † has shown, however, that a thin dielectric layer (or other surface modification) on the wire causes a rapid decrease of the field intensity from the wire outward without a comparable increase in the signal attenuation. Thus, a simple enameled copper wire becomes an effective wave guide in the frequency region above 100 megacycles. The signal may be launched on the

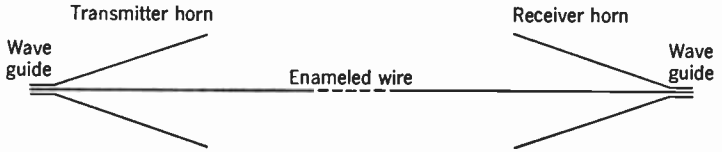


Fig. 15.55. Single-Wire Transmission Line. (Goubau, reference 46.)

wire from a coaxial line by making it the continuation of the inner conductor and flaring the outer conductor into a gently tapered horn with a diameter sufficiently large to embrace the region within which the wire wave fields are of significant magnitude (Fig. 15.55). A similar horn adapts a coaxial line to the wire at the receiver. Figure 15.56 relates the ratio of the horn radius ρ to the wire radius a' leading to a transfer loss of 1, 3, and 6 decibels to the dielectric layer thickness $a'-a$; $\varepsilon_i/\varepsilon$ is the ratio of the dielectric constants of the layer and of free space. Specifically, for a 2-millimeter copper wire with a 0.05-millimeter coating of enamel ($\varepsilon_i/\varepsilon = 3$), 90 percent of the power flow of a 3000-megacycle signal is confined to a radius of 10 centimeters about the wire.

The signal attenuation along dielectric-coated wires is small compared to that of conventional coaxial cables. For a conductor diameter of 1.42 centimeters and a polyethylene coating of 0.34 centimeter the attenuation at 200 megacycles is computed to be approximately 3 decibels per mile; losses at supports are not included in this figure. For thin wires, employing waxed thread supports, they have been found to be very small, however. The effects of rain have also been found to be relatively small.

* See Sommerfeld, reference 44.

† See Goubau, reference 45 and 46.

At very high frequencies, electromagnetic waves can be propagated within tubes even in the absence of a second, inner, conductor. Thus, oscillating electric fields can be maintained between the two wider faces of a rectangular wave guide provided that the wavelength of the oscillation is less than twice the width of the wave guide. The corresponding magnetic field lines are loops parallel to the wide face of the line (Fig. 15.57). Rectangular wave guide with a width twice as great

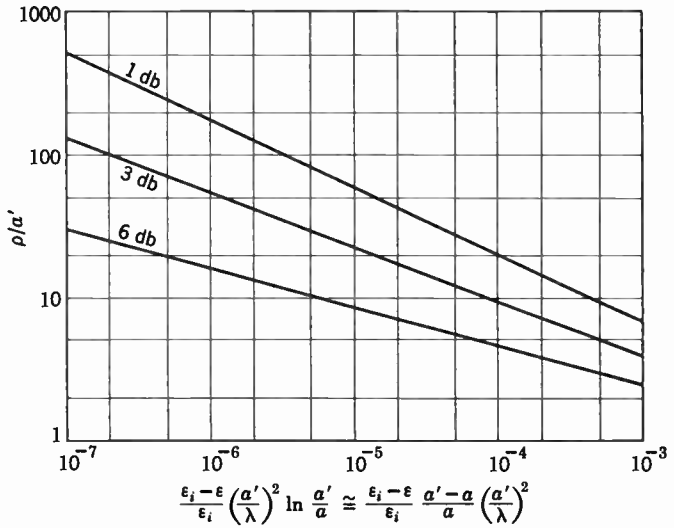


Fig. 15.56. Horn Radius ρ, Giving Indicated Transfer Losses (1, 3, 6 db) as Function of Dielectric Film Thickness $a' - a$. (Goubau, reference 46.) (Courtesy of *Proceedings of the Institute of Radio Engineers.*)

as its height is favored since, for wavelengths not too greatly different from the cutoff wavelength, transmission can take place only in the mode here represented (the "TE10" mode). Rectangular wave guide represents the most efficient means for channeling microwave power and is quite generally used for this purpose in, e.g., television relay equipment.

If a section of wave guide is sealed off by conductors, it represents a resonant cavity, in which electromagnetic oscillations (standing waves) of specific frequencies related to the cavity dimensions can maintain themselves. In fact, any cavity with highly conducting inner walls will act in this manner. The rate at which one of the resonant oscillations, once excited, is damped is determined by the resistivity of the walls or an artificially introduced damping or coupling element.

Examples of resonant cavities have been indicated in the discussion of power tubes for ultra-high frequencies.

Coupling between cavities, wave guides, and coaxial lines may be effected by windows, coupling loops inserted in the guide perpendicular to the magnetic field lines, or probes inserted in the direction of the electric field lines.

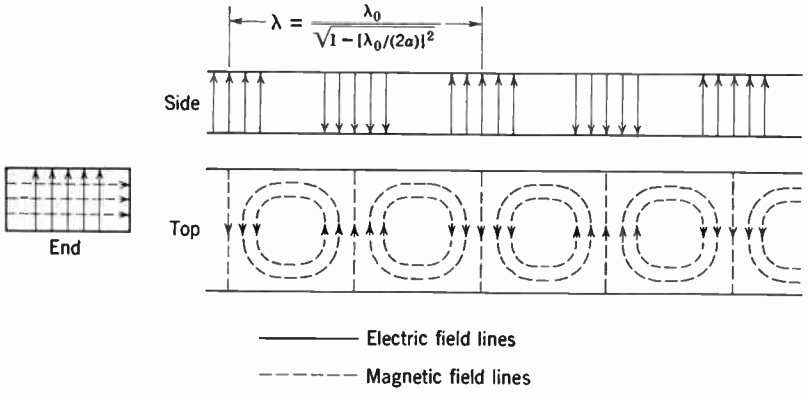


Fig. 15.57. Field Pattern in Rectangular Wave Guide (TE10 Mode).

15.9 The Antenna. The antenna is the means of coupling the output of the transmitter to the medium through which electromagnetic waves are propagated. Its operation follows directly from the Maxwell equations for the electromagnetic field, which predict that an accelerated electric charge will give rise to radiation propagated at the velocity of light. In the case of an antenna, the accelerated charges are the oscillatory current flowing in the antenna conductors.

It can be shown that an element of wire of length dl, carrying a current I, whose frequency is f, will produce a field strength E given by

$$dE = \frac{60\pi}{xc} fI\, dl \cos 2\pi f\left(t - \frac{x}{c}\right) \cos\theta \qquad (15.27)$$

In this equation, x is the distance from the element in meters and θ the angle between the direction of x and the plane normal to dl, E being measured in volts per meter. If the conductor has finite length, the radiation at any point can be determined by adding up the radiation (taking into account magnitude, phase, and direction of the electric vector) from every element of the conductor. It is obvious that the way in which the radiation is distributed about the conductor, or, in

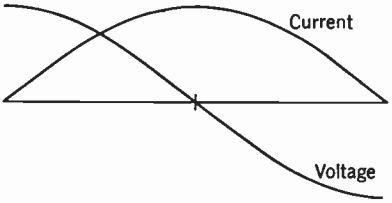
other words, the radiation pattern, depends upon its geometry and current distribution. Determining the radiation pattern is really an interference or diffraction problem.

A simple but very important antenna for high frequencies is that formed by a straight wire, a half-wavelength long, located a number of wavelengths above ground. Such an antenna is known as a dipole. The current and voltage in such a conductor form standing waves, giving the distribution shown in Fig. 15.58. The radiation pattern is included in the figure. The angular factor is here

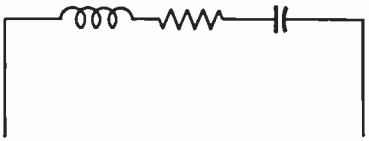
$$\cos\left(\frac{\pi}{2}\sin\theta\right)/\cos\theta$$

which differs but little from the factor $\cos\theta$ for the infinitely short dipole. The electrical properties of a half-wave dipole are essentially those of a resonant circuit. If the radiator is divided in the middle at the current maximum and the two halves are connected to the ends of a transmission line, the termination of the line can be represented as a series circuit consisting of an inductance, a condenser, and a resistance. The value of the resistance is determined by the power which is radiated from the antenna for a given current from the line and is known as the radiation resistance. It is equal to a resistance which will dissipate the radiated power when a current equal to the antenna current flows through it.

The radiation resistance of a half-wave dipole is approximately 73.2 ohms. As the length is increased, the impedance increases, having a large inductive component, which goes through a maximum and then decreases until at one wavelength the antenna is essentially a parallel resonant circuit presenting a large purely resistive impedance. At three-halves wavelengths, the antenna is again series resonant with a radiation resistance in the neighborhood of 150 ohms. Decreasing the length of an antenna from a half wave length increases its impedance by adding a capacitative component. Obviously, a variation in frequency produces an effect on the impedance similar to lengthening or shortening the antenna. This variation in magnitude and phase of the impedance of an antenna with frequency introduces a difficulty when the antenna terminates a transmission line in that it matches the surge impedance of the line at only one frequency. To overcome this difficulty for wide-band transmission, such as is involved in television broadcasting, it is necessary to employ either special matching circuits or an antenna which has a flatter impedance characteristic. In actual practice, both methods are used.



(a) **Current and voltage distribution**

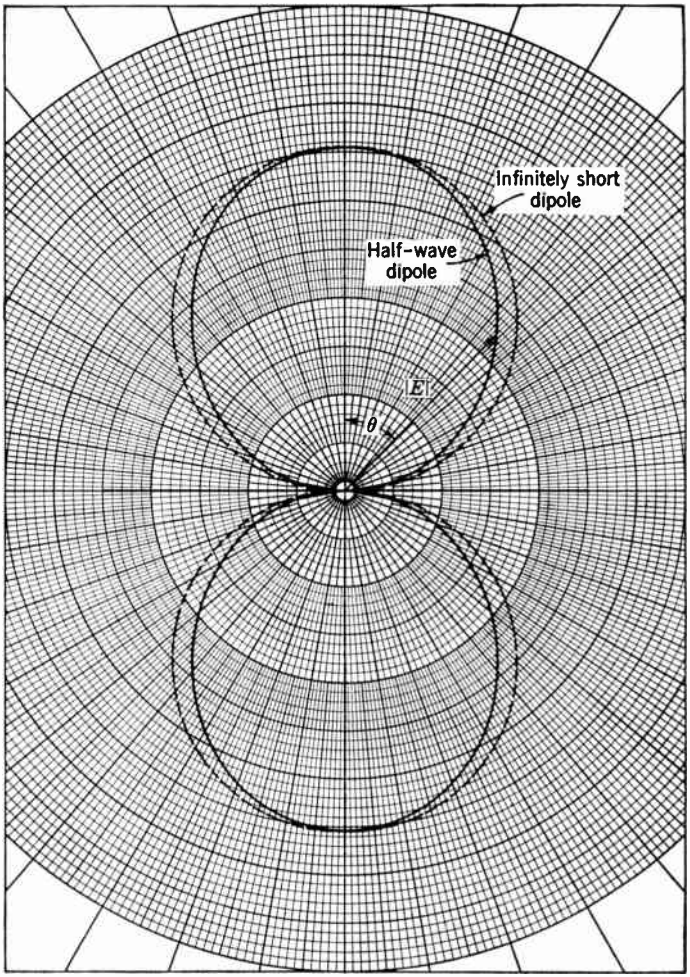


(b) **Equivalent circuit**

Fig. 15.58. Properties of Half-Wave Dipole. (*a*) Current and voltage distribution. (*b*) Equivalent circuit.

This is one consideration affecting antenna design. A second major consideration is the concentration of the radiated power in directions in which receivers to be served are located. In general, this signifies simply the compression of the radiation pattern into a narrow solid angle about the horizontal plane. The ratio of the power radiated per unit solid angle in the preferred direction to the total power radiated, divided by the full solid angle 4π, defines the "power gain" or "directive gain" of the antenna.

Finally, the antenna design depends on the selection of horizontal or vertical polarization for the transmitted signal. In the United States, horizontal polarization has been prescribed by the Federal Communications Commission, upon the recommendation of the National Television System Committee. It possesses the advantage of simpler receiver antenna design, improved signal-to-noise ratio at the receiver for a given radiated power, and reduced difficulties from reflection. Vertical polarization, which possesses the more apparent than real advantage that, with it, a simple dipole transmitter yields an omnidirectional pattern with an appreciable power gain, is employed in Great Britain. The advantage is only apparent since more complex antenna structures, giving a higher power gain, are in any case desirable. Furthermore, practically uniform radiation intensity in all directions in a

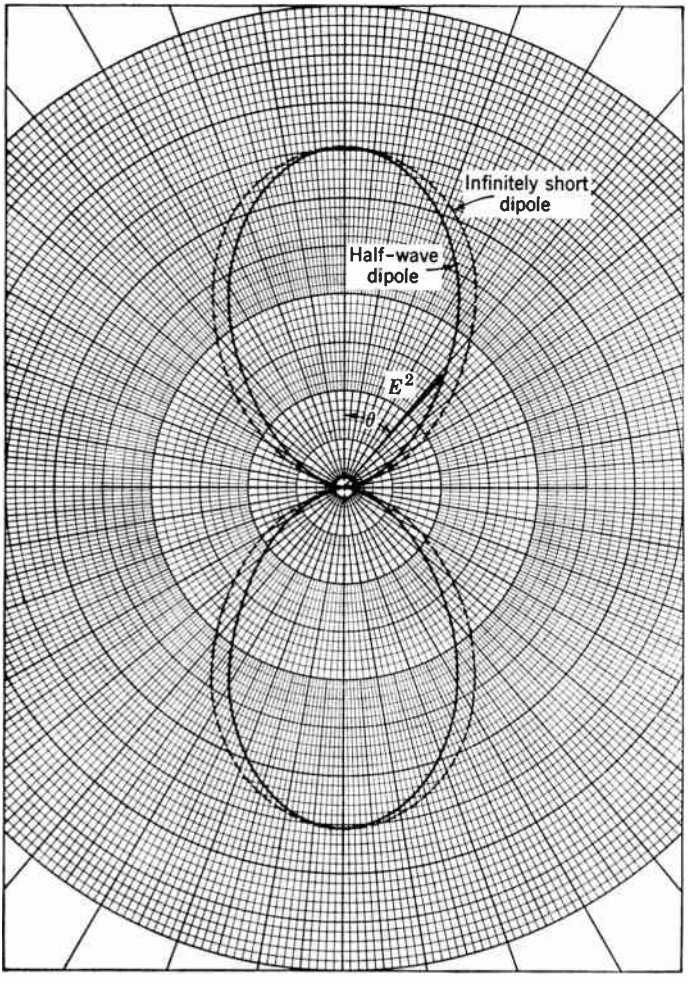


(*c*)

Fig. 15.58 (*continued*). Properties of Half-Wave Dipole. (*c*) Radiation pattern: field strength.

horizontal plane can be readily achieved with horizontal polarization, as will be seen below.

Consider first the feeding of the signal to the antenna. This is generally effected by a single coaxial transmission line. Since the output of the transmitter is double ended or balanced, a balanced-to-unbalanced converter, or balun, is inserted between the coupling loop to the transmitter output and the transmission line leading to the vestigial



(*d*)

Fig. 15.58 (*continued*). Properties of Half-Wave Dipole. (*d*) Radiation pattern: radiated power.

sideband filter. The simplest form of the balun is shown in Fig. 15.59. The quarter-wave sleeve makes the input impedance to ground for the signal applied to the outer conductor of the coaxial line infinite even though, at the output, this conductor is grounded.

The balun shown in Fig. 15.59 is not satisfactory for wide-band operation, since the quarter-wavelength condition can be fulfilled exactly for only one wavelength; for other wavelengths a unilaterally applied

high impedance to ground is added by the circuit. The arrangement in Fig. 15.60 restores balance by applying an equal shunt impedance to both sides of the line.* Additional refinements may be applied to equalize the input impedance of the balun over the entire transmission band.

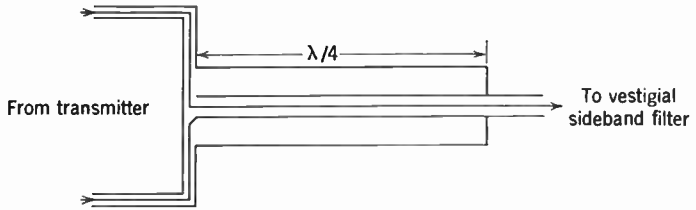


Fig. 15.59. Simplest Balanced-to-Unbalanced Converter. (Balun.)

Circuit elements similar to the balun are also profitably applied at the antenna-feed point. Quarter-wave stubs applied to the terminals of the transmission line insert, in effect, parallel resonant circuits shunting the series resonant circuit representing the antenna; for the two circuits, the change in reactance with deviation from the resonant frequency is in the opposite direction, resulting in compensation over a relatively wide frequency range (Fig. 15.61). The terminations of the stubs are at ground potential and are commonly employed to support the dipole on the mast.

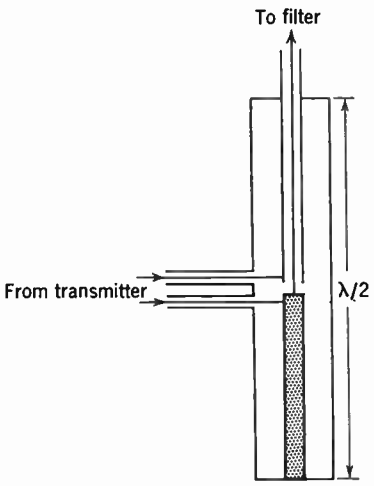


Fig. 15.60. Symmetrical Balun.

Another measure for increasing the frequency range over which the antenna impedance is substantially uniform consists in increasing the thickness of the dipole. A preferred method of doing this is to replace the dipole rods by sheets a half-wavelength wide, fed at the center and grounded at the top and bottom. The current flow is horizontal throughout, so that the sheet is advantageously replaced by a series of horizontal rods, with a resultant reduction in the wind resistance of the antenna. By reducing the rod length at the center, or notching of the flattened dipole, current is concentrated in the top

* See Lindenblad, reference 47.

and bottom rods. The resultant "bat wing" not only has a bandwidth of the order of a quarter of the median operating frequency, but, in addition, a directive gain which is approximately twice that of the simple dipole.

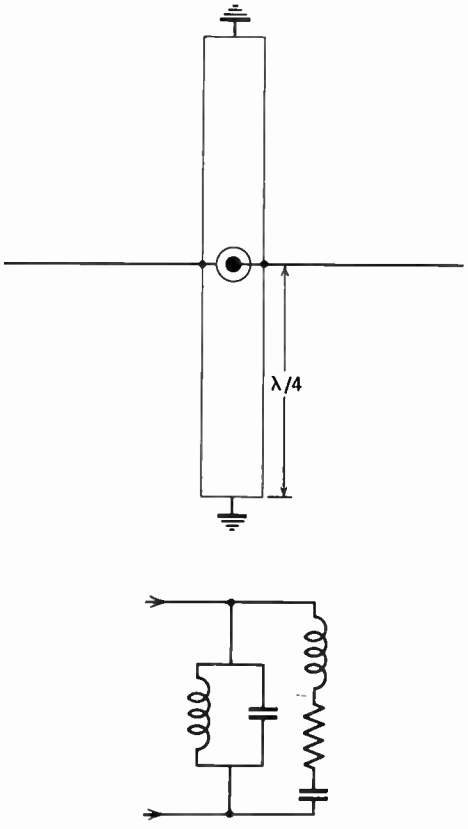


Fig. 15.61. Broad-Band Connection for Dipole Antenna and Equivalent Circuit.

A simple dipole or batwing antenna is generally unsatisfactory, however, since its radiation vanishes in the direction of the dipole axis. It is usually required that the radiated intensity be approximately uniform in a horizontal plane, i.e., the antenna be "omnidirectional." An omnidirectional characteristic is obtained by feeding two dipole antennas, placed at right angles to each other, in quadrature. With such a "turnstile antenna" (Fig. 15.62),* the phase of the radiated

* See Brown, reference 48.

wave varies with horizontal direction, but the radiated power becomes practically independent of azimuth.

The further concentration of radiated power in a narrow solid angle about the horizontal plane which is required is attained by vertical stacking of batwing turnstiles, resulting in a "superturnstile antenna" (Fig. 15.63). The individual turnstile centers or "bays" are separated by a wavelength and are fed, accordingly, in phase. Thus, construc-

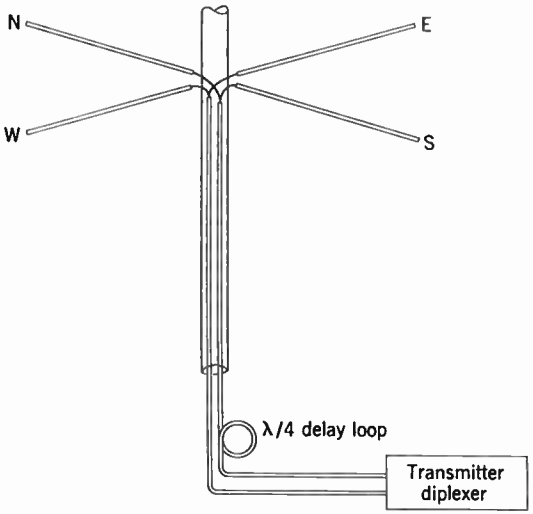


Fig. 15.62. Turnstile Antenna.

tive interference between the radiated waves takes place in a horizontal direction, the sharpness of the intensity peak increasing with the number of bays. The directive gain, compared with a simple dipole, is given by 1.2 times the number of bays, which ranges from 3 to 12.*

Although the superturnstile antenna is probably the most popular form for the very-high-frequency range, numerous other methods of broadbanding and stacking dipoles have been developed.† The supergain antenna ‡ is particularly advantageous whenever, for reasons of mechanical strength and ready accessibility, a steel tower mounting is preferable to a pole mounting. Here the elementary radiators are half-wave dipoles mounted on the four sides of the tower, 0.3 wavelength in front of a conducting shield. The width of the tower is made

* See Wolf, reference 49.

† See, e.g., Gihring, reference 50.

‡ See Wolf, reference 51.

one-half wavelength; in this manner, an adequately circular radiation pattern is achieved without appreciable interference between adjoining radiators. The vertical spacing of the radiators is approximately one wavelength. All the radiators in the East-West and North-South arrays, respectively, are fed from a common junction box by transmission lines of equal length.

An individual radiator is shown in Fig. 15.64. The tapering of the dipole, in conjunction with its spacing from the reflector, is chosen to yield optimum broad-band properties. One of the supporting legs acts as a concentric feed line for the dipole, a second contains a series stub for balancing out the reactive component of the antenna impedance, and the remaining two contain heaters for melting sleet. Shorting bars across the legs at a quarter wavelength from the feed point nullify the electrical effect of the supports on the dipole.

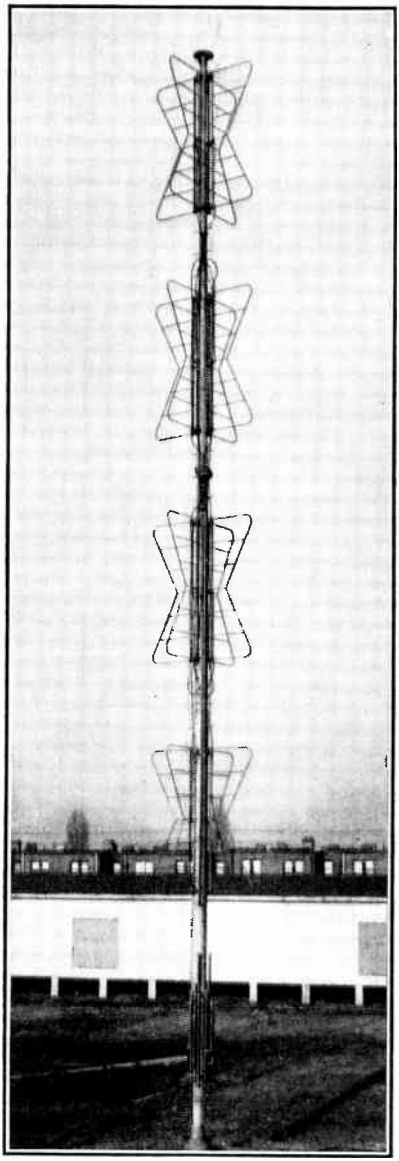


Fig. 15.63. Superturnstile Antenna.

Whereas the broad-band characteristics of the supergain antenna are inferior to those of the batwing superturnstile antenna, its tower mounting adapts it particularly well for systems with high directional gain. For a circular radiation pattern, the vertical angle a between directions for which the radiated power is reduced to half its maximum value is related to the directional gain G by

$$a = 61/G$$

For satisfactory reception, the gain G must be chosen low enough that, under usual wind conditions, $a/2$ remains larger than any tilt angle of the antenna. Since tower manufacturers regard it as possible to hold the deflection normally within $\pm \frac{1}{2}°$, directional gains of 20 or more can safely be planned with the supergain antenna. An additional gain can be achieved by omitting half or even three of the vertical arrays and thus concentrating the radiation in particular directions in which the major service areas happen to be concentrated.

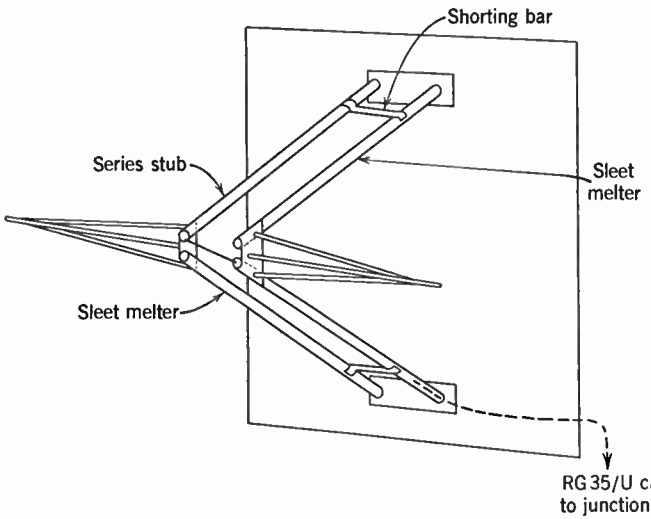


Fig. 15.64. Single Radiator of Supergain Antenna. (Gihring, reference 57.)

The vertical distribution of the radiation pattern can be influenced, furthermore, by feeding the upper half and the lower half of the antenna through separate junction boxes and inserting a phase adjustment in the line feeding the upper or lower section. By introducing a small phase shift at this point, the major radiation lobe can be given a small downward tilt, improving reception close to the antenna. A complete feeder system for a twelve-bay supergain antenna is shown in Fig. 15.65.

In addition to the elements already described, Fig. 15.65 shows a "bridge power equalizer" with a terminating resistor. The function of this component is to render the power delivered to the East-West and North-South antennas equal, even when their impedance does not match the line. This is important in the supergain antenna, since the dipole impedance varies appreciably over the range of a television

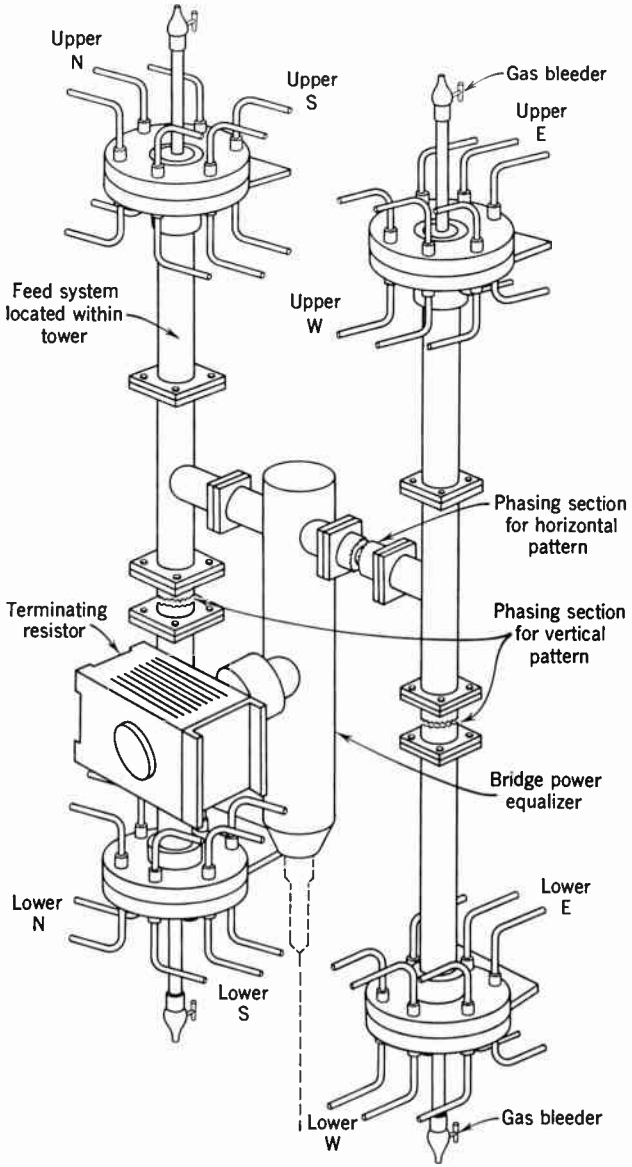


Fig. 15.65. Feed System for Supergain Antenna. (Wolf, reference 51.)

channel, particularly for the low very-high-frequency bands. Suppose, e.g., that for a particular frequency the North-South antenna impedance as seen by the generator is only 85 percent of the line impedance. Then, as the result of impedance transformation by the extra quarter wavelength of transmission line in the feeder of the East-West antenna, the latter will present an impedance which is 118 percent of the line impedance. Thus, the ratio of the power fed to the two antenna systems becomes, for the wavelength in question, $(1.18/85)^2 \cong 2$. This ratio of power would clearly result in a very material distortion of the radiation pattern. In addition, the reflected radiation returning to the antenna with a time delay can give rise to objectionable echo patterns in the picture. To prevent the appearance of secondary images from this cause, the effective reflection coefficient of the antennas should be held below 0.05 within the range of the video channel.

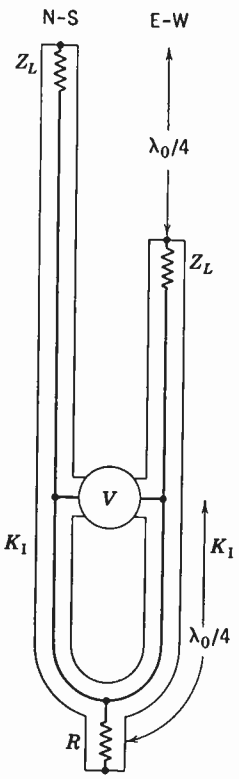


Fig. 15.66. Principle of Bridge Power Equalizer. (Masters, reference 52.)

The principle of the bridge power equalizer, which overcomes these difficulties, is illustrated by Fig. 15.66.* The impedances Z_L represent the antenna loads, λ_0 the midband wavelength. V is the generator which, it will be assumed, feeds the two branches in a push-pull mode. The two quarter-wavelength lines K_1, with the terminating resistance R equal to half the line impedance of K_1 constitute the added power equalization network. Midband frequency waves which travel from the generator to the antennas and are partly reflected return to the generator in phase as the result of the half-wave delay, ignore it accordingly, and pass on to the terminating resistance R, where they are absorbed. Waves traveling from the generator directly to the terminating resistance R, on the other hand, are not affected by its presence, since they arrive at it 180 degrees out of phase. The reflections, hence, do not influence the operation of the generator, which now sees equal impedances (namely, the line impedance) in both branches, and, furthermore, give no rise to secondary images since they are absorbed in R. A more exact analysis shows

* See Masters, reference 52.

that these properties are retained to a close approximation throughout the video band.

In the ultra-high-frequency range, the reduced dimension of the wavelength makes a different arrangement of radiators desirable. In the slotted-coaxial-line or pylon antenna * the batwings effectively form part of the outer conductor of a coaxial line, leaving only a slot a half-wavelength long in the external sheath of the coaxial line. In

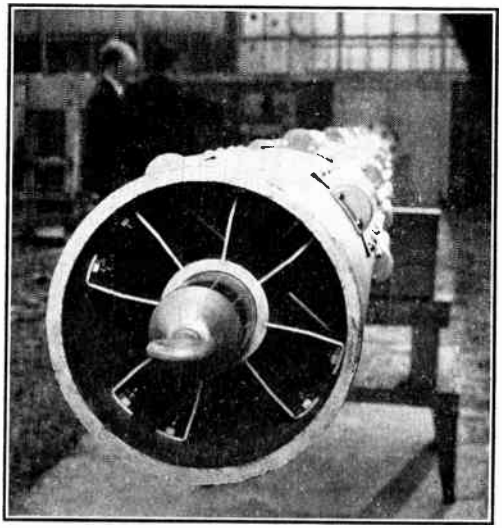


Fig. 15.67. End View of Pylon Antenna.

the antenna employed at the Bridgeport experimental ultra-high-frequency station operating at 529 to 535 megacycles, the outer conductor was a steel cylinder with an outer diameter of 10¾ inches and wall thickness of one-half inch. It provides the mechanical support of the antenna.

The slots are arranged symmetrically about the antenna in groups of four and are spaced vertically by a wavelength, so that all radiate in phase. Probes projecting from the inner conductor to the slot edges determine the amount of radiation from the slot (Fig. 15.67). The probe adjustment is graded so that the intensity of radiation is uniform along the entire length of the antenna.

The antenna is fed from a junction point at its center, the lower half of the inner conductor constituting at the same time the outer

* See Fiet, reference 53. For a discussion of a number of alternative designs, see Epstein, Peterson, and Woodward, reference 54.

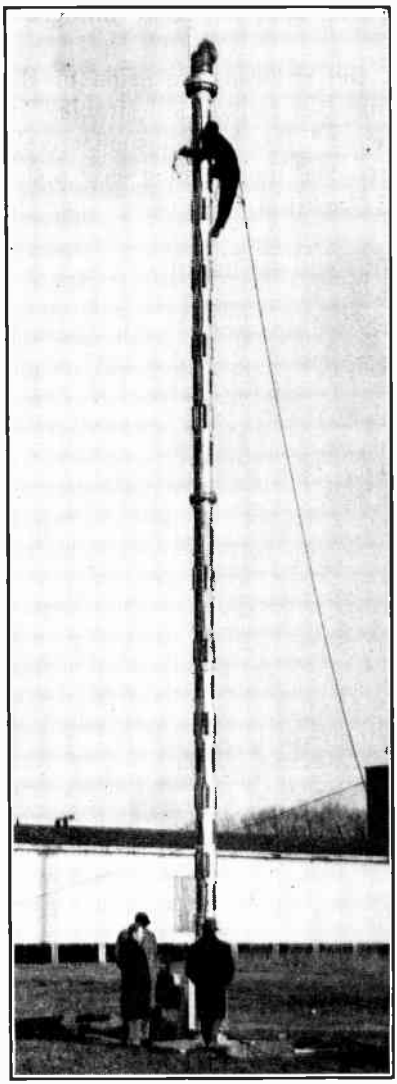


Fig. 15.68. Pylon Antenna Designed for Operation in the 529–535 Mc Television Channel.

conductor of the transmission line from the transmitter. With asymmetrical feed, e.g., from the bottom of the antenna, there would be a tendency for radiation whose wavelength did not coincide exactly with the vertical slot spacing to be concentrated in cones deviating upward and downward from the horizontal direction.

The angular displacement of successive layers of slots by 45 degrees is found essential to suppress undesired modes of propagation within the antenna which may result in a horizontally asymmetrical radiation pattern. The number of layers of slots determines the gain of the antenna, or the concentration of radiation in a horizontal direction. The Bridgeport antenna had 22 such layers, leading to a gain of 17.3 (Fig. 15.68). Its height was 40 feet. The angular spread of the radiation about the horizontal plane was about 5 degrees, which represents a practical minimum if the tilting of the antenna by wind pressure is not to result in excessive signal reduction in some service areas.

A still different style of antenna is required for the super-high-frequency or microwave radiations employed for relaying television signals. Here, the primary objective is the concentration of the radiation into a very narrow beam, directed toward a specific receiver. This result is achieved by placing the radiator, which may be a simple dipole, at the focus of a parabolic

reflector (Fig. 15.69) or "microwave lens," whose diameter is large compared with the wavelength of the radiation.* Diffraction theory indicates that the directional gain achieved with such an arrangement is of the order $k(\pi D/\lambda)^2$, where k is the fraction of the radiation intercepted by the reflector or lens, D is its diameter, and λ is the wavelength of the radiation. Clearly, if the reflector diameter is made very large compared to the wavelength, the directional gain becomes very large, and the power requirements of the microwave transmitter are correspondingly reduced.

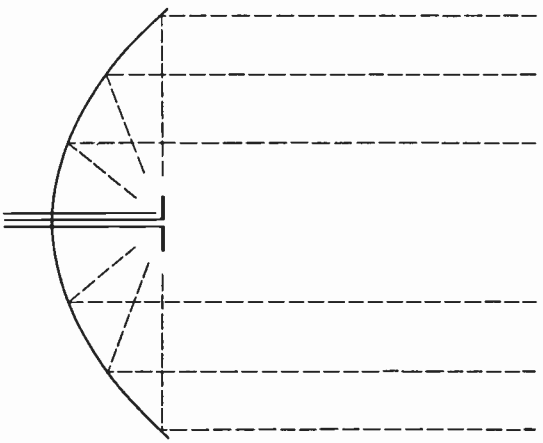


Fig. 15.69. Microwave Radiator with Parabolic Reflector.

Another type of radiator widely employed in microwave practice is the horn antenna, or open termination of a wave guide. The radiation pattern of a horn antenna can be greatly modified by the insertion of shaped rods of low-loss dielectric. Effective directional microwave antennas are obtained by assembling arrays of dielectric rod radiators.†

At this point, it appears proper to make some mention of the sound transmitter, which, so far, has been completely disregarded. The sound transmitter is a frequency modulated transmitter with power comparable to that of the visual transmitter and operating with a carrier 4.5 megacycles above the visual carrier. A separate antenna may be provided for the sound signal. It is entirely practicable, however, and

* See, e.g., Schelkunoff and Friis, reference 7, sections 18 and 19.

† See Mueller and Tyrrell, reference 55.

generally preferred to transmit picture and sound from the same antenna. This transmission requires a "diplexer," i.e., a circuit which applies both the visual and the sound signals to the antenna load with-

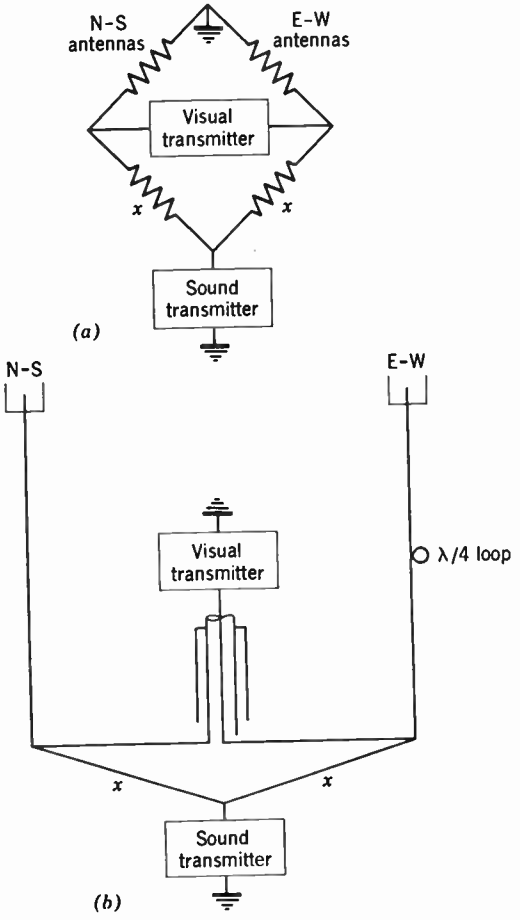


Fig. 15.70. Bridge Diplexer. (Wolf, reference 49.) (*a*) Principle. (*b*) Physical realization. (Lines represent inner conductor of coaxial lines.)

out any reaction of the visual and sound signal generators (transmitters) on each other. As was pointed out in connection with the multiple operation of power tubes, the balanced bridge provides just such a circuit. Figure 15.70 * indicates its use in applying visual and sound

* See Wolf, reference 49.

signals to a turnstile antenna. The visual transmitter, placed in one cross-arm of the bridge, provides push-pull excitation for the North-South and East-West radiators, respectively, whereas the sound transmitter, placed in the other cross-arm, feeds them in parallel. An extra quarter-wave section of transmission line is inserted in the feeder for the East-West radiators to effect the desired 90-degree phase difference between the two sets.

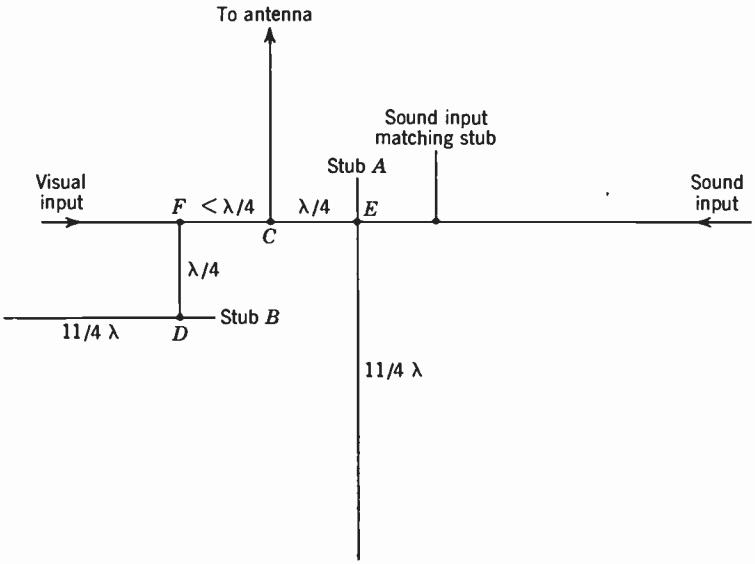


Fig. 15.71. Notch Diplexer. (Fiet, reference 53.) (λ is wavelength of visual carrier; all stubs open circuit.)

A quite different diplexer, suitable for an antenna with a single feed line, is shown in Fig. 15.71.* It is known as a notch diplexer and contains two filters. The filter between the sideband filter and the antenna junction C presents an open circuit for the picture carrier and is tuned, by adjustment of the stubs B and A, for maximum rejection of the sound carrier at the picture input. Conversely, the filter between the sound transmission line and the antenna junction consists of an 11/4 wavelength line for the picture carrier, and, hence, has the effect of rejecting the picture signals at the sound input. The length of the line between junctions F and C is adjusted to obtain a good impedance match at picture frequencies just below the sound carrier, resulting in

*See Fiet, reference 53.

substantially flat response almost up to the sound channel. The notch diplexer does not give as broad-band a response as the bridge diplexer, which is, hence, generally given preference in the very-high-frequency channels.*

It is also possible and, in fact, common practice to utilize a television antenna for a third channel, carrying programs of a frequency-modulated broadcasting station. The mixing circuit employed for this purpose is known as a triplexer.

15.10 The Empire State Television Antenna System. The application of a number of the principles of antenna construction discussed in the last section may fittingly be illustrated by the antenna system erected on the tower of New York's Empire State Building.† Since the joint utilization of any optimal site by a number of stations may prove desirable in many other localities, the experience gained in New York may prove of general value.

The problem to be solved was the accommodation of five television stations operating on Channels 2 (54 to 60 megacycles), 4 (66 to 72 megacycles), 5 (76 to 82 megacycles), 7 (174 to 180 megacycles), and 11 (198 to 204 megacycles) and of three frequency-modulated stations with carriers at 95.5, 97.1, and 101.1 megacycles on a structure with an approximate total height of 200 feet. The structure was to be mounted on the top of the Empire State Building, over 1200 feet above street level. The available height was to be utilized for maximum directional gain consistent with the absence of cross-talk between the antennas. The possibility of diplexing or interleaving television antennas was rejected since each station wished to retain control of the operation of its antenna. These techniques were used, however, to accommodate the frequency-modulated stations. In view of the limited horizontal area available, vertical stacking of the television antennas proved the only practical solution.

The distribution of the available height to the several stations was made in such a way as to equalize the directional gain; 20 percent higher gain was allowed, however, to the two high-frequency stations. In the final structure, the gain for Channels 2, 4, and 5 was approximately 4, that for Channels 7 and 11 was 5. Tests showed that the coupling between adjacent antennas was always less than -30 decibels, even when no spacers were inserted between them. Past experience has indicated that cross-talk is negligible even with -20 decibels.

* See Sayer and deBell, reference 56.

† See Gihring, reference 57.

For ready accessibility and convenience in accommodating feeder lines and other equipment, as well as for reasons of mechanical stability, tower construction was desirable for all except the topmost antenna. Accordingly, supergain antennas were chosen for four of the stations (Channels 2, 5, 7 and 11) and a superturnstile for Channel 4. One of the frequency-modulated stations (97.1 megacycles) was triplexed into the superturnstile. The other two utilized a ring of dipoles interleaved between the two bottom bays of the Channel 2 antenna. The complete antenna structure is shown in Fig. 15.72.

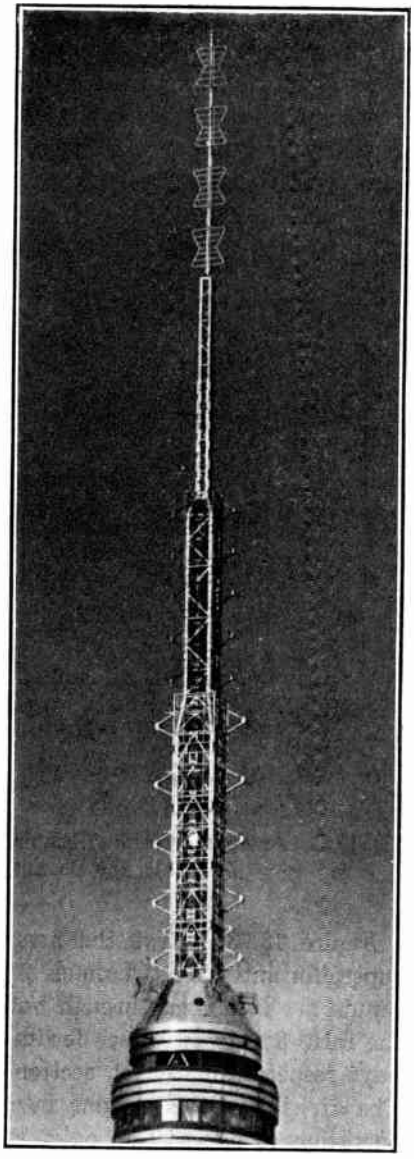


Fig. 15.72. The Empire State Television Antenna System. (Gihring, reference 57.)

Different methods are employed for feeding the signal to the several antennas (Fig. 15.73). A simple bridge diplexer proves adequate for the high-frequency channels (7 and 11) since here the frequency range, in relative measure, is small. The broadband properties of the superturnstile permit a similar system for Channel 4; however, the signal for the FM transmitter is added as indicated in Fig. 15.73*b*. Four double stubs mounted at a quarter-wavelength video carrier from the main antenna line and a quarter FM carrier wavelength apart prevent cross-feed from the FM transmitter to the television transmitter, and vice versa. The lower part of each stub is one-half television wavelength, so that, in conjunc-

tion with the quarter-wave section between it and the antenna line, it presents a high impedance to the television signal. The upper portions of the stubs are adjusted so that the double stubs in the FM line present a high shunt impedance to the FM carrier (and a low impedance to the television frequencies), whereas the remaining double stubs on their quarter-wave lines present a low shunt impedance to the FM carrier.

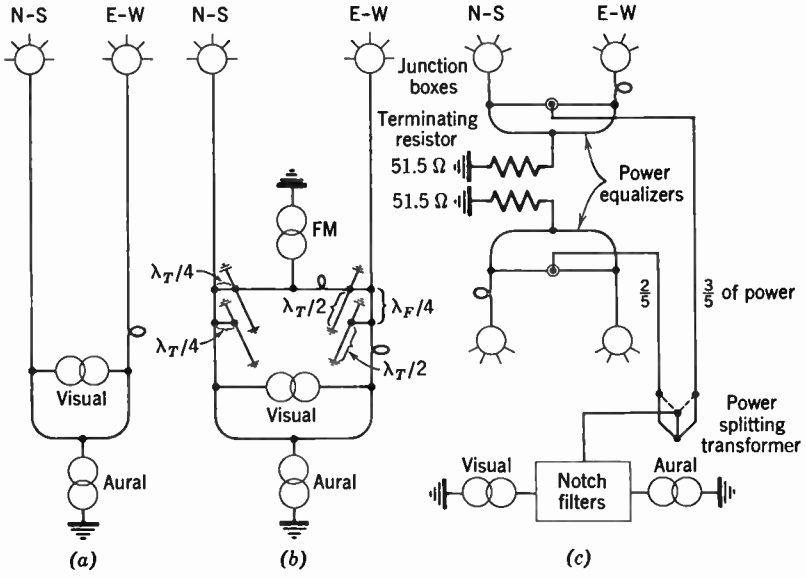


Fig. 15.73. Antenna Circuits of Empire State Television Antennas. (*a*) Channels 7 and 11. (*b*) Channel 4. (*c*) Channels 2 and 5.

Figure 15.73*c* shows the arrangement for the two low-frequency supergain antennas (Channels 2 and 5). Here, the visual and aural inputs are combined through notch diplexers and then distributed in the ratio 3:2 to two lines feeding the upper three and the lower two bays respectively. Each section is provided with a power equalizer. The division of the antenna into two sections permits the temporary operation of one section only, while changes are made in the other, and, in addition, the tilting of the radiation pattern by introducing a phase shift in one of the feeder lines.

The power equalizer was found to be essential. Measurements on the Channel 5 antenna indicated a maximum voltage standing wave

ratio (VSWR) at the visual generator equal to 1.45; with the power equalizer, this was reduced to less than the permitted value 1.1.*

The effective radiated power (ERP) handling capacity prescribed for the antennas was 100 kilowatts; since the directional gain is relatively low (4 or 5), this demanded generously dimensioned junction boxes and feed cables, suitable for average transmitter output of 25 kilowatts.

15.11 Propagation at Very High and Ultra-High Frequencies. Energy leaves the antenna in the form of electromagnetic radiation. This radiation is polarized with its electric vector lying in a plane including the direction of current flow in the antenna. Therefore, the electromagnetic waves propagated from a horizontal dipole are horizontally polarized.†

The field strength E at a distance d from a half-wave dipole in a plane perpendicular to it through its center is given by

$$E = 7\sqrt{W}/d \tag{15.28}$$

Here W is the power radiated in watts, and d and E are measured in meters and volts per meter, respectively. This relation is true only for a radiator in free space, in other words, when no reflected radiation is present. The field strength in other directions varies approximately with the cosine of the angle between a plane normal to the dipole and the line joining the center of the antenna with the point at which the electric vector is measured. This fact has already been mentioned in connection with the radiation pattern from such a radiator in the preceding section.

When the antenna or the point of measurement is close to a ground, this relation is no longer true because of the interference effect between the direct radiation and that reflected by the earth's surface. An approximate expression for the effect of this interference can easily be found. Figure 15.74 represents a transmitting and receiving antenna

* The VSWR is defined as the ratio of the maximum and minimum amplitude of the signal wave along the line. If q is the reflection coefficient, it is given by

$$\frac{1 + |q|}{1 - |q|}, \text{ with } q = \frac{Z_L - Z_c}{Z_L + Z_c}$$

where Z_c is the characteristic impedance of the line and Z_L the effective terminating impedance. The fraction of the power which is reflected is $|q|^2$.

† The convention adopted here is opposite to that customary in optics; the plane of polarization of a light wave is perpendicular to its electric vector.

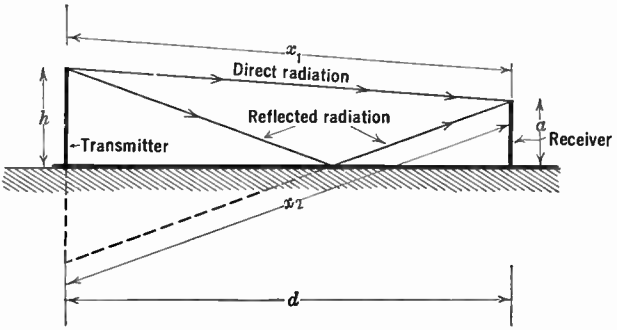


Fig. 15.74. Determination of Signal Intensity at Receiver.

separated by a distance d. The heights of these antennas are h and a, respectively, these heights being small compared with the distance d. The instantaneous field strength at a, due to direct radiation, is

$$e_1 = \frac{7\sqrt{W}}{x_1} \cos\left[2\pi\left(ft - \frac{x_1}{\lambda}\right)\right] \tag{15.29}$$

where x_1 is the path distance. The reflected radiation reaches the antenna over a slightly longer path x_2, and, furthermore, suffers a phase reversal of 180 degrees upon reflection at the earth's surface. Therefore, the field strength due to it is

$$e_2 = -\frac{7\sqrt{W}}{x_2} \cos\left[2\pi\left(ft - \frac{x_2}{\lambda}\right)\right] \tag{15.29a}$$

The resultant field, which is the vector sum of e_1 and e_2, is to a close approximation

$$e = -\frac{7\sqrt{W}}{d} 2 \sin 2\pi\left(ft - \frac{d}{\lambda}\right) \sin \pi\left(\frac{x_2 - x_1}{\lambda}\right)$$

$$= -\frac{7\sqrt{W}}{d} 2\pi \frac{x_2 - x_1}{\lambda} \sin 2\pi\left(ft - \frac{d}{\lambda}\right) \tag{15.30}$$

Since the difference $x_2 - x_1$ is given by

$$\sqrt{d^2 + (h + a)^2} - \sqrt{d^2 + (h - a)^2} \cong \frac{2ah}{d}$$

the amplitude of the field strength at the receiving antenna is

$$E = \frac{28\pi\sqrt{W}ah}{d^2\lambda} \tag{15.31}$$

If a and h are measured in feet, d in miles, E in millivolts per meter, and λ in meters, this becomes

$$E = \frac{0.0032ah\sqrt{W}}{d^2\lambda} \tag{15.31a}$$

It is interesting to note that the field strength varies inversely with the square of the distance separating the antennas, and, furthermore, that it increases with increasing frequency, subject to the implicit condition $2\pi ah/(d\lambda) < 1$. The indicated variation is found to hold very accurately over distances which are within the horizon and over terrain which is free from obstructions. In urban districts, absorption and reflection of the radiation are found to cause marked departures from the values given by the formula. In general, the field strengths for radiation in the neighborhood of 50 megacycles will be found to be 30 to 60 percent lower than the values given by Eq. 15.31a.

In the ultra-high-frequency range, the deviation of the observed field strengths from their theoretical values is much greater. Extensive surveys carried out in the New York * and the Washington † areas are summarized in Fig. 15.75. The deviation proved to be greater in the Washington tests, presumably in part because of the more effective isolation of the New York antenna on the Empire State Building, in part as a result of the less favorable location of the majority of the receiving stations in Washington. In Fig. 15.76, the same data are replotted so as to indicate the deviation of the measured field strengths from the theoretical curve for 67.25 megacycles. This set of curves makes it possible to read off directly the change in the radiated power requirements accompanying a change in carrier frequency. It is clearly impractical to achieve, in the ultra-high-frequency band, as complete coverage of receiver locations as is possible in the very-high-frequency range. An exhaustive field test carried out in the Bridgeport area with station KC2XAK operating on 530.25 megacycles has led to results similar to those obtained in the Washington tests.‡

* See Brown, Epstein, and Peterson, reference 62.
† See Brown, reference 63.
‡ See Guy, reference 64.

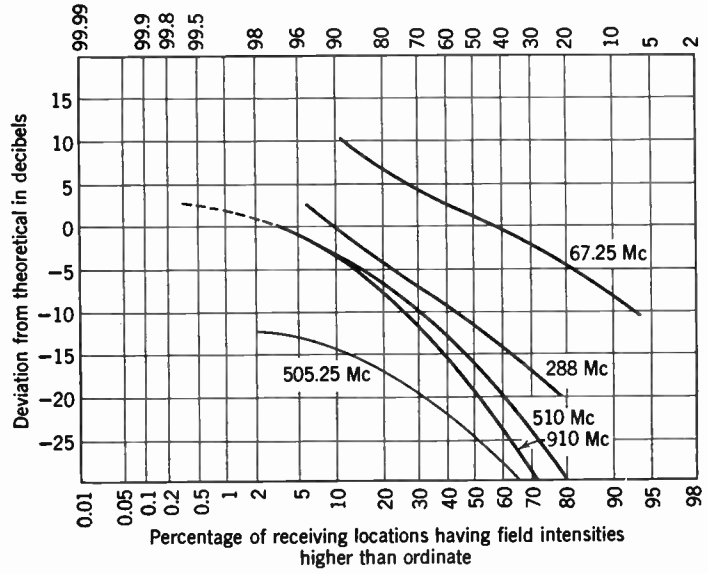


Fig. 15.75. Average Deviation of Field Strength from Theoretically Expected Values at 67.25, 288, 510, and 910 Mc (New York) and at 505.25 Mc (Washington).

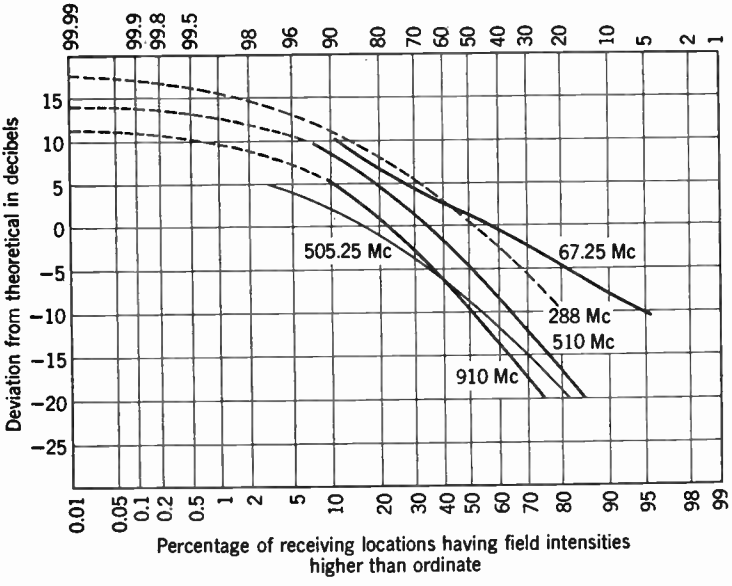


Fig. 15.76. Average Deviation of Field Strength at 288, 510, 910, and 505.25 Mc from That Calculated for 67.25 Mc.

In compensation, ultra-high-frequency reception is less subject to man-made interference than very-high-frequency reception. Furthermore, the possibility of attaining high directional gains with transmitter antennas of reasonable dimensions invites the use of radiation patterns closely fitted to the service area. This is possible by appropriate phasing and spacing of the radiators. Unidirectional antennas may frequently be found more economical to install and quite as effective as the omnidirectional types preferred in the very-high-frequency range.

Beyond the horizon, the field strength decreases more rapidly than the inverse square, owing to the quasi-optical properties of the radiation. At 40 megacycles, the field strength decreases with $1/d^{3.6}$; at 100 megacycles it is given by approximately the inverse fifth power; and at 400 megacycles by the inverse ninth power. As a first approximation, the only radiation reaching points beyond the line of sight is that diffracted by the earth's surface at the horizon. Calculated results obtained on this basis agree fairly well with numerous experimental measurements. The increased tendency of the shorter-wavelength radiations to cast sharp shadows of obstacles is, of course, also responsible for reduced effectiveness in reaching many receiver locations in urban areas.

The effective horizon, determined by the height of the transmitting antenna, is given by

$$l = 1.22\sqrt{h}$$

where h is the antenna height in feet and l is the distance to the horizon in miles. Usually, the value obtained from this expression is taken as the practical service radius of a given transmitter, although reliable reception is often obtained at a considerably greater distance.

In fact, the above formula for the horizon distance takes no account of refraction in the earth's atmosphere. The density of the atmosphere decreases upwards, and the velocity of propagation of electromagnetic waves increases correspondingly. This increase in velocity causes a deflection of the wave normals toward ground. Under normal atmospheric conditions, its effect is such as to increase the effective horizon distance from $1.22\sqrt{h}$ miles to $1.41\sqrt{h}$ miles.* In exceptional weather, the horizon distance may become much larger than this, an important factor in assessing interference effects between stations operating on the same frequency channel. On the other hand, reflection of very-

* See Fink, reference 10.

high-frequency radiations by the ionosphere are very rare and occur only for the longest-wavelength channels. The reception in the United States of pictures transmitted from London is on record;* however, the received pictures are invariably badly distorted as the result of multiple-path propagation.

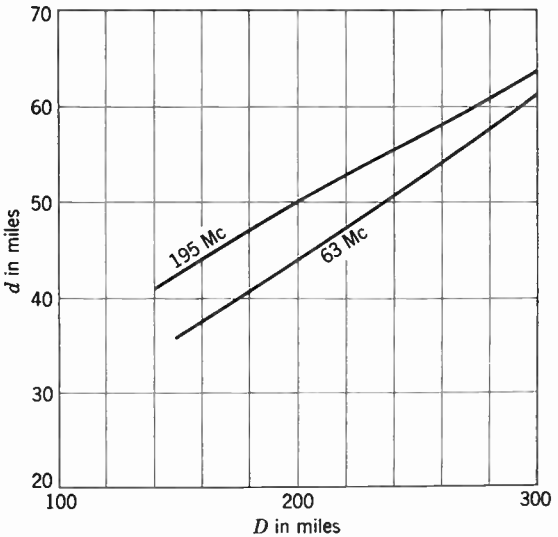


Fig. 15.77. Distance from Transmitter for Which Signal Strength Exceeds Signal from Interfering Station by 30 Db, as Function of the Distance of the Interfering Station, for a Common Carrier Frequency of 63 and 195 Mc, Respectively. (Reference 66.)

Since the service areas are cut off beyond the horizon distance with increasing sharpness as the carrier frequency is increased, interference problems are more serious in the very-high-frequency channels than in the ultra-high-frequency channels. To minimize interference effect between stations operating in the same television channel in fringe areas, it is customary to offset the carriers of such stations relative to each other.† The optimum frequency difference would be half the line frequency (7.875 kilocycles), since in this manner alternate lines of the interfering picture would be superposed in reverse polarity. By making the frequency offset ± 10.5 kilocycles instead, interference may be minimized for three neighboring co-channel stations. With this

* See Goddard, reference 65.

† See RCA Laboratories Division, reference 66.

arrangement, interfering signals may exceed by 15 decibels those which can just be tolerated with nonsynchronous operation of the receivers; in the latter instance, the interference takes the form of broad drifting horizontal bars.

Viewer tests established that a signal difference of approximately 30 decibels was required with offset carrier operation to render interference between two co-channel stations unobjectionable. Figure 15.77 shows the distance d from the desired transmitter for which this signal difference is obtained 90 percent of the time for 50 percent of the locations as function of the distance D between transmitters.

15.12 Special Features of an Ultra-High-Frequency Transmitter. The problems of ultra-high-frequency transmission have been touched upon at several points in the preceding sections. It remains to indicate how a very-high-frequency transmitter may be modified to serve in the ultra-high-frequency range. The TTU-1A Transmitter,* which has been operated for several years in the 529 to 535 megacycle band at Bridgeport, will serve as an example.

The TTU-1A transmitter provides 1 kilowatt of ultra-high-frequency output power and is based on the 500-watt TT-500B very-high-frequency transmitter. The principal new elements consist of a tripler stage added to the carrier generator, a new power amplifier stage, and a cathode-follower modulator direct-coupled through a suitable bucking bias to the video amplifier of the very-high-frequency transmitter. The tripler and the power amplifier stage are physically similar. Both employ eight 4X150A tubes working into a single cavity. Figure 15.78 shows a cross-section diagram of these two stages joined by a coaxial line variable impedance transformer. The tubes are mounted on parallel anode, screen, cathode, and grid plates; the output resonant cavity is formed between the anode and screen plates, the input cavity between the grid and cathode plates. Both screen and cathode plates are radio-frequency grounded, and the field-free space between them is utilized for the distribution of the 60-cycle and d-c potentials to the tube electrodes. Silvered mica sheets placed on the screen-grid and anode plates serve as blocking condensers.

At full power output, the output impedance of the power amplifier is one-half ohm, that of the tripler one-fourth ohm. Since the transmitter works into a 51.5-ohm line and the input impedance of the power amplifier is 100 ohms, matching transformers must be provided for the output of both stages. They consist of two quarter-wave lines

* See Lappin and Bennett, reference 67.

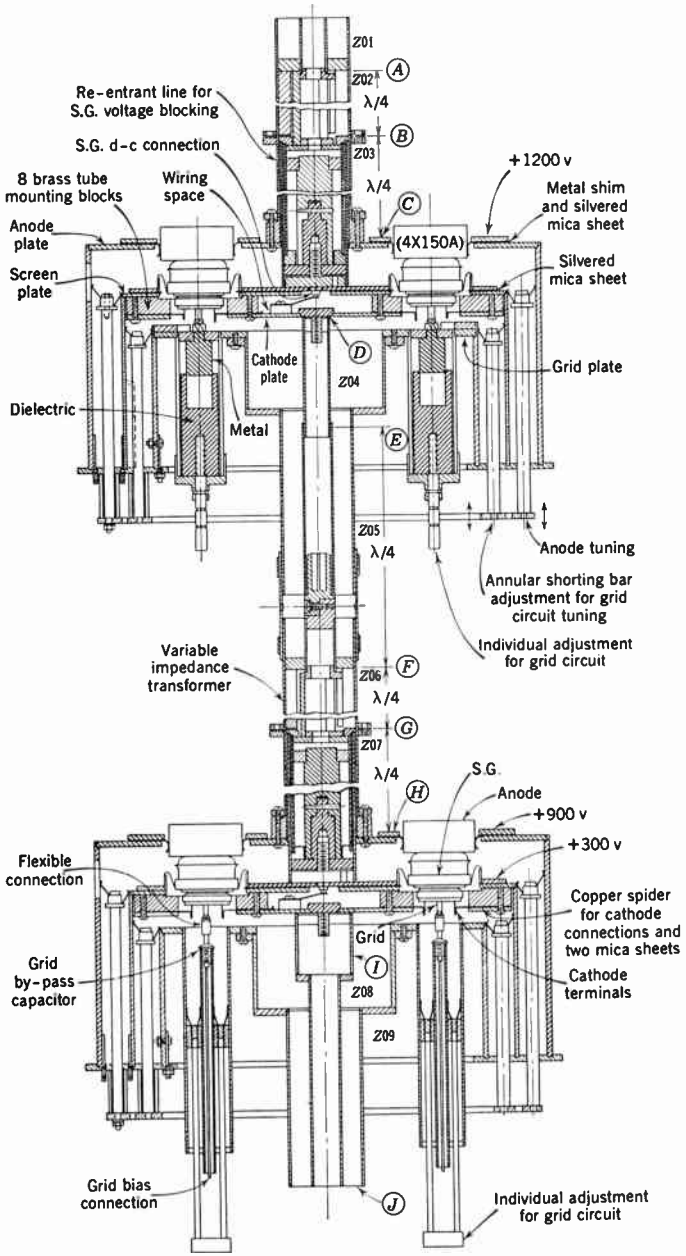


Fig. 15.78. Cross-Section Diagram of Tripler and Power Amplifier Cavities in Ultra-High-Frequency Transmitter RCA TTU-1A.

in series, with inner and outer conductors dimensioned to yield a surge impedance which is the geometric mean of the input impedance and the desired output impedance. In practice, the quarter-wave lines are made asymmetric, so that the transformer ratio may be adjusted by a rotation of the outer and inner conductors relative to each other.

The modulator consists of eight parallel-excited 6L6 tubes in cathode-follower circuits, each controlling the grid voltage of a single power amplifier tube. This individual connection facilitates the balancing of the static plate current in the power amplifier tubes. Furthermore, the cathode follower connection minimizes the required plate current in the modulator tubes and reduces their effective input capacitance, owing to degeneration.

REFERENCES

1. F. E. Terman, *Radio Engineering,* 3rd Ed., McGraw-Hill, New York, 1947.
2. G. E. Anner, *Elements of Television Systems,* Prentice-Hall, New York, 1951.
3. D. G. Fink, *Television Engineering,* 2nd Ed., McGraw-Hill, New York, 1952.
4. H. Pender and K. McIlwain, *Electrical Engineers' Handbook—Electric Communication and Electronics,* 4th Ed., Wiley, New York, 1950.
5. F. E. Terman, *Radio Engineer's Handbook,* McGraw-Hill, New York, 1943.
6. P. Vigoureux and C. F. Booth, *Quartz Vibrators and Their Applications,* His Majesty's Stationery Office, London, 1950.
7. S. A. Schelkunoff and H. T. Friis, *Antennas: Theory and Practice,* Wiley, New York, 1952.
8. E. A. Laport, *Radio Antenna Engineering,* McGraw-Hill, New York, 1952.
9. J. C. Slater, *Microwave Transmission,* McGraw-Hill, New York, 1942.
10. D. G. Fink, *Radar Engineering,* McGraw-Hill, New York, 1947.
11. D. G. Fink (editor), *Television Standards and Practice,* McGraw-Hill, New York, 1943.
12. D. G. Fink, "Design Trends in Television Transmitters," *Electronics,* Vol. 21, pp. 76–82, January, 1948.
13. C. D. Kentner, "TT-5A Television Transmitter," *Broadcast News,* No. 48, pp. 4–17, March, 1948.
14. P. J. Herbst and E. M. Washburn, "Frequency Stability for Television Offset Carrier Operation," *RCA Rev.,* Vol. 13, pp. 95–106, 1952.
15. C. B. Aiken, "Two-Mesh Tuned Coupled Circuit Filters," *Proc. I.R.E.,* Vol. 25, pp. 230–272, 672, 1937.
16. P. Breen, "New 5-KW Television Output Stage," *Tele-Tech,* Vol. 9, pp. 36–38, September, 1950.
17. W. N. Parker, "A Unique Method of Modulation for High-Fidelity Television Transmitters," *Proc. I.R.E.,* Vol. 26, pp. 946–962, 1938.
18. J. S. Donal, Jr., and R. R. Busch, "A Spiral-Beam Method for the Amplitude Modulation of Magnetrons," *Proc. I.R.E.,* Vol. 37, pp. 375–382, 1949.
19. H. Chireix, "High-Power Outphasing Modulation," *Proc. I.R.E.,* Vol. 23, pp. 1370–1392, 1935.

20. W. E. Evans, Jr., "Phase-to-Amplitude Modulation for UHF-TV Transmitters," *Electronics*, Vol. 23, pp. 102–106, September, 1950.
21. E. E. Spitzer, "Grounded-Grid Power Amplifiers," *Electronics*, Vol. 19, pp. 138–141, April, 1946.
22. P. A. T. Bevan, "Earthed Grid Power Amplifiers," *Wireless Eng.*, Vol. 26, pp. 182–192, 235–242, 1949.
23. P. T. Smith and H. R. Hegbar, "Duplex Tetrode UHF Power Tubes," *Proc. I.R.E.*, Vol. 36, pp. 1348–1353, 1948.
24. E. H. Potter, "Quick Changing of 8D21 Tube," *Broadcast News*, No. 48, pp. 18–19, March, 1948.
25. E. L. Ginston, W. R. Hewlett, J. H. Jasberg, and J. D. Noe, "Distributed Amplification," *Proc. I.R.E.*, Vol. 36, pp. 956–969, 1948.
26. G. H. Brown, W. C. Morrison, W. L. Behrend, and J. G. Reddeck, "Method of Multiple Operation of Transmitter Tubes Particularly Adapted for Television Transmission in the Ultra-High-Frequency Range," *RCA Rev.*, Vol. 10, pp. 161–172, 1949.
27. R. R. Law, D. G. Burnside, R. P. Stone, and W. B. Whalley, "Development of Pulse Triodes and Circuit to Give One Megawatt at 600 Megacycles," *RCA Rev.*, Vol. 7, pp. 253–264, 1946.
28. R. R. Law, W. B. Whalley, and R. P. Stone, "Developmental Television Transmitter for 500–900 Megacycles," *RCA Rev.*, Vol. 9, pp. 643–652, 1948.
29. W. W. Salisbury, "The Resnatron," *Electronics*, Vol. 19, pp. 92–97, February, 1946.
30. J. A. Morton and R. M. Ryder, "Design Factors of the Bell Laboratories 1553 Triode," *Bell System Tech. J.*, Vol. 29, pp. 496–530, 1950.
31. R. H. and S. F. Varian, "A High-Frequency Oscillator and Amplifier," *J. Appl. Phys.*, Vol. 10, pp. 321–327, 1939.
32. D. L. Webster, "Cathode-Ray Bunching," *J. Appl. Phys.*, Vol. 10, pp. 501–508, 1939.
33. H. T. Friis, "Microwave Repeater Research," *Bell System Tech. J.*, Vol. 27, pp. 183–246, 1948.
34. R. Warnecke and P. Guenard, "On the Utility of Some Recent UHF Electron Tube Designs in Television," *Ann. radioélec.*, Vol. 3, pp. 259–280, 1948.
35. R. H. Varian, "Recent Developments in Klystrons," *Electronics*, Vol. 25, pp. 112–115, April, 1952.
36. V. Learned and C. Veronda, "Recent Developments in High-Power Klystron Amplifiers," *Proc. I.R.E.*, Vol. 40, pp. 465–469, 1952.
37. P. Guenard, B. Epsztein, and P. Cahour, "5 Kilowatt Broad-Band Klystron Amplifier," *Ann. radioélec.*, Vol. 6, pp. 109–113, 1951.
38. R. Kompfner, "Traveling Wave Valve—New Amplifier for Centimetric Wavelengths," *Wireless World*, Vol. 52, pp. 369–372, 1946.
39. J. R. Pierce, "Theory of Beam-Type Traveling Wave Tube," *Proc. I.R.E.*, Vol. 35, pp. 111–123, 1947.
40. J. R. Pierce, *Traveling Wave Tubes*, Van Nostrand, New York, 1950.
41. J. Brossart and O. Doehler, "On the Properties of Tubes with Constant Magnetic Field," *Ann. radioélec.*, Vol. 3, pp. 328–338, 1948.
42. G. H. Brown, "A Vestigial Sideband Filter," *RCA Rev.*, Vol. 5, pp. 301–326, 1941.

43. E. Bradburd, R. S. Alter, and J. Racker, "Vestigial Sideband Filter Design," *Tele-Tech,* Vol. 8, pp. 38–40, 52, October, 1949 and pp. 44–45, 60–61, November, 1949.
44. A. Sommerfeld, "Propagation of Electrodynamic Waves on a Cylindrical Conductor," *Ann. Phys. u. Chemie,* Vol. 67, pp. 233–290, 1899.
45. G. Goubau, "Surface Waves and Their Application to Transmission Lines," *J. Appl. Phys.,* Vol. 21, pp. 1119–1128, 1950.
46. G. Goubau, "Single-Conductor Surface Transmission Line," *Proc. I.R.E.,* Vol. 39, pp. 619–624, 1951.
47. N. E. Lindenblad, "Antennas and Transmission Lines at the Empire State Television Station," *Communications,* Vol. 20, pp. 13–22, 1940.
48. G. H. Brown, "The Turnstile Antenna," *Electronics,* Vol. 9, pp. 14–17, 48, April, 1936.
49. L. J. Wolf, "Triplex Antenna for Television and FM," *Electronics,* Vol. 20, pp. 88–91, July, 1947.
50. H. E. Gihring, "Practical Considerations in the Use of Television Super-Turnstile and Super-Gain Antennas," *RCA Rev.,* Vol. 12, pp. 159–176, 1951.
51. L. J. Wolf, "High-Gain and Directional Antennas for Television Broadcasting," *Broadcast News,* No. 58, pp. 46–53, March–April, 1950.
52. R. W. Masters, "A Power-Equalizing Network for Antennas," *Proc. I.R.E.,* Vol. 37, pp. 735–738, 1949.
53. O. O. Fiet, "Ultra-High-Frequency Antenna and System for Television Transmission," *RCA Rev.,* Vol. 11, pp. 212–227, 1950.
54. J. Epstein, D. W. Peterson, and O. M. Woodward, Jr., "Some Types of Omnidirectional High-Gain Antennas for Use at Ultra-High-Frequencies," *RCA Rev.,* Vol. 13, pp. 137–162, 1952.
55. G. E. Mueller and W. A. Tyrrell, "Polyrod Antennas," *Bell System Tech. J.,* Vol. 26, pp. 837–851, 1947.
56. W. H. Sayer, Jr., and J. M. deBell, Jr., "Television Antenna Diplexers," *Electronics,* Vol. 23, pp. 74–77, July, 1950.
57. H. E. Gihring, "The Empire State Television Antenna System," *Broadcast News,* No. 70, pp. 24–33, July–August, 1952.
58. H. H. Beverage, "Some Notes on Ultra-High-Frequency Propagation," *RCA Rev.,* Vol. 1, pp. 76–87, January, 1937.
59. C. R. Burrows, L. E. Hunt, and A. Decino, "Ultra-Short-Wave Propagation Mobile Urban Transmission Characteristics," *Bell System Tech. J.,* Vol. 14, pp. 253–272, 1935.
60. R. W. George, "A Study of Ultra-High-Frequency Wide-Band Propagation Characteristics," *Proc. I.R.E.,* Vol. 27, pp. 28–35, 1939.
61. R. S. Holmes and A. H. Turner, "An Urban Field Strength Survey at Thirty to One Hundred Megacycles," *Proc. I.R.E.,* Vol. 24, pp. 755–770, May, 1936.
62. G. H. Brown, J. Epstein, and D. W. Peterson, "Comparative Propagation Measurements; Television Transmitters at 67.25, 288, 510, and 910 Megacycles," *RCA Rev.,* Vol. 9, pp. 177–201, 1948.
63. G. H. Brown, "Field Test of Ultra-High-Frequency Television in the Washington Area," *RCA Rev.,* Vol. 9, pp. 565–584, 1948.
64. R. F. Guy, "Investigation of Ultra-High-Frequency Television Transmission and Reception in the Bridgeport, Connecticut Area," *RCA Rev.,* Vol. 12, pp. 98–142, 1951.

65. D. R. Goddard, "Observations on Sky-Wave Transmission on Frequencies above 40 Megacycles," *RCA Rev.*, Vol. 3, pp. 309–315, 1939.
66. RCA Laboratories Division, "A Study of Co-channel and Adjacent Channel Interference of Television Signals," *RCA Rev.*, Vol. 11, pp. 99–120 (Part I) and pp. 287–395 (Part II), 1950.
67. L. S. Lappin and J. R. Bennett, "A New Ultra-High-Frequency Television Transmitter," *RCA Rev.*, Vol. 11, pp. 190–211, 1950.

16 The Receiver

In this chapter a functional analysis of the various divisions of the monochrome television receiver is given with the purpose of acquainting the reader with its basic operation. No attempt is made to furnish him with a detailed blueprint for the construction of a receiver. In part there is an integration of the preceding chapters, bringing together such elements as the antenna, video amplifier, deflection generator, kinescope, etc., into a coordinated whole. Some divisions including the i-f and r-f amplifiers are discussed only in this chapter.

Second only to obtaining a picture of high quality, the aim of receiver design is dependability and simplicity. Notwithstanding, the modern television receiver is a complex instrument when compared with the average amplitude-modulated broadcast receiver.

16.1 Elements of the Television Receiver. A twofold function is performed by the television receiver, the reception of amplitude-modulated picture signals and generation of a visual image therefrom, and the reception of frequency-modulated aural signals and production of sound. All receivers operate upon the superheterodyne principle whereby the aural and picture carriers and sidebands beat with a local oscillation to shift them to a lower frequency range. Figure 16.1 illustrates in block diagram the two types of receiver which differ only in the treatment of the sound. In the first type, indicated by dashed lines, the sound signal is split off from the picture signal early in the video i-f amplifier and is amplified and detected in a separate frequency-modulated receiving circuit. A different handling of the sound signal termed "intercarrier" sound is shown in dashed-dot lines. Here the sound and picture signals are amplified together in the picture i-f amplifier and are separated in the video amplifier. The older, separate sound method has been largely superseded by the intercarrier sound arrangement.

The order of gain usually provided in the various divisions of a typical receiver are 20 times in the r-f section, 3 times in the conversion process, 20,000 times in the i-f amplifier, and 20 times in the video amplifier.

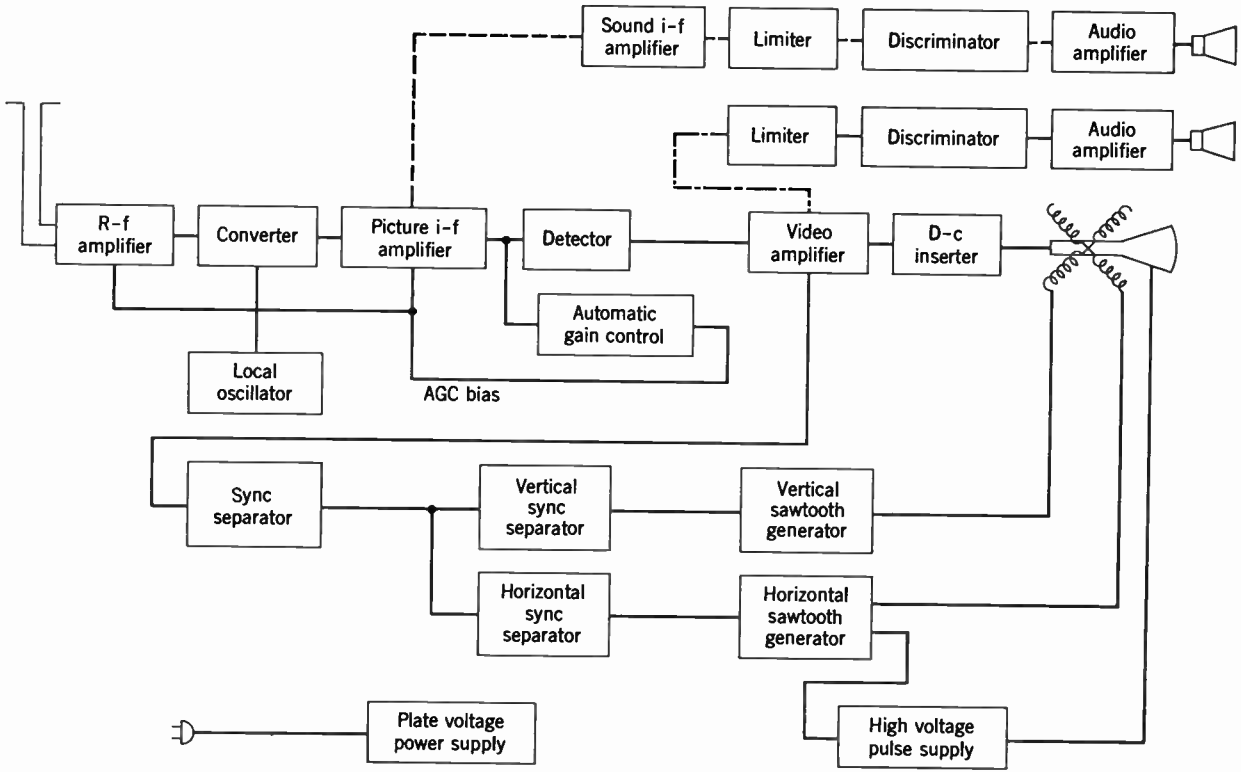


Fig. 16.1. The Receiver. Dashed lines show separate sound system; dashed-dot lines show intercarrier sound system.

Synchronizing signals which occupy the range from 75 to 100 percent amplitude of the detected video signal must first be separated from picture components and then segregated by filters into horizontal and vertical synchronizing pulses. Absolute synchronism of the horizontal and vertical sawtooth generators with the corresponding deflection generators on the camera must be maintained on the average by the timing information contained in the pulses. If the video amplifier is capacitor-coupled, the average brightness information conveyed by the d-c and very low-frequency components of the picture is lost temporarily and must be restored at the grid of the kinescope by a d-c restorer.

16.2 The Receiving Antenna. A high signal-to-interference ratio greatly facilitates the functioning of a television receiver to the end of presenting an acceptable picture. Receiving conditions that influence the ratio include:

1. Signal strength of the direct radiation.
2. Magnitude of any reflected radiation (ghosts) in free space.
3. Magnitude of reflected waves on the transmission line.
4. Presence of impulse noise, diathermy radiation, and other interference created by electrical apparatus in the vicinity of the antenna.
5. Reflected radiation by airplanes overhead.
6. Interference from other communication services such as adjacent channel and co-channel television stations, transoceanic broadcasts, and amateur transmitters.

Two primary factors ordinarily determine the signal input to the receiver: (1) the antenna gain, and (2) the field strength of the direct radiation from the transmitter in the vicinity of the antenna. The quasi-optical properties of high-frequency electromagnetic waves used for television transmission result in pronounced shadow and reflection effects from relatively small obstacles, such as buildings, gas or water tanks, and bridges. Consequently, the direct radiation field strengths may be radically different at points separated by only a few feet. The resultant field may also contain an appreciable reflection component. Where such circumstances are present the location of the antenna must be selected with care, even in a locale where the average field strength is high.

The most important factor influencing the desirability of a possible antenna position is the presence or absence of reflections. These reflections cause a blurring of the image, or even the appearance of multiple images, owing to the delay in time of arrival of the reflected

signal as a consequence of the longer path traversed. At first sight it may appear that the delay caused by a few hundred feet of additional path length would be inconsequential. However, when it is considered that the linear velocity of the spot across a 19-inch picture screen is in the neighborhood of 300,000 inches per second, it will be realized that the spot moves 0.0003 inch per foot of signal path length. Consequently, a reflection path which is only 300 feet longer than the direct path results in a secondary image displaced by 0.09 inch from

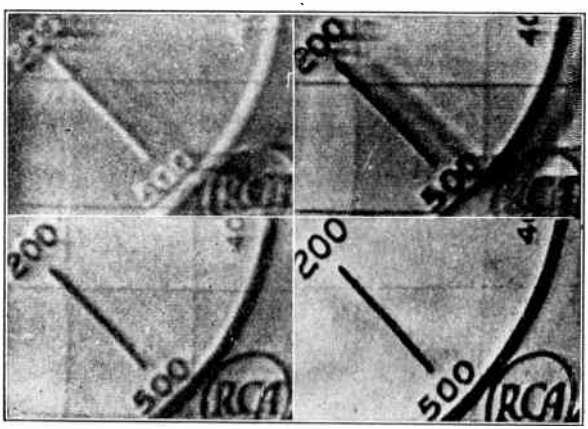


Fig. 16.2. Effect of Reflections on Reproduced Picture. (S. W. Seeley, reference 1.)

the wanted image. The blurring effect of such reflections is clearly visible in the reproduced picture * shown in Fig. 16.2.

Reflections, of course, occur from any conducting surface such as metal roofs and tanks. Large buildings are also potential sources of this type of interference since the difference in dielectric constant between the surrounding air and such building materials as stone, brick, or concrete is sufficient to produce strong reflections. Interference by reflections is avoided in part by properly positioning the antenna, and in part by making use of the directional properties of television antennas.†

The most frequently used television antennas are the half-wave dipole and related forms (Fig. 16.3). This type of antenna has already

* See Seeley, reference 1.

† It should be added that in some circumstances in which the direct radiation cannot be received without the intrusion of reflections, it is preferable to orient the antenna for the reception of a single strong reflection to the exclusion of the direct radiation.

been introduced in the preceding chapter. In the simple form, the dipole consists of a straight-rod conductor divided at the center for feeding a balanced transmission line. When coupled in this way the

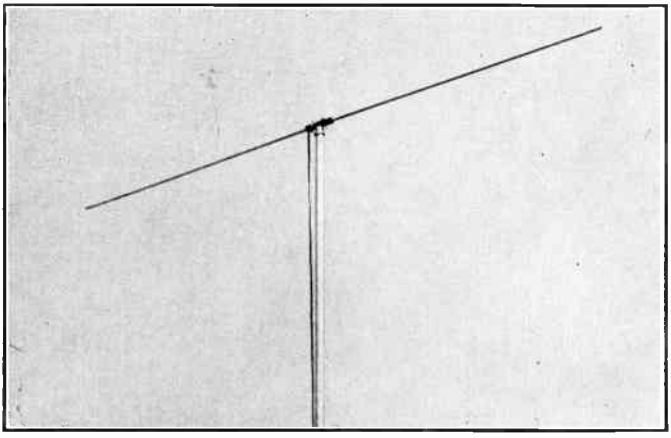


Fig. 16.3. Half-Wave Dipole Antenna.

dipole has a radiation resistance of approximately 72 ohms at resonance. For television reception, the transmission line and its terminating impedance at the receiving end are rarely set equal to the im-

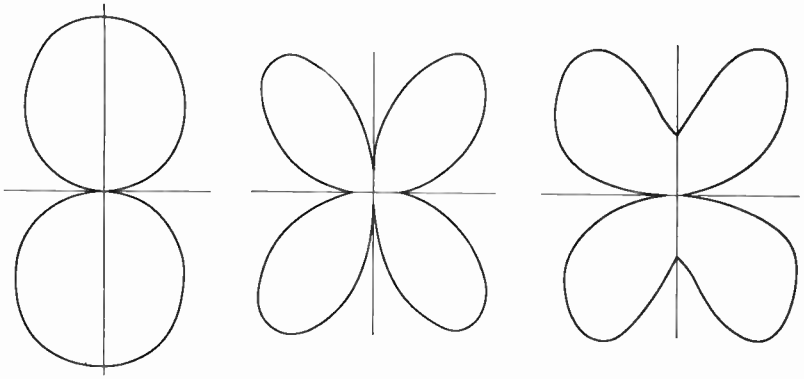


Fig. 16.4. Polar Characteristics of Simple Half-Wave Dipole Antenna.

pedance of the dipole since a matched dipole is sharply resonant at one frequency.

The adopted practice calls for a 300-ohm transmission line and load for which the dipole becomes sufficiently broad-band over the lower

very-high-frequency television band from 54 to 88 megacycles when the length of the dipole corresponds to a half wave in the vicinity of 59 megacycles. However, as the frequency is increased, the sensitivity

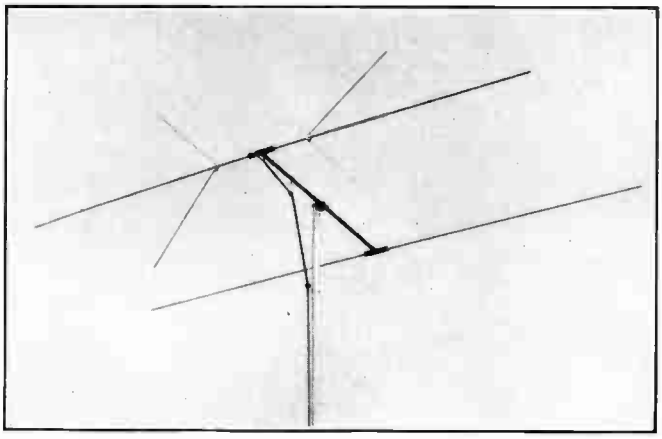


Fig. 16.5. Winged Dipole Antenna with Reflector.

pattern changes as indicated in Fig. 16.4, and in the band 174 to 216 megacycles the pattern of the same dipole is characterized by four

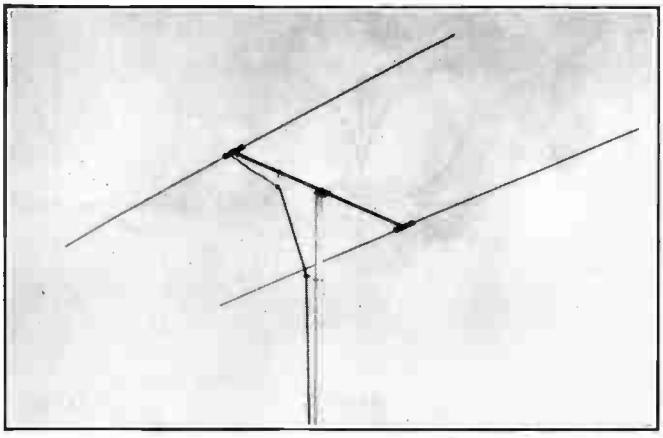


Fig. 16.6. Dipole Antenna with Reflector.

lobes. If the dipole cannot be rotated to utilize the maximum sensitivity of one of the lobes, the voltage resulting at the receiver terminals drops to a low value for the normal orientation.

Various modified dipole arrangements have been devised to restore the pattern to the two-lobed character over the upper bands. One arrangement is the "winged" dipole shown in Fig. 16.5. Here, the addition of rods approximately 0.07λ long, spaced at approximately 0.1λ from the center of the dipole, effectively suppresses additional lobes over the upper ultra-high-frequency channel.

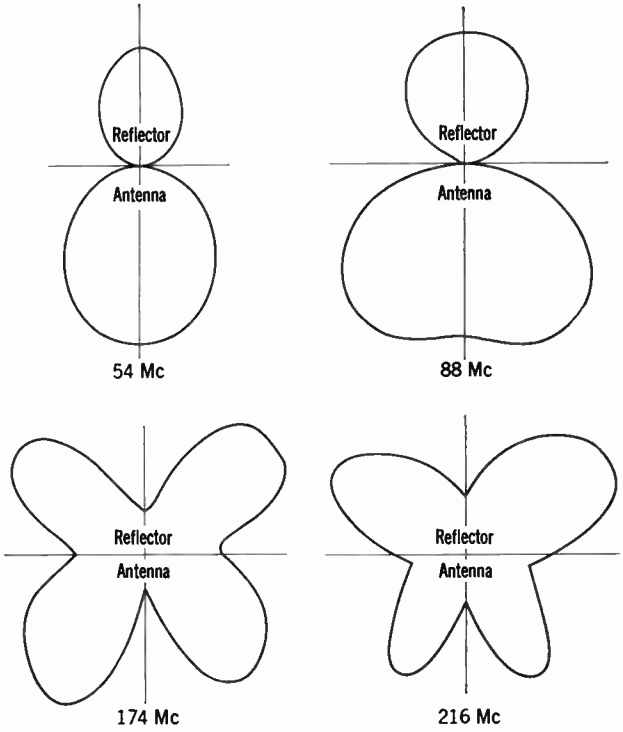


Fig. 16.7. Polar Patterns of Dipole Antenna with Reflector.

An improvement in the signal-to-interference ratio and sensitivity results when a parasitic element or reflector is placed behind the dipole at a distance of approximately $\lambda/4$, as illustrated in Figs. 16.5 and 16.6. The action of the reflector may be explained by noting that the main field induces a current in the reflector which sets up a field in phase opposition to the main field. The field due to the reflector is propagated away in both directions from the reflector. This radiation arrives at the dipole in phase with the main field, and the total current flow in the dipole is thereby reinforced. In a similar manner, the sensitivity to radiation arriving from the backward direction is dimin-

ished by the phase opposition of the current induced in the dipole by the field of the reflector. The directional patterns are shown in Fig. 16.7. Wide-band operation is obtained by the addition of wings to the dipole as in Fig. 16.5 and the interposition of a shorter reflector element spaced approximately midway between the dipole and long reflector.

A variety of other receiving antennas described as rhombic, biconical, vee, folded dipole, yagi, multi-array, etc., are useful for general reception and special purposes.*

16.3 Radio-Frequency Amplifier. The r-f amplifier includes all circuits between the output of the transmission line and the input of the frequency converter. An adequate r-f amplifier performs the following functions satisfactorily:

1. Termination of the transmission line with an impedance which is well matched over the television bands for prevention of reflections on the transmission line and efficient transfer of signal energy. A termination balanced to ground is also desirable for suppression of interference picked up on the transmission line.
2. Provision of adequate selectivity (image rejection) so that signals outside the desired channel are excluded which would otherwise give rise to a frequency spectrum coincident with the desired spectrum after frequency conversion.
3. Low random-noise injection.
4. Uniform amplitude frequency response over the desired channel.
5. High attenuation to the passage of oscillator voltage backward through the r-f amplifier to the antenna. Radiation of energy from the receiving antenna may seriously interfere with the reception at other receiving points.
6. Sufficient gain to over-ride random noise generated in the frequency converter circuit.
7. Economical and trouble-free construction.

In the very-high-frequency range the driven grounded-grid triode circuit or cascode circuit qualifies well as an r-f amplifier.† Pentode amplifiers compare unfavorably with triode amplifiers on account of the higher tube noise of pentodes.

The cascode circuit (Fig. 16.8) is basically a grounded-grid triode amplifier, T_2, driven at the cathode by a second triode, T_1. Amplification in the circuit is supplied by T_2, whereas T_1 makes possible a satis-

* See Bailey, reference 2.

† See Cohen, reference 3, and Murakami, reference 4.

factory impedance match for the transmission line and serves as an efficient driver for T_2. The input impedance of a grounded-grid stage varies with the transconductance of the tube and is, therefore, not easily matched directly to the transmission line.

Feed-through of voltage from the local oscillator is attenuated to about the same extent by the shielding action of the grounded grid in the cascode circuit as by the screen grid of a pentode amplifier.

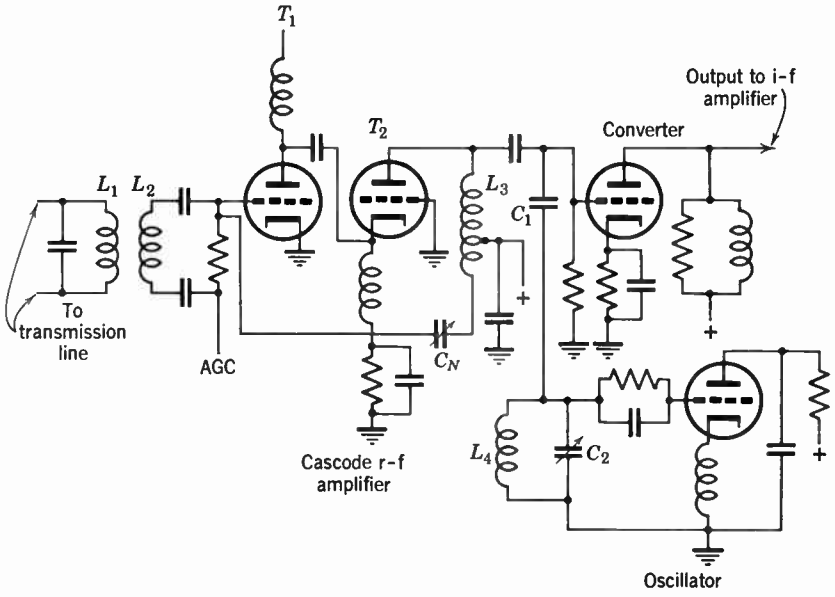


Fig. 16.8. Cascode R-F Amplifier, Converter, and Oscillator.

The tendency for regeneration to occur in the input stage through the plate-to-grid capacitance of T_1 is neutralized by a signal fed back from the second stage through the neutralizing capacitor, C_n.

Individual pretuned inductors, L_1, L_2, and L_3, are switched in for each channel. The values of L_1 and L_2 are determined by the requirement of matching the antenna circuit to the input impedance of T_1 over a 6-megacycle bandwidth with good efficiency.

In the ultra-high-frequency range (470 to 890 megacycles) distributed tuning elements such as the variable transmission line replace the physically familiar coils which no longer may be treated as lumped-circuit constants at the high frequencies involved.*

* See Newton, reference 5.

