RCA REVIEW

A QUARTERLY JOURNAL OF RADIO PROGRESS

VOLUME IV April 1940 NUMBER 4

IN THIS ISSUE

RCA INSTITUTES TECHNICAL PRESS
A Department of RCA Institutes, Inc.
75 Varick Street New York, N. Y.

RCA REVIEW

A Quarterly Journal of Radio Progress

Published in July, October, January and April of Each Year by

RCA INSTITUTES TECHNICAL PRESS

A Department of RCA Institutes, Inc.

75 Varick Street, New York, N. Y.

VOLUME IV — April, 1940 — NUMBER 4

CONTENTS

SUBSCRIPTION:

United States, Canada and Postal Union: One Year $1.50, Two Years $2.50, Three Years $3.50
Other Foreign Countries: One Year $1.85, Two Years $3.20, Three Years $4.55
Single Copies: 50¢ each

Entered as second-class matter July 17, 1936, at the Post Office at New York, New York, under the Act of March 3, 1879.

Printed in U.S.A.

MOBILE FIELD STRENGTH RECORDINGS OF 49.5, 83.5, AND 142 Mc FROM EMPIRE STATE BUILDING, NEW YORK — HORIZONTAL AND VERTICAL POLARIZATION

By

G. S. Wickizer

Engineering Department, R.C.A. Communications, Inc., Riverhead (L. I.), N. Y.

Summary—*The propagation characteristics of horizontally and vertically polarized waves have been studied on 49.5, 83.5, and 142 Mc, in several directions from the Empire State Building in New York City. Continuous mobile recordings of field strength were made from the transmitter location out to the limit of the receiver sensitivity.*

Comparison of average field strength on horizontal and vertical polarization, over the same route, revealed that the horizontal polarization produced a stronger average field than the vertical polarization. Variation in field strength over the same route was found to be greater for horizontal polarization beyond the metropolitan area, and greater for vertical polarization in the city. Average field strength curves for three frequencies, when corrected to the same power, showed the lowest frequency (49.5 Mc) was strongest in the city, and the highest frequency (142 Mc) was strongest in the country, on horizontal polarization. On vertical polarization, the highest frequency produced the highest average field strength in both urban and suburban areas, when the curves were corrected to the same power.

Introduction

THE rapid extension of the useful limits of the ultra-high-frequency spectrum brings with it problems in applying these higher frequencies efficiently. Several factors to be considered are the choice of frequency and polarization to produce the strongest signal over a given area. In general, the most desirable frequency will be determined somewhat by the topography of the area, and the better polarization will be dictated by the electrical characteristics of the earth's surface in the area.

Although these choices might possibly be made from theoretical considerations, the problem is better adapted to an empirical solution, due to the countless irregularities existing in the transmission paths. In making a field strength survey, the system of continuous mobile recording has the advantage of producing a complete record of field strength for the entire route covered, and for this reason is much more thorough than a point-by-point method. Accordingly, mobile recordings of field strength on horizontal and vertical polarization were made on 49.5, 83.5, and 142 Mc, with the transmitters located at the Empire State Building, New York. The same path was traversed in each case, thus providing a direct comparison of field strength on the two polarizations, and also supplying additional information on propagation at these frequencies.

Receiving Equipment

The receiving equipment, with power supplies and antenna, was installed in a passenger two-door sedan. Suitable precautions were taken to prevent mechanical vibration of the equipment when the car was in motion.

The receiving antenna was a short doublet made of two pieces of 5/8-inch diameter, duralumin tubing, supported at a height of ten feet above the ground. The tubing was clamped in a bakelite head, which was attached to a wooden shaft about four feet long. This shaft was mounted on the roof of the car by a mechanical assembly which ex-

Fig. 1—Mobile field strength survey car.

tended through the roof, behind the rear seat. The mechanical fitting on the car was constructed to permit rotation about a vertical axis, and also to allow the antenna to be folded down against the roof when not in use. A small steering wheel and indicator were provided inside the car to assist in setting the bearing of the antenna when receiving horizontal polarization. A picture of the survey car is shown in Figure 1.

The receiver was a triple-detection superheterodyne, equipped with automatic gain control to compress the wide range of field strength to be measured. The direct-current output of the receiver was amplified and applied to a Bristol recording milliammeter. This type of recorder produces a continuous ink record on a paper chart which is drawn under the pen at a known rate.

To associate the record with geographical location, the chart was driven from the car drive shaft, through a suitable reduction gear mechanism. With this arrangement, the recorder chart speed was either five inches per mile or twenty inches per mile, and the charts were numbered consecutively every inch.

Power for the receiver was obtained from six-volt storage batteries, which drive two 250-volt dynamotors. The batteries were connected to the car generator to reduce the battery current drain from fifteen amperes to about seven amperes.

The routes and frequencies covered in the field work are tabulated below.

Direction from New York	Frequencies
Northeast	83.5; 142 Mc.
North	83.5 Mc.
Southwest	49.5; 83.5; 142 Mc.

Fig. 2—Routes covered by survey car.

Additional information on the routes may be obtained from the map shown in Figure 2.

TRANSMITTING EQUIPMENT

A different transmitter was used for each of the three frequencies measured during the field work. On 49.5 Mc, the video channel of the NBC television transmitter was measured. The antenna on this transmitter was a half-wave doublet located on the south side of the building at the 85th floor. The height of this antenna was 1000 feet above the ground, and the radiated power was estimated to be 5 kw. The two higher-frequency transmitters were installed in the tower, and radiated from half-wave doublets mounted outside the tower at the 103rd floor level, a height of 1200 feet above the ground. These doublets were changed from the north to the south side of the building, depending on the direction of the survey trip. The radiated power at 83.5 Mc was 750 watts, and at 142 Mc, the radiated power was 68 watts.

Field Work

On each measuring trip, the observer kept a log correlating the numbers on the recorder chart with important locations along the route. Notations from the log were written on the charts before the analysis was begun. A chart speed of 20 inches per mile was used when re-

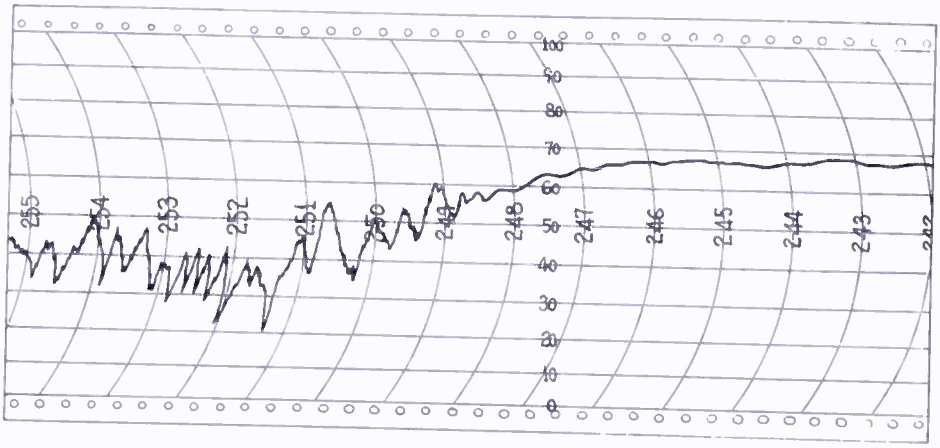


Fig. 3—Sample of mobile recording.

cording within 3 or 4 miles of the transmitter, to show greater detail on the record. At all other times the chart speed was 5 inches per mile.