16.4 Frequency Converter and Local Oscillator. Television receivers use the superheterodyne principle of combining a locally generated sine wave signal with the desired r-f signal spectrum at the grid of a converter tube so that sum and difference spectra are formed in the plate circuit.* The frequency of the local signal is adjusted for each channel so that the first difference spectrum lies in the passband of the i-f amplifier. Picture and sound signal are converted to the lower frequency range simultaneously.

Conversion takes place as a consequence of the wide variation of the grid voltage over the nonlinear parts of the grid voltage-plate current characteristic of the converter tube. Increased useful output is obtained as the amplitude of the local signal is increased up to a point. A ratio of useful output signal to input picture signal of about one-third of the gain of the converter when used as an amplifier is obtained.

Triodes rather than pentodes or multigrid tubes are favored as converters for very-high-frequency receivers since the noise injection of the triode is less.

Simplicity of the crystal triode is an influential factor in a choice of converter for the ultra-high-frequency receiver even when theoretically a suitable tube triode would yield greater conversion gain.

Figure 16.8 illustrates the fundamental arrangement of a converter circuit and the associated local oscillator.

A relatively large signal taken from a Colpitts oscillator through a small coupling capacitor C_1 is combined with the output of the r-f amplifier at the grid of the converter. A different pretuned inductance L_4 is switched into the circuit for each television channel. C_2 is a fine tuning capacitor for accurate adjustment of the positions of the beat frequency picture and sound carriers on the selectivity characteristic of the i-f amplifier.

16.5 Receiver Noise. Small voltages due to the random motion of electrons in circuits and vacuum tubes constitute an ever-present background of noise which produces visible effects on the viewing screen when the gain of the receiver is high for the reception of weak signals.† Noise voltages originating in the antenna circuit, the r-f amplifier, and the converter circuit are the principal concern since the useful signal in these parts of the receiver may be quite low.

If it is assumed that the receiver contains no noise sources whatever, the thermal noise of the antenna radiation resistance would still im-

* See Herold, reference 6.

† For an extensive discussion of receiver noise, see Dome, reference 7, pp. 122–164.

press a noise voltage on the receiver terminals. Hence, a design criterion from a noise standpoint is the noise factor defined by the ratio

$$F = \frac{\text{total noise}}{\text{antenna noise}} = \frac{N_A + N_R}{N_A} \tag{16.1}$$

where N_A represents noise power generated in the antenna and N_R represents noise power generated in the receiver.

The noise factor of a receiver is a measure of how closely the signal-to-noise ratio approaches the theoretical minimum ratio. The minimum ratio applies to the output of the receiver when all noise voltages within the receiver are set equal to zero and only the thermal noise generated by the antenna resistance remains. When all noise voltages are calculated at the same point in a circuit, the noise factor may be taken as the ratio of the total noise voltage and the noise voltage due to the antenna resistance.

Any passive circuit at an absolute temperature T viewed as a two-terminal network gives rise to an rms noise voltage of magnitude

$$E = \left(4kT \int_{f_1}^{f_2} R(f)\, df\right)^{1/2} \tag{16.2}$$

in which k = Boltzmann's constant,

$R(f)$ = resistive component of the two-terminal impedance expressed as a function of frequency over the band.

If $R(f)$ is a constant over a specified band F cycles per second in width, the noise voltage in the band may be represented by a generator having an rms voltage

$$E = (4fTRF)^{1/2} \tag{16.3}$$

acting through a series resistance of R ohms, which is noiseless.

Two types of noise voltage arise in vacuum tubes; shot noise due to the random flow of electrons from cathode to plate and noise originating in the transit time resistance. Shot noise voltage is independent of frequency and, in a triode, corresponds to the thermal noise of an equivalent input resistance given approximately by the formula

$$R_s = \frac{2.5}{g_m}$$

in which g_m is the transconductance of the tube.

Noise voltage due to transit time resistance is approximated by Eq. 16.3 for a temperature which is approximately five times the ambient temperature. $R(f)$ varies here as the square of the frequency.

Application of these principles is illustrated in the following analysis of the cascode r-f amplifier and triode converter shown in Fig. 16.9. Circuit elements not involved in the calculations are omitted.

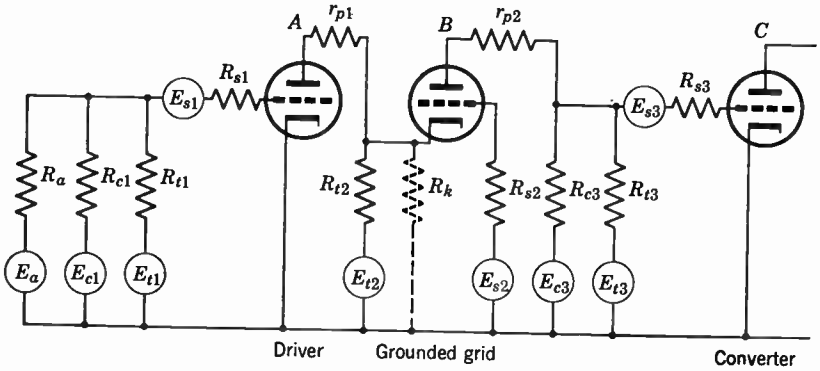


Fig. 16.9. Circuit of Cascode R-F Amplifier and Converter Showing Noise Voltages.

The antenna resistance R_a reflected into the input circuit of the driver tube is assumed to be constant over the bandwidth F. According to Eq. 16.3 the voltage of the equivalent noise generator is

$$E_a = (4kTR_aF)^{1/2}$$

A similar noise voltage E_{c1} arises in the resistance of the input circuit of the driver tube.

Transit time resistance noise is accounted for by a generator having a voltage

$$E_{t1} = (4k \cdot 5T \cdot R_{t1}F)^{1/2}$$

acting in series with a resistance R_{t1}.

Shot noise voltage is represented by a generator

$$E_{s1} = (4kTR_{s1}F)^{1/2}$$

R_{s1} is the equivalent resistance for shot noise.

The four noise voltages mentioned above have the following values at point A of the circuit:

$$(E_a)_A = E_a\mu_1 \frac{R_{c1}R_{t1}}{R_a(R_{c1} + R_{t1}) + R_{c1}R_{t1}}$$

$$(E_{c1})_A = E_{c1}\mu_1 \frac{R_aR_{t1}}{R_{c1}(R_a + R_{t1}) + R_aR_{t1}}$$

$$(E_{t1})_A = E_{t1}\mu_1 \frac{R_aR_{c1}}{R_{t1}(R_a + R_{c1}) + R_aR_{c1}}$$

$$(E_{s1})_A = E_{s1}\mu_1$$

The total noise voltage at A is

$$E_A = [(E_a)_A^2 + (E_{c1})_A^2 + (E_{t1})_A^2 + (E_{s1})_A^2]^{1/2}$$

Therefore, a noise ratio $F_A = E_A/(E_a)_A$ exists at A.

E_A is referred to point B by allowing for the gain G_k of the grounded-grid stage and the voltage dividing effect of the three resistances: r_p, the plate resistance; R_{t2}, the transit time resistance; and R_k, the input resistance viewed from the cathode. There results

$$(E_A)_B = E_AG_k \frac{R_kR_{t2}}{R_kR_{t2} + r_{p1}(R_k + R_{t2})}$$

in which $G_k = \dfrac{R_{c3}(\mu_2 + 1)}{r_{p2} + R_{c3}}$

Noise voltage E_{t2} referred to point B becomes

$$(E_{t2})_B = G_k\sqrt{4k(5T)FR_{t2}} \frac{R_kr_{p1}}{R_kr_{p1} + R_{t2}(R_k + r_{p1})}$$

Noise voltage E_{s2} is referred to B by applying the gain factor G_g for triode operation with a cathode resistor R_e. The result is

$$(E_{s2})_B = G_g[4kR_eTF]^{1/2}$$

in which $G_g = \dfrac{(\mu_2 + 1)R_{c3}}{r_{p2} + R_{c3} + R_e(\mu_2 + 1)}$

$$R_e = \frac{R_{t2}r_{p1}}{R_{t2} + r_{p1}}$$

All noise voltages summed at B yield

$$E_B = [(E_A)_B^2 + (E_{t2})_B^2 + (E_{s2})_B^2]^{1/2}$$

Dividing E_B by $(E_a)_B$, the antenna noise voltage referred to B, gives the noise factor at B.

The calculation of the noise factor is extended to point C at the plate of the converter grid by steps similar to the preceding ones. In general, noise voltages occurring beyond the point C are negligible in comparison with those referred to that point from earlier sources.

Receiver noise factors ranging from about 3 decibels in the low very-high-frequency channels to 6 decibels in the high very-high-frequency channels have been achieved in practice.*

16.6 Intermediate-Frequency Amplifier. Since the i-f amplifier is relied upon to furnish the largest part by far of the overall gain and

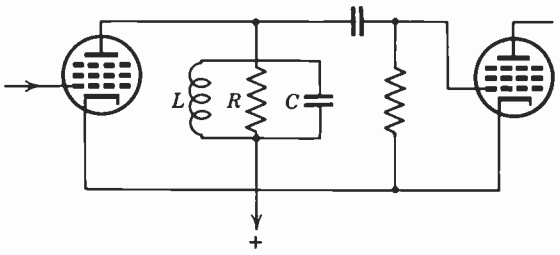


Fig. 16.10. Simple Parallel-Tuned I-F Stage.

adjacent channel selectivity of the receiver, several amplifying tubes in cascade are employed. The networks coupling the successive stages must have an overall frequency response consistent with the requirements of vestigial sideband transmission (Fig. 15.2*c*). There must be sufficient rejection for adjacent channel picture and sound signals. If a separate i-f amplifier is used for sound, the picture i-f amplifier must attenuate the sound i-f carrier after the point of sound take-off.

Coupling networks in the picture i-f amplifier fall into three classifications: (1) stagger-tuned, (2) synchronously tuned with more or less coincident bandwidths, and (3) combination of (1) and (2).

It is well known that multistage amplifiers with identical amplitude and phase characteristics may be designed on any one of the three plans. Two or more simple parallel circuits nonsynchronously tuned and isolated by tubes may have the same transmission characteristic as the multiple-tuned coupled circuit. Factors involved in making a choice are practicability and simplicity of construction, ease of adjustment, and gain per stage.

*W. H. Pan, "Investigation of UHF Television Amplifying Tubes," presented at the Fall Radio Meeting, Toronto, Canada, October 26–28, 1953.

The single parallel-tuned circuit (Fig. 16.10) cannot be repeated exactly through several stages of a multistage amplifier without a serious reduction in the overall bandwidth. However, if the individual resonant frequencies are "staggered" and the individual Q's appropriately chosen, a more nearly rectangular overall response is attained without the sacrifice in bandwidth which would accompany an equal number of identical stages. Defining bandwidth in the usual manner as the width of the frequency response characteristic at the point of 70.7 percent response, the bandwidth for N synchronously tuned circuits is to a good approximation given by *

$$\frac{\text{Bandwidth of single-tuned stage}}{1.2N^{1/2}} = \frac{1/(2\pi RC)}{1.2N^{1/2}}$$

where R is the load resistance and C the total shunt capacitance. For example, in a five-stage amplifier designed for an overall bandwidth of 3.5 megacycles, each individual stage must have a bandwidth of 10 megacycles. By stagger-tuning five stages, an overall bandwidth of 10 megacycles is attainable with the same overall gain.

The frequency response of the N stage stagger-tuned amplifier is $1/\left[1 + \left(2\frac{\delta f}{\Delta f}\right)^{2N}\right]^{1/2}$, where δf is departure of the given frequency from center frequency and Δf is the bandwidth.

Figure 16.11 illustrates the shape of the selectivity characteristic of an amplifier with N staggered stages, $N = 1$ to 5, for equal overall bandwidths of Δf megacycles and maximally flat response. The responses of the individual stages of a five-stage amplifier are shown in Fig. 16.12.

In practice, if a small ripple is allowed in the overall response, a wider bandwidth or greater gain is obtainable. This result is accomplished by narrowing the passband of each tuned circuit, but retaining the same resonant frequency. Considerable narrowing is required to produce appreciable deviations in the overall response of a stagger-tuned amplifier.

It is generally considered that the lowest i-f signal which can be satisfactorily used is 50 to 100 microvolts. With this sensitivity, operation of the diode detector over its full range requires an amplification of about 20,000 times, which is easily achieved by five stages of stagger-tuning.

* See Wallman, reference 8.

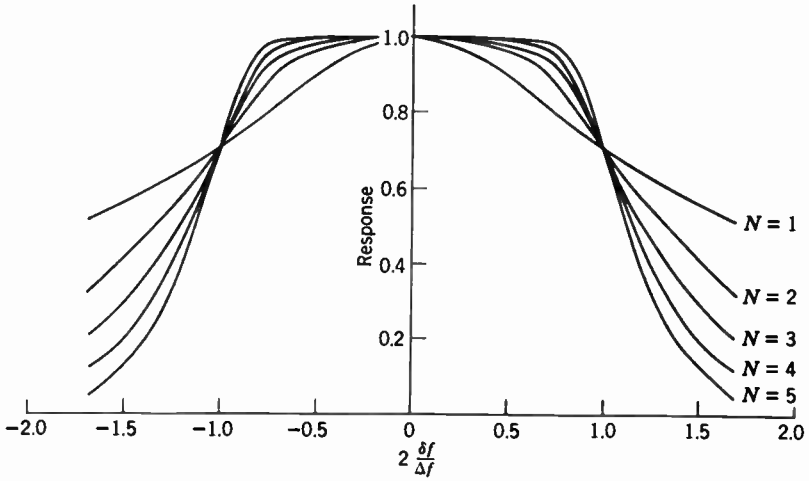


Fig. 16.11. Response of Amplifier with Stagger-Tuned Stages (N = Number of Stages).

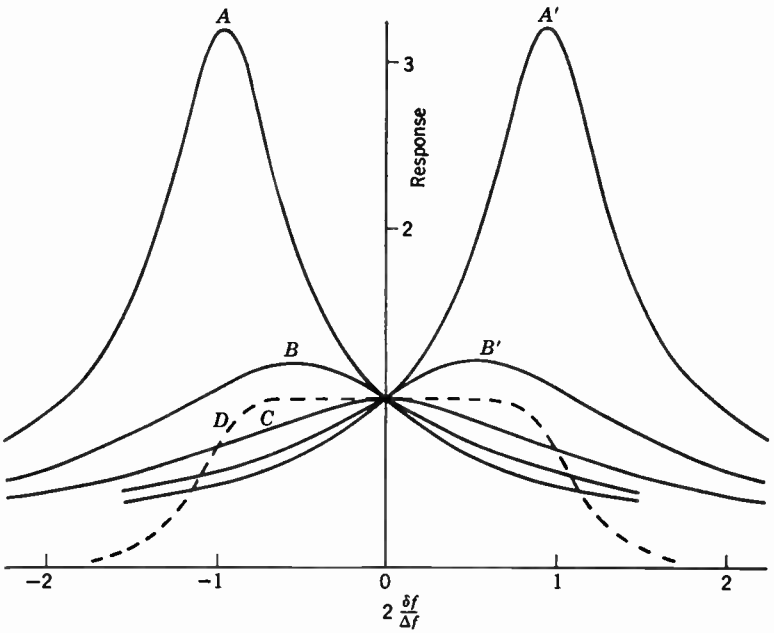


Fig. 16.12. Overall Response (D) of Five Stagger-Tuned Stages A, A', B, B', and C.

Rejection of adjacent sound and picture carriers in a stagger-tuned i-f amplifier is accomplished by simple parallel or series-resonant trap circuits. Figure 16.13*a* shows a parallel-tuned circuit inductively

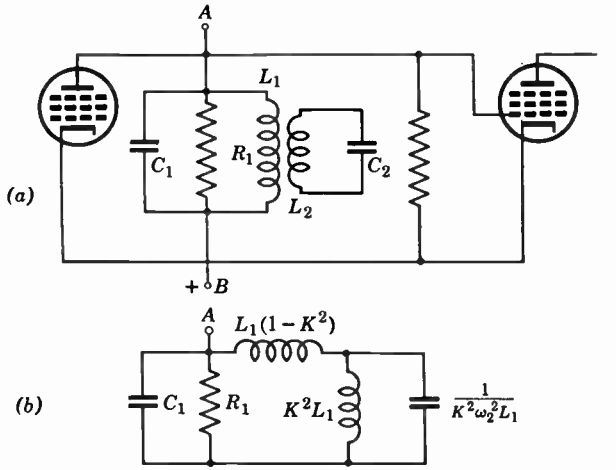


Fig. 16.13. Parallel-Tuned I-F Stage with Inductively Coupled Trap Circuit.

coupled to the tuning inductance of a stagger-tuned stage such that absorption of energy occurs at the trap frequency. The equivalent circuit of the combination is shown in Fig. 16.13*b*.

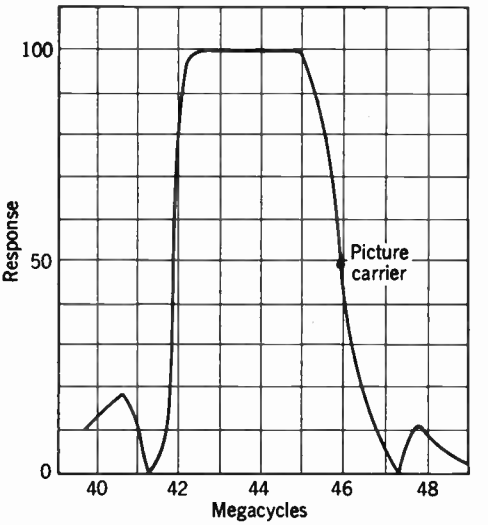


Fig. 16.14. Response of Three Stagger-Tuned Pairs, Including Traps.

The impedance represented by the branch A–B has a zero at the rejected frequency given approximately by

$$f_0 \approx f_2/\sqrt{1 - K^2}$$

in which

$$f_2 = 1/(2\pi\sqrt{L_2C_2})$$

An indication of the overall amplitude characteristic of a stagger-tuned i-f amplifier with absorption traps is given in Fig. 16.14.

16.7 Detection and Video Amplification. Various forms of detection have had limited use in television receiver design, but the preferred detector is the simple diode shown in a typical circuit in Fig. 16.15. Here the diode acts substantially as a peak rectifier of the impressed i-f signal. The capacitance C_1 is charged by positive peaks of signal and discharged by the load resistor R_1. Although R_1 must be large enough to maintain the charge on C_1 from cycle to cycle of i-f signal, it must be small enough to allow the charge to follow the envelope or video modulation. C_1 is usually of the order of 10 micromicrofarads and R_1 approximately 5000 ohms. Inductances L_1 and L_2 constitute series and shunt peaking for compensation of the stray capacitance C_2.

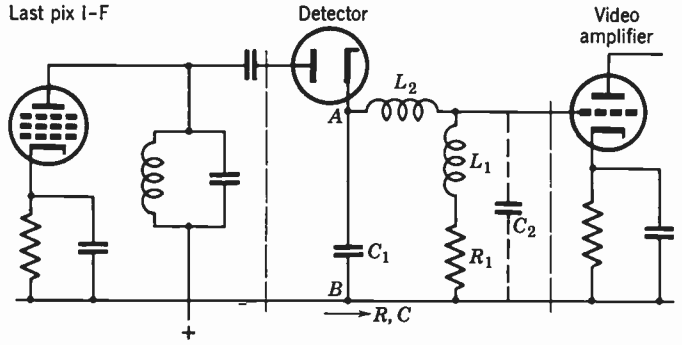


Fig. 16.15. Diode Detector and Coupling Circuit to Video Amplifier.

Since by far the greatest part of the overall amplification in a television receiver is accomplished in the i-f amplifier, the video amplifier is called upon for only a moderate gain of the order of fifty times. A bandwidth of approximately 4 megacycles is usually adequate, although bandwidth per se does not uniquely determine the fidelity of a video amplifier.

Basically, video amplifier circuits are modifications of the familiar resistance-capacitance coupling with inductances added for the extension of the frequency band over which the amplitude is substantially flat. Three types of compensation illustrated in Fig. 16.16 are in general use: shunt, series, and series-shunt. The transmission characteristics of these types are given in Chapter 13.

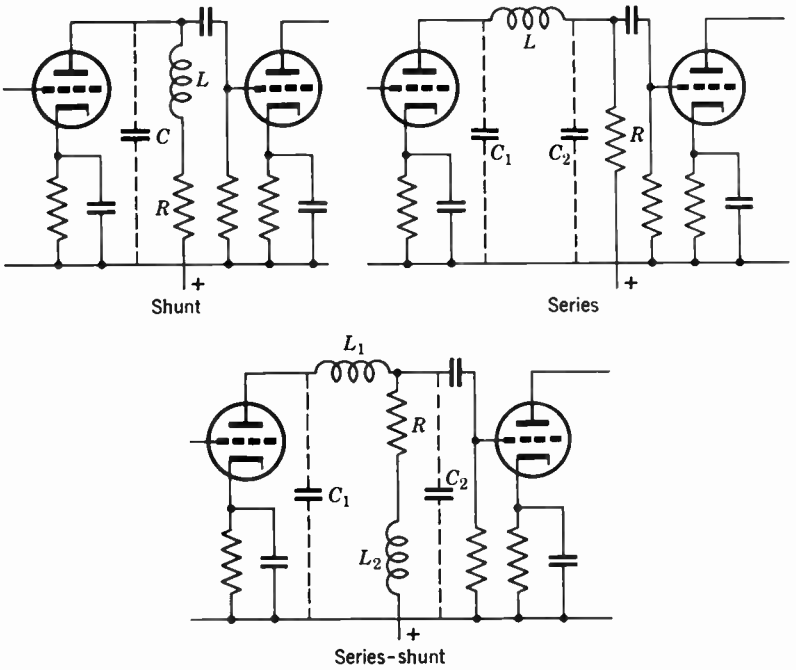


Fig. 16.16. Shunt, Series, and Series-Shunt Video Amplifier Coupling Circuits.

At low frequencies the response of the shunt circuit is determined by R since both C and L are ineffective. At higher frequencies the shunting effect of C tends to be reduced by a resonance action of L, but eventually a frequency is reached beyond which C dominates and the response decreases. Shunt compensation increases the bandwidth of the R-C amplifier about 1.5 times.

Series compensation and series-shunt compensation are four-terminal coupling networks in which the stray shunting capacitances associated with the output and input tubes of the stage are separated by inductive elements. In the series form an inductance L resonates with the stray capacitance C_2 near the end of the useful band, thereby tending to

compensate for the shunting effect of C_1. Series compensation will yield an increase in bandwidth of about two times relative to the uncompensated R-C coupled amplifier.

Series-shunt compensation combines the two types of compensation found in the shunt and series circuits. An increase in bandwidth of approximately 2.5 times over the uncompensated R-C amplifier is possible.

16.8 Synchronizing. Since the broad principles of synchronizing circuits have been considered in Chapter 14, the discussion here will be concerned chiefly with application to the receiver.

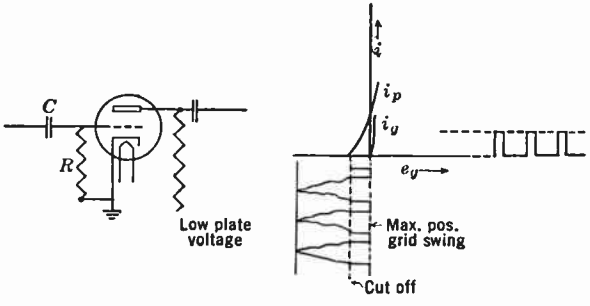


Fig. 16.17. Limiter for Separating Synchronizing from Video Signal.

Before the synchronizing pulses can control the deflection generators they must be separated from the picture signal, and also horizontal and vertical signals must be isolated. The initial separation may be accomplished by the action of a clipper tube which is biased so that the pulses are accepted and picture signals below black level are rejected. A typical grid bias R-C synchronizing clipper is shown in Fig. 16.17. Synchronizing pulses of positive polarity cause grid current to flow at the pulse peaks and rapidly charge C to such a potential that plate conduction ceases. During the ensuing interval, C discharges slowly through R. The tube is operated at a low plate voltage so that cutoff occurs when the grid is only moderately negative. The time constant of the R-C circuit is made fairly large so that only the peaks of the synchronizing pulses carry the grid to zero potential, that is, to the point at which the grid starts drawing current. The amplitude of the synchronizing pulses must be sufficient to swing the grid to cutoff. The video signal, which is more negative than the synchronizing signal, is beyond the cutoff of the tube and, therefore, is not reproduced in the plate current of the limiter.

In the simple form of Fig. 16.17, the R-C separator is vulnerable to impulse noise that simulates synchronizing pulses. Some degree of discrimination against noise is provided by a resistor inserted in series

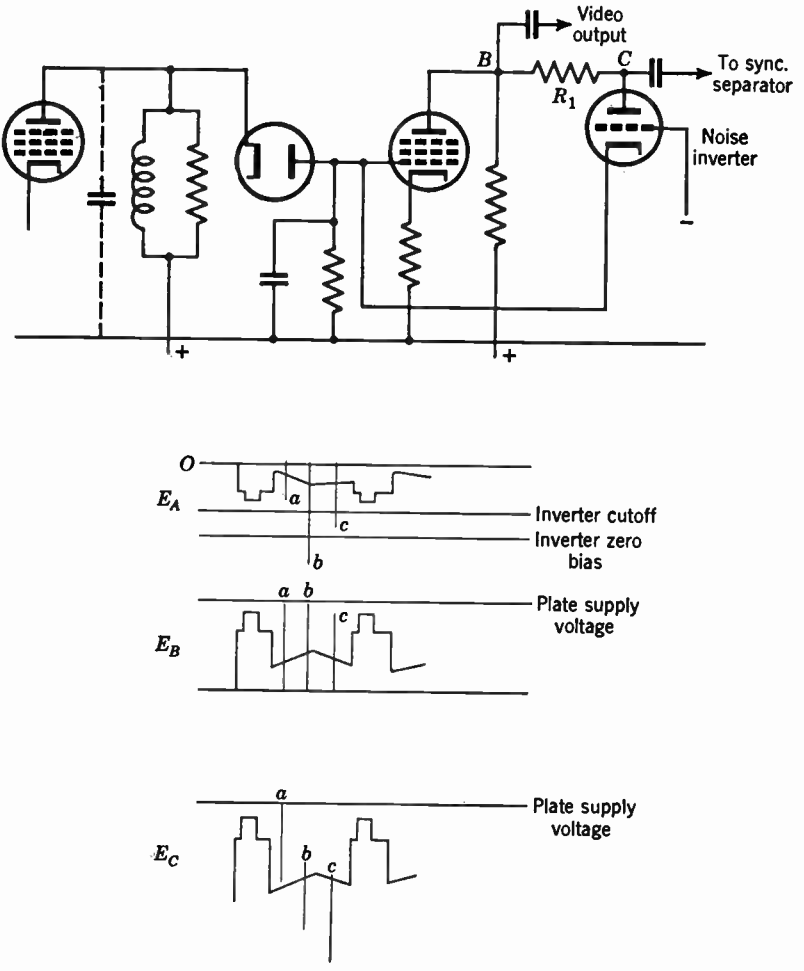


Fig. 16.18. Noise Inverter Circuit.

with C which develops an appreciable voltage drop when high noise peaks are present.

In the main, however, noise immunity in most receivers is obtained in the video amplifier before the signal is applied to the synchronizing separator. This noise immunity is obtained by a tube that limits the

excursions of noise pulses to a level near the top of the synchronizing pulses. Automatic gain control of the signal insures a definite signal amplitude at the input of the noise limiter and thus leads to accurate limiting.

Limiting is of small avail when noise is so intense that the duration and the number of noise pulses obscure the regularity of synchronizing. Although the modern television receiver has been enabled by automatic frequency controlled synchronizing to outperform by far its predecessors in respect to satisfactory synchronization in the presence of noise, further improvements in synchronizing pulse separation can still be utilized.

When many noise peaks widely exceed the synchronizing pulse peaks, the separator tends to pass noise pulses in preference to the synchronizing pulses of lower amplitude. Under such conditions the noise inverter circuit shown in Fig. 16.18 gives a significant improvement in the quality of the separated pulses. The principle of operation is as follows. Picture signal of positive polarity (synchronizing down) is applied to the grid of the video amplifier and the cathode of the noise inverter tube. A negative bias on the inverter tube prevents the flow of plate current whenever the applied signal and noise pulses are less negative than the peaks of synchronization. Consequently, plate current flows in R_1 in response only to noise voltages more negative than synchronization. Since the voltage developed at B by the video amplifier is in phase opposition to the drop across R_1, noise peaks cannot exceed the potential of the plate supply. High peaks such as b and c are depressed at point C into the picture region and exert no effect on synchronizing pulse separation.

16.9 Automatic Gain Control. The purpose of automatic gain control in a television receiver is maintenance of a substantially constant video signal at the input of the detector for a wide variation of r-f signal input to the receiver. Besides the obvious variations in field strength of various transmitters there are also the fluctuations due to the reflected waves from passing airplanes which may detract greatly from satisfactory viewing.

There are many AGC systems, each differing in detail but basically operating according to the same theory. A d-c voltage proportional to the synchronizing peaks of the i-f signal at the detector is developed by a diode and applied as a grid bias to one or more tubes in the r-f and i-f amplifiers. Thus, a weak signal tends to develop less bias voltage and thereby raises the gain of the r-f and i-f amplifier. If the AGC circuit is sufficiently sensitive, a new equilibrium is established

from which the video signal is only slightly less than that corresponding to a stronger input signal. A simple AGC circuit (Fig. 16.19) accomplishing this is essentially a peak-reading voltmeter. The diode rectifier has its cathode connected to output of the i-f amplifier so that conduction occurs during negative peaks of the i-f signal corresponding to synchronizing pulses. The time constant RC is sufficiently large that C is not discharged appreciably in a line period. A time constant of the order of 0.01 second is a typical value. The voltage of C is a measure of the peak amplitude of the i-f signal. An increase in

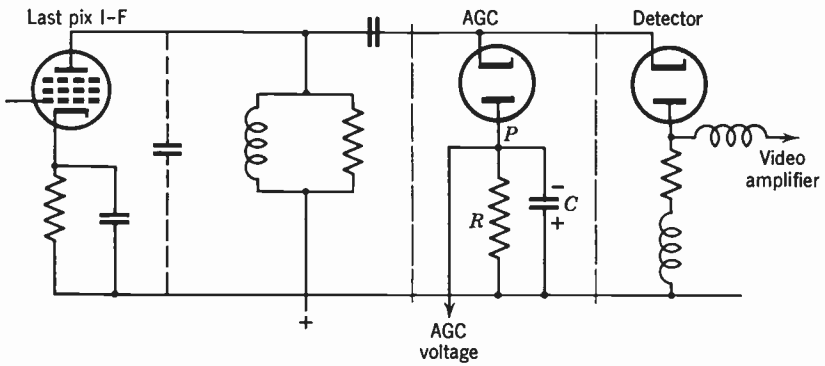


Fig. 16.19. Basic Circuit for Automatic Gain Control.

this amplitude causes a more negative potential to develop at point P which is applied to the variable gain stages to bring about a reduction in the gain of the i-f and r-f amplifiers.

Closer control of the gain may be obtained by amplification of the d-c output of the AGC diode. A bias may also be applied to the AGC to forestall the development of AGC voltage until a certain peak signal is present at the detector. In this manner, the gain of the receiver for weak signals may be held at maximum value.

The simple peak detector described above may respond to impulse noise which rises to the synchronizing level and above and thus generates a false bias. In this manner the effect of noise peaks is more extensive than in the absence of AGC since the bias cannot return to normal until the capacitor discharges through R. Discharge through the high resistance R is much slower than the charge through the low-resistance diode.

Many circuit modifications have been devised for decreasing the effect of noise on the operation of AGC.*

* See Wendt and Schroeder, reference 9.

16.10 Separation of Horizontal and Vertical Synchronizing Pulses. One form of horizontal separator circuit shown in Fig. 16.20*a* is a high-pass filter which has a differentiating action on the pulses applied to the circuit. Comparatively narrow pulses are generated at each abrupt rise and fall of the synchronizing wave form with a positive polarity for a rise and a negative polarity for a fall. Pulses in the positive half of the separator output occur at line frequency, with the exception of

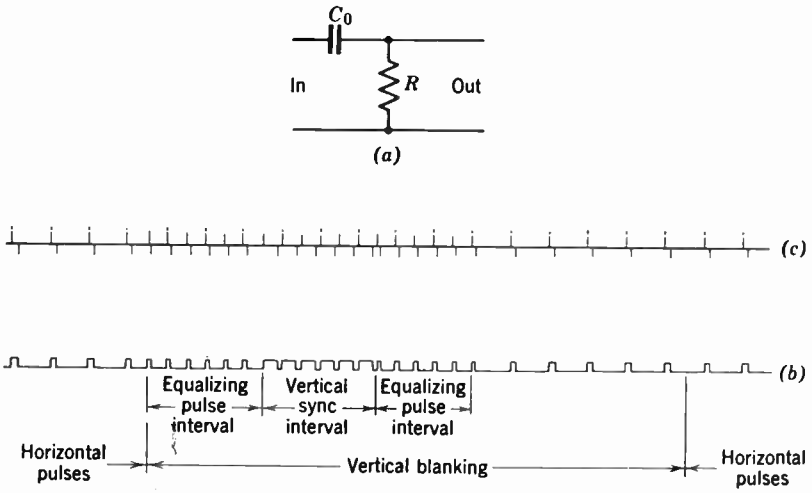


Fig. 16.20. Separation of the Horizontal Synchronizing Pulses from Composite Synchronizing Waveform: (*a*) Basic Circuit; (*b*) Synchronizing Waveform; (*c*) Differentiation of *b*.

extra pulses that occur during the vertical synchronizing period. Only the positive half of the separator output is used for synchronizing. The extra pulses in the output do not correspond to the natural period of the horizontal deflection generator and are, therefore, ineffective.

The vertical separator is a low-pass filter operated as an integrating circuit. Figure 16.21 illustrates schematically the output wave shapes for even and odd fields. Equalizing pulses at two times horizontal line frequency are part of the vertical synchronizing pulse. The purpose of the equalizing signals, as has already been pointed out, is to make the synchronizing pulses corresponding to the odd and even field traversals of the interlaced scanning substantially identical. This precaution, together with those necessary to prevent any cross-talk between horizontal pulses from the deflection generator and the input to

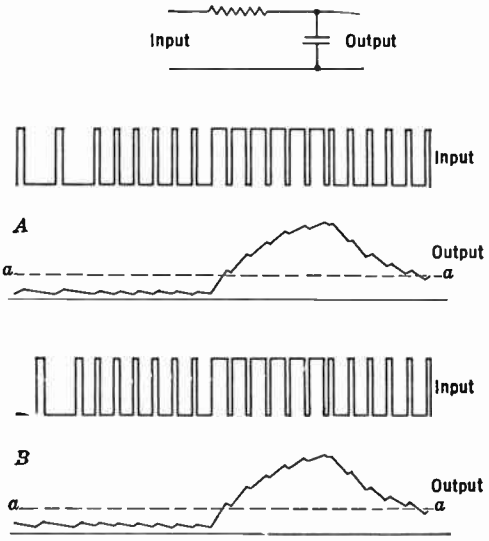


Fig. 16.21. Selection of Vertical Synchronizing Signal (Schematic).

the vertical selector circuit, must be taken if interlacing is to be maintained.

16.11 Horizontal and Vertical Deflection Systems. The synchronized blocking oscillator described in Chapter 14 has long been used as a source of pulse voltage for vertical and horizontal sawtooth generation in television receivers. Figure 14.32 shows a typical blocking oscillator and discharge circuit. Positive pulses of voltage, due to periodic blocking of the oscillator, are applied to the grid of the discharge tube. The resulting plate current is supplied largely by the charge on the capacitor C. As a consequence of the large dropping resistor R, inappreciable current is supplied by the plate supply during this short discharge or flyback interval. After the passage of the pulse, conduction ceases through the discharge tube and the potential of C rises slowly and uniformly throughout the deflection interval as charge is supplied by the plate supply through R. A substantially linear rise in the potential of C is insured by the selection of a large value of R. In this way, use is made of only a small part of the charge cycle which would be completed if it were not prematurely arrested by the opening of the conduction path through the discharge tube. Ample linear sawtooth voltage output, both vertical and horizontal, is easily generated in the foregoing manner for application to the deflection tube.

The horizontal blocking oscillator is a triggered device requiring uninterrupted application of synchronizing signals from a differentiating circuit (Fig. 16.20). Random noise and other forms of interference which would not otherwise seriously impair the reproduction of fine detail may induce irregular triggering of the oscillator, and in this round-about manner actually reduce the resolving power of the receiver. With the introduction of automatic frequency controlled synchronizing (AFC) greatly improved synchronization in the presence of noise was achieved.* An analysis of AFC circuits is given in Chapter 14.

Vertical circuits that have long been serviceable for magnetic deflection of the kinescope are the simple transformer coupling arrangements shown in Figs. 14.18 and 14.19. Since the impedance of the vertical deflection yoke windings is too low for efficient loading of a deflection tube, a transformer is normally employed to raise the impedance level at the plate of the tube. In practice, the yoke currents are also sawtooth in wave form, made possible by linear operation of the deflection tube. The circuits in Figs. 14.18 and 14.19 are well adapted for the requirements of vertical deflection since the range of frequencies from 60 to the order of 1000 cycles does not lead to excessive transient voltages in the windings of the yoke and transformer.

On the other hand, the distributed capacitances and inductances of the iron core circuits exhibit an oscillatory character in the range of horizontal frequencies from 15,750 to 150,000 cycles, especially during the short flyback time. This phenomenon and the various methods of damping are analyzed in Chapter 14.

The basic horizontal circuit employed extensively in home receivers is shown in Fig. 16.22. A modified sawtooth voltage is impressed on the grid-biased deflection tube. The excursion of grid voltage extends from zero grid bias to a point beyond cutoff. An explanation of the operation of the circuit is conveniently started at point t_0 at the end of the conduction interval of the tube. After t_0, the stored energy in the yoke is released to the circuit. The inductance of the yoke and associated shunt capacitance form an oscillatory circuit which swings through a half cycle of oscillation at the resonance frequency of the yoke circuit (inductance L). At the juncture t_1 the plate of the damper diode becomes positive and the stored magnetic energy is permitted to flow into the capacitor C. The oscillatory character of the discharge current is damped completely by the new conduction path

* See Wendt and Fredendall, reference 10.

through the damper tube. At t_2, conduction of the diode ceases. It should be noted that C adds its voltage to the plate supply with which

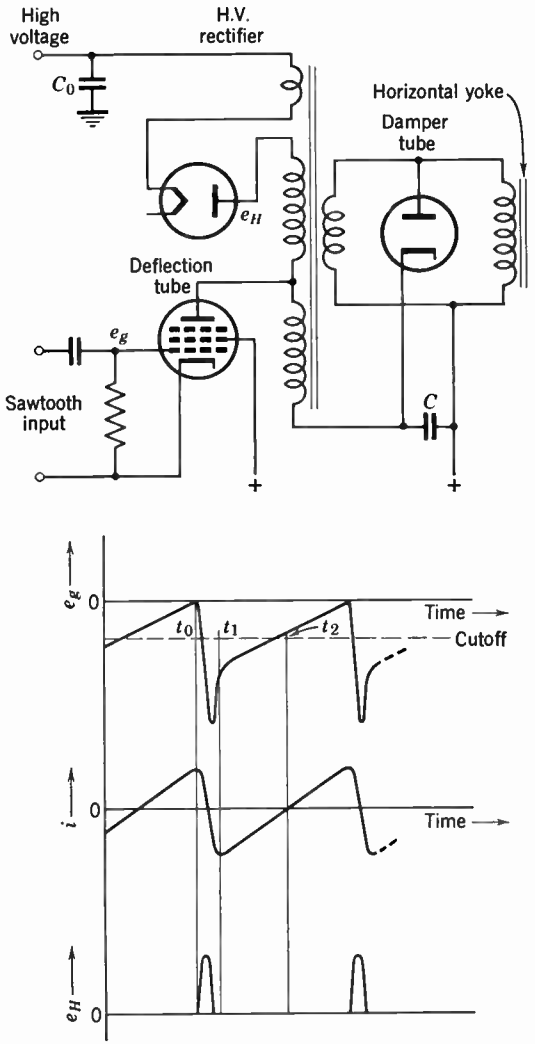


Fig. 16.22. Basic Horizontal Deflection Circuit with Power Feedback.

it is effectively in series. If the circuit had no losses, the energy required for deflection would be periodically returned to the plate supply. However, energy dissipation in the iron core and windings, tube drops, and the high voltage take-off reduces the efficiency.

Additional circuit elements not shown in Fig. 16.22 are incorporated in a working circuit for the control of linearity and flyback time. These have, however, been discussed in Chapter 14. During flyback time very high pulse voltages are developed across the auxiliary winding which are rectified by the H.V. diode in the manner of a peak rectifier. Voltages of the order of 10 kilovolts may be built up across C_0 and used as the anode voltage for the kinescope. This form of high-voltage supply has the advantages of economy of components, insurance against burning of the phosphor screen in the event of the failure of deflection, and small space requirements.*

16.12 Direct-Current Component. Any particular halftone value in a televised scene is represented by a definite level (expressed in percent) of the envelope of the r-f carrier. As illustrated in Fig. 14.46 it is standard that black shall correspond to 75 percent and white to 10 percent of peak carrier. In the output of a diode detector which follows substantially the envelope of the i-f carrier, the relationship of halftone to definite levels is still preserved. If blocking capacitors are used in the subsequent video amplifiers, the d-c component is entirely lost and very low frequency components are attenuated.

Figure 16.23*b* shows the potential at the output of a diode detector representative of all-white and all-black scenes. Very little impedance is offered by C_1 to the abrupt wave fronts of the all-black signal, but as a charge is built up across C_1 the net potential of the grid must diminish until only horizontal synchronizing pulses are passed. The sudden drop in voltage applied to C_1 at the beginning of the all-white picture is transmitted faithfully with reference to magnitude as are subsequent synchronizing and blanking signals, but the absolute levels are lost by the gradual discharge of C_1, as illustrated in *c*. At *d* the voltage is shown subtracted by the capacitor which must be restored to the signal at the grid of the kinescope.

A commonly used d-c restorer is the diode shown in a typical connection in Fig. 16.23*a*. The action of this circuit may be described as follows. When the plate of the diode becomes positive with respect to the cathode during a horizontal synchronizing pulse, current flows until the capacitor C_2 is charged to such an extent that the plate is no longer positive. During the interval between synchronizing pulses, the potential of the plate drops slightly due to discharge current through R_2. If subsequent synchronizing peaks exceed others, more charge is accumulated on the capacitor; and if others are of lower amplitude, a

* See Friend, reference 11, and Schade, reference 12.

smaller charge, insufficient to replace that which has previously leaked away through R_2, is accumulated. The result is that the voltage of the condenser establishes a unidirectional voltage at the synchronizing peaks consisting of a series of small sawteeth occurring at the frequency

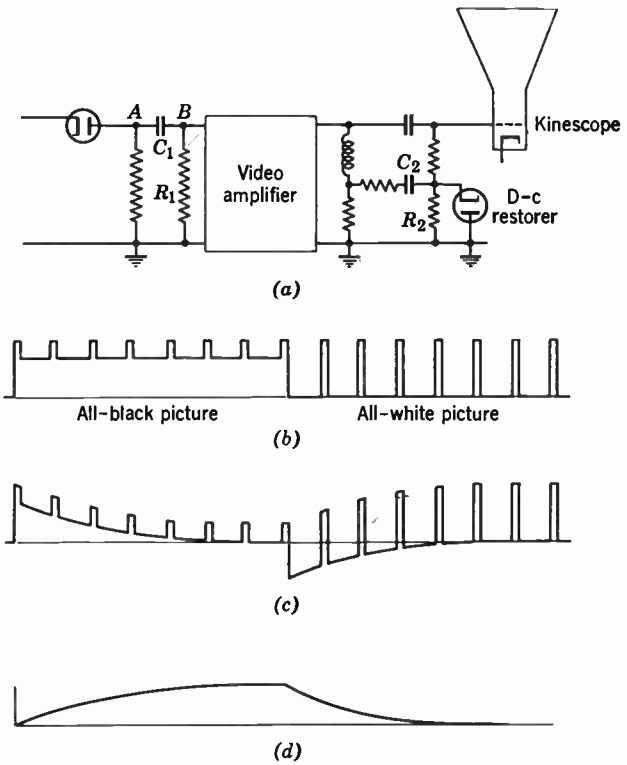


Fig. 16.23. Direct-Current Restoration: (*a*) Basic Circuit for D-C Restoration; (*b*) Output of Detector at Point *A*; (*c*) Potential at Point *B*; (*d*) Voltage Subtracted by C_1.

of horizontal scanning. This component, which is added to the impressed signal, is a good approximation of the component lost by the blocking capacitor C_1.

D-c restoration is unnecessary in the direct-coupled video amplifier which amplifies uniformly all video frequencies down to direct current. In such amplifiers relatively high cathode voltages may be resorted to so that the tubes are properly biased in the absence of blocking capacitors.

16.13 Sound Systems. Two types of sound reception are in general use, the older separate sound system and the intercarrier sound system. Both systems utilize the same frequency-modulated sound signal transmitted on a separate carrier spaced 4.5 megacycles above the picture carrier. The maximum deviation is 25 kilocycles. The audio modulation extends from 50 to 15,000 cycles per second. Preemphasis of the high frequencies is produced by an RC circuit having a time constant of 75 microseconds. The difference in treatment occurs after the heterodyning process in the converter stage of the receiver.

(*a*) *Separate Sound System.* The difference beat between the frequency of the local oscillator and the frequency-modulated sound signal is separated from the picture signal at a point in the initial stages of the picture i-f amplifier and passed to a high-gain narrow-band i-f amplifier (see Fig. 16.1). All the sidebands essential for satisfactory reconstruction of the sound modulation are included in a 100-kilocycle passband. After amplification the sound signal is amplitude limited for the removal of undesirable amplitude modulation and noise and then impressed upon a conventional frequency-modulated discriminator. The output of the discriminator at the a-f level is amplified and applied to the loudspeaker. The position of the sound carrier is held to a tolerance of 0.002 percent by a crystal-controlled oscillator in the sound transmitter. Hence the accuracy of the i-f sound frequency formed in the output of the converter is dependent upon the frequency stability of the local oscillator. Receivers using separate sound commonly include a fine tuning control in the frequency-determining circuits of the oscillator so that drifting of the circuit components during the warm-up period and at other times is compensated manually. The frequency should be held to within about 25 kilocycles of the correct value.

(*b*) *Intercarrier Sound System.* In the intercarrier system of sound reception both picture and sound i-f signals are amplified simultaneously in the picture i-f amplifier. A beat frequency difference between the picture i-f carrier on the one hand and the sound i-f carrier and sidebands on the other is generated by the picture detector (see Fig. 16.1). In effect, the sound i-f carrier and sidebands act as sidebands of the picture carrier. It is desirable that the sound carrier is less than the lowest amplitude of the picture carrier at all times so that the detector does not introduce distortion. In standard transmission, minimum picture carrier corresponding to white has been established as not less than 10 percent of peak carrier during the transmission of the synchronizing pulse. It has been determined that satisfactory sound

signal is obtained if the amplitude response of the i-f amplifier is about 3 percent in the vicinity of the sound carrier (Fig. 16.24).

Absorption traps properly coupled to one or two stages of the i-f amplifier is one method of achieving the required attenuation of the sound signal. The output of the detector contains video components in the range of 0 to 4 megacycles and a sound carrier and sidebands centered at the difference frequency of 4.5 megacycles. As indicated in Fig. 16.1, the sound signal is removed from the composite signal and impressed upon the grid of a tube whose limiting action removes the amplitude variations brought about by the varying amplitude of

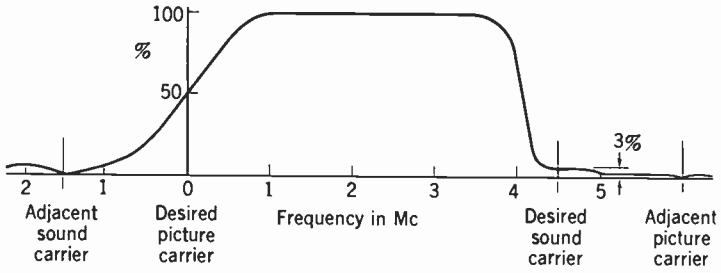


Fig. 16.24. Amplitude Response of the I-F Amplifier for Intercarrier Sound.

the picture carrier used in the demodulation process. After limiting, the signal becomes a true frequency-modulated carrier and is applied to a discriminator centered at 4.5 megacycles for demodulation.

A circuit for intercarrier sound is given in Fig. 16.25. The sound carrier and sidebands centered about a frequency of 4.5 megacycles and the video signal are impressed on the video output tube. On the plate side the impedance of the series tuned circuit is low only in the vicinity of the resonant frequency of 4.5 megacycles. Inductances L_1 and L_2 act as series and shunt compensation. The series circuit simultaneously removes the sound components from the wave form impressed upon the grid of the viewing tube and through the selective action of series resonance builds up appreciable sound voltage at the input of the limiter.

Beyond the point P the sound signal is limited and demodulated in the conventional manner of frequency-modulation reception.

The stability of the difference carrier frequency is dependent upon only the relative constancy of the picture and sound r-f carrier frequencies, which are controlled with great accuracy by crystal oscillators at the transmitter. Therefore, there is no possibility for appreciable movement of the sound carrier beyond the center of the dis-

criminator except as circuit constants of the demodulator may drift and change the tuning, a contingency which is unlikely in a well-designed receiver. The chief contribution of the intercarrier sound principle is that precise setting of the frequency of the local oscillator is not required for distortionless reproduction of sound. In typical cases the frequency is not critical over a range as great as 500 kilocycles.

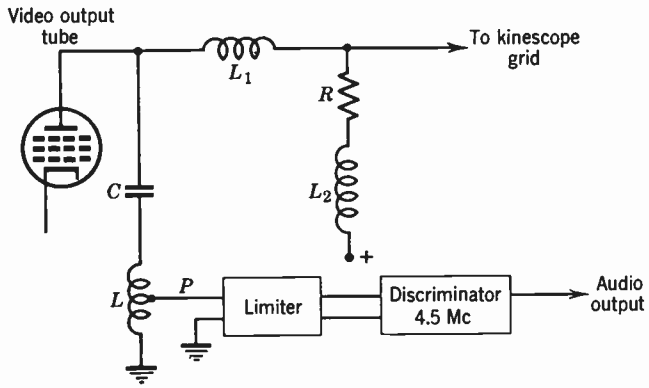


Fig. 16.25. Basic Circuit for Intercarrier Sound.

16.14 Transient Response of a Television Receiver. In a television picture of high fidelity the transitions in halftones corresponding to abrupt transitions in the subject are sharply defined, and there is a minimum of spurious effects such as white edges, ringing, and long black or white smears. A critical test wave form that permits convenient and valid assessment of the fidelity of a television receiver is the step function or its practical equivalent, the square wave. The abrupt transitions in a square wave represent abrupt transitions in picture halftones that involve the transmission of video frequencies of a few hundred kilocycles to 4 megacycles.

When the square-wave response is determined experimentally, a carrier corresponding to one of the standard picture carriers is light-modulated with a square wave and impressed on the antenna terminals of the receiver. The full double sideband signal is used since sideband attenuating means normally employed in television transmitters introduces phase and amplitude distortion. A square-wave period of about 10 microseconds allows the steady state of the response to be attained after each transition of the square wave. Hence, for analytical pur-

poses the response is substantially the same as the response to a step function.

The fidelity of the receiver is revealed by the wave shape of the output voltage of the video amplifier in response to the r-f test signal.* The response shown in Fig. 16.26 is intended to illustrate the types of distortion that occur in improperly aligned receivers.

A damped oscillatory component, owing to the abrupt cutoff displayed by the amplitude characteristic in the vicinity of the sound traps, occurs after the main rise in signal. Cutoff transients appear

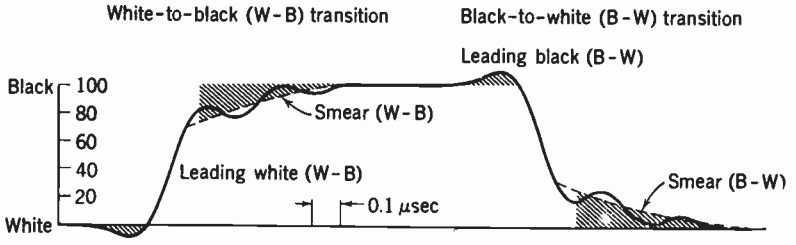


Fig. 16.26. Degradation in Square-Wave Response.

visually as striated patterns near the transition halftone. The frequency of oscillation tends to be approximately the same as the cutoff frequency. A less abrupt cutoff reduces the amplitude of ringing but simultaneously increases the rise time and leads to halftone transitions which are less sharp. Usually a compromise position is taken.

Failure of the axis of the ringing to coincide with the steady-state level at 100 percent response signifies that a smear component exists after each abrupt transition in halftone. A measure of this undesirable component is the deviation of the axis from the reference level either upward or downward, and the length of time required for the axis to attain this level.

Lack of symmetry of the sidebands about the carrier and too rapid roll-off through the carrier position lead to smearing. Operation of the carrier too far below the point of 50 percent response and an excessively rapid roll-off introduce leading white transients in a white to black transition.

The response illustrated in Fig. 16.27 would be termed acceptable since leading white fringes and smear are substantially absent and the frequency of ringing is high enough that the striations produced in the image are not resolved at a normal viewing distance. Overall ampli-

* See Kell and Fredendall, reference 13.

tude and phase characteristics are given as calculated from the transient response.*

Low-frequency square waves (60 cycles) have been used extensively for studying the low-frequency response of a receiver in the range of 60 to 100,000 cycles. The fidelity is judged by the extent of the deviation of the response over the flat top of the square wave as illustrated in Fig. 13.25. Excessive deviation may result from inadequate time

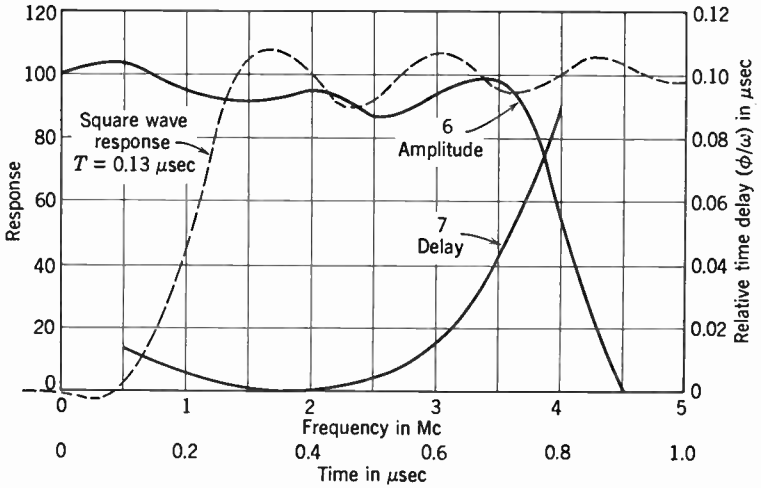


Fig. 16.27. Overall Transient Response of a Television Receiver and Corresponding Amplitude and Phase Characteristics.

constants in plate and screen filters or blocking capacitor and grid lead combinations.

16.15 Community Distribution Systems. The great majority of television viewers obtain satisfactory reception from an individual rooftop antenna or even an antenna built into the television set. There are, however, two greatly different groups who require a more complex installation. They are, on the one hand, dwellers in large apartment houses or hotels in urban centers and, on the other, residents of outlying communities separated by hills or mountain ranges from the nearest television station, or merely great distances away.

The apartment dweller is, in general, not free to employ an individual rooftop antenna because his landlord, for his own protection, will not permit him to do so. Furthermore, the interference between closely spaced antennas serving individual tenants would be such as to make

* See Bedford and Fredendall, reference 14.

satisfactory reception impossible. Built-in antennas are also likely to prove unsatisfactory for a large proportion of apartment dwellers, since the apartment building and adjoining structures may provide effective shields for the desired signal.

The solution of the problem is found in the installation of master antennas on the rooftop, feeding a distribution system with signal outlets in the several apartments.* If the number of outlets is small (e.g., less than six) and the signal strength at roof level is high, a single broadband antenna may be employed, with a 300-ohm transmission line feeding the several outlets. The line is terminated in its characteristic resistance. Resistive networks at each outlet serve, furthermore, to isolate the individual receivers from each other (Fig. 16.28).

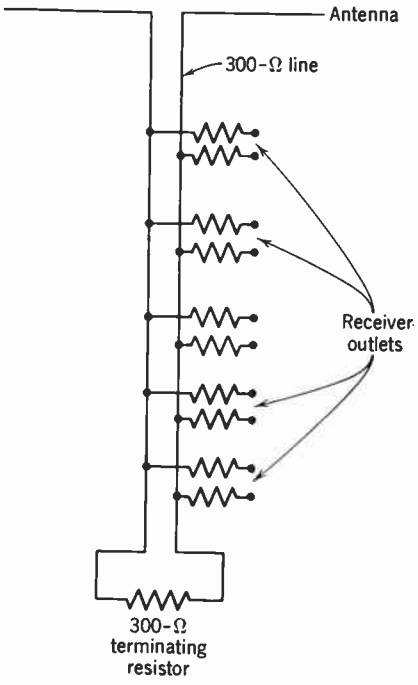


Fig. 16.28. Simple Passive Distribution System.

This simple passive distribution system, obviously, is practical only if there is signal strength to spare. Furthermore, the employment of a single antenna makes it impossible to choose a location and orientation which will minimize "ghosts" from reflected signals for all channels simultaneously. Accordingly, more complex distribution systems are favored for all sizable apartment house and hotel installations.

In master antenna systems such as the RCA Antenaplex System individual antennas are provided for the locally assigned television channels, and eventually for the frequency-modulated, short-wave, and broadcast band as well. Each antenna is placed and oriented so as to provide a strong single-path signal for the selected channel. The outputs of the antennas are fed by 300-ohm cable to individual tuned r-f amplifiers with gains of the order of 70 decibels. This gain is sufficient to compensate losses in distribution cables and in resistive networks placed at the individual outlets to prevent cross-feed between

* See reference 15 and Kamen and Dorf, reference 16.

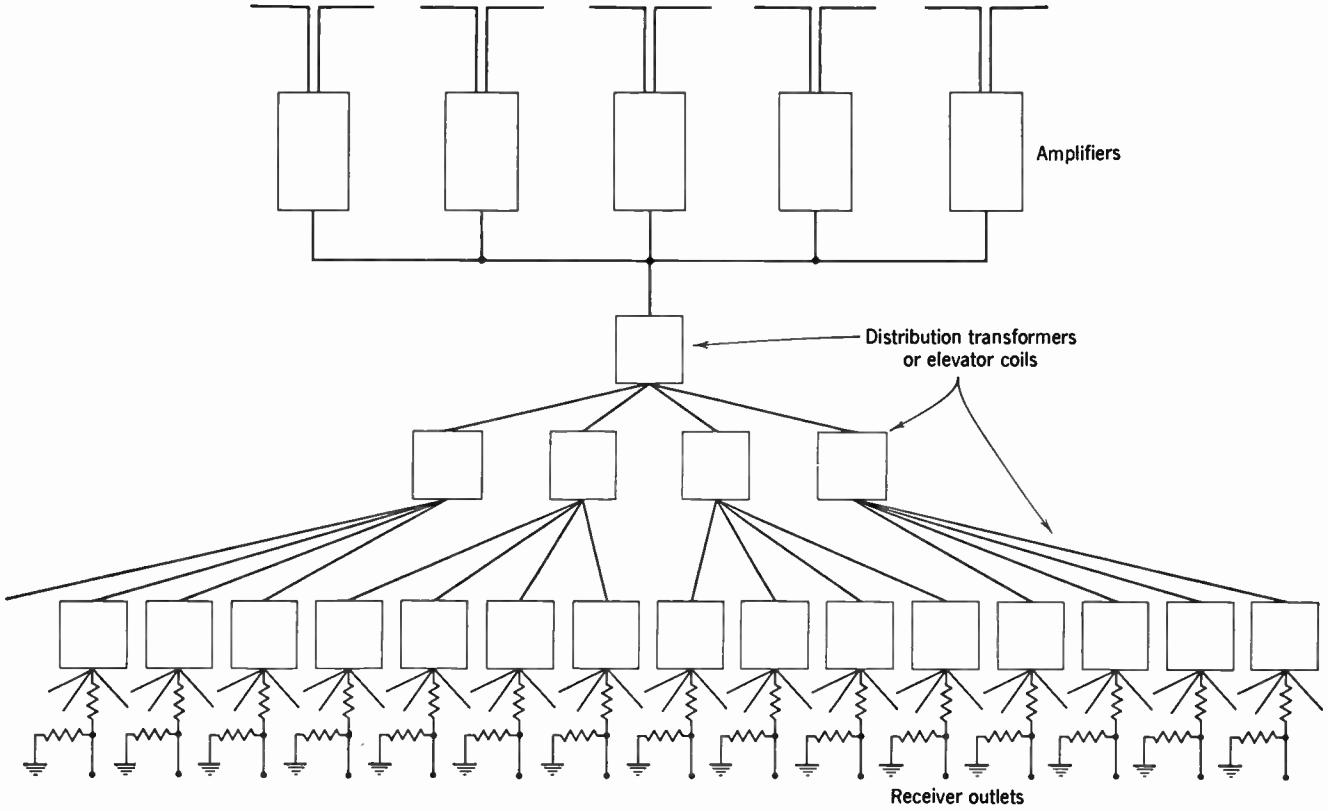


Fig. 16.29. Distribution System for Structure with Broad Horizontal Base.

receivers. Attenuator pads are also inserted in the cables between the antennas and the amplifiers. These pads serve to equalize the signal strength on the several channels.

The arrangement of the distribution system depends on the type of structure which is to be served. For a structure with broad horizontal base a system is appropriate which spreads out fanwise with the aid of a sequence of distribution transformers or equivalent elevator trans-

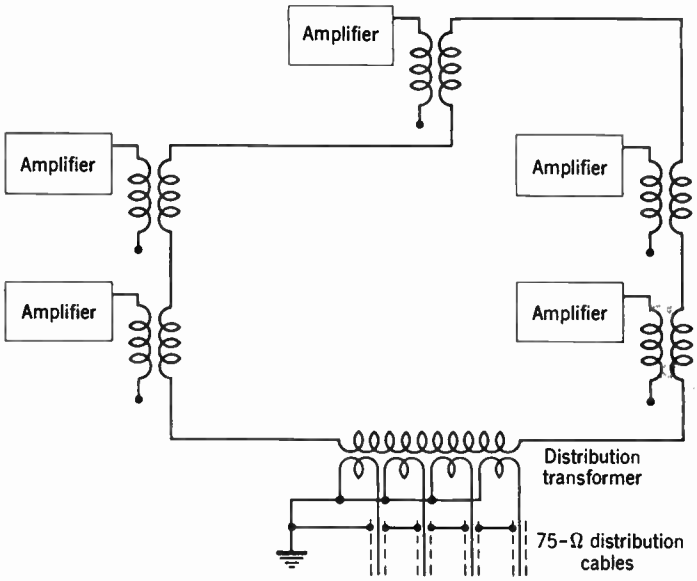


Fig. 16.30. Method of Adding Channel Amplifier Outputs.

former coils, which derive four 75-ohm outputs from a single 75-ohm input (Fig. 16.29). The individual distribution transformer is an r-f transformer with a single primary and four secondary coils, flat over a wide frequency range, such as 50 to 220 megacycles. For the unit following the amplifiers, the outputs of the amplifiers may be added by direct series connection with the primary of the transformer (Fig. 16.30).

An elevator coil distribution unit is illustrated in Fig. 16.31. A match is obtained between the 75-ohm input cable and the four 75-ohm output cables, since the input impedance of each 1:1 transformer is 150 ohms.

For a tall, narrow structure a single signal-splitting unit following the antenna amplifiers is employed, and twenty or more connections are made to each feeder cable (Fig. 16.32). The isolating resistances

for each outlet are placed directly in contact with the inner conductor of the feeder cable, so as to minimize signal reflections at this point.

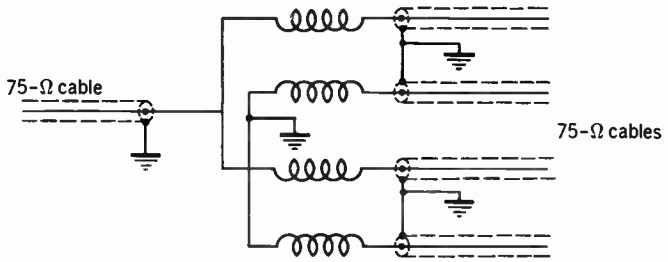


Fig. 16.31. Employment of Elevator Transformer Coils for Signal Splitting.

The resistances are decreased with increasing distance from the distribution center, so as to equalize the signal level at successive outlets. Normally separate outlets are provided for television and frequency-

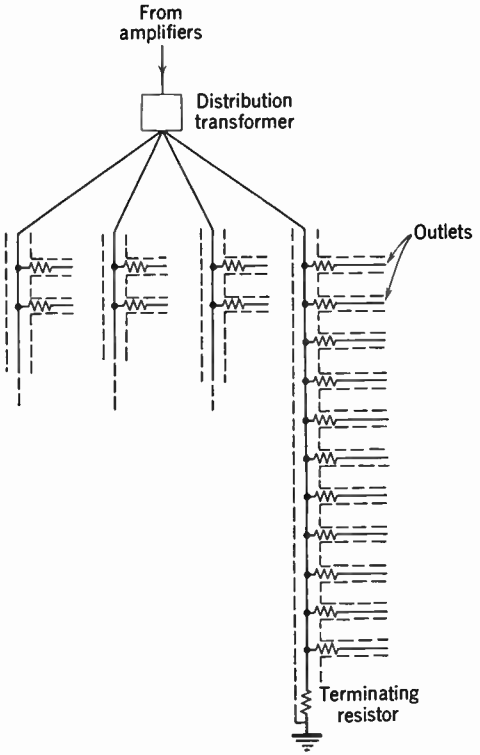


Fig. 16.32. Distribution System for Tall, Narrow Structure.

modulated signals and for broadcast and short-wave signals. The feeder cables, finally, are terminated in their characteristic impedance.

Master antenna systems, with distribution to individual dwellings by coaxial cable, have also proved valuable in providing high-quality television reception for outlying communities, isolated from broadcasting centers by geographical factors. However, here, both the magnitude of the antenna structure and the length of cable required tend to be of a different order. In flat areas antenna towers 500 feet in height may have to be constructed to approach line-of-sight connection with the nearest television transmitter antennas. In mountainous regions, an antenna structure of smaller height, placed on a near-by mountain top, will be satisfactory. At the same time, the greater distance between the antenna and the service area will demand a greater length of cable. All this results in relatively high installation costs, generally exceeding 100 dollars per subscriber. In addition, maintenance requires a service charge of the order of 4 dollars monthly. Even so, by March, 1953, some 70,000 to 85,000 homes had availed themselves of the service provided by community television systems.*

Although community television systems are to be found in all parts of the United States, they were first introduced and enjoy greatest popularity in the mountainous regions of Pennsylvania.† The community television system of Pottsville and Minersville, serving 1650 subscribers (1952), is a good example. Pottsville, ringed by mountains, is 75 air miles from Philadelphia. To obtain reception from Philadelphia stations operating on Channels 3, 6, and 10 a 150-foot twin tower was erected atop 1400-foot high Sharp Mountain. Three selective stacked antennas tuned to these channels are mounted on the top of the tower (Fig. 16.33). The signal from each is amplified in a separate preamplifier. Signals from Channels 6 and 10, combined by a distribution amplifier, as well as the signal from Channel 3 are conducted by two coaxial cables to amplifiers at the foot of the tower. These amplifiers are provided with automatic gain control to equalize the signal strength on the several channels. Furthermore, signal converters shift the Channel 10 signal to Channel 4 and the Channel 3 signal to Channel 2 (Fig. 16.34). By reducing the signal frequencies in this manner to the lowest values consistent with operation on no adjacent channels, cable attenuation and, hence, the number of line amplifiers required are minimized. The converters employ, in general,

* See reference 17.

† See Carroll, reference 18, and Lucas, reference 19.

crystal-controlled local oscillators and frequency multipliers and a vacuum-tube or crystal-mixer stage. Two converters are required for the shift from Channel 3 to Channel 2, since the rejection of the unwanted signals is impracticable when the frequency shift is of the same order as the bandwidth of the desired signal.

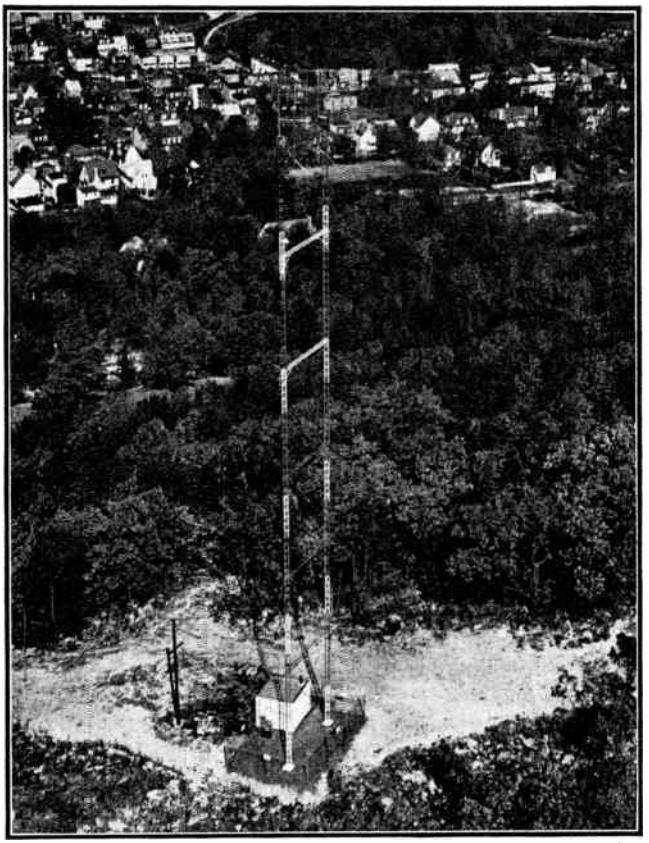


Fig. 16.33. Antenna Installation for Community Television System. (Courtesy of Pottsville Transvideo Corporation.)

The antenna and line amplifiers are essentially similar. They may consist either of a single broadband amplifier utilizing the principle of distributed amplification (see section 15.7) or, more commonly, of three parallel amplifiers, each stagger-tuned with a passband of 6 megacycles and a gain of 50 to 60 decibels. The second solution is generally preferred, since it permits compensation for the different attenuations of the several channels by the cable.

With the parallel amplifier system, the amplifier outputs are recombined, e.g., by a network such as that shown in Fig. 16.35. The use of series-resonant circuits in each output and of connecting half-wave lines prevents signal loss in the load circuits of the paralleled amplifiers.

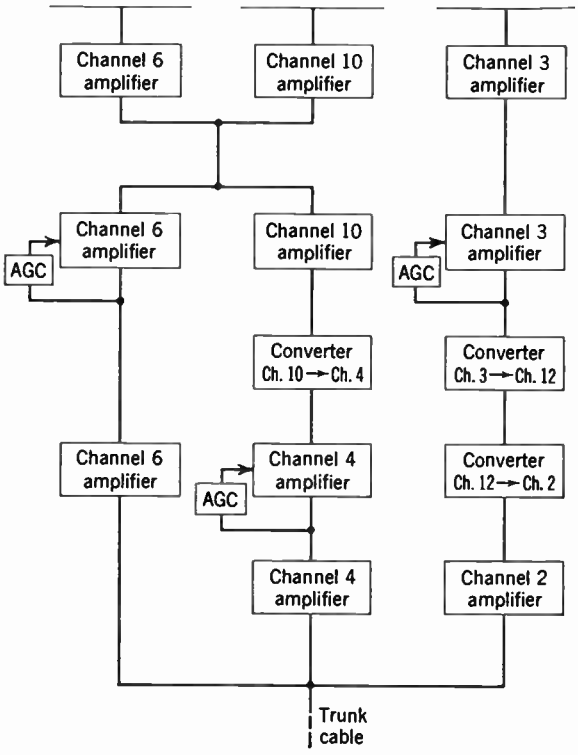


Fig. 16.34. Block Diagram of Antenna Tower Installation for Community Television System. (Carroll, reference 18.)

As the trunk line, supported by power line poles, enters the community which it serves, it is split up into a series of subtrunk lines, and these, in turn, are split up into feeder lines. The splitting may be carried out either by passive or by active networks. The elevator transformer coils shown in Fig. 16.31 represent a suitable passive network. An active network for coupling subtrunk or feeder lines to a trunk line is shown in Fig. 16.36. Each coupled line is fed by a pentode bridged across the trunk line. House connections, finally, are commonly made by tapping a feeder line with an isolation resistance or

capacitance, in the same manner as in the apartment house system shown in Fig. 16.32.

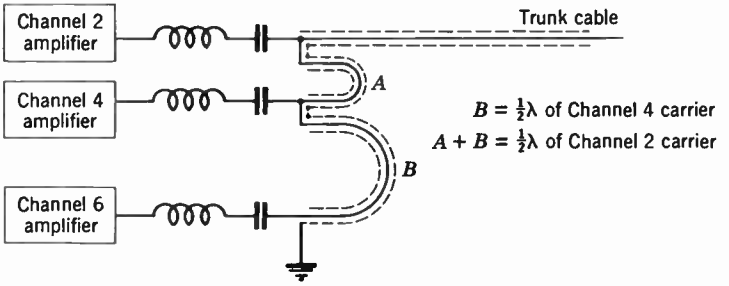


Fig. 16.35. Signal-Adding Network.

The lines on the antenna tower, as well as the trunk and subtrunk lines, generally employ a ½-inch, 75-ohm coaxial cable (RG-11/U). It has a loss of 1.5 decibels per 100 feet at 50 megacycles and a loss

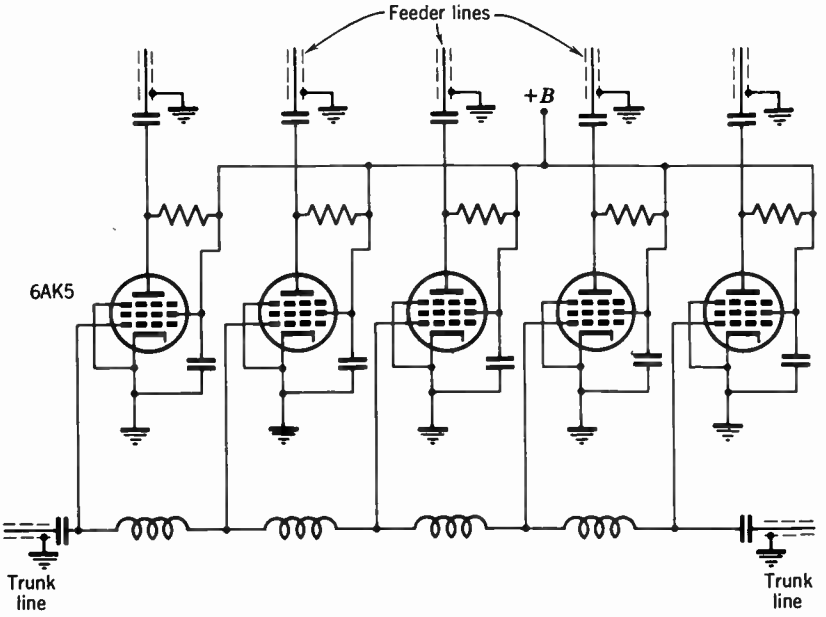


Fig. 16.36. Active Network for Coupling Feeder Lines to Trunk Line.

of 2.9 decibels per hundred feet at 200 megacycles; the loss increases approximately as the square root of the frequency. For the house connections it is customary to use ¼-inch, 73-ohm coaxial cable (RG-

59/U) with a loss of 2.5 decibels per 100 feet at 50 megacycles. Community distribution systems generally aim to provide a minimum signal level of the order of 200 microvolts at the subscriber's set.

REFERENCES

1. S. W. Seeley, "Effect of the Receiving Antennas on Television Reception Fidelity," *RCA Rev.*, Vol. 2, pp. 433–441, 1938.
2. A. B. Bailey, *TV and Other Receiving Antennas*, John F. Rider, Inc., New York.
3. R. M. Cohen, "Use of the New Low-Noise Twin Triode in Television Tuners," *RCA Rev.*, Vol. 12, pp. 3–25, 1951.
4. T. Murakami, "AVHF-UHF Television Turret Tuner," *RCA Rev.*, Vol. 14, pp. 318–340, 1953.
5. A. Newton, "Analysis of UHF Tuner Design," *Electronics*, Vol. 26, pp. 106–111, 1953.
6. E. W. Herold, "The Operation of Frequency Converters and Mixers for Superheterodyne Reception," *Proc. I.R.E.*, Vol. 30, pp. 84–103, 1942.
7. R. B. Dome, *Television Principles*, McGraw-Hill, New York, 1951.
8. H. Wallman, "Stagger-tuned Amplifier Design," *Electronics*, Vol. 21, pp. 100–105, 1948.
9. K. R. Wendt and A. C. Schroeder, "Automatic Gain Controls for Television Receivers," *RCA Rev.*, Vol. 9, pp. 373–393, 1948.
10. K. R. Wendt and G. L. Fredendall, "Automatic Frequency and Phase Controlled Synchronization," *Proc. I.R.E.*, Vol. 31, pp. 7–15, 1943.
11. A. W. Friend, "Television Deflection Circuits," *RCA Rev.*, Vol. 8, pp. 98–138, 1947.
12. O. H. Schade, "Characteristics of High-Efficiency Deflection and High-Voltage Supply Systems for Kinescopes," *RCA Rev.*, Vol. 11, pp. 5–37, 1950.
13. R. D. Kell and G. L. Fredendall, "Standardization of the Transient Response of Television Transmitters," *RCA Rev.*, Vol. 10, pp. 17–34, 1949.
14. A. V. Bedford and G. L. Fredendall, "Analysis, Synthesis, and Evaluation of the Transient Response of Television Apparatus," *Proc. I.R.E.*, Vol. 30, pp. 440–457, 1942.
15. "Television Antennas for Apartments," *Electronics*, Vol. 20, pp. 96–102, May, 1947.
16. I. Kamen and R. H. Dorf, *TV Master Antenna Systems*, John F. Rider, New York, 1951.
17. "Community Television Continues to Expand," *Electronics*, Vol. 26, pp. 6, 8, March, 1953.
18. J. M. Carroll, "Community Antennas Bring TV to Fringe Areas," *Electronics*, Vol. 25, pp. 106–111, December, 1952.
19. E. D. Lucas, Jr., "How Television Came to Panther Valley," *Radio and Television News*, Vol. 45, pp. 31–34, 106–111, March, 1951.

PART 4

Color Television, Industrial Television, and Television Systems

17 Color Television Fundamentals

17.1 Introduction. That color should be added to television is as inevitable as that it should have been added to ordinary photography and to moving pictures. It nevertheless presents a serious technical challenge, and the way in which this challenge has been met is the subject of this and the next two chapters. In addition, color presents many new problems to the program producer and director. The value of the addition of color to natural scenes needs no comment. The additional information obtainable for many sporting events, newsworthy spot pickup, etc., is obviously helpful to the viewer. Also the entertainment and gaiety of many variety shows are enhanced by color. By the same token, it is much more difficult to produce truly artistic pictures in color than in black-and-white. This has been true with moving pictures to the extent that nearly all the great artistic pictures are in black-and-white. Since retakes and interruptions to continuity cannot be tolerated in television, artistic presentation is even more difficult in this new medium.

In earlier chapters, it has been pointed out that television is practical only because of certain characteristics of human vision. The problem of color is similarly simplified because of limitations of both physiological and psychological mechanisms involved in the perception of color. Principal among these limitations which permit the simplification of the color problem are the following. The eye does not resolve the spectrum over the visual range of radiation, but rather synthesizes the sensation of color out of three primary spectral responses. Detail discrimination in color is much less accurate than for brightness variation and, finally, the speed of response of the eye to changes in color is much slower than to changes in brightness.

This chapter discusses the principles of color television in general qualitative terms. It attempts to present the problem and its solution with a minimum of mathematics. Chapter 18 continues the discussion, but takes it up from a more theoretical point of view. Chapter 19, finally, describes the components of a practical system of color tele-

vision. Whereas the present chapter will treat the subject in sufficient detail to give the reader a grounding in the subject, Chapters 18 and 19 may be of value to the specialist and the research worker as a background for advancing the field.

The second section of this chapter deals with the general problem of color, introducing the concepts of tricolor visual stimulation, with its representation in terms of a color triangle. It is followed by a discussion of the two basic systems of color television, the simultaneous and the sequential systems. Separate sections are devoted to various simultaneous and sequential systems, including the three-channel simultaneous system and field, line, and element sequential systems. A simultaneous system derived from the element-sequential system is treated in considerable detail in several sections. The emphasis is placed on this method of color television because it leads to a completely compatible transmission having the full detail of ordinary black-and-white television. In other words, when the transmitted signal is received on an ordinary monochrome receiver, a good black-and-white picture is obtained. The same signal received on a color receiver, of course, reproduces the picture in full color and detail.

The final sections of the chapter discuss the essential components of the terminal equipment required to transmit and receive color television.

17.2 Color Vision. The light from any surface, whether due to its own luminosity or the result of reflecting radiation from some other primary source, consists of radiant energy distributed as a more or less continuous function of wavelength. Therefore, an exact description of the radiation from a colored scene would be a three-dimensional function $R = F(x, y, \lambda)$ varying in time. A television system which required the transmission of the exact description would be technically virtually impossible.

Earlier chapters have described the reduction of the variables x, y, and t into discontinuous elements by the principle of scanning, which takes advantage of certain limitations of human vision. In color television, this simplification is retained, and to it is added the reduction of λ to three discrete components made possible by the nature of color vision.

The sensation of color relates to the spectral quality of the light being perceived. However, it is not a complete description of the spectral distribution of this radiation. As a result of many experiments and tests over a period of years, it has been established that color perception may be described in terms of the stimulation of three types of

light-sensitive receptors in the eye, each having a different spectral response. This tricolor theory of color vision was first proposed by Helmholtz in 1868 and, since then, has been repeatedly verified with minor modifications.

As a consequence of the trichromatic nature of color vision, the color sensation of any spectral distribution can be reproduced by means of three monochromatic sources (e.g., red, green, and blue) which stimulate the three receptors to the same extent as the original distribution. The tristimulus system of color vision and its representation has been the subject of considerable quantitative study. Before taking up the rigorous representation of this system of vision, it might be advantageous to consider it from a qualitative nonrigorous point of view.

Assume that three fairly narrow spectral bands, red (r), green (g), and blue (b), have been selected as the artificial sources with which to match nearly any spectral distribution. A particular distribution $A(\lambda)$ might be matched by a_r, b_g, c_b, where a, b, and c are the intensities of the color bands indicated by the subscripts. If the distribution $A(\lambda)$ is obtained by a white light source illuminating a colored surface or a colored source illuminating a white surface, the color does not change if the source is moved closer or farther away from the surface. However, the surface becomes brighter or less bright. In other words, although the magnitudes of a, b, and c change, the ratio $a:b:c$ is constant. This then distinguishes between brightness, luminosity, or luminance and color, chromaticity, or chrominance.

A special case of matching a spectral distribution is the reproduction of "white" by means of the three color bands r, g, and b. A properly chosen ratio of the three color components, which for the present will be designated as l_r, m_g, and n_b, will be indistinguishable from a continuous spectral distribution with a constant energy per unit wavelength over the visible band, which by definition is "equi-energy white." Again it is sufficient to maintain the ratio $l:m:n$ to give the impression of white.

If an arbitrary surface is again considered (which can be matched by a_r, b_g, c_b), it will be possible in general to describe the color as white plus two color components. In other words, a constant s can be selected such that one of the components

$$(a_r - s \cdot l_r), \quad (b_g - s \cdot m_g), \quad \text{or} \quad (c_b - s \cdot n_b)$$

vanishes, leaving the other components positive or zero. Thus any spectral distribution can be described as two color components plus white. Saturation of a color is the term used to denote the degree

with which a given color approaches a spectrally pure color, or a pure purple formed of a mixture of spectral blue and red. A saturated color is one which contains no white ($s = 0$). The color components which are required to be added to the white define the "hue" of the color.

The tristimulus concept can be put on a more quantitative basis by the following argument, which leads to the color triangle standardized by the International Commission on Illumination in 1931. The bright-

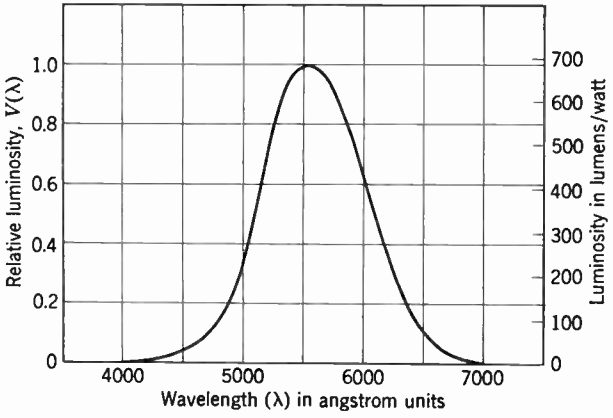


Fig. 17.1. The Luminosity Function.

ness or luminance of a surface emitting light in accordance with the energy distribution function $E(\lambda)$ is given by the integral relation:

$$L = K\int_0^\infty E(\lambda)V(\lambda)\,d\lambda \tag{17.1}$$

where $V(\lambda)$ is the average visibility or luminosity function (Fig. 17.1) and K is a constant giving the photometric equivalence of electromagnetic radiant energy. The visibility function describes the average sensitivity of the human eye as a function of wavelength. It has been derived from measurements on a large number of observers and standardized in 1924 by the International Commission on Illumination. Two surfaces $E_1(\lambda)$ and $E_2(\lambda)$ are said to have equal brightness or luminance if $L_1 = L_2$.

Three analogous functions $\bar{x}(\lambda)$, $\bar{y}(\lambda)$ and $\bar{z}(\lambda)$, the tristimulus values, have been defined for the tricolor response of the eye. These functions, which are related to the responses of the three color receptors of the eye, have been evaluated in detail by extensive measurements on a large number of observers. These tristimulus values are so defined as to be everywhere positive. They are illustrated graphically in Fig. 17.2. By means of these functions, the color of any luminous surface with spectral

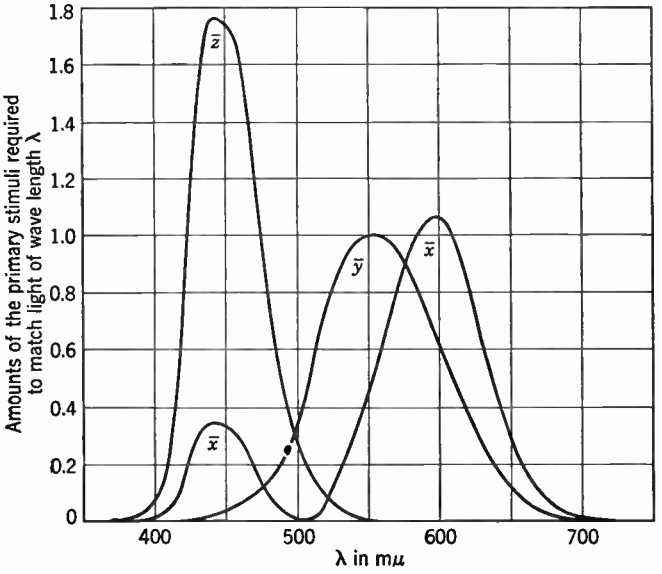


Fig. 17.2. CIE Color Mixture Curves. (From Zworykin and Ramberg, *Photoelectricity*, Wiley, 1949.)

distribution $E(\lambda)$ can be described by the three quantities X, Y, and Z, where

$$X = K\int E(\lambda)\bar{x}(\lambda)\, d\lambda$$

$$Y = K\int E(\lambda)\bar{y}(\lambda)\, d\lambda$$

$$Z = K\int E(\lambda)\bar{z}(\lambda)\, d\lambda$$

Since the color or chromaticity of the surface does not change if it is simply made brighter or less bright by changing X, Y, Z by a multiplying factor, they may be normalized as follows:

$$x = \frac{X}{X + Y + Z}$$

$$y = \frac{Y}{X + Y + Z}$$

$$z = \frac{Z}{X + Y + Z}$$

so that $x + y + z = 1$.

Obviously, with this scheme of representation, all possible colors or chromaticities can be represented by two variables, for example, x and y since $z = 1 - x - y$. Because of the way in which $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ were defined, x and y (and also z) must at all times be positive.

Since only two variables are required to represent all chromaticities, it is possible to represent them on a plane. If x and y are plotted in a rectangular coordinate system, all colors must be within the triangle determined by the $x(y = 0)$ and $y(x = 0)$ axis and the line $x + y = 1$.

The pure spectral (monochromatic) colors can be located in the color triangle by a direct application of the tristimulus value curves. Thus the position of the chromaticity characterized by the single wavelength λ_k is given by

$$x_k = \frac{\bar{x}(\lambda_k)}{\bar{x}(\lambda_k) + \bar{y}(\lambda_k) + \bar{z}(\lambda_k)}$$

$$y_k = \frac{\bar{y}(\lambda_k)}{\bar{x}(\lambda_k) + \bar{y}(\lambda_k) + \bar{z}(\lambda_k)}$$

The location of all the spectral colors of the visible region of the spectrum (from 3800 Å to 7500 Å) is indicated in a color triangle as shown in Fig. 17.3.* Along the line AA' of this figure lie the purples which are not spectral colors but are complementary to the spectral colors from about 5700 Å to 4920 Å.

One of the properties of the color triangle is that the color which results from the mixing of two colors (additively) lies on a straight line connecting the two color points. Thus, if a white surface is illuminated by two sources, one of which when illuminating the surface alone can be represented by the color point a and the other by the color point b, the color resulting from the mixture must lie on the line ab.

On the color triangle (Fig. 17.3) the location of the color daylight white is indicated by C. Complementary colors are a pair of colors which, when added together in proper proportions, produce white. Thus, 6000 Å (d) and 4870 Å (e) are complementary.

As has already been noted, color in television is obtained by the admixture of light from three different color sources. These sources may be a white phosphor seen through the filters of a color disk, or the colored phosphor elements of a three-color viewing tube. Each of the three colors used can be represented by one point on the color tri-

* On the chart, as elsewhere, wavelengths are given in millimicrons instead of angstroms (1 mμ = 10 Å).

angle. Any color which lies within the triangle determined by the three color points can be obtained by adding various relative amounts of the three colors. The points R, G, and B on Fig. 17.3 represent the three color phosphors used in a practical tricolor kinescope. The tri-

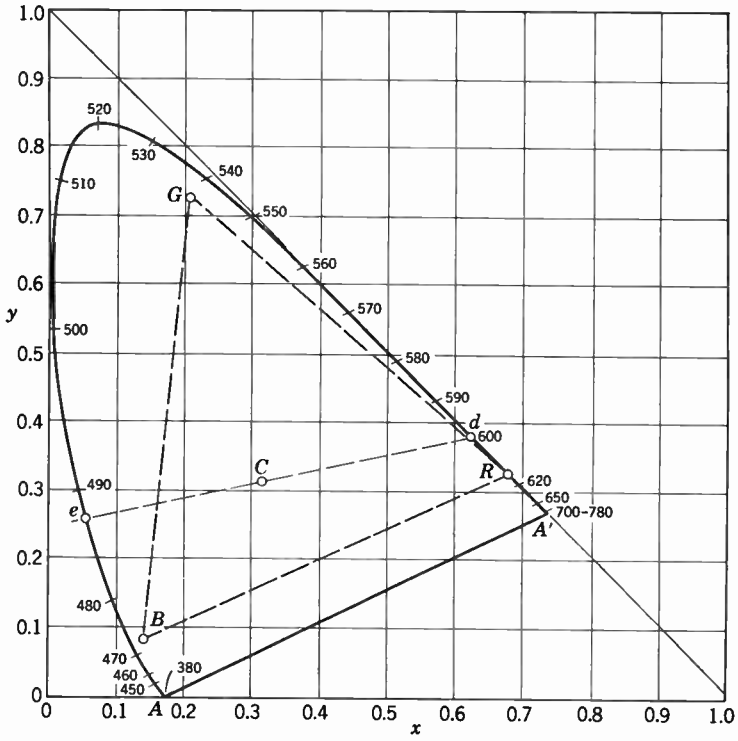


Fig. 17.3. The Color Triangle.

angle bounding all the colors which can be reproduced by such a tube is also shown. It will be noted that nearly all colors except those having a very high degree of saturation, and consequently rarely found in natural objects, can be reproduced.

The luminosities of the three colors used in the kinescope are very different when mixed in proportions to produce white. The luminosities can be determined from the three integrals:

$$L_r = K \int E_r(\lambda) V(\lambda)\, d\lambda$$

$$L_g = K\int E_g(\lambda)V(\lambda)\,d\lambda$$

$$L_b = K\int E_b(\lambda)V(\lambda)\,d\lambda$$

In the particular example considered, the ratios are found to be

$$L_g{:}L_r{:}L_b = 1{:}0.5{:}0.15$$

Still on the subject of color vision, a few other facts which are germane to the problem should be discussed. The resolving power of the eye for differences in chromaticity is very much less than for differences in luminance. In other words, the eye will distinguish between two light areas separated by a dark area where the angle of separation is about a minute of angle, but, in order to resolve two areas of one chromaticity separated by a second color where the luminance is the same, the angle must be considerably greater.

Similarly, the time resolution of the eye for chroma or hue changes is much lower than for changes in luminance. As has been pointed out in Chapter 5, in order to avoid flicker at brightness levels which have been found desirable for television, the field repetition rate must be 60 per second. Flicker due to changes in chromaticity where there is no change in luminance becomes perceptible at a very much lower frequency threshold. In fact, an alternation of only a few cycles per second is sufficient completely to eliminate color flicker. The difference in threshold is the basis of the flicker photometer which compares the luminosity of a colored surface with that of a white standard whose brightness can be controlled. The point at which flicker ceases when the two surfaces are viewed in sequence at an alternation rate of a few cycles per second indicates equality of brightness of the surfaces.

The significance of the lower space and time resolution of the eye for color changes is that color reproduction does not need to be carried into the fine detail of a picture as long as luminance changes are faithfully preserved. Also, spurious alternations in chromaticity, even if at a lower repetition rate than field frequency, will not be visible if unaccompanied by luminance changes.

17.3 Basic Color Television Systems. All color television systems which have been seriously proposed so far may be divided into two broad classes, simultaneous systems and sequential systems. Simultaneous systems are three-color systems in which the red, green, and blue pictures are each transmitted and reconstructed at the same time.

The three pictures are superimposed in registry by electrical or optical means. With sequential systems, as the name implies, the three colors are transmitted at separate times and the superposition is accomplished by the persistence of vision.

There are three principal classes of sequential systems characterized by the fraction of the picture (field, line, or picture element) which is used as the unit of division. All such systems may be represented by the block diagram shown in Fig. 17.4. The color camera $A_r A_g A_b$ continuously produces video signals corresponding to the three color com-

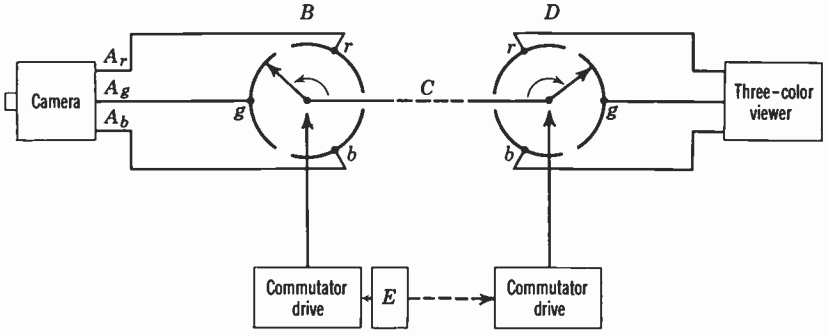


Fig. 17.4. Generalized Sequential Three-Color System.

ponents of the image. Preamplifiers may be associated with these three color channels to raise the signal level above the switching noise of the commutator B where necessary. The commutator connects the communication channel C with the three color components in sequence. The commutator may be a rotating color disk in front of the camera, a mechanically operated switch, or an electron tube commutator. It may even be incorporated in the camera pickup tube. The output from the communication channel feeds a commutator D at the receiver which is synchronized with B through the synchronizing circuit E. The output from the commutator supplies the three color components to the three-color viewing device. Again the commutator may be electrical or may be a color disk in front of a viewing tube. It might be pointed out that, in principle, the commutator may have more than three positions and be so connected that one or more colors is repeated during one cycle. For example, a possible sequence for these commutators might be $g\ r\ g\ b\ g\ r$, etc.

With this block representation, the conventional field sequential color system requires that the commutators have the three positions $r\ g\ b$ and traverse one cycle every three scanning fields. Actually, in

this case, optical commutators are used for both the transmitter and the receiver. A more complete description of this system is given in the next section.

If the commutator completes its cycle in the period corresponding to that required for three horizontal sweeps, the system becomes a line sequential system. This system is a compromise system which has most of the weaknesses of the field sequential and the element sequential systems and very few advantages unique to itself. It is looked upon with the least favor of any of the sequential systems, and it is extremely doubtful that it will ever be widely used. It will be discussed somewhat further, however, in section 17.6.

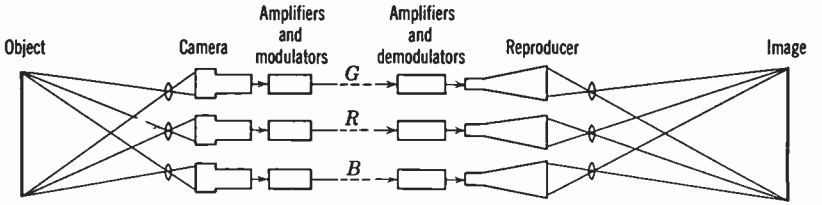


Fig. 17.5. Three-Channel Simultaneous System.

To obtain element-sequential color reproduction, the commutator must complete a cycle in the time required for the beam to traverse three picture elements. This system can be greatly improved by altering it in such a way as to form the basis of the technically most attractive form of color system. Several sections will be devoted to these modified element-sequential systems and to the simultaneous system derived from them.

A schematic diagram of a typical simultaneous system is shown in Fig. 17.5. Here the system is represented by three completely separate television chains, one for each color. Superposition of the three pictures is accomplished optically. It obviously could be accomplished as well or better by means of a tricolor kinescope.

If the full color detail is to be transmitted for each of the three pictures, the bandwidth in the r-f spectrum needed is three times that for a monochrome picture. This is the same whether three carriers and three channels are employed or whether one carrier and one very wide channel carries the information. By taking advantage of the limited color resolution of the eye noted in the previous section, an economy in bandwidth can be effected. For example, the separate color information of only the lower 0.3 megacycle of the video signal for each channel might be transmitted and a mixed high frequency signal (cor-

responding to a monochrome signal) for the remainder of the video signal from 0.3 to 4.1 megacycles could be common to the three channels. With such a compromise, only 4.7-megacycle bandwidth would be needed to transmit a picture which had detail equivalent to a black-and-white 4.1-megacycle bandwidth picture.