On horizontal polarization, the receiving doublet was kept broadside to the transmitter, since the directivity of the doublet was appreciable. This steering operation was accomplished in several ways. At short distances, the Empire State Building was usually visible, and the local geography was well-known. Beyond the metropolitan area, the bearing was obtained from road maps, and from occasional checks of the antenna setting, by noting the position at which the signal was maximum.

Results and Discussion

The recorded charts were analyzed in small sections which could be readily identified on a map. This was necessary to provide a measurement of the airline distance from the transmitter to the middle of each section. The length of these sections was chosen in proportion to the distance, and varied from about half a mile near the transmitter, to 3 or 4 miles at the the far end of the trips.

The sample chart shown in Figure 3 illustrates the two extreme types of recording obtained during the field work. The smooth portion of the trace indicates that very few indirect paths were present; a condition to be expected in clear open country, and on certain wide streets in the city where there are no intervening buildings or overhead wires. The wide irregular trace is caused by indirect rays combining with the direct ray in random phase relation. In some cases, the direct ray may be very weak, with strong indirect rays present; a condition which produces wide local variations or "standing waves".

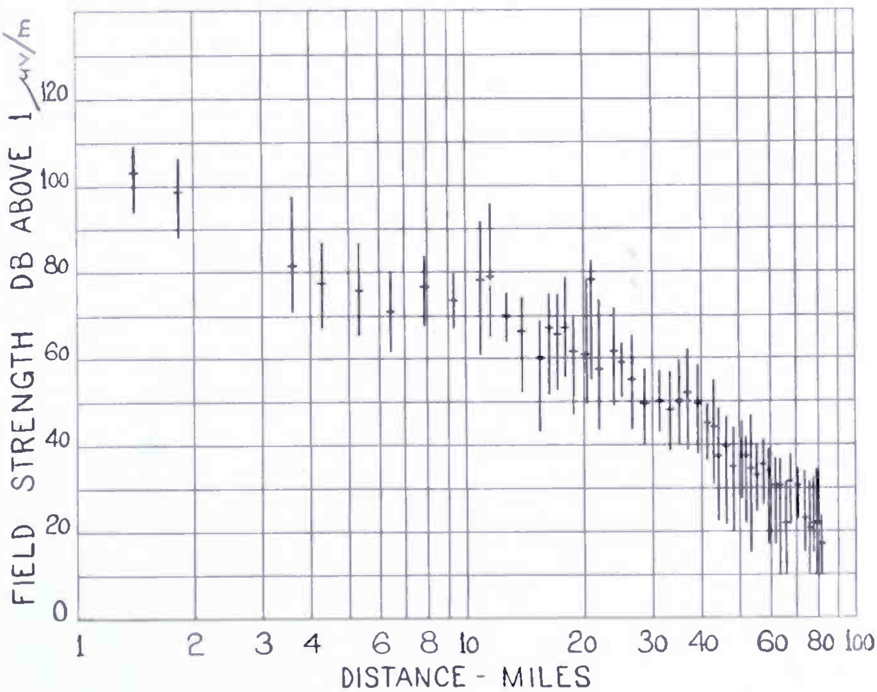


Fig. 4—Summary of field strength record taken between New York City and Camden, N. J. 49.5 Mc, horizontal polarization.

Such fields were often noticed in the city where the transmitter was shielded by high buildings, and the receiver was located near the street level. Thus both the amplitude and the shape of the field strength record contribute information of a general nature regarding the physical surroundings near the receiver location.

The field strength records were summarized by noting the maximum, minimum, and average value of field strength on each section of chart. This summary was then plotted with distance as the abscissa, and field strength as the ordinate. The range of field strength in each short section of chart is represented by a vertical line drawn at the average distance from the transmitter. The average field strength in each section is then indicated by a short horizontal mark crossing the vertical line. It will be noted that this type of graph shows the upper

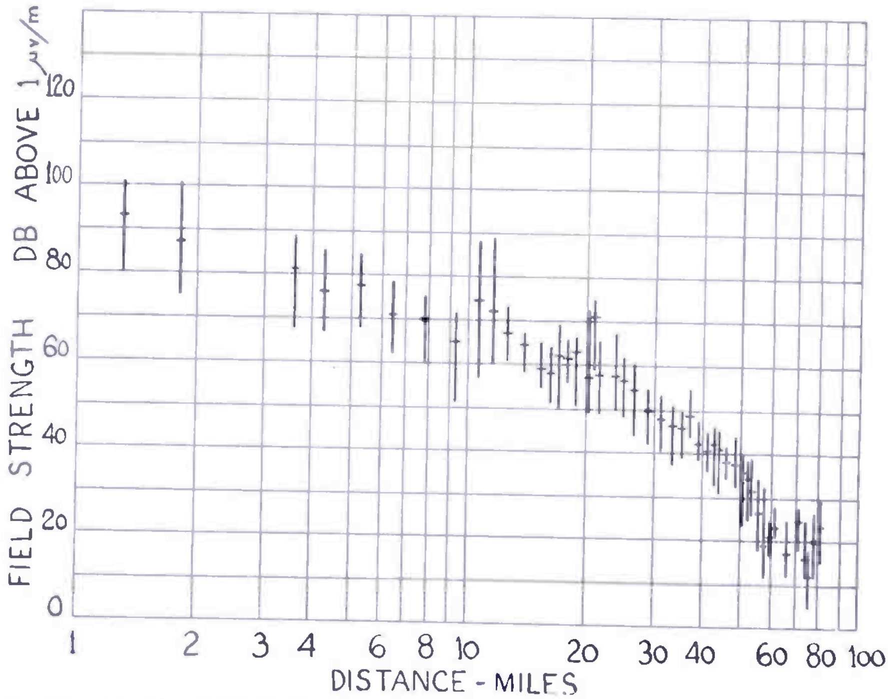


Fig. 5—Summary of field strength record taken between New York City and Camden, N. J. 49.5 Mc, vertical polarization.

and lower limits of field strength, as well as the average value. Figure 4 is the summary of a field strength record taken between New York City and Camden, New Jersey, when using horizontally polarized waves,

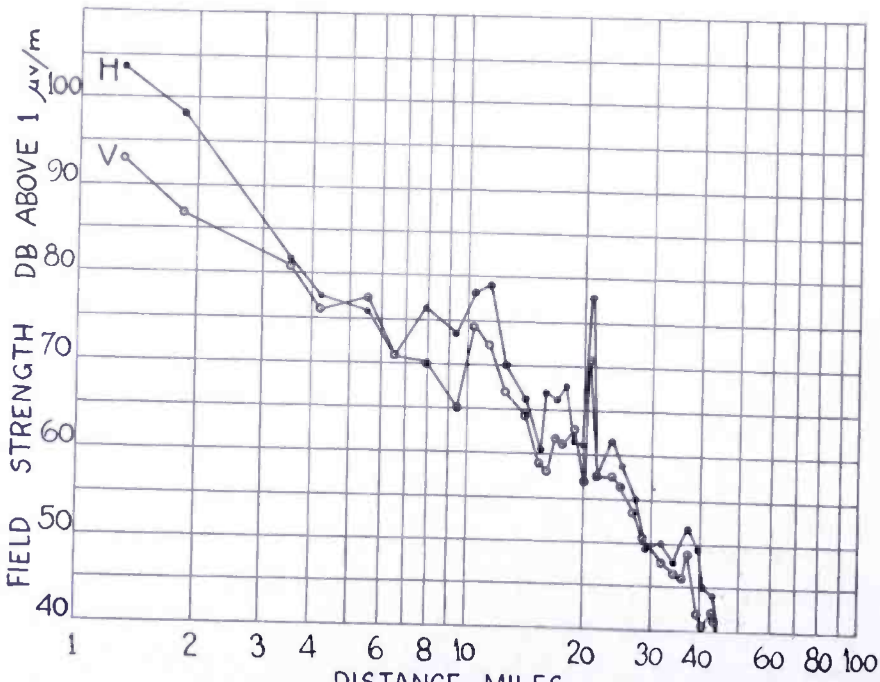


Fig. 6—Comparison of average field strength on horizontal and vertical polarization. New York City to Camden, N. J., 49.5 Mc.

on 49.5 Mc. Figure 5 is the summary of the record taken over the same route, on the same frequency, but with the transmission vertically polarized. The smaller variation in field strength in the country with vertical polarization is apparent without further study.

Comparisons of average field strength on horizontal polarization with average field strength on vertical polarization, are found in Figures 6, 7, and 8. A comparison on each of the three frequencies was chosen to illustrate the consistently stronger average field observed when horizontal polarization was used.

Near the transmitter, the indirect rays reflected from high build-

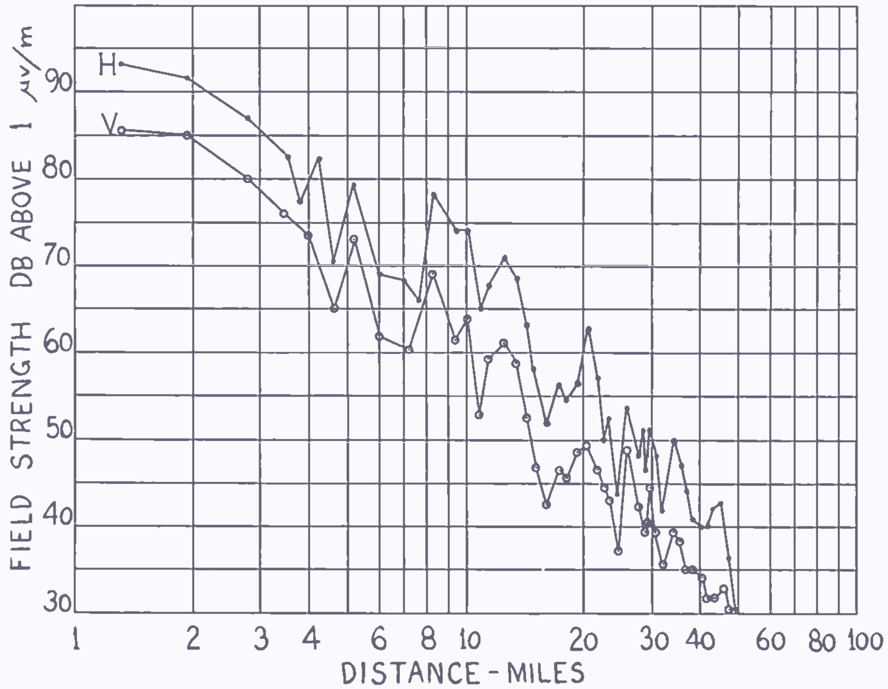


Fig. 7—Comparison of average field strength on horizontal and vertical polarization. New York City to Bridgeport, Conn., 83.5 Mc.

ings down to the street level were often stronger than the direct ray, which was shielded by other buildings. On horizontal polarization, the strength of these off-bearing rays was reduced by the directivity of the receiving doublet, since the doublet was broadside to the transmitter. On vertical polarization, the received signal was the resultant of direct and indirect rays, since the receiving doublet was non-directive. The effect of this condition is not apparent in the average signal comparisons, since only two out of six curves show the vertical polarization to be stronger than the horizontal near the transmitter.

The performance of the three different frequencies was compared by correcting the average field strength curves to an output power of 1 kw. These corrected curves, for horizontal polarization, southwest route, are found in Figure 9. It is evident that the lowest frequency was somewhat the strongest in the city, and the highest frequency was consistently the strongest in the country.

The same kind of a comparison when using vertical polarization indicates that the highest frequency produced the highest average field strength both in the city and in the country, for the same transmitter power. This comparison is found in Figure 10.

A casual comparison of the horizontal and vertical polarization measurements revealed a difference in the maximum-to-minimum variation, over the same route. Graphs comparing the extremes of field

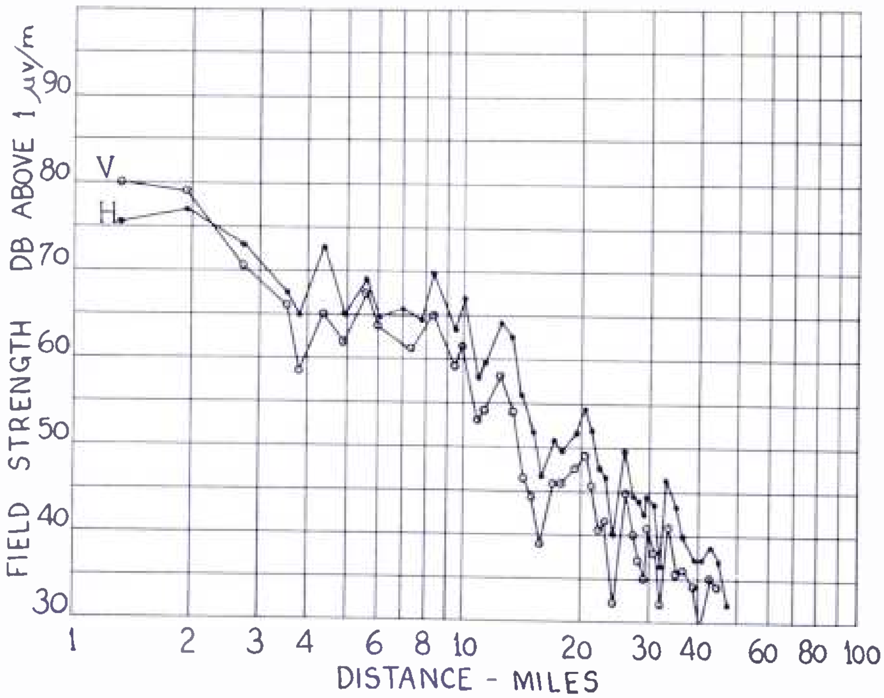


Fig. 8—Comparison of average field strength on horizontal and vertical polarization. New York City to Bridgeport, Conn., 142 Mc.

strength in each section, on horizontal and vertical polarization are shown in Figures 11, 12, and 13. It is apparent that the horizontal polarization was more variable over the greater part of the distance, although the vertical polarization was more variable near the transmitter.

The magnitude of the field strength variation on both polarizations was no doubt influenced by the low receiving antenna height. If the receiving antenna height were increased to thirty or forty feet, the number of obstructions to an optical path would be considerably reduced. Also, irregularities in the topography in the immediate vicinity of the receiving antenna would have less effect on the signal component reflected from the ground.