Actually, as will be discussed in some detail later, this principle of mixed highs is employed in the simultaneous compatible system derived from an application of the element-sequential principle.

17.4 A Three-Channel Simultaneous System. This section is a brief account of a three-channel color system tested at RCA Labora-

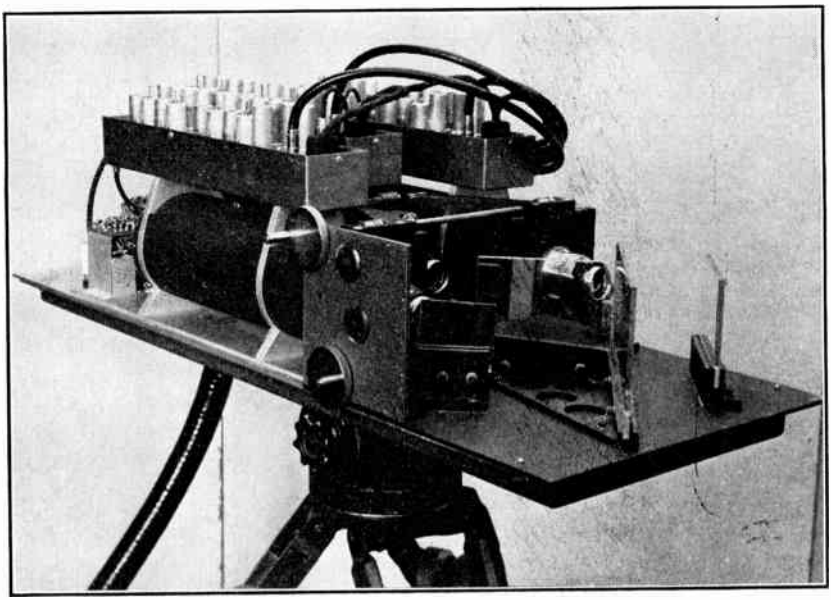


Fig. 17.6. Three-Color Simultaneous Camera (Cover Removed).

tories as part of a general investigation of color television. The system itself is not now suitable for television broadcasting because of the great bandwidth required. However, a great deal of important information of general applicability was obtained from the experiment.

The camera consisted of three image orthicons in very close mechanical and optical register. The optical images focused on the targets of the three tubes were obtained from a single objective with three dichroic mirrors following it to divide the image into its three color components and to direct these separation images onto their respective targets. Figure 17.6 illustrates this camera. Very great care was

taken to insure stability and linearity of deflection in the three tubes, since electrical register of the image is fully as important as optical register. Each camera channel had its own preamplifier and video amplifier into which were inserted the necessary blanking and shading signals.

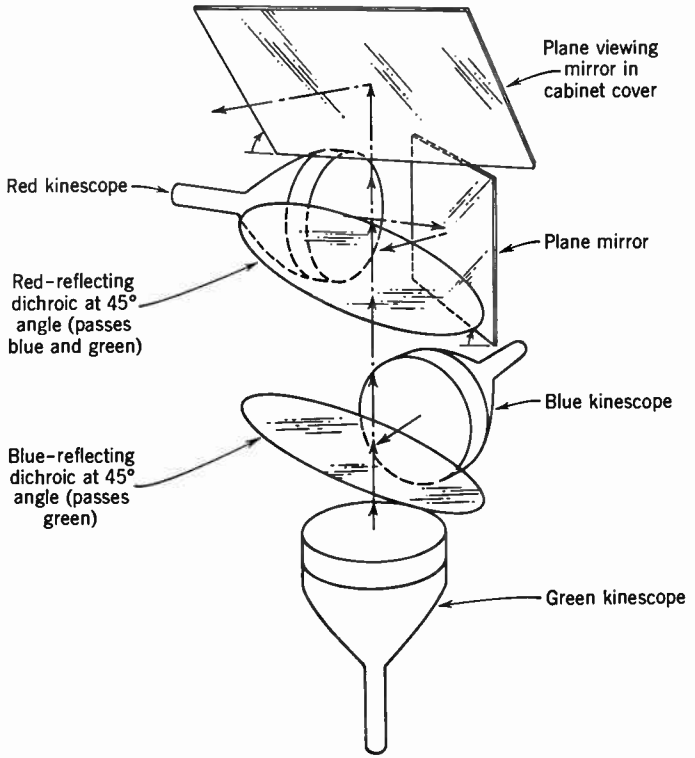


Fig. 17.7. Three-Tube Viewing System.

Various amplifiers, both video and intermediate frequency, could be inserted into the three channels to permit a variety of tests of amplifier characteristic requirements. Complete separation of the three channels was maintained through the system up to and including the viewing tube.

The viewing arrangement, illustrated in Fig. 17.7, used three white screen kinescopes which were viewed through a pair of dichroic mirrors. The arrangement of the system was such that one of the tubes (carrying the green pictures) was viewed directly, while the other two were seen reflected in the mirrors. In one system, two tubes were

mounted so that their axes were parallel to each other but at right angles to the first. One tube faced upward, the other downward. Several tube arrangements were tested in receivers with equal success. Again great pains were taken to make certain that the linearity and the stability of the deflection were as high as possible.

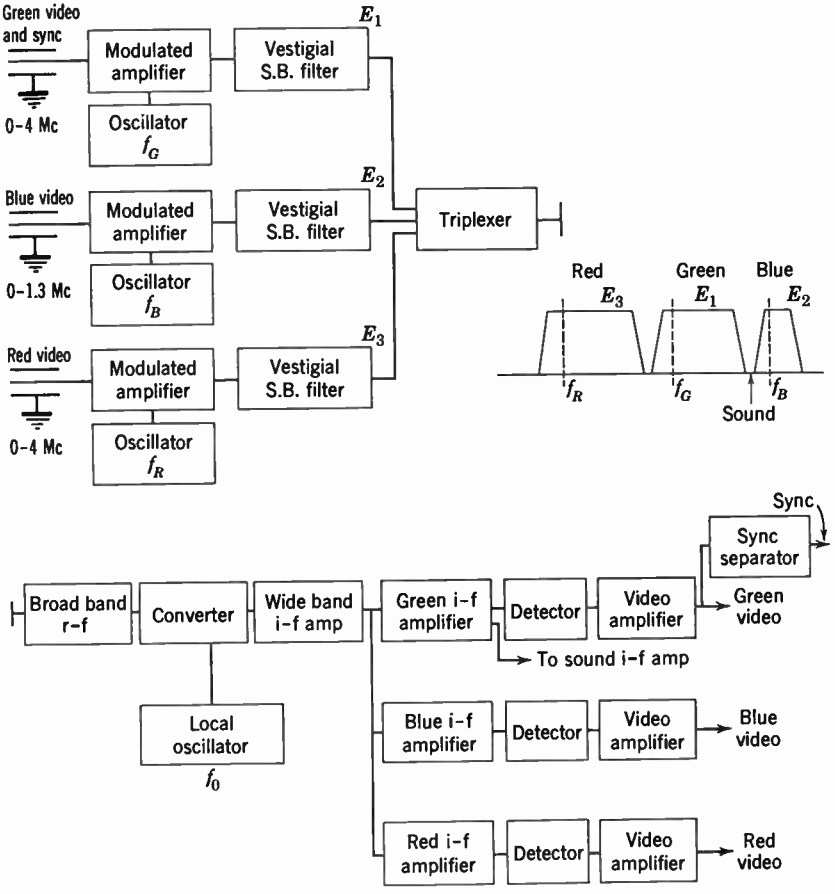


Fig. 17.8. Block Diagram of a Simultaneous System.

A complete discussion of the circuitry of this system is beyond the scope of this chapter. The circuit in detail is very complex, but for the most part follows standard practice. A simplified diagram of a particular circuit arrangement is given in Fig. 17.8. A few specific points relating to the system might be mentioned because of their general importance to color television.

The entire system was made as perfect as was technically practicable. The amplifiers were linear over a wide range, the system gains were matched, and the background levels were correctly set. A comparison of the picture performance under these conditions with that obtained when known changes in linearity, background level, deviations from registry, etc., were made gave quantitative information as to the characteristics needed to generate a picture of any prescribed quality.

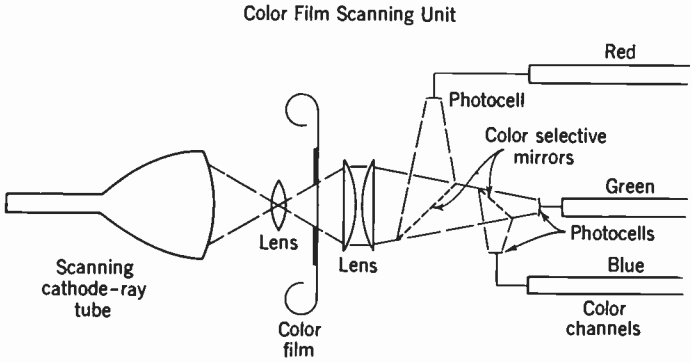


Fig. 17.9. Flying-Spot Pickup System (Schematic).

In addition to the studio camera discussed in the beginning of the section, color pickup from slides was effected with a flying spot scanner and three secondary emission multiplier phototubes. This pickup system is diagrammed in Fig. 17.9. The projection kinescope used to generate the scanning light spot employed a zinc oxide screen. Because the color of the spot must be effectively white, advantage could not be taken of the very fast time constant of the ultraviolet component of the fluorescence from zinc oxide. Therefore, an electrical filter was required to compensate for the decay time of the phosphor. The multiplier phototubes were of the RCA 5819 type, although some had modified spectral responses to extend the spectral response in the red. Excellent results were obtained from this arrangement.

A projection receiver was also tested with this color system. This receiver consisted of three Schmidt projection units mounted so that the three pictures were in register on the screen. A bright picture, 9 by 12 feet, was produced by this receiver. In order to demonstrate overall accuracy of the system, a circuit arrangement was provided to mix the video signals and supply this to the three projection kinescopes in the proper proportions to produce a black-and-white picture. The absence

of color shading and color fringes in the black-and-white picture was very convincing evidence of the perfection of the overall system. Figure 17.10 is a photograph of the color projector.

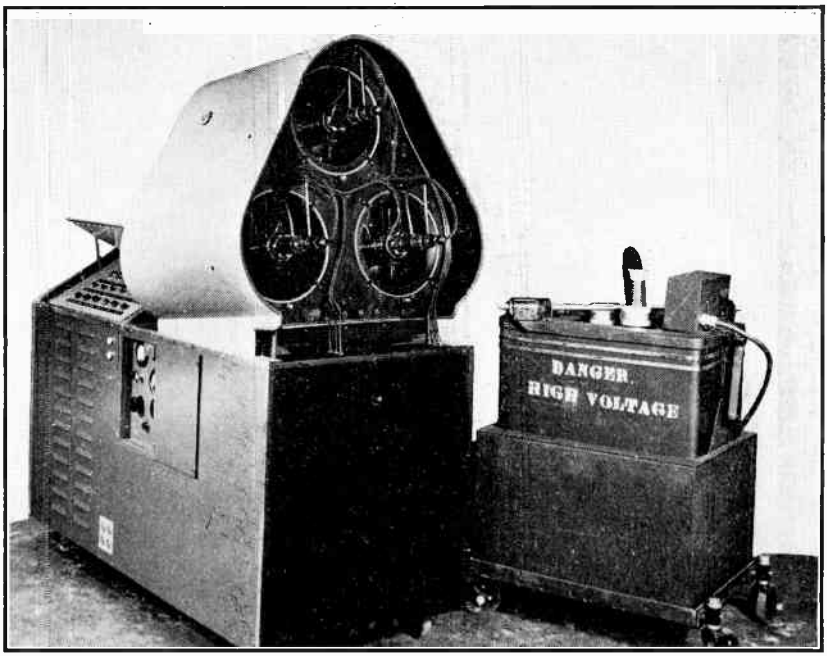


Fig. 17.10. Three-Color Projection Receiver.

17.5 Field-Sequential Color Systems. Color television obtained by rotating a color filter disk in front of both the pickup device and the viewer was the first type of system investigated. It was used by Baird

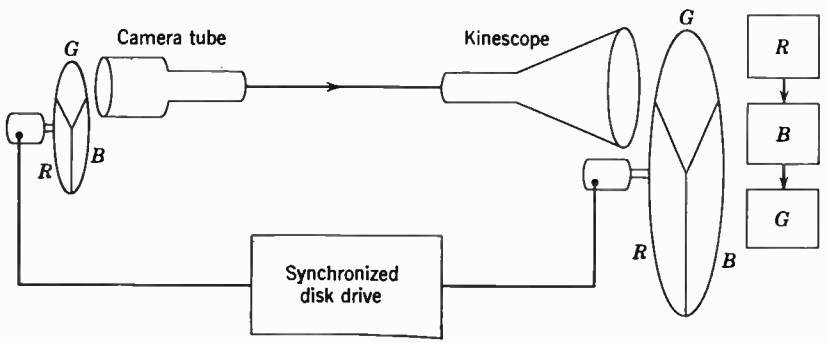


Fig. 17.11. Field-Sequential System.

in England in early experiments in 1928, before electronic television was developed. As applied to an electronic system, the arrangement is shown schematically in Fig. 17.11. The color disks rotate in front of the viewing tube and the pickup device. The rotation of the two disks is kept together and held in phase with the scanning pattern by means of the same synchronizing signal that controls the deflection. This timing signal controls the rotation so that the color filters change with each field sweep of the deflection.

Since the scanning is interlaced, the sequence of colors for three frames which is required for a complete color cycle is as follows:

Red	Frame 1	Field 1	Lines 1, 3, 5, etc.
Green	Frame 1	Field 2	Lines 2, 4, 6
Blue	Frame 2	Field 1	Lines 1, 3, 5
Red	Frame 2	Field 2	Lines 2, 4, 6
Green	Frame 3	Field 1	Lines 1, 3, 5
Blue	Frame 3	Field 2	Lines 2, 4, 6
		etc.	

If it were feasible to use the same frame frequency as that standardized for black-and-white television, this field sequential system would be compatible. In other words, a black-and-white picture could be obtained from the color picture signal. Some unbalance of detail would be present because of the difference in luminosity of the three colors. This probably would be quite unobjectionable.

The difference in luminance of the three component pictures is the factor which makes it impossible to use a 60-cycle field rate with satisfactory picture brightness. It has been noted that, for a black-and-white scene, the green field has twice the luminance of the red and six to ten times that of the blue. It will be seen from the schedule of the color sequence that the green field repetition rate is only one-third the deflection field rate. Therefore, with the normal 60-cycle field frequency, the green repetition rate would only be 20 per second, introducing an entirely intolerable brightness flicker.

In order to reduce the flicker to a tolerable level, it has been found necessary to increase the field rate of field-sequential color television to 144 cycles. Not only does this make the system noncompatible, but it also reduces the information that can be transmitted per frame over the accepted television channels by a factor of more than two. In order to balance the consequent loss of definition between the horizontal and vertical directions, the number of lines per frame is reduced from 525 to 405.

The lack of compatibility and the decreased definition are the two most serious weaknesses of the field-sequential system. They are faults which are fairly fundamental to the system. Both conditions are uneconomical and seem very undesirable.

A less important but nevertheless somewhat annoying defect also is fundamental to the system, namely, color fringing and color breakup.

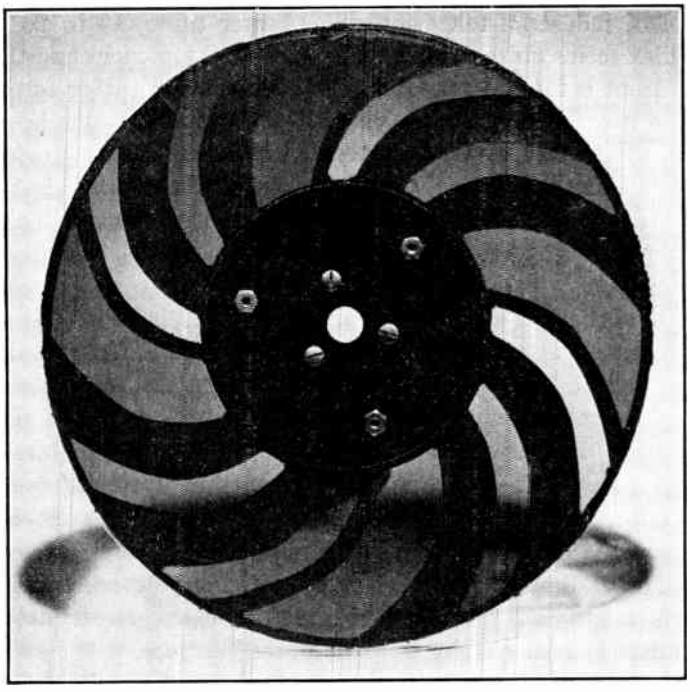


Fig. 17.12. Filter Disk for Pickup Tube.

If an object moves very rapidly in the picture, its image shows leading and trailing color fringes since the successive color images are displaced so far that they are no longer in register. For a similar reason, if the observer's eye moves rapidly, the successive color fields will not be in register at the retina of the eye, and an unpleasant sensation of bright color flashes is experienced.

Most of the work on this type of system has been done with mechanical rotating color disks (or color drums). The most efficient shape for the filter panels have been calculated, taking into account characteristics of the white phosphors used in the viewing tube. A pattern which has been selected as quite suitable is illustrated in Fig. 17.12. The

filters employed with this disk are rotated in front of the pickup tube. Different filters, determined by the spectral response of the pickup tube and the chromaticities of the viewing tube phosphor observed through the receiver filters, are required at the camera and at the receiver.

A rotating disk gives an acceptable picture for special-purpose applications. Receivers with 7- or 10-inch kinescopes which have rotating color disks in front of them would be entirely feasible, particularly in the field of industrial television. For larger tubes, 16 to 20 inches, such a disk seems quite undesirable. A color drum encircling the tube back to front is more practicable. Even this seems rather formidable for home use. It is probable that, if the system should be widely adopted, the picture would be presented on a tricolor kinescope, with three-color guns operating one at a time in sequence, instead of simultaneously as they would operate in the system for which the tube was designed. The sequential operation reduces the brightness correspondingly.

In spite of the limitations noted above, the system is capable of producing an excellent color picture, particularly with the increased bandwidth and field frequency employed in industrial color television chains. The colors are well saturated and natural looking. There is no question of the very high entertainment value of the picture.

Recent investigations on the field-sequential system have taken full advantage of the capabilities of the system and have developed it into a very workable method of obtaining color pictures. Except for three inherent limitations which were noted in the preceding paragraphs, it might well have been regarded as entirely satisfactory for a permanent system of color television broadcasting.

Except for the color disk itself, the circuit element adjustments required to alter the frame and line frequencies, and the means for color synchronization, this television system is identical with normal monochrome television. Hence, no further discussion of the circuits is required.

17.6 Line-Sequential Systems. The line-sequential method of obtaining television pictures in color presents a number of interesting technical problems. Unfortunately, however, it offers little promise. With this system, each line is presented in a different color, the colors being repeated in some sequence such as $g\ r\ b$.

The first consideration is the order of occurrence of the three colors among the lines for successive frames. The type of pattern obtained will depend upon the number of lines per field or frame. Interlaced

scanning is assumed throughout. Where even-line interlace is employed, three possible patterns result, as indicated in Table. 17.1.

TABLE 17.1. SEQUENCE OF COLOR LINES IN LINE-SEQUENTIAL SCANNING PATTERNS WITH EVEN-LINE INTERLACE

	First Field	Second Field	Third Field	Fourth Field	Fifth Field	Sixth Field
Line						
1	*G*		*G*		*G*	
2		*G*		*G*		*G*
3	*R*		*R*		*R*	
4		*R*		*R*		*R*
5	*B*		*B*		*B*	
6		*B*		*B*		*B*
7	*G*		*G*		*G*	
8		*G*		*G*		*G*
Line						
1	*G*		*B*		*R*	
2		*G*		*B*		*R*
3	*R*		*G*		*B*	
4		*R*		*G*		*B*
5	*B*		*R*		*G*	
6		*B*		*R*		*G*
7	*G*		*B*		*R*	
8		*G*		*B*		*R*
Line						
1	*G*		*R*		*B*	
2		*G*		*R*		*B*
3	*R*		*B*		*G*	
4		*R*		*B*		*G*
5	*B*		*G*		*R*	
6		*B*		*G*		*R*
7	*G*		*R*		*B*	
8		*G*		*R*		*B*

These three patterns would be highly unsatisfactory. The first shown in the table is static, with green always at line 1 and 2, and so on. Because of the difference in luminance of lines of different color, the arrangement would show a very pronounced line structure and would have only about one-half the vertical resolution of a normal television picture.

The second and third patterns might be expected to give full vertical detail and be satisfactory from this standpoint. However, it will be noted that, in successive frames, the pattern as a whole moves down,

producing a very distressing impression of crawling of the scanning pattern and causing a subjective loss of detail similar to that noted for the first pattern.

With odd-line interlace, the same phenomena are also present. The patterns are either static or suffer from interline crawling. None of these patterns is acceptable.

This account exhausts the possibilities of patterns made up by a regular repetition of the three primary color lines. A more complicated schedule can be devised for laying down the lines which avoid both static and crawling patterns. Even with such arrangements, which present some technical difficulties, the picture is not wholly satisfactory. Although crawling is absent, the picture will show a shimmer or quiver which at normal brightness levels is very unpleasant.

Since the line frequency is more than 15,000 cycles per second, line-sequential systems do not lend themselves to the use of a mechanical color commutator such as a rotating filter disk. As a consequence, some electronic means must be used to change the colors. The three kinescopes arranged as described in connection with the simultaneous system offer one way of forming the line pattern. In fact, a number of experimental test units have been built up to explore this type of system. A tricolor kinescope offers an alternative. As with the field-sequential system, only one gun at a time would be operating, which would decrease the maximum brightness that could be obtained.

Because the line-sequential method of obtaining color television does not seem to offer any particular advantages over other systems, and has certain objections peculiar to it, this television system will not be discussed further.

17.7 Element-Sequential Color Television. With element-sequential color television, the picture elements along each line go through the sequence of colors. The system in its simplest form is illustrated in Fig. 17.13. Here the maximum frequency which the video channel can accept is assumed to be 4.1 megacycles, and the commutator rotates at 4.1/3 million revolutions per second, giving 4.1 million color elements each second. The commutator is synchronized by means of the leading edge of the horizontal synchronizing signal and controlled in such a way that not only does each line in succession start with a different color in sequence, but also the first line of each frame starts with a different color. Thus, in three frame periods, every picture element will have carried the information of each of the three colors.

The camera for this system may employ three pickup tubes with registered optical images and carefully matched deflections. Each

tube would have associated with it amplifiers with enough gain to raise the video signal level well above commutator noise.

The reproducer might employ three optically registered kinescopes similar to those described in section 17.4. A tricolor kinescope would be an equally suitable alternative.

A moment's consideration will show that the dot color pattern formed by this system will have green and blue (maximum and minimum) vertical or diagonal lines which apparently move as the frames change. Thus this elementary element-sequential system would give an impression of a crawling pattern like that noted for the line-sequential television system.

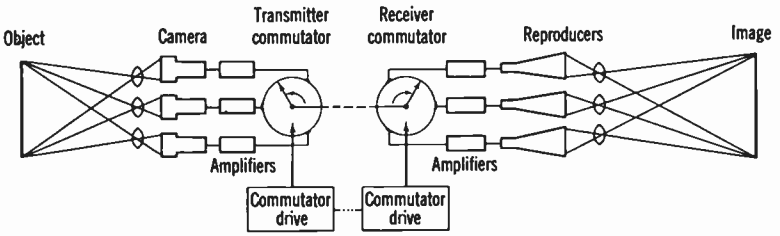


Fig. 17.13. Element-Sequential System.

So far, it has been assumed that each color gun was on only during the time the two commutators connected the channel and pickup tube to it. It is possible to store information in low-pass filters so that each gun stays on until the commutator makes a revolution and returns to the contact which supplies more information. If this is done, it is evident that it entails a considerable loss in definition. This expedient by itself does not solve the difficulty. However, if black-and-white or luminance information corresponding to the higher frequencies can be supplied to all the kinescope guns simultaneously to fill in the fine detail, the system begins to take on many desirable characteristics.

When handled in this way, the fine detail appears in the form of luminance variations, whereas only in the coarse structure is color preserved. This, however, does not detract from the picture quality, as has been noted in section 17.2, since the eye does not resolve fine spatial detail in chromaticity changes, but only in luminance variations. The system is economical, therefore, in the sense that valueless fine detail color information is sacrificed in exchange for the elimination of flicker and crawling which would otherwise be present, as it is in all other sequential systems examined.

This system can be made completely compatible. It will yield a black-and-white picture in an ordinary receiver which is indistinguishable from the picture produced by a monochrome transmission. No changes are required in the deflection or synchronizing circuits.

In this system involving low video frequency color and mixed highs, the concept of a commutator as applied to the electronic color switch loses its meaning. The process can be more appropriately designated as phase and amplitude modulation of a subcarrier. "Sampling" and subcarrier modulation will, in large part, be used interchangeably in the subsequent discussion. Once relieved of the commutator implication, the question arises as to the best frequency for the subcarrier or sampler. It will be pointed out in the next section that a much higher frequency is preferable because it makes simpler the separation of the sidebands associated with the subcarrier frequency from the low-frequency video component. These questions, together with others relating to the frequency band saving which is a fundamental feature of this system, will be taken up in the next two sections.

17.8. A Simultaneous System Derived from Element-Sequential Color. As introduction to this class of system consider a simultaneous system derived from the element-sequential method of obtaining color. In addition to employing a much higher frequency sampler, it has filters in the three color channels to limit the frequency of the video signal and a circuit for bypassing and adding the high-frequency components of the signal. A basic diagram for a sequential sampling system is shown in Fig. 17.14. In discussing this system, the sampler and desampler will be assumed to have a repetition rate of 3.58 megacycles, thus making the time separation between sampling 0.09311 microsecond. The length of time the sample switch is turned on is a very small fraction of the time between colors, a procedure known as "sharp sampling."

In discussing the operation of this system, consider first the output of the sampler as a transient phenomenon. It will be necessary later to treat the frequency spectrum of this signal in order to understand the processes in detail. However, the direct time variation of the signal reveals many of its features.

When the scanning beam of the pickup tube is passing over a saturated red area, the sampler output is zero except at the red contact position. The output signal, therefore, consists of a series of equal pulses at a 3.58-megacycle rate phased to correspond to the red phase angle. This state of affairs is illustrated in Fig. 17.15*a*. When this pulse signal is passed through the low-pass filter having a cutoff fre-

quency of 4.1 megacycles, the wave shape is altered to be a sine wave of 3.58-megacycle frequency and phase corresponding to the red signal

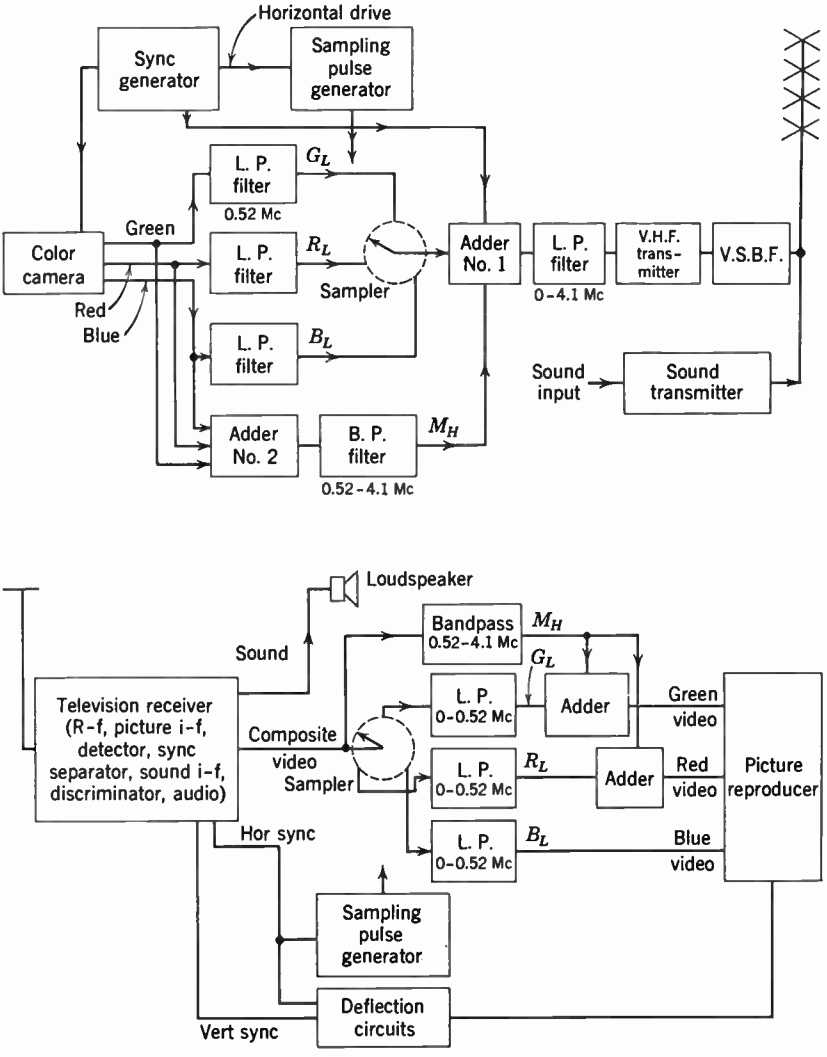


Fig. 17.14. Simultaneous Derived System.

angle. It will be noticed that the sine wave, in addition to having an amplitude A, is displaced by a d-c component B. It may be said in first approximation that the d-c component is a measure of the brightness of the area, and it is this signal which is the video signal used by

a black-and-white receiver in reconstructing a monochrome picture. If, instead of red, the area were colored a saturated green, a similar series of pulses would be formed but would be displaced by a phase angle. The low-pass 4.1-megacycle filter again produces a sine wave

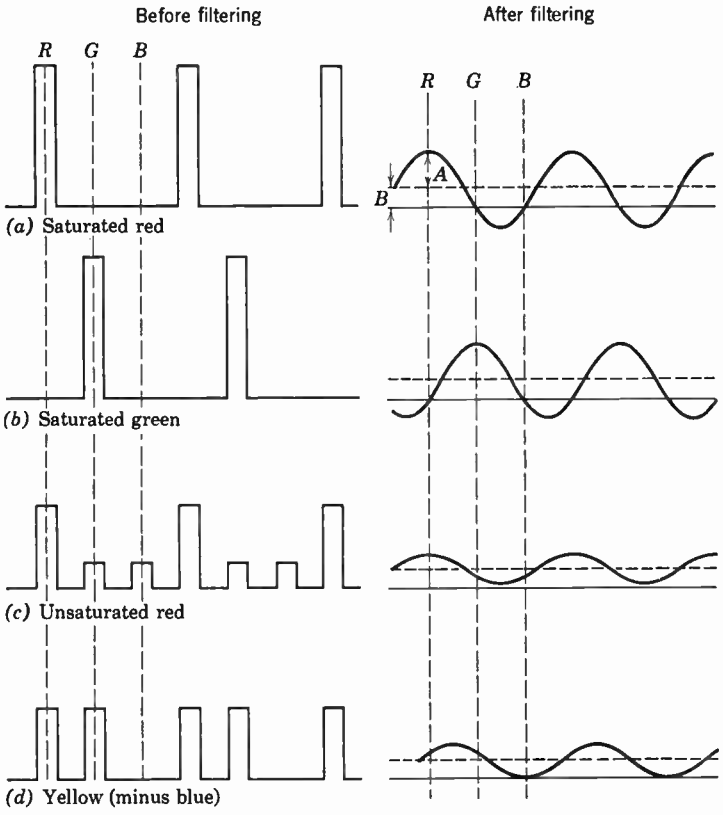


Fig. 17.15. Sampler Output for Colored Areas of Equal Brightness.

plus a d-c component. The amplitude A and displacement B are the same as for the red, but the phase has been shifted to the green position. Thus it is apparent that the phase of the sine wave determines only the hue of the color. The ratio of the amplitude to the d-c component indicates the saturation of the transmitter color.

The desampler is placed in such a way that its contacts occur at the peaks of the corresponding color signal. It will be noticed that when a saturated red is being transmitted, the signal is exactly zero at the positions of the green and blue contacts. Similarly, with a green sig-

nal the transmitted signal is zero at the red and blue contact points, insuring that the reproduced color will be saturated.

Figure 17.15*c* illustrates the wave form for an unsaturated red color. The phase of the sine wave signal from the filter is exactly the same as that of Fig. 17.15*a*, but its amplitude is small. However, at the same time the d-c component B is relatively larger. Thus the signal can easily be seen to be made up of white plus a small amount of red.

A pure white area produces a sampler output consisting of equal pulses at three times the sampling rate. When put through the low-pass filter, no a-c 3.58-megacycle component is found at the output. The only output signal is the d-c output B. This condition gives a high luminance area with a monochrome receiver. With a color receiver, the desampler has equal signals applied to the three contacts so that white is reproduced by activating the three kinescope guns equally.

The final signal to be examined (Fig. 17.15*d*) is from a color region which has only red and green components. The signal from the sampler consists of pairs of pulses at the red and green positions and nothing at the blue. When filtered, the output consists of a sine wave with an amplitude half as large as that in Fig. 17.15*a* and *b* and with its peak half way between the red and green positions. The d-c component is twice the amplitude of that for the pure red and pure green components alone. It will be noted that the sinusoidal variation can be interpreted as a blue signal shifted in phase by 180 degrees—in other words, minus blue. The color may rightfully be thought of as white minus blue, a prediction which could be made from the nature of the color triangle discussed in section 17.2.

The three outputs from the desampler are filtered with low-pass circuits having high-frequency cutoffs at about 0.52 megacycle. They remove all traces of the high frequencies corresponding to the desampler switching and leave only low-frequency changes covering periods of many cycles of the switch. Thus, when a large green area is being scanned, the gun of the green kinescope turns on and stays on until the edge of the area is reached. If the green area has some fine detail in another color, the filters prevent it from carrying through to the color guns. In other words, the whole system is geared to the experimental fact that the eye cannot see color detail. To avoid a possible misunderstanding at this point, it should be explained that the low-frequency color averages will appear in the reproduced picture. For example, suppose that a red area of 100 picture elements in width is the subject of the picture and that in the center of the red field is

a patch 20 elements wide, where red is alternated with blue at one or two element intervals. As far as the color component is concerned, the picture would be reproduced as a red field with a purple (red and blue) patch, 20 elements wide in the center. If there were a difference in luminance between the red and blue, the bypassed mixed highs would cause this difference to reproduce as variations in brightness level in the purple. A simple test with a card colored with the red-blue pattern held at the appropriate distance from the eye will show that the picture reproduced by the system agrees very closely with what the observer sees.

The mixed highs which are included to carry the information about the luminance changes in fine detail are obtained in a fairly conventional manner. The video signal from the three color pickup tubes passes through a low-pass filter with a 0.52-megacycle cutoff before going to the sampler. A fraction of the signal from each tube also goes to an adding circuit, which combines them and is followed by a bandpass filter which transmits the portion of the spectrum between 0.52 and 4.1 megacycles. The output from this filter is added to the output from the filter following the sampler. This combined signal constitutes the complete video signal.

At the receiver, the complete signal is supplied to the desampler. A portion of the video signal is passed through a bandpass filter, which transmits the video spectrum above that transmitted by the desampler filters. This signal is added to the color signal supplied to the three guns, providing the detailed information which will be reproduced as variation in the white content.

In order to answer certain questions about the interaction of various frequency components, as well as to obtain a more quantitative understanding of the operation of the system, it is convenient to take a somewhat different point of view, that is, instead of using the transient wave form approach, to consider the frequency spectrum involved. The input to the sampler is confined to a narrow band of frequencies from 0 to 0.52 megacycle for each channel, as indicated in Fig. 17.16*a*. Where sharp sampling is used, the output is in the form of pulses, which in terms of their Fourier components consist of a fundamental at the sampling frequency together with components of equal amplitude at integral multiples of the fundamental. This pulse output is modulated by the low-frequency video signal. The general form of the signal is

$$P(t) = A(t)(1 + 2\cos 2\pi f_s t + 2\cos 4\pi f_s t + 2\cos 6\pi f_s t \cdots) \quad (17.2)$$

Actually, the relative amplitudes of the higher harmonics depend upon the shape of the sampling pulse. However, this is unimportant here because all but the fundamental is removed by the filter which follows the sampler. Since the amplitude A varies, representing the low-frequency video signal, the Fourier representation shows 0.52-megacycle sidebands about each of the harmonics as indicated in Fig. 17.16*b*.

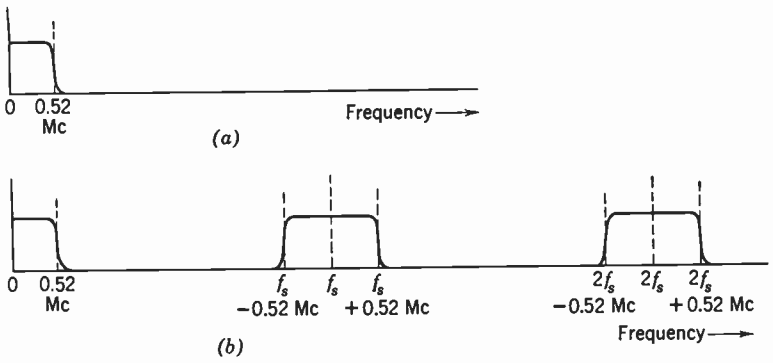


Fig. 17.16. Spectrum of (*a*) Low-Frequency Video Signal and (*b*) Sampler Output.

There are, of course, terms similar to that of, e.g., Eq. 17.2 for each of the three colors, shifted in phase relative to each other by 120 degrees. The form of the total signal for the color information is:

$$S(t) = \tfrac{1}{3}A_r(t) + \tfrac{1}{3}A_g(t) + \tfrac{1}{3}A_b(t) + \tfrac{2}{3}A_r(t)\cos 2\pi f_s t + \tfrac{2}{3}A_b(t)\cos[2\pi f_s t + 120°] + \tfrac{2}{3}A_b(t)\cos[2\pi f_s t + 240°] \quad (17.3)$$

To this signal is added the mixed high-frequency signal. The first three terms which do not vary sinusoidally with a frequency f_s are the portion giving the brightness information of the coarse detail of the picture. This portion, together with the mixed high-frequency signal, constitutes the useful signal for reconstructing the black-and-white picture on a conventional monochrome receiver.

It will be noticed that a spurious signal consisting of the sampling frequency (and its sidebands) may produce a fine-grained unwanted pattern in this picture.

17.9 The Derived Simultaneous System (*continued*). The signal supplied to the receiver may be considered profitably from a slightly different point of view than heretofore. In the preceding sections the

signal to the desampler was considered to be a low-frequency component (B of Fig. 17.15) plus a high-frequency component consisting of the sampling subcarrier and its sidebands (A of Fig. 17.15). The high-frequency brightness component bypassing the sampler as mixed highs is added hereto.

Instead, the entire brightness signal, i.e., the low-frequency part (B) plus the mixed highs, may be bypassed around the desampler. The desampler will then receive only the sampler subcarrier and sidebands (and a portion of the mixed highs signal). A block diagram showing

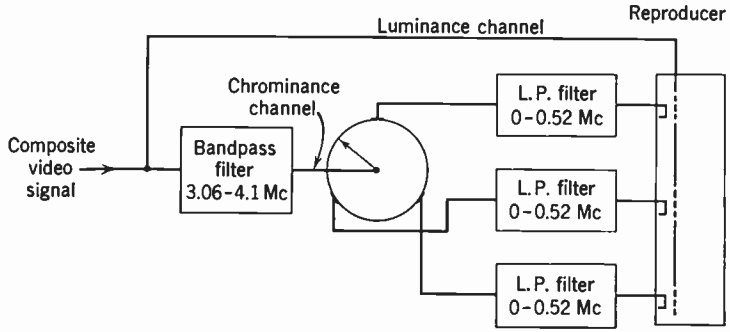


Fig. 17.17. Block Diagram of Receiver with Separate Chrominance and Luminance Channels.

a receiver arranged in this manner is given in Fig. 17.17. Such a receiver is one in which the brightness or luminance signal and the chrominance signals are truly separated. The same result may be obtained by replacing the sampler and desampler by three subcarrier modulators and demodulators phased to match the sampling angles. This second procedure is actually used in practice and will be discussed in detail in the next chapter.

To clarify this procedure, consider the signal illustrated in Fig. 17.18. The color of the area is a saturated yellow (red plus green). The total signal from the sampler is a displaced sine wave with its maximum halfway between red and green, and a minimum of zero at blue. If this complete signal is supplied to the desampler, it is obvious that only the red and green kinescope guns will turn on. With the modified receiver, the signal reaching the desampler will be the sine wave of Fig. 17.18*b* without d-c component. The low-frequency component B will go directly to all the kinescope guns. At the contact positions for red and green, the desampler output is a positive signal which adds to the component B already on the red and green guns. At the blue posi-

tion, the desampler output is a *negative* signal. This signal subtracts from the brightness signal on the blue gun and is just sufficient to cancel that signal. Thus the red and green guns are on while the blue gun remains cut off and the desired yellow is obtained. It is a simple matter to trace through this pattern of operation for other color combinations.

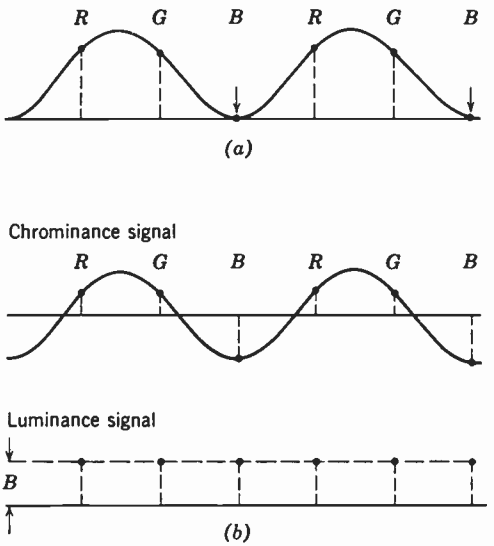


Fig. 17.18. Signals for Derived Simultaneous Systems (*a*) with Bypassed Highs and (*b*) Bypassed Luminance Signals.

It was pointed out in the preceding section that the sampling signal can produce a fine-grained interference in the black-and-white compatible picture. This interference is largely removed by using what is termed horizontal interlace. If the first frame starts with red signal, the second frame starts with blue, but displaced so that the element will lie halfway between the normal first and second element. A moment's consideration will show that this is equivalent to shifting the phase of the sampler by 180 degrees between frame pairs. Although the expedient does not give electrical cancellation of the spurious signal—and indeed it must not since, although it is spurious for the black-and-white picture, it is essential for color—optical cancellation between frames results.

Another type of spurious signal—this time in the color picture—is not eliminated. It is the low-frequency color signal resulting from a beat between the sampling subcarrier f_s and the high-frequency video

components. Since the interfering signal is not in any way synchronized with the subcarrier, the interference does not have a definite color but rather goes through the sequence of red, green, and blue. The reversal of sampling phase in successive frames again produces optical

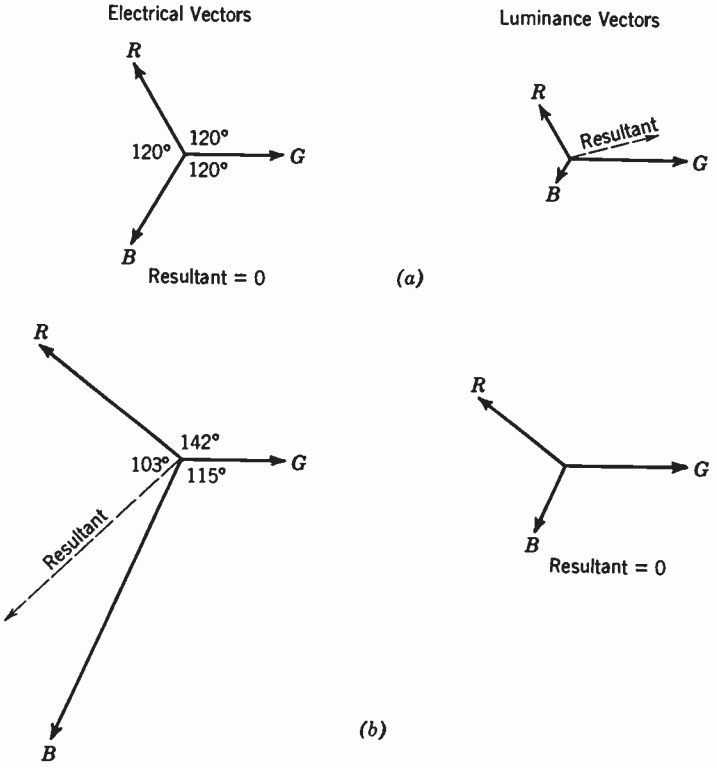


Fig. 17.19. Vector Diagrams for "Constant Amplitude" and "Constant Luminance" Color Systems.

cancellation. The residual color flicker associated with this effect is not readily visible. However, since the luminosities of the colors differ, a brightness flicker or quiver may be present. Although this interference is not prominent, it is undesirable.

It was noted that, if this interference consisted simply of chromaticity variations, it would not be detected and that only the luminosity differences of the colors cause the trouble. This observation led to the proposal of making the bypassed monochrome signal proportional to luminance and modifying the sampling process accordingly. Al-

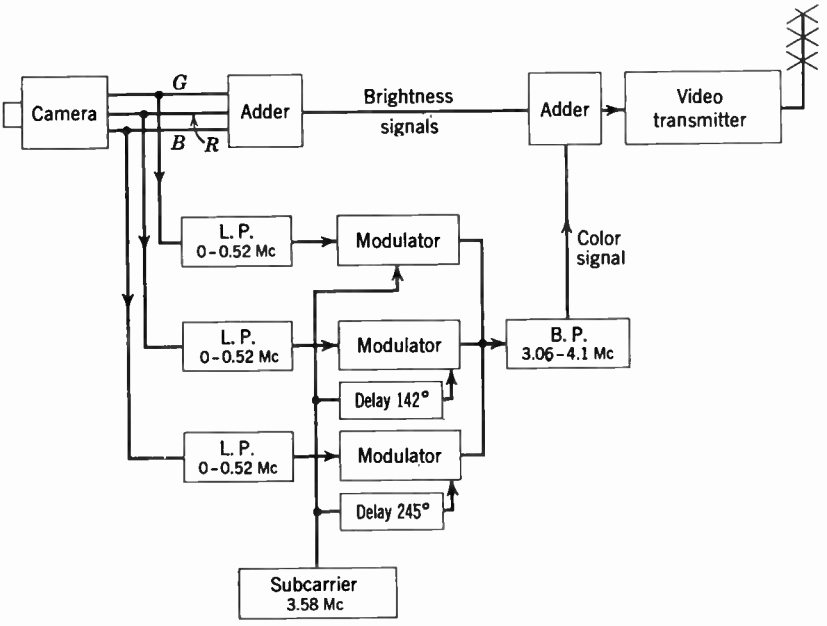


Fig. 17.20. Block Diagram of Transmitter for Constant-Luminance System.

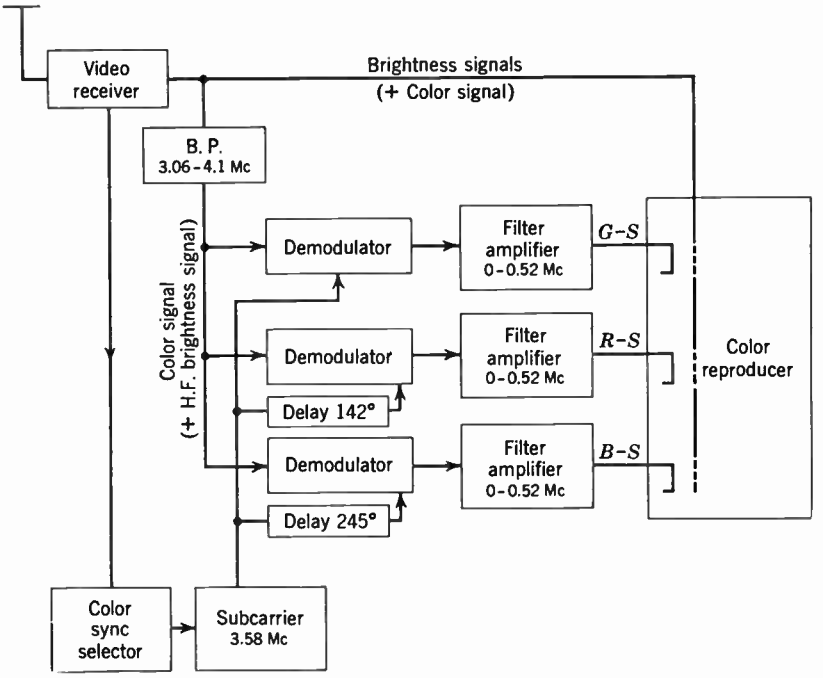


Fig. 17.21. Block Diagram of Receiver for Constant-Luminance System.

though this modification leads to a slight increase in complexity, it largely eliminated the fault under discussion. Figure 17.19 shows how, with constant-luminance sampling, any signal applied to the color channel is without effect on picture luminance.

It is assumed throughout that for a white image the signal for the three colors is equal. A vector diagram of the electrical output of the filters following the desampler, for an arbitrary sinusoidal input signal in the passband range, is shown in Fig. 17.19*a*. In this same figure, the corresponding vector diagram of the luminance is also given. The resultant vector due to the unbalance of the luminance is very apparent, and it is this resultant which leads to the brightness flicker when an interfering video signal close to the sampling frequency is present. By employing unequal gains in the three receiver color channels, a constant luminance chromaticity channel can be devised. In order to avoid the very large inequality of gain that would be required to bring the luminosity of blue up to those of the red and green, it is advantageous to alter also the phase of the sampling. In fact, considerable circuit simplification results from ignoring the luminosity of the blue entirely and balancing the red and green alone, placing these two vectors at 180 degrees with respect to one another. The error from this procedure is small. Figure 17.19*b* shows the luminance vector diagram together with the electrical vector diagram for the signals at the output of the three color channels in the receiver, for a sinusoidal interfering signal. Block diagrams of the transmitter and receiver for effecting this constant luminance color television transmission are shown in Figs. 17.20 and 17.21. Other methods of generating and detecting the chrominance signal in a constant-luminance system will be described in Chapters 18 and 19.

17.10 Color Viewing Tubes. The most important single item in the problem of color television, almost irrespective of the system selected, is the means for reproducing the picture in color. Neither the method of employing a mechanical rotating color disk in front of a kinescope nor that of using three kinescopes with appropriate light filters in optical register is very satisfactory. Direct-view receivers using either of these means, although not completely impractical, would certainly be very awkward. Therefore, the development of a single tube capable of reproducing the three required colors has been of great importance for the wide acceptance of color television.

A great many different methods of obtaining color in a single tube were studied and experimentally investigated in the research program

aimed at solving the problem. A few of them will be discussed in this section, but many must of necessity be omitted.

Perhaps the simplest idea for a three-color tube is one where the phosphor screen is in the form of horizontal lines of colored fluorescent materials. The arrangement is illustrated in Fig. 17.22. Each triplet line (red, green, and blue) corresponds to one line of the raster. The beam is small enough so that it is only on one color at a time. Magnetic deflection is used to produce the scanning pattern, and auxiliary electrostatic plates serve to deflect the beam onto the appropriate color

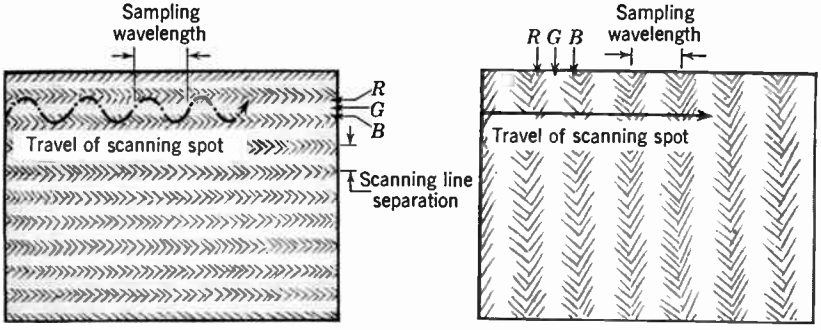


Fig. 17.22. Horizontal and Vertical Phosphor Line Screens for Tricolor Tubes (Phosphor Line Widths Greatly Enlarged).

line. This tube has a number of very real weaknesses. First, the deflection accuracy required is extreme. Second, the very small spot limits the average brightness. Finally, only one color can be used at a time, though this does not apply to similar tubes employing three guns.

Figure 17.23 is a modification of this principle which reduces the accuracy requirements of focus and deflection. With this venetian blind screen having red on one set of the "slats" and blue on the other, while the green is on the tube face, a deflecting voltage is applied between pairs of slats which deflects the beam to the appropriate color. This arrangement is hard to construct, requires considerable power for color deflection, is optically rather unfavorable, and does not give good color purity. Furthermore, it is not suitable for a simultaneous color picture.

A third deflection color selection tube is shown in Fig. 17.24. Here, the beam approaches a metallized coating which is at a slightly negative potential with respect to the gun and, hence, serves as an electron mirror for the beam. The beam is turned back by the mirror onto

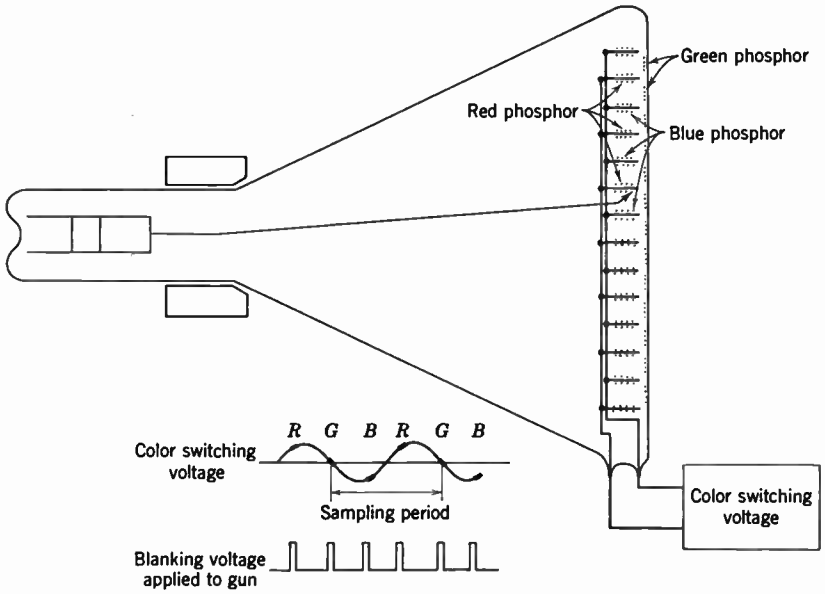


Fig. 17.23. Venetian Blind Color Tube.

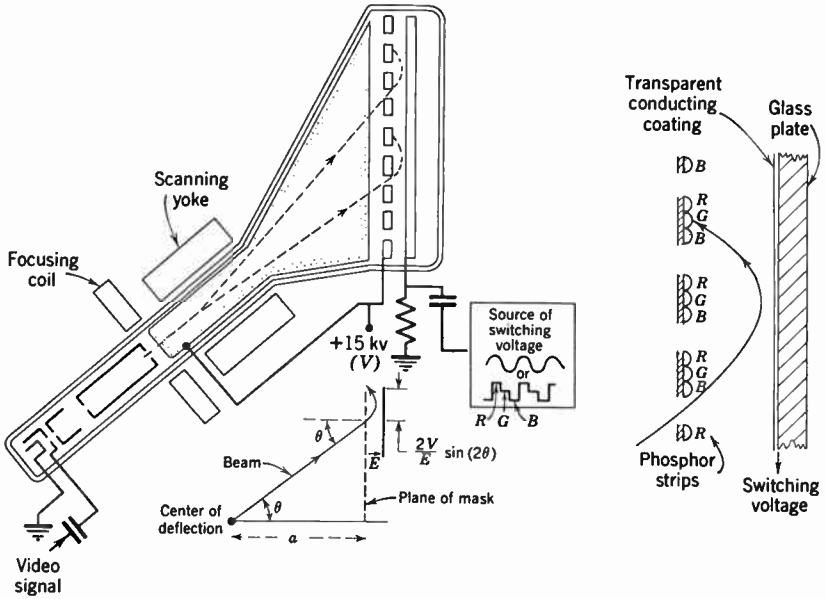


Fig. 17.24. Forty-Five-Degree Reflection-Type Color Kinescope.

strips which have the three color phosphors in bands. The color which the beam strikes depends on the exact mirror potential, and, therefore, the mirror can be used as color selector. Among the drawbacks to this tube are the high voltage differences required in the screen assembly, the accuracy of adjustment required, and the fact that only one color is on at a time.

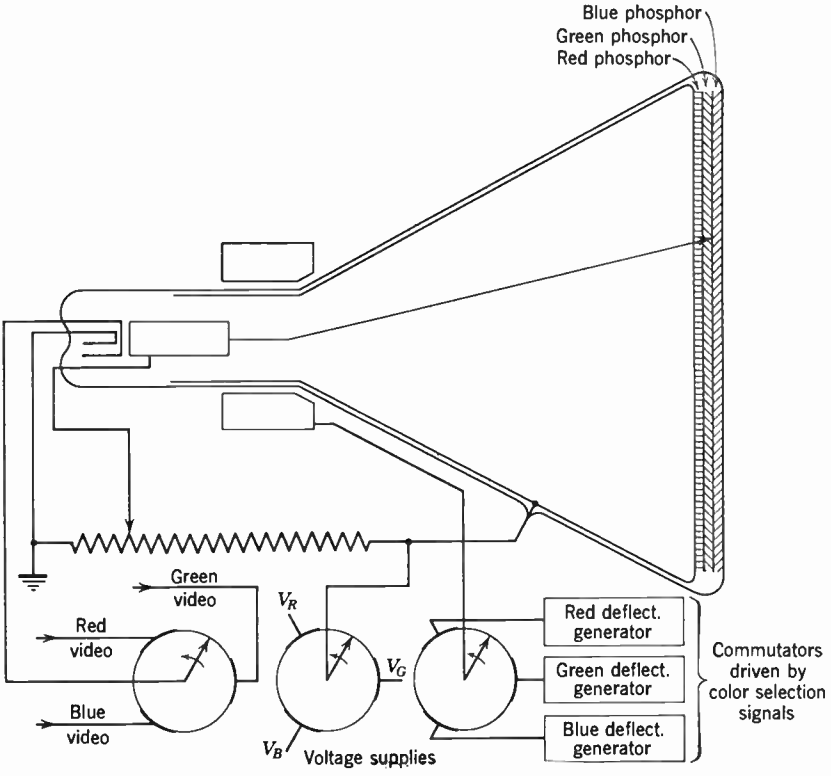


Fig. 17.25. Penetration-Type Color Kinescope.

In an entirely different method of producing color, a screen made up of layers of different colored phosphors is used. Here the penetration of the beam governs the color. The arrangement is illustrated in Fig. 17.25. Lack of color purity, scanning and color control interdependence, and the fact that only one color at a time could be used are all objections to this type of tube.

In order to be able to have all three colors appearing at the same time and independently controllable, a whole class of tubes may be conceived employing three separate guns, one for each color. Figure 17.26

illustrates a screen designed for a three-gun tube. This screen is covered with minute pyramids having their three faces at right angles (cube corners). Each face is coated with a different phosphor, and a beam approaching from a direction at right angles to one set of faces excites only one color. This arrangement has a number of desirable

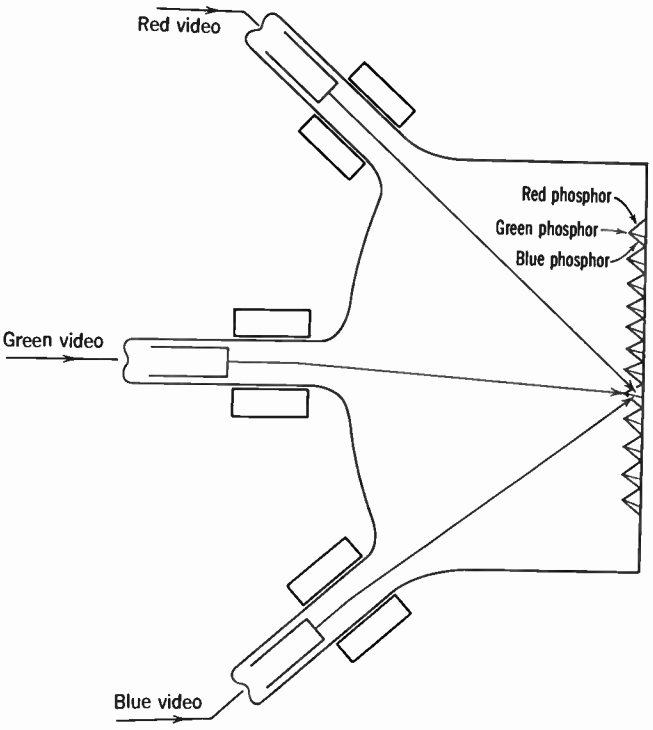


Fig. 17.26. Pyramidal-Screen Color Tube.

features, but requires separate deflections for the three guns, each with very marked "keystone" correction. Under these circumstances the problem of registering the three color images is formidable.

The tricolor kinescope, which was selected as most practicable at the time of writing (1953), employs three guns, a shadow mask to insure color purity, and the colors in the form of a pattern of dots of color phosphors.

A cut-away view of this tube, which will be discussed in greater detail in Chapter 19, is shown in Fig. 17.27. The three scanning beams, forming angles of the order of 1 degree with each other, are deflected together across a thin metal mask perforated in a hexagonal array of

apertures. Three dots on the screen of red, blue, and green phosphor, respectively, are so positioned with reference to each mask aperture and the three beam directions that one beam can strike only red dots, another blue dots, and a third green dots. Thus, as in the tube with the pyramidal screen, the three guns generate, individually and simultaneously, the red, blue, and green partial pictures in superposition, yielding a natural-color rendition of the original. Figure 17.28 shows

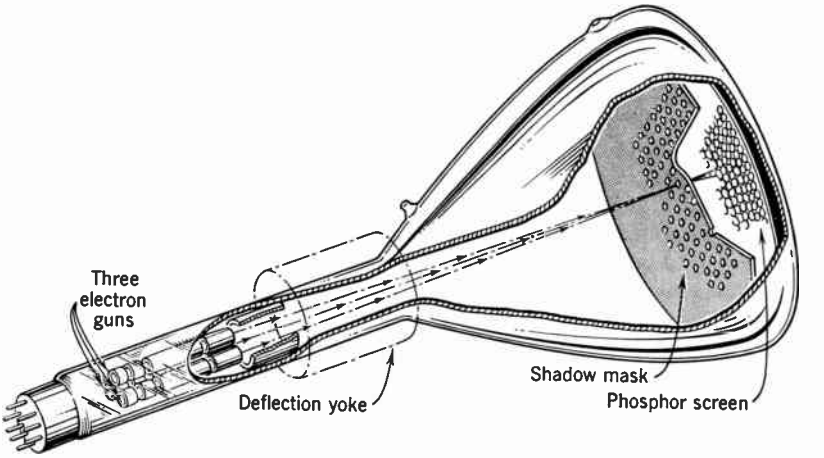


Fig. 17.27. Cut-Away View of Shadow-Mask Tricolor Kinescope.

the external appearance of such a three-gun shadow-mask tricolor kinescope.

Although tubes of this type have been found to yield very satisfactory color pictures, they possess one fundamental drawback. Approximately five-sixths of the beam power is dissipated at the shadow mask. This places a definite limit on the picture brightness which can be obtained, since color separation may be impaired when the energy dissipation at the mask is excessive. Since the transmission of the mask becomes less as the convergence angle between the beams is decreased, it is generally not advantageous to employ very small convergence angles. Small convergence angles, however, do simplify the problem of obtaining perfect superposition of the three partial pictures over large angular scanning fields.

A practical remedy for this difficulty may be found in the focusing action of slits. It may be shown from the Davisson-Calbick formulas in Chapter 3 that, if an accelerating field is applied between a linear mask consisting of parallel wires and a line screen, a parallel electron

beam incident on the mask will be focused into sharp lines on the screen, provided that the screen potential (measured relative to cathode potential) is about four times the mask potential. With this arrangement, the open area of the mask may be greatly increased. It should be noted that the focusing process here described implies beam acceleration after deflection ("postacceleration").

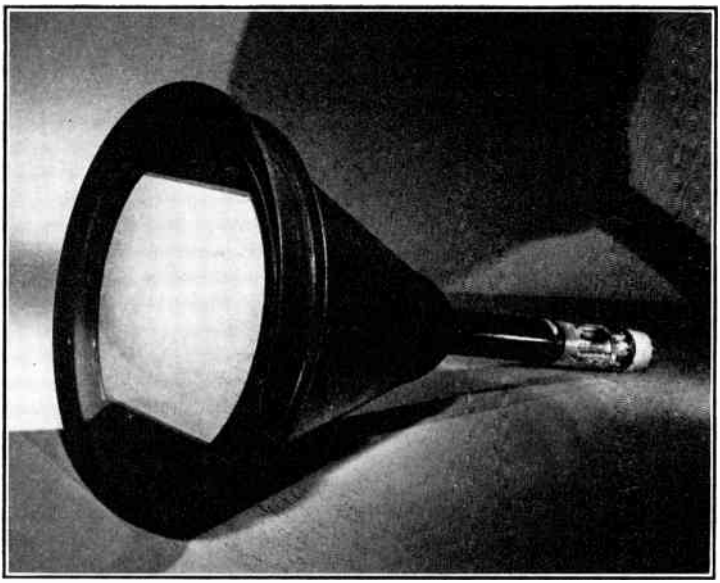


Fig. 17.28. Three-Gun Shadow-Mask Tricolor Kinescope.

A line screen tube with a single beam, employing postacceleration to obtain the desired focusing effect, has been demonstrated.* Sequential color selection after the manner proposed by A. C. Schroeder† was obtained by applying voltage differences at color change frequency between alternate wires (Fig. 17.29). Four hundred 0.008-inch wires, 0.040 inch between centers formed the focusing grill; 400 green, and 200 red and 200 blue phosphor strips were deposited on the screen, 0.4 inch behind the grill. A switching voltage of 440 volts was required to deflect the electrons from the green strips to either the red or the blue strips. The fact of sequential operation, inherent in this

* See D. G. Fink, "Phosphor Strip Tricolor Tubes," *Electronics,* Vol. 24, pp. 89–91, December, 1951.

† See A. C. Schroeder, U. S. Patent 2,446,791 issued August 10, 1948.

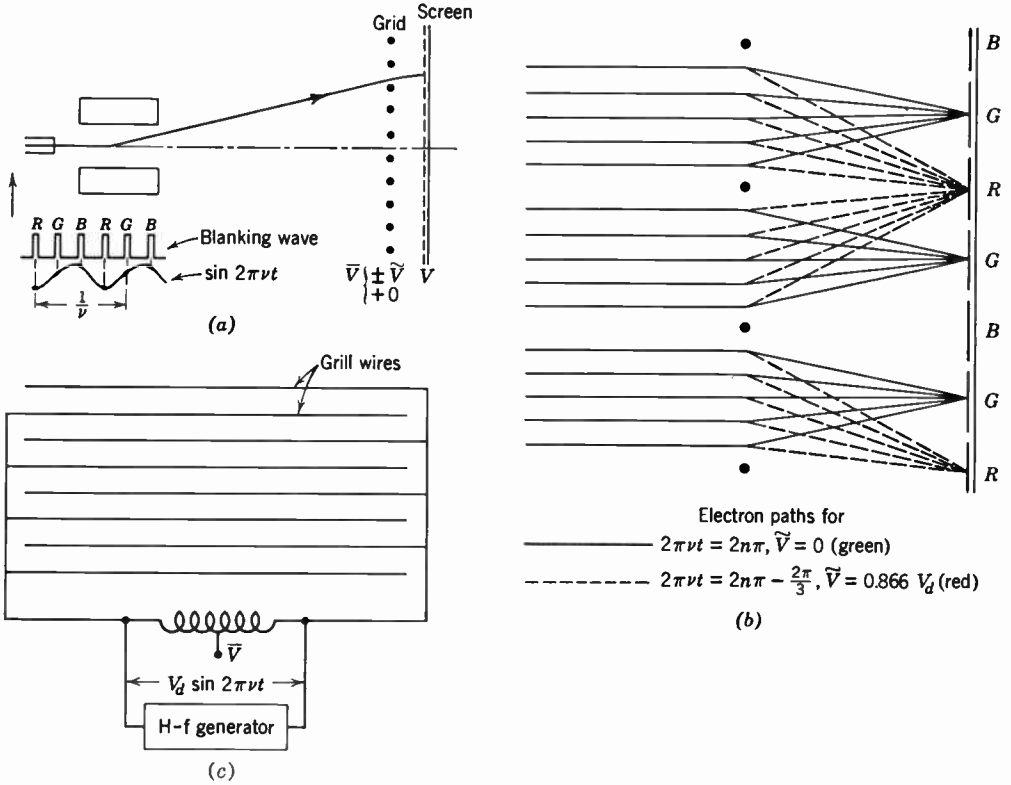


Fig. 17.29. Post-Acceleration Focusing Tube with Color Selection by Grill Modulation. (*a*) Geometry of tube. (*b*) Focusing and deflection action of grill. (*c*) Electrical circuit of grill.

as in most single-beam tubes, tends to compensate the gain in brightness arising from application of the focusing principle.

17.11 The Color Receiver. Through the r-f and i-f portion of the receiver, the instrument for reproducing color does not differ essentially from a conventional monochrome receiver. Care must be taken,

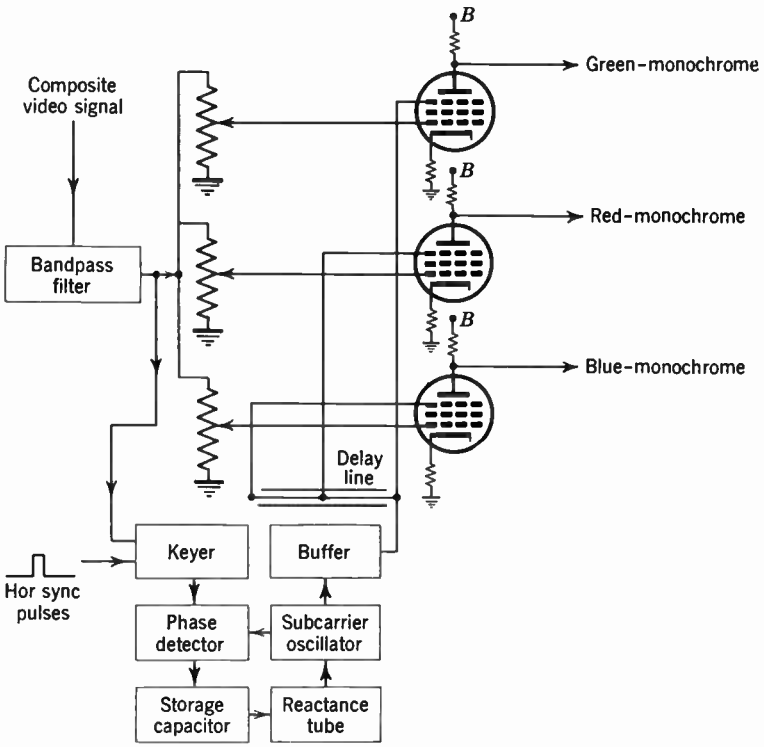


Fig. 17.30. Simplified Diagram of Color Demodulator.

however, that the full width of the video channel is maintained. Also, there must be no serious discontinuities in the phase or amplitude response of the channel. This is obviously so because the important color information is all located at the high-frequency end of the channel, whereas the major part of the brightness or luminance signal is in the low-frequency region.

The desampler which follows the second detector is unique to the color receiver. A simplified diagram of this triple demodulator is shown in Fig. 17.30. The main controlling subcarrier oscillator is brought into synchronism and phase by a special wave train or burst

which follows the standard synchronizing pulse. This oscillator drives the three demodulators.

The outputs from the desampler are filtered by specially designed but more or less conventional filters. They are next amplified by linear video amplifiers and applied to the three appropriate cathodes.

The bypassed brightness signal after passage through a filter is amplified and applied to the grids of the three guns simultaneously.

Separate gun bias controls are provided to regulate the color background. Furthermore, separate gain controls for the three color amplifiers control color balance.

The deflection generators and power supplies follow more or less conventional design. No further discussion will be given of these components.

A block diagram of the complete receiver is given in Fig. 17.21.

17.12 Color Camera and Transmitting Equipment. Active investigation of three-color camera tubes is being carried on. A number of solutions have been proposed and are being studied theoretically and experimentally. It would be premature to discuss these solutions at this time. The problem, of course, is an important one, but it does not rank with the problem of a three-color viewing tube because for every camera there are tens or hundreds of thousands of viewing devices.

The cameras in use to date are based upon one or more conventional pickup tubes, (e.g., image orthicons) and the required filters. For field-sequential systems, the camera uses a single image orthicon and a rotating color disk. In order to take as much advantage as possible of storage, the filter is placed close to the cathode of the tube so that the edge of the filter can follow the vertical movement of the beam. However, full storage cannot be effectively used because of the scanning interlace. For example, if the first field of a frame is red and the second green, some red stored picture will be present on the green field, accumulated during the first field. In order to avoid the loss of saturation that may result, it is necessary to operate at high enough light levels so that only line storage is involved. With an image orthicon, this light level in any case corresponds to normal studio usage. Other considerations for this type of camera relate to good mechanical and electrical practice. The rotating disk must be free from mechanical imperfections which might introduce microphonics, hunting, and other instabilities in the rotation of the disk.

The second type of camera employs three image orthicons in optical register. This type of camera is illustrated in Fig. 17.6. Such a camera can be advantageously used for all the color systems discussed

in this chapter. When used for field-sequential systems, it gives an enormous gain in storage. Each camera can store charges not only during its own field time but also for the two other field periods. This feature results in a direct sensitivity gain which would be important for outdoor and indoor spot pickup. For the element-sequential-derived simultaneous system this camera also takes full advantage of storage so that its sensitivity is high.

The electronic equipment associated with the pickup units will be described in more detail in Chapter 19. A representative block diagram for a derived simultaneous system transmitter has already been given in Fig. 17.20. Again, the sampler or subcarrier modulator is the most characteristic circuit element of the color terminal equipment. In principle, it is like the desampler of the receiver but has greater flexibility and accuracy.

18 Color Television Principles

Chapter 17 provided a general survey of color television. The basic properties of the eye, relevant to the discrimination of color, were outlined and various systems of color television making use of these properties were described. It was found that these systems differed materially both in the effectiveness with which they are adapted to the peculiarities of the eye and in the degree to which they meet compatibility requirements, permitting the reception of color transmissions and monochrome transmissions on both color receivers and monochrome receivers. In both respects, the "derived simultaneous" or "dot-simultaneous" system was found to be outstanding.

In this chapter, the principles of colorimetry and their implications for color transmission systems are treated somewhat more quantitatively. Furthermore, the operation of the dot-simultaneous system is analyzed and several variations designed to overcome specific shortcomings are described. This discussion provides a basis for an outline of the standards ultimately adopted by the Federal Communications Commission.

A description of the practical implementation of the dot-simultaneous system and of experience gained in its operation is reserved for Chapter 19.

18.1 Principles of Colorimetry as Applied to Color Television. It was brought out in the last chapter that the human eye behaves *as though* it contained, in the color-sensitive portion of the retina, just three types of receptors with different spectral responses. Whereas the nature of these receptors is unknown and there is no direct proof of their existence, tricolor processes of color reproduction have a firm basis in the following experimental facts:

1. Any light sample may be matched by any three primary lights. Here, the *primary* lights are defined as three lights of different color (e.g., green, red, and blue) such that none of the three can be matched by combinations of the other two. A match may be effected by illuminating one perfectly reflecting diffusing surface with the sample

light, an adjoining similar surface with the primary lights, and adjusting the intensities of the latter until the two surfaces appear identical. The match may, eventually, require the addition of one or two of the primaries to the sample rather than to the comparison field. In this case, the sample is matched not simply by the addition of primary lights, but by their addition and subtraction.

2. A match effected at one intensity level is valid within a wide range if the intensities of the components are scaled up or down in proportion.

3. Two lights matching a third, match each other.

4. If one matching pair of lights is added to another matching pair, the resultant combinations match each other.

5. The relative proportions of primaries effecting a match for a given sample light vary little for different normal-sighted observers. Furthermore, since luminance, the objective measure of brightness as carried out by a photometer, is an additive property—i.e., the luminance of a surface illuminated by two different light sources is equal to the sum of the luminances when illuminated by either source alone—the judgment of the relative luminance of the two lights also varies little among normal-sighted observers.

The last fact permits an objective description of light samples in terms of the reaction of a "standard observer," who represents a statistical average of a large number of normal-sighted observers. In this description it is convenient to separate luminance or brightness and chromaticity or chrominance as given by the *relative* intensity of the three matching primaries. The term "chrominance" has been introduced, finally, to describe the difference between a given light sample and a prescribed white of equal luminance.

As has already been pointed out in the last chapter, an objective measure of the brightness is the luminance L of the sample, which may be established with the aid of the luminosity function $V(\lambda)$ (Fig. 17.1), provided that the spectral distribution of the radiance of the sample $E(\lambda)$ (radiant energy in unit wavelength range emitted by unit area of the source per unit time into unit solid angle) is known:

$$L = k \int E(\lambda) V(\lambda)\, d\lambda \tag{18.1}$$

The luminosity function indicates the relative luminance of monochromatic sources of equal radiance, referred to a monochromatic source of 5550 Å, for which the luminance is a maximum. For this

latter source the ratio of luminance to radiance is given by the constant k and is equal to 680 lumens per watt. The unit of luminance, the candle per square centimeter or lumen per steradian per square centimeter, has been fixed with relation to the luminance of freezing platinum, set at 60 candles per square centimeter. The more commonly employed unit of luminance, the foot-lambert, is defined as $1/\pi$ candle per square foot or $1/\pi$ lumen per steradian per square foot.

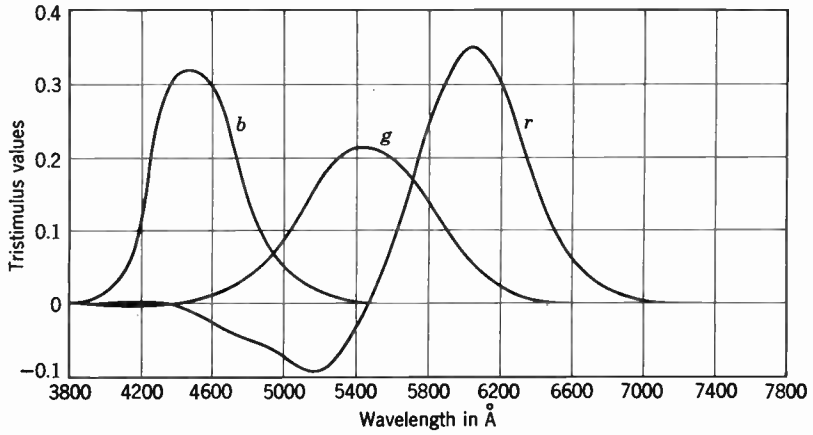


Fig. 18.1. Color Mixture Curves for the Spectral Primaries at 7000, 5461, and 4358 Å, with Equi-Energy White as Reference White. (Wintringham, reference 5.) (Courtesy of *Proceedings of the Institute of Radio Engineers*.)

The luminance of a sample of known spectral composition may thus be determined by summing the luminances of its spectral components. Its chromaticity may, in similar fashion, be specified in terms of the relative intensities of three primaries required to match it. Assume, to begin with, that these primaries are monochromatic sources at 7000 Å (red), 5461 Å (green), and 4358 Å (blue). Furthermore, a new unit of intensity of each of the primary sources is established, such that a white light sample is matched by equal amounts of each primary intensity. It is convenient to choose for this standard white sample a so-called equi-energy white, i.e., one whose spectral energy distribution is uniform. If the luminances of the three unit primary sources are now compared, they are found to be related by the ratios $L_r : L_g : L_b = 1 : 4.4907 : 0.0601$.

As a next step monochromatic sources of equal radiance are matched by the three primary sources. In this manner the *color mixture curves* $\bar{r}(\lambda)$, $\bar{g}(\lambda)$, and $\bar{b}(\lambda)$ shown in Fig. 18.1 are obtained. The curves in

Fig. 18.1 represent a statistical average derived from the efforts of many observers to match, in hue, saturation and brightness, monochromatic sources of equal energy with the accepted three primary sources, just as the luminosity function of Fig. 17.1 represents a statistical average of efforts to match fields of different monochromatic colors in brightness.

With the aid of the color-mixture curves, the color components or trichromatic coefficients R, G, B of an arbitrary light sample of known spectral composition are obtained by summing the chromaticities of the spectral components of the light sample:

$$R = \int E(\lambda)\bar{r}(\lambda)\, d\lambda$$

$$G = \int E(\lambda)\bar{g}(\lambda)\, d\lambda \qquad (18.2)$$

$$B = \int E(\lambda)\bar{b}(\lambda)\, d\lambda$$

For equi-energy white, these sums are given simply by the areas under the three curves in Fig. 18.1. Since the units of intensity of the primaries were chosen so that the chromaticities of equi-energy white would be equal, it follows that the areas under these three curves, also, are equal.

The chromaticity, as against the luminance, of a sample is given by the *relative* magnitudes of the three trichromatic coefficients, R, G, B or the relative trichromatic coefficients r, g, b obtained by dividing R, G, B by their sum:

$$r = \frac{R}{R+G+B}, \qquad g = \frac{G}{R+G+B}, \qquad b = \frac{B}{R+G+B} \qquad (18.3)$$

Since, now, $b = 1 - r - g$, r and g alone suffice to specify the chromaticity of the sample. Any chromaticity can thus be represented by a point on a two-dimensional chart on which r and g represent the abscissas and ordinates. The locus of the spectral chromaticities on this chart (Fig. 18.2) is a horseshoe-shaped curve which, together with the straight line connecting its two endpoints, encloses all known chromaticities.

It follows from the definition of the coordinates of the color points that the color point corresponding to the mixture of two lights lies between the color points of the two components, on the straight line joining them. Specifically, if R_1, G_1, B_1 represent the trichromatic

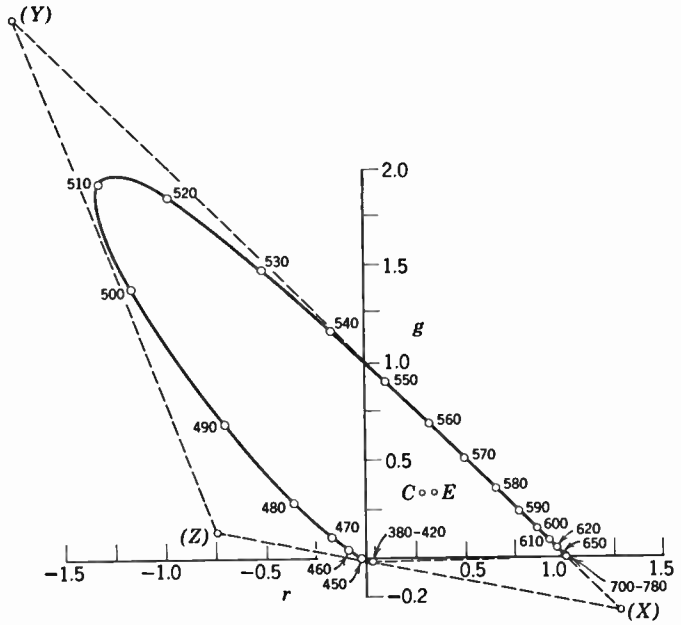


Fig. 18.2. Color Triangle for the Spectral Primaries at 7000, 5461, and 4358 Å. with Equi-Energy White as Reference White. (Wintringham, reference 5.) (Courtesy of *Proceedings of the Institute of Radio Engineers.*)

coefficients of one of the combining lights for unit radiance and R_2, G_2, B_2 are the coefficients for the other light, also for unit radiance, the chromaticity coordinates for the mixture of the two lights with radiance a_1 and a_2 become, respectively,

$$r = \frac{a_1R_1 + a_2R_2}{a_1(R_1 + G_1 + B_1) + a_2(R_2 + G_2 + B_2)} = Kr_1 + (1 - K)r_2$$

$$g = \frac{a_1G_1 + a_2G_2}{a_1(R_1 + G_1 + B_1) + a_2(R_2 + G_2 + B_2)} = Kg_1 + (1 - K)g_2$$

with $K = \dfrac{a_1(R_1 + G_1 + B_1)}{a_1(R_1 + G_1 + B_1) + a_2(R_2 + G_2 + B_2)}$. Since K is inherently positive and less than one, the point (r, g) lies between (r_1, g_1) and (r_2, g_2), on the line joining them.

An immediate consequence of this relation between the point representing the mixture and the points representing the components is that all physically realizable colors lie within the spectral horseshoe curve,

since they may all be obtained by the addition of spectrally pure colors. Furthermore, all chromaticities that may be produced by combining the chosen primaries lie within the triangle with the apices (0, 0), (0, 1), (1, 0). Point E, at ($\frac{1}{3}$, $\frac{1}{3}$), represents equi-energy white, whereas point C indicates so-called artificial daylight. The latter approximates the chromaticity of a black body at 6500°K. The "C illuminant" is readily obtained by employing suitable filters in conjunction with an incandescent lamp at a color temperature of 2848°K.* Whereas the curved part of the perimeter of the horseshoe-shaped figure represents the spectral chromaticities, the bottom boundary corresponds to the purples, obtained by the mixture of violet and red.

The spectral primaries used above are not necessarily most convenient for color specification; so-called filtered primaries, obtained by combining illuminant C with Wratten filters 25, 58, and 47 of prescribed transmission to provide a red, green, and blue source, respectively, are advantageous in a number of respects. In particular, they are more easily matched by the receiver primaries, i.e., the colors of the red, green, and blue partial pictures which, by superposition, yield the reproduced natural-color picture.

The system of primaries which, by international convention, has been adopted for the specification of colors, differs from both the above systems. Both the spectral and the filtered primaries have the drawback that the color-mixture curves (Fig. 18.1) take on negative values. This must be so, since the spectral colors lie outside the triangle defined by the trichromatic coefficients of the primaries and hence cannot be matched by any additive combination of the primaries. The presence of the negative loops increases the errors which may arise in the calculation of the trichromatic coefficients from the spectral distribution of a sample (e.g., by Eq. 18.2) and constitutes, hence, a material drawback for color specification.

The primaries selected by the International Commission on Illumination (CIE primaries),† which are free from this drawback, hence cannot have a physical existence. To further simplify calculations, they are chosen so that one of the trichromatic coefficients (Y) specifies the luminance of the sample, that the spectral colors in the red and yellow are mixtures as nearly as possible of two primaries only (X and Y), and that the triangle (0, 0), (0, 1), (1, 0)—namely, the color points of the primaries—while fully enclosing the horseshoe curve representing the spectral chromaticities, contains as little area outside of

* See Judd, reference 1.

† In the earlier literature the abbreviation ICI commonly replaces CIE.

it as possible, subject to the preceding conditions. In terms of the spectral trichromatic coefficients, the CIE trichromatic coefficients are given by

$$\begin{aligned} X &= 2.7690R + 1.7518G + 1.1300B \\ Y &= 1.0000R + 4.5907G + 0.0601B \\ Z &= 0.0000R + 0.0565G + 5.5943B \end{aligned} \tag{18.4}$$

The CIE tristimulus values $\bar{x}$, $\bar{y}$, $\bar{z}$ for equi-energy monochromatic light samples are plotted in Fig. 17.2; since the horseshoe curve is wholly enclosed in the primary triangle, the tristimulus values are positive throughout. Furthermore, over a considerable range in the red and yellow, one tristimulus value, $\bar{z}$, is zero.

The CIE color chart itself is shown in Fig. 18.3. On it are shown not only the chromaticities of equi-energy white, and that of the C-

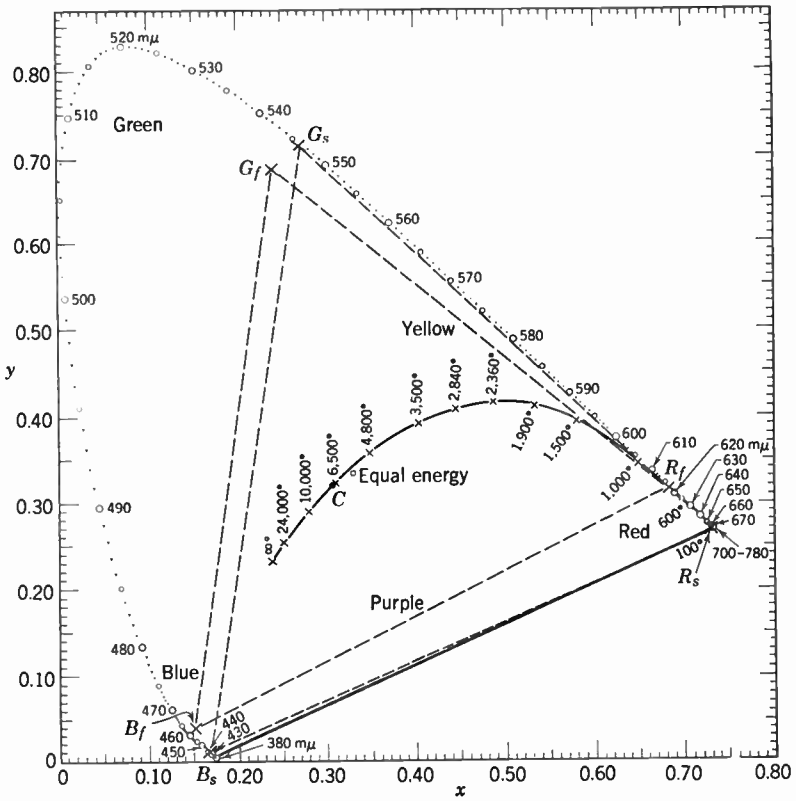


Fig. 18.3. The CIE Color Chart.

illuminant (daylight), C, but also chromaticities and the corresponding triangles of the spectral primaries and the filtered primaries. Finally, the locus of black-body sources at different temperatures is indicated.

The triangle defined by the filtered primaries (R_f, G_f, B_f) is of particular interest, since they correspond closely to the receiver primaries employed in color television systems.* At first sight, it might seem that this triangle encloses only a relatively small portion of physically

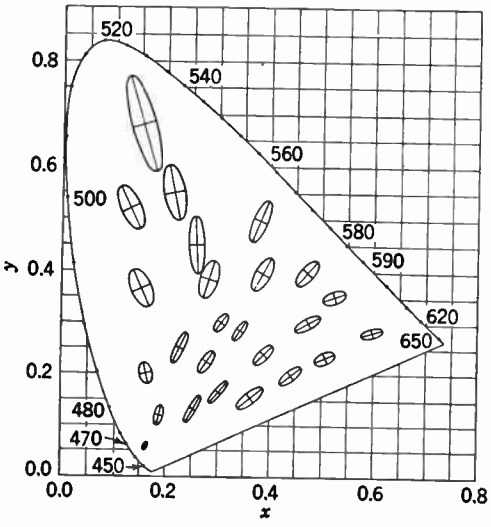


Fig. 18.4. Variation of Chromaticity Discrimination on the CIE Color Chart. (MacAdam, reference 4.) (Courtesy of *Journal of the Optical Society of America.*)

realizable color values, particularly in the range of the greens and blue-greens. However, the steps of chromaticity discrimination are particularly large in this region; a large shift in the chromaticity point represents only a small change in color. This is shown by the chart in Fig. 18.4. Here the radius vectors to the ellipses from their centers represent an equal number of just perceptible steps of difference in chromaticity.

Another way of showing the adequacy of the color range covered by the filtered primaries is by plotting the full range of chromaticities reproduced by pigments and printers' inks in color catalogues on the same chart as the filtered primary triangle (Fig. 18.5). The coverage by the filtered primaries is very evidently more extensive. Finally, it

* The primaries corresponding to the phosphors employed in Tri-Color Kinescopes are indicated on Figs. 17.3 and 18.36.

should be mentioned that the colors which fall outside the triangle represent a high degree of spectral purity. This means that colored surfaces which give rise to them under white illumination reflect in a

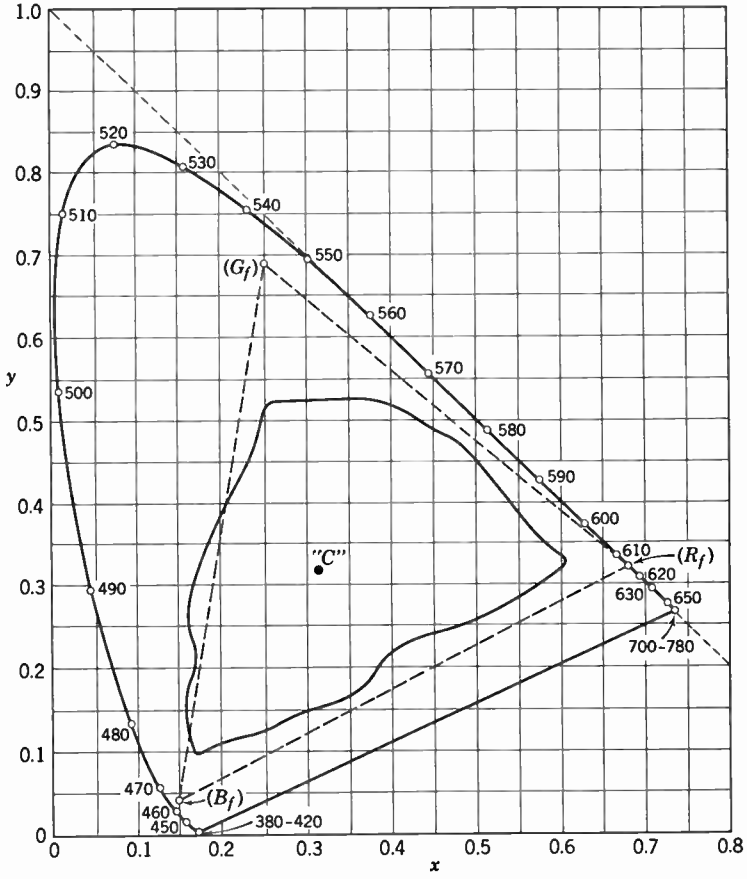


Fig. 18.5. Range of Colors Reproduced by Pigments and Printers' Inks in Color Catalogues. (Wintringham, reference 5.) (Courtesy of *Proceedings of the Institute of Radio Engineers.*)

very narrow spectral range and have hence, necessarily, low luminance. The areas which cannot be accurately reproduced in color are areas of low brightness and, hence, play a small role in the picture.

Two methods may be indicated which permit the determination of the CIE trichromatic coefficients of a light sample without requiring a knowledge of the spectral distribution of the sample. In the first, the

sample is matched with three physical primaries, such as the spectral primaries, and the values of R, G, B so obtained are employed to find X, Y, Z by substitution in Eq. 18.4, or an analogous equation if the selected physical primaries differ from the spectral primaries. In the second, even more direct, method, a tristimulus photometer is employed. The light is distributed by a beam splitter to three multiplier phototubes provided with filters which give them overall spectral responses approximating the three curves $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ in Fig. 17.2. Under these circumstances the meters indicating the photocurrents may be calibrated to yield directly the trichromatic coefficients X, Y, Z. In practice, the spectral response of the $\bar{x}$-multiplier phototube takes account of the major (long-wave) loop of the $\bar{x}$-curve only; the contribution of the short-wave loop to X is obtained by unidirectional cross-feed from the output of the $\bar{z}$-multiplier phototube.* Procedures for transforming, more generally, the trichromatic coefficients from one set of primaries to another set can also readily be inferred from the preceding discussion. Obviously, if r'_r, r'_g, r'_b etc. denote the trichromatic coefficients of the unprimed primaries in terms of the primed primaries, the general transformation equations between an unprimed and a primed system of primaries become

$$\begin{aligned} R' &= r'_r R + r'_g G + r'_b B \\ G' &= g'_r R + g'_g G + g'_b B \\ B' &= b'_r R + b'_g G + b'_b B \end{aligned} \tag{18.5}$$

This set of equations may be solved for R, G, B, leading to

$$R = r_{r'} R' + r_{g'} G' + r_{b'} B', \text{ etc.} \tag{18.6}$$

with $r_{r'} = (g'_g b'_b - b'_g g'_b)/D$,

$r_{g'} = (b'_g r'_b - r'_g b'_b)/D$,

$r_{b'} = (r'_g g'_b - g'_g r'_b)/D$, and

$$D = \begin{vmatrix} r'_r & r'_g & r'_b \\ g'_r & g'_g & g'_b \\ b'_r & b'_g & b'_b \end{vmatrix}$$

These relations permit the determination of the new set of trichromatic coefficients, provided that the coefficients of the new primaries in terms of the old primaries are known.

* See Sziklai, reference 7.

The problem is rendered somewhat more complex if the luminance and the relative trichromatic coefficients are given for an arbitrary system of primaries and if, furthermore, the reference white (the color corresponding to equal intensities of the three primaries) is changed. Procedures for carrying out transformations in such instances may be found in the literature.*

The conventional description of a light sample by the luminance (Y) and two relative CIE trichromatic coefficients (x and y) may be modified by the introduction on the color chart of polar coordinates with origin at the point corresponding to neutral white (e.g., the point C). The radial coordinate is now given by the ratio of the distance of the color point from C to the length of the same radius vector extended to the horseshoe-shaped curve, whereas the azimuth is given by the wavelength corresponding to the intersection of this radius vector with the spectral locus. For points in the purple sector, the wavelength of the intersection with the spectral locus of the radius vector extended backwards is specified, and the radial coordinate is given a negative sign (Fig. 18.6).

Whereas the description of color by the trichromatic coefficients is best suited for colorimetric computations, it can readily be seen that the description given by the polar diagram corresponds more closely to our instinctive classification of colors. Thus, for fixed azimuth, all colors are produced by mixing a spectrally pure color (or a purple) with different amounts of white. All, e.g., a delicate pink, a rich pink, and a full red, are recognized as having the same *hue*. All the colors along the radius vector extended in the opposite direction have the *complementary hue*, in the above example a particular shade of green, with various admixtures of white. It follows from the properties of the color chart that two colors of complementary hues, added in appropriate proportions, yield the neutral white corresponding to the origin of the polar diagram. The hue of any color is specified by the wavelength of the spectral color of the same hue or, for the purples, by the wavelength of the spectral color of the complementary hue.

The radial coordinate indicates the relative proportion of the pure spectral color (or pure purple) and white, which, on addition, produce the color in question. It is called the "saturation" or "chroma." A saturation 1 corresponds to a pure spectral color or pure purple; a saturation 0, of course, to neutral white.

* See, e.g., Wintringham, reference 5.

The parameters, brightness, hue, and saturation (also called purity or chroma) not only correspond to our instinctive description of light values, but are also found to have their definite counterparts in the picture signal transmitted in the dot-simultaneous system of color television. The similar parameters of value (corresponding to the

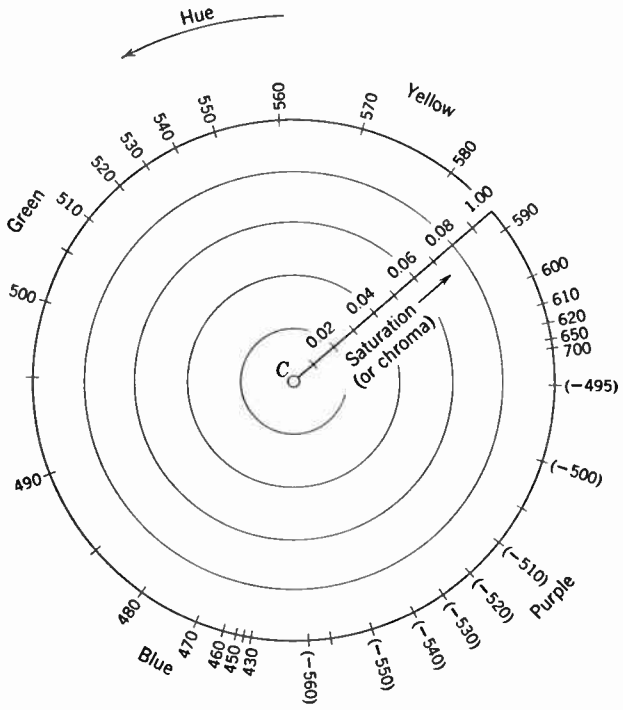


Fig. 18.6. Specification of Chrominance in Polar Coordinates.

square root of the relative luminance), hue, and chroma are employed by Munsell to describe the reflectivities of his well-known system of color cards.

18.2 Color Reproduction by a Television System. The basic conditions for faithful color reproduction by a television system can be laid down without reference to the mechanism employed to effect the reproduction. Similarly, it is possible to determine quite generally the nature of the color distortions which may result from failure to observe these conditions. The only assumption regarding the system of color television will be that it includes a triple pickup system, with three different spectral responses, and a reproducing system in which

the picture signal generated by each spectral response forms a partial picture of the color of one of three receiver primaries (Fig. 18.7). It is a matter of indifference whether the partial pictures are presented simultaneously or in temporal sequence, as well as whether the different color signals utilize three separate channels or, by some system of multiplexing, utilize the same channel.

It is evident from the start that the reproduction of a large gamut of colors demands the use of receiver primaries, i.e., phosphors or phosphor-filter combinations producing the red, green, and blue component pictures, which include as large an area as possible of the color

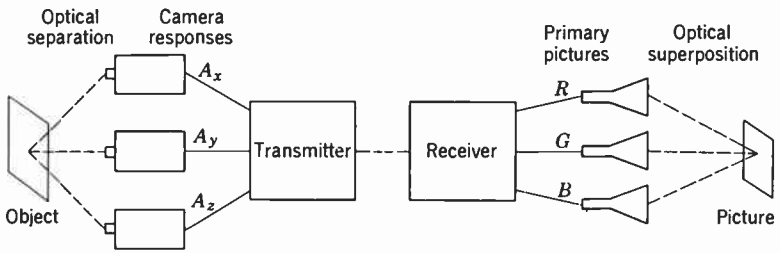


Fig. 18.7. General Color Television System.

triangle. It has been mentioned, furthermore, that the so-called filtered primaries, represented by R_f, G_f, and B_f in Fig. 18.3, are quite satisfactory from this point of view and, in addition, can be realized physically to a close approximation.

It is also easy to lay down the basic requirement for the transmission channel. If color values are to be undistorted, the transmission channel must be linear: the brightness of any one partial image must be directly proportional to the light falling on the corresponding pickup element. Suppose that this is not fulfilled, that, e.g., the picture brightness increases more rapidly than linearly with the light entering the pickup system. Then an orange tone, matched by a small green and a large red trichromatic component, will appear reddened in the reproduction, a yellow-green, greener than in the original. Quite generally, all colors will tend to be pushed toward their nearest primary colors. Conversely, with subnormal contrast, the colors will be shifted away from the primaries, giving a washed-out appearance to the picture.

The appropriate spectral response of the pickup system or camera remain to be established. For this purpose, consider the condition for the faithful reproduction of a spectrally pure color of unit radiance. If

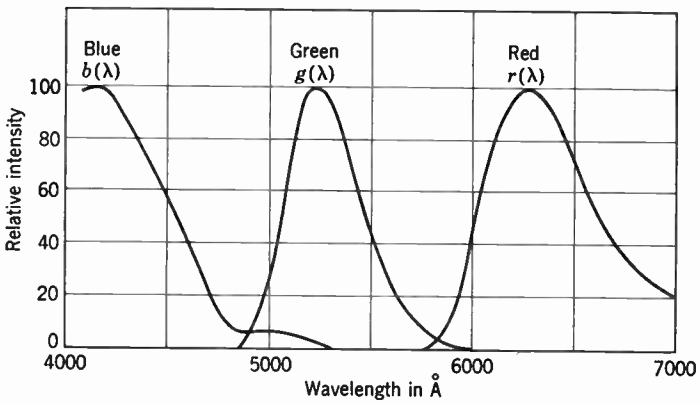


Fig. 18.8. Receiver Primaries Employed in RCA's November, 1949, Demonstrations of a Dot-Simultaneous Color Television System.

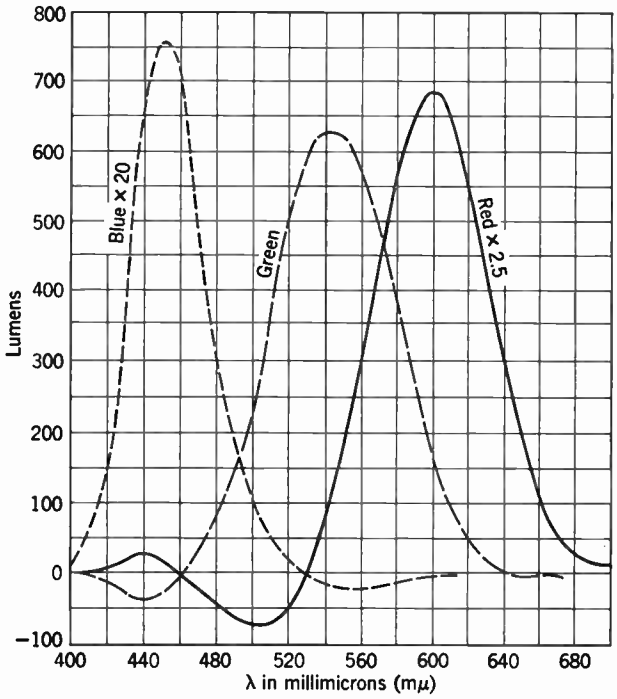


Fig. 18.9. Color-Mixture Curves for the Receiver Primaries in Fig. 18.8. (Epstein, reference 11.)

all spectrally pure colors are properly reproduced by the system, any color whatsoever will also be reproduced properly since, physically, it is formed by a combination of spectrally pure radiations. Now, any spectrally pure color of unit radiance is matched by the three receiver primaries if their amplitudes are adjusted to the ordinates of the cor-

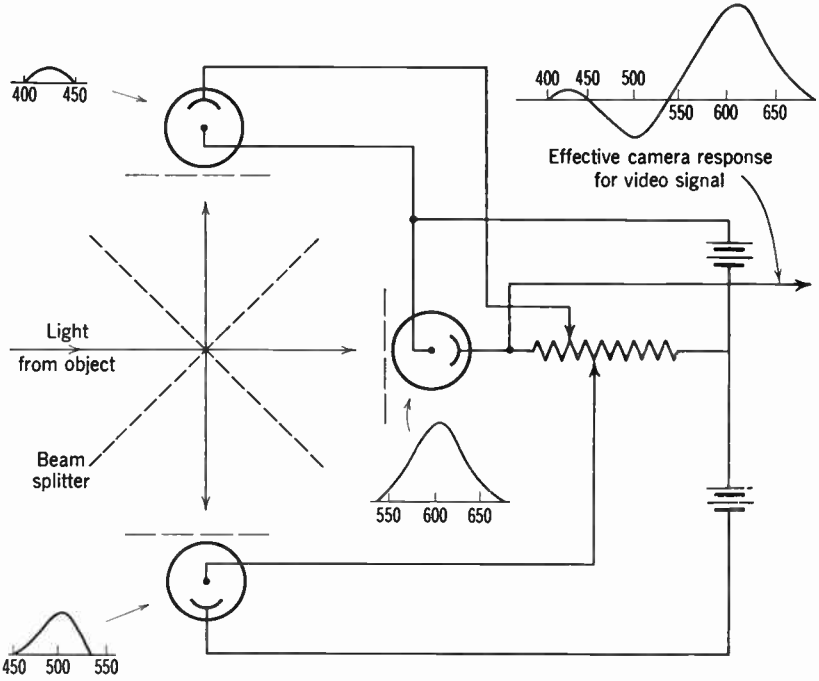


Fig. 18.10. Addition of Outputs from Several Photosensitive Devices to Match Positive and Negative Loops of Color-Mixture Curves.

responding color-mixture curves. In brief, faithful color reproduction demands that the spectral responses of the pickup system be given by the color-mixture curves for the receiver primaries.

The receiver primaries employed in a particular dot-simultaneous television system are shown in Fig. 18.8, the corresponding color-mixture curves in Fig. 18.9.* These primaries correspond quite closely to the filtered primaries mentioned in the last section. It is noted that the color-mixture curves possess, in addition to their principal positive peak, smaller negative loops. At first sight, it might appear that this would make it impossible to realize the prescribed spectral responses

* See Epstein, reference 11.

since photocurrents are, normally, unidirectional.* However, by adding the outputs of several photosensitive devices with spectral responses corresponding to the positive and negative loops, respectively, and a polarity corresponding to the polarity of the loop (Fig. 18.10), an arbitrary response curve can be matched.

As an alternative † a pickup system may be employed which has spectral responses that are positive throughout, as exemplified by the color-mixture curves for the CIE primaries. Each of the CIE primaries can be matched by a certain known combination of the receiver primaries, in positive and negative amounts. In terms of the notation of Eqs. 18.5 and 18.6, these amounts are the coefficients r_x, g_x, b_x; r_y, g_y, b_y appearing in the transformation equations

$$\begin{aligned} R &= r_x X + r_y Y + r_z Z \\ G &= g_x X + g_y Y + g_z Z \\ B &= b_x X + b_y Y + b_z Z \end{aligned} \tag{18.7}$$

The same coefficients may be expressed in terms of the CIE trichromatic coefficients of the receiver primaries x_r, y_r, z_r, etc., by the scheme of Eq. 18.6:

$$r_x = (y_g z_b - z_g y_b)/D, \qquad r_y = (z_g x_b - x_g z_b)/D, \qquad D = \begin{vmatrix} x_r & x_g & x_b \\ y_r & y_g & y_b \\ z_r & z_g & z_b \end{vmatrix}$$

Specifically, for the spectral primaries 7000 Å, 5461 Å, and 4358 Å with equi-energy white as reference white, a solution of Eq. 18.4 leads to

$$\begin{aligned} R &= 0.41847X - 0.09117Y + 0.00092Z \\ G &= -0.15867X + 0.25243Y - 0.00255Z \\ B &= -0.08283X + 0.01571Y + 0.17860Z \end{aligned} \tag{18.8}$$

If now the three camera signals, which correspond to the CIE tristimulus values for the transmitted picture, are distributed in accord with Eq. 18.7 or Eq. 18.8 to the three receiver channels, the receiver primaries are excited in the proper ratio to reproduce the original color (Fig. 18.11). It is evident that this procedure is restricted to simultaneous systems of signal generation.

* The quenching of dark current in photoconductors by radiations within certain wavelength ranges suggests that, at least in principle, it might be possible to match the required responses.

† See Cherry, reference 12.

More precisely, both the last systems make it possible, in principle, to reproduce exactly all colors within the triangle defined by the receiver primaries. For points outside of this range, the signal applied to at least one of the red, green, and blue reproducing organs is driven negative, beyond cutoff. This has the effect of displacing the corresponding chromaticities in the reproduction to the periphery of the receiver primary triangle.

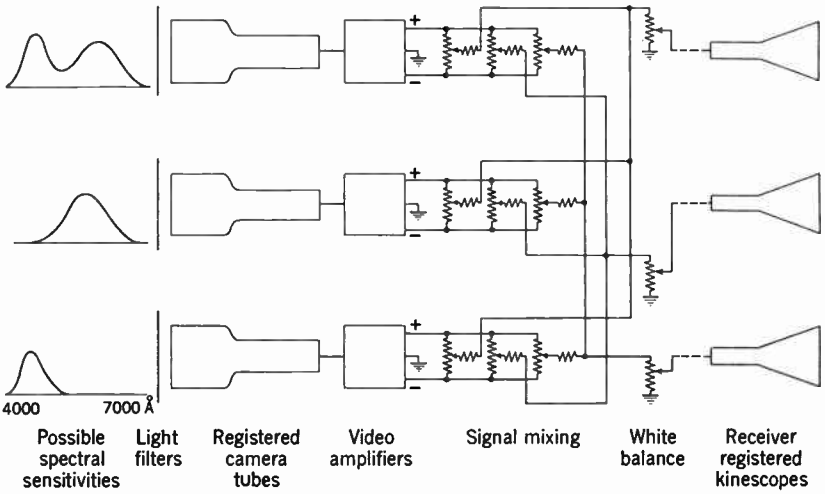


Fig. 18.11. Attainment of Correct Color Reproduction by Forming Appropriate Linear Combinations of the Camera Outputs. (Cherry, reference 12.) (Courtesy of *Journal of the Society of Motion Picture and Television Engineers.*)

The intrinsic possibility of perfect color reproduction resident in the methods of multiple photosensitive units in the camera and of signal mixing outlined above is offset by practical difficulties. They are primarily inefficient light utilization in the first instance and the necessity of a high degree of stability and careful adjustment of electronic circuits as well as a lowered signal-to-noise ratio in the second instance.

For these reasons, it has been general practice to fit the camera sensitivities as closely as possible to the principal positive loops of the color-mixture curves for the receiver primaries and to neglect the smaller negative loops. Thus the camera sensitivities employed in the color television system mentioned on page 813 are shown in Fig. 18.12. They are to be compared with the color-mixture curves for the receiver primaries shown in Fig. 18.9. If the amplifications of the three color channels are adjusted so that a white surface is reproduced correctly,

the other colors will be reproduced with a fidelity depending on the deviation of the camera sensitivities from the color-mixture curves for the receiver primaries. The color changes obtained with a linear transmission system are represented on the color chart in Fig. 18.13. The arrows indicate the displacement of the color points in the reproduction as compared with the original. The point E, representing the studio illuminant (equi-energy white) is not shifted since the channel amplifications have been adjusted to reproduce it correctly.

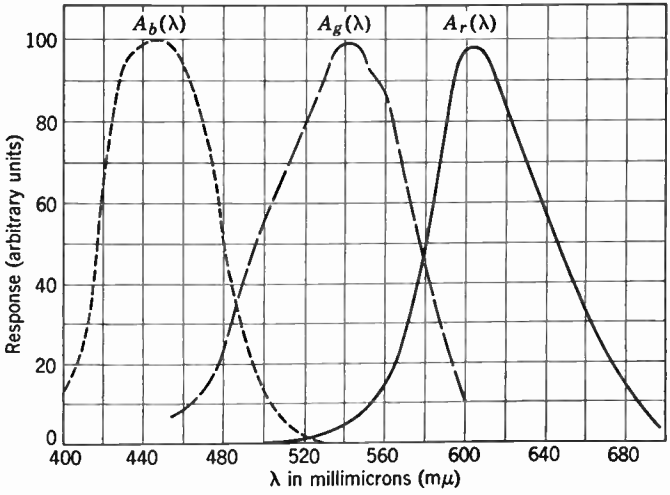


Fig. 18.12. Camera Responses Employed with Receiver Primaries Shown in Fig. 18.8.

The color reproduction shown is least faithful in the blue, if the minimum perceptible color difference is taken as unit of the color shift. The error is much less for the red and the green and becomes negligible, of course, for colors of low saturation. The degree of color fidelity shown is probably adequate for entertainment purposes. There are, however, examples of the use of television for scientific and educational purposes where greater fidelity might well be desirable. It then becomes necessary to resort to the use of multiple pickups or signal mixing to increase the precision of color reproduction.

The viewpoint accepted throughout in the preceding discussion is that the most faithful color reproduction is also the most desirable color reproduction. This is not true in all cases. For example, tests have shown that a majority of judges tend to reject the exact reproduc-

tion of facial coloration as unsatisfactory, preferring paler tints.* Apart from this, the tolerance of deviations from the original color (or its preferred reproduction) varies greatly with the familiarity of the subject matter as well as the brightness and dimensions of the colored areas reproduced. Thus, again, the viewer tends to be more

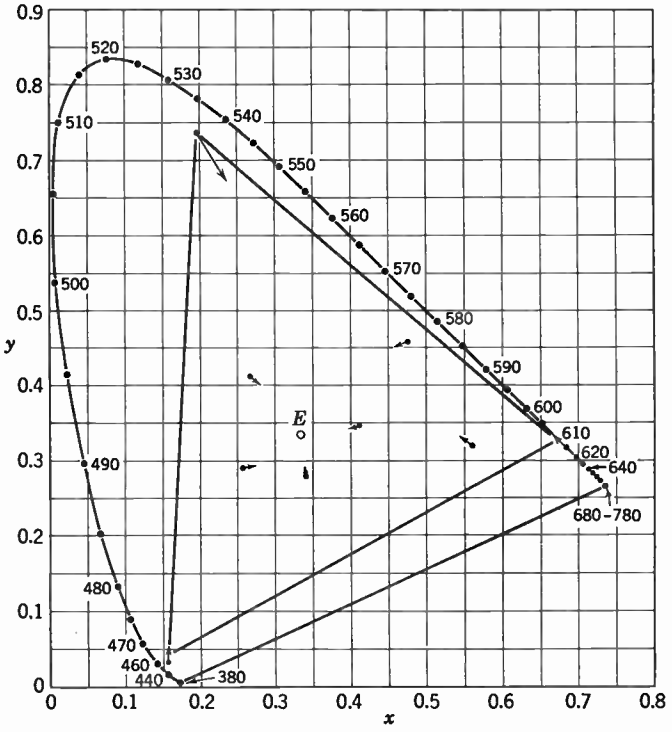


Fig. 18.13. Errors in Color Reproduction Resulting from Use of Camera Responses in Fig. 18.12 with Receiver Primaries in Fig. 18.8. (Epstein, reference 11.)

critical of flesh tints than of other portions of the picture. More generally, a large amount of data will have to be amassed before a rational system of tolerable limits of deviation can be set up.

18.3 The Discrimination of Chromaticity Variations and the Mixed-Highs Principle. It is one of the significant points of superiority of simultaneous color television systems over sequential systems that they can utilize the reduced discrimination of the eye for chromaticity differences to enhance the detail in a color picture which may be trans-

* See MacAdam, reference 4.

mitted with a given bandwidth. This fact has already been brought out in the last chapter. It remains to indicate the method of determining the difference in the discrimination of chromaticity variations and luminance variations and to show once more how the findings are applied in specific instances.

In the tests devised by A. V. Bedford * observers were presented a small divided field with patterns of the type shown in Fig. 18.14, sur-

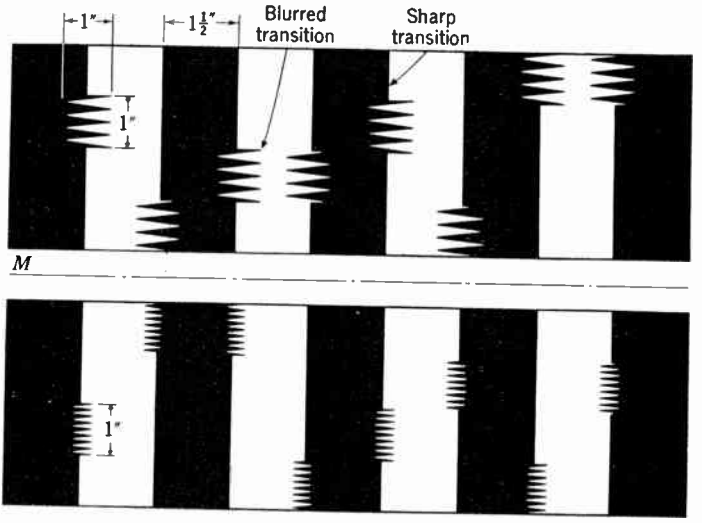


Fig. 18.14. Test Patterns Employed for Determining the Relative Acuity of the Eye for Luminance and Chrominance Differences. (Bedford, reference 8.) (Courtesy of *Proceedings of the Institute of Radio Engineers.*)

rounded by a uniformly illuminated white background with a brightness of 10 foot-lamberts and were asked to locate the (unresolved) blurred regions at several different viewing distances. The reciprocal of the angle which the pattern subtended at a distance for which the identification was 50 percent successful was taken as measure of the acuity of the eye for the transition between two fields A and B. In the first tests, area A was made black and area B either white (20 foot-lamberts) or green, red, or blue with a luminance equal to that required to produce a white-light brightness of 20 foot-lamberts by addition; this luminance was arbitrarily denoted as 100 percent. In the second set of tests, area A was made either red or blue and area B green or red, with the intensity adjusted so as to minimize the apparent

* See Bedford, reference 8.

difference between the two fields. The results of the tests are summarized in Table 18.1.

TABLE 18.1. RELATIVE VISUAL ACUITIES

Color Combination	Relative Brightness Values Used in Tests	Acuity in Percent
White-black	White = 100 percent	100
Green-black	Green = 100 percent	94
Red-black	Red = 100 percent	90
Blue-black	Blue = 100 percent	26
Green-red	Red = 100, green = 49 percent	40
Green-blue	Blue = 100, green = 3.3 percent	19
Red-blue	Blue = 100, red = 6.3 percent	23

The most important conclusions of these measurements are (1) that the acuity of the eye for chromaticity differences is less than half as great as for luminance differences, and (2) that detail in the blue partial picture is not perceived unless it is approximately four times as coarse as the finest black-and-white detail that can be resolved. These results may also be interpreted in the following manner. If signals giving the luminance information and the chrominance information are transmitted separately and the passband of the luminance signal is adjusted relative to the viewing distance of the observer so that the finest horizontal detail which is transmitted is barely resolved, the passband of the chrominance signal may be made less than half that of the luminance signal and the passband of the blue signal one-fourth that of the luminance signal, without detracting in any way from the observed picture.

The relationship between the required passband and the acuity tests may be made plainer by the following consideration. Figure 18.15 shows the difference between the sharp transition and the blurred transition and the Fourier representation of this difference. As might be expected, the Fourier representation has a rather sharp peak at the point where the width of diffusion corresponds to one wavelength. If D refers to the retinal image of the transition and the cutoff wavelength λ_c of the aperture admittance of the eye (in the sense of Chapter 5) is appreciably greater than $D(D/\lambda_c \leqq K < 1)$, the eye will not perceive a difference between the blurred and sharp transition. The tests in effect measure the relative value of the cutoff wavelength λ_c of the aperture admittance of the eye for different brightness and color transitions, K being fixed by the requirement that the difference between the sharp and the blurred transition just disappears.

It was shown in Chapter 5 that the bandwidth requirements of a television system were fixed so as to match the frequency response of the television system to the aperture admittance of the eye at normal viewing distance. The relation between wavelength and frequency is

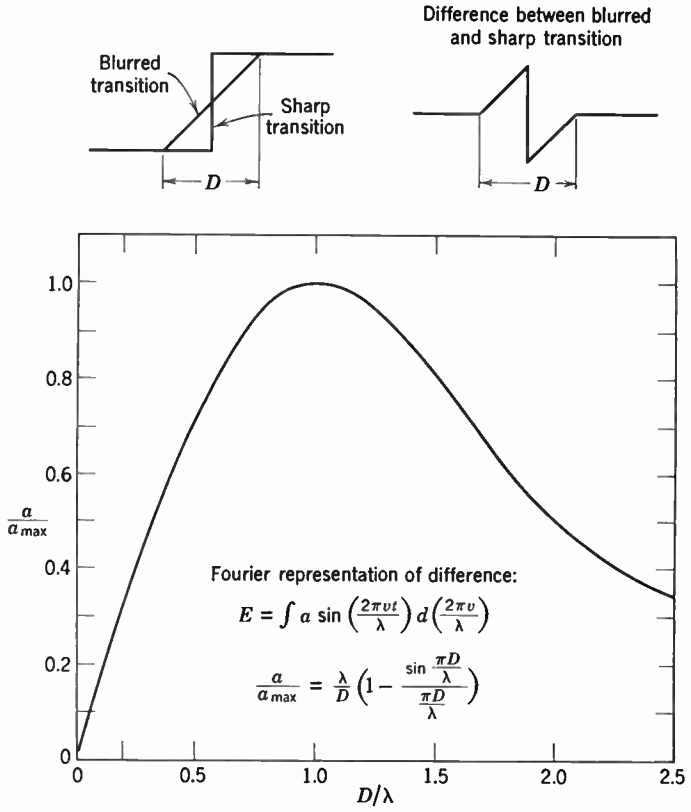


Fig. 18.15. Difference between Blurred and Sharp Transition and Frequency Spectrum of This Difference.

here of course $1/\lambda = v/v$, where v is the linear scanning speed. Extending the passband appreciably beyond the limits set by the cutoff wavelength of the eye would be uneconomic and would not contribute to a perceptible picture improvement. Similarly, since the cutoff wavelength for chromaticity differences was found to be over twice as great as that for luminance differences, no advantage is gained by making the passband for the chrominance signal more than half that for the luminance signal.

A three-channel simultaneous system which makes full use of the resulting permissible bandwidth economy is shown in Fig. 18.16. It is here assumed that the three camera responses correspond to the conventional "filtered receiver primaries" and that the individual camera channel amplifications have been adjusted so that equal signal amplitudes correspond to white. Then, according to Bedford's measure-

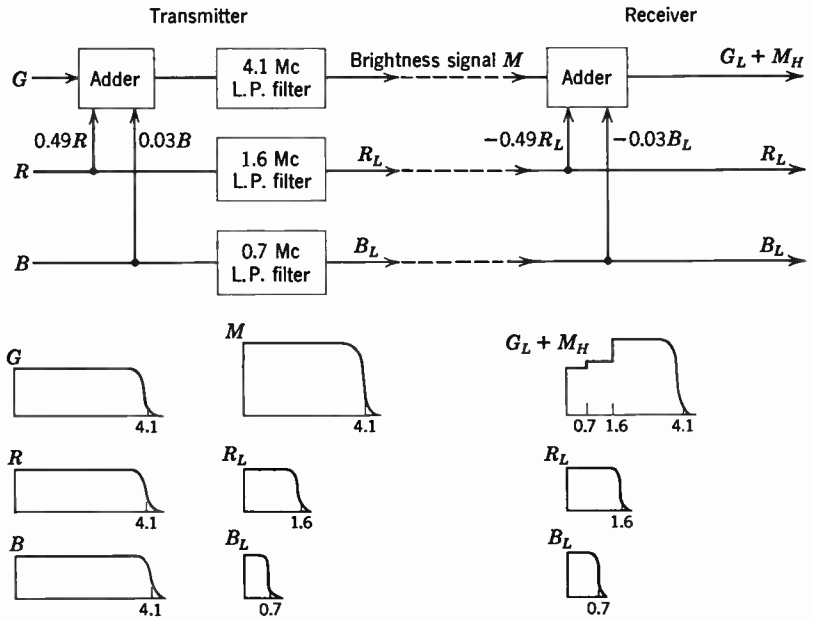


Fig. 18.16. Three-Channel Simultaneous System Making Full Use of Mixed-Highs Principle.

ments, the contribution to picture brightness of equal green, red, and blue signals is as 100:49:3.1. Accordingly, by adding to the green signal G 49 percent of the red signal R and 3 percent of the blue signal B, a picture signal M ("monochrome") is obtained, which corresponds to the luminance variations in the picture. This is transmitted with the full video bandwidth of 4 megacycles. If applied to a black-and-white receiver, it will reproduce the original scene with proper tonal gradation and unimpaired resolution. The red signal, on the other hand, will be passed through a 1.6-megacycle low-pass filter and the blue signal through a 0.7-megacycle low-pass filter before transmission. In a simultaneous receiver appropriate fractions (49 and 3 percent, respectively) of the red and blue signals are subtracted from the mono-

chrome signal to recover the green signal, to which now the *mixed-high*-frequency components M_H of all three channels are added. In the natural color picture obtained by the superposition of the green, red, and blue partial images, color is accurately reproduced for areas exceeding about six picture elements in a horizontal direction; variations in the blue component which cover smaller areas are replaced by variations in the green component of equal brightness; similarly, brightness variations of the red component covering only one or two picture elements are reproduced as equal brightness variations of the green components, whereas red brightness variations covering appreciably larger areas are correctly reproduced. In brief, by rendering intensity fluctuations for which the eye is color-blind in monochrome, the total video bandwidth required for transmitting the color picture without deterioration has been reduced from 12 to 6.3 megacycles. More recent observations indicate that even materially smaller transmission bands for the red and the blue component are permissible, as has already been noted in the last chapter.

The method of application of the mixed-highs principle is not unique. Thus, the contribution of the blue component to the high-frequency brightness fluctuations is generally so small that it may be neglected. Furthermore, the mixed-high signal may be applied to both the green and the red channels, so that it represents a fluctuation in the white content of the picture.* Figure 18.17 shows a system in which this is carried out. Here, the mixed highs, derived from the green and red camera channels, are added to the green signal, which, like the red and blue signals, is passed through a low-pass filter. In the receiver, the mixed highs, selected from the green signal by a high-pass filter, are added to the red signal. The green signal with the mixed highs generally approximates the complete luminance signal sufficiently closely to yield a high-quality monochrome reproduction of the scene.

The application of the mixed-highs principle to a dot-sequential system of color television is indicated in Figs. 18.18 and 18.19. In the

* Here, fluctuations in white are reproduced exactly; detail fluctuations in the brightness of any one color will be reproduced as fluctuations in chroma (high brightness = low chroma) as well as brightness. In the system shown in Fig. 18.16, fluctuations in green will be reproduced exactly; detail fluctuations in the brightness of white will appear as fluctuations between (low-brightness) purplish white and (high-brightness) greenish white. Similarly, detail fluctuations in red will be reproduced as fluctuations between (low-intensity) purplish red and (high-intensity) orange red. As long as these detail fluctuations are confined to a range in which the eye does not discriminate colors, these differences are, of course, entirely immaterial.

transmitter, the individual camera signals are passed through low-pass filters and sampled. The sampler output (after filtering) then consists of the sum of the low-frequency brightness signal plus a sampling frequency component which is phase- and amplitude-modulated by the

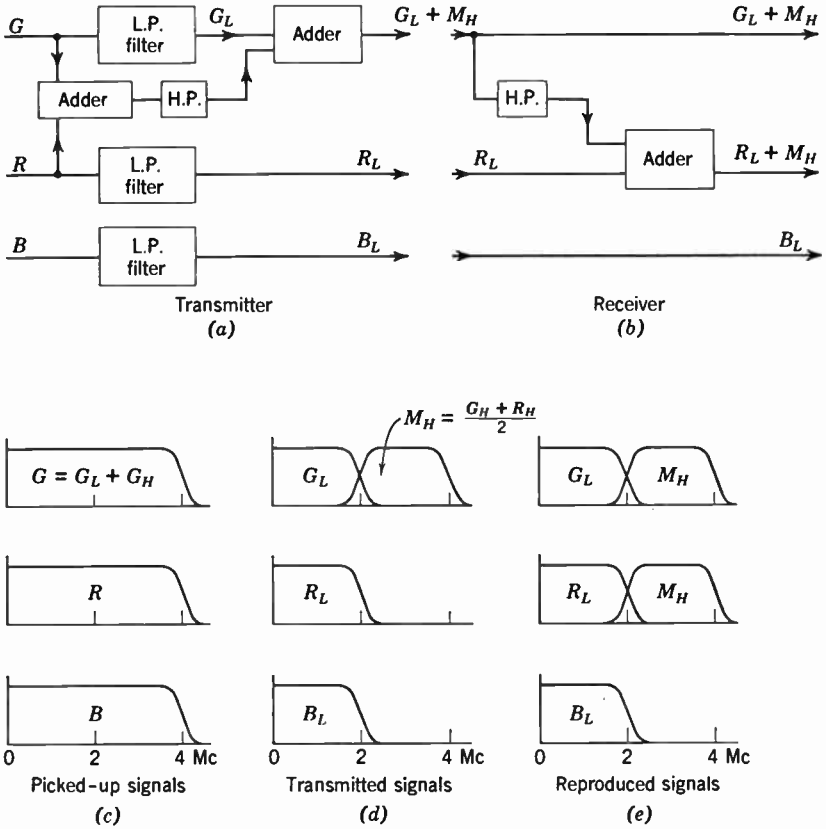


Fig. 18.17. Color Television System with Mixed Highs Reproduced as High-Frequency Variations in White Content of Picture. (Bedford, reference 8.) (Courtesy of *Proceedings of the Institute of Radio Engineers.*)

chrominance information contained in the picture. The mixed-highs signal, obtained by adding the camera signals and passing them through a bandpass filter, supplements the sampler output to provide the complete video signal. In the receiver (Fig. 18.19), the low-frequency components are recovered by passing the complete video signal through a desampler and removing the high-frequency components of the desampler output by filtering; the mixed-highs signals, selected by a

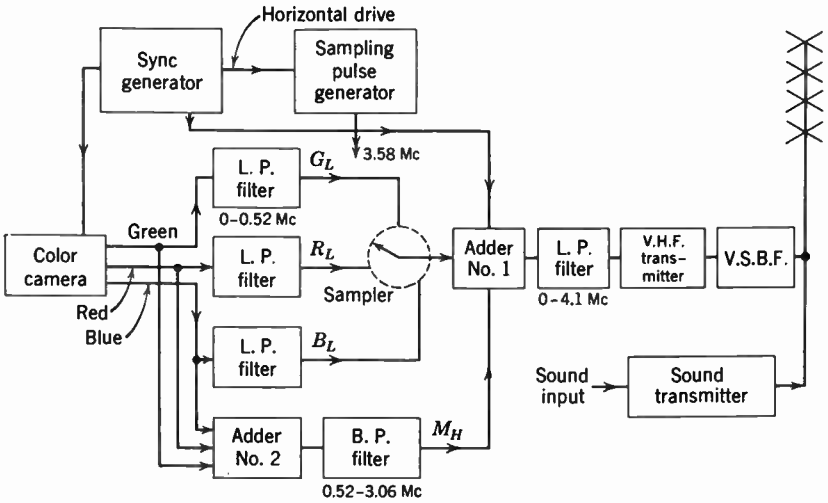


Fig. 18.18. Dot-Sequential Transmitter with Mixed Highs.

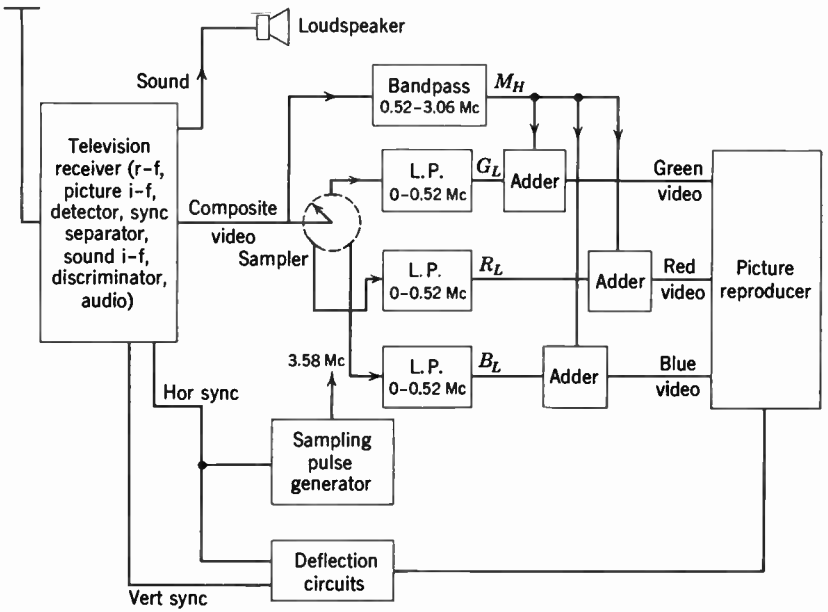


Fig. 18.19. Dot-Sequential Receiver with Bypassed Highs.

bandpass filter from the complete signal, are added to the red and green channels to reproduce fine detail. Apart from the appearance of spurious signals, which will be discussed later, the application of the mixed-highs principle is here the same as in the three-channel simultaneous system of Fig. 18.17.

In the bypassed monochrome system, as exemplified by Figs. 17.20 and 17.21, the brightness and color signals are completely separated, so that the bandwidth saving obtained through application of the mixed-highs principle is achieved by transmitting the brightness signal with the full video passband and limiting the chrominance signal by suitable filters in the modulator and demodulator outputs and inputs. This

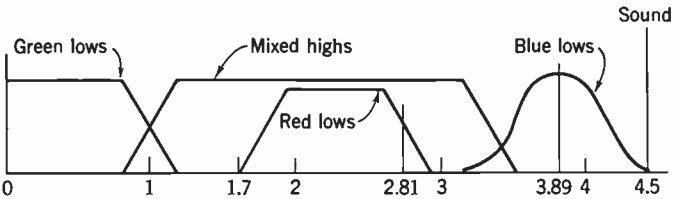


Fig. 18.20. Frequency Interlace Color Television System. Disposition of subcarriers and sidebands.

system, in its various modifications, will form the subject of most of the remainder of this chapter.

In one form of frequency interlace system, finally, two color subcarriers, for the red and the blue low-frequency components, respectively, are provided. Sideband filters prevent overlap between the two color signals (Fig. 18.20). However, the red color signal overlaps the mixed-highs (i.e., green and red high-frequency component) signal which is added to the low-frequency green channel. In the receiver, the mixed-highs signal is added to the demodulated red and blue signals.

In conclusion, it may be noted that the permissible reduction in the bandwidth for the color signal indicated by Bedford's tests has been found to be conservative. Viewer tests of television pictures indicate that the chrominance bandwidth may be reduced to one megacycle without perceptible picture degradation.

18.4 Dot-Sequential and Dot-Simultaneous Color Television. In the following, the character of the signal obtained with a simple dot-sequential system such as is illustrated in Fig. 17.13 will be considered in a more quantitative fashion. It will be assumed that the preamplifier gains in the camera have been adjusted so that the three signals G, R, B delivered to the sampler are equal in amplitude for portions of the scene which correspond to some selected standard white; the

usual choice is the CIE C illuminant, approximating average daylight or a 6500°K source. With infinitely narrow sampling with a frequency $\omega_s/(2\pi)$ at phase angles 0, θ_r, θ_b for the green, red, and blue channels and for the specific sinusoidal camera signals

$$G = A_g \cos \omega_g t, \qquad R = A_r \cos \omega_r t, \qquad B = A_b \cos \omega_b t \tag{18.9}$$

the sampler output becomes

$$\begin{aligned} S &= A_g \cos \omega_g t(1 + 2 \cos \omega_s t + 2 \cos 2\omega_s t + \cdots) + \\ &\quad A_r \cos \omega_r t(1 + 2 \cos (\omega_s t + \theta_r) + 2 \cos 2(\omega_s t + \theta_r) + \cdots) + \\ &\quad A_b \cos \omega_b t(1 + 2 \cos (\omega_s t + \theta_b) + 2 \cos 2(\omega_s t + \theta_b) + \cdots) \\ &= A_g\{\cos \omega_g t + \cos (\omega_s - \omega_g)t + \cos (\omega_s + \omega_g)t + \cos (2\omega_s - \omega_g)t \\ &\quad + \cos (2\omega_s + \omega_g)t + \cdots\} + A_r\{\cos \omega_r t + \cos [(\omega_s - \omega_r)t + \\ &\quad \theta_r] + \cos [(\omega_s + \omega_r)t + \theta_r] + \cos [(2\omega_s - \omega_r)t + 2\theta_r] \\ &\quad + \cos [(2\omega_s + \omega_r)t + 2\theta_r] + \cdots\} + A_b\{\cdots\} \end{aligned} \tag{18.10}$$

This signal is filtered by the video low-pass filter with a cutoff frequency $\omega_c/(2\pi)$ (corresponding, e.g., to 4.1 megacycles), modulates a carrier, and is transmitted. In the receiver, the filtered signal is recovered at the second detector and is applied to the receiver sampler (or desampler), which is synchronized with the transmitter sampler. For reasons which will appear presently, the sampling frequency $\omega_s/(2\pi)$ is always chosen high enough and the passband of the low-pass filters preceding the transmitter sampler, which determine the upper limit of ω_g, ω_r, and ω_b, small enough, that the terms with $2\omega_s$ and higher multiples of ω_s in Eq. 18.10 are removed by the video filter.

The green signal obtained from the desampler then becomes

$$\begin{aligned} G_1 &= KS(1 + 2 \cos \omega_s t + 2 \cos 2\omega_s t + \cdots) \\ &= KA_g\{\cos \omega_g t + \cos (\omega_s - \omega_g)t + \cos (\omega_s + \omega_g)t + \cdots \\ &\quad + \cos \omega_g t + \cos (\omega_s - \omega_g)t + \cos (\omega_s + \omega_g)t + \cdots + \\ &\quad \cos \omega_g t + \cos (\omega_s - \omega_g)t + \cos (\omega_s + \omega_g)t + \cdots\} \\ &\quad + KA_r\{\cos \omega_r t + \cos (\omega_s - \omega_r)t + \cos (\omega_s + \omega_r)t + \cdots \\ &\quad + \cos (\omega_r t - \theta_r) + \cos [(\omega_s - \omega_r)t + \theta_r] + \\ &\quad \cos [(\omega_s + \omega_r)t - \theta_r] \cdots + \cos (\omega_r t + \theta_r) + \\ &\quad \cos [(\omega_s - \omega_r)t - \theta_r] + \cos [(\omega_s + \omega_r)t + \theta_r] \cdots\} \\ &\quad + KA_b\{\cdots\} \end{aligned} \tag{18.11}$$

Here K is a general constant which represents the gain of the transmission channel. The dots generally represent terms with $2\omega_s$, etc., which would be removed by a low-pass filter with cutoff $\omega_c/(2\pi)$ between desampler and reproducer; in the last line they replace a term completely analogous to that multiplying KA_r. Finally, successive lines in Eq. 18.11 correspond to successive terms in the expression for S (Eq. 18.10), omitting terms removed by the video filter.

It is seen immediately that if a filter between desampler and reproducer removes the terms in Eq. 18.11 which contain ω_s, the first term in this equation reduces simply to $3KG$, i.e., becomes identical with the green camera signal except for a constant multiplier representing channel gain. The removal of the unwanted terms without loss of the signal itself is possible only if ω_g is less than $\omega_s - \omega_g$. This leads to the following condition for the low-pass filter preceding the transmitter sampler:

$$(1) \qquad \omega_g \leq \omega_s/2 \tag{18.12}$$

Since the resolution in the picture is limited by the maximum value of ω_g (and the identical maximum value for ω_r and ω_b), it is evidently desirable to make the sampling frequency as high as possible. An upper limit is set, however, by the requirement that, e.g., the red and blue camera signals do not add spurious terms to the green signal G_1 at the receiver. Since (with the symmetrical sampling here assumed) $\theta_r = 120°$, $\theta_b = 240°$, the terms in Eq. 18.11 which are multiplied with A_r and A_b, respectively, cancel:

$$\cos a + \cos(a - \theta) + \cos(a + \theta)$$
$$= \cos a(1 + 2\cos\theta) = 0 \text{ for } \theta = 120° \text{ or } 240°$$

This is obvious, since it represents simply the projection on the x-axis of the resultant of three unit vectors forming 120 degrees with each other (Fig. 18.21). However, if one or two of the terms are removed by the video filter, the sum will no longer be equal to zero, and the red and blue camera channels will contribute to the intensity distribution in the green partial image. This is prevented only if the sum of the sampling frequency and the maximum values of ω_r, ω_b, and ω_g is less than the cutoff frequency:

$$(2) \quad \omega_s + \omega_{g\,\max} = 3\omega_s/2 \leq \omega_c; \qquad \omega_{s\,\max} = 2\omega_c/3;$$
$$\omega_{g\,\max} = \omega_c/3 \tag{18.13}$$

Thus, the maximum horizontal resolution which can be obtained with this simple system is just one-third that obtainable with a standard

monochrome television system with the same frequency bandwidth, and the optimum sampling frequency is two-thirds of the cutoff frequency. If the cutoff frequency is 4.1 megacycles, the sampling frequency is 2.73 megacycles, and the passband for the three individual color channels, 1.37 megacycles. This is exactly the performance to be expected; the resolution is the same as for a field- or line-sequential system with the same number of lines and color frames per second.

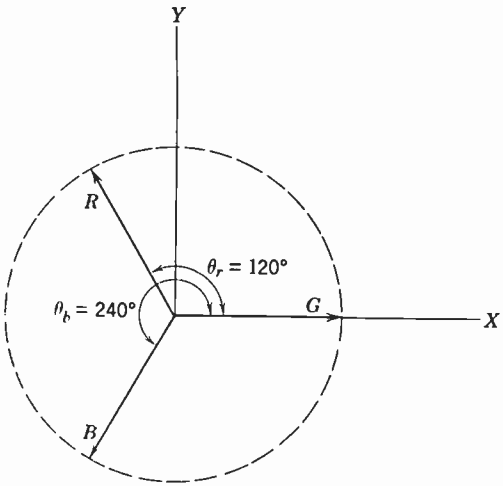


Fig. 18.21. Vector Diagram for Symmetrical Sampling.

The dot-sequential system differs from, e.g., the field-sequential system, however, in permitting an improved distribution of bandwidth in accord with the mixed-highs principle (Figs. 18.18 and 18.19). To this end, the sampling frequency is increased to, e.g., 3.58 megacycles, and low-pass filters with a cutoff frequency of 0.52 megacycle are inserted between the camera outputs and the sampler and between the desampler and the reproducer. This system thus transmits the low-frequency signals, G_L, R_L, B_L, up to 0.52 megacycle without any color cross-talk or other error. In addition, the three camera signals G, R, and B are added in proportion to the luminance of the three primaries required to produce standard white and are passed through a bandpass filter from 0.52 to 3.06 megacycles, yielding the mixed-highs signal

$$M_H = (L_G G_H + L_R R_H + L_B B_H)/(L_G + L_R + L_B) \qquad (18.14)$$

Typical values of the coefficients for a set of phosphor primaries are $L_G : L_R : L_B = 0.59 : 0.30 : 0.11$, whereas in Bedford's visual acuity tests, these ratios had the values 0.66:0.32:0.02.

The mixed-highs signal M_H is added to the sampler output S_L to form the complete video output. Since the two signals do not overlap in their frequency ranges (S_L:0 − 0.52; 3.06 − 4.1 megacycles; M_H:0.52 − 3.06 megacycles), they can be completely separated in the receiver. The sampler with its low-pass filters generates simply the low-frequency signals G_L, R_L, B_L, whereas a bandpass filter selects out the mixed-highs signal, which is simply added to all three reproducer channels. The relative gain of the mixed-highs and the sampler channels are adjusted so that, but for a common multiplying constant, the signals applied to the reproducer are

$$G_1 = G_L + M_H, \qquad R_1 = R_L + M_H, \qquad B_1 = B_L + M_H \quad (18.15)$$

The picture will give exact color reproduction up to frequencies of 0.52 megacycle, corresponding to horizontal detail about eight times as coarse as that which can be resolved in monochrome television. The brightness reproduction, on the other hand, is true up to 3.06 megacycles, giving three-fourths the horizontal resolution of the standard monochrome system. This can be shown in the following manner. The luminance of the picture (in relative units) is obtained by multiplying G_1, R_1, and B_1 by L_G, L_R, and L_B, respectively, and forming the sum of the products. Substitution of M_H from Eq. 18.14 then leads to the simple expression

$$\begin{aligned} L &= L_G G_1 + L_R R_1 + L_B B_1 \\ &= L_G(G_L + G_H) + L_R(R_L + R_H) + L_B(B_L + B_H) \end{aligned} \quad (18.16)$$

This is simply the luminance of the original scene, with detail corresponding to frequencies above 3.06 megacycles omitted.

Consider now the reproduction obtained when the transmitted signal is applied to a standard monochrome receiver. It will be assumed, to begin with, that the video passband of the receiver has been modified to provide cutoff at 3.06 megacycles. The luminance of the picture on the kinescope is now simply given by the transmitted video signal, $S_L + M_H$. This is not proportional to the expression in Eq. 18.16 and hence does not yield correct brightness gradation; in particular, blue parts of the picture will appear too light, green parts too dark. A simple modification, however, will yield proper gradation in the black-and-white picture without altering the reproduction of the color picture. It is only necessary to make the camera channel gains such that the output signals are proportional to the *luminance* of the three partial pictures instead of to the ratio of their luminance to that which,

in combination, produces white. This implies that the components of the sampler output which depend on G_L, R_L, B_L are multiplied by $(L_G, L_R, L_B)/(L_G + L_R + L_B)$. In compensation, the amplification of the desampler outputs is made proportional to the reciprocal factors $(L_G + L_R + L_B)/L_G$, $(L_G + L_R + L_B)/L_R$, $(L_G + L_R + L_B)/L_B$ (Fig. 18.22). The transmitted video signal is now, apart from the terms with frequencies above 3.06 megacycles (corresponding to $\omega_s - \omega_{g\,\max} = 2\omega_s - \omega_c$), a measure of the luminance in the picture.

The utilization of the video passband with this system is seen to be the following. In the color receiver, the brightness is reproduced for the full 3.06-megacycle range, whereas the chromaticity information, corresponding to two quantities, hue and chroma, is added over 0.52 megacycle, adding up to the full passband of 4.1 megacycles. In the monochrome receiver, on the other hand, only 3.06 megacycles, or three-fourths of the available information is utilized.

The system as outlined, however, presupposes a reduction in the receiver bandwidth. Suppose, next, that the passband of both the band filters in the color transmitter and receiver and the passband of the monochrome receiver are extended to 4.1 megacycles. Now the mixed-highs signal and the signal from the sampler overlap and cannot be separated by filters. On the one hand, the sampler signal frequency components between 3.06 and 4.1 megacycles are added to the mixed-highs signal in this range and tend to falsify the brightness variation in this range; on the other, the mixed-highs signal in the same frequency range is converted by the desampler into spurious signals in the 0 to 0.52-megacycle range which tend to falsify color values.

The effect of the spurious signals is minimized in practice by making the sampling frequency a half-integer multiple of the line frequency, e.g., $\omega_s/(2\pi) = 3{,}583{,}175 = 227.5 \cdot 15{,}750$ second^{-1}.* This has the effect of reversing the sign of the spurious signal on alternate frames, so that, averaged over two successive frames or $\frac{1}{15}$ second, the spurious signals cancel for a stationary picture. The visual effect of these spurious signals will cancel also provided that (*a*) the reproducer has a linear characteristic and (*b*) the eye integrates stimuli over periods of $\frac{1}{15}$ second or longer. Neither condition is fully satisfied. Hence, it is desirable to take special measures to minimize the visibility of the spurious signals.

* As will be seen later, the subcarrier frequency 3,579,545 second^{-1} is adopted in practice to reduce sound-carrier interference, and the line frequency is reduced slightly to maintain the half-integer relationship.

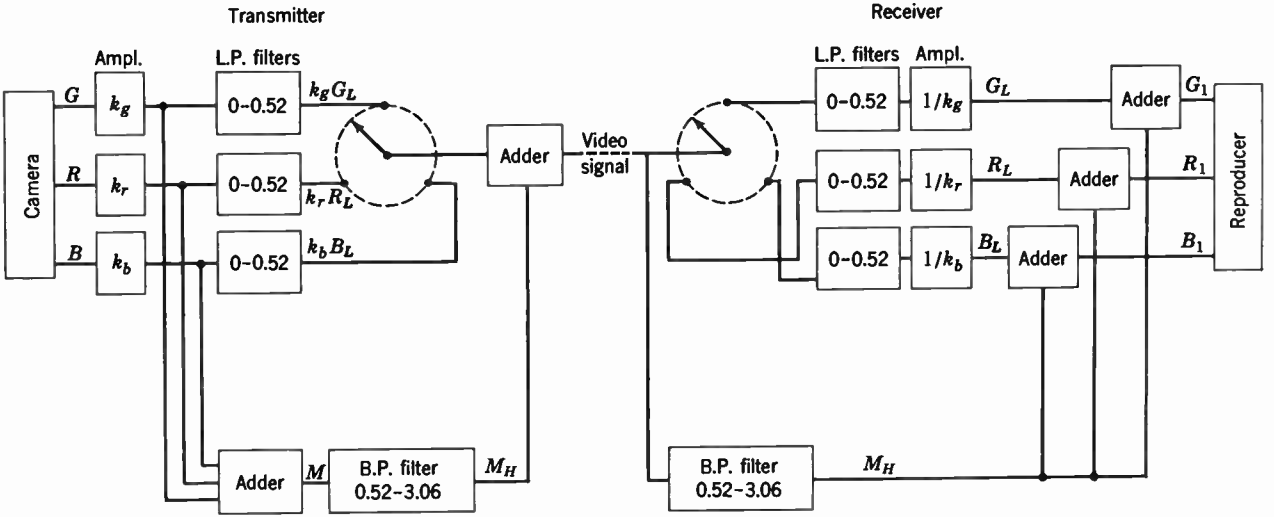


Fig. 18.22. Color Television System, Giving Faithful Brightness Reproduction on Monochrome Receiver (Linearity Assumed).

In greater detail, the reversal of the spurious signal terms can be seen by reference to Eqs. 18.10 and 18.11. Let T be a frame time and H a horizontal period, so that $T = 525H = (2n + 1)H$; at the same time, $\omega_s/(2\pi) = (m + 1/2)/H$, n and m being integers. Then, the high-frequency components of the sampler output which are added to the mixed-highs signal may be written (eventually with the addition of constant factors such as given in Eq. 18.14):

$$S_H(t) = 2\{G(t) \cos \omega_s t + R(t) \cos (\omega_s t + \theta_r) + B(t) \cos \omega_s t + \theta_b)\} \quad (18.17)$$

A frame time later, at $t + T$, the camera signals take on the same values for a stationary picture:

$$G(t + T) = G(t), \qquad R(t + T) = R(t), \qquad B(t + T) = B(t)$$

On the other hand,

$$\begin{aligned} \omega_s(t + T) &= \omega_s t + 2\pi(2n + 1)(m + 1/2) \\ &= \omega_s t + 2\pi(2nm + m + n) + \pi \end{aligned}$$

Thus, the argument of all three cosine terms changes by π, apart from an immaterial integer multiple of 2π. Since this is equivalent to a change in sign of the original terms,

$$S_H(t + T) = -S_H(t)$$

Consider now the effect of mixed-highs signals in the range between 3.06 and 4.1 megacycles on the signals delivered by the desampler and low-pass filter. For a stationary picture, any mixed-highs component may be written $a \cos (\omega_H t + \phi)$, where ω_H is an integer multiple of $2\pi/T$. Substitution of this component for S in Eq. 18.11 results in low-frequency spurious terms,

$$Ka \cos [(\omega_s - \omega_H)t - \phi]$$

Substitution of $t + T$ for t results again in the addition of an odd integer multiple of π to the argument of the cosine, which is equivalent to a change in sign of the original term. Thus here, also, the spurious signals average to zero over two successive frame times, or $\frac{1}{15}$ second.

It is also of interest to consider the effect of the spurious signal on the brightness distribution in a single frame. Assume that, as suggested on page 829, the transmitted signal is a luminance signal and that the gain in the three channels beyond the desampler is made inversely proportional to the luminance of unit intensity of the three

primaries. The contribution to luminance of the three component signals becomes then, for a linear reproducer, directly proportional to the three sampler signal outputs. Specifically for any mixed-highs component in the 3.06 to 4.1 megacycles range, this leads to a brightness contribution

$$\begin{aligned} L_H &= Ka\{\cos[(\omega_s - \omega_H)t - \phi] + \cos[(\omega_s - \omega_H)t - \phi + \theta_r] \\ &\quad + \cos[(\omega_s - \omega_H)t - \phi + \theta_b] \\ &= 0 \text{ for } \theta_r = 2\pi/3,\ \theta_b = 4\pi/3 \end{aligned}$$

Generally, with this receiver, no interfering high-frequency term entering the sampler affects the brightness distribution of the image. Since, as noted before, the sensitivity to flicker as well as the acuity of the eye is much less for chromaticity differences than for luminance differences, the visibility of the spurious low-frequency terms is minimized.

The effect on the picture of the spurious high-frequency terms added to the mixed-highs signal is the same in a color receiver and in a monochrome receiver. It is most readily visualized by reference to the sampling order for a sampling frequency equal to a half-integer multiple of the horizontal frequency (Fig. 18.23). The letters G, R, and B represent the positions of green, red, and blue dots in the three-color image obtained without low-pass filters between sampler and reproducer. Making the sampling frequency a half-integer multiple of the line frequency causes a 180-degree shift in the sampling phase for every alternate line (i.e., lines succeeding each other in time) and an equal shift in phase for alternate frames. The first circumstance causes the dots of a single color to be aligned neither vertically nor horizontally, whereas the second circumstance causes the dots of the second frame (third and fourth field) to be inserted midway between, or *interlaced* with, the dots of the first frame (first and second field). For the last reason, making the sampling frequency equal to a half-integer multiple of the line frequency is commonly designated as "dot interlacing."

The interfering signals are phased in exactly the same manner as the sampling frequency. Hence, insofar as they become visible at all, they produce a dot pattern similar to that of, e.g., all the dots marked G in Fig. 18.23, representing the sampling pattern; for an extended area of uniform color even the dot spacings will be identical. The dot spacing becomes smaller as the sampling frequency is increased; thus a high sampling frequency is desirable either to make the dot pattern less visible or to make possible its removal altogether, with a minimum

loss in resolution, by the insertion of a low-pass filter in the mixed-highs channel.

It is seen that the system described is fully compatible. The transmitted signal, applied to a monochrome receiver with a linear characteristic, gives the same horizontal resolution as a standard black-and-

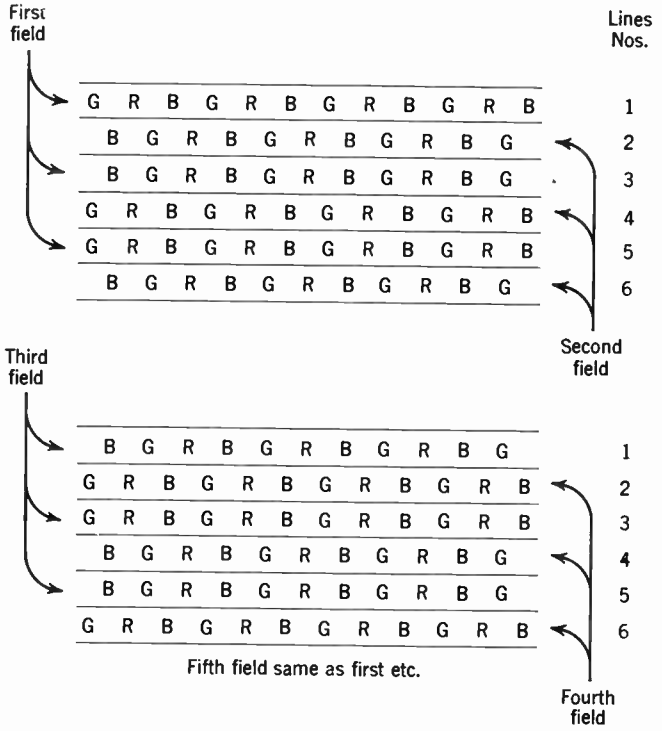


Fig. 18.23. Relative Positions of Dot Centers in Four Successive Fields of a Dot-Interlaced Color Television System.

white signal. The only difference is that, for picture components above $2\omega_s - \omega_c = 3.06$ megacycles, the frame period is increased from $\frac{1}{30}$ to $\frac{1}{15}$ second. Since these components correspond to very fine detail, the resulting flicker ("dot crawl") is relatively unimportant. It is seen that 3.06 megacycles + $1/2 \cdot 1.04$ megacycles = 3.58 megacycles of the signal are utilized for reproducing the monochrome picture; the factor 1/2 arises from the fact that the upper luminance components are reproduced at only half the normal frame rate. Thus, seven-eighths of the passband is utilized for reconstructing the monochrome picture with the same horizontal resolution as for a black-and-white transmis-

sion with a full 4.1-megacycle passband; more generally the ratio 7/8 is replaced by the ratio of the sampling frequency to the cutoff frequency, ω_s/ω_c.

In the color picture, the chrominance information (to $\omega_c - \omega_s = 0.52$ megacycle) is reproduced correctly at a frame rate of 15 second^{-1}, for which the spurious signals cancel out. The luminance is reproduced exactly up to $2\omega_s - \omega_c = 3.06$ megacycles at a frame rate of 30 second^{-1} and at a frame rate of 15 second^{-1} for the remainder of the passband. The passband utilization is here

$$\frac{1}{2} \cdot 2(\omega_c - \omega_s) + 2\omega_s - \omega_c + \frac{1}{2}(\omega_c - (2\omega_s - \omega_c)) = \omega_c$$

In other words, the passband is fully utilized, as was to be expected.

It has been brought out repeatedly already that the dot-sequential system is, in fact, not sequential at all. The signal supplied by the sampler in conjunction with the video filter following it is (from Eq. 18.10)

$$\begin{aligned} S &= k_g G_L(1+2\cos\omega_s t)+k_r R_L[1+2\cos(\omega_s t+2\pi/3)] \\ &\quad +k_b B_L[1+2\cos(\omega_s t+4\pi/3)] \\ &= M_L+(3k_g G_L-M_L)\cos\omega_s t-\sqrt{3}(k_r R_L-k_b B_L)\sin\omega_s t \\ &= M_L+2\sqrt{(k_g G_L+k_r R_L+k_b B_L)^2-3(k_g k_r G_L R_L+k_r k_b R_L B_L+k_b k_g B_L G_L)} \\ &\quad \cos(\omega_s t+\phi) \end{aligned} \tag{18.18}$$

Here,

$$\begin{aligned} k_g, k_r, k_b &= (L_G, L_R, L_B)/(L_G + L_R + L_B) \\ M_L &= k_g G_L + k_r R_L + k_b B_L \end{aligned} \tag{18.19}$$

and

$$\phi = \arctan \frac{\sqrt{3}(k_r R_L - k_b B_L)}{3k_g G_L - M_L}$$

Through the addition of the mixed-highs signal M_H, the total video signal becomes simply the sum of the complete monochrome signal $M = M_L + M_H$ and a subcarrier of frequency $\omega_s/(2\pi)$ which is modulated both in phase and amplitude. This subcarrier term is the chrominance signal. It is seen that the chrominance signal can be obtained by modulating two sinusoidal subcarrier oscillations in quadrature with suitable linear combinations of the camera signals; the camera

signals themselves are suppressed by the use of balanced modulators. Similar simplifications are possible in the receiver. Whereas the character of the monochrome signal appears to be fixed by the requirement of compatibility, it would seem that the dependence of the amplitude and phase of the chrominance signal on the camera outputs can be varied within wide limits. The same applies to the methods of signal generation and detection. Both will, in general, be selected on the basis of circuit simplicity, immunity to interference, and the attainment of high signal-to-noise ratios in the picture.

In particular, the signal selection represented by Eq. 18.18, which appeared logical from the point of view of the bypassed-highs system discussed so far, ceases to be so when the whole monochrome signal is bypassed. Thus, the interpretation of the chrominance signal becomes much simpler if k_g, k_r, k_b are all replaced by 1/3. Then the amplitude of the chromaticity signal vanishes for black-and-white portions of the picture ($G = R = B$) and the ratio of the amplitude to the signal $G + R + B$ may be taken as a measure of the chroma, the phase as a measure of the hue. This condition corresponds precisely to the condition described in the last chapter in relation to the dot-sequential system. This condition of operation has the incidental advantage that the dot pattern is completely absent for black-and-white portions of the picture. With the signal in Eq. 18.18, the same occurs for areas of a relatively highly saturated blue tone. It will be seen later that the vanishing of the chrominance signal for black-and-white portions of the picture can be achieved also with a monochrome signal strictly proportional to luminance.

A final remark is appropriate with respect to the name which might be given to the system just described. In the last chapter, it was called a "derived simultaneous system." Perhaps the most descriptive name is "phase-amplitude modulated," or "PAM, simultaneous system." Other terms which have been favored are "dot-simultaneous" and "dot-multiplex," the prefix "dot" distinguishing it from the three-channel simultaneous system and from field- or line-sequential systems, respectively.

18.5 Receiver Nonlinearity and Selection of the Transmitted Signal. The preceding discussion has been based, in the main, on an entirely linear transmission system, with linear pickup and reproducing elements. Whereas certain camera tubes, such as the image orthicon operated below the knee of its characteristic, supply a signal which is very nearly a linear function of the photocathode illumination, the con-

verse is true of kinescopes. The average "gamma," * or measure of contrast enhancement, of a kinescope is usually taken to lie between 2 and 3, e.g., at 2.2, instead of unity. Since the overall response of the system must be linear to provide true color reproduction—in particular unchanged chromaticities when the brightness of the picture alone is altered—nonlinear "rooter" circuits must be inserted somewhere in the transmission channel to compensate the kinescope nonlinearity and, eventually, camera tube nonlinearity.

The placement of the rooter circuits to produce accurate rendition of both brightness and chromaticity is quite unique. It is shown in Fig. 18.24 for a system employing a bypassed brightness signal. The camera compensation circuits must be placed between the camera tubes and any subsequent signal-adding or mixing circuits, and the kinescope compensation circuits must be placed directly ahead of the signal input electrodes of the reproducer, in particular beyond the point where the brightness signal and the signals giving the difference between the three color signals and the brightness signal are added.

The system in Fig. 18.24 yields optimum chromaticity and brightness rendition as well as the best possible cancellation of spurious signals. It is essentially the system assumed in the discussion of the preceding section. At the same time, it has two very serious drawbacks. The first is a very material increase in receiver cost, resulting from the addition of the compensating circuits and signal adders; the second is a certain violation of the requirement of compatibility. Since standard monochrome receivers are not provided with rooter circuits, the black-and-white picture obtained from such linear color transmissions would show excessive contrasts. Standard monochrome camera chains generally utilize a nonlinear portion of the pickup tube characteristic for television transmissions and thus achieve a certain compensation of kinescope gamma or employ gamma-correcting circuits to attain the same end.

A final disadvantage of the linear system in Fig. 18.24 as compared with the system employed in practice, in which the compensating cir-

* "Gamma" may be defined generally by the formula

$$\gamma = \frac{d(\log S_o)}{d(\log S_i)} = \frac{S_i}{S_o}\frac{dS_o}{dS_i}$$

where S_i and S_o are the input and output, respectively, of the device under consideration. For the camera tube, S_i is the photocathode illumination and S_o the video signal at the camera tube output, whereas, for the reproducer, S_i is the video signal on the kinescope grid and S_o is the luminance of the reproduced picture.

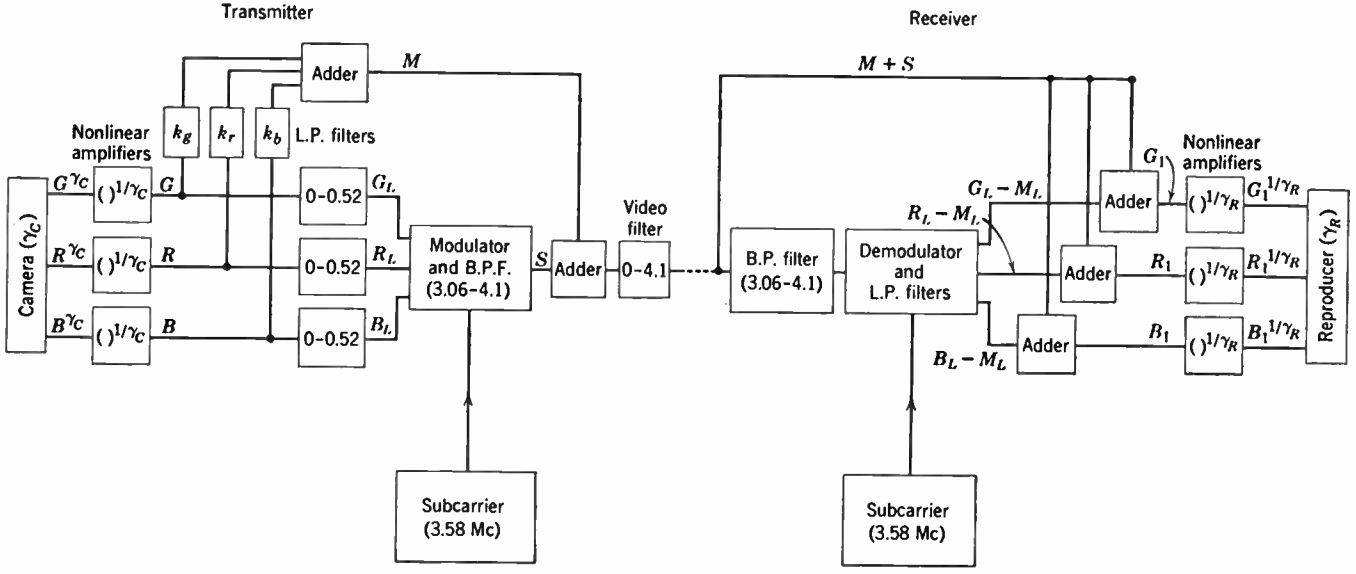


Fig. 18.24. Color Television System with Complete Nonlinearity Compensation.

cuits are all incorporated in the transmitter (Fig. 18.25), is a reduced signal-to-noise ratio in picture areas where the video signal is small. For a given total range of video signal, rooter circuits compensating receiver nonlinearity at the transmitter expand the low-level signals, so that signal steps become large in this range in comparison with noise fluctuations at the second detector of the receiver. The kinescope characteristic, conversely, compresses the noise at low levels as it restores the proper brightness gradation in the picture.

The placement of the compensating or rooter circuits at the transmitter places certain limits on the type of signal which can be radiated as well as on the picture fidelity which can be achieved. The most reasonable requirement consistent with this arrangement is accurate color and brightness rendition of the low-frequency terms of the picture, permitting deviation from precise rendition in the case of the mixed-highs components. This can be obtained very nearly with the system shown in Fig. 18.25. As long as no mixed-highs signals are present, the rendition is seen to be precisely given by $[G^{1/\gamma}]_L{}^{\gamma}$, $[R^{1/\gamma}]_L{}^{\gamma}$, $[B^{1/\gamma}]_L{}^{\gamma}$, which differ slightly from the desired values G_L, R_L, B_L, since the low-pass filters preceding the sampler remove some of the higher harmonics of the low-frequency components generated in the rooter circuits.

The coefficients of the bypassed monochrome signal

$$M = k_2 G^{1/\gamma} + k_1 R^{1/\gamma} + k_3 B^{1/\gamma} \tag{18.20}$$

may be fixed by the requirement that a small interfering high-frequency signal $A \sin \omega_p t$ applied to the receiver demodulators have no effect on the luminance of a picture with low saturation. It can readily be shown that this condition is met if

$$k_2,\ k_1,\ k_3 = k_g,\ k_r,\ k_b = (L_G,\ L_R,\ L_B)/(L_G + L_R + L_B) \tag{18.21}$$

Specifically, the condition is satisfied by the signal given by Eq. 18.18. With modulators and demodulators in the form of sharp symmetrical samplers the low-frequency portion of the signal output becomes here

$$G_1 = G_L{}^{1/\gamma} + \frac{A}{k_g} \cos (\omega_s - \omega_p)t \tag{18.22a}$$

$$R_1 = R_L{}^{1/\gamma} - \frac{A}{2k_r} \cos (\omega_s - \omega_p)t \tag{18.22b}$$

$$B_1 = B_L{}^{1/\gamma} - \frac{A}{2k_b} \cos (\omega_s - \omega_p)t \tag{18.22c}$$

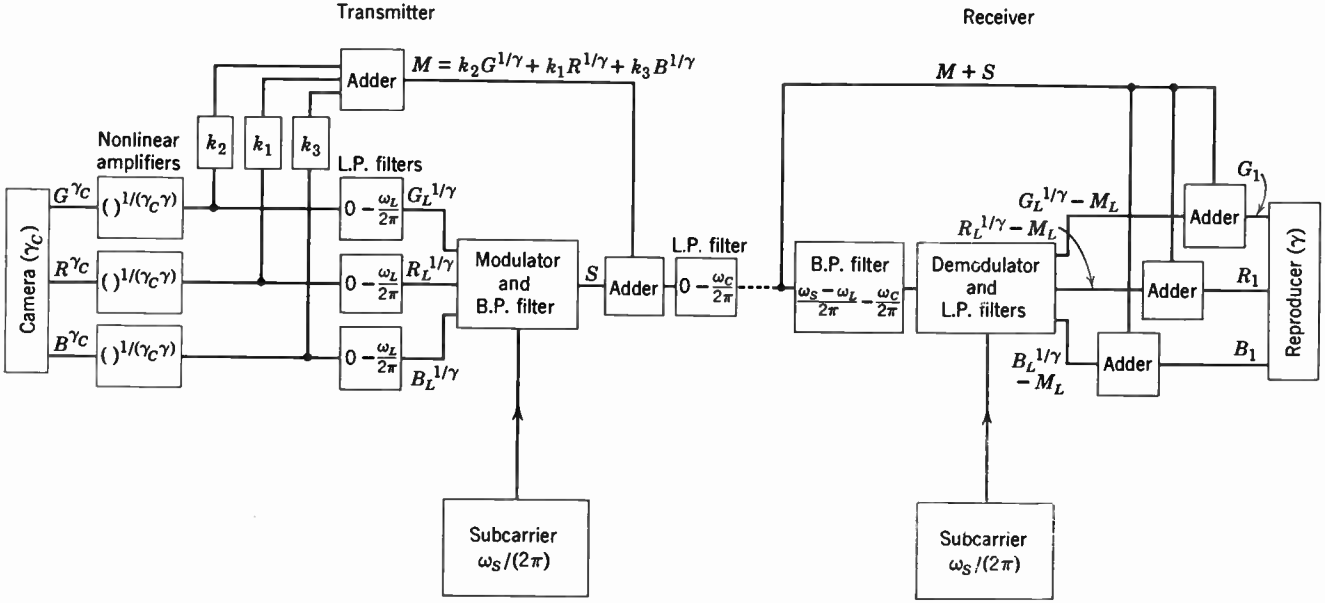


Fig. 18.25. Color Television System with Receiver Nonlinearity Compensation at the Transmitter.

Here $G_L^{1/\gamma}$ has been set for $G^{1/\gamma}{}_L$. This is justified if the harmonics introduced by the rooter and subsequently removed by the low-pass filter contribute little to the signal.

The luminance is given by

$$L = L_G G_1^{\gamma} + L_R R_1^{\gamma} + L_B B_1^{\gamma} \tag{18.23}$$

If terms of higher order than the first in A are omitted,

$$L = L_G G_L + L_R R_L + L_B B_L + \gamma(L_G + L_R + L_B)$$
$$A \cos(\omega_s - \omega_p)t\,[G_L^{(\gamma-1)/\gamma} - \tfrac{1}{2}R_L^{(\gamma-1)/\gamma} - \tfrac{1}{2}B_L^{(\gamma-1)/\gamma}] \tag{18.24}$$

Since, by assumption, the saturation of the picture is low,

$$G_L \cong R_L \cong B_L \tag{18.25}$$

the interference term does not contribute to the luminance. For a linear system ($\gamma = 1$) the contribution to the luminance would have been zero for a field of arbitrary saturation. It will be shown later (Eq. 18.34) that the effect of signals from the chrominance channel on the luminance is generally minimized for a luminance signal employed with a chrominance signal vanishing for a black-and-white picture.

It can easily be shown that an application of the bypassed monochrome signal M to a black-and-white receiver does not lead to a correct tonal rendition of different colors, although correct contrasts are obtained for a monochrome picture. The brightness in this instance is given by

$$L = (k_g G^{1/\gamma} + k_r R^{1/\gamma} + k_b B^{1/\gamma})^{\gamma} \tag{18.26}$$

with, e.g., $k_g = 0.59$, $k_r = 0.30$, $k_b = 0.11$, and $\gamma = 2.2$.

Consider green, red, blue, and white fields of equal luminance:

$$k_g G + k_r R + k_b B = 1$$

The relative brightness of the reproduced fields becomes:

green	0.53
red	0.24
blue	0.071
white	1.00

More generally, colors of low saturation are reproduced with greater luminance than colors close to the primaries, and the primaries of high luminosity give rise to greater brightness than primaries of low lumi-

nosity. In general, this imperfect tonal rendition in monochrome as well as the imperfect rendition in intensity of the mixed-highs in the color picture are not noticeable as drawbacks.

Of considerably greater importance is the effect of the nonlinearity on the visibility of spurious signals, since it prevents their cancellation. The large gamma of the kinescope causes positive swings of these signals to add more to the brightness of the picture than the negative swings subtract from it. This applies both to the interference of the chrominance signal with the monochrome signal, resulting in a visible dot pattern, and to the interference of the high-frequency monochrome components, resulting in low-frequency color fluctuations. The first may be made small relative to the second—by reducing the amplitude of the chrominance signal in comparison to the monochrome signal in the radiated video signal—or vice versa. They cannot both be reduced beyond a certain limit.

It has been seen that the nonlinearity of the receiver has a restricting effect on the selection of the transmitted signal, which will be considered next. Specifically, the camera channels feeding the transmitter sampler must deliver signals corresponding to the color coordinates of the picture based on the receiver primaries. There is, however, a considerable latitude both in the character of the transmitted signal and the method of its generation and detection *unless* a number of restricting conditions are imposed.

The following considerations will be limited to systems with bypassed luminance signal—more precisely, bypassed monochrome signal—since the amplitude of this signal need not be proportional to the luminance of the picture. This signal will be denoted by

$$M = k_2G + k_1R + k_3B, \qquad k_1 + k_2 + k_3 = 1 \tag{18.27}$$

If $k_2 = k_g$, $k_1 = k_r$, $k_3 = k_b$, where k_g, k_r, k_b represent the relative luminances of unit intensity of the three receiver primaries (assumed to be equal to 0.59, 0.30, and 0.11, respectively, in the following), M is a measure of the luminance of the picture.

It has been seen that the chrominance signal can be expressed quite generally by

$$S = A \cos (\omega_s t + \phi) \tag{18.28}$$

where both the amplitude A and the phase ϕ are functions of G, R, and B; for simplicity, the subscript L will be omitted in the following treat-

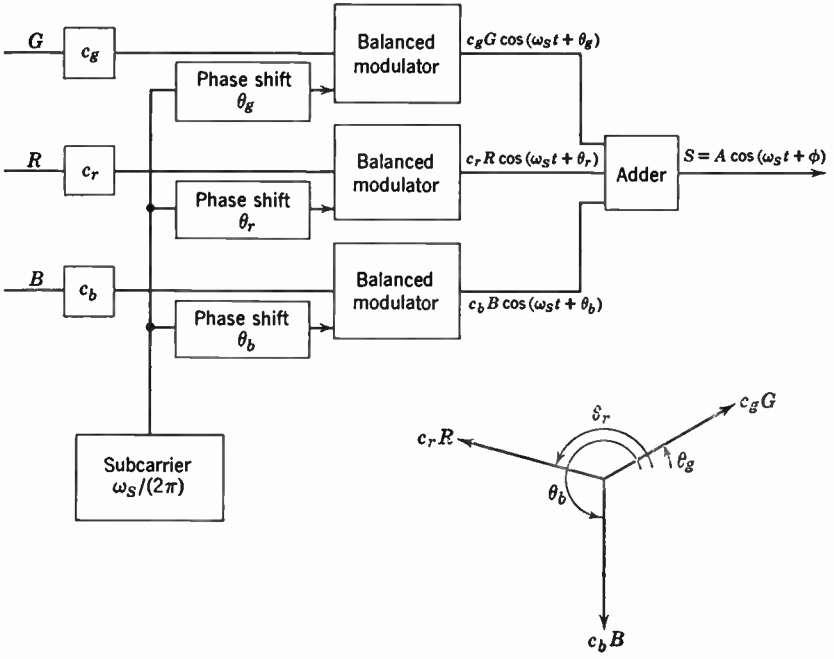


Fig. 18.26. Generation of Chrominance Signal by Three Modulators with Odd-Phase Angles.

ment. A chrominance signal in which amplitude and phase are specific functions of G, R, and B may be obtained in unique manner by two different procedures, which are distinguished by their simplicity:

(a) By inserting gains c_g, c_r, and c_b in the camera channels feeding the modulators and modulating at odd angles θ_g, θ_r, θ_b (Fig. 18.26); the expression for the amplitude and phase becomes

$$A = \sqrt{\left\{\begin{array}{l} c_g{}^2G^2 + c_r{}^2R^2 + c_b{}^2B^2 + 2[c_gc_rGR\cos(\theta_g - \theta_r) \\ \quad + c_rc_bRB\cos(\theta_r - \theta_b) + c_bc_gBG\cos(\theta_b - \theta_g)] \end{array}\right\}} \tag{18.29}$$

$$\tan\phi = \frac{c_gG\sin\theta_g + c_rR\sin\theta_r + c_bB\sin\theta_b}{c_gG\cos\theta_g + c_rR\cos\theta_r + c_bB\cos\theta_b}$$

(b) By forming suitable linear combinations of the three camera signals G, R, B and letting these modulate two subcarrier oscillations

cos $\omega_s t$ and sin $\omega_s t$ in quadrature (Fig. 18.27); the two modulating signals which will produce the chromaticity signal given in (*a*) are equal to

$$\begin{aligned} A \cos \phi &= c_g G \cos \theta_g + c_r R \cos \theta_r + c_b B \cos \theta_b \\ A \sin \phi &= c_g G \sin \theta_g + c_r R \sin \theta_r + c_b B \sin \theta_b \end{aligned} \tag{18.30}$$

In both instances, it is assumed that the modulations effect simply a multiplication of the low-frequency video signals and the carrier

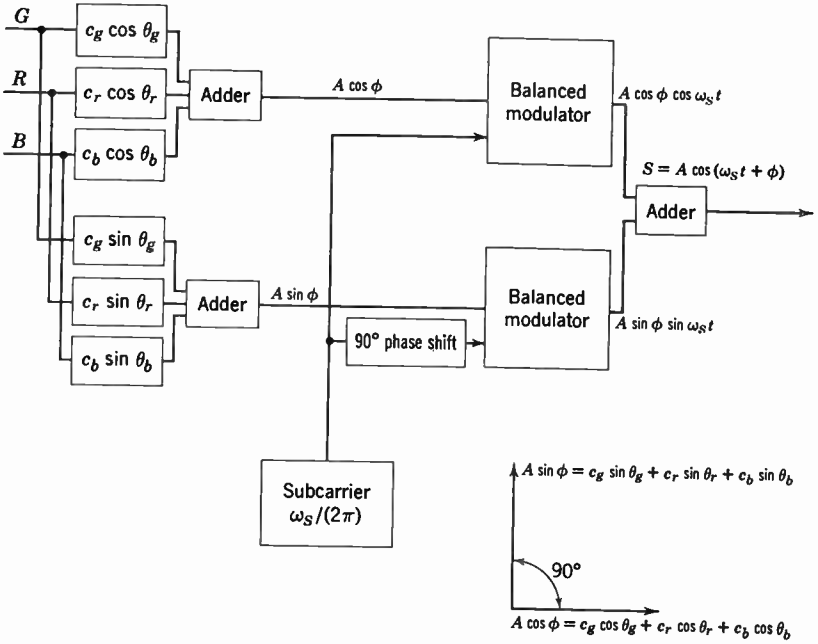


Fig. 18.27. Generation of Chrominance Signal by Two Modulators in Quadrature.

oscillation cos $(\omega_s t + \theta)$. This is attained by the employment of balanced modulators. Other methods of generating the chrominance signal are, in general, more complex.

At the receiver, the chrominance signal is demodulated at desampling angles θ'_g, θ'_r, θ'_b and the three derived signals are multiplied by the gains c'_g, c'_r, c'_b to yield signals *G–M*, *R–M*, and *B–M*. Upon addition of the monochrome signal, the original camera signals *G*, *R*, and *B* are recovered and yield a faithful reproduction of the original scene. Again, as an alternative, two demodulators in quadrature may be

employed and the color difference signal obtained by forming linear combinations of the demodulator outputs. The condition for faithful reproduction is obtained by writing down three equations of the type

$$G - M = c'_g A \cos(\omega_s t + \phi) \cos(\omega_s t + \theta'_g) = c'_g(A/2 \cos(\phi - \theta'_g) \cdots$$

retaining only the low-frequency terms, and equating the coefficients of G, R, and B in each. This leads to the system of equations:

$$\begin{aligned}
\tfrac{1}{2}c_g c'_g \cos(\theta_g - \theta'_g) + k_2 &= 1 \quad \tfrac{1}{2}c_r c'_g \cos(\theta_r - \theta'_g) + k_1 = 0 \\
\tfrac{1}{2}c_g c'_r \cos(\theta_g - \theta'_r) + k_2 &= 0 \quad \tfrac{1}{2}c_r c'_r \cos(\theta_r - \theta'_r) + k_1 = 1 \\
\tfrac{1}{2}c_g c'_b \cos(\theta_g - \theta'_b) + k_2 &= 0 \quad \tfrac{1}{2}c_r c'_b \cos(\theta_r - \theta'_b) + k_1 = 0 \\
\tfrac{1}{2}c_b c'_g \cos(\theta_b - \theta'_g) + k_3 &= 0 \\
\tfrac{1}{2}c_b c'_r \cos(\theta_b - \theta'_r) + k_3 &= 0 \\
\tfrac{1}{2}c_b c'_b \cos(\theta_b - \theta'_b) + k_3 &= 1
\end{aligned} \tag{18.31}$$

The further condition that the chromaticity signal should vanish for a black-and-white field ($G = R = B$) may be written, by Eq. 18.30,

$$\begin{aligned}
c_g \cos\theta_g + c_r \cos\theta_r + c_b \cos\theta_b &= 0 \\
c_g \sin\theta_g + c_r \sin\theta_r + c_b \sin\theta_b &= 0
\end{aligned} \tag{18.32}$$

Addition of the horizontal rows in Eqs. 18.31 shows, however, that this condition is satisfied automatically, since $k_1 + k_2 + k_3 = 1$.

Equations 18.31 thus represent nine equations for twelve unknowns; in other words, there are many different sets of parameters which will satisfy them. If a number are fixed from the point of view of convenience or simplicity, the remaining ones may be evaluated from Eqs. 18.31.

One of the systems proposed by RCA at an early date is obtained by setting all the gain coefficients equal to each other and making the sampling angles also the same in transmitter and receiver; with $k_1 = k_2 = k_3 = 1/3$. The conditions of Eqs. 18.31 are then satisfied, provided that all the coefficients are made equal to $2/\sqrt{3}$ and the sampling angles are $\theta_g = \theta'_g = 0$, $\theta_r = \theta'_r = 2\pi/3$, $\theta_b = \theta'_b = 4\pi/3$.

This system (Fig. 18.28) supplies a monochrome signal which differs considerably from the luminance and hence cannot give correct luminance rendition on a monochrome receiver. However, receiver non-

linearity prevents correct rendition in any case and, in view of it, the reproduction of color in black-and-white tones is not noticeably worse than with a luminance signal. The relative apparent brightness of

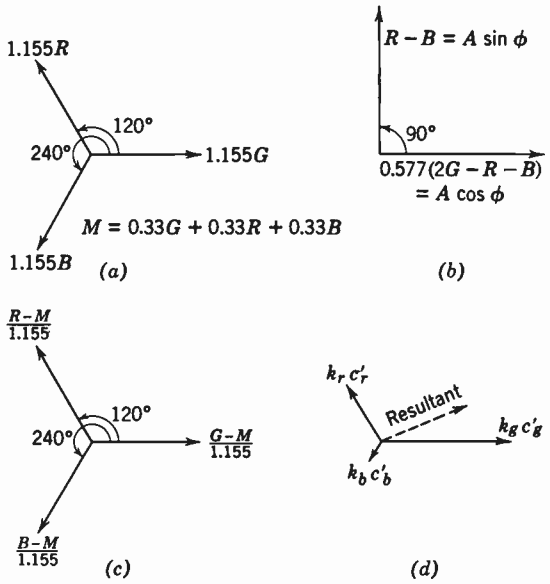


Fig. 18.28. "Constant-Amplitude" System with Symmetrical Sampling. (*a*) Generation of signal by three modulators. (*b*) Generation by quadrature modulators. (*c*) Recovery of color. (*d*) Luminance vector for interfering signal.

green, red, blue, and white fields of equal luminance becomes, for $\gamma = 2.2$,

green	0.15
red	0.30
blue	0.81
white	1.00

These values may be compared with the values for a luminance signal given on page 841.

It will be noted that the sum of the signals applied by the chrominance channels to the kinescope guns vanishes in this system. In particular, for a signal component $A \cos \omega_f t$ the three output signals are given generally by

$$A\{c'_g \cos [(\omega_s - \omega_f)t + \theta'_g],\ c'_r \cos [(\omega_s - \omega_f)t + \theta'_r],$$
$$c'_b \cos [(\omega_s - \omega_f)t + \theta'_b]\}$$

If each component is indicated by a vector of amplitude Ac'_g, Ac'_r, and Ac'_b and with azimuth θ_g, θ_r, and θ_b, respectively, the instantaneous amplitude is represented by the projection of the vectors on a line through the origin, rotating with the angular speed $\omega_s - \omega_f$. In particular, if all the coefficients are equal and the angles differ by 120 degrees $(2\pi/3)$, the resultant of the vectors vanishes, so that the chromaticity signal does not contribute to the sum of the signal amplitudes applied to the reproducer. For this reason such a system has been called a "constant amplitude system."

The sum of the signal amplitudes has, however, no practical significance. On the other hand, if each amplitude is multiplied by the relative luminosity of the corresponding primary (k_g, k_r, k_b), the contribution of that particular channel to the picture brightness is obtained, provided that the reproducer is linear. Figure 18.28*d* shows that the contribution of the chrominance signal to the brightness or luminance of the picture is by no means zero for a constant amplitude system. On the other hand, the condition for the vanishing of the luminance vector can be readily written down. It is

$$\begin{aligned} k_g c'_g \cos \theta'_g + k_r c'_r \cos \theta'_r + k_b c'_b \cos \theta'_b &= 0 \\ k_g c'_g \sin \theta'_g + k_r c'_r \sin \theta'_r + k_b c'_b \sin \theta'_b &= 0 \end{aligned} \tag{18.33}$$

A system which fulfills this condition is called a "constant luminance system." It has the advantage that (for a linear receiver) both noise in the chrominance channel and interference from the high-frequency components of the monochrome signal appear only as chromaticity variations in the picture, without accompanying luminance variations, and are hence much less noticeable to the observer. This permits, in turn, the transmission of the chrominance signal at a smaller amplitude relative to the monochrome signal, resulting in a reduction in the prominence of the dot pattern representing the interference of the chrominance signal with the monochrome signal.

It is true that, in view of the nonlinearity of the receiver, the chrominance signal does have an effect on picture brightness. This effect is, however, greatly reduced, particularly if the signals derived from the chrominance channel are small compared to the monochrome signal. If the complete signal entering the chrominance channel is written in the form $A \cos (\omega_s t + \phi)$, the picture brightness is now

$$L = k_g\left[M + \frac{Ac_{g'}}{2}\cos(\phi - \theta'_g)\right]^\gamma + k_r\left[M + \frac{Ac_{r'}}{2}\cos(\phi - \theta'_r)\right]^\gamma + k_b\left[M + \frac{Ac_{b'}}{2}\cos(\phi - \theta'_b)\right]^\gamma$$

$$= M^\gamma\left\{1 + \frac{A^2\gamma(\gamma-1)}{8M^2}[k_g c'_g{}^2\cos^2(\phi - \theta'_g) + k_r c'_r{}^2\cos^2(\phi - \theta'_r) + k_b c'_b{}^2\cos^2(\phi - \theta'_b)]\cdots\right\} \quad (18.34)$$

It is seen that the chrominance signal contributes to the brightness only through terms of the second and higher order in the ratio of the chrominance amplitude to the monochrome signal amplitude.

If the distribution of amplification of the chromaticity signal between transmitter and receiver is fixed in some manner, Eqs. 18.31 and 18.33 supply just eleven equations to determine eleven unknowns. The symmetry of the problem suggests that a solution may be obtained by setting the coefficients in Eqs. 18.32 and 18.33 equal to each other ($c_g = k_g c'_g$, etc.) and making the sampling angles in transmitter and receiver identical ($\theta_g = \theta'_g$). This is in fact the case. For $k_g = 0.59$, $k_r = 0.30$, $k_b = 0.11$ the sampling angles and channel gains become

$\theta_g = \theta'_g = 0$	$\theta_r = \theta'_r = 141°45'$	$\theta_b = \theta'_b = 245°03'$
$c_g = 0.696$	$c_r = 0.648$	$c_b = 0.443$
$c'_g = 1.18$	$c'_r = 2.16$	$c'_b = 4.03$

The amplitude and luminance vector diagrams for the constant luminance system as well as the mixing constants required for quadrature modulations are shown in Fig. 18.29.

Other systems have been proposed with the primary objective of simplifying receiver design. An example is the zero-luminance-blue system, which comes close to satisfying the constant luminance requirement and at the same time derives the three color difference signals in very simple fashion from two demodulators in quadrature at the receiver. Here, the small contribution to the luminance of the blue component is neglected, and the monochrome signal is set equal to

$$M = 0.667G + 0.333R$$

If the contribution to the brightness of the blue component is neglected and, as implied by the above equation, the relative luminance of the green and red primaries is as 2:1, a disturbing signal in the chroma-

ticity channel will have no effect on the luminance, provided that the desampling angles for the green and the red signals are 180 degrees

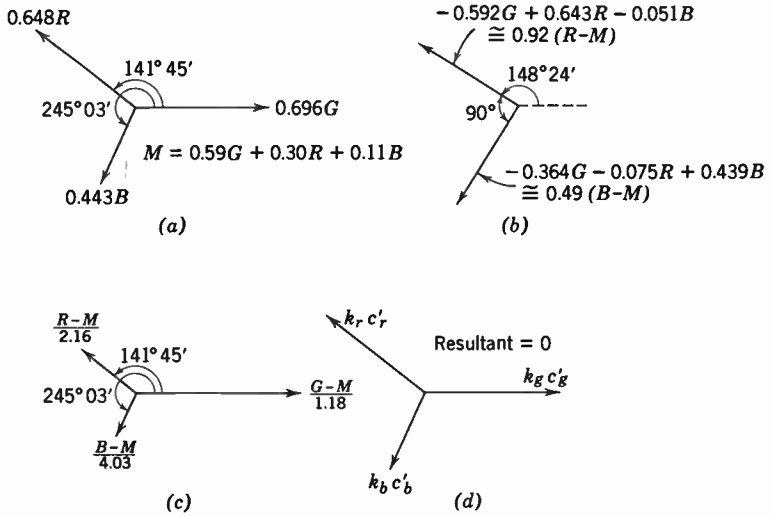


Fig. 18.29. "Constant-Luminance" System (*a*, *b*, *c*, *d* as in Fig. 18.28).

apart and the gain in these two channels is as 1:2. The desampling angle for the blue signal is then conveniently set at 270 degrees (Fig. 18.30), at right angles to the other two. The blue receiver channel

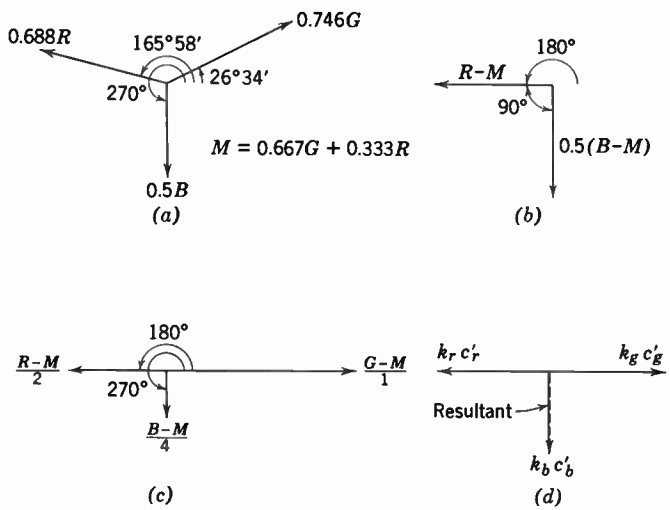


Fig. 18.30. "Zero-Luminance-Blue" System (*a*, *b*, *c*, *d* as in Fig. 18.28).

gain will, furthermore, be set equal to 4, for convenient comparison of the results with those for the exact constant luminance system. It will be noted that only one demodulator is needed for the green and red channels, since the red signal may be obtained from the green signal by simple phase inversion.

With the receiver parameters thus determined, the transmitter parameters may be obtained from Eqs. 18.31. The complete set, for the zero-luminance-blue system, is given below:

$$\theta_g = 26°34' \qquad \theta_r = 165°58' \qquad \theta_b = 270°$$
$$\theta'_g = 0 \qquad \theta'_r = 180° \qquad \theta'_b = 270°$$
$$c_g = 0.746 \qquad c_r = 0.688 \qquad c_b = 0.5$$
$$c'_g = 1.000 \qquad c'_r = 2.000 \qquad c'_b = 4.00$$

The corresponding vector diagrams are shown in Fig. 18.30.

In another system using two demodulators the green signal is transmitted as monochrome signal ($M = G$; $k_2 = 1$, $k_1 = k_3 = 0$). Specifically, if the desampling angles θ'_r and θ'_b are put equal to 180 and 270 degrees, Eqs. 18.31 yield as a set of parameters for this system:

$$\theta_g = 45° \qquad \theta_r = 180° \qquad \theta_b = 270°$$
$$\theta'_r = 180° \qquad \theta'_b = 270°$$
$$c_g = 1.41 \qquad c_r = 1.00 \qquad c_b = 1.00$$
$$c'_g = 0 \qquad c'_r = 2.00 \qquad c'_b = 2.00$$

The vector diagrams appropriate to this system are represented, once more, in Fig. 18.31. This is in no sense a constant-luminance system, and the gain in simplicity is obtained at the expense of poor compatibility in the form of improper rendition of differently colored objects in a black-and-white picture. As compared with the constant-amplitude system (as well as the exact and approximate constant-luminance systems), it has the drawback of demanding a much greater amplitude range of the radiated signal (monochrome plus chrominance signal) for a given receiver amplification and receiver sensitivity to interference in the chrominance channel. Thus, for a purely green picture, the monochrome signal must be three times as large for the "green" system as for the constant-amplitude system, and the chrominance signal must be almost twice as large. Since the amplitude range is fixed by the capacity of the transmitter, this implies that the signal-

to-noise ratio is smaller by a factor between 2 and 3 for the "green" system as compared with the constant-amplitude system.