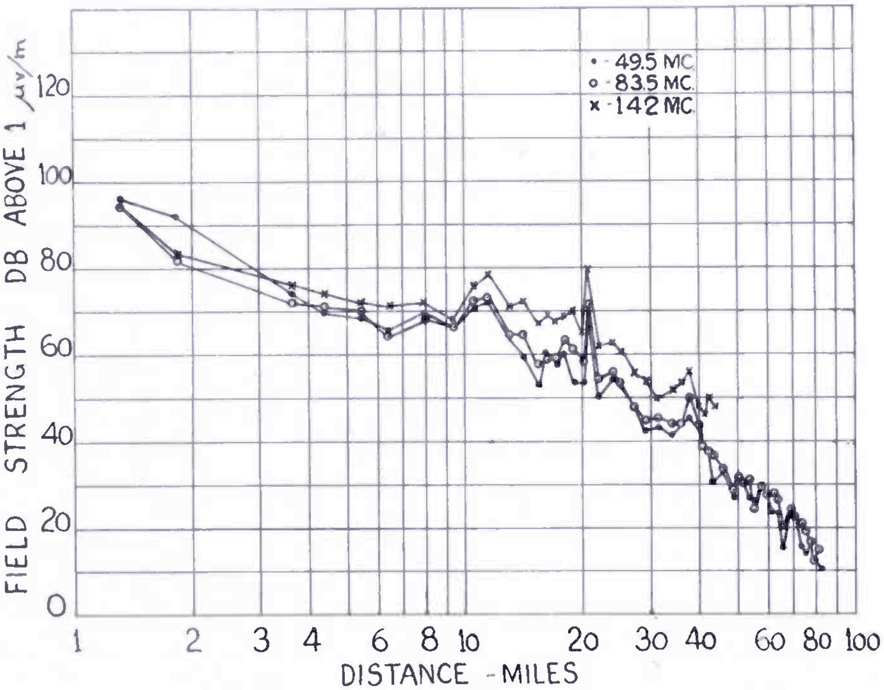


Fig. 9—Comparison of average field strength on 49.5, 83.5, and 142 Mc, corrected to 1 kw antenna power. New York City to Camden, N. J., horizontal polarization.

The trend revealed in the variation curves raised the question of whether the maximum or the minimum values of field strength were

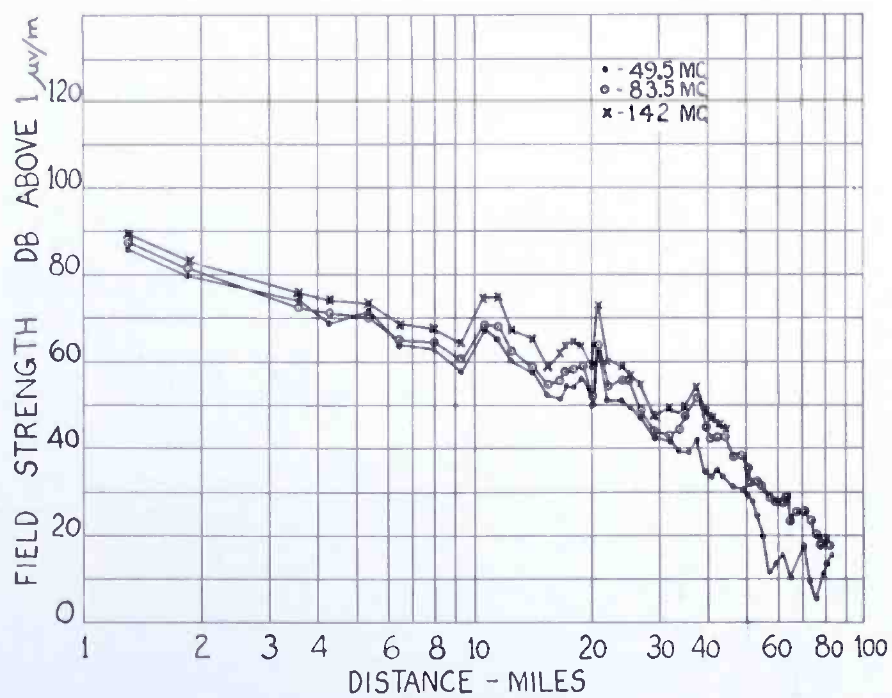


Fig. 10—Comparison of average field strength on 49.5, 83.5, and 142 Mc, corrected to 1 kw antenna power. New York City to Camden, N. J., vertical polarization.

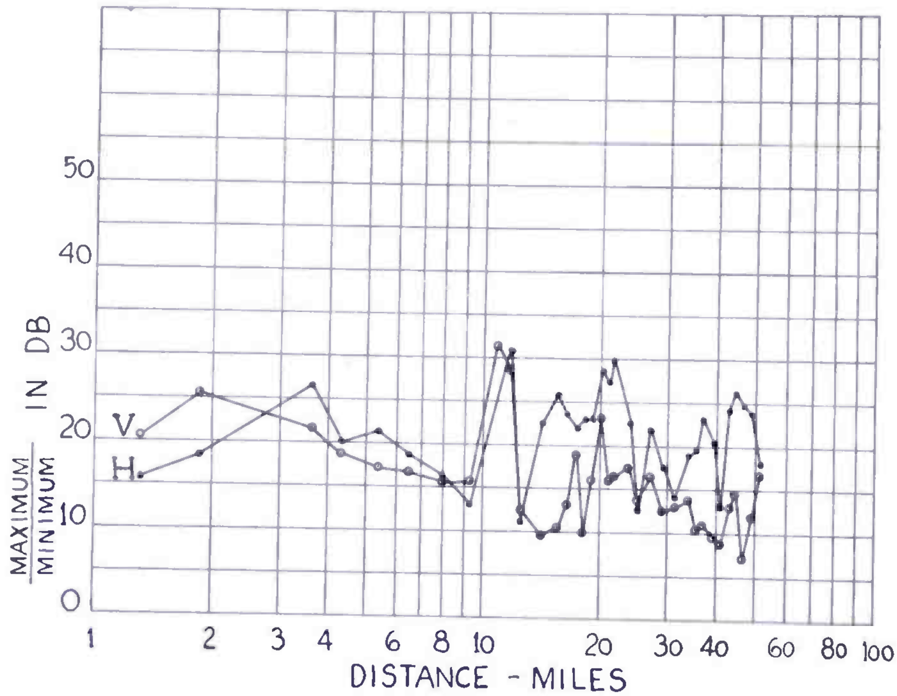


Fig. 11—Variation of field strength on horizontal and vertical polarization. New York City to Allentown, N. J., 49.5 Mc.

showing more variation. The maximums on horizontal polarization were compared with the maximums on vertical polarization for each route, and the same comparison was made between the minimums on

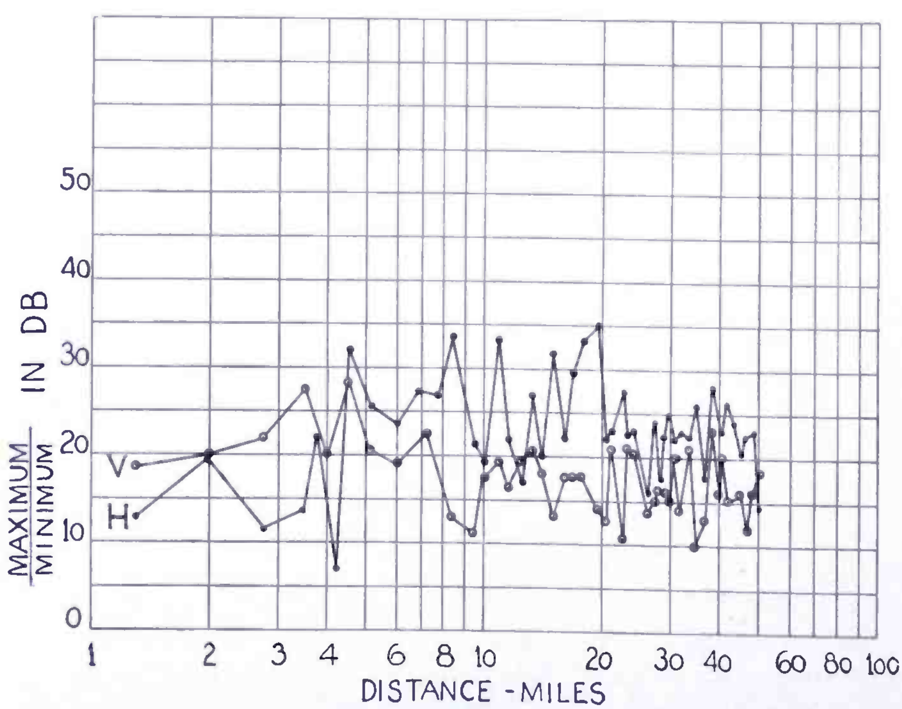


Fig. 12—Variation of field strength on horizontal and vertical polarization. New York City to Bridgeport, Conn., 83.5 Mc.