Finally, it seems appropriate to indicate some reasons for preferring the form of the bypassed-highs system realized by the bypassed monochrome signal system. One obvious reason is that with the bypassed monochrome signal a bandpass filter in the mixed-highs channel care-

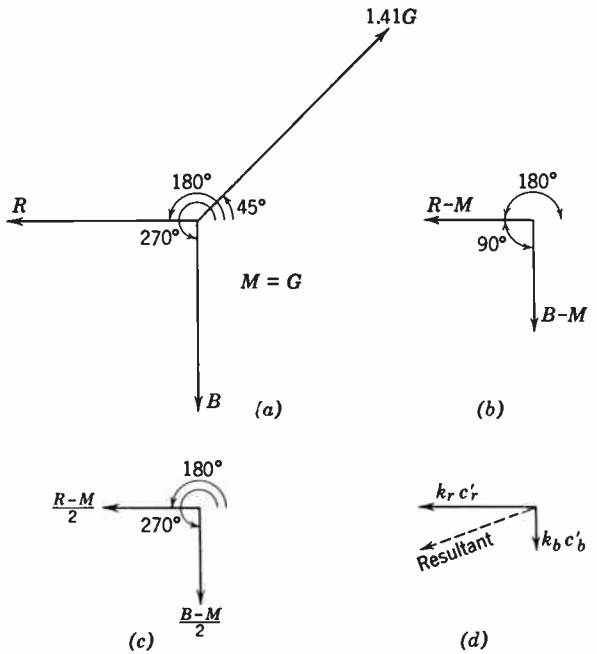


Fig. 18.31. Green Monochrome Signal System (*a*, *b*, *c*, *d* as in Fig. 18.28).

fully matched to the low-pass filters in the color channels becomes unnecessary. Unless the whole monochrome signal is bypassed, an imperfect match results in a noticeable irregularity in the frequency characteristic of the system. These matched filters are required in the receiver as well as in the transmitter. In addition, the generation of the chrominance signal with balanced modulators is easier than that of the composite chromaticity and low-frequency brightness signal with proper amplitude relation demanded when the low-frequency monochrome signals pass through the modulators; narrow samplers, which accomplish this effectively, are inherently noisy.

It is not possible to take advantage of the constant-luminance feature in the last instance to reduce the dot pattern without, at the same

time, reducing the signal-to-noise ratio of the low-frequency brightness components in the picture, since the low-frequency brightness components and the chrominance signals pass through the same channels. A strong dot pattern is objectionable even when unresolved by the eye, since, with it, nonlinearity of the kinescope response may easily lead to lack of color fidelity; for example, in a relatively bright orange field, the red phosphor may saturate more than the green phosphor, resulting in a shift from orange to yellow.

In summary, of the signals considered so far, the most favorable results are obtained with systems incorporating:

1. A bypassed luminance signal M.
2. A chrominance signal derived from the color coordinates of the picture in terms of the receiver primaries corrected for the nonlinearity of the receiver ($G^{1/\gamma}$, $R^{1/\gamma}$, $B^{1/\gamma}$).
3. A transmitter subcarrier modulation system forming a chrominance signal which, when detected in a *constant-luminance* color receiver, supplies the three signals $G^{1/\gamma} - M$, $R^{1/\gamma} - M$, $B^{1/\gamma} - M$.

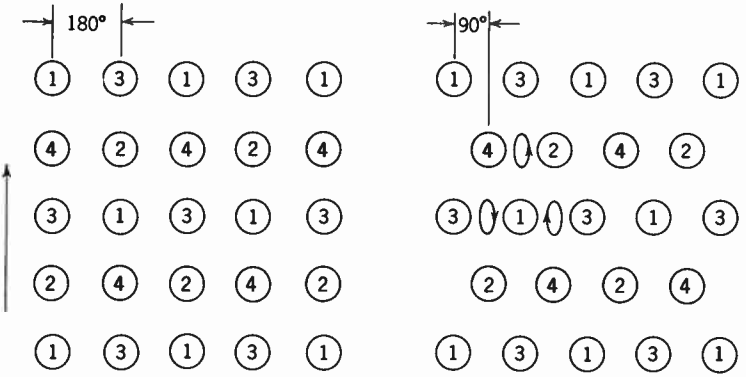
18.6 Correction of Residual Picture Defects. A number of improvements have been proposed and tested which have the objective of minimizing residual defects in the color picture. In particular, they seek to reduce "dot crawl," to lessen the interference of high-frequency signals in the chrominance channel, and to extend the frequency range of faithful color reproduction. Even though these proposals have not been incorporated in the standards described in section 18.9, they are of interest as illustrating alternative approaches to the problem of obtaining satisfactory color reproduction.

It has been found that dot crawl—an apparent upward motion of the dot pattern in the picture resulting from the sequence in which dot patterns are laid down—can be avoided by imposing a 90-degree shift on the sampling frequency for each successive field. In Fig. 18.32, the dot patterns for monochrome fields are shown without and with this 90-degree shift. The circles represent points of maximum intensity and the numbers within them the field to which they correspond. The time sequence 1–2–3–4 which the eye tends to follow corresponds to a continuous upward motion without the shift; with the shift, it follows narrowly confined circular paths or zigzag paths, which lead to much less readily perceived flicker effects.

A proposal to minimize low-frequency color cross-talk resulting from the admixture of high-frequency monochrome components to the chrominance signal utilizes modulation of the monochrome components

at twice subcarrier frequency. The high-frequency components so generated are added to the brightness signal and generate compensating signals in the receiver.

The functioning of this procedure is most obvious in the zero-luminance-blue system (Fig. 18.33). The monochrome signal is first passed through a low-pass filter with cutoff (more precisely, 50 percent reduction in response) at the subcarrier frequency. It then is sampled at twice subcarrier frequency, the sampling oscillation being obtained by



(a) Odd multiple of one-half line frequency

(b) Same, plus 90° shift on every other field

Fig. 18.32. Dot Patterns for Monochrome Fields without (*a*) and with (*b*) 90-Degree Phase Shift in Alternate Fields. (Loughlin, reference 18.) (Courtesy of *Proceedings of the Institute of Radio Engineers.*)

frequency multiplication of the source for the blue transmitter modulator. The low-frequency sideband (up to $\omega_c/(2\pi)$) resulting from this sampling is thus added to the video signal, consisting of the monochrome signal and the chromaticity signal.

The effect of the added frequency components between the subcarrier frequency and the video cutoff frequency is the following. An arbitrary frequency component $A \cos(\omega_f t + \phi)$, with ω_f between $2\omega_s - \omega_c$ and ω_s, has added to it a term

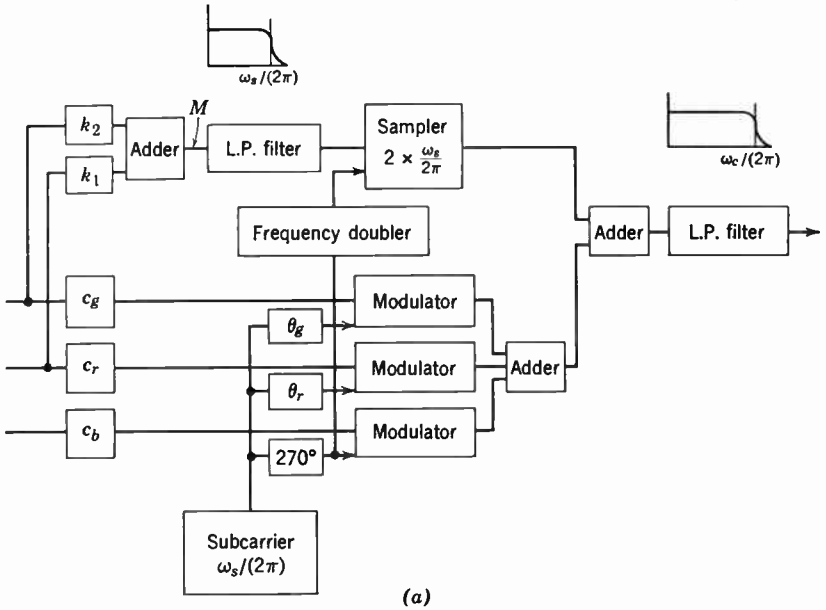
$$A \cos(2[\omega_s t + 2\pi/2] - \omega_f t - \phi) = -A \cos(2\omega_s t - \omega_f t - \phi)$$

In the green and red demodulators, with $\theta' = 0$, 180 degrees, the two terms are converted into

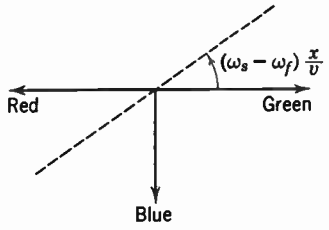
$$\pm \left\{ \frac{A}{2} \cos[(\omega_s - \omega_f)t - \phi] - \frac{A}{2} \cos[(\omega_s - \omega_f)t - \phi] \right\} = 0$$

Here, the added terms hence cancel the spurious chromaticity terms resulting from the interference of the high-frequency brightness components. In the blue channel, on the other hand, the result is

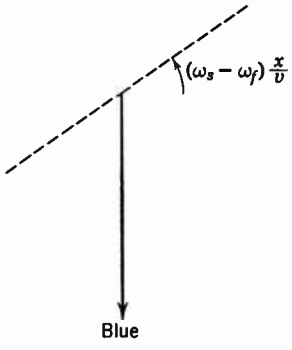
$$-\frac{A}{2}\sin[(\omega_s-\omega_f)t-\phi]-\frac{A}{2}\sin[(\omega_s-\omega_f)t-\phi]$$
$$=-A\sin[(\omega_s-\omega_f)t-\phi]$$



(a)



(b)



(c)

Fig. 18.33. Reduction of Spurious Chrominance Signals by Modulation of High-Frequency Components in Zero-Luminance-Blue System. (*a*) Block diagram. (*b*) Signals without modulation. (*c*) Signals with modulation.

The spurious terms are doubled by the added signals. The interference of the high-frequency terms in the chromaticity channel is hence shifted entirely to the blue picture, where their presence is least objectionable. Furthermore, in view of the low acuity of the eye for the blue, the passband for the blue channel can be greatly reduced, eliminating the higher-frequency spurious components altogether.

In addition to shifting the spurious signals into the blue channel, the procedure outlined restricts the horizontal resolution by shifting the effective cutoff frequency downward to the subcarrier frequency. Finally, the added terms produce a fine-grained brightness variation in the picture; with the 90-degree shift in alternate fields, this variation (i.e., a dot pattern for a particular frequency component) is out of phase in successive lines and hence has minimum visibility. Without the shift, the pattern becomes a pattern of vertical lines. Dot interlacing is not effective in canceling the signals in the course of $\frac{1}{15}$ second, since the frequency multiplication of the subcarrier converts a 180-degree shift into a 360-degree shift.

The last, and most important, improvement to be considered is the reversal of the sampling order in successive scanning fields. "Color phase alternation" overcomes, in particular, color cross-talk, or defects in color rendition, which occurs when the low-pass filters in the chrominance channels pass frequency components in excess of $(\omega_c - \omega_s)/(2\pi)$, the difference between the video cutoff frequency and the subcarrier frequency. Color cross-talk is particularly serious if the subcarrier frequency is chosen close to the upper limit of the video band. This choice renders the dot pattern resulting from the presence of the chrominance signal in the brightness channel less visible by making it more fine-grained and, in view of the attenuation of most receivers near video cutoff, reducing its amplitude. At the same time, it reduces the frequency range over which proper color rendition is obtained unless the remedy offered by color phase alternation is applied. The effect of reversing the color sequence in successive fields is to make the color defect in each even line of the picture opposite in sign to that of the preceding odd line, so that, in view of the reduced flicker sensitivity and acuity of the eye for chromaticity differences, the defects cancel out visually.

Both the color error and its compensation by color phase alternation can be demonstrated most simply for modulation and demodulation of two subcarrier components in quadrature. Figures 18.29 and 18.30 show that, for constant-luminance systems, the modulating functions are then very closely approximated by $K_1(R - M)$ and $K_1K_2(B - M)$,

where K_1 and K_2 are constants. Assume that the components of frequency $\omega_f/(2\pi)$ in the signals $R - M$ and $B - M$ are $r_f \cos(\omega_f t + \phi_{rf})$ and $b_f \cos(\omega_f t + \phi_{bf})$ respectively. Furthermore, it is necessary to introduce the amplitude response of the video circuits, with the nominal cutoff $\omega_c/(2\pi)$, which will be designated by $F(\omega)$. Then the chrominance signal for the component in question is given by multiplying the two terms with $\cos \omega_s t$ and $\sin \omega_s t$ and modifying the resultant components by the appropriate factor $F(\omega)$:

$$\begin{aligned} S_f = \frac{K_1}{2} \{ & F(\omega_s - \omega_f)[r \cos[(\omega_s - \omega_f)t - \phi_{rf}] \\ & + K_2 b \sin[(\omega_s - \omega_f)t - \phi_{bf}]] \\ & + F(\omega_s + \omega_f)[r \cos[(\omega_s + \omega_f)t + \phi_{rf}] \\ & + K_2 b \sin[(\omega_s + \omega_f)t + \phi_{bf}]] \} \end{aligned} \tag{18.35}$$

The output signals of the demodulator are obtained from this expression by multiplying it with $(2/K_1) \cos \omega_s t$ and $(2/(K_1K_2)) \sin \omega_s t$, respectively, and dropping the high-frequency terms:

$$\begin{aligned} (R_1 - M)_f = & \frac{r}{2}(F(\omega_s - \omega_f) + F(\omega_s + \omega_f)) \cos(\omega_f t + \phi_{rf}) \\ & - K_2 \frac{b}{2}(F(\omega_s - \omega_f) \\ & - F(\omega_s + \omega_f)) \sin(\omega_f t + \phi_{bf}) \end{aligned} \tag{18.36a}$$

$$\begin{aligned} (B_1 - M)_f = & \frac{1}{2K_2} r(F(\omega_s - \omega_f) - F(\omega_s + \omega_f)) \sin(\omega_f t + \phi_{rf}) \\ & + \frac{b}{2}(F(\omega_s - \omega_f) \\ & + F(\omega_s + \omega_f)) \cos(\omega_f t + \phi_{bf}) \end{aligned} \tag{18.36b}$$

$$(G_1 - M)_f = -\frac{k_1}{k_2}(R_1 - M) - \frac{k_3}{k_2}(B_1 - M) \tag{18.36c}$$

If both sidebands for the frequency $\omega_f/(2\pi)$ are transmitted without attenuation, i.e., $F(\omega_s + \omega_f) = F(\omega_s - \omega_f) = 1$, the original signals applied to the transmitter modulator, namely, $r \cos(\omega_f t + \phi_{rf})$ and $b \cos$

($\omega_f t + \phi_{bf}$), are recovered without any distortion. On the other hand, if only one sideband is transmitted for frequencies in excess of some fixed value (e.g., $\omega_f > \omega_c - \omega_s$), the amplitude of these components is reduced to half, and, in addition, spurious cross-talk terms appear which are in quadrature with the principal terms. Figure 18.34 shows their effect on the transmission of a narrow yellow (minus blue) bar in a

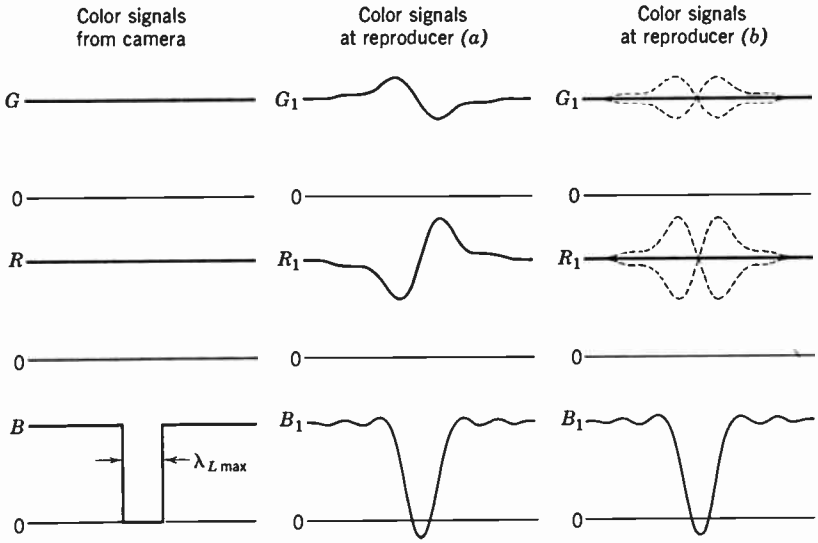


Fig. 18.34. Reproduction of a Narrow Yellow Bar in a White Field in the Zero-Luminance-Blue System (*a*) without Color Phase Alternation and (*b*) with Color Phase Alternation. (Average for two successive lines.)

white field. The chrominance signal is here zero except during the scanning of the yellow bar, and should give rise exclusively to a reduction in the blue signal. However, the cutting off of the upper sideband of the subcarrier leads to the appearance of quadrature terms of opposite phase in the red and green channels. These make one edge of the reproduced yellow bar appear greenish, the other edge orange.

Consider now the next line of the picture, in the succeeding field, and assume that the color sequence here is reversed. This reversal may be accomplished by replacing $\cos \omega_s t$ in the transmitter modulator and the receiver demodulator by $-\cos \omega_s t$ and leaving the sign of $\sin \omega_s t$ unaltered. If the amplitude and phase of the chromaticity signal components on the new line are distinguished by primes, Eqs. 18.36*a* and 18.36*b* are now replaced by

$$(R_1 - M)'_f = \frac{r'}{2}(F(\omega_s - \omega_f) + F(\omega_s + \omega_f)) \cos(\omega_f t + \phi'_{rf}) + K_2 \frac{b'}{2}(F(\omega_s - \omega_f) - F(\omega_s + \omega_f)) \sin(\omega_f t + \phi'_{bf}) \quad (18.37a)$$

$$(B_1 - M)'_f = \frac{-1}{2K_2} r'(F(\omega_s - \omega_f) - F(\omega_s + \omega_f)) \sin(\omega_f t + \phi'_{rf}) + \frac{b'}{2}(F(\omega_s - \omega_f) + F(\omega_s + \omega_f)) \cos(\omega_f t + \phi'_{bf}) \quad (18.37b)$$

In brief, the signs of the spurious cross-talk terms are reversed, whereas the remaining terms are unaltered. Assume now that the video response is made antisymmetrical about the subcarrier frequency or, in

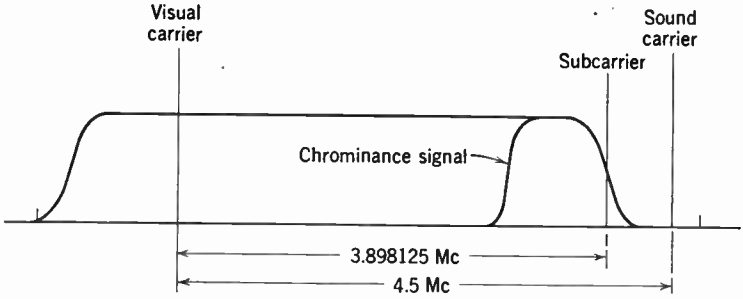


Fig. 18.35. Placement of Subcarrier for Vestigial Sideband Transmission.

other words, that the subcarrier frequency is placed at the point at which the video response has dropped to half its value at low frequencies (Fig. 18.35). Then

$$F(\omega_s + \omega_f) = 1 - F(\omega_s - \omega_f) \quad (18.38)$$

If, furthermore, the signals in adjoining lines are assumed to be substantially unchanged and the amplification of the chrominance signal is doubled, the wanted terms attain exactly the right magnitude with this vestigial sideband transmission of the subcarrier, and the distortion terms appear with opposite signs in successive lines:

$$R_1 - M = {R_L}^{1/\gamma} - M_L \mp D_b \qquad B_1 - M = {B_L}^{1/\gamma} - M_L \pm D_r$$

$$G_1 - M = {G_L}^{1/\gamma} - M_L \pm \frac{k_1}{k_2} D_b \mp \frac{k_3}{k_2} D_r \tag{18.39}$$

$$D_b = K_2 \sum_f b(1 - 2F(\omega_s + \omega_f)) \sin(\omega_f t + \phi_{bf})$$

$$D_r = (1/K_2) \sum_f r(1 - 2F(\omega_s + \omega_f)) \sin(\omega_f t + \phi_{rf})$$

This result is obtained by summing over all the frequency terms (ω_f) passed by the low-pass filters; the subscript L indicates the corresponding low-frequency portion of the signals and γ represents the kinescope nonlinearity. For $\gamma = 1$, the brightness of the component pictures becomes directly proportional to the applied signals, R_1, B_1, G_1 so that, in this case, the spurious terms D_b and D_r cancel out completely for a pair of lines and in the course of two successive fields ($\frac{1}{30}$ second). If $\gamma \neq 1$, terms of even order in the ratio of the quadrature terms to the wanted terms lead to some false coloration in the neighborhood of sharp transitions in chromaticity. For instance, the brightness of the red partial image is given by

$$\overline{k_1 {R_1}^{\gamma}} = k_1({R_L}^{1/\gamma} + M_H) \left\{ 1 + \frac{\gamma(\gamma - 1){D_b}^2}{2({R_L}^{1/\gamma} + M_H)^2} + \cdots \right\}$$

the bar denoting averaging over two successive lines. It should be noted that the very low frequency terms in the camera signals do not contribute to the distortion terms, since here $1 - 2F(\omega_s + \omega_f) = 0$.

Color phase alternation also serves to correct for large-area color errors resulting either from phase distortion of the transmitted chromaticity signal or inaccurate phasing of the subcarrier reinserted in the receiver. Since the subcarrier is close to the cutoff frequency of the video band, some phase distortion may easily occur in either transmitter or receiver. If the phase distortion at the subcarrier frequency is $\Delta\theta$ and $\omega_f \ll \omega_s$ so that both phase and amplitude response are practically equal for the upper and lower sidebands for this particular frequency, the corresponding chrominance signal may be written (see Eq. 18.35):

$$\begin{aligned} S_f = \frac{K_1}{2} \{ \pm r[&\cos[(\omega_s - \omega_f)t - \phi_{rf} + \Delta\theta] \\ &+ \cos[(\omega_s + \omega_f)t + \phi_{rf} + \Delta\theta]] \\ &+ K_2 b[\sin[(\omega_s - \omega_f)t - \phi_{bf} + \Delta\theta] \\ &+ \sin[(\omega_s + \omega_f)t + \phi_{bf} + \Delta\theta]] \} \end{aligned} \tag{18.40}$$

The corresponding red output signal, obtained by multiplication with $\pm (2/K_1) \cos \omega_s t$, is

$$(R_1 - M)_f = \frac{r}{2} [\cos (\omega_f t + \phi_{rf} - \Delta\theta) + \cos (\omega_f t + \phi_{rf} + \Delta\theta)]$$

$$\mp \frac{K_2 b}{2} [\sin (\omega_f + \phi_{bf} - \Delta\theta) - \sin (\omega_f t + \phi_{bf} + \Delta\theta)]$$

$$= r \cos \Delta\theta \cos (\omega_f t + \phi_{rf})$$

$$\pm K_2 b \sin \Delta\theta \cos (\omega_f t + \phi_{bf}) \qquad (18.41)$$

The phase error hence has the effect, with color phase alternation, of reducing the chrominance signals (and hence the saturation of the reproduced colors) by a factor $\cos \Delta\theta$. In addition, it gives rise to cross-talk terms, which cancel exactly for two successive lines if the kinescope response is linear. With the actual, nonlinear, kinescope response, cancellation is incomplete. There are residual cross-talk terms proportional to $\gamma(\gamma - 1) \tan^2 \Delta\theta$. It has been found experimentally that acceptable color rendition can be secured with phase errors as large as 20 to 30 degrees; however, under these circumstances the large-area, 30-cycle, color flicker resulting from the oscillation of the color sequence becomes annoying, so that it is desirable in practice to keep the phase error below 15 or 20 degrees.

If both phase errors and unequal transmission for the two sidebands (as for the higher-frequency components with vestigial subcarrier transmission) are present, the cancellation of color cross-talk is incomplete even for a linear reproducing system.

Incidentally, a color-transmission system with equal modulation and demodulation angles, such as the constant-luminance system represented in Fig. 18.29, possesses special advantages when employed with a nonlinear receiver. Substitution of the expressions of the parameters c'_g, c'_r, c'_b and θ'_g, θ'_r, θ'_b in terms of k_g, k_r, k_b in Eq. 18.34 shows that the quadratic term in the luminance expression reduces here to $\gamma(\gamma - 1)A^2/(4M^2)$, i.e., is independent of the phase of the chrominance signal. It follows that for such a system, the luminance flicker at edges occurring with color phase alternation vanishes for $\gamma = 2$ and is greatly reduced generally.

It should be mentioned that other picture defects may occur which can be minimized simply by a suitable choice of standards. An example is the visibility of sound carrier interference in the picture. Unless the attenuation of the visual i-f stages of the receiver for the sound signal is very high, a beat note corresponding to the frequency difference be-

tween the sound signals and the chrominance subcarrier is generated in the subsequent nonlinear stages of the receiver. If the subcarrier is placed at about 3.9 megacycles and the sound carrier at 4.5 megacycles, the beat note is approximately 600 kilocycles. This corresponds to about 30 maxima and minima per horizontal line and is hence easily recognized. It gives rise to a series of bars in the picture whose slope depends on the exact choice of the sound carrier frequency. However, their visibility can be minimized by reversing the polarity of the bars every $\frac{1}{30}$ second; this is accomplished by making the sound carrier an integer multiple of the line frequency, causing the beat note (like the color subcarrier itself) to be a half-integer multiple of the line frequency.

18.7 Orange-Cyan Wideband System. By far the most serious picture defect discussed in the preceding section is color cross-talk at edges resulting from the attenuation of the upper sideband of the chrominance signal. Though color phase alternation is very effective in removing the resulting color fringes, it has the drawback of an added complication in receiver design and the introduction of low-frequency (30-cycle) color flicker at edges. Color fringes can also be removed without these drawbacks, at the expense of relatively unimportant departures from color fidelity at transitions. This is accomplished by restricting tricolor representation of the scene to a frequency range for which double-sideband transmission of the chrominance signal is possible and employing an optimal two-color presentation in the single-sideband range.

For example, the blue color difference signal may be restricted to the frequency range for which both sidebands of the chrominance signal are passed without attenuation. With $\omega_s/(2\pi) = 3.89$ megacycles this may be 0.3 megacycle. The red difference signal may then be transmitted without cross-talk out to, e.g., 1.5 megacycles, the amplitude attenuation of the upper sideband in this range being compensated by giving the low-pass filter ahead of the red modulator a rising characteristic $F'(\omega_f)$ between 0.3 and, e.g., 1.1 megacycles. Equation 18.36 is now replaced in the intermediate-frequency range simply by

$$(R_1 - M)_f = \frac{r}{2} F'(\omega_f)(F(\omega_s - \omega_f) + F(\omega_s + \omega_f)) \cos(\omega_f t + \phi_{rf}) \tag{18.42a}$$

$$(B_1 - M)_f = 0 \tag{18.42b}$$

$$(G_1 - M)_f = -\frac{k_1}{k_2}(R_1 - M) \tag{18.42c}$$

Thus, with this system, all colors are reproduced faithfully for a frequency range up to 0.3 megacycle; furthermore, the red color difference signal is reproduced accurately for the intermediate-frequency

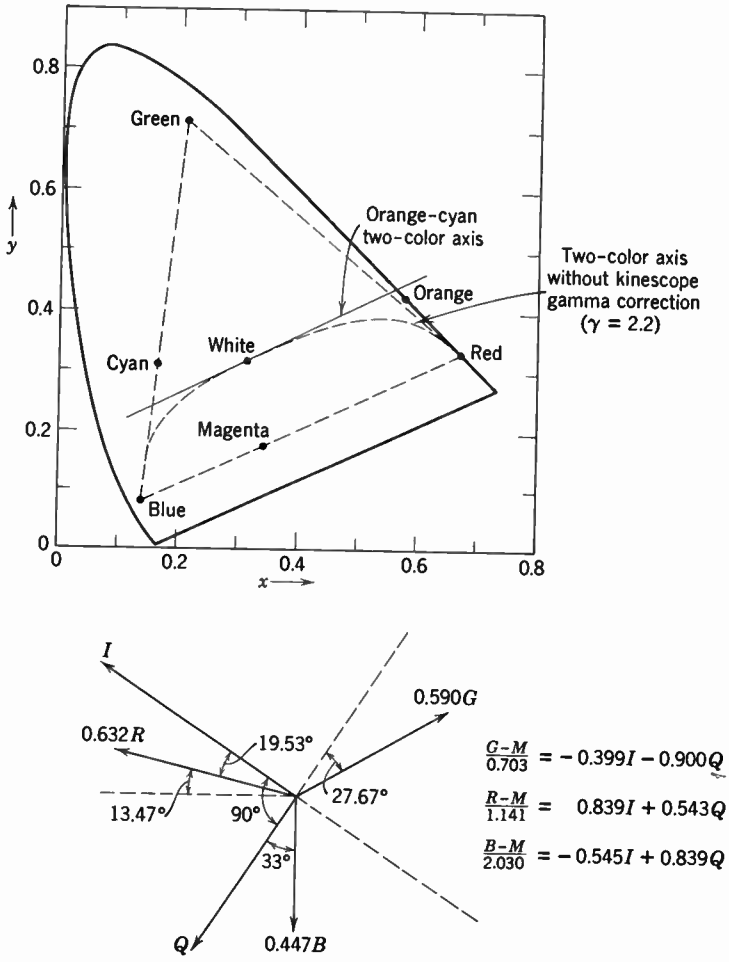


Fig. 18.36. Location of Orange-Cyan Axis on Color Triangle and Relationship of I and Q and G, R, and B in Orange-Cyan Wideband System.

range provided that within this range the low-pass filter characteristic is made such that $(\frac{1}{2})F'(\omega_f)\,(F(\omega_s - \omega_f) + F(\omega_s + \omega_f)) = 1$. In this range, the faithfulness of color reproduction increases in proportion as the relative luminosity of the blue primary is made smaller. Here, the system is a two-color, rather than a three-color, system.

In practice, a lower subcarrier frequency, $\omega_s/(2\pi) = 3.58$ megacycles, serves to extend the range of faithful color presentation. At the same time, increasing the separation between cutoff frequency and subcarrier

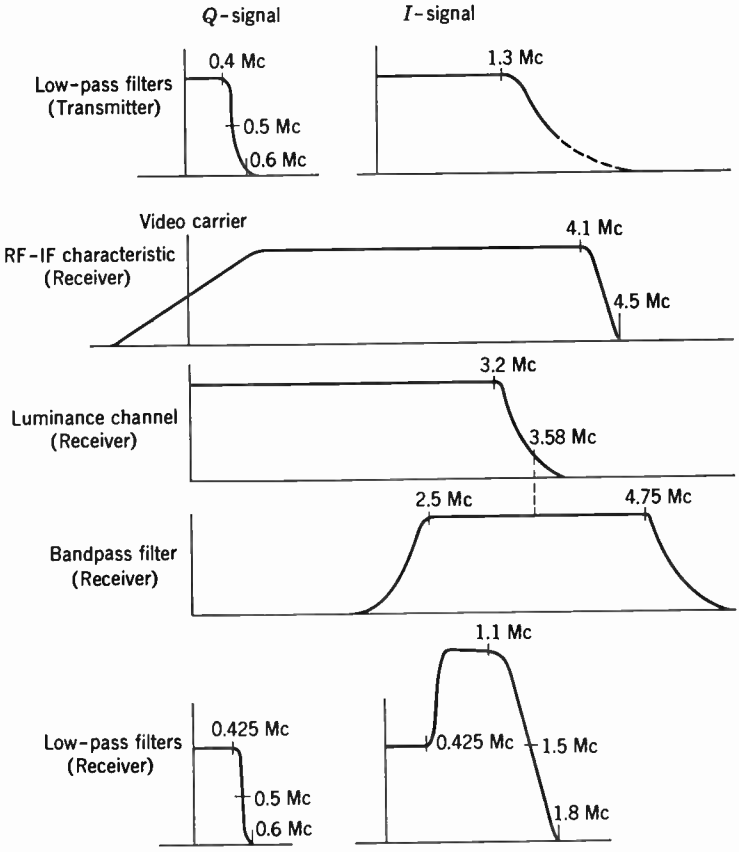


Fig. 18.37. Transmission Characteristics of Orange-Cyan Wideband System.

simplifies the problem of phase and amplitude response in the neighborhood of the subcarrier. Furthermore, hue changes at edges are minimized by choosing orange and its complementary hue cyan to determine the color axis for accurate transmission throughout the intermediate-frequency range. The corresponding color difference signal is denoted by I:

$$I = -0.274G + 0.596R - 0.322B$$

The second color difference signal Q modulating the subcarrier in quadrature with I and transmitted up to 0.5 megacycle corresponds to colors in the range between magenta and its complementary green:

$$Q = -0.522G + 0.211R + 0.311B$$

Figure 18.36 shows the location of the orange-cyan axis on the color triangle as well as the vector diagram indicating the relationship between the I and Q signals and the camera signals for a constant-lumi-

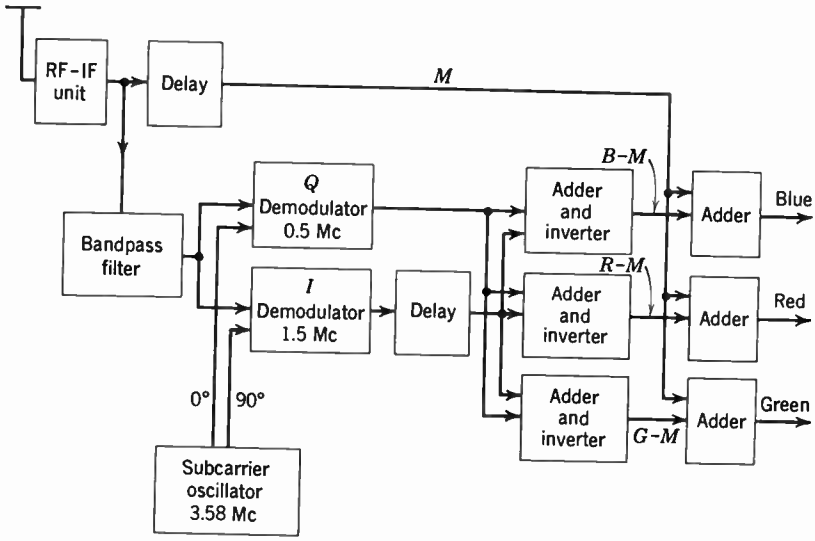


Fig. 18.38. Block Diagram of Receiver for Orange-Cyan Wideband System.

nance system as well as the formulas for the recovery of the three-color signals at the receiver. The transmission characteristics of the several channels of the system are shown in Fig. 18.37. It will be noted that the amplitude compensation of the I-signal in the single-sideband region is carried out in the receiver rather than in the transmitter. The converse procedure would tend to accentuate dot structure in the picture resulting from the superposition of the chrominance signal on the luminance signal.

Finally, a block diagram of a receiver for the system described is given in Fig. 18.38. An added delay is inserted in the I-signal channel to equalize the overall delays in the I- and Q-channels.

18.8 Frequency Interlace Color Television System. It was shown in Chapter 5 that the frequency spectrum of the video signal derived

from a stationary picture consisted of clusters of components differing by the frame frequency about integer multiples of the line frequency (Fig. 18.39). It can be inferred herefrom that signal components halfway between these clusters could be added without affecting appreciably the appearance of the picture averaged over the time. This is exactly what happens when a chrominance signal, obtained by modulating a subcarrier which is a half-integer multiple of the line frequency by camera signals, is added to the usual video signal. But for nonlinear effects and the retentivity of the eye, the effects of the

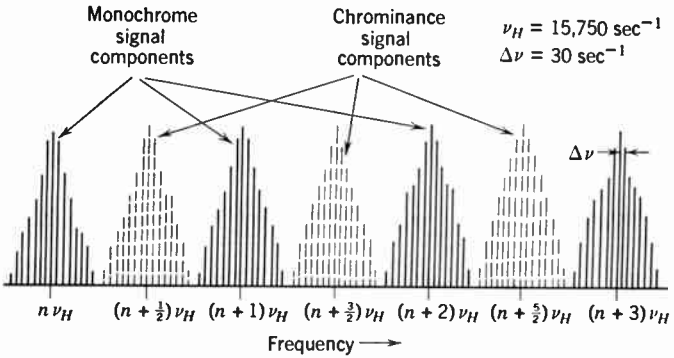


Fig. 18.39. Distribution of Monochrome Signal Components and Chrominance Signal Components in the Frequency Spectrum.

chrominance signal, applied to a black-and-white receiver, cancel out in the course of twice the frame time. The phase-amplitude-modulated subcarrier system with dot interlacing, which has formed the subject of the preceding sections, can thus be regarded as a specialized example of a frequency interlace television system.

In the phase-amplitude-modulated subcarrier system the two quadrature components of the subcarrier are modulated independently, which is equivalent to a simultaneous modulation of amplitude and phase, as implied by the name. With this system both the frequency and the phase of the subcarrier reinserted in the receiver must be accurately controlled by signals from the transmitter.

If, instead, two separate subcarriers, both half-integer multiples of the line frequency, are employed to transmit the low-frequency color information given by, e.g., the red and blue camera units respectively, whereas the main carrier is modulated directly by the low-frequency green components and the mixed-highs (Fig. 18.20), the phase synchronization of the subcarriers in transmitter and receiver becomes

unnecessary. Filters select out the two chromaticity signals in the receiver; they are then demodulated by the appropriate local subcarriers to provide the desired red and blue signals (Fig. 18.40).

The most serious disadvantage of this system as compared with the PAM subcarrier system is that beat notes between the red and blue chromaticity signals, generated in nonlinear elements of the trans-

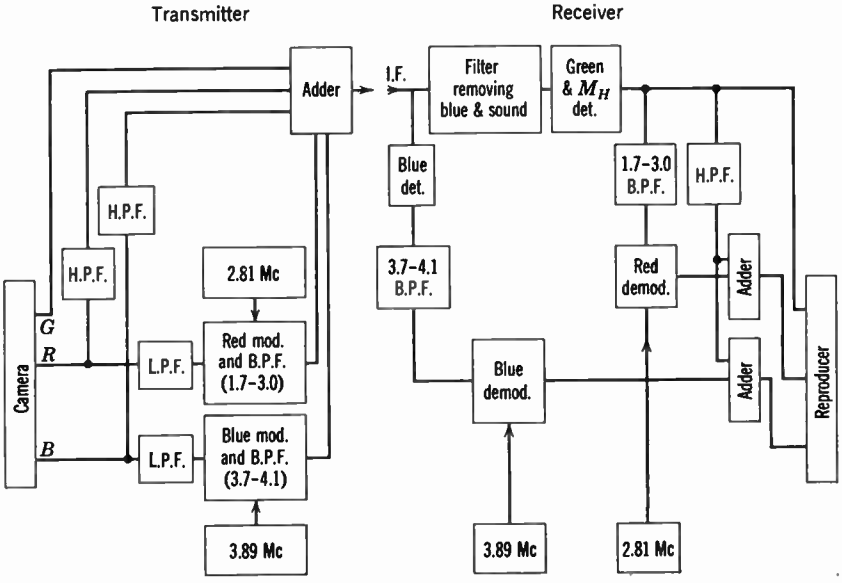


Fig. 18.40. Frequency Interlace System with Two Subcarriers.

mitter and receiver, are not canceled in the course of a $\frac{1}{15}$-second period and may hence be very disturbing.

Several modifications of the frequency interlace system have been proposed. One of these, the "alternating highs" system represented in Fig. 18.41, avoids color cross-talk by completely separating the color channels. Adequate horizontal resolution is obtained even so by transmitting different portions of the mixed-highs spectrum in alternate fields or frames. It is thus obtained at the expense of increased detail flicker and, eventually, reduced vertical resolution. The difficulty of visible beat notes between the color carriers exists here as in the previously described system.

In a second system, the "alternating lows" system indicated in Fig. 18.42, the red and blue low-frequency signals modulate a quadrature component of the principal picture carrier on alternate lines (or red

modulation takes place on two lines, blue modulation on every third line). This system requires an accurate phase relation between the picture carrier in the transmitter and receiver and is thus likely to

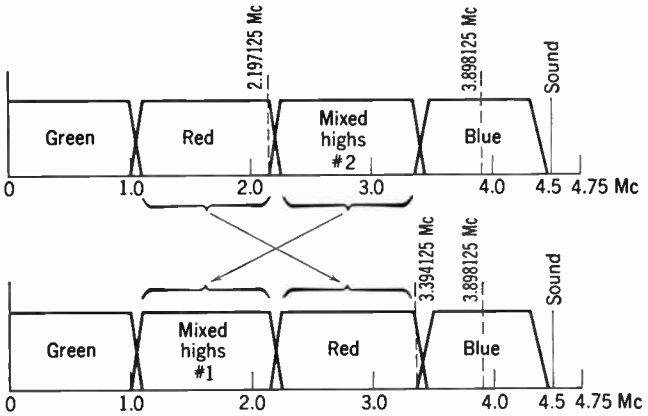


Fig. 18.41. Alternating Highs System: Channel Utilization in Successive Periods. (R. B. Dome, reference 20.) (Courtesy of *Proceedings of the Institute of Radio Engineers.*)

violate the compatibility requirement. For a given frequency range of color information, the information transmitted on the monochrome channel is limited by the requirement of transmitting both sidebands of the color signals to avoid color cross-talk. Apart from this, the

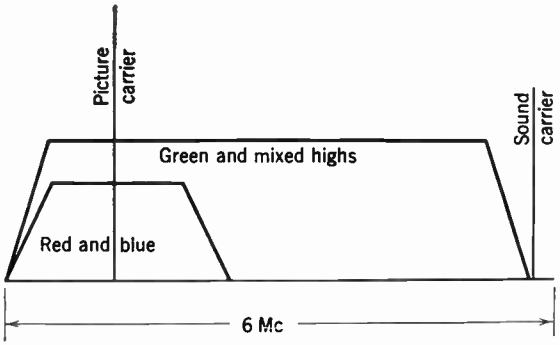


Fig. 18.42. Alternating Lows System: Channel Utilization. (R. B. Dome, reference 20.) (Courtesy of *Proceedings of the Institute of Radio Engineers.*)

transmission of the red and blue signals with the same subcarrier is made possible, of course, by reducing the vertical resolution for these components.

18.9 Standardization of the Color Television Signal. In the face of a great variety of different proposals for the transmission of natural-color pictures by television, the television industry, through the National Television System Committee (NTSC), set itself the task of devising tentative standards which would receive the endorsement of the industry as a whole and could be recommended to the Federal Communications Commission as a basis for color television transmissions. These standards were approved by the FCC in December, 1953. It was found that agreement on the principal features of the proposed standards could be achieved without much difficulty. The standards had to permit the operation of a system which was compatible, had to yield pictures with detail resolution of the same order as that obtained with standard black-and-white transmissions, and be free from noticeable flicker effects under normal conditions of viewing. Apart from questions of picture quality, cost considerations, particularly with reference to receiver construction, entered into the final decisions. The major features of the standards, which fit the previously described orange-cyan wideband system, are enumerated below:

1. The transmitted signal consists of a phase-amplitude-modulated subcarrier chrominance signal added to a luminance signal.

2. The nonlinearity of the receiver is compensated at the transmitter, ahead of the subcarrier modulation stages. The value for the assumed contrast gain of the receiver, γ, is 2.2.

3. The luminance signal is given by $E_Y = k_1E_R + k_2E_G + k_3E_B$, equal signals E_R, E_G, E_B applied to the red, green, and blue channels of the reproducer yielding a white picture with a tone corresponding to the CIE "C illuminant" (daylight). The values of k_1, k_2, and k_3 have been set at $k_1 = 0.30$, $k_2 = 0.59$, $k_3 = 0.11$. The k's correspond to the relative luminous efficiencies of a certain set of phosphors. The index Y indicates that the signal in question is a measure of luminance, designated by Y in the CIE system of colorimetry. The primes indicate that gamma correction for the receiver characteristic has been applied.

4. The chrominance signal has the form

$$S = E'_Q \sin(\omega_s t + 33°) + E'_I \cos(\omega_s t + 33°)$$

with

$$E'_Q = 0.41(E'_B - E'_Y) + 0.48(E'_R - E'_Y)$$

$$E'_I = -0.27(E'_B - E'_Y) + 0.74(E'_R - E'_Y)$$

and

$$\omega_s/(2\pi) = 3{,}579{,}545 \text{ sec}^{-1} = 227\tfrac{1}{2} \cdot 15{,}734.264 \text{ sec}^{-1}$$

the horizontal frequency being set at 15,734.264 second^{-1}.

The Q-signal is passed through a low-pass filter which is flat within 2 decibels to 400 kilocycles, attenuates by less than 6 decibels at 500 kilocycles, and by more than 6 decibels beyond 600 kilocycles. The I-signal is passed through a low-pass filter which attenuates by less than 2 decibels at 1.3 megacycles and by more than 20 decibels at 3.6 megacycles. In the low-frequency region (to about 500 kilocycles), the chrominance signal has the form

$$S = K_1((E'_R - E'_Y) \cos \omega_s t + K_2(E'_B - E'_Y) \sin \omega_s t)$$

This signal is essentially that for the constant-luminance system indicated in Fig. 18.29, since $K_2 = 1/1.78$. The value $K_1 = 1/1.14$ has been chosen small enough that the chrominance signal superposed on the brightness signal does not drive the kinescope gun beyond cutoff under any circumstances. Otherwise grid rectification would severely limit the cancellation of the chrominance signal obtained with dot interlacing.

It will be recognized that the chrominance signal specified eliminates color fringing by providing for full three-color rendition only in the range of double-sideband transmission of the chrominance signal, i.e., approximately 500 kilocycles, and for two-color rendition along an orange-white-cyan axis for an intermediate frequency range, to beyond 1.3 megacycles, as indicated in section 18.7.

5. The primary colors to which E'_R, E'_G, and E'_B refer have the following CIE color coordinates:

R	$x = 0.67$	$y = 0.33$
G	$x = 0.21$	$y = 0.71$
B	$x = 0.14$	$y = 0.08$

With this selection, the chromaticity produced when the chrominance signal vanishes corresponds to illuminant C ($x = 0.310$, $y = 0.316$).

6. The separation of the sound carrier and the picture carrier is set at 4.5 megacycles $\pm$ 1000 cycles. For the specified position of the color subcarrier (3,579,545 second^{-1}), the frequency difference between the sound carrier and the color subcarrier then becomes just 58½ times the horizontal frequency. As a result, interference patterns of this frequency in the picture tend to be canceled visually in successive frames.

Additional specifications relate to the synchronizing signal and signal delays. Some of these matters will be considered in the next chapter, which will discuss the practical realization of color television systems.

The establishment of standards for color television along the above lines was recommended to the FCC by RCA in a petition of June 25, 1953. The recommendation of the NTSC was submitted a month later, on July 23, 1953. Approval by the FCC took place on December 17, 1953.

REFERENCES

1. D. B. Judd, "The 1931 Standard Observer and Coordinate System for Colorimetry," *J. Opt. Soc. Amer.*, Vol. 23, pp. 359–374, 1933.
2. A. C. Hardy and F. C. Wurzburg, Jr., "The Theory of Three-Color Reproduction," *J. Opt. Soc. Amer.*, Vol. 27, pp. 227–240, 1937.
3. D. L. MacAdam, "Visual Sensitivity to Color Differences in Daylight," *J. Opt. Soc. Amer.*, Vol. 32, pp. 247–274, 1942.
4. D. L. MacAdam, "Quality of Color Reproduction," *Proc. I.R.E.*, Vol. 39, pp. 468–485, 1951.
5. W. T. Wintringham, "Color Television and Colorimetry," *Proc. I.R.E.*, Vol. 39, pp. 1135–1172, 1951.
6. D. G. Fink, "Color Fundamentals for TV Engineers," *Electronics*, pp. 88–93, December, 1950; pp. 78–83, January, 1951; and pp. 104–109, February, 1951.
7. G. C. Sziklai, "A Tristimulus Color Photometer," *J. Opt. Soc. Amer.*, Vol. 41, pp. 321–323, 1951.
8. A. V. Bedford, "Mixed Highs in Color Television," *Proc. I.R.E.*, Vol. 38, pp. 1003–1009, 1950.
9. M. W. Baldwin, Jr., "Subjective Sharpness of Additive Color Pictures," *Proc. I.R.E.*, Vol. 39, pp. 1173–1176, 1951.
10. J. E. Benson, "A Survey of the Methods and Colorimetric Principles of Colour Television," *A.W.A. Tech. Rev.*, Vol. 9, pp. 99–157, 1951.
11. D. W. Epstein, "Colorimetric Analysis of RCA Color Television System," *RCA Rev.*, Vol. 14, pp. 227–258, 1953.
12. W. H. Cherry, "Colorimetry in Television," *Proc. Soc. Motion Picture Engrs.*, Vol. 51, pp. 613–642, 1948.
13. "Six-Megacycle Compatible High Definition Color Television System," *RCA Rev.*, Vol. 10, pp. 504–524, 1949.
14. W. Boothroyd, "Dot Systems of Color Television," *Electronics*, Vol. 22, pp. 88–92, December, 1949, and Vol. 23, pp. 96–99, January, 1950.
15. A. V. Loughren and C. J. Hirsch, "Comparative Analysis of Color Television Systems," *Electronics*, Vol. 24, pp. 92–96, February, 1951.
16. "An Analysis of the Sampling Principles of the Dot-Sequential Color Television System," *RCA Rev.*, Vol. 11, pp. 255–286, 1950.
17. N. Marchand, H. R. Holloway, and M. Leifer, "Analysis of Dot-Sequential Color Television," *Proc. I.R.E.*, Vol. 39, pp. 1280–1287, 1951.
18. D. B. Loughlin, "Recent Improvements in Band-Shared Simultaneous Color-Television Systems," *Proc. I.R.E.*, Vol. 39, pp. 1264–1279, 1951.

19. R. B. Dome, "Frequency-Interlaced Color Television," *Electronics,* Vol. 23, pp. 70–75, September, 1950.
20. R. B. Dome, "Spectrum Utilization in Color Television," *Proc. I.R.E.,* Vol. 39, pp. 1323–1331, 1951.
21. R. D. Kell and A. C. Schroeder, "Optimum Utilization of the Radio Frequency Channel for Color Television," *RCA Rev.,* Vol. 14, pp. 133–143, 1953.
22. G. H. Brown and D. G. C. Luck, "Principles and Development of Color Television Systems," *RCA Rev.,* Vol. 14, pp. 144–204, 1953.
23. "Technical Signal Specifications Proposed as Standards for Color Television," *RCA Rev.,* Vol. 14, pp. 358–366, 1953.

19 Practical Color Television

Chapter 18 stressed the principles on which practical compatible color television systems are based. To clarify the presentation, the characteristics of the components of such systems were frequently idealized. This chapter discusses the form which these components have taken in actual field tests and gives a brief review of the special circuitry involved in the transmission and reception of color pictures. The last section gives an outline of the field-sequential system which for an interim period was made standard by a ruling of the Federal Communications Commission in October, 1950, and finds practical application in industrial television systems.

19.1 Generation of Color Signals. Fundamentally, any pickup system applicable to black-and-white television may be adapted for use as a component of a simultaneous color television camera. The two systems which have found special favor are the flying-spot scanner for the transmission of transparencies and the image orthicon camera for live pickup. The first has the advantage of the complete absence of registration difficulties and a well-defined (linear) transfer characteristic, the second that of great sensitivity combined with a constant transfer characteristic when operated below the "knee" (Fig. 10.11).

The general arrangement of the flying-spot slide pickup has already been shown in Fig. 17.9; Fig. 19.1 shows a particular physical realization. Dotted lines indicate the position of the flying-spot tube as well as the ray paths from the scanning pattern on the tube face to its image on the transparency and from the transparency by way of two dichroic mirrors to the photocathodes of the three multiplier phototubes generating the color signals.

The dichroic mirrors,* which represent a constituent of all present-day simultaneous color pickup apparatus, are essentially interference filters with color-selective reflection and transmission characteristics. By appropriate adjustment of the layer thicknesses of the interference filter, a wide range of spectral responses may be attained. Figure 19.2

* See Dimmick, reference 3.

shows typical response curves for a particular pair of dichroic mirrors. Camera responses such as those shown in Fig. 18.12 are obtained by the multiplication of the individual spectral responses of the photosensitive camera tube with the ordinates of these curves, eventually modified by the addition of auxiliary filters.

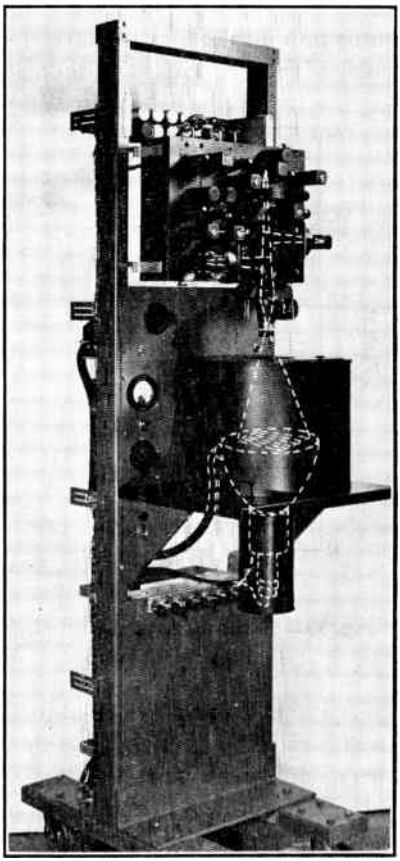


Fig. 19.1. Flying-Spot Slide Scanner.

The lack of registration difficulties in the flying-spot system arises from the fact that the scanning pattern is projected on the original colored picture, separation of the three color signals taking place beyond this point. It is purchased by the relatively low sensitivity characteristic of nonstorage systems—since the red component of the light emission of the flying-spot tube is relatively weak, noise tends to be visible in the red partial picture—and the necessity of compensating for the finite decay period of the light emission of the phosphor. The effect of the gradual decay of the light emission is equivalent, to a first approximation, to an attenuation of the high-frequency components of the picture and may hence be compensated by high-frequency peaking circuits in the camera preamplifiers. Figure 19.3 shows a preamplifier circuit suitable for a flying-spot pickup system as well as the effect of the peaking circuits in restoring the proper signal shape for a square light pulse.

The image orthicon camera is free of this last difficulty, but requires the solution of other problems instead. Here, the green, red, and blue partial pictures of the identical scene must be formed on three separate image orthicon photocathodes. This could be accomplished most conveniently by introducing a pair of dichroic mirrors between the common objective lens and the photocathodes, as shown in Figs. 19.4*a* and

b. Both systems are generally unacceptable, however, since they do not permit the use of the short-focus objectives necessary for covering a large angular field. System *a*, which is preferable from the point of view of noninterference between the magnetic fields from the three focusing coils as well as with regard to the quality of the partial pictures produced, is particularly restricted in this respect. The employment of crossed dichroics, indicated in *b*, considerably shortens the re-

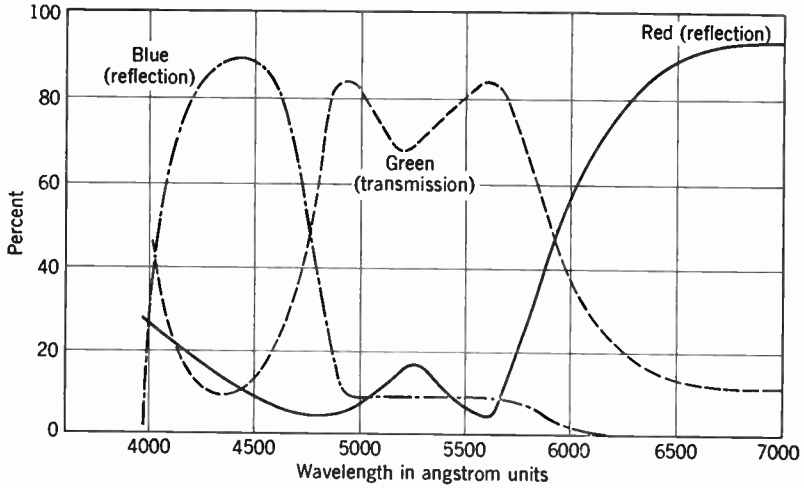


Fig. 19.2. Transmission and Reflection Characteristics of a Pair of Dichroic Mirrors.

quired separation between objective and photocathodes. However, the irregularities introduced into the optical paths by the region where the two mirrors meet mar the image reproduction to some extent. Although this effect, also, can be removed in principle by immersing the dichroics in a medium of the same index as the supporting plates (e.g., by employing four right-angle prisms forming a solid cube), this solution gives rise to other difficulties.

To obtain an adequate field angle, early simultaneous color cameras adopted, therefore, the optical arrangement shown in Fig. 19.5. Three matched objectives, mounted on a common focusing carriage, are placed between the beam-splitting system and the individual image orthicons. Here, also, physical dimensions limit the permissible field angles and lens apertures, but the limits so imposed are less restrictive than for the single-lens systems in Fig. 19.4. Furthermore, with the three-lens system a right-angle arrangement with crossed dichroics leads to less

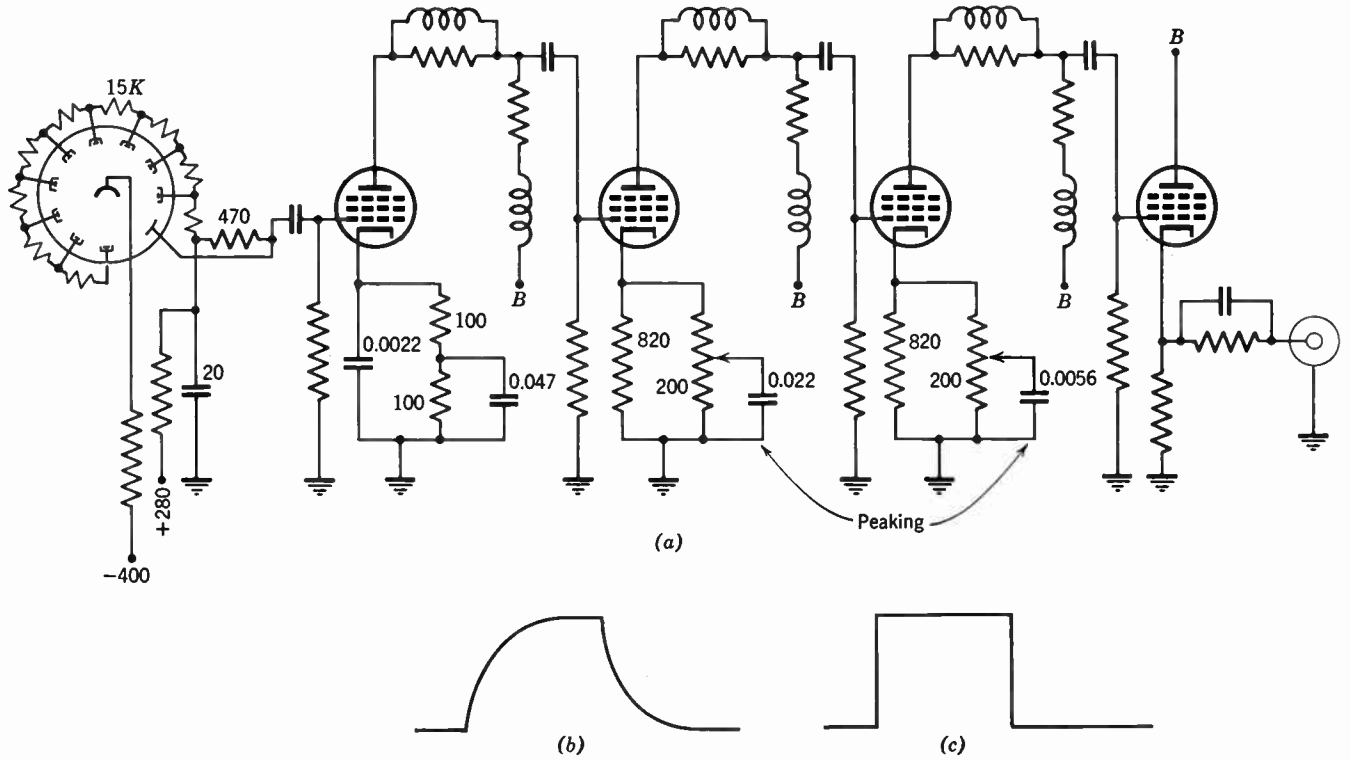


Fig. 19.3. Flying-Spot Scanner Preamplifier with High-Peaking Circuits. (*a*) Circuit. Response for narrow vertical bar (*b*) without and (*c*) with high-peaking correction.

pronounced picture defects resulting from the mirror junction and reduced magnetic interference as compared with the system of Fig. 19.4*b*.

Although excellent results have been obtained with cameras employing the optical arrangement shown in Fig. 19.5, they have not proved

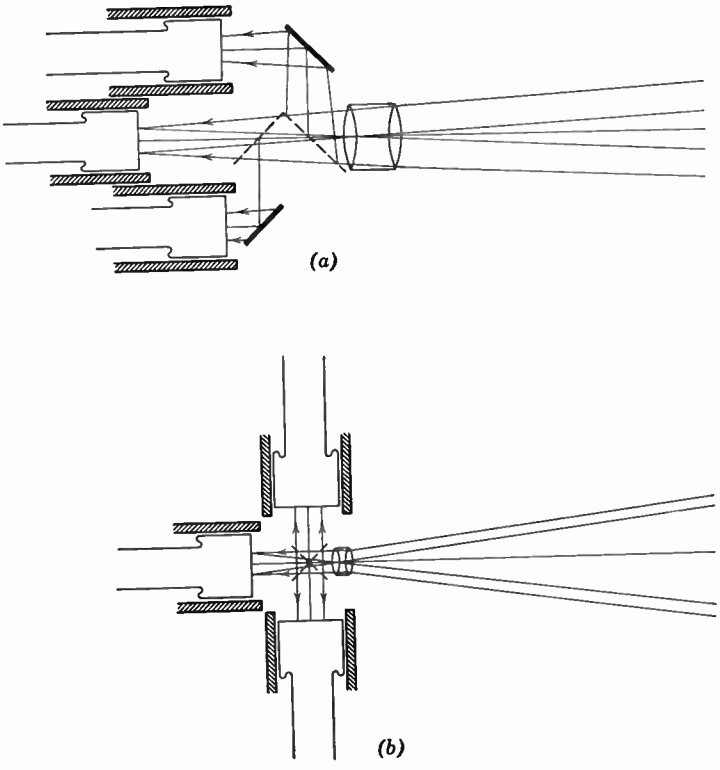


Fig. 19.4. Image Orthicon Cameras with Single Objective and (*a*) Sequentially Arranged or (*b*) Crossed Dichroics.

generally acceptable in view of the difficulty of interchanging lenses. In particular, they exclude the possibility of employing a conventional lens turret.

The employment of a turret becomes possible if a fixed, long-focus relay objective is employed to reimage the picture produced by the objective on the photocathodes of the three image orthicons. This provides ample space for the insertion of the beam-splitting system between the relay objective and the photocathodes.

The optical arrangement of a practical color camera based on this principle is shown in Fig. 19.6. The principal optical elements which are added in this system are the relay objective, G, H, consisting of two 9½-inch focal length $f/4.0$ photographic lenses, and the field lenses C. The objective A images the scene in the plane of the biconvex field

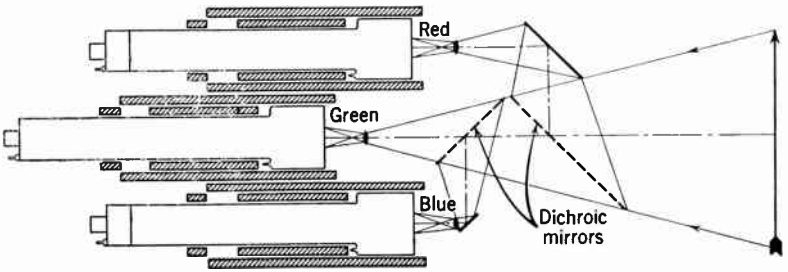


Fig. 19.5. Optics of Image Orthicon Camera with Three Separate Objectives.

lens, which in turn images the exit pupil of the objective A into the entrance pupil of the relay objective G, H. The relay objective, finally, reproduces the original image on the image orthicon photocathodes with unity magnification. Since the field lenses must be matched to the objectives A, they are mounted on a frame rotating with the objective turret.

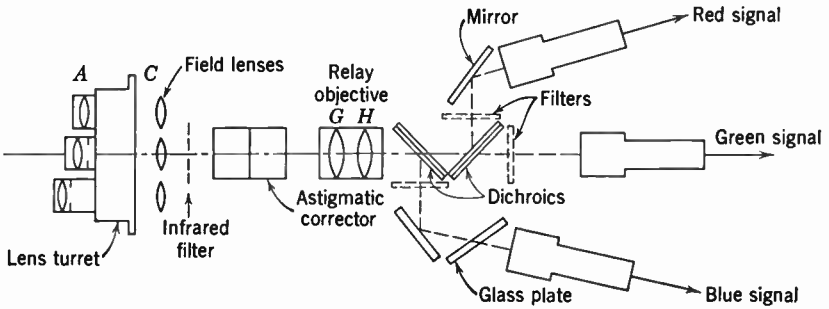


Fig. 19.6. Optics of Image Orthicon Camera with Relay Lens System. (Sachtleben, Parker, Allee, and Kornstein, reference 4.)

The dichroic films are sandwiched between identical glass plates. A glass plate is inserted, furthermore, in the blue light path to equalize the length of optical path through glass for the three images. Furthermore, the astigmatism introduced by passage of the rays through the glass plates at an angle is compensated by the insertion, ahead of the relay objective, of two glass plates equal in thickness to the dichroic

sandwiches. Like the dichroic mirrors, the compensating plates are inclined approximately 45 degrees with respect to the optic axis. However, they are rotated with respect to the axis by 90 degrees, so as to make the glass paths similar for two mutually perpendicular planes through the axis.

Both additional gelatin color filters, to attain the desired spectral responses for the three color channels, and neutral filters are attached

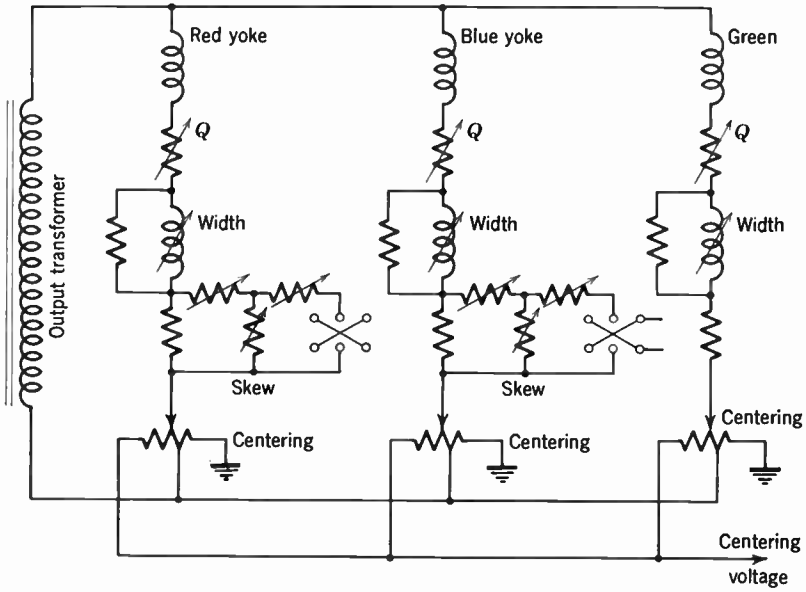


Fig. 19.7. Horizontal Deflection Circuits for Image Orthicon Camera.

to the mount for the dichroic mirrors. Cementing of the filters between low-reflection-coated glass plates minimizes light loss. The neutral filters serve to bring the operating range for all three image orthicons to the same point, just below the knee of the image orthicon characteristic.

The camera circuits are basically similar to those for a black-and-white image orthicon camera, but require additional adjustments for obtaining and maintaining accurate registration between the three partial pictures. Each image orthicon is provided with two pairs of mutually perpendicular alignment coils, which permit electrical adjustment of both the orientation and the strength of the alignment field. Both the three horizontal deflection coils and the vertical deflection coils are connected in parallel and supplied by a single deflection

generator. The horizontal circuits, which contain most of the registration adjustments, are sketched in Fig. 19.7. The inductive width adjustments are coupled with changes in the series resistance Q so as to leave the effective Q of the circuit, and hence the scanning linearity adjustment, unaltered. Vertical deflection voltage, in either polarity, is added to the horizontal voltage by means of the "skew controls," so

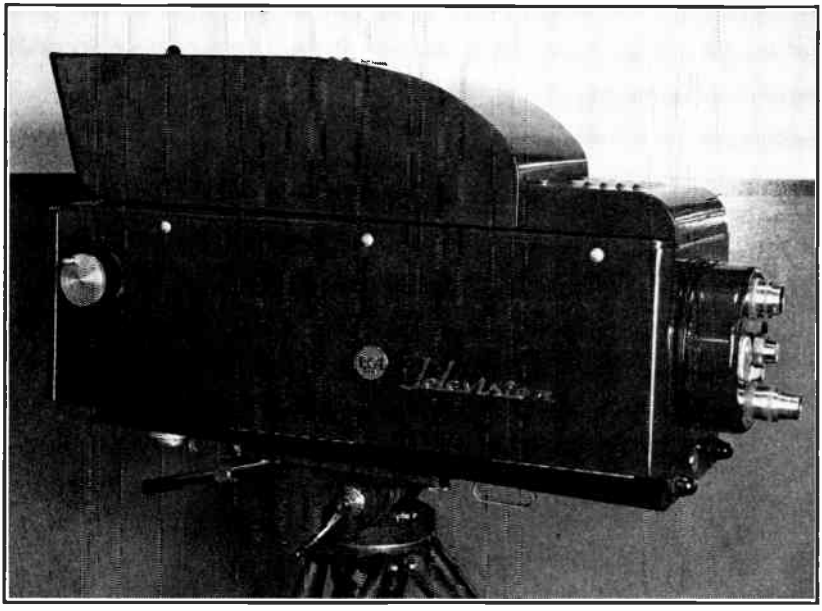


Fig. 19.8. External Appearance of Image Orthicon Camera. (Spradlin, reference 5.)

as to rectify a small degree of skewness of the scanning pattern, which may prevent accurate registration. The individual width controls (and similar height controls in the vertical generator) serve registration purposes only. Overall size is adjusted by varying the sawtooth voltages in the two deflection generators.

The camera contains, in addition to the optical system, the camera tubes, and the deflection circuits, three plug-in six-stage preamplifiers and a cathode-ray tube view finder, on which the operator may observe either the green picture (in black and white) alone or two or three partial pictures superimposed. This enables him not only to frame the scene but also to check optical and electrical focus and registration. In Fig. 19.8 the view finder and its circuits are located under the curved cover on top of the camera.

The preamplified video signals are applied by cable to a fixed video control amplifier. This contains, in triplicate, circuits such as those outlined in block diagram in Fig. 19.9. At the input of the amplifier, horizontal shading signals are added to the video signals. The shading signals are adjustable combinations of sawtooth and parabolic voltage variations designed to compensate variations in sensitivity over the area of the photocathodes of the individual image orthicons; unless

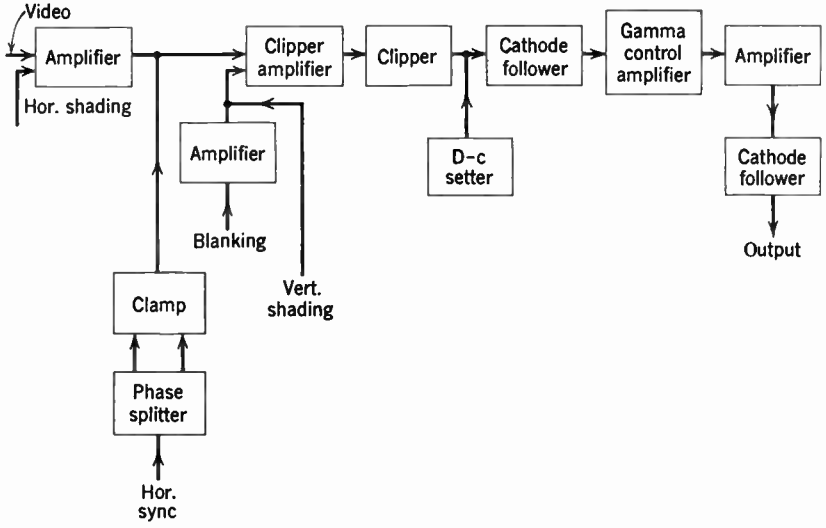


Fig. 19.9. Block Diagram of Video Control Amplifier Circuit for One Color Channel.

corrected in this manner, the variations in sensitivity may give rise to objectionable nonuniformity in the color of the background.

Clamping circuits keyed by the horizontal drive pulses re-establish the black level of signals, which thereupon are mixed with the blanking pulses in a clipper amplifier. The vertical shading signals, applied to the cathode of the clipper amplifier, effectively establish the clipping level. They cannot be introduced ahead of the clamp, since they would otherwise be eliminated by it. The channel gain is adjusted by varying the bias of the input stage and the background by changing the bias of the voltage regulators setting the clipping level.

The gamma control unit, "rooter" circuit, or nonlinear compensating amplifier, is of particular interest. Figure 19.10 shows a circuit employed in connection with the flying-spot scanner. Here, the characteristic of the pickup system is linear, so that the rooter circuit must

convert an applied signal S into $kS^{1/\gamma}$, where γ describes the receiver characteristic. The circuit operates in the following way. As the signal current increases, corresponding to decreasing picture brightness, the diodes in the cathode follower output circuits are made to conduct, reducing the resistance in series with the output resistance $R_2 + R_3$ in

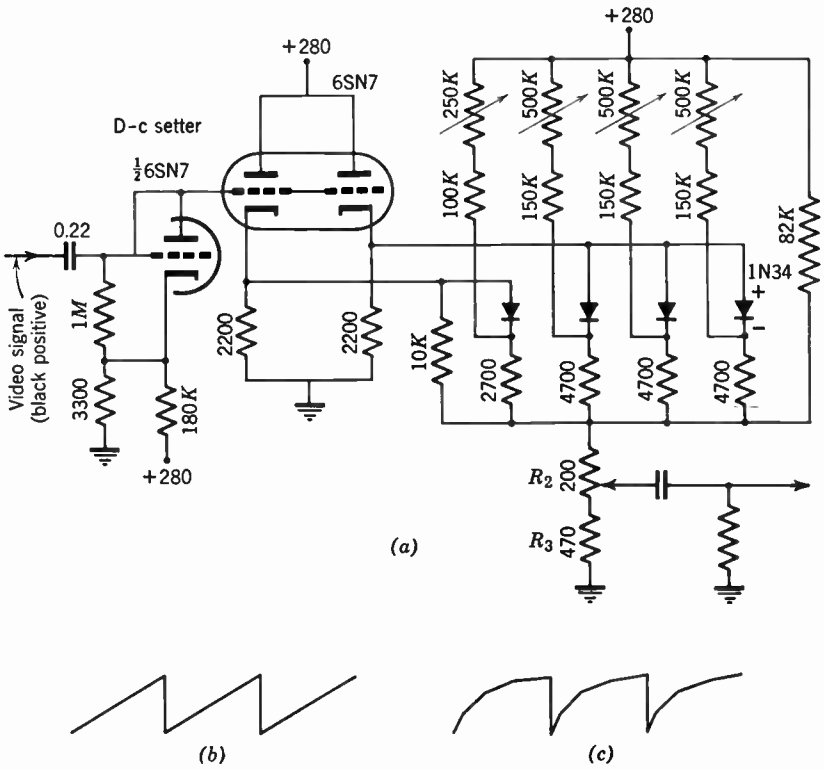


Fig. 19.10. Gamma Control Circuit for Flying-Spot Scanner. (*a*) Circuit. (*b*) Sawtooth wave. (*c*) Output of circuit for sawtooth input.

steps and increasing the rate of change of the output voltage correspondingly. By adjustment of the several resistance values, an arbitrary nonlinear distortion of the input signal may be achieved, as indicated by the action of the network on a sawtooth wave. In the present instance, the variation shown in Fig. 19.10*c* is made the inverse of the kinescope characteristic of the receiver. More generally, the desired transfer characteristic can be approximated by just as many straight-line segments as there are diodes in the circuit. The resistance values in Fig. 19.10*a* correspond to $\gamma = 2.75$. The transfer characteristic of

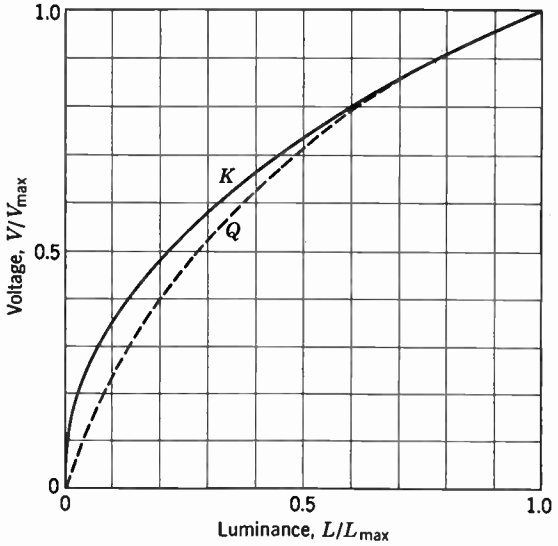


Fig. 19.11. Typical Transfer Characteristics of Image Orthicon Operated Just below the Knee (Q) and of Kinescope with $\gamma = 2.2$ (K). (W. F. Davidson.)

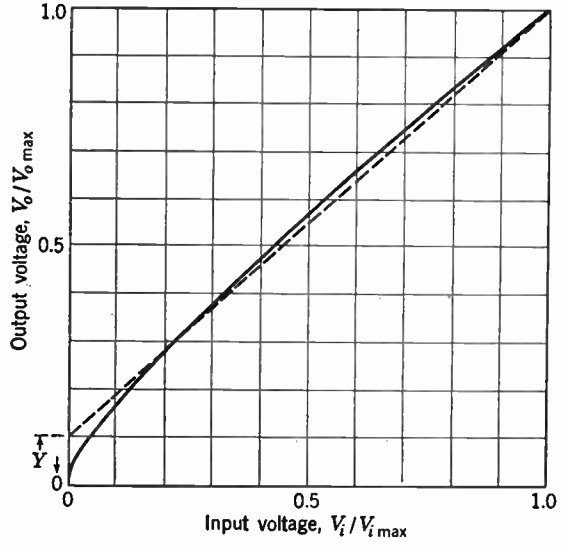


Fig. 19.12. Desired Transfer Characteristic of Gamma Control Circuit and Its Approximation by a Straight Line. (W. F. Davidson.)

tricolor kinescopes is approximated more closely by $\gamma = 2.2$, the value assumed in the color television standards established by the Federal Communications Commission in December, 1953.

Typical characteristics for an image orthicon operated at the knee and just below the knee, as well as the assumed characteristic of the kinescope ($\gamma = 2.2$) modified by the presence of ambient illumination equal to 0.02 of maximum brightness, are shown in Fig. 19.11. The desired transfer characteristic, of the correcting amplifier, is obtained by plotting the ordinates of curve K against the ordinates of curve Q for equal luminance, as shown in Fig. 19.12. The dotted line shows here how the transfer characteristic may be approximated by a linear response with the addition of a constant signal Y. The inaccuracy at very low voltages (corresponding to low brightnesses) is of little importance. The constant signal Y must be added by background adjustment. This is generally accomplished in the receiver rather than in the transmitter, so as to permit the largest possible signal-to-noise ratio. A more accurate rendition of color values is achieved with a nonlinear amplifier such as that in Fig. 19.10, permitting the fitting of the ideal transfer curve by four straight segments.

19.2 Formation of the Chrominance and Monochrome Signals. Figure 19.13 shows, in the form of a block diagram, circuits for the generation of the chrominance and monochrome signals in accord with the standards established by the FCC. As a first step, the (approximate) luminance signal M is formed by the addition of the red, blue, and green camera signals (modified by gamma correction) in proper proportion. A similar process of addition (and subtraction) of the camera signals with the aid of matrix units and phase inverters results in the two signals I and Q specified in section 18.9. These signals, after having passed through low-pass filters appropriate to the "orange-cyan wideband" system, are converted into push-pull signals by phase splitters. Clamps keyed by the synchronizing signals establish the background level of the four signals obtained in this manner. These signals are then applied to balanced modulators, whose 3.58-megacycle carriers are in quadrature. The signals derived from the two balanced modulators and from a third similar modulator keyed by the "burst flag" so as to provide the color synchronization signal are added together. In two further adders, the properly delayed luminance signal and the pedestal are added to the chrominance signal. An output clamp restores the proper d-c level and a cathode follower acts as low-impedance output stage for the unit.

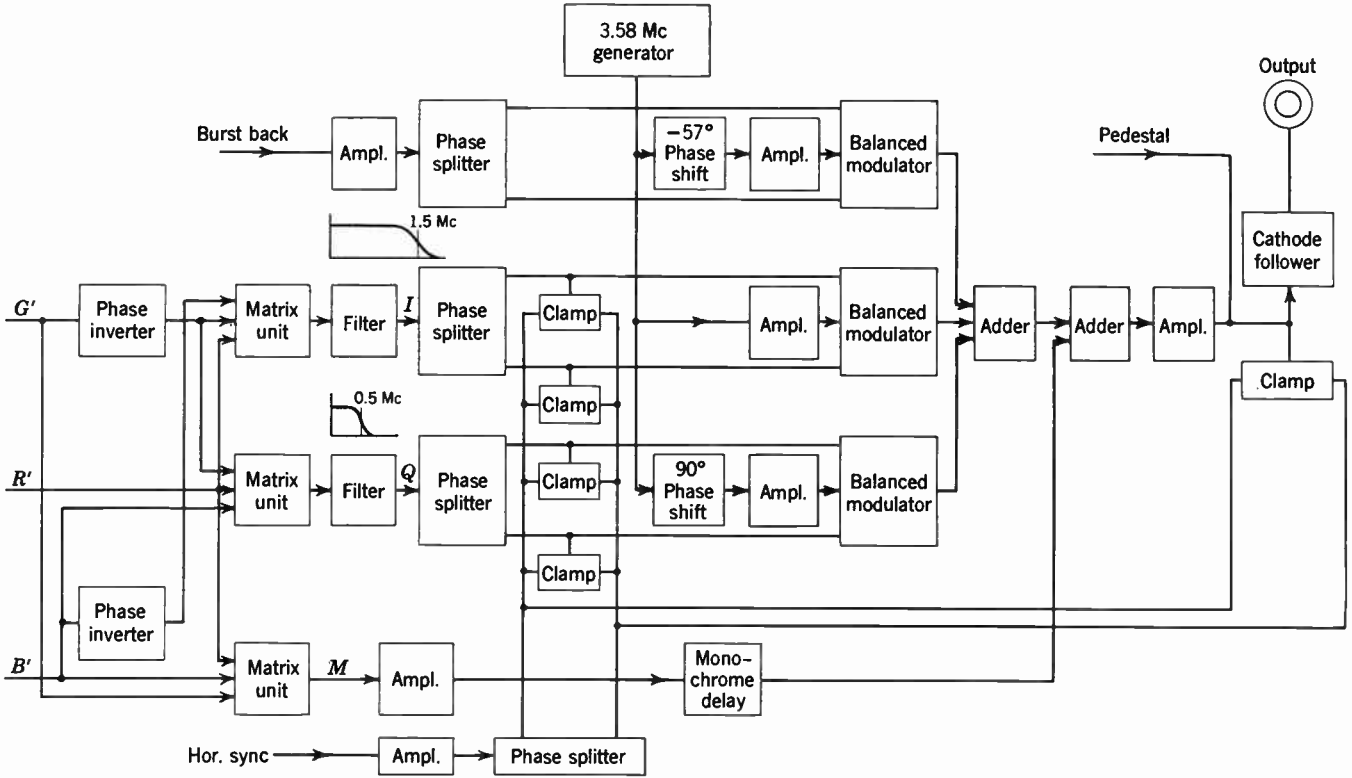


Fig. 19.13. Block Diagram of Circuits for Generation of Chrominance and Monochrome Signals.

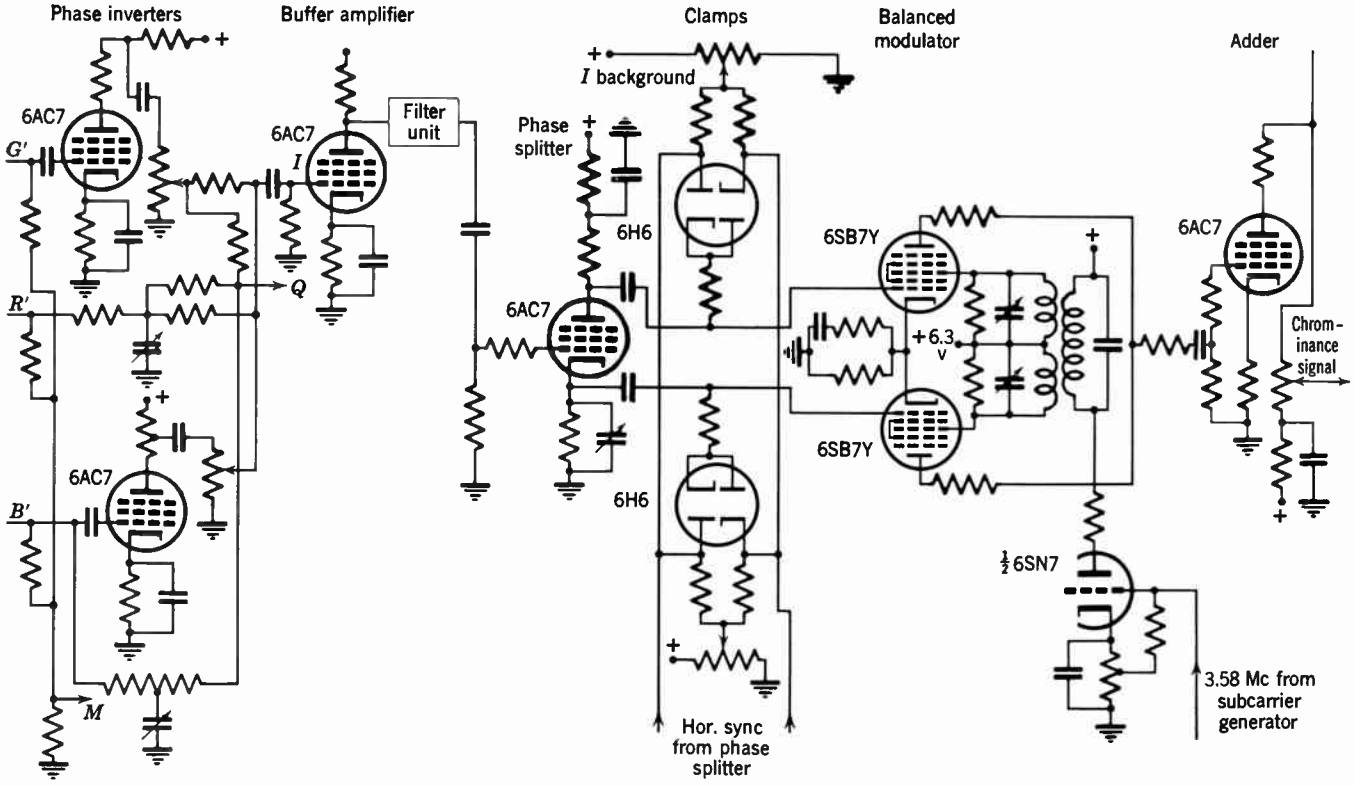


Fig. 19.14. Generation of the Chrominance Signal: *I* Color Channel.

Figure 19.14 shows the actual circuit for the I color channel. Addition of the signals, partly inverted in phase, is carried out by resistance networks. A plug-in filter unit is inserted in the output circuit of a buffer amplifier stage. The balanced modulator employs two pentagrid converter tubes (6SB7Y). Since the balanced modulators generate only the sidebands of the odd harmonics of the carrier, the modulating signal is absent from the output, and no bandpass filters are required for the chrominance signal.

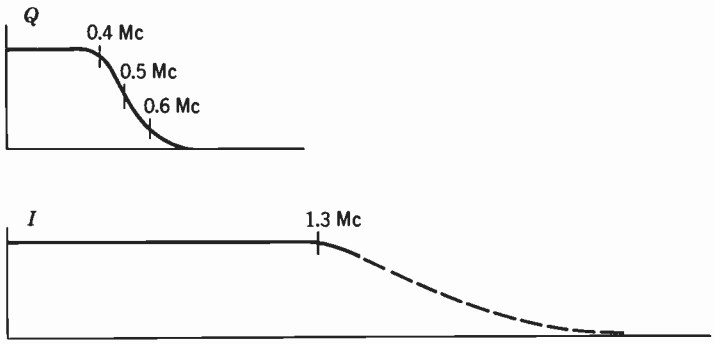


Fig. 19.15. Transmitter Low-Pass Filter Characteristics.

The characteristics of the low-pass filters for the I- and Q-signals are shown in Fig. 19.15. The Q-signal extends over the frequency range for which double-sideband transmission is obtained. The I-signal is flat to approximately 1.3 megacycles. Compensation for single-sideband transmission resulting from the increasing attenuation of the video channel beyond 4.1 megacycles takes place in the receiver. As mentioned in section 18.7, the orange-cyan wideband system here employed to eliminate the color fringing at edges has proved more satisfactory, from the point of view of stability, simplicity of adjustment, economy, and freedom from edge flicker, than color phase alternation (CPA).

19.3 Synchronization. The synchronizing signal employed to keep the color subcarrier in the receiver in step with that in the transmitter is shown in Fig. 19.16. It consists of a "burst" of sampling frequency lasting a fraction of horizontal flyback time (e.g., nine subcarrier periods) which is placed on the back porch of the horizontal synchronizing signal. In view of the fact that the subcarrier frequency is an odd multiple of half the line frequency, the phase of this signal is reversed for successive lines.

An overall block diagram indicating the generation of the subcarrier oscillation and the derivation, from it, of the 31.5-kilocycle master

oscillation for the conventional synchronizing generator, as well as the formation, finally, of the burst pulse is shown in Fig. 19.17.

The source of the subcarrier signal is a crystal ground for 3,579,545 second^{-1} at 75°C mounted in a crystal oven which maintains the

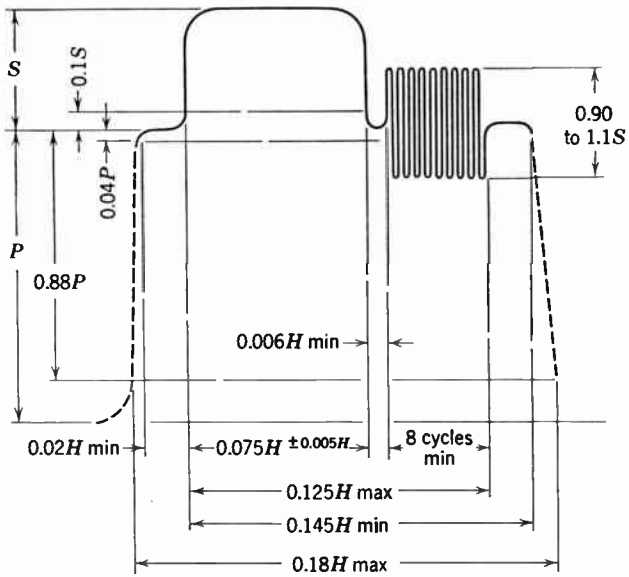


Fig. 19.16. The Color Synchronization Signal. *Notes:* 1. The radiated signal envelope shall correspond to the modulating signal of the above figure, as modified to comply with a television channel width of 6 Mc. 2. The burst frequency shall be the frequency specified for the chrominance subcarrier. The tolerance on the frequency shall be ±0.0003 percent with a maximum rate of change of frequency not to exceed one-tenth cycle per second per second. 3. The horizontal scanning frequency shall be $\frac{2}{455}$ times the burst frequency. 4. Burst follows each horizontal pulse, but is omitted following the equalizing pulses and during the broad vertical pulses. 5. Vertical blanking 0.07 to 0.08V. 6. The dimensions specified for the burst determine the times of starting and stopping the burst, but not its phase. 7. Dimension P represents the peak-to-peak excursion of the luminance signal, but does not include the chrominance signal.

temperature constant within ±2 degrees; this range corresponds to a maximum frequency variation of ±12 second^{-1}. The crystal forms part of a conventional crystal oscillator, whose output is amplified by a buffer amplifier to 20 volts peak-to-peak across a 75-ohm load. The amplified subcarrier signal may be applied by way of distribution amplifiers to the color and synchronizing modulators. It is also applied to a sequence of three frequency divider stages effecting a change in

frequency by the factors $\frac{1}{5}$, $\frac{1}{7}$, and $\frac{1}{13}$ and terminating in an oscillation with a frequency of 31,468.528 second^{-1}, which serves as master oscillation for the conventional synchronizing generator which follows it. In addition, the gating pulse ("burst flag pulse") for the burst is derived from the horizontal drive pulses in conjunction with a delay multivibrator; the vertical drive pulses are utilized to suppress the burst signal during the portion of vertical return time devoted to the

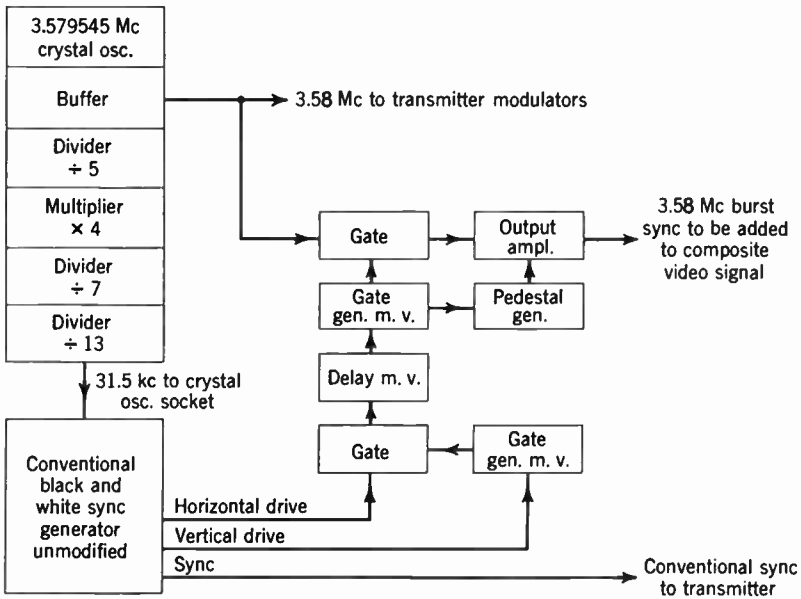


Fig. 19.17. Block Diagram of Burst Pulse Generator.

transmission of equalizing pulses. This prevents interference of the color synchronizing signal with the proper operation of the vertical synchronizing pulses as well as with the proper setting of the d-c level in the receiver.

A preferred form of the frequency divider is the regenerative divider illustrated in Fig. 19.18. It is here assumed that the desired frequency multiplication factor is 1/5. The unit consists of a regeneratively coupled mixer and multiplier stage. The output of the mixer is tuned to one-fifth the input frequency, that of the multiplier to four times the output frequency of the mixer. As the input oscillation is applied, the output circuit of the mixer is excited by it. The resulting low-frequency oscillation is applied to the grid of the multiplier which, thus, generates an oscillation with four-fifths the input frequency. In the

mixer this oscillation regenerates, by interaction with the input, the low-frequency oscillation of the mixer output, completing the cycle. The loop gain is generally chosen just low enough to suppress free oscillations.

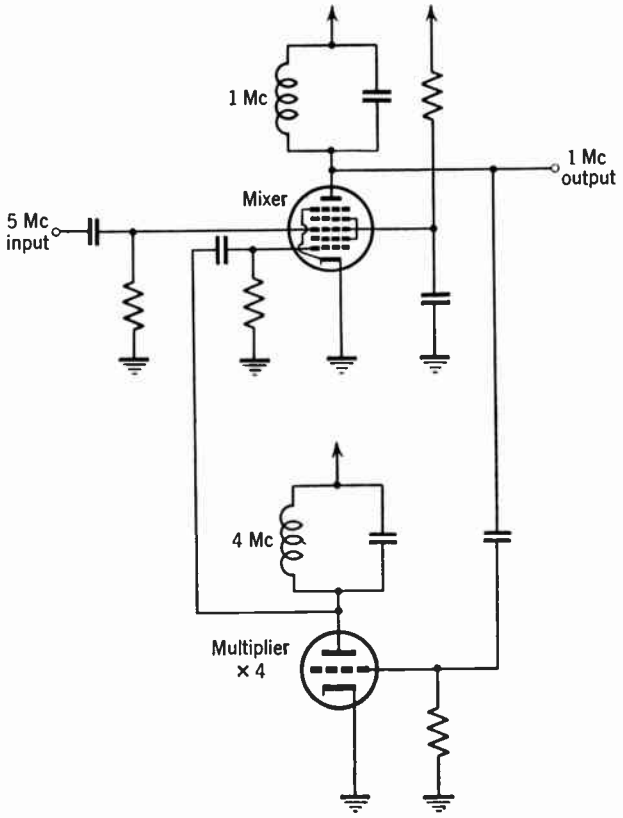


Fig. 19.18. Regenerative Frequency Divider. (T. O. Stanley, N. D. Larky, and D. D. Holmes.)

In the receiver, the phase of the bursts may be compared with that of a local subcarrier oscillation and the resulting error signal employed to modify the frequency of the local oscillation in such a direction as to achieve synchronism. Figure 19.19 shows how this is accomplished in detail. A gating pulse is derived from the horizontal synchronizing signals which coincides with the duration of the bursts. It selects the bursts from the composite video signal in a gated amplifier, whereupon they are applied to a phase detector along with the output of the sub-

carrier oscillator. An error signal corresponding to the phase difference is integrated and modifies the reactance presented by a reactance tube in the resonant circuit of the subcarrier oscillator.

The several components of this circuit admit of a wide variety of realizations. Figure 19.20 shows a particular example which has proved satisfactory in operation. Proceeding from the left to the right, the horizontal synchronizing pulses trigger a one-shot multivibrator which generates gating pulses of the desired width. They are applied, along with the composite video signal, to the grids of a two-stage gated

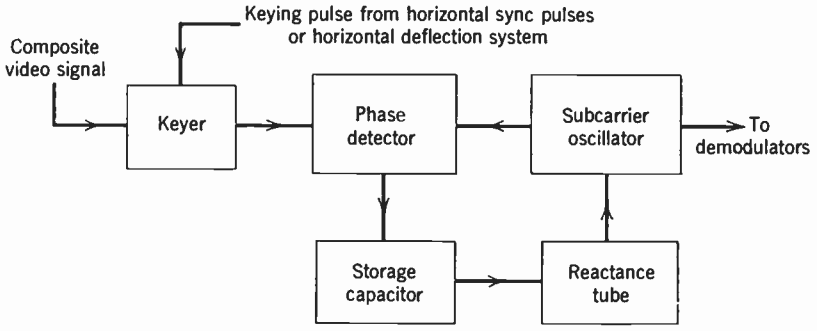


Fig. 19.19. Block Diagram of Color Synchronization Circuit in Receiver.

amplifier. The d-c level of the gating pulses is maintained by d-c coupling. The inductance L_N and the coupling capacitance of 2.7 micromicrofarads serve to neutralize the two stages.

The push-pull output of the gated amplifier, consisting of the synchronizing bursts, is applied to one anode and cathode of the phase-detector double diode, whereas the local oscillation is applied to the remaining anode and cathode in parallel. When the two oscillations are out of phase, current is fed into the storage capacitor, changing the grid bias of the reactance tube; the time constant of the storage circuit is such that integration of the error signal takes place over a period of four or five lines.

The bias produced across the storage capacitance determines the amount of quadrature signal generated across the inductance L_R which is added by the reactance tube section to the output of the oscillator or, in brief, the shift in resonant frequency from that determined by the output tank circuit. The shift is, of course, of such polarity as to correct the existing phase difference between the burst and the local subcarrier signal.