both polarizations. The averages of these individual comparisons revealed that the maximum values of field strength were greater on horizontal polarization on all recordings. The averages of the comparison of minimums were not so consistent; horizontal polarization was stronger to the north and northeast, while vertical polarization was stronger to the southwest. Most of the roads to the southwest of New York were relatively narrow, with horizontal open wires present for a considerable part of the distance. The shielding effect of horizontal wires on horizontally polarized waves is very noticeable when

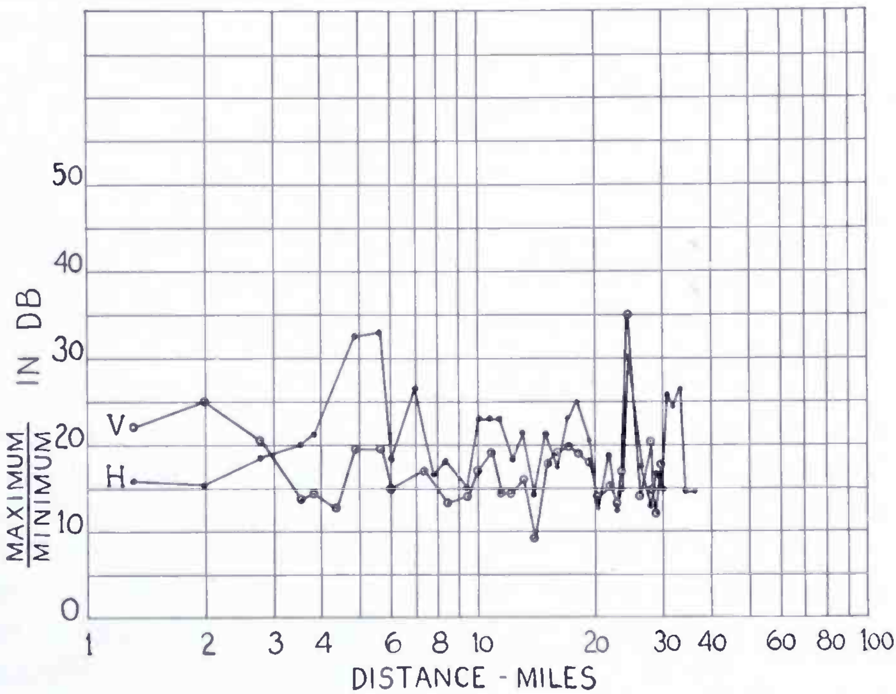


Fig. 13—Variation of field strength on horizontal and vertical polarization. New York City to Darien, Conn., 142 Mc.

recording, causing the minimums to drop to lower values on horizontal polarization.

TABLE I

Summary of field strength comparisons on horizontal and vertical polarization.

Freq. Mc	Direction	Average H/V in DB Variation	Maxima	Minima	Average Field Strength
83.5	Northeast	+6.0	+9.8	+3.8	+8.1
142	Northeast	+2.7	+5.1	+2.4	+4.7
83.5	North	+2.1	+4.8	+2.7	+4.7
49.5	Southwest	+5.7	+5.7	0	+3.1
83.5	Southwest	+7.1	+4.3	−2.8	+1.9
142	Southwest	+6.9	+5.4	−1.5	+3.5

The results of all comparisons have been summarized in Table I, which indicates that, for the territory covered by the survey, horizontal polarization (1) varies over a wider range, (2) has higher maximum values, and (3) produces a higher average field strength than vertical polarization, other things being equal. The values shown in this tabulation are arithmetic averages, which correspond to a geometric average of field strength.

It should be remembered that these results and conclusions apply only to the two general directions covered by the mobile measurements. A comparison made over a salt-water path, or high-conductivity ground, might show the vertical polarization to be considerably more effective than the horizontal polarization.

Conclusion

On the basis of this mobile survey, the following general conclusions regarding horizontal and vertial polarization may be drawn:—

1. Horizontal polarization produced a stronger average field. Average ratio for all frequencies and routes was 4.3 db.

2. Horizontal polarization was more variable over a given distance in the country.

3. Vertical polarization was usually more variable in the city.

4. The greater variation on horizontal polarization was due to higher maximum values.

5. With horizontal polarization and equal power, the lowest frequency (49.5 Mc) produced the strongest average signal in the city, and the highest frequency (142 Mc) produced the strongest average signal in the country.

6. With vertical polarization and equal power, the highest frequency (142 Mc) was strongest both in the city and the country.

Acknowledgment

The field work described in this report was made possible by the cooperation of the Development Group of the National Broadcasting Company, and the Transmitter and Receiver Laboratories of R.C.A. Communications, Inc.

TELEVISION STUDIO TECHNIC*

By

Albert W. Protzman

National Broadcasting Company

Summary—*The studio operating technic as practiced in the NBC television studios today is discussed and comparisons are made, where possible, to motion picture technic. Preliminary investigations conducted to derive a television operating technic revealed that both the theater and the motion picture could contribute certain practices.*

The problems of lighting, scenic design, background projection, and make-up are discussed, with special emphasis on the difficulties and differences that make television studio practice unique.

An explanation is given of the functioning of a special circuit used in television sound pick-up to aid in the creation of the illusion of close-up and long-shot sound perspective without impracticable amount of microphone movement. The paper concludes with a typical television production routine showing the coördination and timing of personnel and equipment required in producing a television program.

IF ONE were forced to name the first requirement of television operating technic and found himself limited to a single word, that word would undoubtedly be "timing." Accurate timing of devices and split-second movements of cameras are the essentials of television operation. Personnel must function with rigid coördination. Mistakes are costly—they must not happen—there are no second chances.

Why such speed and coördination? Television catches action at the instant of its occurrence. Television does not allow us to shoot one scene today and another tomorrow, to view rushes or resort to the cutting room for editing. Everything must be done as a unit, correct and exact at the time of the "takes"—otherwise, there is no television show.

Now, to discuss some preliminary investigations conducted before production was attempted, and to describe the equipment and technic used in meeting these production requirements. Technical details are deliberately omitted. Wherever possible, we shall compare phases of television operation with their counterparts in motion picture production.

For so new a medium as television it is, of course, an impossibility to present a complete and permanently valid exposition. Television technic and apparatus constantly advance. Some technic now current may be outmoded in a day or a month. We have only to recall the early days of motion picture production, when slow-speed film and

* Reprinted from the *Journal of the Society of Motion Picture Engineers*, July, 1939.

inferior lenses were a constant limitation. So, with television, it is already possible to envision more sensitive pick-up tubes that will permit the use of smaller lenses of much shorter focal length, thus eliminating many of today's operating difficulties.

Production Technic Investigations

In May, 1935, the Radio Corporation of American released television from its research laboratories for actual field and studio tests. Long before the first program was produced in the middle of 1936, plans were laid, based on extensive research into the established entertainment fields, for the purpose of determining in advance what technics might be adaptable to the new medium of television. From the stage came the formula of continuity of action, an inherent basic requirement of television. This meant memorized lines and long rehearsals. Prompting could not be considered, for, as you know, the sensitive microphone which is as much present in television as it is in sound motion picture production, does not discriminate between dialog and prompting.

From the motion picture studio came many ideas and technics. If television is a combination of pictures with sound, and it is, no matter what viewpoint is taken, the result spells in part and for many types of programs, a motion picture technic at the production end. However, enough has already been said about the peculiarities of television presentation to justify saying that the movie technics do not supply the final answer. There remained the major problem of preserving program continuity without losing too much of motion picture production's flexibility. Our present technic allows no time for adjustments or retakes. Any mistake immediately becomes the property of the audience. The result of the entire investigation led to what we think is at least a partial answer to the problem. This technic, we hope, will assist considerably in bringing television out of the experimental laboratory and into the field of home education and entertainment.

General Layout of Facilities

In order to present a clearer view of our problems, we shall give a brief description of our operating plant. The present television installation at the National Broadcasting Company's headquarters in the RCA Building, New York, N. Y., consists of three studios, a technical laboratory, machine and carpenter shops, and a scenic paint shop. Our transmitter is located on the 85th floor of the Empire State Building. The antenna system for both sight and sound is about 1300 feet above the street level. Both the picture and sound signals are relayed from the Radio City Studios to the video and sound transmitters either by coaxial cable or over a special radio link transmitter.

One of the studios is devoted exclusively to televising motion picture film, another to programs involving live talent, and the third for special effects. It is the operation of the live-talent studio with which we are concerned in this paper.

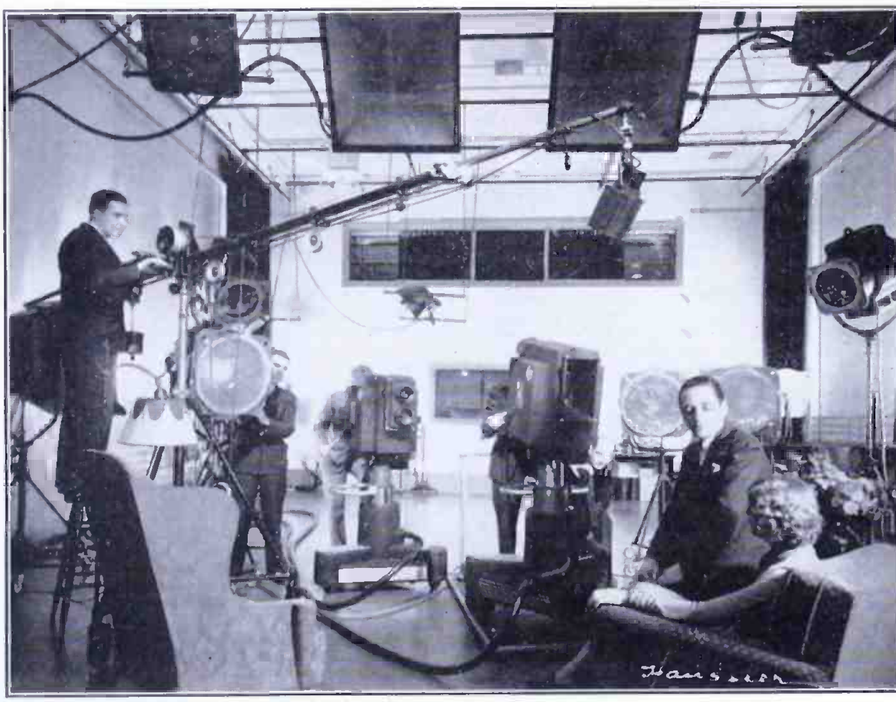


Fig. 1*(a)*—General layout of live-talent studio; control room at upper rear.

Description of Live-Talent Studio

Figure 1 (*a*) shows the general layout of the live-talent studio. The studio is 30 feet wide, 50 feet long, and 18 feet high. Such a size should not be considered a recommendation as to the desired size and proportions of a television studio. The studio was formerly a regular radio broadcasting studio, not especially designed for television. To anyone familiar with the large sound stages on the motion picture lots, this size may seem small (Figure 1 (*b*)). Yet, in spite of our limited space, some involved multi-set pick-ups have been successfully achieved by careful planning. Sets, or scenes, are usually placed at one end of the studio. Control facilities are located at the opposite end in an elevated booth, affording full view of the studio for the control room staff. Any small sets supplementing the main set are placed along the side walls as near the main set as possible, and in such position as to minimize camera movement. At all times, we reserve as much of the floor space as possible for camera operations and such floor lights as are absolutely essential. At the base of the walls and also on the ceiling are scattered numerous light-power outlets to minimize the length of lighting cables. At the rear of the studio is a permanent projection room for background projection.

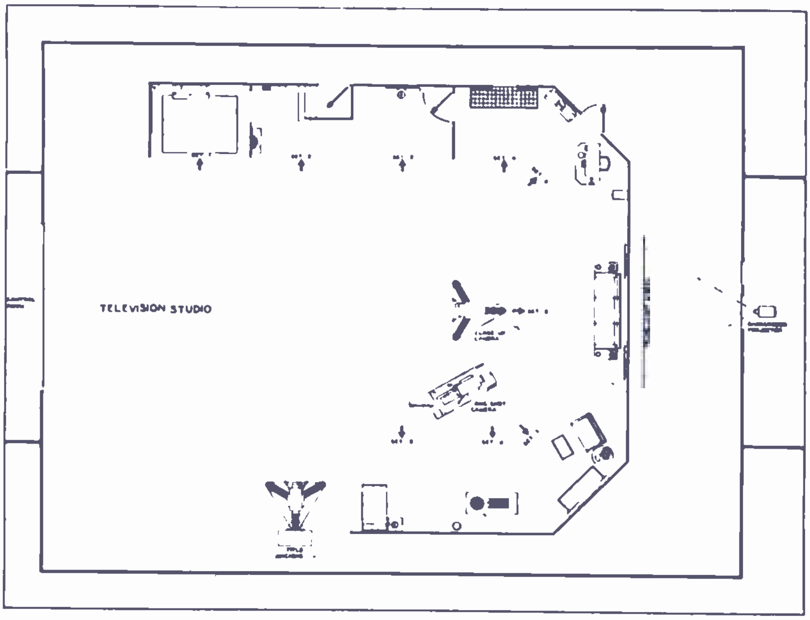


Fig. 1(*b*)—Television studio floor plan.

Camera Equipment

The studio is at present fitted for three cameras. To each camera is connected a cable. This cable is about two inches in diameter and fifty feet long; it contains 32 conductors including the well known coaxial cable over which the video signal is transmitted to the camera's associated equipment in the control room. The remainder of the conductors carry the necessary scanning voltages and current supplies for the camera amplifiers, interphone system, signal lights, etc. From this description, it is apparent that adding another camera in a television studio involves a much greater problem than that of moving an extra camera into a motion picture studio. In television, it is necessary to add an extra rack of equipment in the control room for each additional camera.