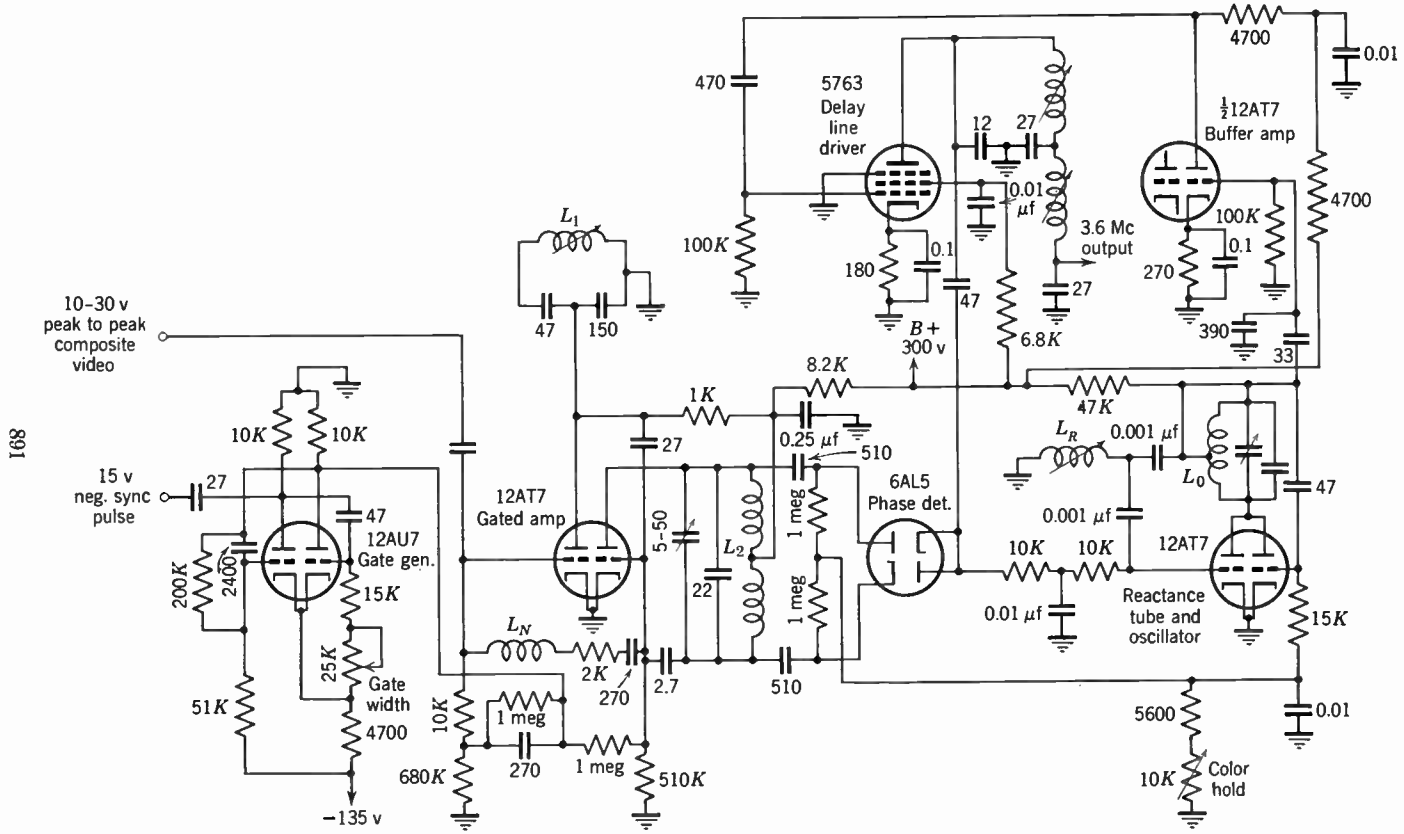


Fig. 19.20. AFC Color Synchronization Circuit. (E. O. Keizer.)

The synchronizing bursts may be employed in several other ways to assure a local subcarrier oscillation in phase with the transmitter subcarrier. A relatively straightforward procedure is to excite a crystal or high-Q tank circuit tuned to the subcarrier frequency by the bursts. The oscillation is continued, with damping which may be minimized by the addition of sharply tuned amplifier stages, through the interval between bursts.

A particular "crystal ringing" circuit for generating the local subcarrier is shown in Fig. 19.21. A small trimmer capacitance in series

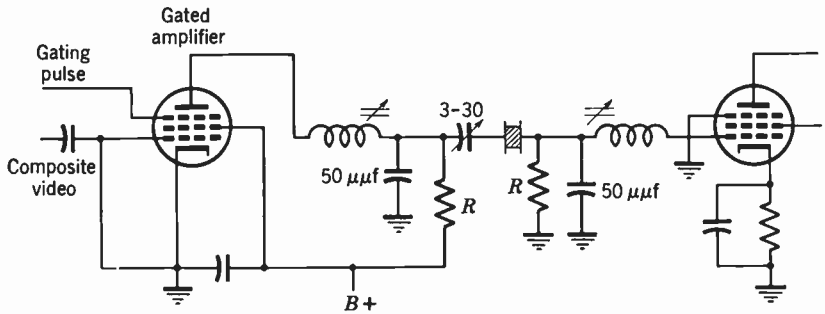


Fig. 19.21. Crystal Ringing Circuit for Color Synchronization. (J. Avins.)

with the crystal permits adjustment of the resonant frequency. The filter half sections on either side serve as impedance transformers; it is desirable to have the crystal work into relatively low impedances—of the order of 100 to 1000 ohms—so as to emphasize its "series mode" of oscillation. Within this range, the larger values of resistance are to be preferred insofar as they lead to smaller phase errors if the crystal circuit is detuned. With a load resistance of 1000 ohms, a frequency difference of ± 80 cycles leads to phase errors of ± 5 degrees. Crystal ringing circuits must be carefully adjusted to maintain the proper subcarrier phase.

The color synchronization circuit shown in Fig. 19.22 is in many respects similar to the automatic-frequency control circuit first described. It is applicable when the desired phase difference in the subcarrier oscillation applied, e.g., to the blue demodulator and the subcarrier oscillation transmitted in the bursts is 90 degrees. This circumstance is utilized to eliminate the subcarrier oscillation from the color hold circuits. Demodulation converts the burst in the blue demodulator to a constant signal which, for the correct 90-degree phase difference, is equal to the d-c level of the blue $(B - M)$ color difference

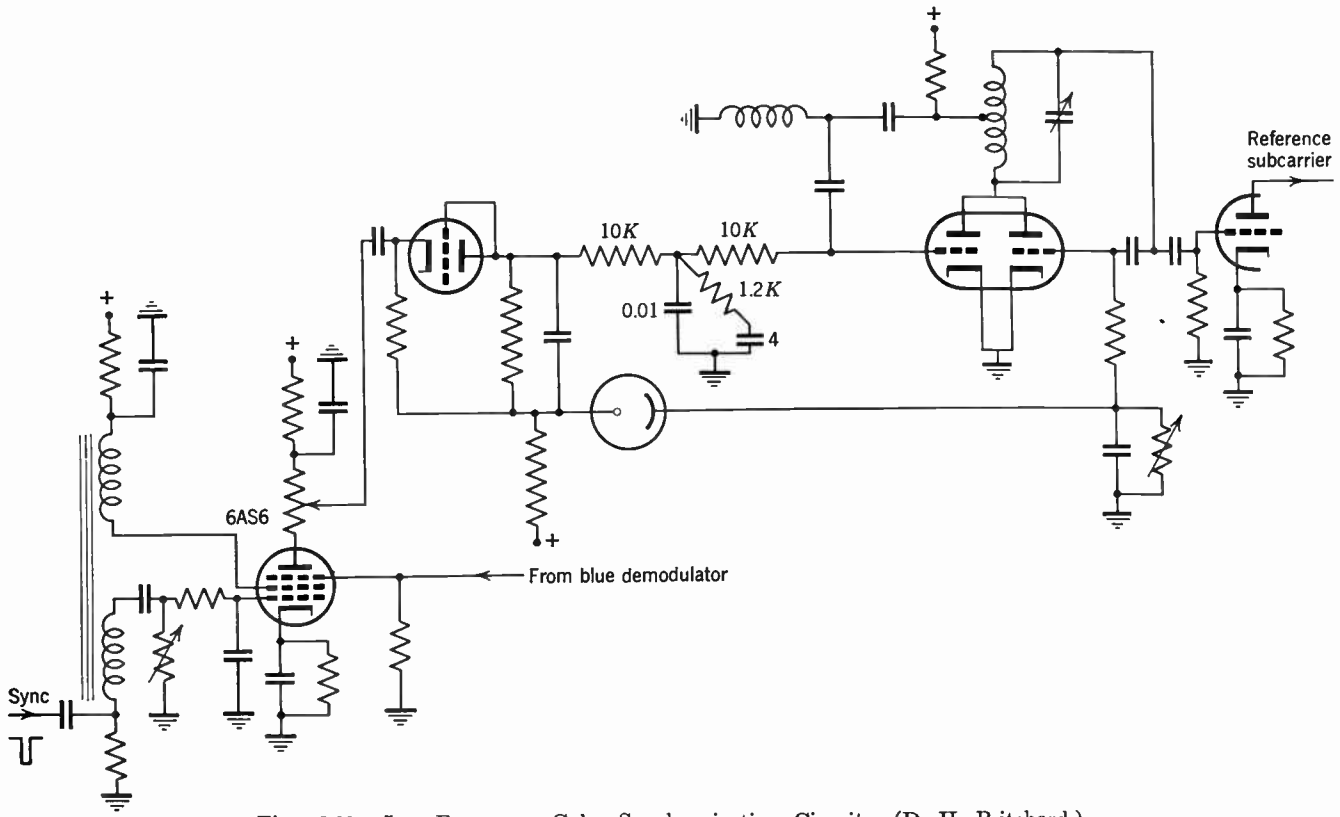


Fig. 19.22. Low-Frequency Color Synchronization Circuit. (D. H. Pritchard.)

signal. A deviation in phase results in a change in the height of this signal which varies linearly with the phase deviation. The circuit in Fig. 19.22 employs a pentode to act both as blocking oscillator to produce a gating pulse for the burst duration and as gated amplifier for the blue signal. The amplitude of the negative pulses obtained in this manner measures the phase deviation; they are rectified by a diode and integrated over several line intervals by the storage capacitance which establishes the grid voltage of the reactance tube. A neon regulator tube holds the d-c bias of the reactance tube constant.

19.4 Recovery of the Color Signals. The broad principles of the recovery of the color signals from the composite video signal were discussed in Chapter 18. This section will confine discussion to a prac-

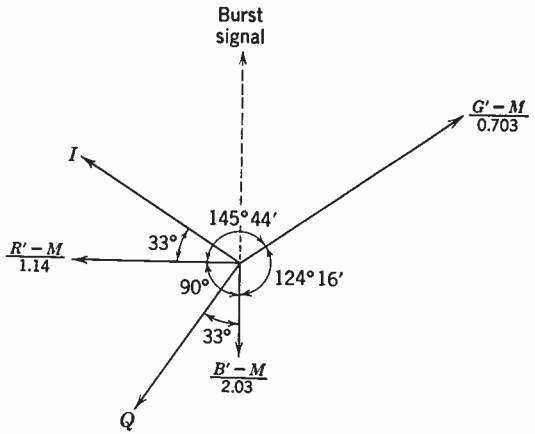


Fig. 19.23. Demodulator Phase Angles and Channel Amplifications in Orange-Cyan Wideband System (FCC Standards).

tical receiver designed for the standards approved by the FCC. As a matter of recapitulation, Fig. 19.23 shows the required demodulator phase angles for the I- and Q-signals and the relationship between them and the three color difference signals. They determine the matrix units effecting the conversion $I, Q \rightarrow B' - M, R' - M, G' - M$. Only two demodulators in quadrature are required. The receiver circuits may take the form shown in block diagram in Fig. 19.24.

Only minor changes are required from a monochrome receiver in the stages ahead of the video amplifier; these are primarily concerned with eliminating irregularities in the phase response in the neighborhood of the subcarrier frequency and making the amplitude response flat to a

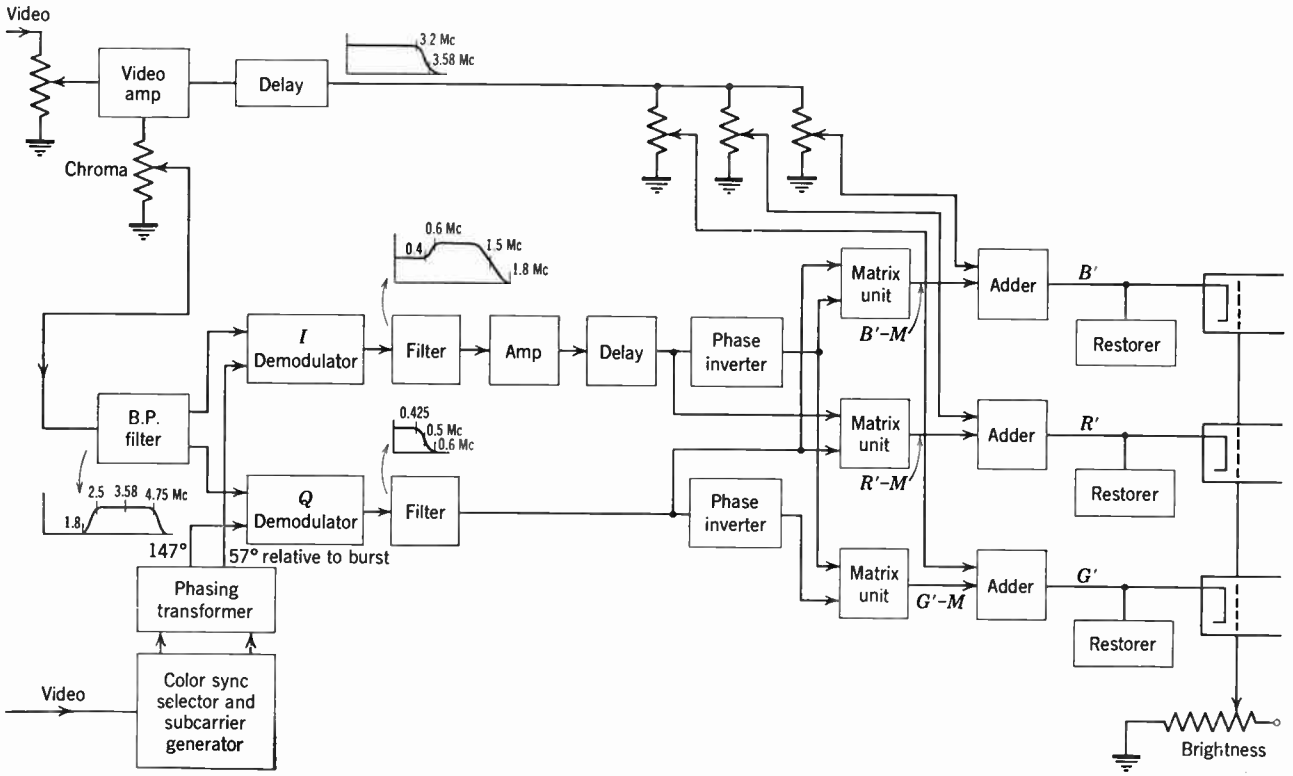


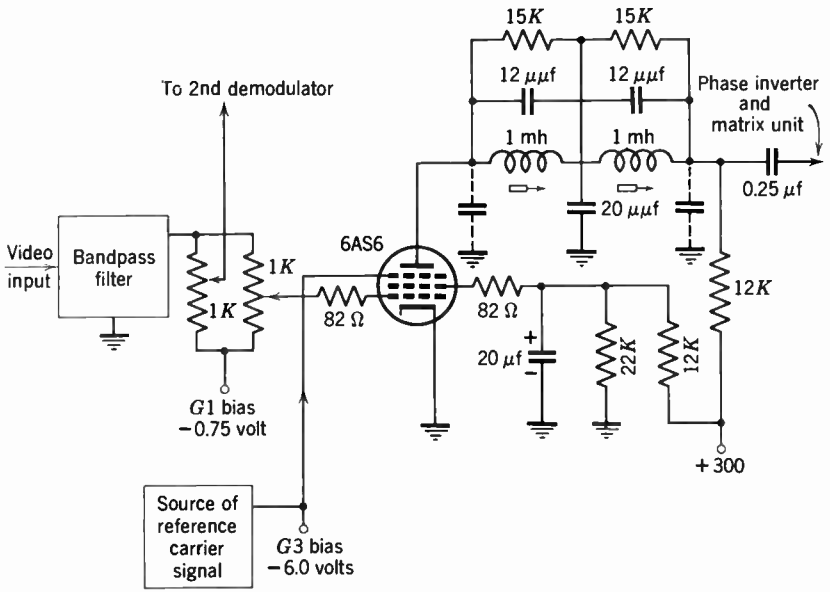
Fig. 19.24. Block Diagram of Circuits for the Recovery of the Color Signals. (D. H. Pritchard.)

frequency of 4.0 to 4.2 megacycles above the video carrier. At the video amplifier, however, the composite video signal is distributed to two channels. In the monochrome channel, the signal is strongly attenuated at the subcarrier frequency to reduce the visibility of dot patterns and passes through a delay line to equalize the total delay experienced by the monochrome signal and the color difference signals, to which it is subsequently added to produce the color signals. The chrominance signal, on the other hand, is selected by a bandpass filter and generates the signals I and Q by acting upon a pair of unbalanced demodulators followed by low-pass filters with appropriate transmission characteristics. A subcarrier generator controlled by the burst signals supplies the carrier to both demodulators, by way of phasing transformers. The color difference signals are obtained from I and Q by means of matrix units and the color signals themselves, finally, by the addition of the monochrome signal. Three d-c restorers serve to re-establish black level before the color signals are applied to the reproducer.

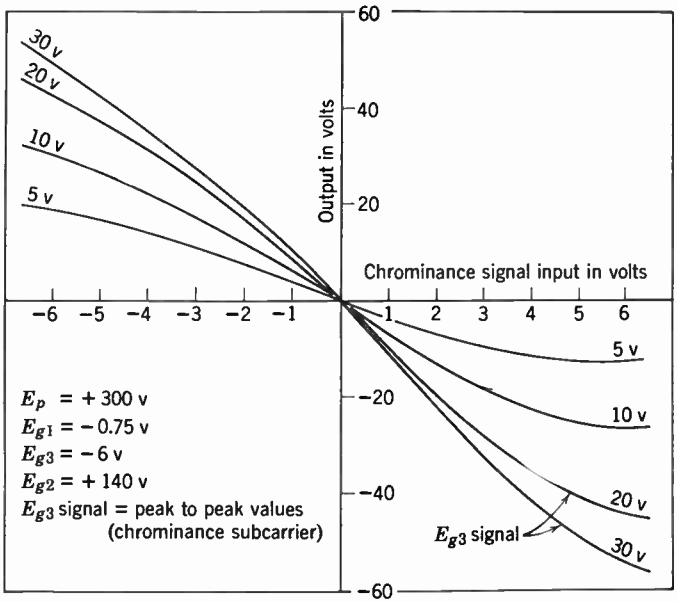
The recovery of the I- and Q-signals from the chrominance signal can be carried out in a number of ways. Balanced and unbalanced demodulator circuits, employing conversion elements ranging from diodes to pentagrid converters, may be employed. Unbalanced demodulators are generally preferred. Although the balanced demodulator renders the bandpass filter in the receiver superfluous, this saving is outbalanced by the added sources of cross-talk resulting from the application of the complete video signal to the demodulator and increased shielding and adjustment difficulties.

The single-ended circuit shown in Fig. 19.25*a* is advantageous from the point of view of economy, ease of adjustment, and compactness. The miniature 6AS6 pentode, developed as a gate tube, requires a relatively low level of reference carrier to control the beam current by means of the suppressor grid; at the same time, it has a high output signal level, good conversion gain, and a favorable ratio of conversion gain to amplifier gain.

The bandpass filter shown in Fig. 19.25*a* is a low-impedance (500-ohm) constant-k filter, flat from 2.5 to 4.75 megacycles; the low impedance minimizes interaction of the two demodulator circuits. At the output relatively high-impedance filters, differing in their passbands for the I- and Q-signals as indicated in Fig. 19.24, are employed to provide high signal level. The 82-ohm resistances in the control and screen-grid leads serve to suppress parasitic oscillations. Considerable



(a)



(b)

Fig. 19.25. Demodulator Circuit. (a) Circuit. (b) Demodulation characteristics. (Pritchard and Rhodes, reference 6.)

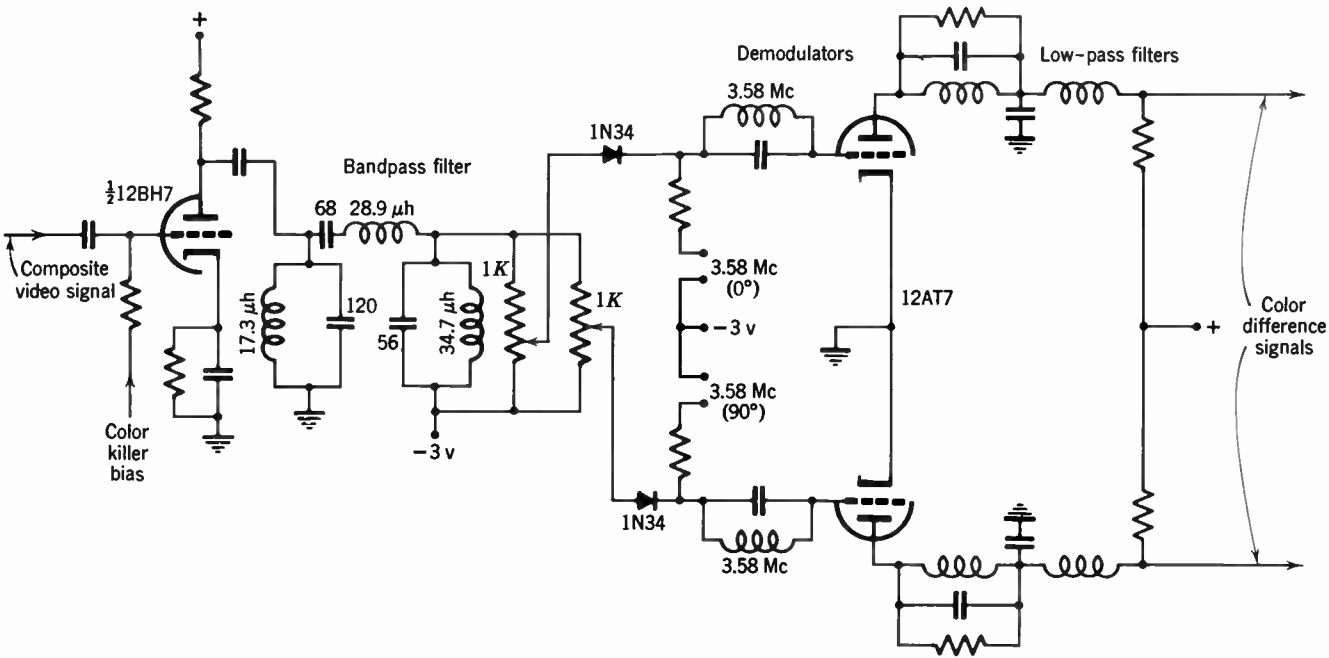


Fig. 19.26. Diode Demodulator Circuits for the Recovery of the Color Difference Signals. (D. H. Pritchard.)

care must be taken in the mutual shielding of the two demodulators to prevent the appearance of reference carrier oscillations in the wrong phase in the demodulator. The demodulator characteristics, for suitable grid-bias values, are shown in Fig. 19.25*b*. The curve parameter is the peak-to-peak voltage of the reference carrier applied to the suppressor grid. It is seen that, for 20 volts peak-to-peak reference carrier, a substantially linear response covering a range of ± 25 volts is obtained for an input signal range of ± 3 volts.

A diode demodulator circuit is shown in Fig. 19.26. Here, the composite signal is applied to a triode amplifier, which is blocked by the "color killer bias" whenever the set is tuned to a monochrome transmission. This is advisable in sets with AFC-controlled local subcarrier, as shown in Figs. 19.20 and 19.22, to prevent interference patterns resulting from the presence of the drifting subcarrier. The color killer bias is derived from the gated burst pulse amplifier and becomes negative in the absence of the burst.

Beyond the amplifier, the signal passes through the bandpass filter for the elimination of the low-frequency brightness components and is applied to two diode demodulators followed by triode amplifiers. The two quadrature components of the subcarrier oscillation are introduced at four or five times the amplitude of the chrominance signal so as to suppress harmonics of the color-difference signals in the output of the demodulator. Parallel-resonant circuits block the subcarrier oscillation from the amplifier grid. Low-pass filters select the low-frequency color difference signals (e.g., the signals I and Q) from the output of the amplifier.

The composite signal, consisting primarily of the monochrome signal M, is also applied to an artificial delay line consisting of low-pass filter sections (Fig. 19.27) with a characteristic impedance of 1000 ohms. Its output is distributed to one side of three double triodes functioning as adders and the three-color difference signals are applied to the other side. Addition takes place in the cathode resistance. The outputs of the double triodes are coupled through a frequency-compensating network to the reproducer cathodes; the d-c level is set by the low-frequency components of the synchronizing signals with the aid of a diode in each channel.

It should be repeated that the circuits described above must be regarded simply as examples of practical circuits. The functions which they must fulfill can be discharged in various ways whose adoption is governed by economic considerations.

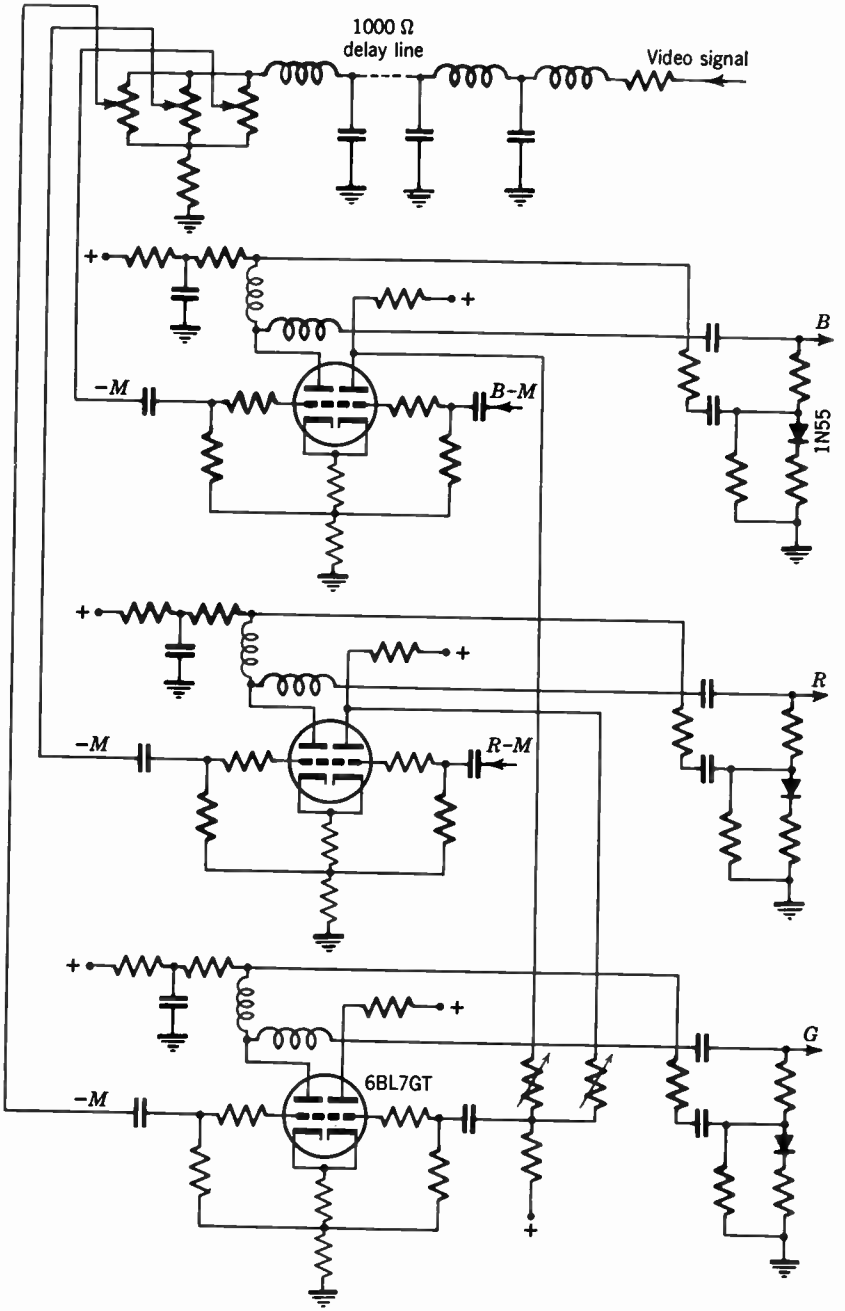


Fig. 19.27. Addition Circuits for the Three Color Channels. (D. H. Pritchard.)

19.5 Reproduction of the Picture. The picture may be reproduced on three separate kinescopes with green, red, and blue phosphors and viewed through dichroic mirrors (Fig. 17.7); it may be projected in superposition on a screen from three such kinescopes (Fig. 17.10); or, finally, it may be formed directly on the screen of a tricolor kinescope. The first form is employed today only in color monitors; such a "trinescope" monitor (Fig. 19.28) usually consists of three kinescopes symmetrically placed about a crossed dichroic block. The projection system, with three separate reflective projection units, finds application in

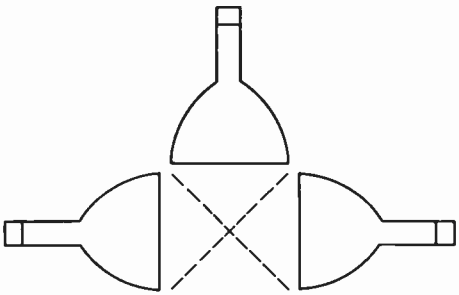


Fig. 19.28. Color Monitor with Crossed Dichroics.

theater television. For home television, the tricolor kinescope appears to present the most practical solution.

In Chapter 17 a variety of types of color kinescopes were described. At the time of writing, the three-gun shadow-mask tricolor kinescope had been produced in considerable numbers. It will hence form the subject of the present, more detailed discussion. Figure 19.29 shows a diagram of the complete tricolor kinescope, as well as the relationship between the aperture mask and the phosphor dot screen.

In the shadow-mask tricolor kinescope,* three electron beams, modulated by the green, red, and blue color signals respectively, scan a shadow mask placed in front of the screen in such fashion that, at any instance, they converge upon the same point on the mask. The mask is a thin sheet of supernickel (70 percent copper, 30 percent nickel) alloy perforated in a regular hexagonal array of some 195,000 apertures for a 15-inch tube. If a line is extended from the center of beam deflection through the center of any one mask aperture to the screen, a red, a green, and a blue phosphor dot are grouped symmetrically about the point of incidence of the line. The dot centers lie on lines drawn

* See Law, reference 7.

from the three beam centers in the plane of deflection through the center of the mask aperture and the spacing of the mask apertures, the distance between mask and screen, and the convergence angle of the three beams are so related that the dot spacing on the screen is

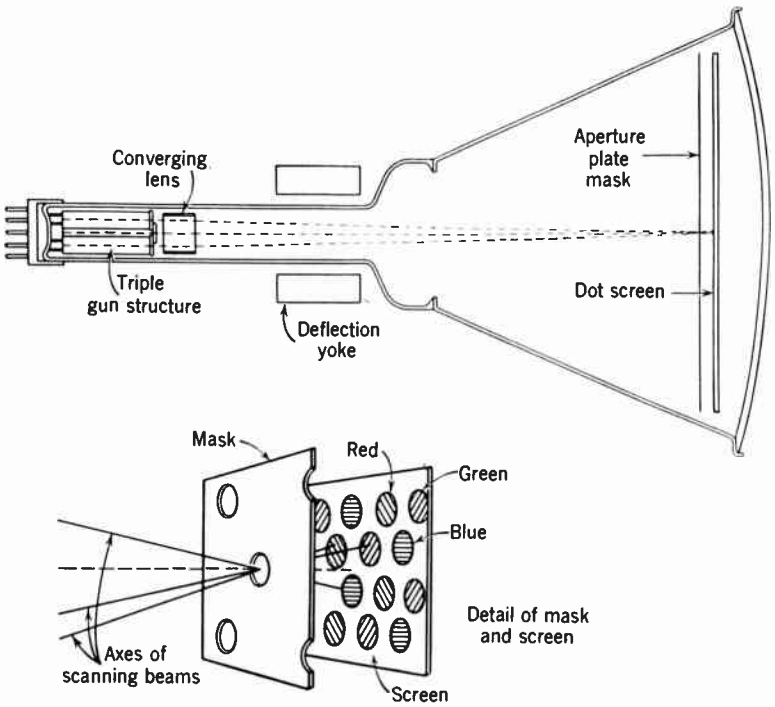


Fig. 19.29. Sectional Diagram of Three-Gun Shadow-Mask Tricolor Kinescope.

uniform throughout. Characteristic values for a 15-inch tricolor kinescope are

Distance between deflection plane and phosphor screen	14½ in.
Distance from mask to screen	⅜ in.
Convergence angle	1°14′
Distance of beam centers from axis in deflection plane	0.3 in.
Spacing between aperture centers, a	0.023 in.
Diameter of mask apertures, B	0.009 in.

The mask efficiency, or the fraction of electron beam current transmitted by the mask, is equal to the ratio of the aperture area to the total area, or $\pi B^2/(2\sqrt{3}\,a^2) = 0.12$ for the numerical values given above. As indicated in Fig. 19.30, the maximum permissible ratio B/a,

which determines the mask efficiency, depends on the ratio of the beam diameter M in the deflection plane to the distance S of the beam center from the axis of the tube; B must be chosen small enough so that

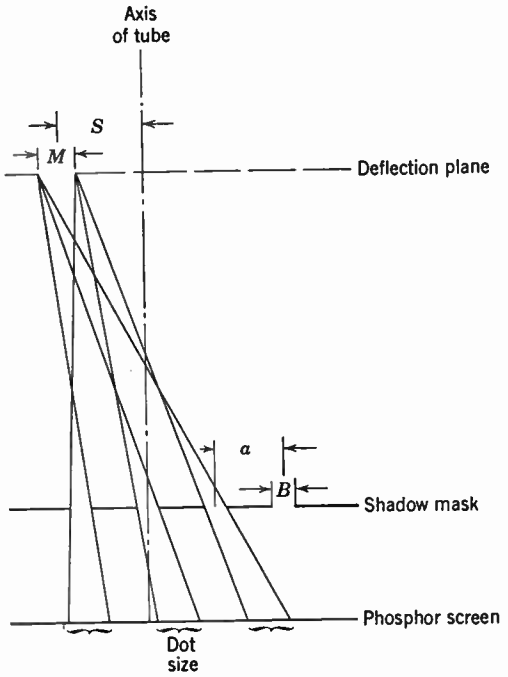


Fig. 19.30. Geometrical Relationship between Aperture Size B, Beam Diameter M, and Beam Spacing S. (H. B. Law, reference 7.)

there is no overlap between the electron spots produced on the screen by the three beams. This leads to a maximum value

$$B = \frac{a}{3}\left(\sqrt{3} - \frac{M}{S}\right) \tag{19.1}$$

corresponding to a current reaching the screen equal to

$$I_S = I_M \frac{\pi}{18\sqrt{3}}\left(\sqrt{3} - \frac{M}{S}\right)^2 \tag{19.2}$$

where I_M is the beam current. I_M, also, depends on the magnitude of the beam diameter M in the deflection plane. Assume that the current density which can be attained at the center of the beam in the deflec-

tion plane is ρ_0 and, furthermore, that the current distribution follows an error curve, the density falling to a fraction $1/\epsilon$ at a distance b from the axis. Then

$$I_M = \pi\rho_0 b^2(1 - \epsilon^{-M^2/(4b^2)}) \tag{19.3}$$

is the beam current confined within a diameter M in the deflection plane, selected from the total beam by a suitably placed aperture. With Eq. 19.3 substituted in Eq. 19.2, it is seen that, for any given ratio S/b, there is a particular ratio S/M which yields a maximum current I_S reaching the screen. For a wide beam ($b \to \infty$), the optimum ratio S/M is seen to be $2/\sqrt{3}$, corresponding to a mask efficiency of 7.5 percent.

More generally the optimum ratio S/M, as function of S/b, is obtained by differentiating Eq. 19.2 with respect to M:

$$\frac{S}{M} = \frac{1}{\sqrt{3}}\left[1 + \frac{4b^2}{M^2}(\epsilon^{M^2/(4b^2)} - 1)\right] \tag{19.4}$$

Finally, if S/b is substituted (in terms of S/M) from this relation in Eq. 19.2, the optimum current utilization $I_S/I_0 (I_0 = \pi\rho_0 b^2)$ is obtained as function of S/M. The optimum beam current utilization, the corresponding mask efficiency, and the relation between S/b and $(S/M)_{opt}$ are all plotted in Fig. 19.31. In practice S and b will be given by the gun design. The optimum value of M to be used as well as the best efficiency that can be obtained with the system and the fraction of the current intercepted by the mask can then be read off from the curves.

It is seen that, from the point of view of obtaining large screen currents, it is desirable to employ large spacings S between the guns and the axis. On the other hand, the employment of such large spacings and correspondingly large convergence angles greatly increases the difficulty of obtaining perfect convergence of the three beams throughout the field. The compromise employed places the convergence angle in the neighborhood of 1 degree, with corresponding mask efficiencies of the order of 10 percent.

After these general preliminary considerations the shadow-mask tricolor kinescope will be described in somewhat greater detail.

Figure 19.32 shows the gun structure of the tube.* It consists essentially of three parallel cylindrical guns comprising individual cathodes, control grids, accelerating electrodes, and focusing electrodes as well as "converging electrodes" issuing into a common large diameter cup.

*See Moodey and Van Ormer, reference 8.

Leads to the cathodes, control electrodes, and accelerating electrodes are brought out separately to permit independent modulation of the three beams and individual adjustment of the transfer characteristics (gamma) over the working range. Focusing of the beams on the screen is brought about primarily by the three lenses formed between the

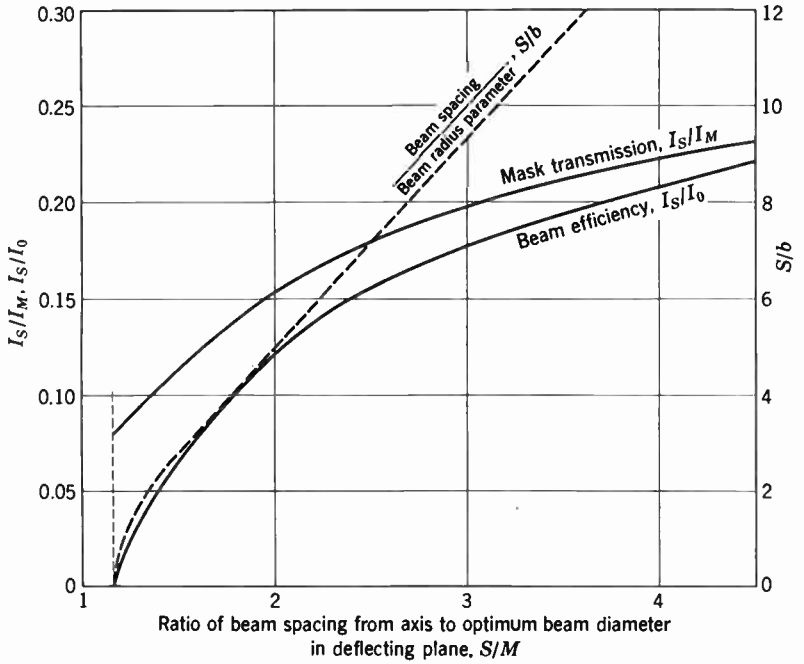


Fig. 19.31. Optimum Beam Current Utilization I_S/I_0 and Corresponding Mask Efficiency Obtainable for a Given Beam Spacing S and Beam Current Distribution Constant b. (H. B. Law, reference 7.)

focusing electrodes and the convergence electrodes. Convergence between the three beams is effected by the lens formed between the common convergence cylinder and the conducting coating on the tube neck. The astigmatism of this last lens, which will cause overconvergence near the edges of the picture, can be compensated by modulating the voltage on the convergence electrode in synchronism with the deflection (dynamic focusing). Slight mechanical misalignments of the individual guns are compensated by the displacement of three small permanent magnets outside the tube neck.

Several variations of the above design have been employed successfully. In one of these the convergence lens is omitted and convergence

is achieved by giving the three guns the proper degree of mechanical tilt. In this instance, dynamic convergence is attained by an added magnetic lens.

In another modification, designed for obtaining very small convergence angles, single-grid, accelerating electrode, and focusing electrode

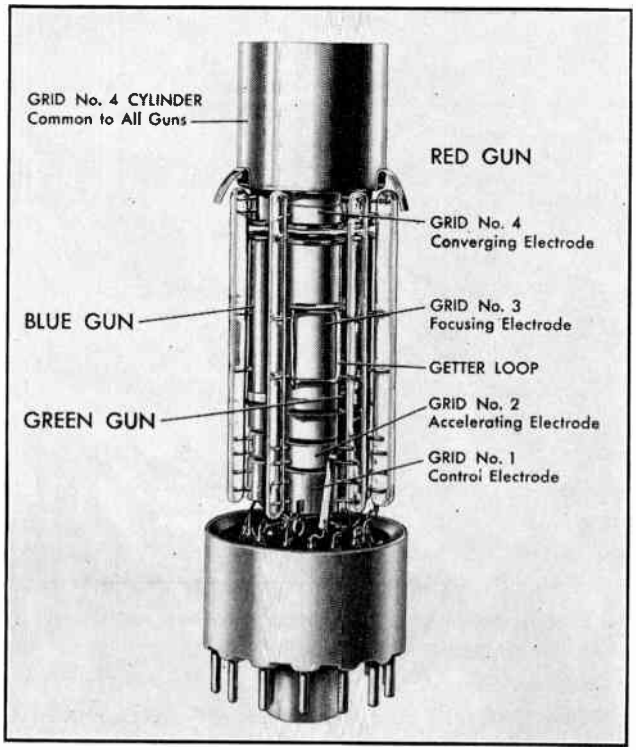


Fig. 19.32. Triple Electron Gun for Tricolor Kinescope. (Moodey and Van Ormer, reference 8.) (Courtesy of *Proceedings of the Institute of Radio Engineers.*)

cylinders are employed with three individual cathodes. The electrode cylinders are terminated by disks with three apertures apiece. This structure has, of course, less flexibility with respect to the adjustment of the transfer characteristics for the three beams.

The three beams supplied by the triple gun must be deflected by the deflection yoke in such fashion that they remain both converged and focused over the entire screen area. The design problem for the de-

flection yoke is very nearly the same as that of deflecting a beam sufficiently wide to encompass all three individual beams across the screen of a kinescope without appreciable defocusing.

The problem was solved in two stages: (1) a yoke giving an essentially uniform magnetic deflecting field was designed. This gave uniformly increasing overconvergence (curvature of field) from the center of the field outward. (2) A correcting potential depending on the instantaneous angular inclination of the beam to the axis was applied to

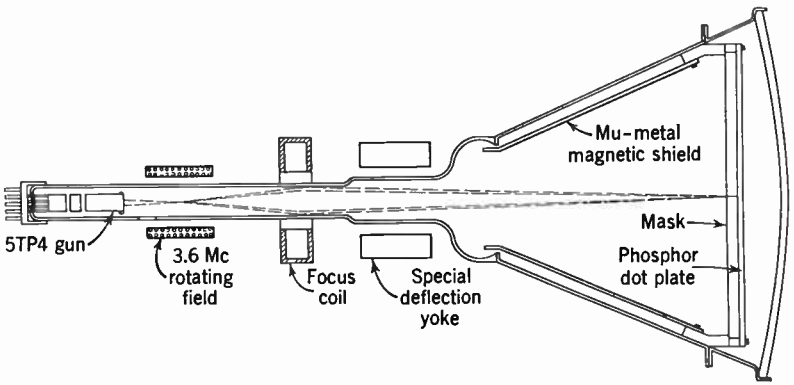


Fig. 19.33. Schematic Section of Single-Gun Shadow-Mask Tricolor Kinescope.

the convergence lens to balance this overconvergence (dynamic focusing).

A kinescope with a conical beam produced by a predeflection system similar to that employed in the single-gun shadow-mask tricolor kinescope * (Fig. 19.33) was employed to test the properties of empirical yokes and to derive an optimum coil design.† Multisection coils, with individually adjustable currents, were used in conjunction with an external ferrite ring core. Four ferromagnetic tabs, finally, provided proper shielding for the fields derived from the turned-up coil ends of the vertical coils. When the beams were sharply converged at the center of the field, such an anastigmatic core yielded slightly elliptical ring patterns with the conical beam when deflection currents were applied. Application of the appropriate convergence voltages to the convergence electrode contracted the ellipses to sharp dots, indicating convergence over the entire field. Deflection coil designs of this type were

* See Law, reference 9.

† See Friend, reference 10.

developed for total deflection angles of both 45 degrees (corresponding to the original 15-inch tricolor kinescopes) and 65 degrees. The appearance of one of the yokes is shown in Fig. 19.34.

The dynamic convergence voltage needed to obtain complete convergence correction is derived from the deflection circuits by integra-

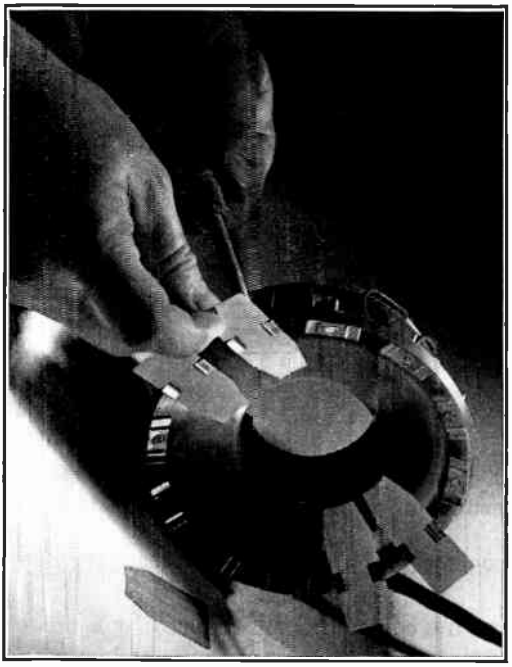


Fig 19.34. Sixty-Five-Degree Deflection Yoke with Four Compensating Tabs. (Friend, reference 10.) (Courtesy of *Proceedings of the Institute of Radio Engineers.*)

tion of the sawtooth, since the needed correction voltage is a parabolic function of deflection angle. In practice, the required voltage variation may be derived directly from the cathode circuits of the horizontal and vertical output tubes, if these circuits are given appropriate time constants. Figure 19.35 shows how the voltages are added, amplified, and applied to both the convergence electrode and the focusing electrodes. In practice (for 45-degree deflection) the maximum correction voltage required is approximately 10 percent of the voltage difference between convergence electrode and anode, or 1000 volts for a 10-kilovolt lens voltage. The voltage to be applied to the focusing electrodes to give sharp focus throughout the field is about a quarter of this.

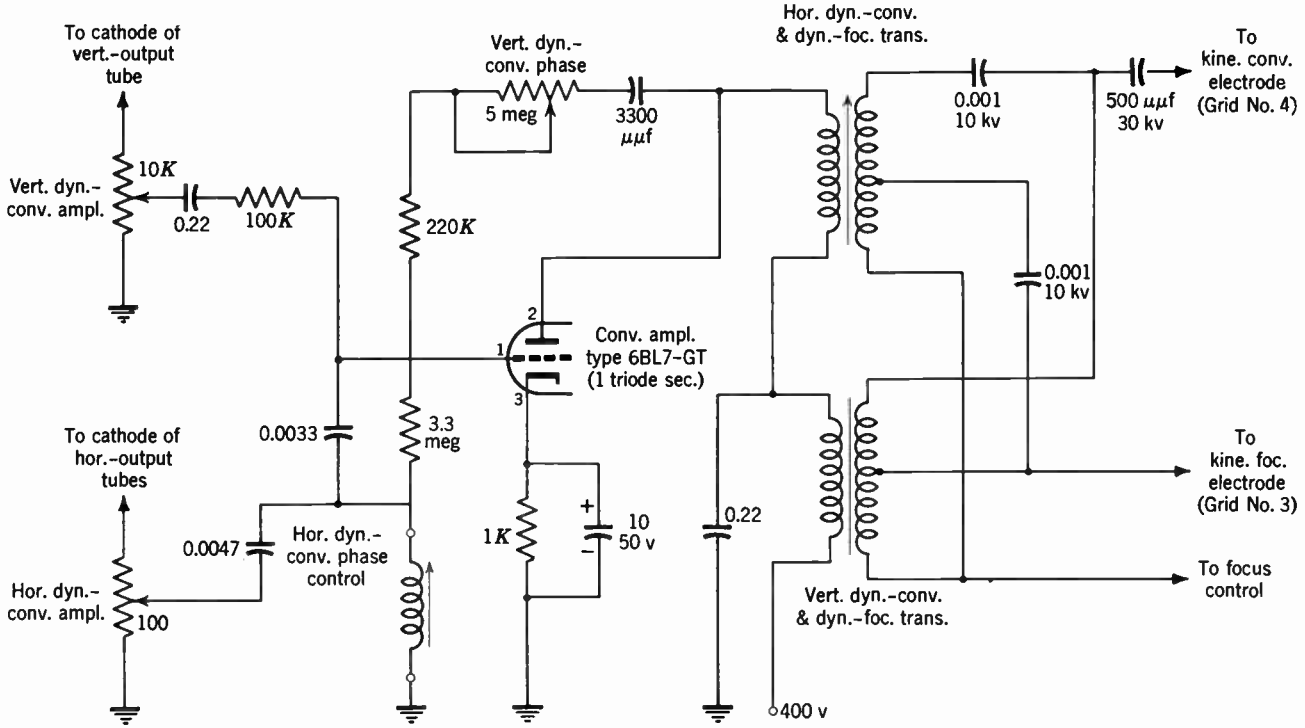


Fig. 19.35. Dynamic Convergence Circuit.

In the horizontal circuit, the sawtooth is integrated and an adjustable amount of sawtooth voltage added to the resulting parabolic voltage variation. A step-up transformer increases the output voltage to the level required at the convergence electrode. In the vertical circuit, double integration is employed since, as a matter of economy, a differentiating transformer is used instead of a step-up transformer.

As an alternative, magnetic astigmatic lenses may be employed to effect dynamic convergence correction. Other components concerned with beam deflection are the external mu-metal shield, protecting the tube interior from outside magnetic sources, and the color purity coil. The latter applies an adjustable transverse magnetic field at the triple gun and serves to shift the axis of the three beams toward the kinescope axis so as to correct the effects of small mechanical misalignments. The color purity coil, in conjunction with the deflection centering voltages, can also be employed to compensate small errors in the mechanical alignment of the mask-screen structure. Details of the yoke mounting and the auxiliary alignment elements can be gleaned from Fig. 19.36.

Some of the most interesting problems in the design of the shadow-mask color kinescope revolved about the construction of the mask-screen assembly, with some 600,000 phosphor dots (for a 15-inch tube) on a glass plate accurately aligned (within about 0.001 inch) with the 200,000 apertures in the mask.

The original negative for the mask was prepared by making a contact print of a fine wire grill on a photographic plate after two exposures, with an intervening 60-degree rotation of the grill.* By appropriate photographic processing and copying, it is possible to convert the initially diamond-shaped dots on the negative into nearly circular dots. A thin supernickel sheet coated with photoengravers' enamel is then exposed through this negative, the unexposed enamel washed away, and the exposed metal etched to form the desired apertures with a sharp feather edge.

The mask is "hot-blocked" on a steel frame, i.e., fastened to the supporting frame at a high temperature, in an expanded condition, so that it will not buckle when heated by bombardment in the tube. The glass plate on which the phosphor dots are to be deposited is fixed in position with respect to the frame as follows. Three straight-sided pins fitting in collets equipped with locking nuts are located on the periphery of the frame, 120 degrees apart. The heads of the pins carry a

* See Law, reference 7.

hemisphere which is located off the axis of the pin. These heads mate with three V grooves in the glass plate. In this manner allowance is made for differential thermal expansion.

The phosphor dot patterns are printed from a gelatin stencil.* The first step in the preparation of this stencil is the exposure of a Kodalith plate, replacing the screen, through the mask, a point source of light

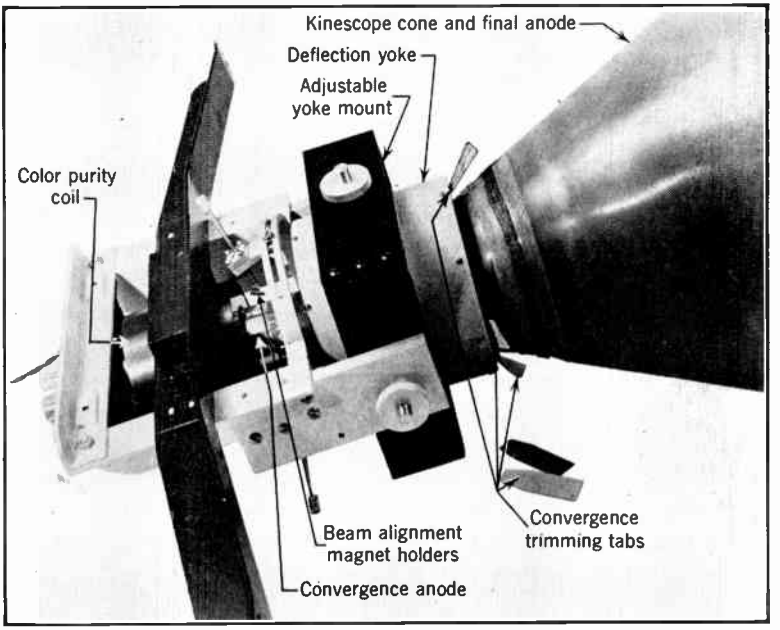


Fig. 19.36. Assembly of Deflection and Convergence Components on Neck of Tricolor Kinescope. (Friend, reference 10.) (Courtesy of *Proceedings of the Institute of Radio Engineers.*)

being placed at the deflection center for one of the three beams. The pattern obtained in this manner serves as master for all three groups of phosphor dots, the stencil being displaced relative to the plate by a dot separation between printings.

After the Kodalith plate has been developed, a gelatin emulsion with an intervening protective film of lacquer and wax is squeegeed onto the Kodalith emulsion and exposed through it to the light of a carbon arc lamp. The unexposed gelatin is then washed away with lukewarm water, and the wet stencil, after immersion in a temperature-controlled water bath, is placed on the printing mesh, where it is dried.

* See Freedman and McLaughlin, reference 11.

The phosphor printing medium, e.g., ethylcellulose in amyl alcohol, in which the phosphor powder is suspended, is applied to the glass plate through the stencil with a squeegee. Contrary to usual printing techniques, the stencil is held in contact with the glass plate during printing, so as to prevent distortion resulting from the squeegee pressure. Registry marks on the gelatin stencil fix its position relative to the glass plate within 0.001 inch. After the first printing, the stencil is raised straight upward from the glass and then removed for cleaning with amyl acetate. After the first printing has been dried, the stencil is once again placed in registry on the table, with the proper displacement for printing the next set of dots. The process is repeated three times.

The binder (ethylcellulose) is removed by baking the phosphor plate in air at 425°C for about one-half hour. The phosphor may then be bonded by spraying with a fine mist of potassium silicate solution. As a final step, a nitrocellulose "blanket" is floated onto the surface of the plate and aluminum evaporated to form a conducting reflective film. A final bake in air removes the nitrocellulose blanket. With certain phosphors, such as zinc-cadmium selenide, it is advisable to carry out only one bake in air, *after* deposition of the aluminum, since they deteriorate in a bake when in direct contact with air.

The three phosphors employed in the first tricolor kinescopes were green willemite, Zn_2SiO_4:Mn; blue calcium–magnesium silicate, $CaMg(SiO_3)_2$:Ti; and red-orange cadmium borate, $2CdO \cdot B_2O_3$:Mn. A sheet of didymium glass was placed in front of the kinescope to shift the color coordinates of the last primary component further toward true red. In later tubes, the position of the primaries in the color triangle was considerably improved by replacement of the borate phosphor with a zinc phosphate and admixture of zinc sulfide to the blue calcium–magnesium silicate. The properties of the new primaries are given below:

Primary	Relative Color Coordinates	Relative Luminous Efficiency	Relative Luminosity to Produce 7,300°K White
Red $Zn_3(PO_4)_2$:Mn	$x = 0.678$, $y = 0.322$	25.3	82.5
Green Zn_2SiO_4:Mn	$x = 0.204$, $y = 0.732$	100.0	100.0
Blue $CaMg(SiO_3)_2$:Ti + ZnS:Ag	$x = 0.146$, $y = 0.088$	26.6	40.0

Figure 19.37 shows the spectral distribution of the radiation emitted by these phosphors under electron bombardment, and Fig. 19.38 the location of the color coordinates in the color triangle.

To obtain identical transfer characteristics on all three channels, so that, for proper adjustment, white will be reproduced as white over

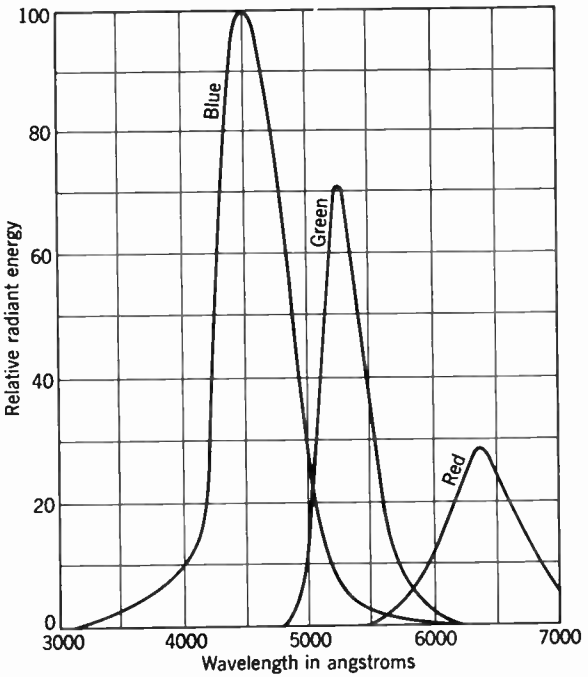


Fig. 19.37. Spectral Distribution of Phosphors Listed on Page 912.

the full operating range of the kinescope, it is desirable that equal grid voltages will be required on the three guns to reproduce white. For the primaries given above, this is far from being fulfilled. If equal signals are applied to all three grids (with identical screen-grid adjustment), the red picture will possess less than one-third the required brightness and the blue picture two-thirds the desired brightness.

The spectral responses for the standardized group phosphor P22, shown in Fig. 19.39, show considerable improvement over those in Fig. 19.37. Inspection indicates that, here, the relative red response lags by less than 50 percent behind the ideal value, the blue response, by less than 20 percent.

The sealing of the tube presents special problems since during it the mask-screen assembly, with relative position tolerances of the order of 0.001 inch, must not be disturbed.* In the metal tube, the metal

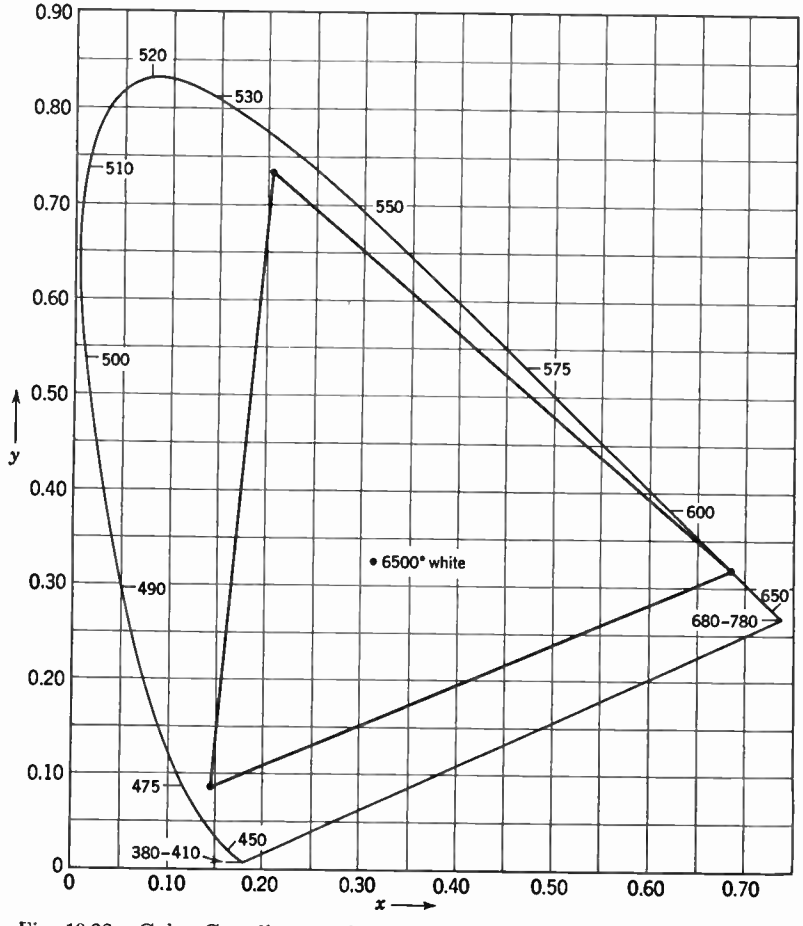


Fig. 19.38. Color Coordinates of Phosphors Listed in Table on Page 912.

cone of a 16AP4 kinescope was prepared in two sections, terminating in well-fitting flat flanges (Fig. 19.40). The mask-screen assembly was mounted on mounting posts on the lower section and the face plate sealed to the upper section. Then the flanges were joined together by a Heliarc weld. Strain on the face plate seal was minimized during

* See Barnes and Faulkner, reference 12.

this process by clamping heavy copper rings to the flanges so as to provide effective cooling. Essentially similar techniques are employed in the sealing of the all-glass tube shown in Fig. 19.41.

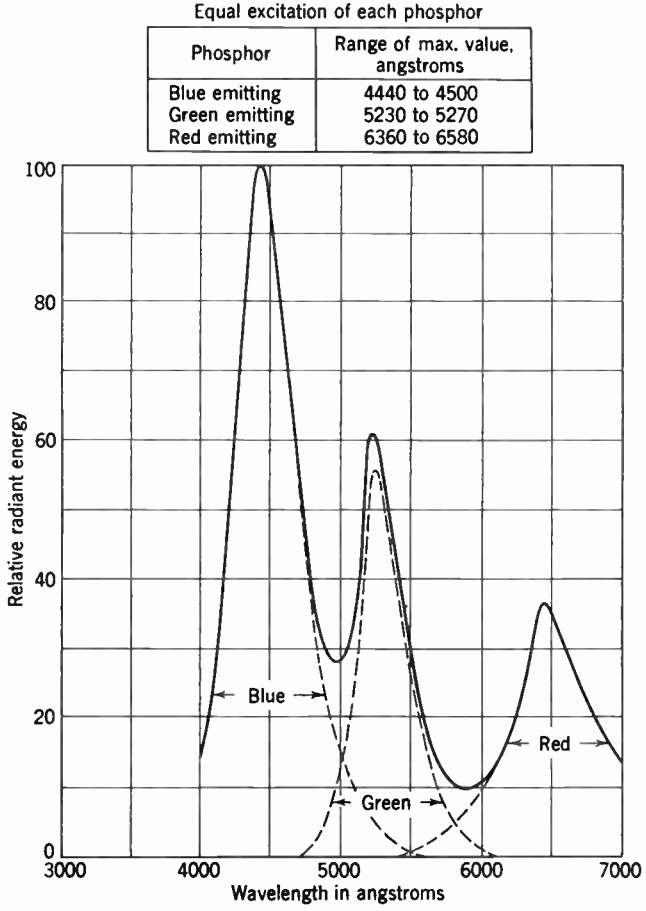


Fig. 19.39. Spectral Distribution of Group Phosphor P22.

The tubes described above have given very satisfactory pictures with a high-light brightness in excess of 20 foot-lamberts. With scanning in the preferred direction (Fig. 19.42), minimizing the separation of successive rows of apertures parallel to the scanning lines, "moiré effects," resulting from the interference between the scanning pattern and mask pattern, are scarcely detectable. Moiré effects prove disturbing only

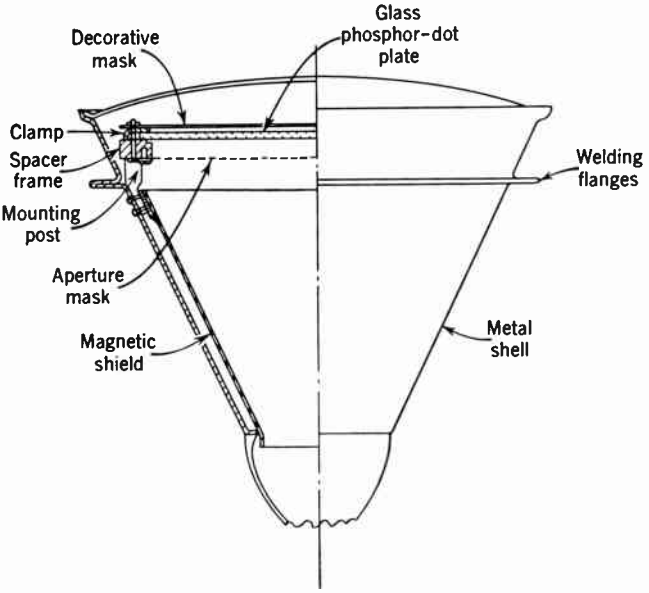


Fig. 19.40. Mechanical Construction of Shadow-Mask Tricolor Kinescope. (Barnes and Faulkner, reference 12.) (Courtesy of *Proceedings of the Institute of Radio Engineers.*)

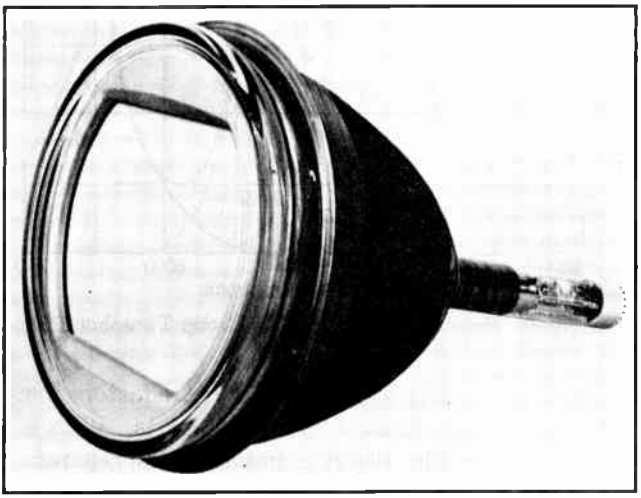


Fig. 19.41. RCA Glass Envelope Tricolor Kinescope.

when the scanning direction deviates from the preferred orientation by several degrees.*

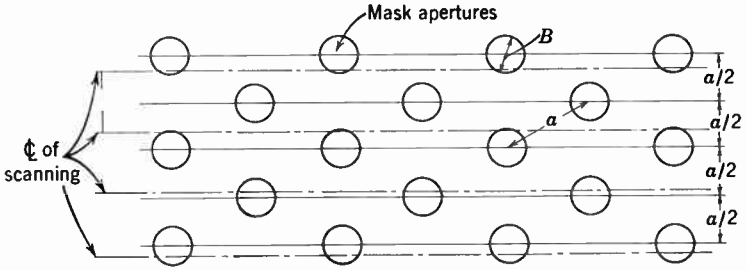


Fig. 19.42. Preferred Direction of Scanning in Shadow-Mask Kinescopes.

Fig. 19.43. A Color Receiver for Compatible Color Television.

The power supplies required by the color receiver are essentially the same as those needed by a monochrome receiver, making due allowance for the increased demand resulting from the larger number of

*See Ramberg, reference 13.

tubes in the set. The high-voltage supply is a conventional regulated high-voltage supply delivering 1 milliampere at 18 kilovolts, deriving its power from a high-frequency oscillator. Figure 19.43 shows the external appearance of a complete color receiver.

19.6 Special Problems in Color Transmission. There are a number of special problems in the transmission of color television programs which may be expected to become of little importance as the art de-

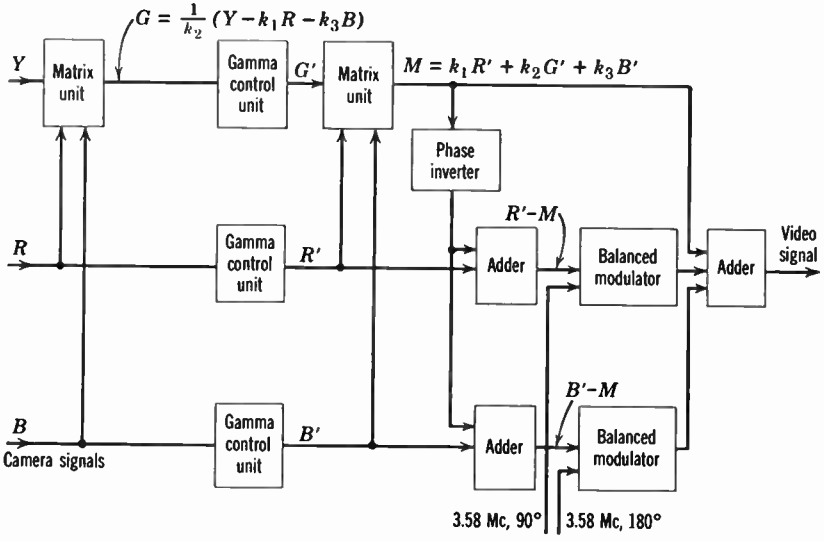


Fig. 19.44. Transmitter with One Camera Channel Providing the Luminance Signal Y, to Reduce Effects of Camera Misregistration.

velops. One of these concerns minimizing the effects of misregistration in the camera. This problem may be expected to disappear with the development of a single-tube color pickup camera.

Effects of camera misregistration can be minimized in the picture received by a monochrome receiver and a color receiver as well, by letting one camera tube deliver the monochrome or luminance signal and deriving the chrominance signal from it and the red and blue sensitive camera tubes (Fig. 19.44). Faithful color reproduction demands, it is true, that the monochrome signal be obtained by first developing the green signal with the aid of a matrix unit and then rooting this along with the red and blue signal and combining the three in a second matrix unit, rather than from the monochrome camera tube directly. In a completely linear color television system, monochrome reproduction would be obtained, by this device, without any effect of misregis-

tration in the camera; in color reproductions, the misregistration would result in chromaticity errors at transitions, but in no luminance errors. Nonlinear transmission results in misregistration fringes with a small luminance error in both cases. It is found empirically, however, that

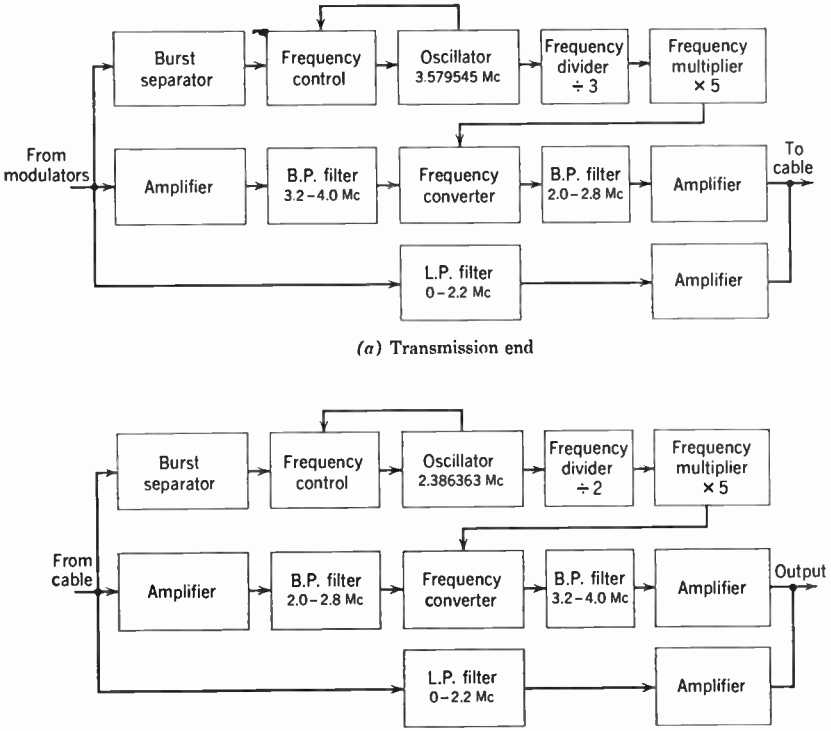


Fig. 19.45. Modification of Color Television Signals for Transmission over Channel of Limited Passband.

these misregistration effects are very small compared to those observed when a green camera channel replaces the luminance channel.

Another problem is the transmission of color television signals over cables with a passband cutting off below the accepted subcarrier frequency, a condition prevailing in some of the older coaxial cable circuits. For some of these, the cutoff frequency is at 2.8 megacycles. It then becomes necessary to convert the subcarrier from its normal value (which, in the present illustration, will be chosen as 3.58 megacycles) to a lower, e.g., 2.39 megacycles, at the entrance to the cable

and to reconvert it to its original value, 3.58 megacycles, at the exit so that the signal can be handled by regular color receivers. Figure 19.45 shows the process in greater detail. The burst component of the output from the transmitter subcarrier oscillator is utilized to phase-control a 3.579545-megacycle oscillator from which a 5.965908-megacycle oscillation is derived by frequency division and multiplication. This oscillation is beaten with the subcarrier modulator output to produce the new 2.386363-megacycle oscillation with its sidebands, which are selected out by a bandpass filter. The shunted monochrome signal, limited in frequency to 2.2 megacycles, is added to the color signal. In the receiver the process is simply reversed, the burst signal, which is now heterodyned down to 2.39 megacycles, serving once more to control the local oscillator. Obviously, the detail in the picture is now limited as the result of the narrower passband of 2.2 megacycles. Like cable-transmitted black-and-white pictures, the color picture is found acceptable.

19.7 Field-Sequential Color Television. A field-sequential system of color television, with a repetition rate of 24 color frames per second and 405 lines per frame, was established as standard in the United States by an order of the Federal Communications Commission of October 10, 1950, which has since been superseded. Some of the drawbacks of this system were pointed out in Chapter 17. They have prevented its successful use in commercial broadcasting. One of the most serious objections to the system, namely, its lack of compatibility, is, however, of minor interest in industrial television, where, in fact, field-sequential chains designed by the Columbia Broadcasting System have been introduced by Remington-Rand under the name of Vericolor. In closed chains, furthermore, it is permissible to modify the standards by an increase of the frequency passband, the line frequency, and the color-frame repetition rate, so as to reduce flicker and improve resolution. This course was taken by DuMont in the development of their industrial color television chains (see Chapter 20).

In view of these practical applications, it appears appropriate to return at this point to a brief consideration of the equipment that has been employed by the Columbia Broadcasting System * for field-sequential color transmission and reception.

The principal changes required in a monochrome television system to convert it to a field-sequential color system based on the former FCC standards are the following:

* See Goldmark, Christensen, and Reeves, reference 14.

At the camera, a color disk must be inserted in front of the camera tube face; deflection circuits must be modified to operate at a vertical frequency of 144 second^{-1} and a horizontal frequency of 29,160 second^{-1}; a color mixer unit must be added permitting separate control of gain and background for each color channel; and a "color pulse" must be inserted between the first two equalizing pulses of the vertical

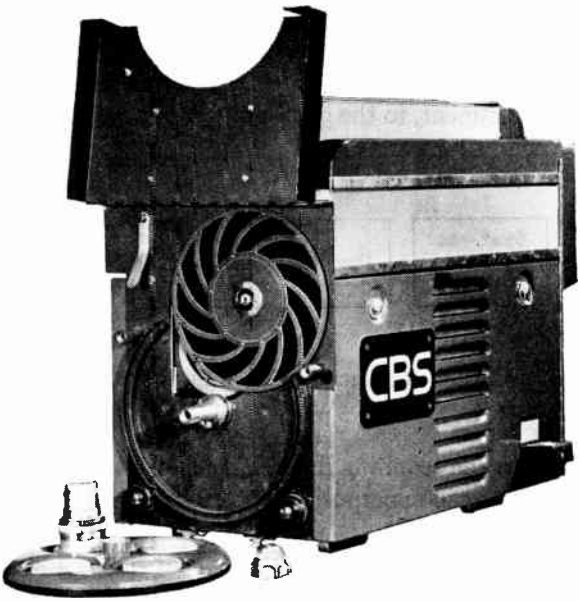


Fig. 19.46. RCA Monochrome Camera Modified for Field-Sequential Color Operation. (Goldmark, Christensen, and Reeves, reference 14.) (Courtesy of *Proceedings of the Institute of Radio Engineers.*)

synchronizing signal in every red field to permit proper color synchronization between transmitter and receiver.

At the receiver, a disk or drum whose drive is synchronized with the color disk at the camera is required, and modifications in the deflection circuits corresponding to those mentioned in connection with the camera are needed. Special care must be taken in the filtering of power supplies and the shielding of sensitive components to prevent the appearance of 60-cycle hum in the deflection circuits since it results in serious picture deterioration. If the same receiver is to serve for the reception of both color and monochrome transmissions, it is necessary to design the deflection circuits in such a way that, by actuation of a multiple switch, they can be shifted from operation on color standards

to operation on monochrome standards. Furthermore, provision must be made for removing the filter disk from the region in front of the viewing tube face.

Figure 19.46 shows an RCA monochrome studio camera modified for field-sequential color operation. The twelve-filter disk is driven at 720 rpm by a synchronous motor, which, in turn, is actuated by a 48-cps, 115-volt amplifier. The motor drive is derived from the 48-cps color drive pulse as indicated in Fig. 19.47. The drive pulses generate a 48-cycle square wave in a multivibrator. This square wave is filtered to produce a sine wave, which is coupled through a selsyn, to permit phase adjustment, to the push-pull output amplifier. The phase

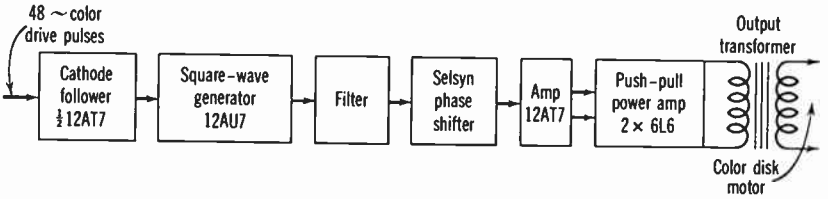


Fig. 19.47. Drive Circuits for Camera Disk Motor. (Goldmark, Christensen, and Reeves, reference 14.)

is adjusted so that the scanning lines coincide at all times with a spoke of the filter. Thus the superposition of signals from successive color fields is prevented.

Figure 19.48 shows schematically the color-mixer unit. A color gating pulse generator in the nature of a ring multivibrator actuated by the vertical drive pulses and color drive pulses gates, in turn, the red, blue, and green amplifier tubes in synchronism with the transmission of the red, blue, and green fields. The resultant, added to the composite blanking pulses, forms the video signal. The camera signal is "rooted" ahead of the color-mixer unit to compensate receiver nonlinearity.

The much more compact industrial color television camera is shown in Fig. 19.49. The image orthicon is here placed at right angles to the lens axis, the light path being broken by a 45-degree mirror placed within a 2½-inch filter drum. All camera controls are operated from the control console.

In the combination receiver, a six-section filter disk rotates at 1440 rpm in front of the viewing tube face; for a 10-inch tube, the disk diameter is 22½ inches. The apparent size of the picture may then be increased to 12½ inches by use of an enlarging lens. To permit a

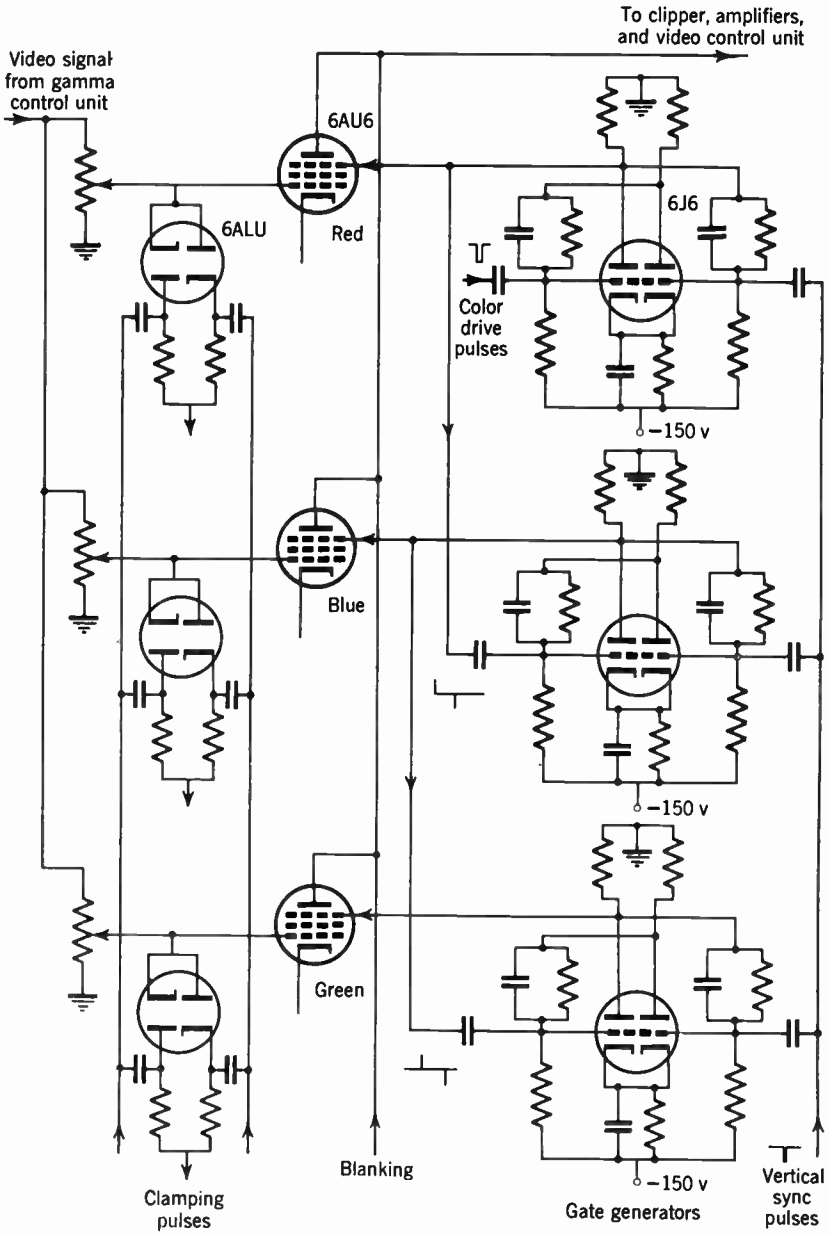


Fig. 19.48. Schematic Diagram of Color Mixer.

rapid shift from color transmissions to monochrome transmissions, the filter disk is made in two sections; one of these is driven, whereas the other floats freely on the disk axis. As the disk is driven, air drag causes the floating section just to fill the clear portion of the driven section, so that the compound disk acts as a single six-section filter disk. When the color-to-monochrome switch is actuated, the drive is reversed, and the floating disk is stopped by a centrifugal latch. The rotating section continues to rotate backwards until the clear portions

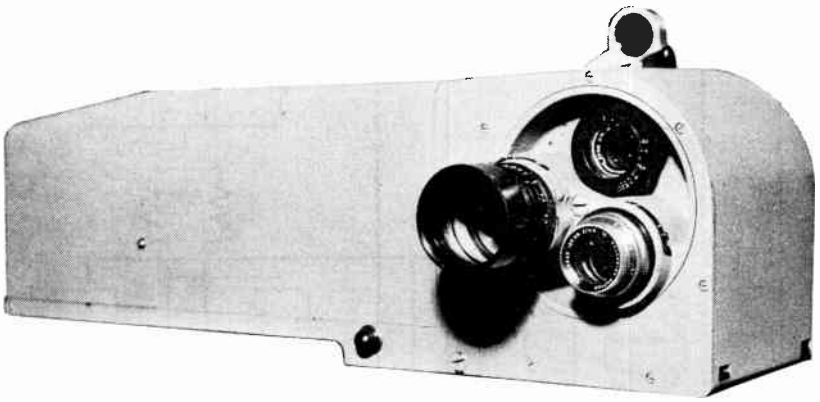


Fig. 19.49. Industrial Color Television Camera. (Goldmark, Christensen, and Reeves, reference 14.) (Courtesy of *Proceedings of the Institute of Radio Engineers.*)

of the two disks are superposed in front of the tube, at which point a microswitch stops the motor.

A reluctance-type generator is mounted on the disk shaft. During operation, it produces a sharp 144-cycle sawtooth wave which is applied, along with the vertical synchronizing signals, to a phase detector. The d-c voltage corresponding to the phase difference between the two waves developed in this manner controls the saturation of a saturable reactor in series with the windings of an induction motor driving the color disk. For example, if the sawtooth wave lags behind the vertical synchronizing signals, the saturation of the reactor is increased and the motor is speeded up. The proper disk phase can be maintained within 2 degrees in this manner. Anti-hunting circuits prevent oscillation about the proper adjustment. A push button puts the synchronizing circuit temporarily out of control when the disk has locked in on the wrong color field.

As in the camera, the deflection circuits have to be designed so that they can operate both on the former FCC sequential color standards and on monochrome standards. This requires, e.g., in the horizontal circuit, nine-ganged single-pole double-throw switches. With "slave

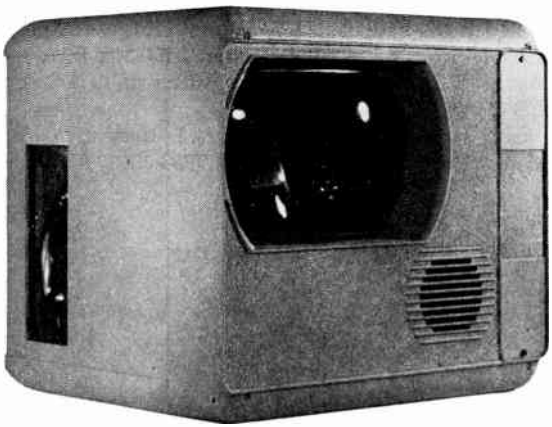


Fig. 19.50. Industrial Color Monitor. (Goldmark, Christensen, and Reeves, reference 14.) (Courtesy of *Proceedings of the Institute of Radio Engineers.*)

units" designed for attachment to monochrome receivers, as well as industrial monitors designed for color exclusively, these complications are absent. Figure 19.50 shows an industrial color monitor unit with an intercommunication telephone mounted at the rear.

REFERENCES

1. R. D. Kell, G. C. Sziklai, R. C. Ballard, A. C. Schroeder, K. R. Wendt, and G. L. Fredendall, "An Experimental Simultaneous Color-Television System," *Proc. I.R.E.*, Vol. 35, pp. 861–875, 1947.
2. V. K. Zworykin, *Television,* J. Franklin Inst., Vol. 244, pp. 131–145, 1947.
3. G. L. Dimmick, "A New Dichroic Reflector and Its Application to Photocell Monitoring Systems," *J. Soc. Motion Picture Engrs.*, Vol. 38, pp. 36–55, 1942.
4. L. T. Sachtleben, D. J. Parker, G. L. Allee and E. Kornstein, "Image Orthicon Color Television Camera Optical System," *RCA Rev.*, Vol. 13, pp. 27–33, 1952.
5. J. D. Spradlin, "The RCA Color Television Camera Chain," *RCA Rev.*, Vol. 13, pp. 11–26, 1952.
6. D. H. Pritchard and R. N. Rhodes, "Color Television Signal Receiver Demodulators," *RCA Rev.*, Vol. 14, pp. 205–226, 1953.
7. H. B. Law, "A Three-Gun Shadow-Mask Color Kinescope," *Proc. I.R.E.*, Vol. 39, pp. 1186–1194, 1951.
8. H. C. Moodey and D. D. Van Ormer, "Three-Beam Guns for Color Kinescopes," *Proc. I.R.E.*, Vol. 39, pp. 1236–1241, 1950.

9. R. R. Law, "A One-Gun Shadow-Mask Color Kinescope," *Proc. I.R.E.,* Vol. 39, pp. 1194–1201, 1951.
10. A. W. Friend, "Deflection and Convergence in Color Kinescopes," *Proc. I.R.E.,* Vol. 39, pp. 1249–1263, 1951.
11. N. S. Freedman and K. M. McLaughlin, "Phosphor Screen Application in Color Kinescopes," *Proc. I.R.E.,* Vol. 39, pp. 1230–1236, 1951.
12. B. E. Barnes and R. D. Faulkner, "Mechanical Design of Aperture-Mask Tri-Color Kinescopes," *Proc. I.R.E.,* Vol. 39, pp. 1241–1245, 1951.
13. E. G. Ramberg, "Elimination of Moiré Effects in Tri-Color Kinescopes," *Proc. I.R.E.,* Vol. 40, pp. 916–923, 1952.
14. P. C. Goldmark, J. W. Christensen, and J. J. Reeves, "Color Television—U.S.A. Standard," *Proc. I.R.E.,* Vol. 39, pp. 1288–1313, 1951.

20 Industrial Television

In its broadest aspect television signifies an extension of human vision. The word implies that the impediment which this extension seeks to overcome is distance. The same technique may be employed, however, to overcome other impediments to vision as well. These are, to mention a few, hazards to the observer from high or low temperatures, violent reactions, and atmospheres unsuitable for the maintenance of life; observer locations inaccessible to human beings because of weight and space requirements; the impossibility of simultaneous observation of a single event by a group from the same vantage point as well as the simultaneous observation of events separated in space by a single observer; and the insensitivity of the human retina to radiations in the ultraviolet and infrared portions of the electromagnetic spectrum. These potentialities map out for television a sphere of usefulness which is far wider than the dissemination of entertainment and information from central broadcasting stations. It has become customary to refer to the uses of television outside the broadcast field collectively as industrial television.

20.1 Applications of Industrial Television. It is a distinctive property of industrial television that it deals with closed television systems. Unlike broadcast television, it lays no claim to the public domain of the "air waves." Consequently, it is free of public regulation and of externally imposed standards. In principle, every industrial television system may be tailored to fit a specific application. In practice, economic factors impose limits on this freedom. Low unit cost of the equipment can be achieved only if it is designed to meet the needs of a large range of applications.

An examination of the potential applications of industrial television constitutes therefore a first step in laying down the design principles for a general-purpose industrial television system. At the time of writing, the field-testing of industrial television equipment has progressed sufficiently far that, here, the "potential applications" may be largely replaced by actual applications.

It is convenient to consider in sequence the uses of industrial television in industry, in education, in commerce, and in research. Finally, there are various military applications, which claimed the major portion of early efforts in this field.

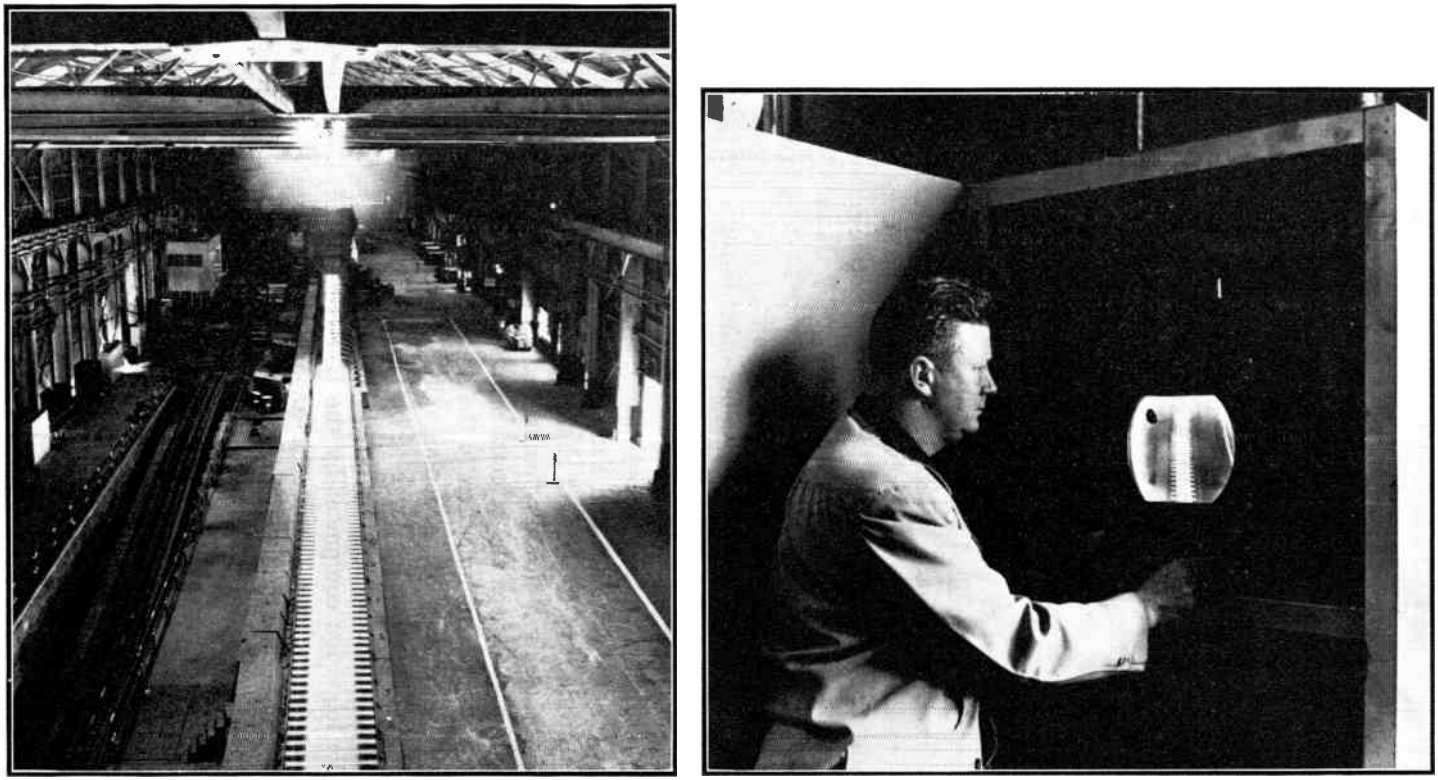
Industrial television has found particularly extensive use in the utilities and in the steel industries. In both places the greater heat tolerance of the television camera, as compared with that of a human observer, plays a role. In addition, television makes possible the observation and control of a number of related, spatially separated, processes from a convenient central point.

In the utility industry television chains have been used principally so far for transferring pictures of the water-level indicator of large-capacity boilers and of the flame condition in furnaces to the control room, which may be located some five floors below the boiler. The observation of the water-level gauge by television presents no special problem. For observing the furnace flames, on the other hand, various methods have been developed. In a pulverized-coal burning furnace of the Long Island Lighting Company * the camera is mounted directly at the top of the furnace pit, above a water-cooled window, giving a full view of the pilot flames, the light-oil burners, and the general flame condition after the injection of the pulverized coal. Imperfect operation of any part, excessively rich mixtures which might give rise to explosion hazards and objectionable smoke from the stacks, etc., are immediately detected. In another installation † a mirror system directs the images of six windows placed in the front wall of the furnace below flame level onto the photocathode of the camera tube.

Industrial television aids the manufacture of steel by showing the operator in the control room the proper moment for shifting steel plate from a conveyor belt into reheating furnaces as well as the instant for interrupting the pouring process in making large castings.‡ The proper filling of the forms is essential to prevent spongy spots and air inclusions in the castings. In a Fisher Body Plant near Pittsburgh a television monitor in a control room within the plant shows the operator the loading of gondola cars on a railroad siding with scrap steel bales from a baling machine; when the car is filled he manipulates controls which shift an empty car in the place of the full one.§

There are many examples of the use of industrial television for the monitoring of dangerous operations. The control by television of the disassembly of explosive items by remote-control instruments operating

* See Exley, reference 1.
† See Bice, reference 2.
‡ See reference 3.
§ See reference 4.



(*a*) (*b*)

Fig. 20.1. (*a*) Television Camera Mounted 48 Feet High on Rafters Scans 395 Feet of Hot Steel Strip. (*b*) Operator Checks Strip on Passage from Finishing Stand to Coilers on Television Monitor.

communications and surveillance. Fashion shows and other exhibits of special interest can be transferred, by this means, to a number of strategic locations in a large store. In banks, a television camera in a central file room can transmit signature and account cards to receivers in the tellers' cages upon request, speeding up the cashing or certification of customers' checks (Fig. 20.4). A similar technique, for giving rapid access to the information stored in library stacks, is under

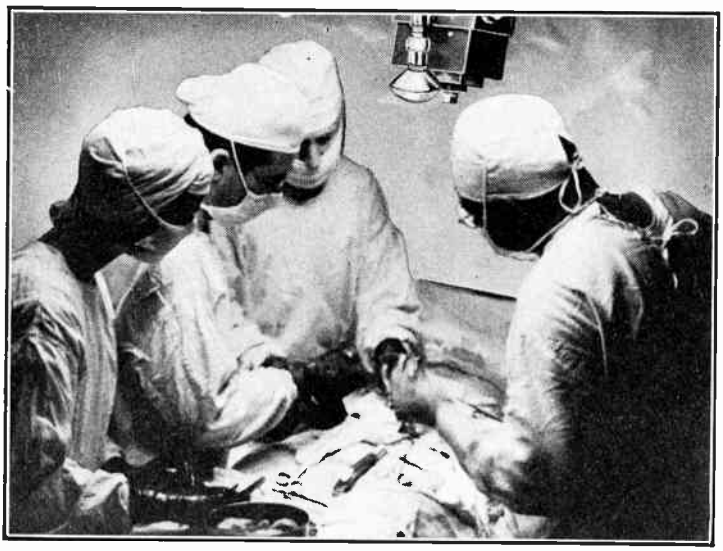


Fig. 20.3. A Stereoscopic Television Camera in Use during a Surgical Operation.

investigation by the Office of Naval Research cooperating with the Library of Congress.*

Industrial television has important research applications as substitute observer in dangerous and inaccessible locations; a number of instances have already been noted. Its use in astronomy, to provide sidereal images free from atmospheric effects by mounting the camera on a guided balloon, has also been considered.†

In the television microscope, to be treated in greater detail in a later section, television serves the quite different function of making possible the continuous observation of specimens illuminated by radiation to which the eye is insensitive and, furthermore, of employing electronic controls for modifying and enhancing picture contrast. Other types of

* See reference 11.

† See Zworykin, reference 15.

specialized television systems serve for making cell counts and mapping body potentials indicative of heart and brain functioning.*

Underwater television was used to aid divers in their work as early as 1947, in connection with the atomic bomb demonstrations at the Bikini Atoll. Its use for ship salvage has already been mentioned. The application of television to marine research promises to be even more important. With its aid, the fauna and flora of the ocean can be

Fig. 20.4. Checking Signatures on Drafts in a Bank.

studied in its natural habitat, hundreds of feet below the surface. The scientist can see events in the proximity of the submerged camera directly on the screen of the monitor. He need not rely on the verbal reports of scientifically untrained divers or their selection of scenes for underwater photography. Furthermore, the television camera, unlike a human diver, can remain submerged for long periods of time and be raised and lowered at a rapid rate.

The equipment required for underwater television is naturally somewhat more complex than standard industrial television equipment. A pressure-proof housing is an obvious requirement. In addition, a lighting unit controllable from the surface must be provided. Lens focus, iris, and mechanical orientation of the camera must also be remotely controlled, in addition to electrical camera adjustments. However,

* See Zworykin and Flory, reference 16.

these complications are slight when compared with the difficulties inherent in deep-sea studies by conventional methods.

The original military application of industrial television was the placement of a television camera in the nose of a guided missile, to transmit to the control point information with respect to its aim.* Many other uses have been added since. A few of these were noted in connection with the applications in industry and education.

20.2 Requirements of a General-Purpose Industrial Television System. The preceding survey indicates a wide variety in the applications of industrial television. The one common feature appears to be that the number of cameras which may be required is of the same order as the number of receivers. This is in sharp contrast to broadcast television, where the ratio of these two numbers is approximately 1:10,000. The standard unit of industrial television is not the receiver or the camera, but the television chain, consisting of a camera, a receiver, a connecting cable and power supplies and deflection control units for both camera and receiver. Obviously, low cost and simplicity of operation of the camera is of much greater importance in industrial television than in broadcast television. More generally, the following may be listed as desirable characteristics of a general-purpose industrial television system:

1. Low overall cost of the television chain.
2. Simplicity of adjustment.
3. Durability of components.
4. High sensitivity.
5. High resolution.
6. Light weight and small bulk of camera.
7. Remote camera control.
8. Overall portability.
9. Standards permitting addition of commercial receivers to chain.

The relative importance of these several factors will differ, depending on the type of application emphasized. For instance, equipment designed primarily for the surveillance of water-level gauges of boilers requires neither high resolution nor high sensitivity; a powerful spotlight can always be employed to raise the light level to the required point. On the other hand, great ruggedness and low temperature dependence are here decidedly desirable. In most research and educa-

* See Zworykin, reference 17; Kell and Sziklai, reference 18; and Sanders, reference 19.

tional applications, on the other hand, high resolution and sensitivity are factors of much greater importance than extremely long tube life and tolerance of excessive temperatures. In still other instances questions of cost, bulk, and complexity are secondary in comparison with the demand for the highest possible picture quality. Here, the conventional studio camera chain, eventually modified to permit the use of greater video bandwidth, provides the logical answer.

Since studio equipment is described elsewhere, the following account is concerned primarily with systems which differ materially from it—in cost, simplicity, and other aspects which fit them for their specific purpose.

20.3 Industrial Television Systems. The first system to be considered is the Utiliscope developed by the Capehart-Farnsworth Corporation for the Diamond Power Specialty Corporation.* It employs an image dissector with translucent cathode as camera tube. This choice marks out the field of application of the Utiliscope. The image dissector does not meet studio requirements with respect to either sensitivity or resolution; the figure cited for the second factor is 300 lines. The first factor implies the need for high-intensity illumination of the subject.

In compensation, the image dissector, if properly treated, has an almost indefinite life since it contains no hot cathode; like the iconoscope and other tubes with photoemissive targets it tolerates large variations in temperature; and, finally, its 11-stage silver-magnesium dynode multiplier provides a high-level video output (one-half volt across 1000 ohms), minimizing additional amplification requirements.

It follows that the Utiliscope is designed primarily for industrial surveillance applications, where the object to be observed presents relatively little detail and either high-level illumination is readily provided or the object is self-luminous. The long-life characteristics of the pickup tube permit the placement of the camera at relatively inaccessible locations and minimize the amount of servicing required.

The complete Utiliscope equipment (without connecting cables) is shown in Fig. 20.5. It consists of a camera with a 90-millimeter $f/1.4$ lens and video amplifier circuits; a power supply including deflection and blanking circuits for the camera and synchronizing signal generators for the monitor; and a monitor with a 10-inch picture tube. Apart from the pickup and viewing tubes the equipment contains only 15 vacuum tubes. A multiconductor cable, which may be up to 25

* See Sanders, reference 20.

feet in length, connects the power supply with the camera; three cables transmit, separately, the video signals (with blanking) and horizontal and vertical synchronizing pulses to the monitor, which may be located as much as 1000 feet away.

The scanning pattern has the standard 4:3 aspect ratio, but scanning is sequential rather than interlaced, with a horizontal frequency of 21.5 kilocycles. The scanning frequencies are internally controlled, by blocking oscillators, and are independent of the power-line alternations.

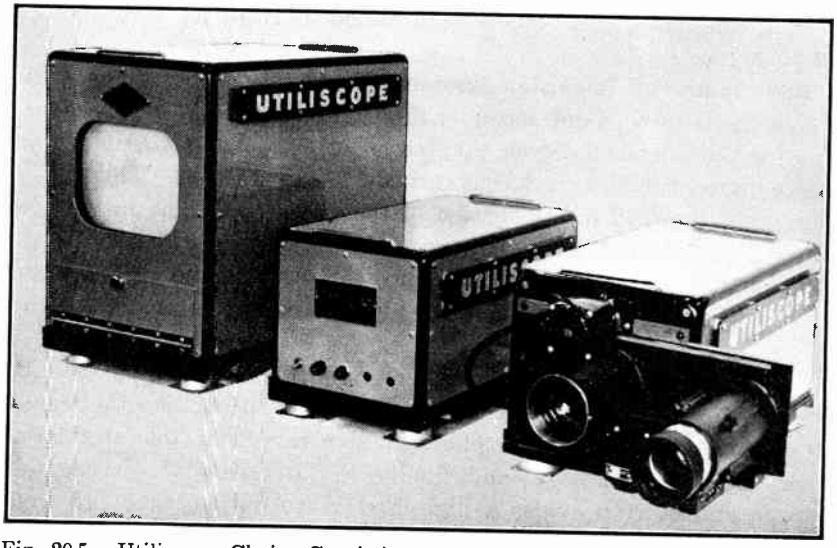


Fig. 20.5. Utiliscope Chain, Consisting of Monitor, Power Unit, and Camera. (Courtesy of Diamond Power Specialty Corporation.)

The simplified 5527 iconoscope, with electrostatic deflection and translucent mosaic, is the camera tube in an early industrial television system * serving similar purposes as the Utiliscope; this tube has the advantage of much greater sensitivity as compared with the dissector. In addition, it is extremely cheap and requires an operating voltage of only 600 to 800 volts. In compensation it has a certain amount of spurious shading, of which the dissector is entirely free, and demands a number of preamplifier stages to bring the video signal up to the level of the order of one volt.

As in the Utiliscope, the video signals and synchronizing signals are transmitted separately. Furthermore, sequential scanning is employed,

* See Barrett and Goodman, reference 21.

with a frame frequency of 60 and a line frequency of 15,000 second^{-1}, yielding a 250-line picture. Again, no attempt is made to lock the deflection in with the power-line frequency.

The next system to be described * comes far closer to satisfying all the requirements set forth for a general-purpose television system than either of the preceding ones. By the use of the Vidicon as camera tube, this system achieves extreme compactness along with high resolution (400 lines) and high sensitivity; only the image orthicon is superior to the Vidicon in the last respect. In addition, the Vidicon is free from the type of redistribution shading which characterizes the iconoscope and yields a relatively high-level signal output. Its high signal-to-noise ratio permits complete correction for aperture distortion in the horizontal direction, resulting in a picture with the full horizontal resolution admitted by the transmission channel.

Unlike the two other systems, the RCA Industrial Television System incorporating the Vidicon employs standard 525-line, 60-frame interlaced scanning, with the scanning frequency locked in with the local a-c power system. This makes possible the reception of the picture on any standard commercial television receiver, which may be connected to the monitor by a cable up to a mile in length. The complete video signal modulates a carrier corresponding to a television channel 3 to 6 in a modulated stage provided in the monitor; the cable transmits the signal from a jack connection on the monitor to the antenna terminals of the remote receiver.

The monitor itself, which may be connected to the camera by a 500-foot cable, contains complete control circuits and power supplies for the camera, including optical focusing control, blanking, beam current and voltage (focusing) control, and horizontal and vertical deflection current generators. In addition, it contains, of course, the monitor kinescope controls and circuits as well as the aforementioned carrier modulator for remote receivers.

The complete system, shown in Fig. 20.6, thus consists of only two units, a camera weighing only 7½ pounds and a control monitor with a 10-inch viewing tube weighing 70 pounds. The camera contains, in addition to the Vidicon with its permanent-magnet focusing system and its deflection coils and a motor at the rear for rotating the lens focusing mount, only two miniature tubes and associated circuits (Fig. 20.7). One tube is a pentode serving as preamplifier, the other a double triode, of which one half is a cathode follower for matching the

* See Webb and Morgan, reference 22.

video output to the 50-ohm cable, whereas the other half serves as blanking amplifier.

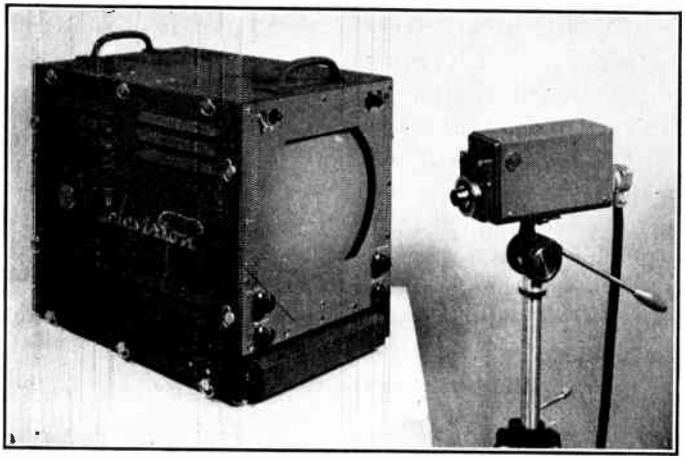


Fig. 20.6. RCA Industrial Television System with Vidicon Camera.

The Vidicon focusing system is shown in Fig. 20.8. It consists of four Alnico rods set in a pair of square soft-iron pole pieces. Soft-iron shunting sleeves on the rods may be displaced to obtain optimum alignment between the Vidicon gun and the external magnetic field.

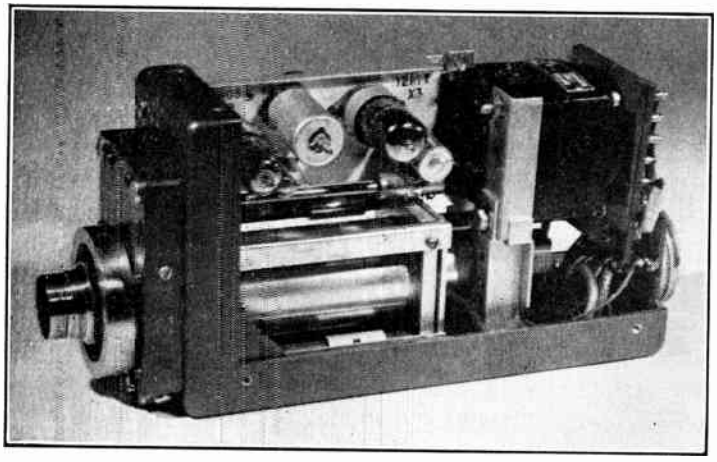


Fig. 20.7. Vidicon Camera, with Top and Sides Removed.

The circuits of the system incorporate a number of novel features. The four-stage video amplifier in the monitor, peaked to compensate

for the effect of Vidicon capacitance and with blanking inserted in the penultimate stage, is essentially conventional. The plate output of

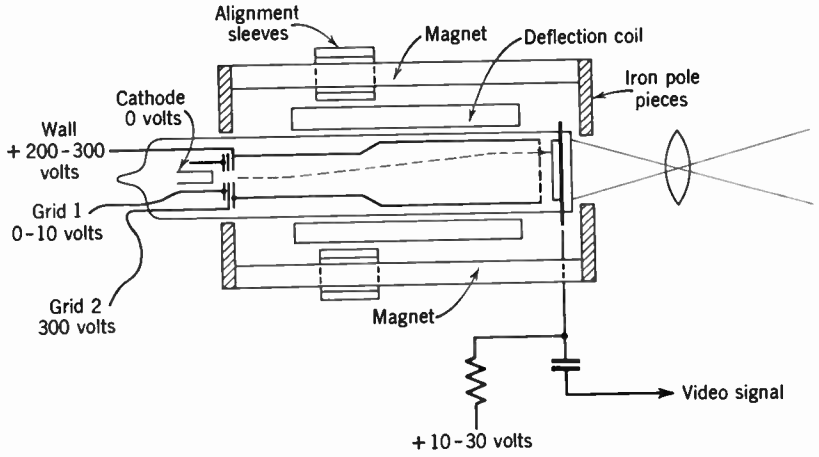


Fig. 20.8. Focusing System of the Vidicon.

the final stage drives the kinescope, while the cathode output is employed to modulate a very-high-frequency carrier provided by a Hartley oscillator (Fig. 20.9). In this manner an r-f signal for remote

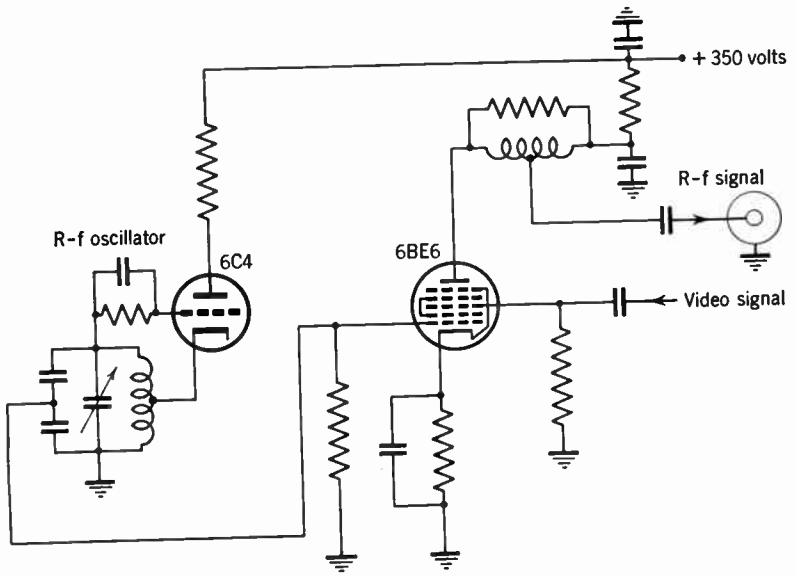


Fig. 20.9. Generator of R-F Signal for Remote Receivers.

receivers is generated. Synchronization is provided by the blanking pulses, whose amplitude is always made sufficiently great so that they correspond to "blacker-than-black" signals.

The synchronizing generator consists essentially of five blocking oscillators. The first of these is tuned to twice horizontal frequency

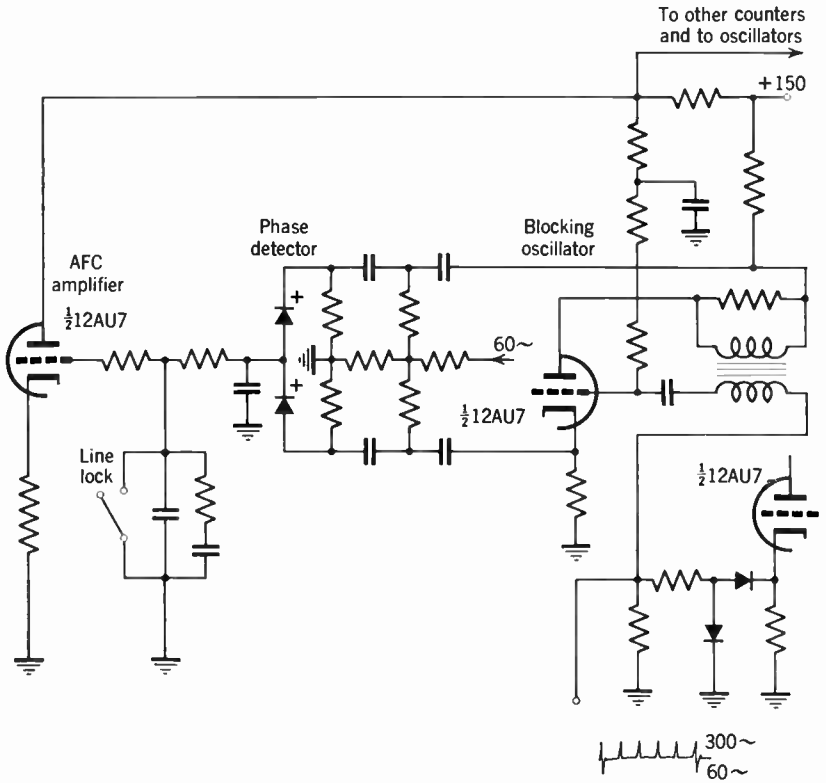


Fig. 20.10. Last Frequency Divider Stage of the Synchronizing Generator and Phase Detector.

(31.5 kilocycles). It controls, on the one hand, through a buffer amplifier, the horizontal pulse generator, on the other, a series of blocking oscillators tuned to one-fifteenth, one-seventh, and one-fifth the frequency of the preceding oscillator. Thus the final oscillator generates pulses at 60 cycles, the vertical frequency. The triggering pulses are derived from the cathode resistor of the preceding stage and are applied to the grid resistor; feedback is prevented by a pair of crystal rectifiers connected with the proper polarity (Fig. 20.10). Finally, a

phase detector consisting of two selenium crystals generates a d-c bias for the preceding oscillators, which depends on the phase difference between the vertical pulses and the power-line oscillations. This bias modifies the oscillator frequency so as to bring the vertical frequency into synchronism with the power-line frequency.

Vertical and horizontal blanking pulses are derived from the outputs of the respective blocking oscillators, mixed in a mixer stage, and applied, as already mentioned, to the video signal in the next-to-the-last video amplifier stage as well as, by cable, to the blanking amplifier

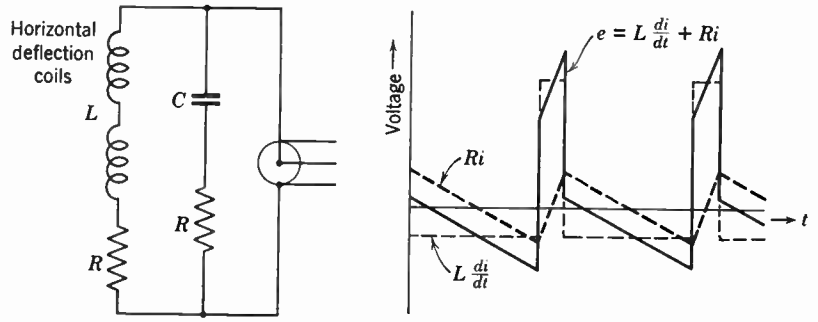


Fig. 20.11. Constant-Resistance Network Incorporating Horizontal Deflection Coil in Camera and Required Deflection Voltage Transmitted over Cable.

in the camera. The original pulses also control vertical and horizontal deflection generators for both kinescope and camera, the high voltage for the kinescope being obtained in the usual manner from the flyback pulse in the horizontal generator. However, the distortion-free transmission of the horizontal pulses over the long cable connecting the camera and the monitor demands that the cable be terminated by its characteristic impedance for the entire frequency range represented in the horizontal pulses. This may be realized by making the horizontal deflection coil of the camera, with inductance L, part of the constant resistance terminating network shown in Fig. 20.11. Here the resistance R is made equal to the characteristic impedance of the cable (approximately 50 ohms), and the capacitance C is adjusted so that $\sqrt{L/C} = R$. The voltage pulse which must be applied to the cable to yield sawtooth current in the coil is now, in view of the presence of the series resistance R, not a square pulse, but a sawtooth pulse superposed on the square pulse.

The total tube complement of the Industrial Television System just described is 23, *including* the camera tube and kinescope as well as

regulator tubes in the voltage supply. Various accessories are provided to adapt it for special applications. Thus a water-cooled lens protects the camera from extreme furnace temperatures; an explosion-proof housing permits operation in explosive atmospheres; a weather-proof housing provided with windshield wipers protects the camera from the elements in outdoor use; and a pan and tilt unit permits a panoramic survey of large angular viewing fields.

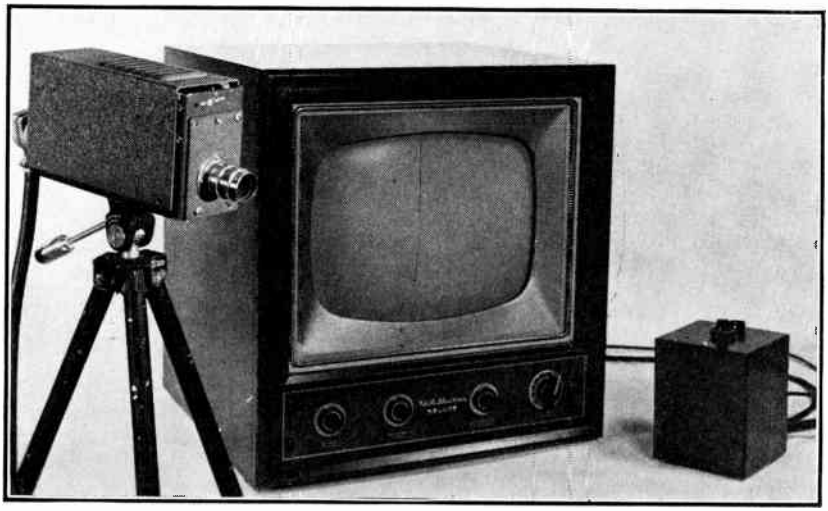


Fig. 20.12. Standard Television Receiver with Camera Attachment and Control Box.

Even the system last described is too costly, however, to give industrial television the wide distribution which its potential usefulness merits. A further step has been taken to bring it within reach of the classroom, the small business establishment, and, eventually, the home. This step consists essentially in converting the Vidicon camera into an adjunct of the standard home television receiver, from which it derives deflection signals, blanking pulses, and operating voltages.* The complete system, including a control box for the camera voltages, is shown in Fig. 20.12. The control box is connected with the receiver by means of a series of tube-base adapters and plugs, the camera to the control box by a cable which may be several hundred feet in length.

The amplification which must be provided in the camera is minimized by transmitting the signal on a regular television channel, so

* See Zworykin, Flory, Pike, and Gray, reference 23.

that the r-f and i-f amplifier, as well as the video amplifier, of the receiver may be utilized. The regular station selector knob of the receiver thus permits instantaneous switching from a broadcast program to the closed-circuit camera, and vice versa.

Figure 20.13 shows the camera circuit. The first two double tubes amplify the Vidicon signal to a level of the order of 2 volts, whereas

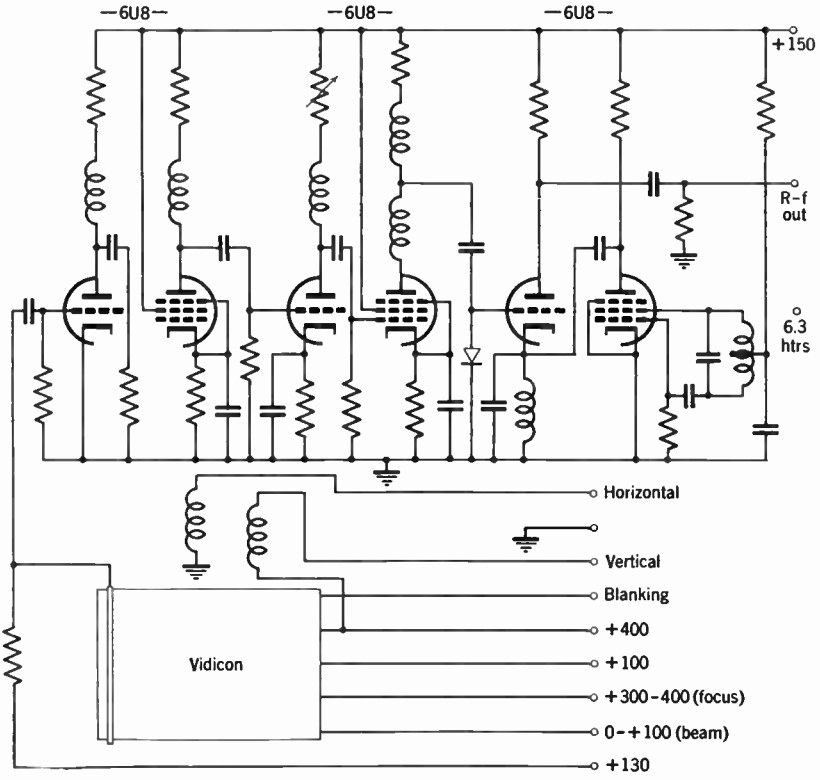


Fig. 20.13. Circuit of Camera Attachment.

the last tube contains an oscillator electron-coupled to a modulator section; the oscillator is tuned to the video carrier frequency of a television channel unused in the reception area in question. The modulated carrier generated in this fashion is coupled to the antenna circuit of the receiver through an attenuating pad to prevent interference with the normal operation of the receiver.

The control box connections are shown in Fig. 20.14. Vertical deflection is provided by simply connecting the camera coils in series with

the low side of the receiver coils. Horizontal deflection is obtained from a transformer on the low side of the receiver horizontal coils, permitting a single-wire connection, even though the horizontal coils of the receiver are commonly not returned to ground.

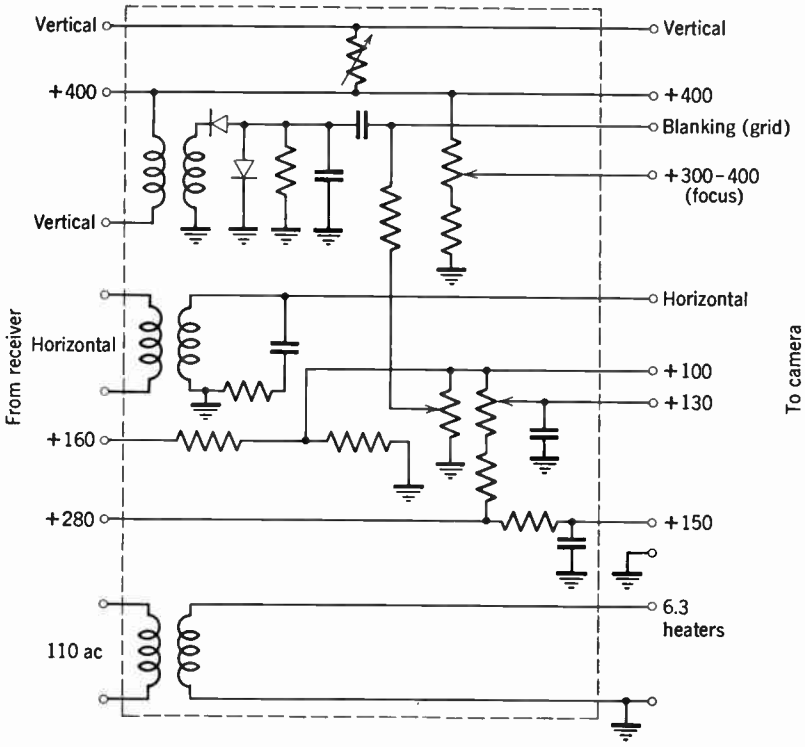


Fig. 20.14. Circuit of Control Box.

Horizontal blanking is obtained from a pulse of the order of 10 volts which appears across the camera horizontal coils; this pulse, which is made positive, is applied to the cathode of the Vidicon. Vertical blanking is generated in the control box by using a blocking oscillator transformer with a pulse stretching and clipping circuit as shown in Fig. 20.14. The negative pulses so obtained are sent to the camera on the Vidicon grid lead. A small amount of these blanking signals may also be mixed into the video signal to provide a means of synchronizing additional receivers.

Although Fig. 20.14 represents a control box for a single camera attachment, the circuits can readily be modified for the operation of

a number of cameras in conjunction with a single receiver. In this instance individual beam current and focus controls are provided for each Vidicon camera; in addition, a selector switch serves to connect power and deflection to the desired camera. The possible uses to which such a multiple camera system might be put in the home are represented graphically in Fig. 20.15. The extraordinary simplification of

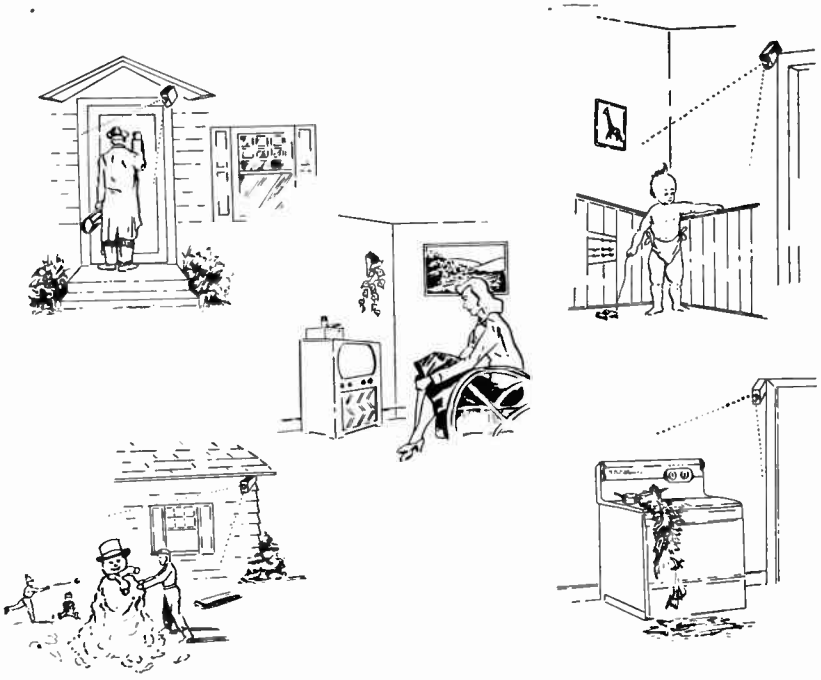


Fig. 20.15. Multiple-Camera Installation in the Home.

the system, which demands only three double tubes in addition to the Vidicon for each camera, makes closed-circuit television available for a wide range of applications which appeared out of the question heretofore.

A system intermediate to the two just described is the RCA TV Eye. Its camera is essentially identical with the one described in connection with the camera attachment for a receiver. The power and deflection for the camera, however, are derived from a small unit operating from the power line. The output from the camera is, as before, an r-f signal which is applied to the antenna terminals of a receiver serving as monitor. The system has the drawback of higher cost, arising from

the added power and deflection circuits, with the compensating advantage that it can be used with any receiver, without thought of special adapters.

A unit similar to the TV Eye, also utilizing a standard receiver as monitor, is the Dage Television Camera.* Here, however, the power supply and deflection circuits, as well as all controls, are located on the camera. In addition, a small monitor kinescope is mounted on the camera for adjustment purposes.

20.4 The Television Microscope. The effectiveness of conventional light microscopy is limited not only by the finite resolving power of the instrument, set by the wavelength of the illuminating radiation, but also by inadequate contrast between different structural elements of the specimen. This fact applies particularly to the study of organic materials at high magnification. Selective staining and phase microscopy frequently aid in the attainment of the desired differentiation. They are not universally effective, however, and leave ample room for other methods.

The television microscope can provide enhanced contrast in two ways: first, by electronic contrast enhancement and, second, by the employment of invisible radiations, particularly in the near ultraviolet, to which the eye is insensitive. Electronic contrast enhancement in combination with greater ease of viewing and convenient adjustment of the brightness level of the image make it possible to follow closely changes in living cells observed at high magnification. Professor A. K. Parpart † has successfully used this method to clear up questions of internal structure in red blood corpuscles. The great advantages which the same technique offers in the teaching of bacteriology, histology, and other branches of microbiology are obvious.

The second method of contrast enhancement depends on the fact that many organic materials have strong distinctive absorption bands in the near ultraviolet, facilitating differentiation. Photography with high-contrast emulsions or plates sensitive to invisible radiations can, it is true, accomplish the same ultimate end. However, the time delay implicit in the photographic process prevents continuous observation and greatly slows the examination of extended specimens. The observation of a fluorescent screen on which the invisible image is projected, on the other hand, is unsatisfactory with respect to resolution and brightness.

* Dage Electronics Corporation, Beech Grove, Indiana.

† See Parpart, reference 27.

The earliest realization of a television microscope—for the purpose of studying specimens illuminated by ultraviolet radiation—employed an iconoscope with ultraviolet-transmissive window.* At a much later date, R. C. Webb developed a flying-spot microscope, in which the face of a flying-spot tube was imaged on a microspecimen through the microscope, and the transmitted light was collected by a multiplier phototube (Fig. 20.16).† Young and Roberts,‡ in England, have obtained good results with a system of this type.

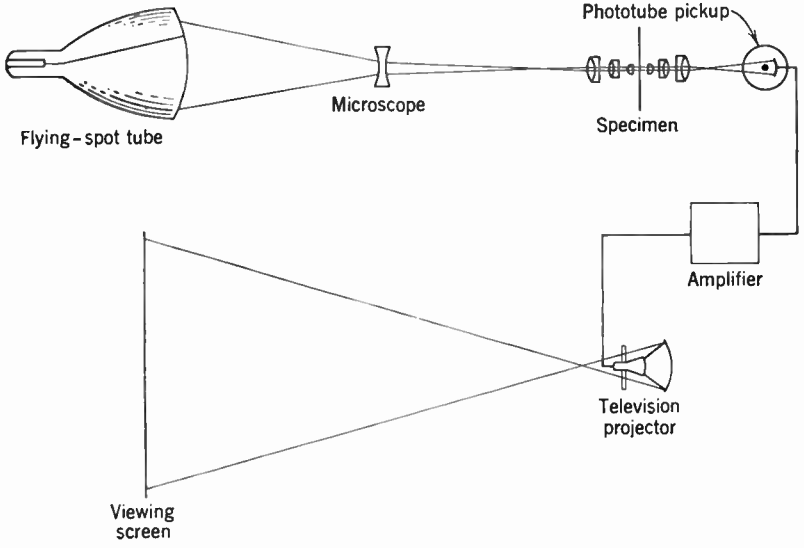


Fig. 20.16. Flying-Spot Microscope (Schematic).

Although a flying-spot microscope is very efficient in the utilization of the light which reaches the specimen, it is limited in the intensity of illumination, particularly when the illumination is to be restricted to a selected wavelength range. Noise-free images at high magnifications are much more readily achieved with a storage-type camera tube such as the Vidicon, employed in conjunction with a high-intensity source, exemplified by a mercury arc.

The industrial television equipment described at the end of the last section is readily combined with a standard microscope to form a highly practical television microscope (Fig. 20.17). If the standard optics of

* See Zworykin, reference 24.

† See Zworykin and Ramberg, reference 25, p. 383.

‡ See references 26 and 35.

the microscope are replaced by quartz-fluorite and/or reflective optics and a simple quartz prism monochromator with a medium-pressure mercury arc source is used as illuminator, it becomes possible to compare images obtained at different selected wavelengths in the ultraviolet and visible (Fig. 20.18). The only modification in the Vidicon camera is the replacement of the standard Vidicon with a tube pro-

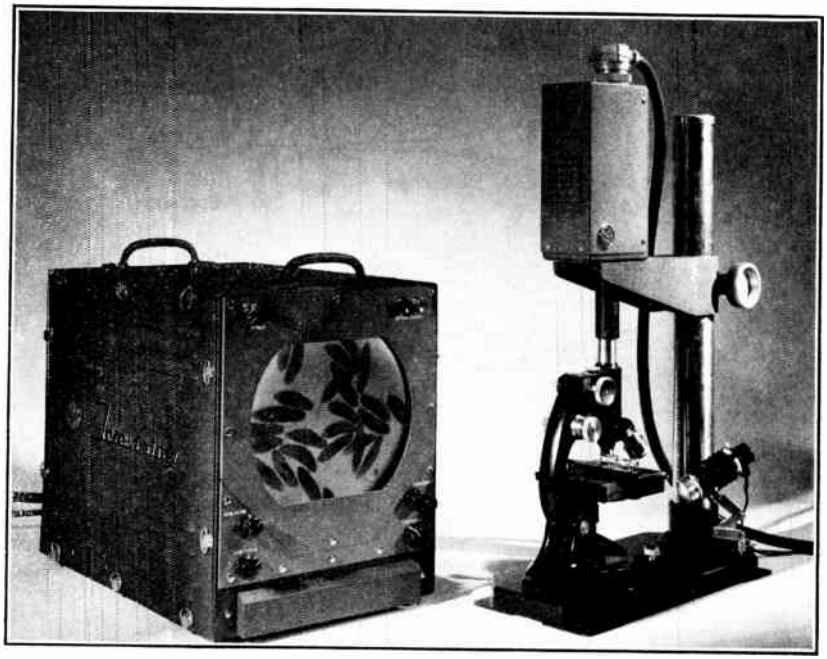


Fig. 20.17. Television Microscope.

vided with an ultraviolet-transmitting face plate and a selenium target. Figure 20.19 shows the spectral response of such a tube in the ultraviolet. The difference in the appearance of a specimen of unstained kidney tissue illuminated with 4000 Å radiation and 2537 Å radiation is indicated in Fig. 20.20. The strong absorption of nucleic acid at the shorter wavelength causes the cell nuclei to stand out sharply in the second picture.

Ultraviolet television microscopy has also been employed effectively for the visualization of changes in cancer cells produced by irradiation and therapeutic agents by Dr. G. Z. Williams of the Virginia Medical College and the National Institutes of Health.* Television microscopy

* Papers presented before the American Association for Cancer Research and the Federation of Biological Societies in Chicago, April, 1953.

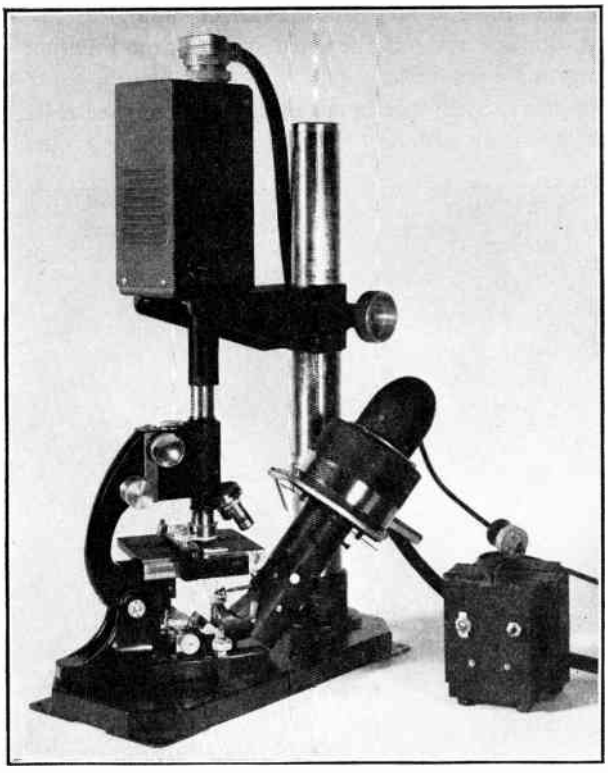


Fig. 20.18. Television Microscope for the Ultraviolet.

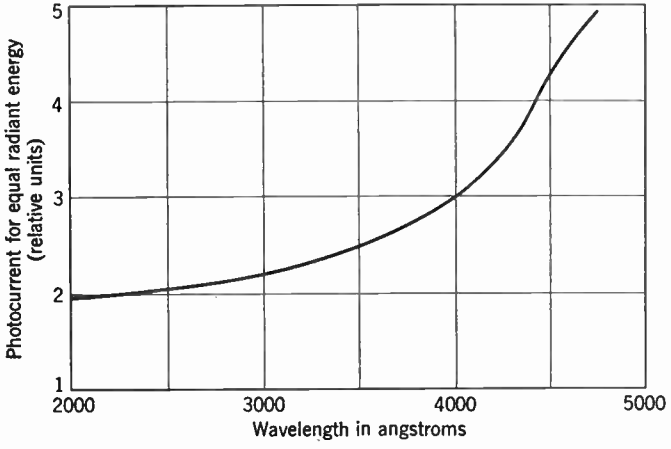
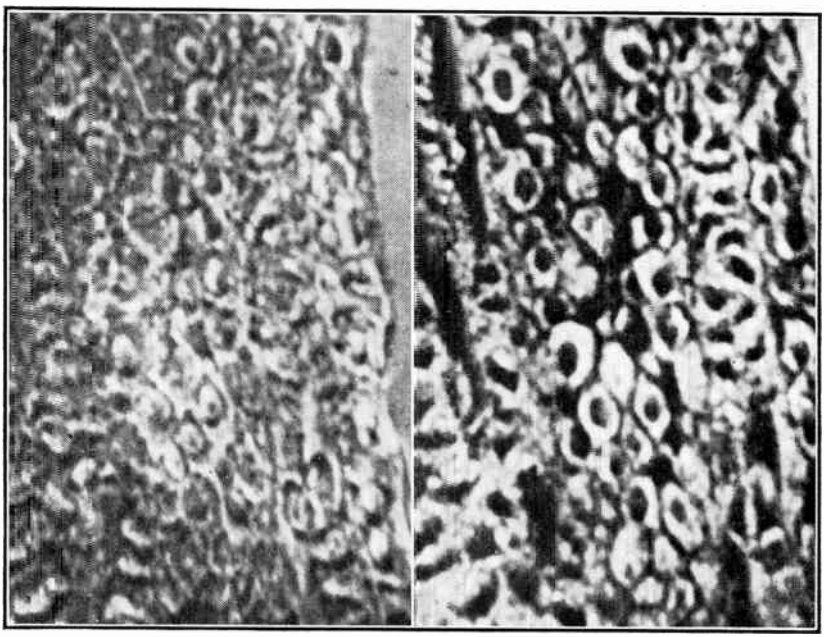


Fig. 20.19. Spectral Response in the Ultraviolet for an Ultraviolet-Sensitive Vidicon.

permits the examination of great masses of biological materials with ultraviolet radiation without the great loss of time inherent in conventional photographic methods.

The differential absorption in the ultraviolet may be recognized more directly by assigning different colors to different wavelengths in the



(*a*) (*b*)

Fig. 20.20. Photographs of Identical Sections of Unstained Kidney Tissue Illuminated at 4000 Å (*a*) and 2537 Å (*b*) Obtained from Kinescope Screen of Television Microscope.

ultraviolet and superposing, e.g., three partial images formed with three wavelengths in their corresponding colors. Such a color representation of microspecimens has been realized photographically by Land and his associates.* Apparatus achieving a similar result with the television microscope is shown in block diagram in Fig. 20.21. A mercury arc light source and gates shifting the video signal from one gun of a tricolor kinescope to the next are pulsed in synchronism with the rotation of a wave-length selector in the illuminating system of the microscope. The selector consists of a three-sector mirror, the parallel sectors being recessed by different amounts so that, for each, radiation of a particular

* See Land et al., reference 28.

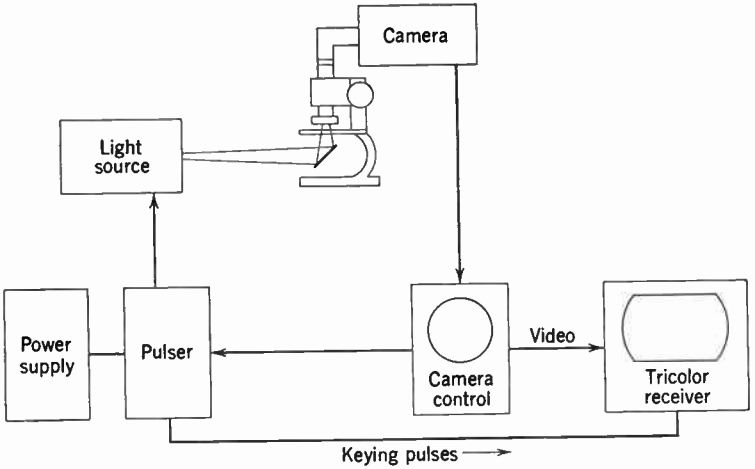


Fig. 20.21. Block Diagram of Television Microscope Translating Ultraviolet Absorption Differences into Color Differences.

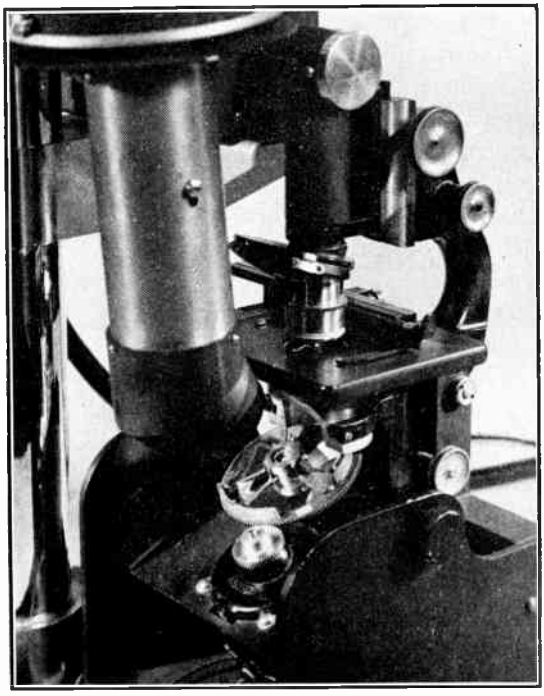


Fig. 20.22. Detail of Ultraviolet "Color" Microscope Showing Wavelength-Selecting Mirror Disk.

wavelength is concentrated on the exit slit of the illuminator (Fig. 20.22). Although in the experimental system the color alternation frequency was too low to give an even reasonably flicker-free picture, the effectiveness of the method in indicating differential absorptions in the ultraviolet could be readily demonstrated.* It should be noted that the above application requires the use of a reflective microscope objective, since even quartz-fluorite achromats are adequately corrected for only two wavelengths in the ultraviolet.

Through the addition of some auxiliary circuits, the television microscope can be transformed into an effective blood cell counter.† A standard counting chamber containing the cell suspension is observed at a fixed magnification with dark-field illumination in the television microscope. At the same time, the video signal is applied to an electronic counter, which counts the pulses corresponding to transits of the scanning spot through individual cell images. Since a large cell will register more counts than a small one, and the average cell size differs for different samples, the count-rate meter is adjusted for average cell size, determined by comparing a variable-length standard pulse with the measured pulses and locating the point of maximum coincidence.

20.5 Special Forms of Industrial Television Equipment. Standard industrial television equipment can be modified in various ways to adapt it for special purposes. Thus, field-sequential color television has been utilized for the presentation of surgical operations, the transmission of fashion shows in department stores, and for other purposes. The inadequate resolution and marked flicker effects observed with this system operating under broadcasting standards are overcome by increasing the bandwidth of the transmission and the frame frequency, respectively. Compatibility is, of course, of little importance with closed systems. The Vericolor chains presented by Remington-Rand in cooperation with the Columbia Broadcasting System and discussed in some detail in section 19.7 employ a bandwidth of 10 megacycles, whereas Telecolor chains designed by DuMont utilize 18 megacycles. The Telecolor system ‡ is a 525-line, 180 fields-per-second system consisting of five portable units (Fig. 20.23), color disks driven by synchronous motors being provided in both the image-orthicon camera and the 7-inch color monitor. The color signals are separated in a control unit, to permit individual gain adjustment for the three partial pictures, and are then recombined, along with blanking, shading, and

* See Zworykin, Flory, and Shrader, reference 29.

† See Zworykin and Flory, reference 16, and Flory and Pike, reference 36.

‡ See Smith, Olson, and Cotellessa, reference 30.

synchronizing signals, before application to the distribution amplifier (for additional receivers) and the color monitor. It should be noted that the adoption of field-sequential methods in these early examples of industrial color television has been based primarily on lower costs. It may be anticipated that, with the advance of the art, preference will be given here to simultaneous color systems just as in color television.

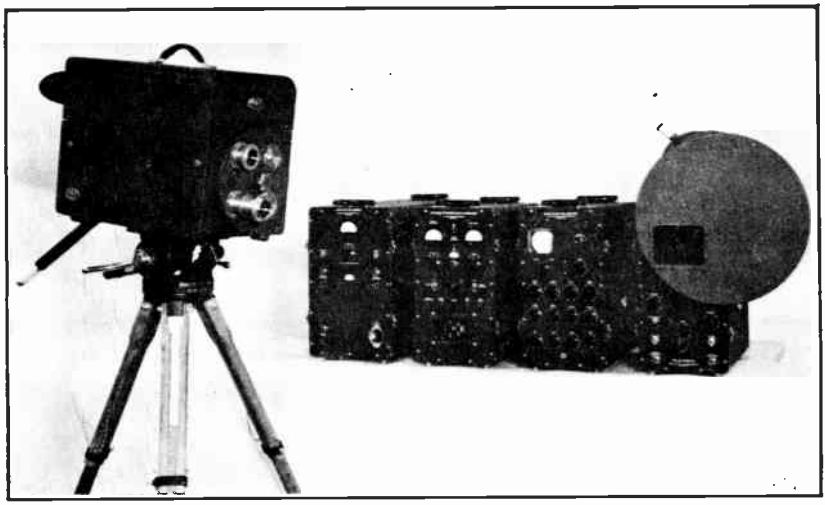


Fig. 20.23. DuMont Telecolor Chain Incorporating Camera, Camera Control Unit, Synchronizing Generator, Low-Voltage Power Supply, and Color Monitor. (Courtesy of Allen B. DuMont Laboratories, Inc.)

A detailed study of stereo television techniques was carried out by Johnston, Hermanson, and Hull of the Argonne Laboratories Remote Control Engineering Division, employing a chain provided by DuMont.* The problem consists of presenting to each eye of the observer the image which it would see if the observer were stationed at the camera location. Either a sequential or a simultaneous method may be employed for this purpose. In the sequential method moving mirrors and shutters are used to translate the effective position of the camera lens and to mask one or the other eye of the observer in synchronism. In the simultaneous method the two images are transmitted simultaneously and observed individually by each eye through some type of viewing stereoscope.

The sequential method proved unsatisfactory because of flicker and lag effects. Consequently, a simultaneous method was adopted, em-

* See reference 6.

ploying a mirror system to project the images formed by two objectives side by side on the photocathode of the image orthicon (Fig. 20.24).

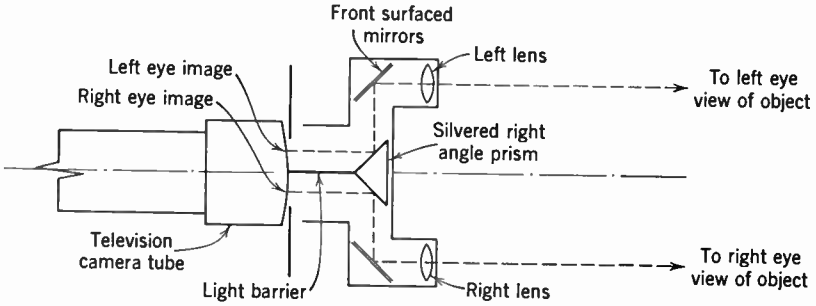


Fig. 20.24. Projection of Two Images on Photocathode of Camera Tube in Stereo System Employing Single Image Orthicon. (Johnston, Hermanson, and Hull, reference 6.) (Courtesy of *Electrical Engineering*.)

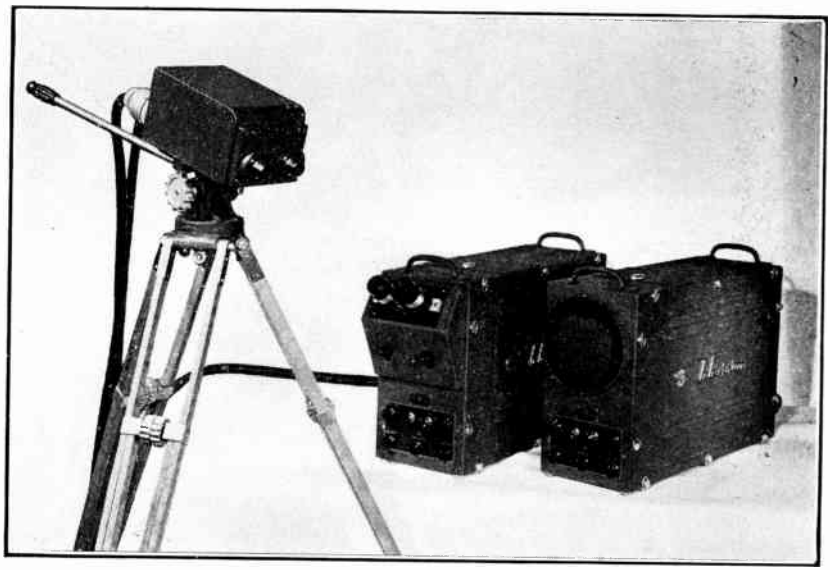


Fig. 20.25. RCA Stereo Television Chain Consisting of Camera, Stereo Monitor, and Control Unit.

In the RCA system * (Fig. 20.25) the resolving capabilities of the industrial television chain are utilized more fully by providing two Vidicons and two parallel video channels. The figure shows a stereo

* See Zworykin, reference 31.

monitor employing two miniature kinescopes along with the camera and control unit. The three-dimensional image is viewed preferably through polaroid spectacles on a receiver in which the two kinescope images, filtered by complementary polarizing screens, are superposed by a semitransparent mirror. A relatively compact receiver of this type is shown in Fig. 20.26.

Finally, there are numerous applications of industrial television equipment for which a cable connection proves impractical. In all

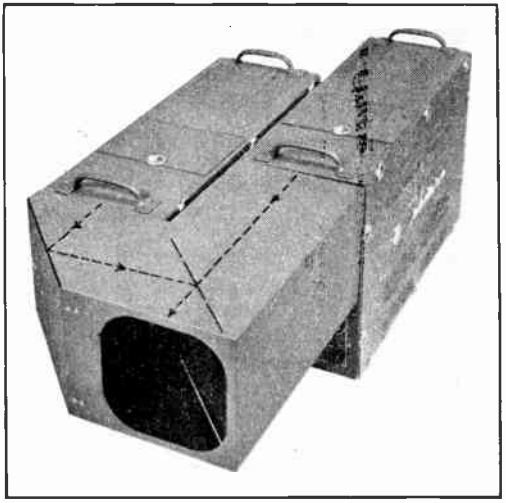


Fig. 20.26. A Table Model Stereo Receiver.

these cases a transmitter must form part of the camera and monitor. A portable television station suited for such applications is described in Chapter 21.* It is a logical further development of RCA's general-purpose television equipment described in section 20.3.

REFERENCES

1. L. M. Exley, "Television in Power Stations for Direct Viewing of Furnace Conditions," *Mech. Eng.*, Vol. 73, pp. 1008–1009, 1951.
2. G. W. Bice, "Television Proves Itself in the Plant," *Power*, Vol. 93, pp. 120–121, September, 1949.
3. "Utiliscope," *Mech. Eng.*, Vol. 73, p. 510, 1951.
4. "Television Increases Efficiency of New Fisher Body Plant," *Automotive Industries*, Vol. 103, pp. 47, 92, 96, Dec. 15, 1950.
5. "Industrial Television," *Mech. Eng.*, Vol. 71, p. 596, 1949.

*See Flory, Pike, Dilley, and Morgan, reference 32.

6. H. R. Johnston, C. A. Hermanson, and H. L. Hull, "Stereo Television in Remote Control," *Elec. Eng.*, Vol. 69, pp. 1058–1062, 1950.
7. "Television Monitors Dangerous Operation," *Electronics*, Vol. 21, p. 168, April, 1948.
8. D. G. Fink and C. L. Engleman, "Electronics at Bikini," *Electronics*, Vol. 19, pp. 84–89, November, 1946.
9. "Underwater Television," *J. Brit. Inst. Radio Engrs.*, Vol. 12, p. 266, 1952.
10. "Television System for Industrial Applications," *Electronics*, Vol. 20, pp. 138, 140, 156, May, 1947.
11. "Can Industry Use Television?," *Fortune*, Vol. 44, pp. 120–123, September, 1951.
12. "Army Television as Training Aid," *Electronics*, Vol. 24, pp. 148, 150, November, 1951.
13. "Televised Microscopy," *J. Franklin Inst.*, Vol. 251, pp. 411–413, 1951.
14. "Doctors See Operating Table Technique by Color Television," *Elec. Eng.*, Vol. 69, p. 1134, 1950.
15. V. K. Zworykin, "TV Camera Tubes," *J. Soc. Motion Picture Television Engrs.*, Vol. 55, pp. 227–242, 1950.
16. V. K. Zworykin and L. E. Flory, "Television in Medicine and Biology," *Elec. Eng.*, Vol. 71, pp. 40–45, 1952.
17. V. K. Zworykin, "Flying Torpedo with an Electric Eye," *RCA Rev.*, Vol. 7, pp. 293–302, 1946.
18. R. D. Kell and G. C. Sziklai, "Miniature Airborne Television Equipment," *RCA Rev.*, Vol. 7, pp. 338–357, 1946.
19. R. W. Sanders, "Industrial Television," Radio Electronic Eng. Ed. of *Radio and Television News*, Vol. 12, pp. 3–6, 30–31, February, 1949.
20. R. W. Sanders, "Closed-Circuit Industrial Television," *Electronics*, Vol. 23, pp. 88–92, July, 1950.
21. R. E. Barrett and M. M. Goodman, "Simplified Television for Industry," *Electronics*, Vol. 20, pp. 120–123, June, 1947.
22. R. C. Webb and J. M. Morgan, "Simplified Television for Industry," *Electronics*, Vol. 23, pp. 70–73, June, 1950.
23. L. E. Flory, W. S. Pike, and G. W. Gray, "A Vidicon Camera Adapted for Television Receivers," *Electronics*, Vol. 25, pp. 141–143, January, 1954.
24. V. K. Zworykin, *Electric Microscope*, 1° Congresso Internazional di Elettro-Radio-Biologia (Venice, Italy), Vol. 1, pp. 672–686, 1934.
25. V. K. Zworykin and E. G. Ramberg, *Photoelectricity and Its Application*, Wiley, New York, 1949.
26. J. Z. Young and F. Roberts, "A Flying-Spot Microscope," *Nature*, Vol. 167, p. 231, 1951.
27. A. K. Parpart, "Televised Microscopy in Biological Research," *Science*, Vol. 113, pp. 483–484, 1951.
28. E. H. Land, E. R. Blout, D. S. Grey, M. S. Flower, H. Husek, R. C. Jones, C. H. Matz, and D. P. Merrill, "Color Translating Ultraviolet Microscope," *Science*, Vol. 109, pp. 371–374, 1949.
29. V. K. Zworykin, L. E. Flory, and R. E. Shrader, "Television Microscopy in the Ultraviolet," *Electronics*, Vol. 24, pp. 150–152, September, 1952.
30. H. R. Smith, A. L. Olson, and R. F. Cotellessa, "A Color Television System for Industry," *Elec. Eng.*, Vol. 70, p. 517, 1951.

31. V. K. Zworykin, "Industrial Television and the Vidicon," *Elec. Eng.*, Vol. 69, pp. 624–627, 1950.
32. L. E. Flory, W. S. Pike, J. E. Dilley, and J. M. Morgan, "A Developmental Portable Television Pickup Station," *RCA Rev.*, Vol. 13, pp. 58–70, 1952.
33. J. E. Telfer, "A Survey of Some Applications of Television in Industry, Scientific Research, and Education," *Proc. I.R.E. (Aust.)*, Vol. 13, pp. 407–425, 1952.
34. R. C. G. Williams, "Industrial and Professional Applications of Television Technique," *Proc. Inst. Elec. Engrs.*, Vol. 99, Part III-A, pp. 651–664, 1952.
35. F. Roberts and J. Z. Young, "The Flying-Spot Microscope," *Proc. Inst. Elec. Engrs.*, Vol. 99, Part III-A, pp. 747–757, 1952.
36. L. E. Flory and W. S. Pike, "Particle Counting by Television Techniques," *RCA Rev.*, Vol. 14, pp. 546–556, 1953.

21 Practical Television Systems

21.1 Television as a Major Industry. During the decade following 1940 television in the United States experienced a phenomenal growth, even in spite of the retardation which was the inevitable result of the Second World War. As spectacular as was this growth, it was by no means unpredicted. Many of those working in the field were well aware of the potentialities of the medium and have not been surprised by the way in which the industry has developed.

In 1940 only half a dozen transmitters, serving at most a few hundred receivers, were in operation. At that time the transmission was on an experimental basis and limited to only a few hours a week. It might well have been noted at the time, however, that the programs broadcast were by no means gauged to the smallness of the audience. Both the scope of material and the equipment employed for outdoor, moving picture and studio transmissions indicated that they were in reality research programs to develop means and methods suitable for the present vast spectator group. Excellently produced and directed stage plays, variety shows, concerts, and educational programs were transmitted. Spot pickup included sporting events such as baseball, football, and boxing as well as such newsworthy events as political conventions, military exercises, and interviews with people in the national spotlight.

At the time of the entry of the United States into the war, television was just starting to be put on a commercial basis. During the next five years, all the engineering and industrial effort which might have gone into developing the field was, of course, diverted to the grim business of the military. During this time, television broadcasting continued, but only on a scale sufficient to keep the facilities from deteriorating.

At the close of the war, the normal expansion of television began. The following table is a guide to its growth:

Year	Receivers in Use at End of Year (U. S.)
1947	186,000
1948	1,000,000
1949	3,400,000
1950	10,500,000
1951	15,200,000
1952	20,400,000
1953	28,000,000

The enormous rate of growth which is indicated in the table has, of course, been a gratifying vindication of the earlier effort which went into developing the system. It also means that television has already become an important and valuable aspect of our economic life. However, a rate of growth such as this is not without its disadvantages. For one thing, it has forced a number of decisions, both technical and nontechnical, to be made before adequate data were available upon which to base the decision. As illustration, the example might be cited of the decision on the part of industry to standardize on a 10-inch viewing tube. At the time it was made, it seemed very reasonable from an economic and manufacturing standpoint. Data on the public preference were not available. Fortunately, our form of economic system corrects for this type of error very rapidly and with a minimum of overall inconvenience. In less than a year, public pressure reversed the decision, and viewing tube sizes were expanded as rapidly as technology permitted. There is every reason to feel that when such errors are made, they will be similarly corrected as long as television is a free, healthy industry.

Television has reached such proportions that it is of considerable economic importance. Both because of the type of service rendered and the history of the development of television, television manufacture and broadcasting are closely tied to the radio industry. However, their magnitude has become such that the new industry overshadows its parent. At the time of writing (the end of 1953) the plant investment for production of television equipment far exceeds the corresponding investment for radio broadcasting and receiving equipment. The total output of the television industry in 1953 has been estimated at 1.5 billion dollars, approximately twice that of the (civilian) radio industry.*

The television broadcasting service represents an important industry all by itself. Here many thousands of people are employed in

* See reference 1.

handling the very complex problem of picture production and transmission. Each program requires the services of a large group of technically trained personnel to operate cameras, lights, sound and video monitoring equipment associated with the point of origin of the program, as well as the engineers necessary for the transmitter. In addition, each production requires producers, performers, and attendants. When all these activities are multiplied by the number of programs produced each year, and have added to them the formidable amount of time required for rehearsals, it will be seen that full-time employment of a large group of well-trained and talented individuals is involved.

In relation to the economic aspect of television, the employment opportunities which have been opened up to meet the servicing requirements of home receivers should not be neglected. It is difficult at this time to estimate the dollar volume of this business. However, it is quite evident from the earlier chapters that the television receiver is quite a complex piece of equipment and can be repaired and adjusted only by men who not only have the requisite training but who also have the necessary equipment. Therefore, the equivalent manpower back of each television servicing job is greater than that required to service a radio broadcast receiver. With some 28 million sets in use at the end of 1953, the total servicing work may well exceed that for the present existing broadcast receivers.

The indirect economic effect of television, namely, its effect as an advertising medium, is also extremely important. Almost without exception, the advertisers who have used this facility have had immediate and excellent results. One consequence has been that the supply of available video time for this purpose is not nearly adequate to meet the demands. The rapid increase in the total expenditure for television advertising, reaching 580 million dollars * in 1952, is a clear indication of the high value placed upon this medium of advertising. A vivid example of its effectiveness was given when, as a result of a series of video programs, the sponsor received so many orders for his product that his production facilities were entirely inadequate for the demand. He was, therefore, forced to suspend his program, pending the time required to expand his plant and catch up on the backlog of orders.

All the industrial and economic activity outlined in the preceding paragraph was possible only through the expenditure of large amounts on engineering and research. Public demand for television is such

* Quoted in *Television Factbook*, No. 17, July 15, 1953, from *Printers' Ink*.

that increased activity in both fields is inevitable. This means that there is and will continue to be a large demand for men who by training and inclination are suited for work on television development. Opportunities in this field are almost unlimited.

21.2 The Studio. The television studio is the point of origin of a large fraction of all television programs. The studio set is one of the two focal points of the entire complex system whose fundamentals have been outlined in the preceding pages. The other focal point is, of course, the home receiver which was described in an earlier chapter. In principle, perhaps, the whole system should be subservient to the studio set. The studio set in action should tell the story it was created to tell or, even better, be this story quite independent of the television terminal equipment which surrounds it. It would then be the duty of the cameras, the microphones, the lighting, and the rest of the system to pick up this story and relay it to the viewer in the best possible manner. If necessary, engineering details of the system might have to be altered to meet the particular needs of the program. In practice, of course, this cannot be strictly realized. Both the set and the acting must take positive cognizance of the means by which the program reaches the audience. The scenery must be designed to compensate for or minimize the limitations of the cameras. Similarly, the action and speech must be tailored to fit the shortcomings of the system as a whole.

Since the primary objective of this book is to present the basic principles of television rather than the engineering details, it would be out of place to attempt a complete description of the television studio or its equipment. However, a qualitative description of the studio and the manner in which it functions is presented to give the reader a better feeling for the practical equipment which represents the physical embodiment of the principles discussed in the preceding chapters.

The Studio Room. It is not possible, of course, to generalize the description of the room in which the studio is located so as to cover all or even a large fraction of practical cases. In general, the room selected will be governed in its size and architecture by the requirements of the broadcaster, the available accommodations, and the economics of the situation.

The large broadcaster will have a number of studios primarily so that one production can be on the air while others are being set up or taken down. Under these circumstances, the studios do not have to be as flexible as those for a small broadcaster, who has perhaps only one room available for the purpose. The room or rooms selected are

usually large enough to accommodate a number of sets. In general, an attempt is made to put all the sets of the production in one studio. This reduces the number of cameras required to put on the program, since they may move from one set to another. It also simplifies the problem of coordinating the action. In order to accomplish this, the studio room must be arranged in such a way that temporary partitions may be put up and removed with ease and rapidity. The wall and ceiling of the room used must be acoustically treated in order to avoid excessive reverberation and other undesirable spurious sound effects. If possible, the room should be air-conditioned since, frequently, the lighting required for the production produces a large quantity of heat. Normal precautions to avoid fire and electrical hazards must, of course, be taken.

There is generally a control room associated with each studio. This room is slightly elevated with respect to the studio and is partitioned from it with a large glass window so that the people in the control room may view the studio. During the performance, very little use is made of the direct view of the studio and almost entire reliance is placed on the pictures transmitted from the studio to the control monitors. However, during the setting up of the scenery and rehearsing of the program a good deal of use is made of the fact that the entire layout may be seen through the control room window.

Another essential component of a television studio installation of any size is a generously dimensioned storage room for the many props required in television broadcasting. This, by itself, makes the space requirements of television broadcasting radically different from those of sound broadcasting.

The walls and ceiling of the studio room must be provided with the requisite brackets, tracks, hooks, etc., to support the required lighting. Also, the floor must be firm and smooth and sturdy enough to withstand the floor loading of the heavy camera dollies which are moved across it. Needless to say, the studio must be provided with a complex of electrical circuits and wiring to provide the power required for lights, cameras, and other electrical equipment.

Studio System. Even a small television station has a number of sources of program material at its disposal. Smooth, effective program presentation makes it essential that these be carefully coordinated. Figure 21.1 shows a layout suitable for a station of intermediate size. Here the nerve center of the station, the master control console, is placed in the transmitter control room. Picture and sound signals from all available sources, be they local or remote, terminate at the master

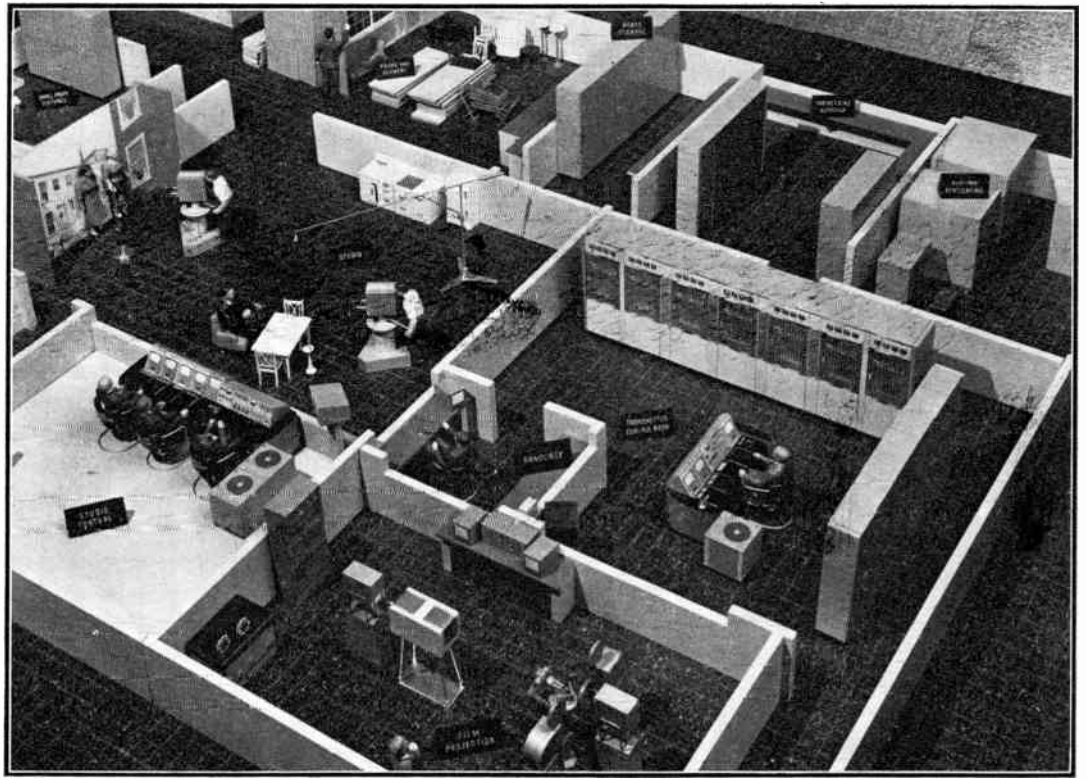


Fig. 21.1. Diagram of Television Studio Installation for Station of Intermediate Size. (Anderson and Hadlock, reference 2.)

control console and may be switched on and off air at will. Monitors give the transmitter engineer a view of the picture being broadcast as well as a preview of the next presentation. Guided by oscilloscopes and meters, he is responsible for maintaining the level of audio and video signals within the proper limits.

In the example shown in Fig. 21.1 the local sources of program material are a "live" studio, a film projection room, and an announce booth. The remote sources may be network programs relayed by coaxial cable or microwave links, spot pickup from a mobile unit, and auditoriums and theaters in other parts of the city connected by wire or radio with the station.

Two cameras, mounted on standard dollies, are shown in the live studio. Larger studios are commonly equipped, in addition, with a camera mounted on a large crane-type dolly for angle shots and shots made from an elevation. A boom-mounted microphone and a stand microphone provide for sound pickup. For news, disk jockey, and announce shows one of the cameras is pointed through the announce booth window.

Seated at the studio control console, flanked by the video operator at the left and the audio operator on the right, the program director has a clear view of the studio as well as the monitors. The two camera control units at the extreme left (Fig. 21.2), with a 10-inch kinescope and a 5-inch oscilloscope, are continuously monitored by the video operator, who adjusts background, gain, and shading. The program director controls the switching unit, which selects the camera signals fed to the master control console for transmission. In addition to the switching monitor, which shows the program on the line at the moment, a preview monitor is provided, which shows the next picture to be switched in. The remote-control section, for operating the film and slide projectors and cameras and switching in the monoscope channel, is commonly a duplicate of a similar section on the master control console. Finally, the audio operator at the consolette "rides" the gain of the audio amplifiers, guided by sound level meters, so as to maintain the sound at the proper level as the actors move about the set. He is also in charge of the turntables for recorded music and transcribed announcements.

Two film camera chains are provided in the film projection room. One of the film cameras accepts pictures from two 16-millimeter projectors and a slide projector through an optical multiplexer. Film projectors are always employed in pairs to permit smooth change-over

from one reel to the next. The other film camera is employed, in the example shown in Fig. 21.1, with a Gray "Telop," a special projector for large slides, newstape, small objects, etc. As an alternative, it may be used in conjunction with a pair of 35-millimeter projectors or a second pair of 16-millimeter projectors. Whereas the projectors are loaded and adjusted by the film technician, they are started and switched by remote control from the master or studio control console. In larger stations, with more than two film camera chains, a separate film control console is provided.

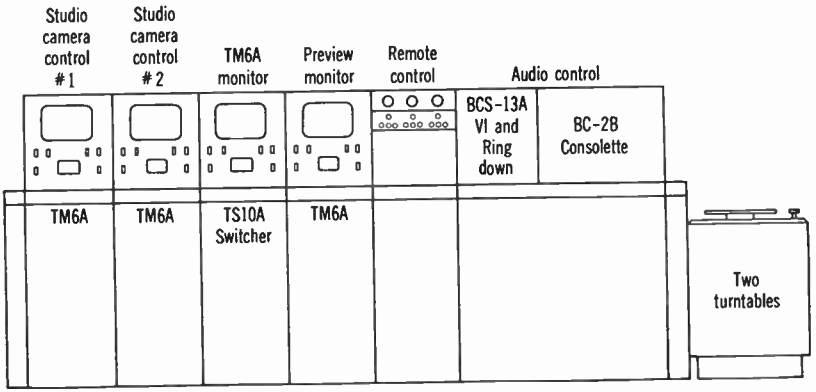


Fig. 21.2. Arrangement of Studio Control Console. (Anderson and Hadlock, reference 2.)

Obviously, for large stations with many studios the control system is considerably more complex, and the functions of the operating personnel are further subdivided. Conversely, small stations, which may have no live studio whatever, have a considerably simplified layout.

21.3 Television Cameras. The most important item of equipment in any television studio or, for that matter, any point of origin of television material is, of course, the pickup camera. In the preceding chapters, detailed descriptions have been given of the camera tubes, associated deflection and focusing equipment, and the amplifiers required to raise the signals to the level of their final use. In the camera these elements are integrated into a unit. There are, of course, many different types of cameras designed for specific purposes. However, a description of three types will suffice to give a general background for this element of the television system. The three to be discussed are a general-purpose studio and field camera, a small portable camera, and a film pickup unit.

The studio and field camera finds employment wherever weight and size are not the most important consideration. This camera includes all the refinements needed for the transmission of as nearly perfect a picture as possible. Some of these refinements must, of necessity, be sacrificed to portability in the small hand-carried camera.

The most widely used television cameras of today (1954) employ an image orthicon as pickup tube. They, therefore, require much less light than the earlier iconoscope cameras. Under normal lighting conditions of 1 to 200 foot-candles an excellent picture may be obtained. A usable picture may be transmitted even at light levels of one foot-candle. This is an important consideration for outdoor pickup, where adverse lighting conditions cannot always be avoided.

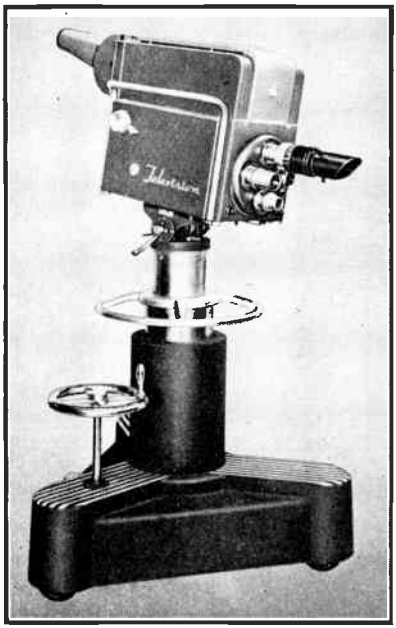


Fig. 21.3. Camera Mounted on Studio Pedestal. (Reisz, reference 3.)

For studio use, the camera is mounted on a heavy pedestal equipped with easy-moving, silent, heavy-duty casters (Fig. 21.3). The casters are arranged so that the camera can be moved in any direction or rotated on the floor and, also, so that it can be locked into position. The pedestal is equipped with an elevator which allows raising and lowering of the camera box. For field use, a lightweight aluminum tripod takes the place of the studio dolly. Here, also, arrangement is made for raising and lowering the camera (Fig. 21.4). In both instances the camera mounting resembles that of a movie camera insofar as controls permit both turning the camera and tilting it.

The camera tube, together with its focusing coils and deflection yoke, is mounted in the lower part of the camera box. A four-position lens turret, manipulated by a handle at the rear of the camera, permits ready control of the angle of view of the camera. Television camera lenses range from a 35-millimeter, $f/3.3$ wide angle to a 25-inch, $f/5.0$ lens for close-ups of distant objects. A typical lens complement for a studio camera consists of a 50-millimeter, $f/1.9$, a 90-millimeter, $f/3.5$,

a 135-millimeter, $f/3.8$, and an 8½-inch, $f/3.9$ lens, giving a range of over 4:1 in the angular field covered. The longer focal length lenses are designed for field use.

In addition to these standard lenses, a number of special-purpose lenses have been developed, such as the 7-inch, $f/1.3$ Television Balow-

Fig. 21.4. Camera Mounted on Aluminum Tripod, Viewed from Cameraman's Position. (Reisz, reference 3.)

star, for extremely low light levels; the 40-inch, $f/8$ Television Reflectar, for long-distance shots; and the Television Zoomar lenses, for continuously varying the focal length from 3 to 13 inches (for studio use) or from 5 to 22 inches (for field use). With the zoom lenses it is possible, e.g., to follow a player across a football field without changing his apparent size or loosing track of him at any moment. The focal length of the Zoomar is controlled by a rod passing through the turret axis.

Focusing of these lenses is accomplished not by moving the lenses themselves but by moving the camera tube and its associated coils with respect to the lens. A focusing knob on the side of the camera

box serves this purpose. The iris opening of the lens, actuated by a small motor and gear train mounted on the lens, is remotely controlled from the camera control position.

The video amplifier and camera deflection chassis are located on either side of the camera tube, hinged side doors giving ready accessi-

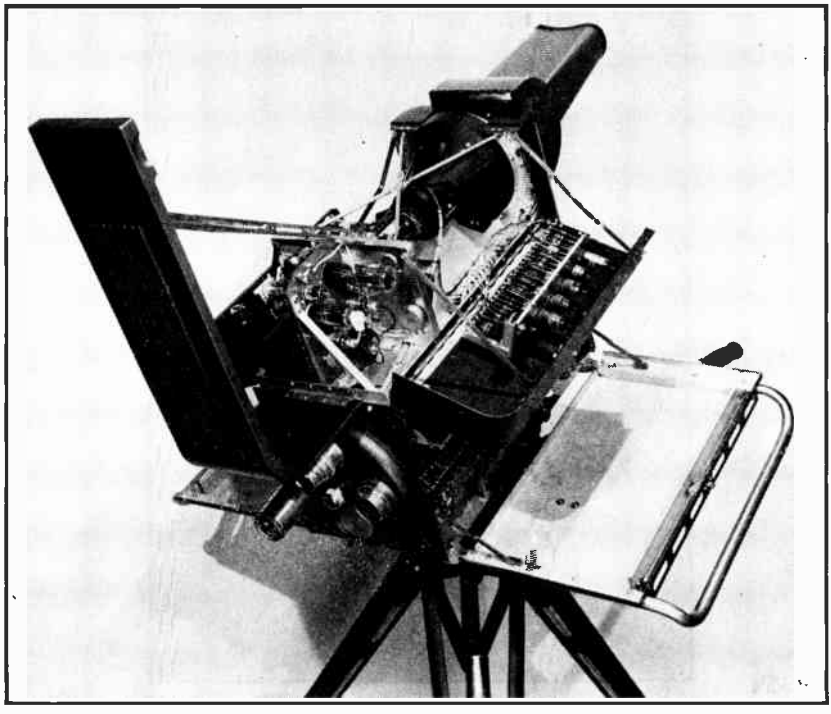


Fig. 21.5. View of Interior of Camera, Showing Viewfinder Compartment at Top. (Reisz, reference 3.)

bility. A response flat to 8.5 megacycles permits full utilization of the image orthicon resolution. A pulse-type high-voltage supply generates the high voltages necessary in the operation of the camera.

Above the main part of the camera and in the form of a plug-in unit is the view finder (Fig. 21.5). The view finder consists of a 7-inch kinescope with a flat aluminized screen which shows the camera man the picture transmitted by the camera tube. A detachable hood shields the tube face from stray light. The necessary kinescope controls are just outside the hood mounting. The electronic view finder enables the cameraman to check exactly the framing and focus of the picture.

The principal camera controls are placed in two vertical rows on the back panel of the camera box, in position readily accessible to the cameraman. They include beam focus, image focus, alignment, centering, and size controls as well as horizontal shading. Additional controls, which require less frequent adjustment, are placed inside the hinged side doors. Tally lights on the front of the camera, in full view of the actors, as well as on the rear, indicate whether the picture is being transmitted from this particular camera.

All electrical connections of the camera are carried by a single cable connected to the camera control unit. In the studio, the control unit is mounted on the studio control console, whereas in the field it is a suitcase unit supplementing the field camera. A portable power supply completes the field camera equipment.

A dual earphone system enables the cameraman to listen in either on the action of the production or to receive instructions from the program director. In the major studios at least one camera is mounted on a crane-type dolly instead of the pedestal described above. The crane-type dolly, as the name implies, has the camera mounted on the end of a massive, movable boom. The camera operator sits on a seat which is also mounted on the boom close to the camera. The boom may be raised or lowered at will to permit a wide variety of angle shots. Again, the dolly which carries the camera is made extremely maneuverable and is operated by a second man who is directed either from the cameraman's position or by the program director in the control room.

A camera unit has been developed that is much smaller and lighter in weight than the general-purpose studio and field equipment described above. It is a truly portable instrument. The camera, its entire control accessories, a microwave transmitter for relaying the picture and sound signals to a base station, and batteries sufficient for an hour and a half of operation can easily be carried by one man.* This type of camera uses a Vidicon pickup tube which can be made very small and light in weight. Its sensitivity, though less than that of an image orthicon, is sufficient for picture transmission under most outdoor conditions. In the model shown in Fig. 21.6 † the electronic view finder is mounted directly above the camera tube, so that the whole camera resembles a small telescope through which the operator looks to observe the scene before the camera.

* See Flory, Pike, Dilley, and Morgan, reference 4.

† See Ohler, reference 5.

A block diagram of the somewhat modified units employed at the National Political Conventions in Chicago in 1952 is shown in Fig. 21.7.* The back pack contains a battery-fed power supply, a simplified synchronizing generator with a 94.5-kilocycle crystal master oscillator, audio and video amplifiers, a modulator in which video and

Fig. 21.6. Vidicon Camera with Backpack Transmitter. (Ohler, reference 5.)

(sound-modulated) synchronizing signals are combined, and an amplitude-modulated 2000-megacycle cavity oscillator with a pencil triode, generating the signal applied to the transmitting antenna. The sound is added to the signal by width modulation of the synchronizing pulses.

At the base station, which may be as much as one-half mile away, the signal is applied to a microwave receiver by a paraboloidal antenna. The sound signal is recovered from the synchronizing pulses, which are replaced by standard synchronization signals from a syn-

* See Ohler, reference 5.

chronizing generator matched in phase to the received synchronizing signals. The resulting video and sound signals can then be utilized for television transmission over regular channels.

The final type of camera to be described is the film pickup camera. For film pickup, the iconoscope has generally been given preference

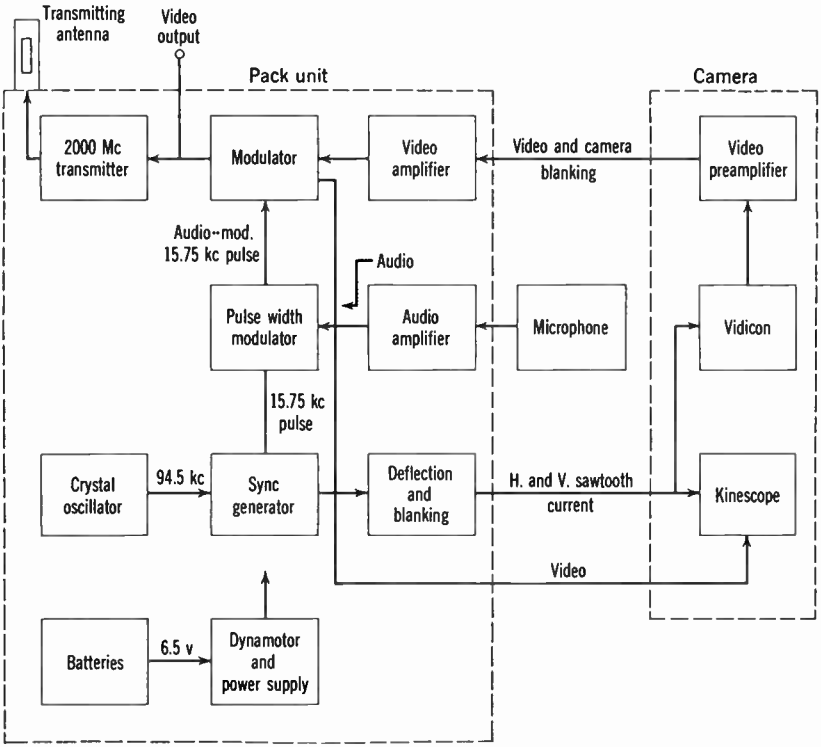


Fig. 21.7. Block Diagram of the Pack Unit and Camera Showing the Interconnections and Signal Path. (Ohler, reference 5.)

over the image orthicon, in view of its excellent resolution and favorable contrast characteristics. More recently (since 1953) the Vidicon has found favor as camera tube for film transmissions.

The film camera proper is relatively simple, since it is stationary and does not contain the optics required for forming the picture on the camera tube target. An iconoscope camera, such as that shown in Fig. 21.8, does require, however, carefully adjusted back-lighting and edge-lighting sources (Fig. 21.9) to eliminate the need of operating shading controls during program transmission.

The equipment for translating the film, timing the exposure, and forming the optical image is part of the film projector. In principle, several solutions are possible. The most straightforward procedure would confine the displacement of the film to retrace time and project alternate motion-picture frames for two and three field periods on the camera tube target. This would match the standard frame frequency of motion pictures (24 per second) to the television field frequency (60 per second) and permit the use of any type of (storage or nonstorage) pickup equipment. On the other hand, completing pulldown of the film in approximately $\frac{1}{1000}$ second would place extreme demands on the mechanical construction of the projector and result in excessive wear of projected film. Consequently, the preferred solution is to flash the picture on the target of the tube during vertical retrace time only and to rely on the excellent storage properties of the iconoscope (or Vidicon) to generate the picture signals during the succeeding field period. The time now available for pulldown depends on whether the intervals between pulldown are made equal ($2\frac{1}{2}$ field periods) or are alternately two and three field periods (Fig. 21.10).

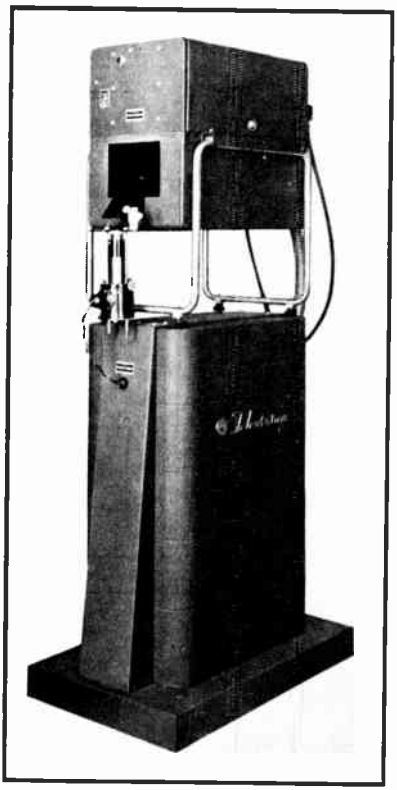


Fig. 21.8. Iconoscope Film Camera. (Herbst, reference 6.)

Both systems are used in practice. The less elaborate projectors, most widely employed, particularly, in smaller studios, have equal-interval pulldown. With $\frac{1}{1200}$ second allowed for exposure, at most $\frac{3}{400}$ second is then available for the pulldown period. With a 3:2 interval arrangement, almost a field period (more precisely, $\frac{19}{1200}$ second) becomes available for this purpose. This reduces film wear and is favored in more elaborate equipment, particularly for 35-millimeter film projectors.

A widely used 16-millimeter projector with equal-interval pulldown is shown in Fig. 21.11. A slot in an 18-inch disk rotating at 3600

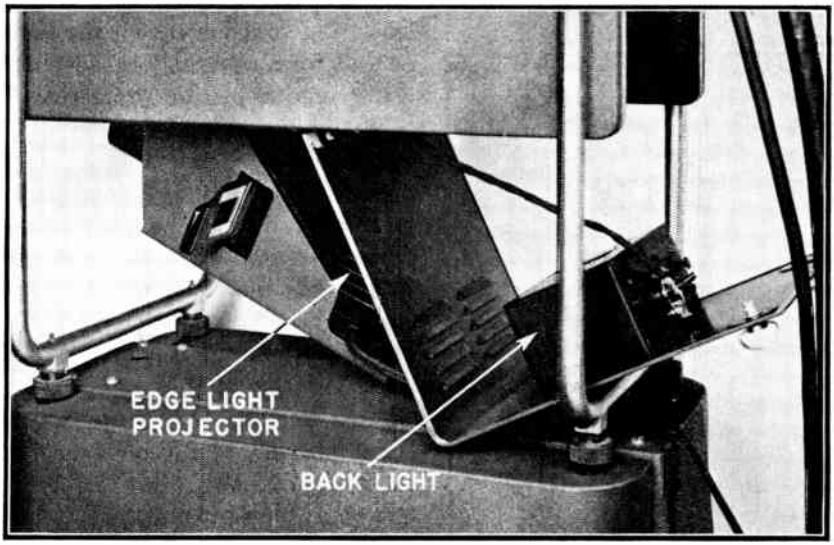


Fig. 21.9. Edge and Back Lighting Units in Film Camera. (Herbst, reference 6.)

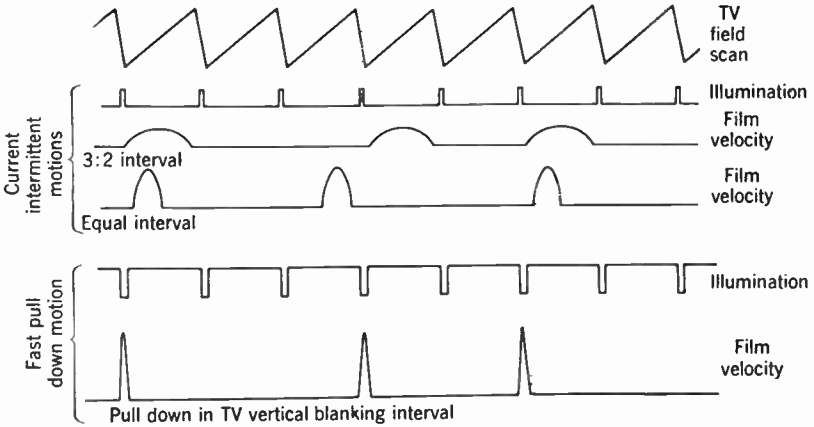


Fig. 21.10. Timing Relationships in Television Intermittent Projector. (Herbst, reference 6.)

rpm admits the light from an airblast-cooled 1000-watt incandescent lamp to the film gate for $\frac{1}{1200}$ second at $\frac{1}{60}$-second intervals. The disk is driven by a three-phase synchronous motor controlled by the

same power source as the synchronizing generator. An automatic phasing adjustment makes the exposure period coincide with retrace time. The picture is projected on the iconoscope target by a 3-inch lens. The sound system is the same as that for a conventional 16-millimeter motion-picture projector.

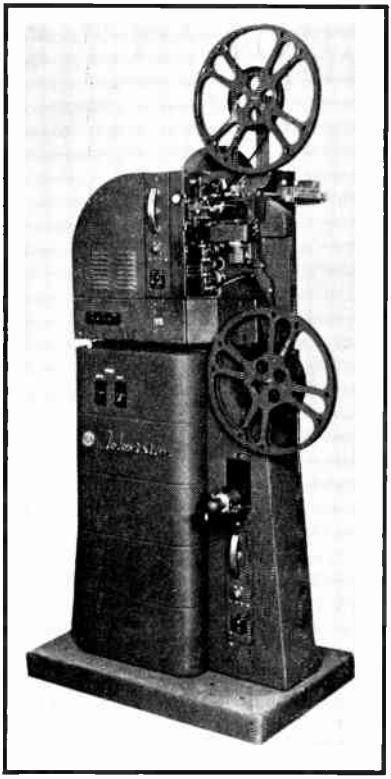


Fig. 21.11. RCA TP-16D Film Projector. (Herbst, reference 6.)

In the 35-millimeter projector a pulsed light source replaces the shutter disk, resulting in extremely quiet operation. The slower pull-down permitted by the 3:2 interval system has the same effect.

As shown in Fig. 21.1, two film projectors are usually placed so that, by way of a pair of front-surface mirrors, they have access simultaneously to a film camera; in addition, a small slide projector mounted above the mirrors can project station announcements on the iconoscope target. The device consisting of the mirrors and the slide projector mounted on a pedestal is called a "film multiplexer."

21.4 Motion-Picture Recording of Television Material. Essentially all the program material produced by the larger television installations is recorded on motion-picture film. It includes the material sent directly to the transmitter, that which is put on cable, and program material produced but not transmitted until later from the film recordings. To record all this material makes necessary the preparing and processing of an enormous amount of motion-picture film.

Any one of the larger television centers probably handles as much or more film in a given length of time as the entire moving picture industry in Hollywood. The National Broadcasting Company system alone, in order to handle its recording, uses about a million and a half feet of film per month. Since NBC records everything in duplicate to reduce the danger of loss of program records due to operational

error or equipment failure, only about two-thirds of this film or a million feet per month should be considered directly useful.

There are a number of reasons why kinescope recordings of all television programs are made. First, these recordings serve as a historical record of the material and as legal documentation of the subject matter. Second, the films of the program material may be used for delayed broadcasting or to repeat the program at a later time. The fact that

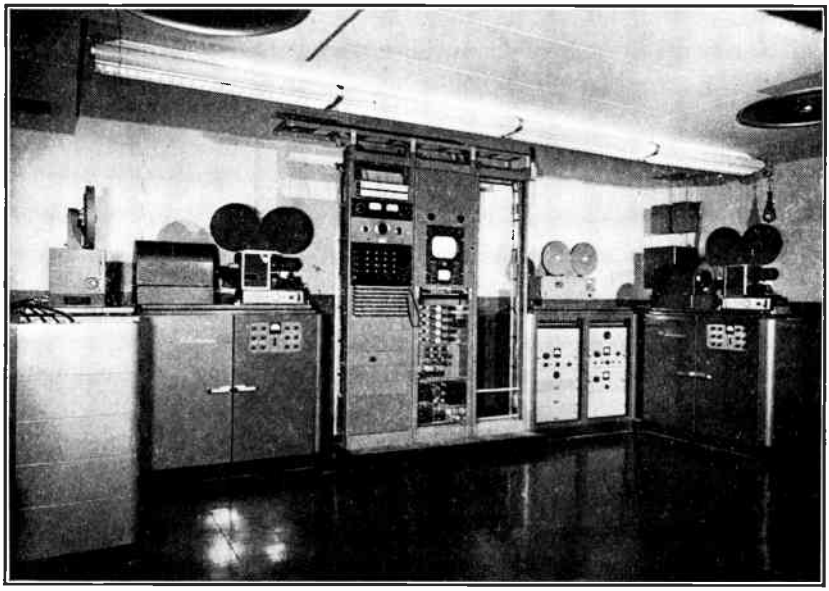


Fig. 21.12. Kinescope Recording Room of NBC Hollywood Television Studio.

there is a time difference of three hours between the East Coast and the West Coast of the United States is by itself sufficient to require repeat broadcasts of important programs, so as to present them during the most favorable listening hours in each location. The films may also be distributed to other stations, which would otherwise have no access to the material; e.g., network programs for stations that are not serviced by the coaxial cable or microwave relay system. Finally, the television records are useful for such internal uses as the sale of programs, auditioning, etc.

Two recording units are shown in the overall view of a television recording room in Fig. 21.12. The recordings are made by reproducing the television image on a small, bright, projection kinescope and photographing these images directly from the face of the tube by means of

a modified moving picture camera. The camera is equipped with a special shutter, which will be explained later, and is accurately synchronized with the television picture. Present-day recordings are made exclusively on 16-millimeter film rather than on standard 35-millimeter photographic material. The 16-millimeter recordings are found to be just adequate in quality so that they do not degrade the record, and the cost is only about one-third as much as for the same recording period on 35-millimeter film. There is the added reason that all 16-millimeter film is acetate-based and consequently relatively fireproof. Thirty-five-millimeter film may be either nitrate- or acetate-based. Since nitrate is extremely inflammable, the Underwriters rules for 35-millimeter cameras are very severe and require the provision of fire walls and other safety equipment, even though the user plans to employ only acetate films. Finally the cost of 16-millimeter camera and projection equipment itself is much lower than that for the equivalent 35-millimeter equipment. As has already been pointed out, the present-day standard 525-line picture corresponds to 483 actual lines in the vertical direction. A good studio picture has a resolution of somewhat over 600 television lines horizontally and a detail contrast range of 10 to 1 or less. To resolve 600 television lines on 16-millimeter film, the emulsion must be capable of resolving 42 lines per millimeter. Standard Super-X panchromatic emulsion is rated at 55 lines per millimeter. However, this standard is determined by the use of a very high contrast ratio chart, and observing the point where the contrast ratio on the film is just greater than unity. Fifty-five lines per millimeter by this type of rating is not adequate resolution for a good television picture. Experimentally, it has been found that the film emulsion must have a rating of 90 lines per millimeter in order not to degrade the television picture. Suitable emulsions with resolving power in excess of this figure and having both sensitivity and gamma suitable for kinescope recording can, however, be obtained.