Movement of Cameras

One camera, usually the long-shot camera using a short-focal length lens, is mounted on a regular motion picture type dolly to insure stable movements. The handling of the dolly is done by a technician assisting the camera operator. It is impracticable to lay tracks for dolly shots as is often the motion picture practice, because usually each camera must be moved frequently in all directions during the televising of a studio show. Naturally, dolly tracks would limit such movement. The other television cameras utilize a specially de-

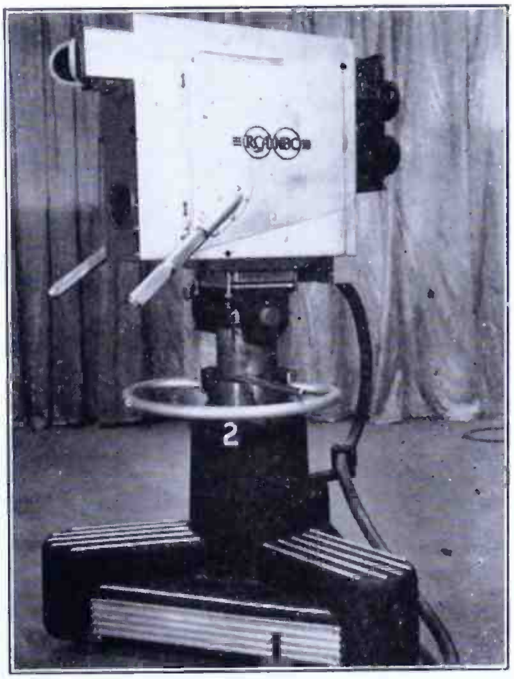


Fig. 2—Studio camera.

signed mobile pedestal (Figure 2). Cameras mounted on these pedestals are very flexible and may be moved in and out of position by the camera operators themselves. Built into the pedestals are motors which elevate or lower the camera; this action is controlled with push-buttons by the camera operators. A panning head, similar to those used for motion picture cameras, is also a part of the pedestal. It is perhaps needless to stress here that one of the strict requirements of a television camera is that it must be silent in operation. In the electronic camera proper there are no moving parts other than those used for focusing adjustments; hence, it is a negligible source of noise. When camera pedestals were first used they were the source of both mechanical noise and electrical disturbance when the camera-elevating motor was in use. Since then this problem has been overcome, and it can be stated that the entire camera unit is now free of objectionable mechanical noise or electrical surges.

Lens Complement

Each camera is equipped with an assembly of two identical lenses displaced 6 inches vertically. The upper lens focuses the image of the scene on a ground-glass which is viewed by the camera operator. The lower lens focuses the image on the "mosaic," the Iconoscope's light-sensitive plate. This plate has for its movie counterpart the film in a motion picture camera. The lens housings are demountable and interchangeable. Lenses with focal lengths from 6½ to 18 inches are used at present. Lenses of shorter focal length or wider angle of pick-up can not be used since the distance between the mosaic and the glass envelope of the Iconoscope is approximately 6 inches. Lens changes can not be effected as fast as on a motion picture camera, since a turret arrangement for the lenses is mechanically impracticable at present. However, it is probably safe to say that future advances in camera and Iconoscope design will incorporate some type of lens turret. Ordinarily, one camera utilizes a 6½-inch focal length lens with a 36-degree angle, for long shots, while the others use lenses of longer focal lengths for close-up shots. Due to its large aperture, the optical system used at present has considerably less depth of focus than those used in motion pictures, making it essential for camera operators to follow focus continuously and with the greatest care. This limitation will probably be of short duration, since more sensitive Iconoscopes will permit the use of optical systems of far greater depths of focus.

It is desirable here to point out a difference in focusing technic between motion picture cameras and television cameras. "Follow-focus" in motion pictures occurs practically only in making dolly shots. For all fixed shots, the lens focus is set, the depth of focus being sufficient to carry the action. Also, it is the duty of the assistant cameraman to do the focusing. This relieves the cameraman of that responsibility and allows him to concentrate on composition, action, and lighting. In television, the camera operator must do the focusing for fixed shots and dolly shots alike. This added operation, at times, is quite fatiguing.

Vertical parallax between the view finder lens and the Iconoscope lens is compensated for by a specially designed framing device at the ground-glass that works automatically in conjunction with the lens-focusing control. It may be of interest to note here that at first the television camera had no framing device. This meant that images, in addition to being inverted as they are in an ordinary view-finder, were also out of frame. The camera operator had to use his judgment in correcting the parallax. With this new framing device, the operator now knows exactly the composition of the picture being focused on the mosaic in his camera. The framing device can be quickly adjusted to accommodate any lens between 6½ and 18 inches focal length.

Because of the fact that several cameras are often trained on the same scene from various angles, and because all cameras are silent in operation, performers must be informed sometimes—such as when they are speaking directly to the television audience—which camera is active at the moment. Two large green bull's-eye signal-lamps

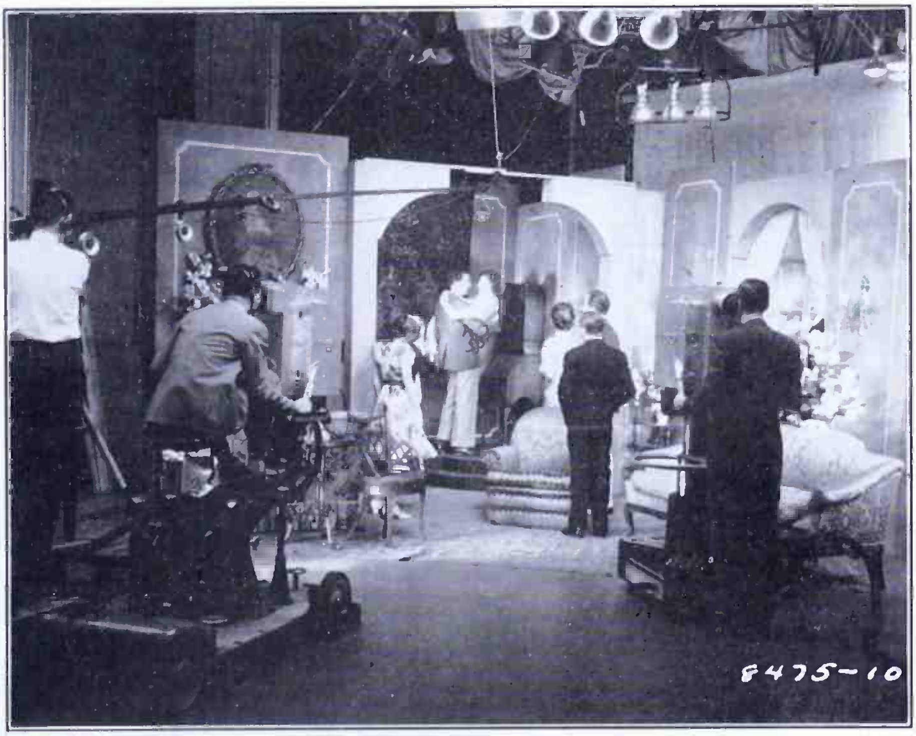


Fig. 3—Typical television set.

mounted below the lens assembly are lighted when the particular camera is switched "on the air."

SET LIGHTING

There are two outstanding differences between television lighting and motion picture lighting. A much greater amount of key light is required in television than in motion pictures. Also, a television set must be lighted in such a way that all the camera angles are anticipated and properly lighted at one time. Floor light is held to a minimum to conserve space in assuring maximum flexibility and speed of camera movements. Great care must also be taken to shield stray light from all camera lenses. This task is not always easy, since, during a half-hour performance, each camera may make as many as twenty different shots. Just as excessive leak-light striking the lens will ruin motion picture film, it has a definitely injurious effect upon the photo-sensitive mosaic and upon the electrical characteristics of the Iconoscope. A direct beam of high-intensity light may temporarily paralyze a tube, thus rendering it useless for the moment.