The television picture derived from a well-made 16-millimeter kinescope recording is essentially indistinguishable from original live studio program material. It has been found expedient to record television pictures on film with a 24-per-second frame rate so as to permit the transmission of television recordings with standard 16-millimeter projection equipment. It is, therefore, necessary to modify the recording camera in such a way as to convert the 30-frame rate of the television picture to the 24 frames per second of the recording. This is done in a somewhat different manner from that described in connection with the pickup camera for moving picture film. The recording camera is

equipped with a shutter which has an open angle of 288 degrees and a closed segment of 72 degrees. The shutter is driven by a 60-cycle synchronous motor geared down to give a shutter speed of 24 revolutions per second. The shutter is phased with the television synchronizing signal in such a way that the shutter is open for the two fields of the first television frame. The shutter then closes for one-half a

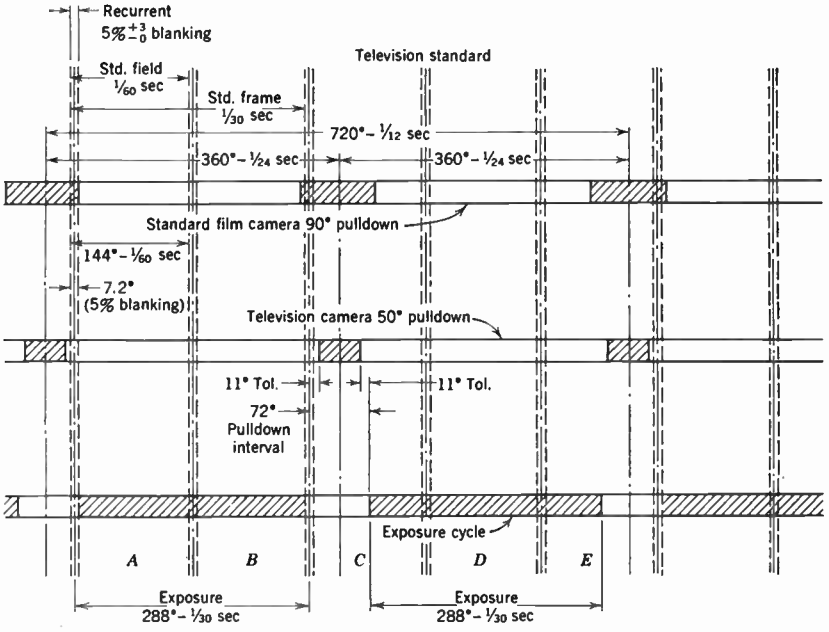


Fig. 21.13. Comparison of Exposure and Pulldown Timing in a Standard Film Camera, a Television Film Projector with Equal-Interval Pulldown, and a Television Recording Camera. (Little, reference 8.)

field, during which time the pulldown occurs and opens for the second half field. It remains open for the next field and the following half field, when it closes again and pulldown occurs. The shutter then opens for the second field of the third television frame. It will be seen that a full cycle of this operation involves five television frames and four motion-picture frames. This gives the required conversion from 30 cycles to the 24 cycles. The accompanying diagram will help to clarify the timing sequence involved (Fig. 21.13).

In order to carry out this recording operation successfully it is obviously necessary to have very carefully and well-designed equipment capable of the control and stability that are needed. The camera

must be extremely accurately positioned in respect to the kinescope and mounted so that there is no vibration which might blur the image. The motors driving the camera are driven from the same power source that produces the television synchronizing signal so that no phase shift occurs between the television system and the camera. Furthermore, the camera is driven by motors having d-c field magnets so that the correct phase is automatically assumed. In order to avoid any shadow bars from the shutter, the camera employs a fairly large aperture lens. It has been found that, if the lens opening is smaller than $f/5.6$, shadow bars may be visible. In practice the camera is equipped with an $f/1.6$ lens diaphragmed down to $f/2$. Great care must be taken to insure that there is no vibration.

In order to avoid vibration the shutter may be driven by a separate synchronous motor especially designed for the purpose, and a mechanical coupling may be provided between the main camera motor and the shutter to help bring the shutter up to speed in starting, but this coupling floats when the equipment is at running speed. A camera so designed by Eastman Kodak for television recording is shown in Fig. 21.14. This camera employs an eight-tooth sprocket pulldown actuated by an accelerated Geneva motion. The pulldown angle is approximately 57 degrees. In order to avoid the possibility of emulsion pile-up at the film gate, nylon pressure pads are utilized. Focusing and framing are done by means of a magnifying right-angle viewer. When in use, this unit is clamped on in place of the normal pressure pad. A great deal of work has been done in the development of this type of camera, which has been brought to a high state of perfection.

The kinescope used for recording purposes is essentially a projection kinescope with a blue zinc sulphide screen. It has been found that the P11 phosphor is very satisfactory for this purpose. In order to obtain sufficient brightness, the kinescope must be operated at 25 to 30 kilovolts. This need imposes no particular problems or restrictions since these voltages are commonly used in television projection. The 5WP11, a modified five-inch projection kinescope, has been developed specifically for kinescope recording.

The television picture reproduced on the recording kinescope may be either positive or negative. This gives the recorder the option of photographing a negative picture on the kinescope and obtaining a positive reproduction directly. Alternatively, he may photograph a positive on the kinescope and obtain a negative in his camera. Where the picture is made for record only, it most frequently will be recorded as a positive. If, however, a number of copies are to be made, it is obvi-

ously advantageous to make a negative recording from which at a later time positive prints can be made. It should be pointed out that when a negative is made from the kinescope, the direction of horizontal deflection should be reversed. This follows from the fact that it is advantageous, both in the camera and in the projector, to have the

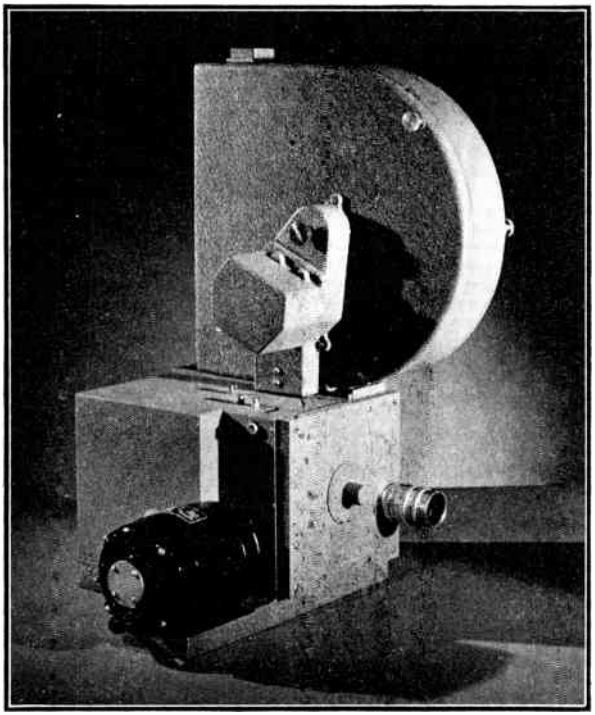


Fig. 21.14. Television Recording Camera for 16-Mm Film with Single Storage and Take-Up Magazine.

emulsion side of the film towards the lens. When a print is made from the negative film, there is a reversal from left to right. To compensate for this reversal, the deflection must be reversed.

The audio portion of the telecast is also recorded at the time the kinescope recording is made. Usually sound is recorded on a separate film by separate equipment, particularly when the recording is made as a negative. Separate recording permits the most favorable processing for each record. The sound track is then combined with the film when the positive print is made. In order exactly to synchronize the audio and video recording, a special 350-cycle synchronizing signal is re-

corded both on the kinescope picture and on the sound track. It is very easy to line up this synchronizing signal within a fraction of one cycle.

As has already been mentioned, the present recording techniques are good enough so that it is very difficult to distinguish a well-made kinescope recording from a live studio production. When, as unfortunately is frequently the case, the kinescope recording appears inferior to the original program material, this is not due to fundamental reasons but to some slight operational negligence. However, the improvement in picture quality of television has advanced and will continue to advance fairly rapidly. It is within the definitely foreseeable future that 16-millimeter film, unless its characteristics are correspondingly improved, will no longer be adequate for kinescope recordings. In spite of the severe economic burden which will be caused by this shift, it is almost certain that from that time on kinescope recordings will be made on 35-millimeter film.

When color television comes into general use, the recording of received programs in color will present a major problem. In principle, either a single-color film or three black-and-white color separation films may be prepared. There is some reason to believe that in spite of the equipment complexity and the fact that three times as much film is used, recording on three separate films will be found to be the more economic method. However, the experimental and developmental work required to give a definite answer to this problem has not been concluded so that the final answer remains for the future. Furthermore, the recording of television programs on tape offers a third alternative, which, in many respects, appears more attractive than either of the two other methods.

21.5 Video Tape Recording. The recording of video signals on magnetic tape, of the type familiar from sound recording, has a number of obvious advantages. Magnetic tape requires no processing whatsoever, permits immediate playback, and can be reused at will after demagnetization. There is little question that, with further development, tape recording of television programs will be economic and desired.

The tape itself is a plastic ribbon with a magnetically treated surface; at demonstrations given in December, 1953, the video tape for monochrome recording was one-fourth inch wide and the tape for color recording was one-half inch wide. For recording and playback the tape is moved past the extremely narrow gap of a horseshoe electromagnet. In order to prepare the tape for recording and to cancel any

previously recorded signals, a relatively large constant current may be passed through the magnet coil, causing the magnetic material of the tape to be saturated in the direction of motion. For recording, the coil current consists of a component proportional to the video signal superposed on a suitable biasing current. This causes a modulation of the magnetization of the tape determined by the magnitude of the video signal. During playback, finally, the variations in magnetic field at the surface of the tape record induce currents in the coil on the horseshoe core. They serve to generate the recovered video signal.

At the demonstrations of tape-recorded pictures at the RCA Laboratories in December, 1953, the speed of the tape motion was 30 feet per second. With this speed, a 17-inch reel was required to record a 4-minute television program. However, reductions in the required speed and improvements in tape characteristics are expected to result in considerable savings in material and storage space.

It will be noted, incidentally, that even at 30 feet per second a picture element corresponds to only a fraction of a ten-thousandth of an inch along the tape record. A whole picture line is recorded in a segment 0.023 inch in length. It is obvious, under the circumstances, that extreme precision is required in the tape transport to prevent distortions in the reproduced picture.

The quarter-inch tape employed for monochrome video programs carries two records, one for the video signal, including synchronization signals, and one for the accompanying sound. On the half-inch tape employed for color recording five separate tracks are recorded, one for each of the three primary color signals (red, green, and blue), one for the synchronizing signal, and one for the sound signal.

The excellent results obtained with tape recording in the aforementioned demonstrations make it appear certain that it will play an important role in the television industry in the future.

21.6 Staging of Studio Programs. It might be of interest to describe briefly the equipment and operation of one of the large studios to be found at a key broadcast point of a nationwide network. Studio 8-H in Radio City, part of the National Broadcasting Company system, is an example in point. This studio is one of the largest studios in the country, having an overall volume of some 300,000 cubic feet. It is designed to accommodate a great many sets at once. Figure 21.15 is a phantom view of the studio.

The control rooms for video, audio, and lights are along one end of the studio raised above the studio floor and equipped with windows through which the directors may observe the proceedings in the area

below. At the opposite end are dressing rooms and rooms for sound effects. Along one of the side walls is an observation gallery capable of accommodating a number of visitors. The gallery is not large enough, however, to contain a large studio audience. Most of the lights, both for base lighting and for special effects, are hung on pipes which are individually counterweighted so that they may be raised or

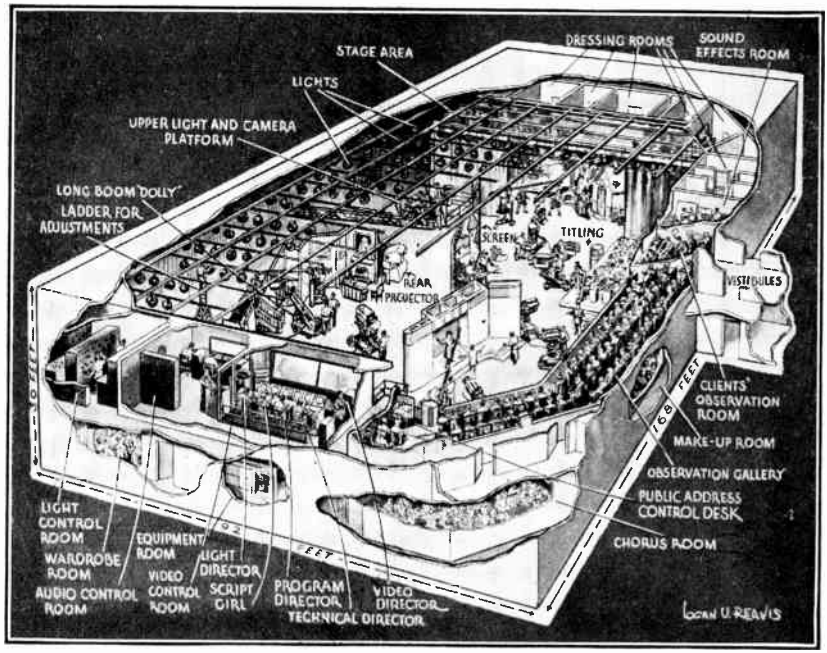


Fig. 21.15. Phantom View of Studio 8H. (National Broadcasting Co., New York.)

lowered conveniently. A total of some 1000 lights is used; each of them can be controlled from the lighting-control room. The engineering detail and equipment involved in the lighting alone are enormous. It might be pointed out here that there are in general two broad methods of studio lighting for television work. One is to put in essentially special lighting for each performance, this lighting being of a temporary nature so that it can be easily removed and installed for another production. The second is to provide enough lights so that the requirements of any production which may be put on in the future can be met. The former is economical from the standpoint of equipment but expensive from the standpoint of labor. The second duplicates equipment many times over in order to reduce labor costs. The

second procedure is found to be the most advantageous, in general, in large metropolitan centers.

The video control room is actually one story above the floor of the studio, the studio being three stories high (Fig. 21.16). The directing staff in the control room has nine viewing monitors to work from.

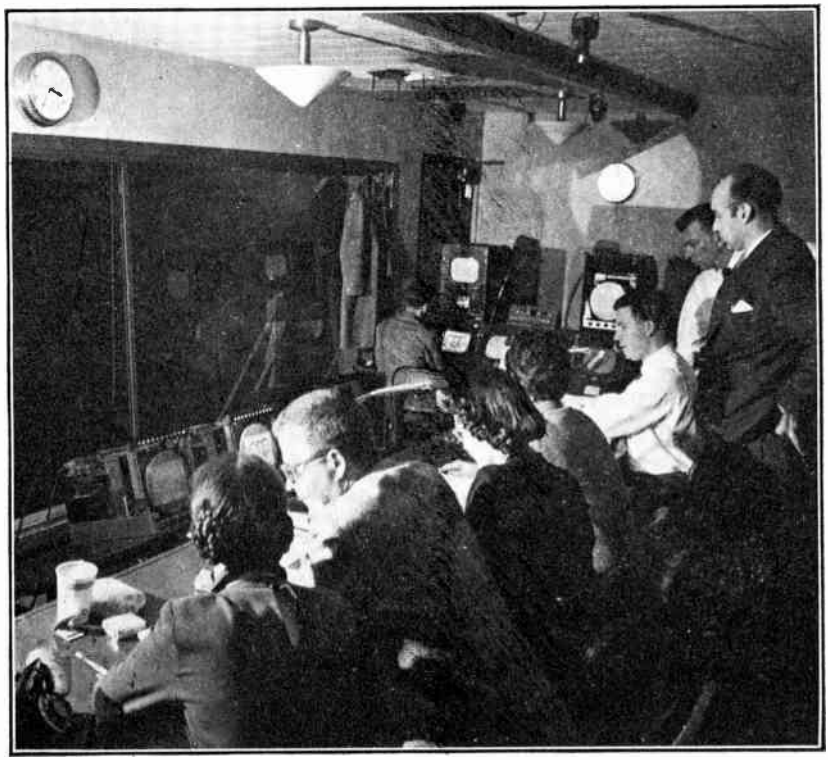


Fig. 21.16. A Video Control Room. (Studio 8H, WNBT, New York.)

They include one for each of six cameras, a preview monitor, and an on-the-air monitor. The ninth viewing tube shows possible portions of the program which may originate elsewhere, as from a film studio. The on-the-air monitor is larger than the others and mounted above them so that it is clearly visible to all.

An interesting innovation is made in this studio in that a platform is provided just below ceiling level on which a high camera can be mounted to give special effects (Fig. 21.17). Of course, a full complement of cameras on various mounts as discussed in section 21.3 is provided on the floor of the studio as are a variety of stand microphones,

boom microphones, etc. The studio is equipped to be sufficiently versatile so that either variety shows or straight dramatic presentations can be given. It is arranged so that a standard stage about 30 feet deep can be set up with a specially built proscenium. Since this stage is

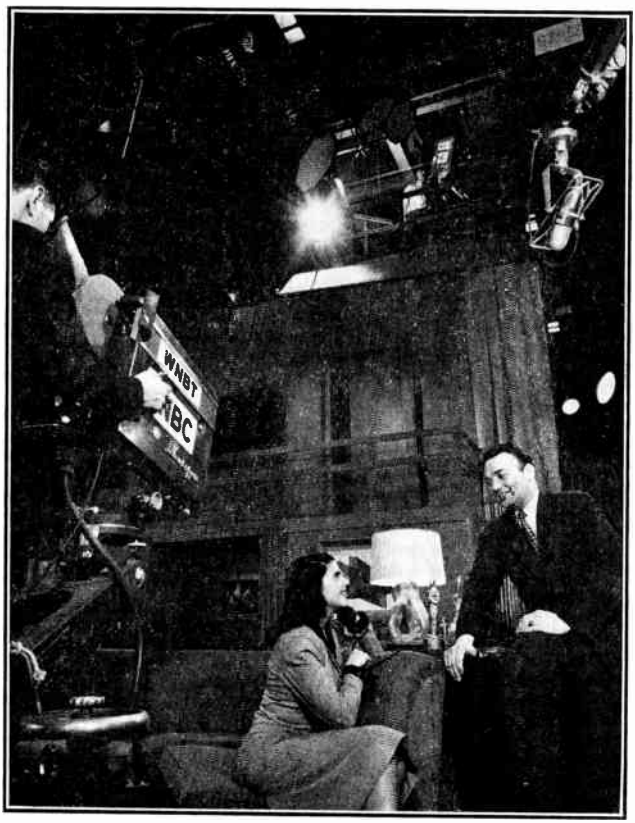


Fig. 21.17. A Studio Set in Studio 8H. (Note camera on balcony.)

not elevated, the cameras are not restricted to platforms as they frequently are in theatre-type studios. Various new devices are provided for special effects such as a motor-driven transitor, which is a device to provide moving patterns suitable, e.g., in dream sequences; a combination hand- and motor-driven "crawl" for titles; and a motor-driven revolving display table which is similar to a "lazy Susan" in its action. Furthermore, one of the largest rear-projection screens in the country, measuring 15 by 20 feet, is available. A smaller screen, of course, can also be used. When in full operation, this studio requires

a technical crew of some twelve to fifteen men in addition to a staging crew of two or more stagehands.

NBC's Center Theater in New York is another elaborate and very large point of origin of television programs. Here a theater-type arrangement rather than studio arrangement is employed, with a raised stage and an audience seating capacity of some 3000 viewers. The control room here is located at the back of the orchestra, the seats for

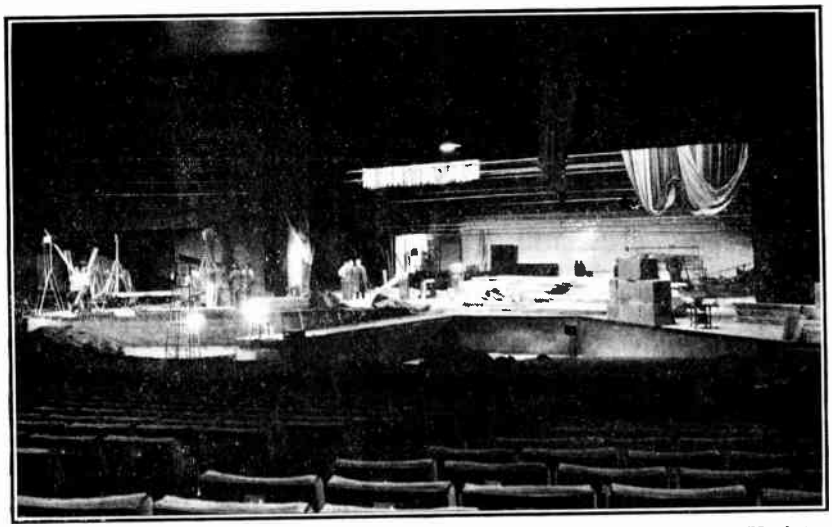


Fig. 21.18. Preparing the Stage at the Center Theater. (NBC, New York.)

the audience being in front and on both sides of it. In this location it has a clear view of the entire stage. In order to provide adequate camera travel, a ramp has been extended from the front of the stage (Fig. 21.18). Like the studio described above, it is provided with a full complement of cameras, lights, special-effect equipment, etc.

Studio installations of comparable complexity and capabilities may be found at the principal broadcast stations of any major television system.

When it is realized that a single system such as NBC may put on the air as many as 125 live programs each week plus a great many film programs as well and that each of these programs in itself is a production, the complexity of the total operation becomes quite evident. First of all, the programs themselves cannot be scheduled on a hit-or-miss basis without enormous economic waste and inefficiency of utilization of equipment. The programming of the productions must

be run on a schedule resembling a railroad timetable if complete chaos is to be avoided. Each program must be scheduled about two weeks before the first rehearsal, and plans for it must be integrated into the other productions which will be occurring at the time.

A week before the rehearsal a complete list of costumes and properties must have been formulated and the construction started for stage sets and special equipment. Because of the great variety of items required, the problem of maintaining a stock of properties is in itself formidable. Special shops are required to produce such unusual items as cannot be obtained through normal channels. They are constructed of papier-mâché, balsa wood, plaster, etc. They may include tree trunks, rocks, antique weapons, musical instruments, grave stones, enlarged packages or containers of commercial products as well as various hand props that require special treatment. Another line of special equipment is composed of the trick items, such as breakaway chairs that collapse easily and bottles, bowls, and vases that can be broken, if necessary, over performers' heads without injuring them.

When the production itself has been scheduled and the props and special equipment have been arranged, it is necessary to organize the individual program. The program is in the charge of the program director who is responsible for the performance. Assisting him he has, in major programs at least, an assistant director, script girls, and other aides. The technical director is responsible to the program director for all the technical aspects of the performance. The technical director has full control of the camera operators and the sound men. Using the script as a guide, the technical director will analyze the technical problems involved so that during the dry rehearsal he will be able to work effectively with the producer and director. The technical director works closely with the program director, on the one hand, and the cameramen and other technical television people on the other. He cues the latter group concerning the best lighting, lenses to be used, camera positions, following action, etc. He also performs the necessary switching function between cameras, superimposition, and dissolves. In other words, he has direct control of the cameras from the control room and indirect supervision through his direction of the technical personnel.

Assisting him is the video control engineer. He is responsible for the proper electronic adjustment of the camera. The picture level, contrast, shading, and special effects are under his direct control. Of course he must coordinate his actions with those of the lighting engineer in order to obtain an optimum picture.

The audio engineer also assists the technical director and has the responsibility for all the sound in connection with the picture. He controls the microphones, turntables, and such special sound effects as may be needed. He also controls the audio signal and cues the men on the microphone booms during operations.

The cameraman is responsible for the particular camera to which he has been assigned. During the production, he moves the camera into the correct position, maintains its optical focus, and follows the action under the instructions of the technical director. When the camera is mounted on a dolly, he directs the dollyman by hand signals on cue from the technical director. The dollyman, of course, must take care of the movement of the camera so that it will be smooth and uninterrupted. He must avoid cable and other obstructions which might lie in the path of the camera. It is sometimes necessary to employ cablemen to move cables that are in the camera's path.

The lighting director has charge of all the light sources of the set. He is directly responsible to the technical director, but, of course, the lighting effect must, in the final analysis, be analyzed and selected by the program director. It is, however, his knowledge, together with that of the technical director, which makes possible the achieving of the various lighting effects that the artistry of the production requires. The lighting director of most shows is located in the control room and gives cues for light changes on the communication system to the operators at the switch and dimmer board.

Rehearsals for a television program, therefore, are not only for the benefit of the artists and performers but also for the benefit of the technical people, all of whom must perform smoothly and accurately on cue. Because of the precision with which a television performance must be executed and because of the complexity of the operation, the ratio of rehearsal time to air time is very large. For the average live program, this ratio is probably of the order of eight. In some of the elaborate programs, such as the "Hit Parade," the ratio may run as high as twenty. The objective of the technical personnel is to put on a performance which, to the viewer, appears to have no technical problems at all. In fact, the viewer should not be aware of the technical aspects in any way. The infrequency of intrusion of this aspect of television upon the viewer through some technical error or failure is an outstanding tribute to the work done by the technical planning and operating groups.

21.7 The Television Station and Transmitter. The previous sections have dealt in some detail with the television studio, its equip-

ment, and its operation. In an expanding examination of the field, the next item to consider is the television station itself, including the transmitter. The planning and design of a television station is a very complex matter. In addition to the technological problem, which in itself is formidable, there are other aspects of equal or greater importance which must also be considered. They include the physical and political geography of the locale being considered, the economics of the construction and operation of the station, and many sociological problems which govern the type of material to be transmitted, the matter of personnel available to build and man the station, and the working environment of the transmitter. To indicate the magnitude of these problems, the nine basic steps in logically planning a station listed by J. Herold * are given below:

1. Determine area of market, radius of coverage required to serve the area, television channel availabilities, and estimated potential income.
2. Select site, determine antenna height, antenna gain, and transmitter power.
3. Determine source of program material, program policies, outline tentative program schedules, and plan extent of programming facilities.
4. Estimate total capital investment.
5. Estimate yearly operating expense.
6. File application for construction permit with Federal Communications Commission.
7. Project probable future expansion.
8. Determine personnel requirements and begin training.
9. Plan the building and the design, construction, and installation of technical equipment.

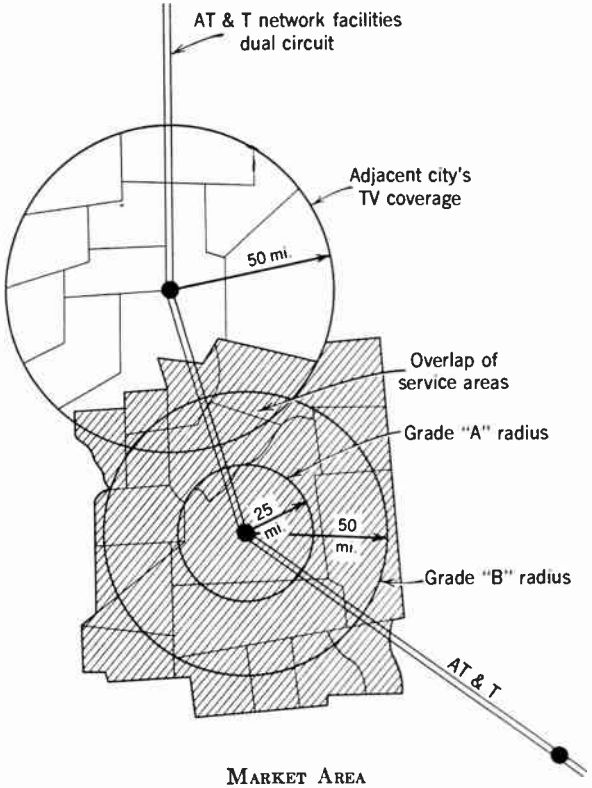
The economic analysis pertaining to a television station does not lie within the province of this book. Figure 21.19 indicates some of the relevant considerations. Once the analysis has been made, a decision can be reached regarding the size of the television station. For convenience, television stations are frequently placed in the following four categories, depending upon the size and power of the installation:

Group A. Installations of power outputs up to 2 kilowatts with equipment for film and slide transmission and network programs. A

* See reference 11.

group A′ designates a similar installation, but includes a small studio for handling simple live programs.

Group B. Includes transmitters with power output up to 10 kilowatts. These stations are equipped to handle, in addition to film, slides, and network programs, live programs from one studio.



MARKET AREA

Important Considerations: A. Potential TV households in coverage area. *B.* Existing TV receiver installations. *C.* Retail sales. *D.* Program sources. *E.* Duplication of coverage from existing and proposed TV stations. *F.* Rates for Class A time. *G.* Competition from other media.

Fig. 21.19. Considerations in Planning a Television Station. (Herold, reference 11.)

Group C. Designates stations with transmitter powers up to 25 kilowatts, with two or more live studios, remote pickup, and, of course, equipped to handle network and film programs.

Group D. Is constituted of television stations radiating more than 25 kilowatts. Installations belonging in this class would generally in-

clude two or more rather elaborate studios with their control rooms, a master control room, facilities for remote pickup, and, of course, network and film program equipment.

The general size and nature of the installation having been selected, the next important question is its location. The primary considera-

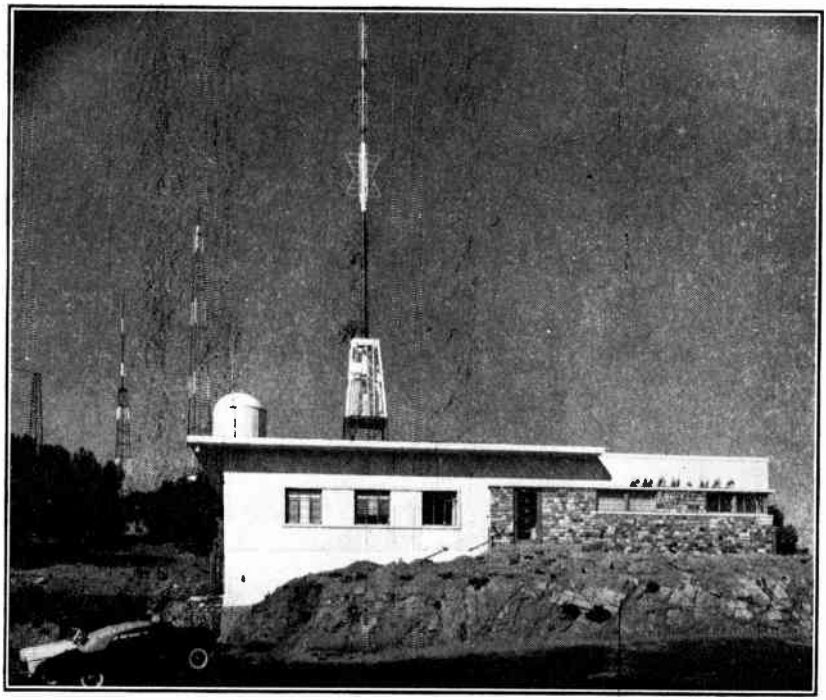


Fig. 21.20. NBC Hollywood Transmitter on Mt. Wilson.

tion in choosing the site of the television station is the location of the antenna itself. If possible, it should be near the center of the areas to be served. It must be as high as is physically possible. If it can be located on top of a natural feature of the terrain, such as a mountain (Fig. 21.20) or the top of a tall building, there is an obvious structural advantage. The antenna also should be well away from the sources of interference. The transmitter must be located close to the antenna to minimize power losses in the coaxial lines or wave guides feeding the antenna. The studio preferably should be in the same building with the transmitter. This arrangement makes for economy of operation and ease of coordinating transmitter activities with pro-

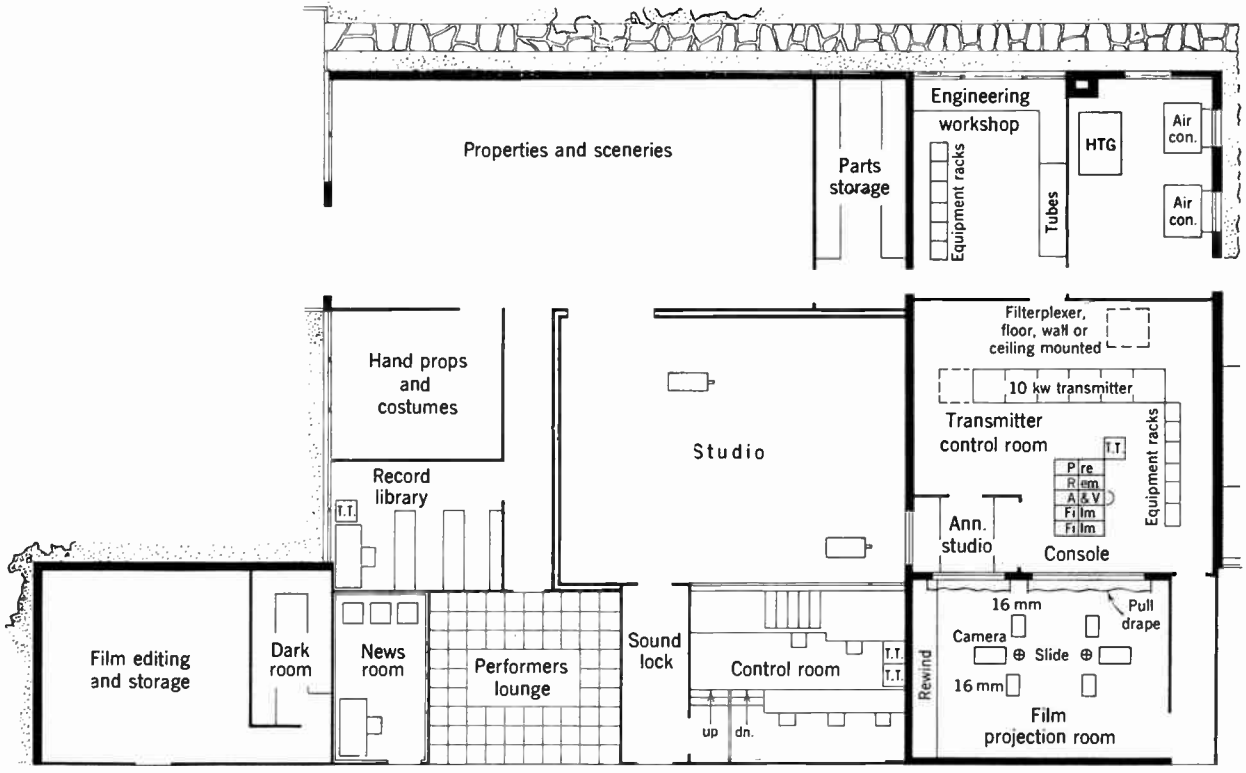


Fig. 21.21. Typical Floor Plan of Group B Television Station. (Anderson and Hadlock, reference 2.)

gram production. However, this often proves to be physically impossible, and the studio building has to be located at some distance from the transmitter. When this is necessary, the signals transmitted at the studio are sent to the transmitter by means of a small radio relay link or over a cable.

As was pointed out in Chapter 15, the antenna construction is very important in determining the effective radiated power of a transmitter. By properly shaping and arranging the radiating elements of an antenna, it can be made to radiate more power in a horizontal plane and less vertically, where it would be wasted. For example, a superturnstile antenna with three sections will deliver 3½ times more power in a horizontal plane than a dipole radiator fed with the same power.

As an illustration, let it be assumed that steps 1 and 2 in the planning of the station have been completed and that a Group B transmitter has been found to be best suited to the needs of the particular situation. A suitable site has been found which will permit the inclusion of the transmitter and studios in a single building. The building is then planned to permit efficient operation of the entire station, including studios and transmitter. In designing the building, allowance should be made for possible future expansion. Costly errors in this direction are not infrequent where a station will be so thoroughly designed to perform with its initial size that expansion is virtually impossible without completely tearing out the initial installation. A typical floor plan of a Group B television station which might be laid out for the problem at hand is shown in Fig. 21.21. The transmitter is at the center right of the building. Adjoining it on two sides are the live-talent studio and the film projection room. A maintenance and test shop and the air-conditioning equipment adjoin the third side of the transmitter room. The control room and a large properties room are on either side of the studio. Provision is also made for film editing and storage, darkroom work, news reporting, and storage of records for sound broadcasting and musical accompaniment. A block diagram of the video system for this installation is shown in Fig. 21.22. The transmitter for this station might be designed to deliver 3 kilowatts of power into the transmission line feeding the antenna. In this case, the video portion of the transmitter might be a crystal-controlled oscillator followed by several Class C narrow-band r-f amplifier stages supplying a final power stage using an RCA 8D21 water-cooled tetrode operated as a modulated amplifier. This final stage would be grid modulated by the amplified video signal. The audio system

might, for example, consist of an RCA frequency-modulated exciter, followed by several amplifier stages and a final power amplifier employing the type of tube used in the video section. For satisfactory operation very carefully regulated power supplies should be provided to insure stability. The equipment must be thoroughly interlocked to avoid possible danger to the operator, and starting and overload relays should be provided to protect the equipment from damage.

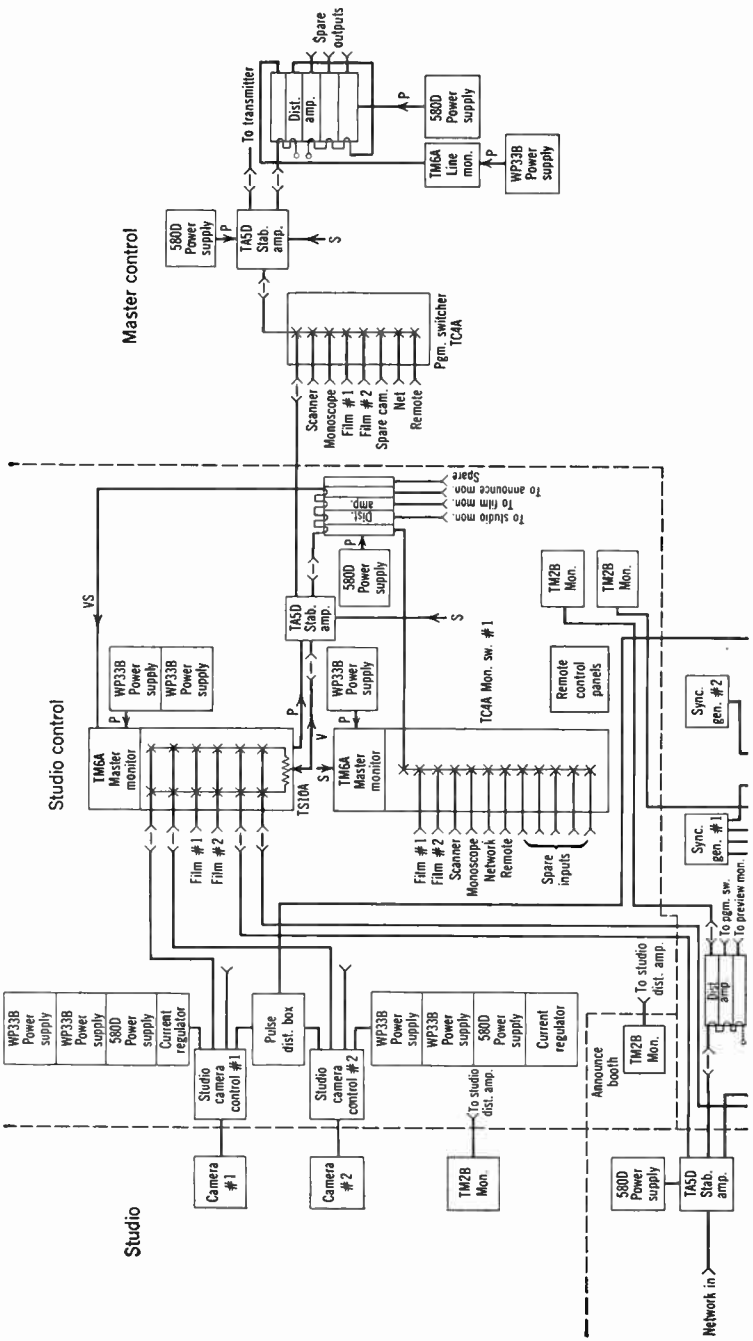
Before feeding the transmitter output to the antenna, it would go through a vestigial sideband filter so that the output would conform to FCC requirements. The antenna, which is mounted on a steel tower approximately 400 feet high, located close to the transmitter room, is a three-section super-turnstile. A photograph of such an antenna is shown in Fig. 15.63. In the channel allotted to the particular station in question the antenna has a power gain of approximately 4. The height of the antenna itself is approximately 40 feet, and its weight about 1½ tons.

A television station of the type described above would have a high-quality service radius of 10 to 15 miles and a useful service radius up to about 25 miles. Larger television stations require the solution of technical problems similar to those discussed above but very considerably increased in complexity. Apart from larger and more elaborate live studio and film projection rooms, a mobile unit here becomes essential equipment.

The mobile unit (Fig. 21.23) permits remote spot pickups which are relayed from the unit to the station by microwave relay or cable. Its interior is divided into a storage section and an operating section. The storage section houses cameras, tripods, dollies, cable reels, and the microwave transmitter in a manner which makes them readily accessible for use on location. The operating section is the analogue of the studio control room, with seats for the program director and the operating engineers. Three camera controls and a standard complement of monitors may be provided. The roof of the unit is reinforced for supporting a camera and cameraman as well as the microwave transmitter.

A more powerful transmitter generally parallels the more elaborate studio installation. A typical power output stage for a larger station is the 20- to 25-kilowatt 7-tube unit illustrated in Fig. 15.36.

A short section such as this can at best outline some of the steps involved in the complex problem of setting up a television transmitter and station. This phase of television is almost a field in itself with



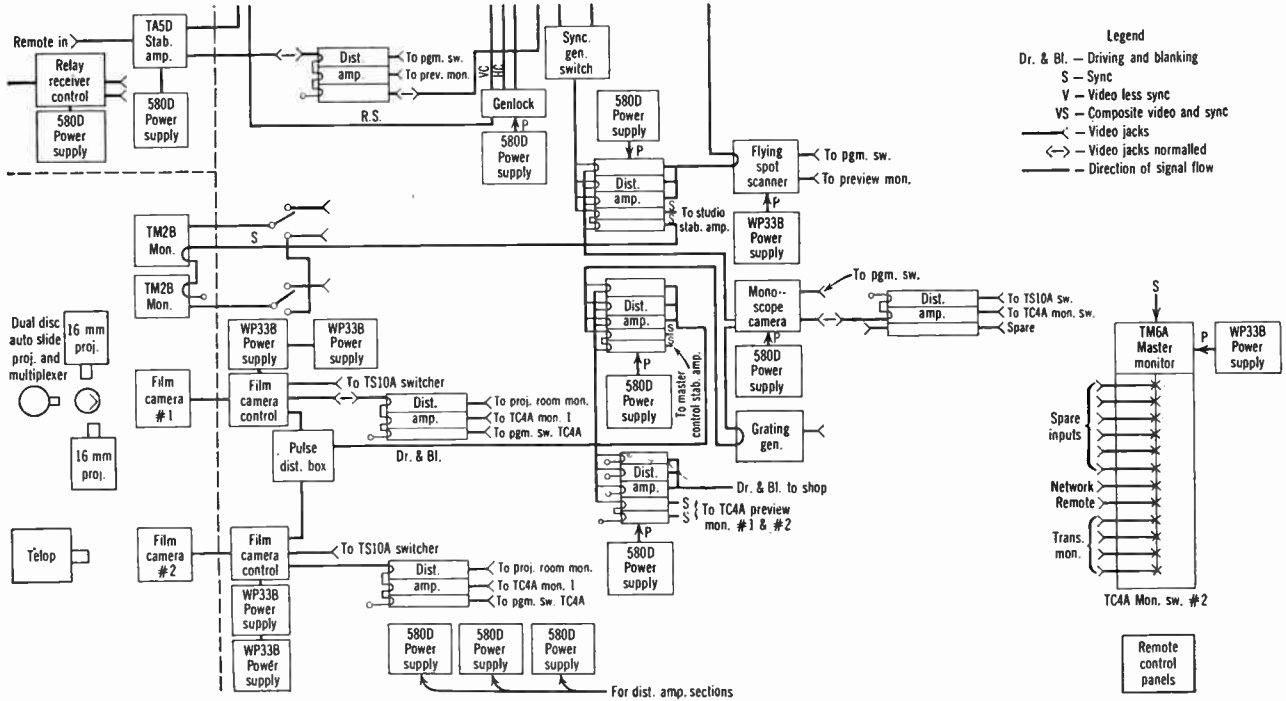


Fig. 21.22. Block Diagram of Video System for a Group B Television Station. (Anderson and Hadlock, reference 2.)

its own literature and specialists that must be called upon to solve specific problems.

Fig. 21.23. A Mobile Unit for Remote Pickup.

21.8 The Television Network. The limited service area of the individual television station as well as the high cost of television productions creates an obvious need for the interconnection of television stations. Networking, alone, makes possible the instantaneous witnessing of events of national interest throughout the country. Furthermore, networking gives access to high-grade entertainment originating in the studios of New York and Hollywood to cities and villages throughout the land. Economically, it extends financial support for television programs by increasing the audience of the advertiser. Finally, interconnection simplifies the problems of the station manager by greatly increasing the amount of high-quality program material available to him.

In view of the above facts it is not surprising that experiments with long-distance point-to-point transmission actually antecede the first regular television broadcasts. In the United States the burden of research in this field, as well as of the eventual establishment of network routes, has been carried by the Bell Telephone System. In England and other countries with publicly owned communication systems, the Post Office Department has been charged with this func-

tion. This close linkage between the telephone system and the television network is natural, since the means employed for transmitting television programs can, and generally do, transmit large numbers of long-distance telephone messages at the same time.

Two different types of transmitting systems are employed for interconnecting television stations: the coaxial cable and the microwave relay link. The two services have developed in parallel, and, at the close of 1953, the total mileage of coaxial cable and that of microwave relays in the United States were approximately equal.* The preference for one or the other type of transmission depends largely on the terrain. Thus, in mountainous regions and where large bodies of water must be crossed, the radio relay is likely to be advantageous. In relatively flat areas, on the other hand, the cost of the cable may easily be exceeded by that of the relatively complex radio relay repeater stations, particularly in view of their higher maintenance and power requirements. In city areas factors such as the ease of providing cable ducts and the availability of convenient line-of-sight routes may dictate the choice.

Coaxial Cable Networks. The advantage of the coaxial cable over other types of wire lines consists in its immunity to interference and its low attenuation over a wide frequency band. This immunity to interference applies, however, only to high-frequency signals, in the range above, e.g., 50 kilocycles, where the skin effect prevents penetration of the external sheath by interfering signals. For the direct transmission of the video signal, whose frequency spectrum extends down to about 60 cycles, "shielded pairs," such as are shown in Fig. 21.24, are preferred to coaxial cables. The shielded pair consists of two insulated wires, balanced to ground, enclosed in a conducting sheath. Low-frequency interference induces equal signals in the two wires, which cancel in the output.

Shielded pairs are widely employed for local networks, connecting studios and theaters with the master control room of the television station. For long-distance lines, on the other hand, the lower attenuation of the coaxial cable outbalances the extra cost of converters required to shift the video band to higher frequencies and to render, in this manner, the video signal immune to interference.

A typical long-distance cable, with eight coaxials and thirty-two additional paper-insulated wire lines, is shown in Fig. 21.25. The coaxial cable itself consists of an external copper sheath approximately

* See Loomis, reference 12.

0.012 inch thick, with an internal diameter of 0.375 inch and a central conductor held in place by polyethylene disks 0.085 inch thick and approximately 1 inch apart. The ratio of the inner diameter of the sheath to the outer diameter of the central conductor is made 3.6. As has been shown in section 15.8, this corresponds to a characteristic impedance of about 75 ohms and leads to minimum attenuation for a cable of given external diameter employing loss-free insulation with a dielectric constant 1.

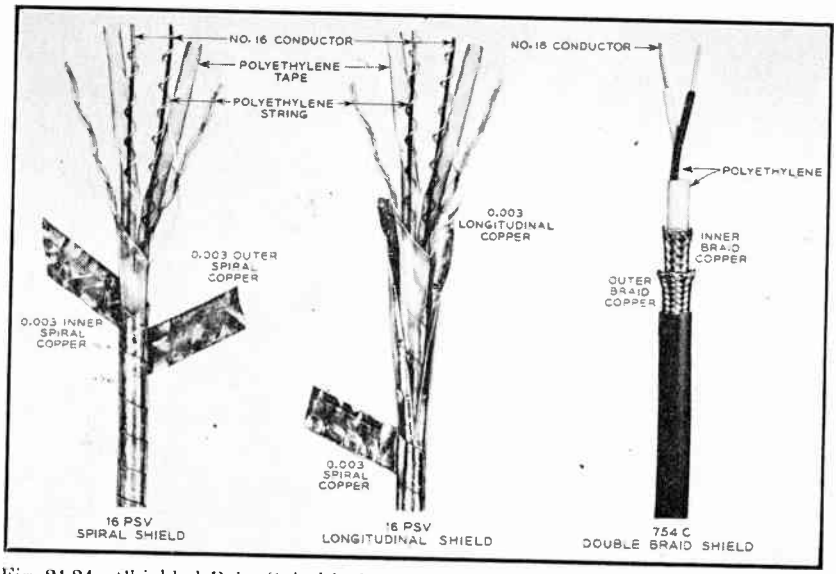
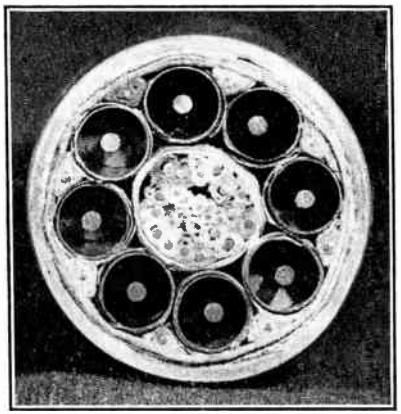


Fig. 21.24. Shielded Pairs Suitable for the Transmission of Video Signals. (Courtesy of American Telephone and Telegraph Co.)

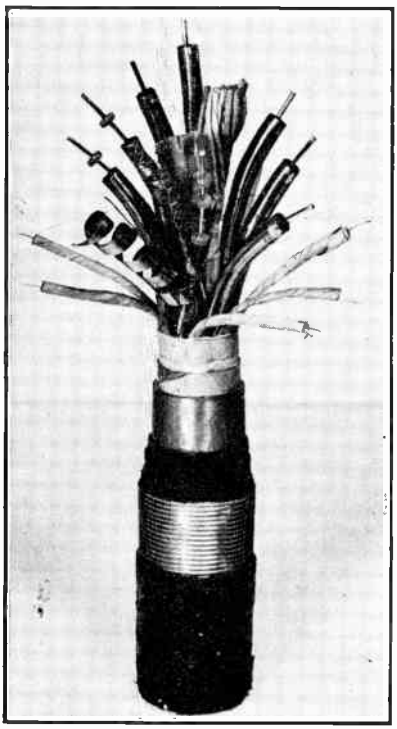
The external copper sheath is generally formed by copper tape bent into a tube with a longitudinal seam. It is enveloped by helically wound steel tape for mechanical protection. Identifying paper tape may be wound on the exterior of each coaxial pair, serving at the same time as insulation between the several units. Finally, the eight coaxial pairs are assembled together with the individual paper-insulated telephone wires in a lead envelope.

A single two-way circuit requires four coaxial pairs; two of them represent spare circuits, to which the signals are transferred in case of failures on the main circuit. The additional telephone wires carry warning signals in case of malfunctioning of any part of the system and are utilized in other ways for system maintenance. A pressure

of dry nitrogen or sulphur hexafluoride is maintained within the cable to prevent the entrance of moisture and to aid the detection of faults in the cable sheath. The complete cable is generally plowed into the ground by a cable-laying machine.



(*a*)



(*b*)

Fig. 21.25. Cable with Eight Coaxial Pairs. (*a*) Cross section. (*b*) Fanned-out end of cable. (Courtesy of American Telephone and Telegraph Co.)

The attenuation of the coaxial cable increases approximately as the square root of the frequency. For this reason, it is desirable to make the frequency shift to which the video signal is subjected relatively small. Even so, Fig. 21.26 shows that the attenuation of only 100 miles of cable is about 375 decibels at 1 megacycle and twice this value at 4 megacycles. A more detailed examination shows that there are deviations from proportionality with the square root of the frequency which, in the range from 70 kilocycles to 4 megacycles, have maximum values of the order of 0.2 decibel per mile.*

The high values of attenuation along the cable necessitate the installation of amplifiers, or repeaters, along the cable. The gain char-

* See Gould, reference 13.

acteristic of each line amplifier must be such as to compensate the attenuation of the succeeding length of cable, i.e., increase approximately with the square root of the frequency. The line amplifiers must be spaced sufficiently closely that the required gain can be attained without distortion with low-power components. Furthermore, the signal level at the input of each amplifier must be sufficiently high to over-ride the cumulative amplifier noise. A simple analysis shows

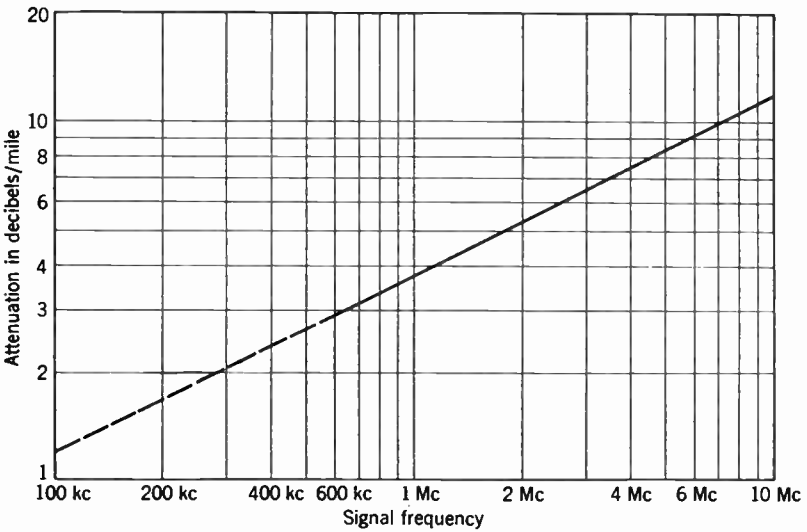


Fig. 21.26. Attenuation of a Typical ⅜-Inch Coaxial Pair as Function of Frequency.

that, with properly adjusted amplifier gain, the input signal power required for a given signal-to-noise ratio at the output is simply proportional to the length of the cable route and independent of the amplifier spacing.

The cable attenuation and the repeater gain also depend, in rather complex fashion, on the ambient temperature, so that it is desirable to control the amplifier gain automatically to compensate temperature and aging effects. Thus, to make certain that the signal level remains, in successive repeaters, sufficiently high to over-ride the noise and, at the same time, sufficiently low to remain within the range of linear amplification of the amplifiers, a number of "pilot tones" are transmitted along with the television signal. A pilot tone is a fixed frequency oscillation, which is filtered out at the repeater and is utilized to control the repeater gain. In practice, several pilot tones are em-

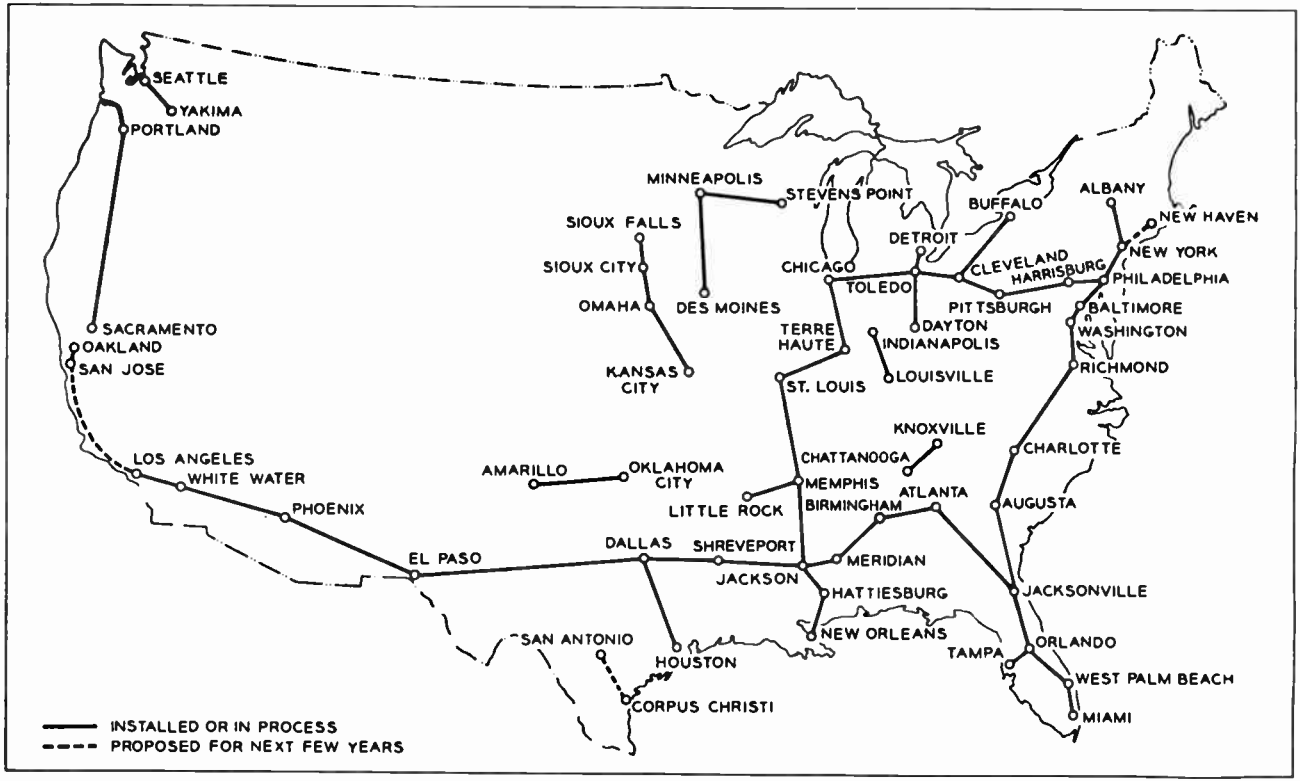


Fig. 21.27. Map of Coaxial Cable Routes for Television Relaying (1953). (Elmendorf, Ehbar, Klie, and Grossman, reference 19.) (Courtesy of *Bell System Technical Journal.*)

ployed, making possible a differential gain adjustment at different frequencies of the television signal.

Finally, the cutoff characteristics of the amplifiers introduce nonuniform signal delays which are cumulative. Hence, "delay equalizers" must be inserted for their compensation, along with "gain equalizers" which flatten the overall frequency characteristic by compensating the imperfect match between the cable attenuation and the line amplifier gain.

This brief account of the problems faced in the design of coaxial cable networks for television transmission may serve as an introduction to the extensive literature on this subject.* The following description of coaxial cable systems in the United States is, at best, very incomplete.

Whereas the first coaxial cables for television transmission, such as the New York–Philadelphia cable over which 240-line pictures were sent in 1937, employed ¼-inch coaxial pairs, the more recent cables utilize ⅜-inch coaxial pairs of the type described on page 997. The repeater system for the 1937 cable was designed for a video bandwidth of somewhat less than 1 megacycle.† With the arrival of commercial television coaxial cable routes, denoted as the "L1 System," ‡ were put into service which transmitted a video band of 2.8 megacycles. A map of the L1-System Routes in 1953 is shown in Fig. 21.27. A few of these—in particular, the stretch from Phoenix to Dallas—were, at the time, used only for telephone transmissions. Figure 21.28 shows a typical cable-laying operation in a desert area of the United States.

In 1953, the conversion of these routes to the L3 coaxial cable system, § providing for a video bandwidth of 4.2 megacycles with additional space for 600 message channels, had been begun. The L3 System differs from the L1 System principally in the utilization of more precise equalization techniques, improved components, and refined system analysis. It employs a repeater spacing of 4 miles as compared with 8 miles in the L1 System and has an overall bandwidth of 8 megacycles as compared with 3 megacycles. The cable itself and a considerable portion of the terminal equipment are the same in the two systems.

Figure 21.29 shows the frequency bands allocated for television transmission in the original New York–Philadelphia test of 1937 and in the L1 and the L3 Systems. In the first two instances, vestigial sideband transmission about a low-frequency carrier, at 144 and

* See references 13 to 19.

† See Strieby, reference 14.

‡ See references 15 to 18.

§ See reference 19.

311.27 kilocycles, respectively, is employed. The lower, vestigial, sideband extends only to 120 and 200 kilocycles, respectively, so that the bandwidth of the displaced signal is scarcely any greater than that of

Fig. 21.28. A Cable-Laying Crew in the Desert. (Courtesy of American Telephone and Telegraph Co.)

the original signal. The desired signal is obtained by successive modulation with the video signal of two high-frequency carriers (7944.72 and 8256 kilocycles in the L1 System) whose difference is equal to the low-frequency carrier of the transmitted signal. In the L1 System pilot frequencies are located at 64, 556, 2064, and 3096 kilocycles and the sound is transmitted on an 80-kilocycle carrier. The position of the three upper pilot frequencies is chosen so as to minimize the visi-

bility of interference of the pilot frequencies with the television carrier.

In the L3 System only a single modulation, with a carrier at 4139 kilocycles, is required. The upper sideband extends to 8500 kilocycles, and the vestigial lower sideband to 3639 kilocycles. The space between 564 and 3084 kilocycles is allocated to 600 telephone message channels. Pilot frequencies are placed at 308, 556, 2064, 3096, 7266,

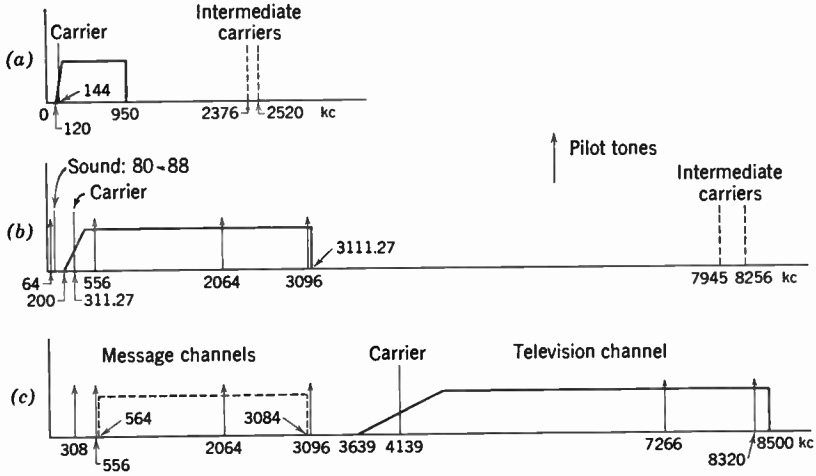


Fig. 21.29. Frequency Allocations in Coaxial Cable Systems. (*a*) New York-Philadelphia Cable (1937). (*b*) L1 System. (*c*) L3 System.

and 8320 kilocycles. The modulator is double-balanced so as to eliminate both the video signals and the carrier from the output. The amplitude range of the transmitted signal is reduced to half the value for 100 percent modulation by subtraction of the carrier in such amplitude as to make the signal amplitude equal for the synchronizing signal (blacker-than-black) and for white portions of the picture. The carrier phase is, of course, opposite in these two portions of the signal. "Homodyne" detection at the receiver, employing a demodulator with a carrier which agrees with the carrier component of the signal wave both in phase and frequency, permits the recovery of the undistorted video signal. At the receiver both the carrier and the pilot tones are separated out from the signal by very sharp crystal filters.

The general layout of the L3 repeaters is shown in Fig. 21.30. Except for the spacing of the repeaters, the layout of the L1 System is

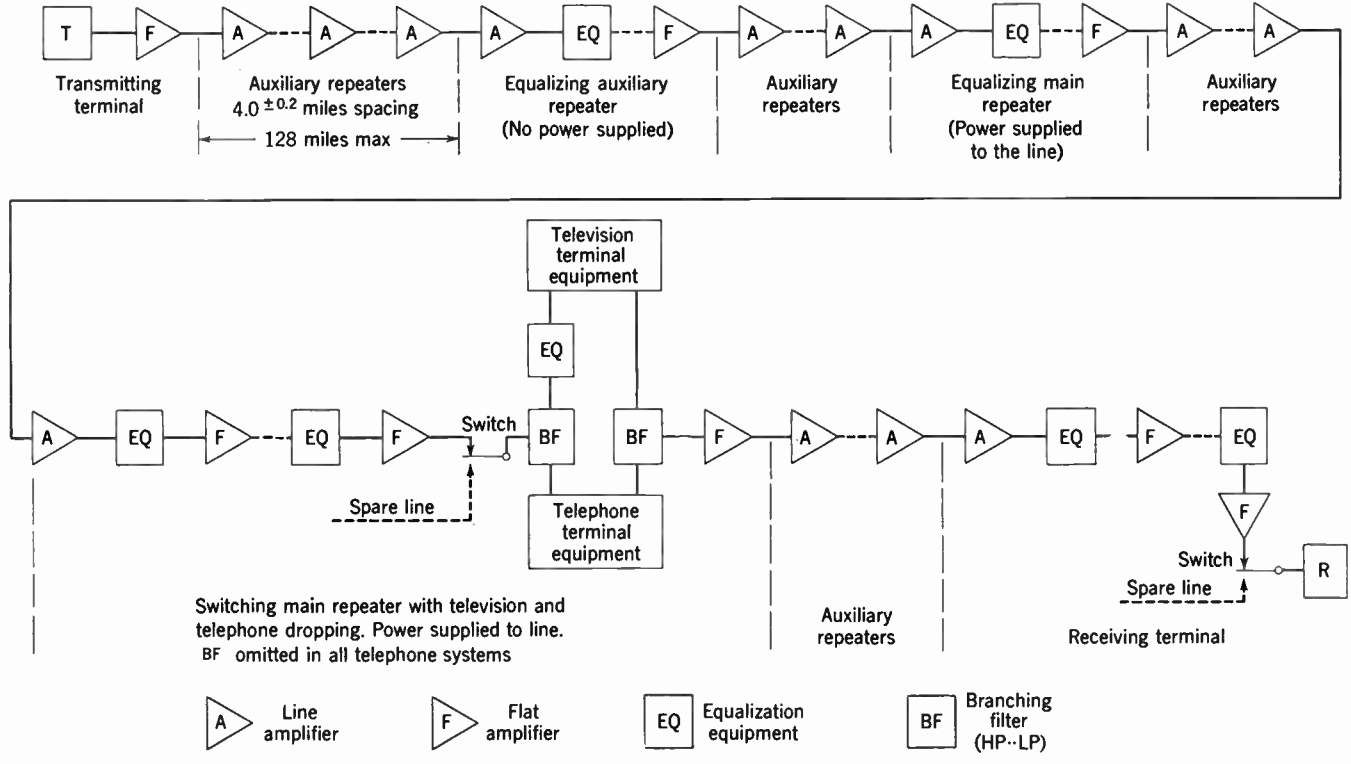


Fig. 21.30. General Layout of L3 Repeaters. (Elmendorf, Ehbar, Klie, and Grossman, reference 19.) (Courtesy of *Bell System Technical Journal*.)

basically similar. The 60-cycle power is carried by the central conductors of two coaxial pairs and is supplied at main repeaters spaced approximately 100 miles apart. At the auxiliary repeaters, it is separated out by power separation filters and is applied to regulated power supplies for the amplifier. Line loading coils at each repeater point keep the 60-cycle current equal throughout the system. Standby equipment in the form of batteries and engine-driven generators takes the place of the external power supply in case of power failure at one of the main repeaters.

The auxiliary repeaters are generally housed in small unattended concrete huts. They contain an amplifier compensating the line loss. In the L1 system this is regulated by a thermistor controlled by the 2064-kilocycle pilot tone; in the L3 system the regulating thermistor responds either to the 7266-kilocycle pilot tone or simply to the cable temperature, the variation of cable loss with temperature being known accurately. The main repeaters are housed in regular telephone buildings and contain, in addition to line amplifiers, equalizers controlled by several pilot tones, manual equalizers, and flat amplifiers compensating the equalizer loss, servicing equipment for testing and checking components. Cable damage and excessive deviation from normal of the amplitude of a pilot tone at an auxiliary repeater or other evidence of malfunctioning along the line are signaled automatically to the nearest main repeater. The same applies to calls for assistance from line crews making routine checks.

Switching main repeaters are usually also branching or terminal points. Here, the transmission may be switched to a spare line if necessary repairs make this desirable, and the television signal may be supplied to or from a local station. In the L3 System the television terminals, with demodulation and modulation of the signal, are placed in tandem. Between terminals 4000 route miles apart, this process may be repeated as many as twenty times. Satisfactory picture reception is made possible by extremely careful elimination of distortion terms in the modulation and demodulation processes.

Microwave Networks. Whereas the attenuation characteristics of cables dictate the use of a low-frequency carrier in coaxial networks, point-to-point transmission of television signals by radio is advantageously carried out on the highest-frequency carriers which can be generated with reasonable efficiency and are not strongly absorbed by the atmosphere. Elementary diffraction theory shows that if a small signal source is placed at the focus of a lens or reflector of area A, the

major part of the radiated energy is distributed, at a distance d from the transmitter, over an area approximately equal to $3.7\lambda^2 d^2/A$, where λ is the wavelength of the radiation. If the receiver detector is placed at the focus of a second lens or reflector of area A, the signal strength transmitted is seen to be proportional to $A^2/(d\lambda)^2$. Thus, for equal radio link attenuation, the required antenna area is proportional to the wavelength or inversely proportional to the frequency, and, for equal antenna area, the repeater spacing is proportional to the frequency (provided that no obstacles interfere with the transmission).

Thus, to economize in antenna and relay station installations, as well as to utilize the large frequency range available in the microwave region of the radio spectrum, radio relay systems for long-distance television transmission quite generally employ microwaves in the neighborhood of 4000, 6000, and 7000 megacycles. In this range (λ = 7.5 to 4.3 centimeters) the condition that ground configuration has little influence on the propagation, namely, that the "Fresnel region" including paths which deviate from the direct path between transmitter and receiver by a half wavelength be free from obstacles, can be fulfilled with moderate elevations of the two antennas. Ordinary rough terrain does not give rise to regular reflections, though water surfaces and the desert will do so. Atmospheric refraction, on the average, causes a bending of the microwave beam with a radius of curvature four times as great as the radius of the earth. In addition, atmospheric disturbances may cause rapid fading effects which may vary the received signal within a range of -20 to $+10$ decibels, measured relative to the average value.* This fading must be compensated at the repeater stations by automatic volume control of the transmitted signal.

A map of the microwave relay system employed, in 1953, for television relaying in the United States is shown in Fig. 21.31. The repeater stations on the Bell System routes are placed approximately 25 or 30 miles apart along each route. Most of these stations are unattended auxiliary stations. Main repeater stations with facilities for switching circuits so as to permit repairs on inoperative circuits and with limited test equipment are generally several hundred miles apart. The same applies to alarm and maintenance centers placed at accessible points to serve an "alarm section" of some twelve repeaters.

* See Friis, reference 20.

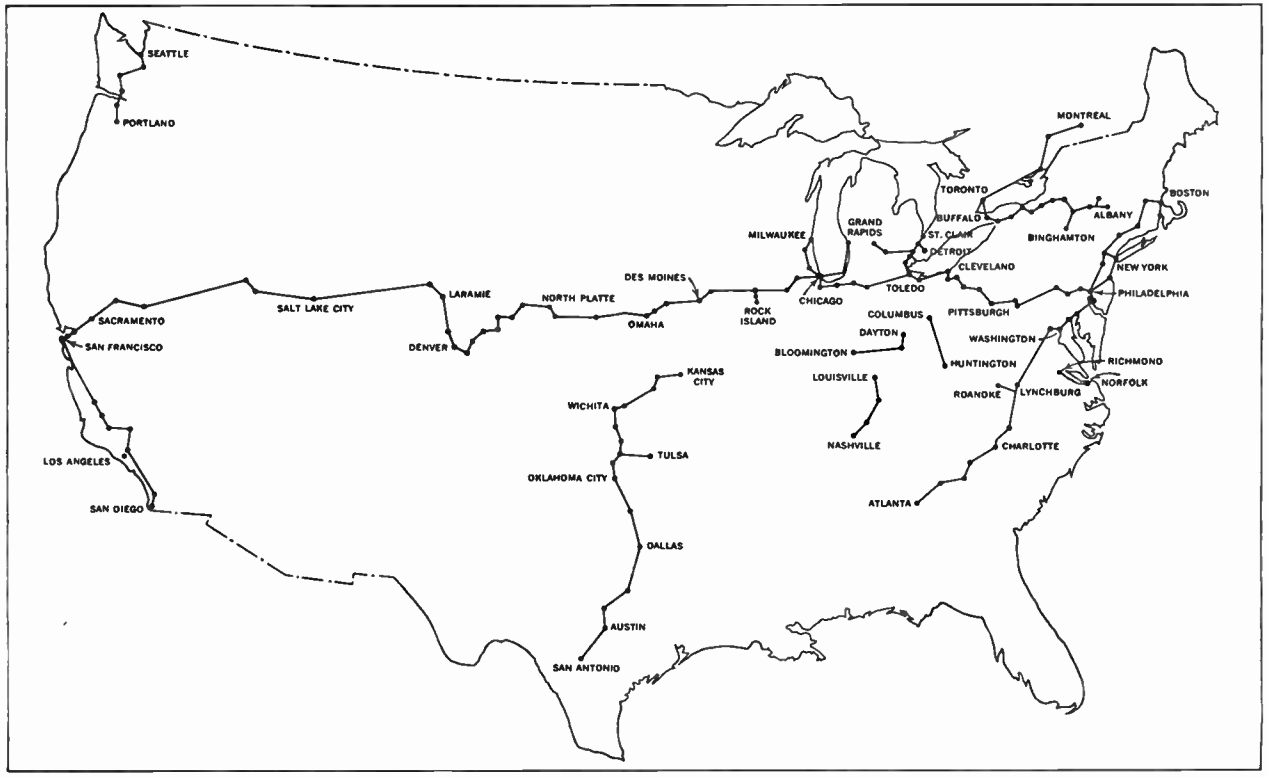


Fig. 21.31. Map of Microwave Routes for Television Relaying (1953). (After *Tele-Tech*, Vol. 12, February, 1953, by permission.)

The individual repeater units consist of concrete (Fig. 21.32) or steel towers (Fig. 21.33). In the second instance the housing for the electrical equipment is generally at the base. Four directional antennas, in the form of waveguide horns placed at the focus of microwave lenses or paraboloidal reflectors, are mounted on the top of the tower. The antennas constitute the receiving and transmitting antennas for signals traveling in the two opposite directions along the route. The microwave lenses are commonly made up out of slabs of polystyrene foam in which thin metal strips have been inserted in such fashion as to create the analogue of a refractive medium for light waves.* Such lenses are superior to paraboloids in ease of weather protection, reduced reflection of radiation into the horn, and lessened cross-talk with the antenna placed back-to-back with respect to it.

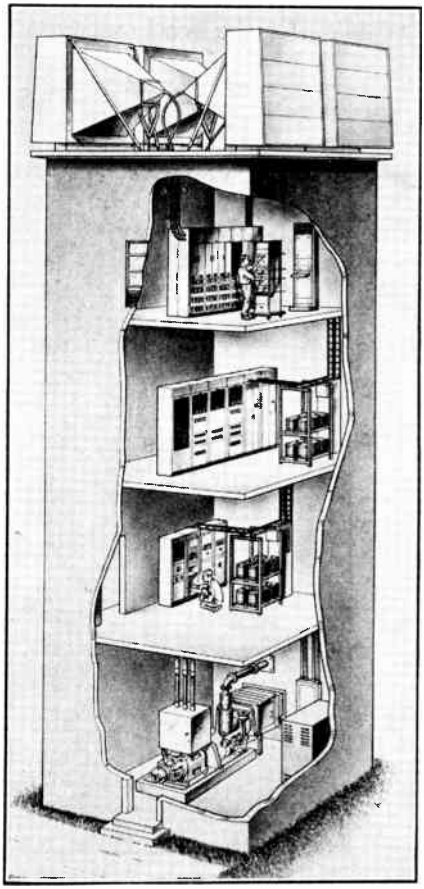


Fig. 21.32. Concrete Microwave Repeater Tower, Indicating Internal Layout. (Courtesy of American Telephone and Telegraph Co.)

Since the passband of both free space and the antennas employed in the radio relay system is practically unlimited, the entire 3700 to 4200 megacycle common carrier band can be handled as a unit. This band is broken up into twelve broadband channels, six in each direction, and spaced 40 megacycles apart, at the repeater † (Fig. 21.34). Furthermore, the frequency of each channel is shifted by 40 megacycles at the repeater, so as to eliminate feedback from the transmitter to the receiver for the same channel.

* See Kock, reference 21.

† See Roetken, Smith, and Friis, reference 22.

Figure 21.35 is a schematic diagram of the entire transmission system. At the transmitter terminal the video signal, at a signal level of 1 volt, frequency modulates a 70-megacycle carrier, the frequency excursion for the synchronizing signal and white portions of the pic-

Fig. 21.33. Steel Repeater Tower with Equipment Housed at Base. (Courtesy of American Telephone and Telegraph Co.)

ture being held to ±4 megacycles, respectively. In order to obtain a highly linear response, the frequency modulation is carried out originally at 4280 megacycles, by modulating the repeller voltage of a reflex klystron, and the resulting signal is converted to 66 to 74 megacycles with a 4210-megacycle oscillation. The 70-megacycle i-f signal modulates a selected microwave carrier, which is generated by a quartz crystal tuned to about 18 megacycles followed by six frequency multipliers. Modulation occurs in a 416-A microwave triode similar to the BTL 1553 triode described in section 15.7. The modu-

lator is succeeded by three amplifier stages employing the same tube type. The coupling between the input and output cavities of the successive stages is such as to provide a passband of 20 megacycles between points 0.1 decibel down. The signals from several transmission channels are passed through channel separation filters and combined in the wave guide feeding the transmitter antenna.

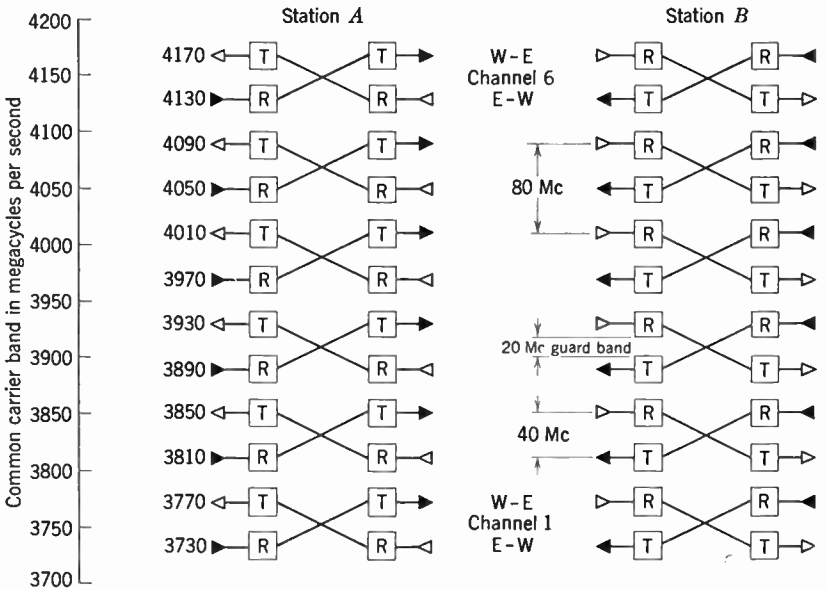


Fig. 21.34. Transmission Channels in the TD-2 Microwave System. (Roetken, Smith, and Friis, reference 22.) (Courtesy of *Bell System Technical Journal*.)

At the receiver the signal from the antenna passes once more through a system of channel separation filters. At their output the i-f signal is recovered by conversion with a microwave oscillation in a crystal converter stage. The i-f signal is amplified by 12 decibels in a preamplifier and an additional 65 decibels in a main amplifier, which is provided with automatic volume control to keep the amplitude of the synchronization signal at the output within narrow limits (1 decibel for 30-decibel variation in input power). Provision is also made for delay equalization. The amplified i-f signal modulates a microwave oscillation differing from that of the receiver microwave generator by 40 megacycles. The resulting microwave signal is then again amplified and combined with the microwave signals from the remaining channels as at the transmitter terminal.

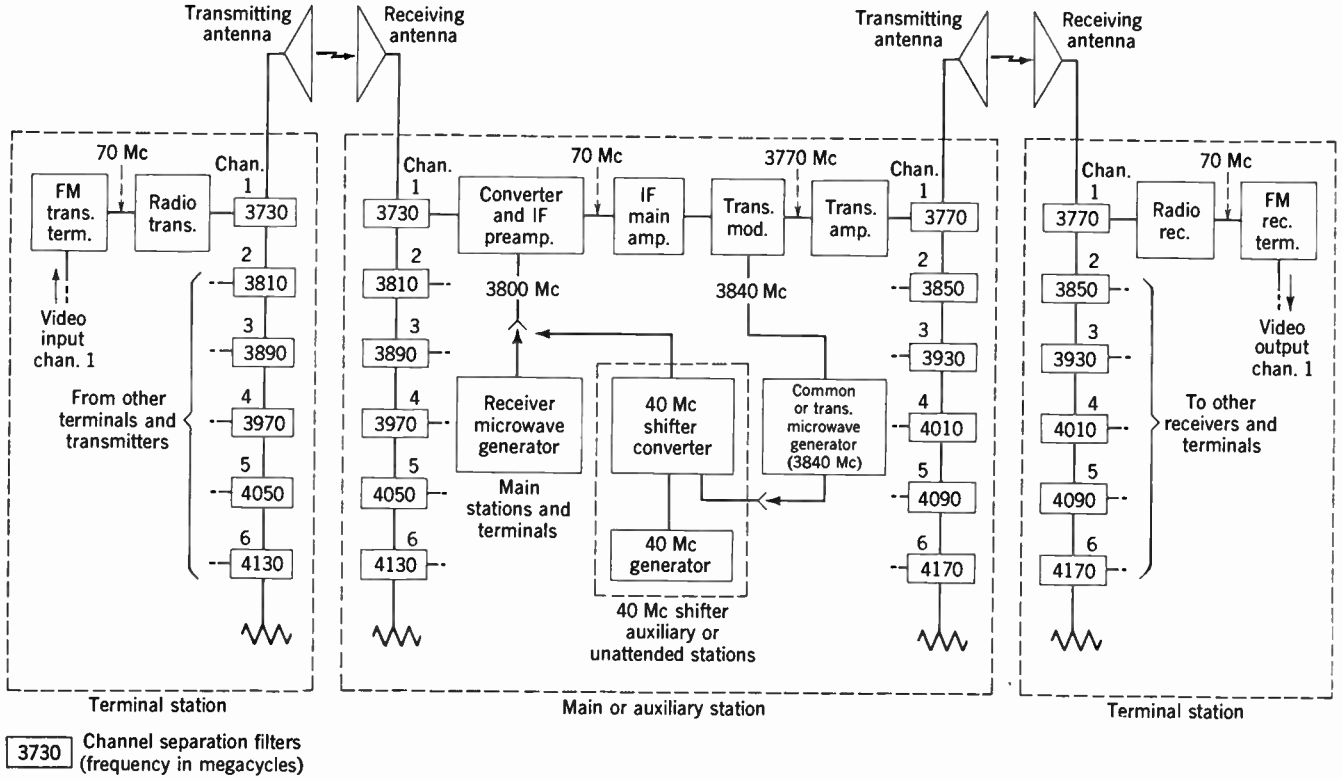


Fig. 21.35. Block Diagram of TD-2 System. (Roetken, Smith, and Friis, reference 22.) (Courtesy of *Bell System Technical Journal*.)

In the auxiliary repeaters the receiver microwave oscillation is obtained simply by mixing the transmitter microwave oscillation with a 40-megacycle signal. Thus, the frequency shift of the output depends only on the stability of a relatively low-frequency generator. In main repeaters the two microwave generators must be highly stable and independent of each other, so as to permit switching of signals from one circuit to another.

The intermediate frequency is chosen with relation to the channel spacings so that the image frequencies generated in the converter fall

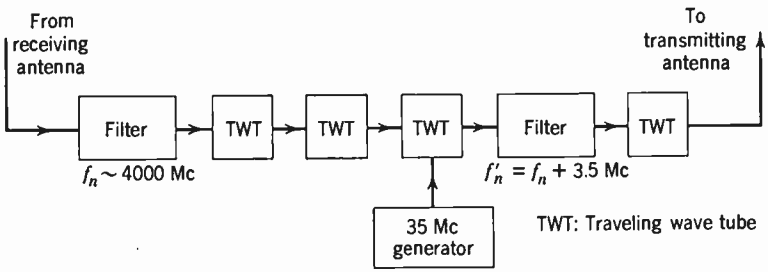


Fig. 21.36. Proposed Traveling-Wave Tube Repeater System without Intermediate Frequency Stages. (Bray, reference 24.)

midway between the channel centers. The reason for carrying out most of the amplification at the i-f level is the unavailability of sufficiently noise-free microwave amplifiers.

At the receiver terminal the i-f signal is applied to a typical frequency-modulated receiver incorporating a limiter and discriminator.

Looking toward the future, traveling-wave tubes appear particularly promising in microwave relay applications, since they will readily handle a passband of, e.g., 500 megacycles at 4000 megacycles and higher frequencies. This would permit microwave amplification without signal separation. Traveling-wave tubes are now employed as transmitter amplifiers in the Manchester–Kirk o'Shotts radio relay link in Great Britain,* together with conventional i-f amplifier circuits. Repeater construction can be enormously simplified when the noise properties of traveling-wave tubes have been improved to the point that the stepdown to the i-f level becomes unnecessary. A block diagram of a proposed repeater of this type is shown in Fig. 21.36.† It is seen that, apart from the 35-megacycle shift frequency generator,

* See reference 23.

† See Bray, reference 24.

the total tube complement of the repeater consists of four traveling-wave tubes.

Numerous shorter microwave television relay routes are in operation which differ from those described above in various details. Most of them employ, like the earlier Bell System routes, klystrons as transmitter amplifiers.* Paraboloidal antennas, eventually mounted near ground level and combined with tower-mounted reflectors, are widely employed. Together with the coaxial routes, microwave relay links may soon be expected to provide network connections for practically all the television stations in the United States. Similar developments are taking place in most of the countries of the world with several television stations.

21.9 Television Abroad. The preceding discussion has been confined, in the main, to the status of television in the United States. Table 21.1 gives a bird's-eye view of parallel television developments in the rest of the world. It is seen that, judged by the number of sets in use, Great Britain far outstrips any country apart from the United States. This is not surprising in view of the fact that the first regular public television service was instituted in England. The next point of note is the extensive television development in the Americas outside the United States, which accounts for two-thirds of the television sets outside of the United States and Great Britain.

By comparison, the distribution of receivers in continental Europe is very small. To a considerable extent, this may be attributed to the physical destruction and the economic setback effected by the Second World War. The relatively large number of stations in operation, under construction, and on the planning boards makes a large expansion in the immediate future likely.

An examination of the standards indicates that Great Britain alone employs 405 lines and 25 frames and that France and its dependencies (as well as French-speaking Belgium and Vatican City) utilize 819 lines and 25 frames. The rest of the world is split up between the 525 line–30 frame standard dominant in the Western hemisphere and the 625 line–25 frame standard adopted throughout the Eastern hemisphere. Since the line frequencies in these two systems are practically the same, relatively slight modifications permit the use of receivers manufactured for one system in the other system. This is particularly true since the two standards match each other in other respects, such as the use of negative modulation and frequency-modulated sound as well. Great Britain, France, and its dependencies, Belgium

*See Forster, reference 25, as well as reference 26.

TABLE 21.1. DISTRIBUTION OF TELEVISION STATIONS AND RECEIVING SETS THROUGHOUT THE WORLD IN MAY, 1953

(Quoted from *Tele-Tech*, Vol. 12, p. 146, June, 1953, by permission)

Country	Standards	Channel Width, Mc	Stations under				Estimated Number of Receivers
			Oper.	Constr.	Planned	Exp.	
Africa							
Morocco	819/25	14		1	2		
America							
Argentina	625/25	6	1	1	4		15,000
Brazil	525/30	6	4	3	9		60,000
Canada	525/30	6	2	1	4		250,000
Chile	525/30	6			2		
Colombia	525/30	6			1		
Cuba	525/30	6	8	20			120,000
Dominican Republic	525/30	6	1				500
Guatemala	525/30	6			3		
Haiti	525/30	6			1		
Mexico	525/30	6	6	3	9		50,000
Uruguay					3		
United States	525/30	6	167	349	627	23	23,738,000
Venezuela	626/25	6	1	2	2		
Asia							
Japan	525/30	6			9	2	4,500
Thailand						1	
Turkey	625/25	7			4	1	
Australia							
Australia	625/25	7.5			2		
Europe							
Belgium	819/25	14	1		4		4,500
	625/25	7	1		4		
Czechoslovakia	625/25	8				1	
Denmark	625/25	7			1	1	1,000
France	819/25	14	2	4	30		70,000
	441/25	9	1				
Germany (East)	625/25		1				
Germany (West)	625/25	7	7	5	10		5,000
Great Britain	405/25	5	5	2	3		2,300,000
Italy	625/25	7	2	2	5		5,500
Monaco	819/25	14			2		
Netherlands	625/25	7	1	2	2	1	8,500
Saar	625/25	7			1	1	
Spain	625/25	7		2		2	300
Sweden	625/25	7			4	2	200
Switzerland	625/25	7		1		2	
USSR	625/25	8	4				160,000
Vatican City	819/25	14		1			
Yugoslavia	625/25	7	1		2		

and Turkey, employ amplitude-modulated sound. As has already been mentioned, the standard for the field frequency is quite generally equated to the frequency of the local power system, though the possibility of obtaining satisfactory reception with a field frequency which deviates from the power frequency has been amply demonstrated.

Coaxial and microwave links are in operation in several of the countries which have more than one operating television station, such as Great Britain, West Germany, and France.

21.10 Prospects of Television. Since, in 1954, over half the homes of the United States were equipped with television receivers, it requires scarcely any extrapolation to state that television will shortly assume, or already has assumed, the role of the major medium of entertainment in this country. The introduction of color can only accelerate this trend. Furthermore, it can confidently be expected that the remaining countries of the world will follow suit, with a speed dictated largely by local economic conditions.

It is equally certain that entertainment will be only one of the functions of the enormous investment in television sets and broadcasting systems. Even at the time of writing (1954), television has played a major role in familiarizing the public with men and issues in political campaigns to an extent which has been impossible heretofore; it has served the cause of adult education through countless forums, science programs, and cultural presentations; and it has been employed locally to reduce the impact of interruptions in regular school programs. Such purely educational functions are bound to increase greatly as the television set becomes a practically universal part of home equipment. Considerable progress has already been made in the use of television as a teaching aid at universities, an entire campus being wired for television in at least one instance.

At the same time, the establishment of television networks in most of the countries of the world will invite international program exchanges. Such exchanges have already been effected successfully between France and England.* In order to obtain a (British) 405-line, 25-frame television signal from a (French) 819-line, 25-frame television transmission, the 819-line picture formed on a kinescope screen was projected on the target of a storage pickup tube with a 405-line scanning pattern, which supplied the desired converted signal.

This method is not applicable, however, for conversion between television systems with greatly differing frame rates; exchanges be-

* See Lord, reference 27.

tween a 30 frame-per-second area, such as the United States, and a 25 frame-per-second area, such as Europe, would result in large-area low-frequency flicker which would detract seriously from the entertainment value of the programs.* The remedy appears to lie in the development of improved types of storage tubes. In particular, if the writing portion of a signal converter storage tube realizes the principle of the storage kinescope (section 8.6) the effect of differences in frame rate on stationary pictures is completely eliminated. It is probable that the storage-tube problem will have been solved by the time that a system of television relays joins the European and American continents.

Several proposals have been made for the realization of the last objective. Both floating relay transmitters and planes circling at high altitudes have been considered. One of the most attractive suggestions † foresees a chain of land-based microwave relay stations girdling the globe. Such a plan is rendered feasible by the fact that the largest separation between land masses which must be spanned—the distance between Iceland and the Faroe Islands—is only 290 miles.

The practical realization of the storage kinescope could have consequences which would go far beyond signal conversion for program exchange. This was fully realized by Schröter. ‡ First of all, the resulting elimination of flicker would permit a reduction of the field frequency to a quarter of its present value, which suffices to create an impression of continuity of motion. The necessary video bandwidth would be reduced by the same factor. As a second step, methods might be developed for transmitting picture signals corresponding to the difference of successive frames. Since, normally, these differences encompass only a small fraction of the entire picture field, the number of picture elements which must be transmitted to achieve a prescribed picture resolution is greatly decreased in this manner. Accordingly, picture difference transmission could be utilized for an additional great reduction in the video bandwidth required. The possibility of further savings in bandwidth is suggested by a statistical analysis of picture content. §

The reduction of the necessary video bandwidth to a small fraction of its present value would make entirely new services feasible, as, for

* See Zworykin and Ramberg, reference 28.

† See Forman, reference 29.

‡ See articles by F. Schröter in Leighäuser and Winckel, reference 30.

§ See Kretzmer, reference 31, and Harrison, reference 32.

instance, a video telephone system in which the home receiver would function as one terminal.

The development of television systems of very high resolution is another avenue along which progress may be anticipated. It is well known that terminal equipment can be built with much higher resolution capabilities than can be transmitted over broadcast channels by present techniques. When such equipment is placed into service, and the necessary transmission channels are made available, it may be anticipated that television spot pickup will replace news photography not only for television reception in the home, but also for the newspaper. The picture will be derived directly from the television signals at the newspaper press.

Various uses of the home receiver for which the engineering basis has been created even now have already been detailed in Chapter 20. An example was the possible role of the receiver as an aid in home management, through the employment of strategically placed camera attachments. The same cameras, in conjunction with a video tape recorder, may take the place of the home movie camera. The complete absence of processing makes the magnetic tape technique particularly attractive to the amateur. Furthermore, playback on the home television receiver is free from the inconveniences of setting up the screen and projector and darkening the room, which characterize conventional home movie presentations.

Finally, suppose that the major obstacles to interplanetary travel have been solved—that, at last, a space ship has been designed which can reach the surface of the moon intact. It is quite certain that such a ship, on its first successful flight, will be launched without human observers aboard. It is equally certain, however, that, among the many instruments assembled in it to bring us news of the voyage, the television camera will play a prominent role. Our first close view of the moon and the planets will, undoubtedly, be through the eyes of television, opening up vistas in new worlds as it has transformed the old.

REFERENCES

1. "1953–54 Statistics of the TV-Radio-Electronic Industries," *Tele-Tech*, Vol. 13, pp. 78–79, January, 1954.
2. L. E. Anderson and W. O. Hadlock, "Four Versatile TV Station Equipment Plans for VHF and UHF," *Broadcast News*, Vol. 73, pp. 31–71, March-April, 1953.
3. A. Reisz, "A New Television Camera for Studio and Field Use," *Broadcast News*, Vol. 68, pp. 30–41, March-April, 1952.

4. L. E. Flory, W. S. Pike, J. E. Dilley, and J. M. Morgan, "A Developmental Portable Television Pickup Station," *RCA Rev.*, Vol. 13, pp. 58–70, 1952.
5. A. E. Ohler, "The 'Walkie Lookie,' a Miniaturized TV Camera Custom Built for Use by NBC at National Political Conventions," *Broadcast News*, Vol. 71, pp. 8–15, September-October, 1952.
6. P. J. Herbst, "Televised Films," *Broadcast News*, Vol. 69, pp. 56–63, May-June, 1952.
7. R. V. Little, Jr., "Film Projectors for Television," *J. Soc. Motion Picture Engrs.*, Vol. 48, pp. 93–110, 1947.
8. R. V. Little, Jr., "Television Film Recording," *Broadcast News*, Vol. 54, pp. 32–36, April, 1949.
9. J. L. Boon, W. Feldman, and J. Stoiber, "Television Recording Camera," *J. Soc. Motion Picture Engrs.*, Vol. 51, pp. 117–126, 1948.
10. J. Herold, "TV Station Operating Costs," *Broadcast News*, Vol. 68, pp. 50–54, March-April, 1952.
11. J. Herold, "Considerations in the Early Planning of TV Stations," *Broadcast News*, Vol. 69, pp. 24–37, May-June, 1952.
12. F. Loomis, "Bell System Plans for Broadband Network Facilities," *Tele-Tech*, Vol. 12, pp. 78–80, 184–185, April, 1953.
13. K. E. Gould, "Equalization of Coaxial Lines," *Trans. Am. Inst. Elec. Engrs.*, Vol. 68, pp. 1187–1199, 1949.
14. M. E. Strieby, "Coaxial Cable System for Television Transmission," *Bell System Tech. J.*, Vol. 17, pp. 438–457, 1938.
15. L. G. Abraham, "Progress in Coaxial Telephone and Television Systems," *Trans. Am. Inst. Elec. Engrs.*, Vol. 67, pp. 1520–1527, 1948.
16. W. R. Lundry, "Attenuation and Delay Equalizers for Coaxial Lines," *Trans. Am. Inst. Elec. Engrs.*, Vol. 68, pp. 1174–1179, 1949.
17. J. P. Kinzer, "Stability of Tandem Regulators in the L1 Carrier System," *Trans. Am. Inst. Elec. Engrs.*, Vol. 68, pp. 1179–1186, 1949.
18. L. W. Morrison, Jr., "Television Terminals for Coaxial Systems," *Trans. Am. Inst. Elec. Engrs.*, Vol. 68, pp. 1193–1199, 1949.
19. "The L3 Coaxial System," *Bell System Tech. J.*, Vol. 32, pp. 779–1005, July, 1953. C. H. Elmendorf, R. D. Ehbar, R. H. Klie, and A. J. Grossman, "System Design," pp. 781–832. R. W. Ketchledge and T. R. Finch, "Equalization and Regulation," pp. 833–878. L. H. Morris, G. H. Lovell, and F. R. Dickinson, "Amplifiers," pp. 879–914. J. W. Ricke and R. S. Graham, "Television Terminals," pp. 915–942.
20. H. T. Friis, "Microwave Repeater Research," *Bell System Tech. J.*, Vol. 27, pp. 183–246, 1948.
21. W. E. Kock, "Metallic Delay Lenses," *Bell System Tech. J.*, Vol. 27, pp. 58–82, 1948.
22. A. A. Roetken, K. D. Smith, and R. W. Friis, "The TD-2 Microwave Relay System," *Bell System Tech. J.*, Vol. 30, pp. 1041–1077, 1951.
23. "British Television Relay Network," *Electrical Communication*, Vol. 29, pp. 171–178, 1952.
24. W. J. Bray, "The Travelling Wave Valve as a Microwave Phase Modulator and Frequency Shifter," *Proc. Inst. Elec. Engrs.*, Vol. 99, part 3, pp. 15–20, 1952.

25. W. H. Forster, "6000-Mc Television Relay System," *Electronics*, Vol. 22, pp. 80–85, January, 1949.
26. "Microwave Relay Link for Television," *Electrical Communication*, Vol. 30, pp. 3–8, 1953.
27. A. V. Lord, "Standards Converter for International Television," *Electronics*, Vol. 26, pp. 144–147, August, 1953.
28. V. K. Zworykin and E. G. Ramberg, "Standards Conversion of Television Signals," *Electronics*, Vol. 25, pp. 86–91, January, 1952.
29. A. J. Forman, "Global Microwave System for TV and Communications," *Tele-Tech*, Vol. 11, pp. 40–41, 118, 122, 1952.
30. G. Leithäuser and F. Winckel, *Fernsehen*, Springer Verlag, Berlin, 1953.
31. E. R. Kretzmer, "Statistics of Television Signals," *Bell System Tech. J.*, Vol. 31, pp. 751–763, 1952.
32. C. W. Harrison, "Experiments with Linear Prediction in Television," *Bell System Tech. J.*, Vol. 31, pp. 764–783, 1952.

AUTHOR INDEX

SUBJECT INDEX

TELEVISION

V. K. ZWORYKIN
G. A. MORTON

TELEVISION

V. K. ZWORYKIN
G. A. MORTON