SETS

Television sets (Figure 3) are usually painted in shades of gray. Since television reproduction is in black and white, color in sets is

Fig. 4—Background projection window shot.

relatively unimportant. Chalky whites are generally avoided because it is not always possible to keep "hot lights" from these highly reflective surfaces which cause a "bloom" in the picture. This, in turn, limits the contrast range of the system. Due to the fact that the resolution of the all-electronic system is quite high, television sets must be rendered in considerable detail, much more, in fact, than for a corresponding stage production. As in motion picture production, general construction must be as real and genuine as possible; a marked difference, for instance, can be detected between a painted door and a real door. On the legitimate stage, a canvas door may be painted with fixed highlights; that is, a fixed perspective, because the lighting remains practically constant, and the viewing angle is approximately the same from any point in the audience. But, in television the perspective changes from one camera shot to another. Painted perspectives would therefore be out of harmony with a realistic appearance. This is also true in motion picture work. Sets must also be designed so that they can be struck quickly with a minimum effort and noise because it is often necessary to change scenes in one part of a studio while the show is going on in another part. At present, we find it desirable to construct television sets in portable and lightweight sections without sacrificing sturdiness.

Background Projection

The problems of background projection in television differ somewhat from those encountered in motion pictures. More light is necessary because of the proportionately greater incident light used on the sets proper (Figure 4).

Considering the center of a rear-screen projection as zero angle, we must make it possible to make television shots within angles of at least 20 degrees on either side of zero without appreciable loss of picture brightness. This requirement calls for the use of a special screen having a broader viewing angle than those used in making motion picture process shots. Also, in motion pictures, the size of the picture on the screen can be varied to the proper relation to the foreground for long shots or close-ups. For television, the background picture size can not be changed once the program starts. Our background subject matter must also be sharp in detail and high in contrast for good results.

At present, only glass slides are used. A self-circulating water-cell is used to absorb some of the radiant heat from the high-intensity arc. Also both sides of the slide are air-cooled. These precautions permit the use of slides for approximately 30-minute periods without damage.

Make-Up

This may be a suitable time to correct some erroneous impressions concerning the type of make-up used in television. It has never been necessary to use gruesome make-up for the modern all-electronic-RCA television system. At present, No. 26 panchromatic base, similar to that used for panchromatic film, and dark red lipstick is being used satisfactorily. From the very beginning, we have made tests to determine the proper color and shades of make-up, keeping in mind that a color closely approximating the pigmentation of the human skin is most desirable from the actor's psychological standpoint.

The Control Room

Now, a few words about the operations in the studio control room during a televised production (Figure 5). All camera operators in the studio wear head-phones through which they receive instructions from the control room. Directions are relayed over this circuit by the video engineer or the production director. Here the televised images are observed on special Kinescope monitors and necessary electrical adjustments are made. Alongside each of these monitoring

Fig. 5—The television control room. Note the two Kinescope monitors in the upper left corner.

Kinescopes is a cathode-ray oscilloscope which shows the electrical equivalent of the actual picture. Two monitors are provided in order that one may be reserved for the picture that is actually on the air, while the other shows the succeeding shot as picked up by a second or third camera. This enables the video engineer to make any necessary electrical adjustments before a picture goes on the air.

Seated immediately to the left of the video engineer is the production director whose responsibility corresponds to that of the director of a motion picture. He selects the shots and gives necessary cues to the video engineer for switching any of the cameras into the outgoing channel. The production director has, of course, previously rehearsed the performance and set camera routines in conjunction with the camera operators and the engineering staff. The camera operator has no control to switch his camera on the air. All camera switches, which are instantaneous, are made by electrical relays controlled by buttons in the control room. At present, the video engineer's counterpart in motion picture work is the editor and the film processing laboratory.

To the left of the production director sits the audio control engineer whose responsibility is entirely separate from that of the video engineer. He also is in a position to view the monitor, and may communicate by telephone with the engineer on the microphone boom. The audio engineer is responsible for sound effects, some of which are dubbed in from records. His job is somewhat similar to that of the head sound engineer on a motion picture production. Thus, we have the control room staff—three men who have final responsibility for the success of the completed show.

An assistant production man is also required on the studio floor. Wearing headphones on a long extension cord, he is able to move to any part of the studio while still maintaining contact with the production director in the control room during a performance on the air. Actors require starting cues, titles require proper timing, and properties and even an occasional piece of scenery must be moved. The assistant director supervises these operations and sees that the instructions of the production director are properly carried out.

Members of the studio personnel also to be mentioned include lighting technicians, the property man, and scene shifters, whose responsibilities parallel those of their motion picture counterparts. Specially trained men are also needed for operating title machines. In the future all titling will undoubtedly be done in a separate studio inasmuch as operating space in a television studio is at a premium. Today, however, title machines do operate in the studio and require the utmost care in handling. Types of titles used include dissolves and wipes similar to those used in moving pictures.

Sound Reproduction

As in motion picture work, a microphone boom is used in television production, and is operated in a similar way. Perspective in motion picture sound is accomplished by keeping the microphone, during a long shot, just out of the picture and moving it down closer to the action as the camera moves in for a close-up, thus simulating a natural change in perspective. In television this is not always possible because there are always three cameras to consider. This same condition prevailed in the early days of motion pictures when it was thought desirable to take a complete scene, shooting both long-shot and close-up cameras, at one time. In the television studio at least one camera is always set for a long shot while the others are in position for closer shots. If the microphone is placed in such a position as to afford a "natural" perspective for close-ups, the succeeding switch to a long shot would reveal the microphone in the shot. You in motion pictures can order a retake; in television broadcasting we can not rectify the mistake. It is quite obvious, therefore, that the man on the boom can not lower his microphone to the "natural" position for each camera shot. We therefore place the microphone in a position just out of range of the long shot. In order to accomplish some sense of perspective between long and close-up shots, a variable equalizer that drops the high and low ends of the spectrum is automatically cut into the audio circuits when the long-shot camera is on the air. In this operation, sufficient change in quality and level is introduced to aid the illusion of long-shot sound perspective. Of course, when a close-up camera is switched in, the audio returns to the close-up perspective quality once more. This may be called remote control sound perspective.

Special sound effects, music, etc., from the studio picked up from recordings are mixed in the control room. In motion pictures, some of the effects and most of the music are dubbed in after the actual shooting of the scene.

The general acoustical problems in a television studio are similar to those in a motion picture sound-stage. Walls and ceiling should be designed for maximum absorption to permit faithful exterior speech pick-up. A stage or studio must be designed to enable presentation of an exterior or an interior scene. With the studio designed for maximum absorption, illusions of exterior sound characteristics can be created. For interiors, the hard surfaces of the sets and props offer sufficiently reflective surfaces to create the indoor effect.

Typical Production Routine

After the foregoing discussion of the equipment and personnel, it may be interesting to follow an actual production from the beginning of rehearsal to its final presentation. For this example, assume that we are to produce a playlet (Figure 6). When the scenery has been erected, the first rehearsals begin without the use of cameras or lights. Besides familiarizing the actors with their lines, the rehearsals afford the production director and the head camera operator an opportunity to map out the action of the play. All action, including camera shots, cues, and timing, is noted on a master script which thereafter becomes the "bible" of the production. Timing is very important because of the necessity of having a particular act time in with the other acts or film subject.

After several hours of rehearsing, the first equipment rehearsal is called. Cameras are checked electrically and mechanically. Focus controls and framing devices are lined up so that correct focus on the ground-glass is also correct focus on the mosaic plate. This completed, the cameras are ready for rehearsal. With the scene properly lighted, the camera operators begin working out movements to pick up the

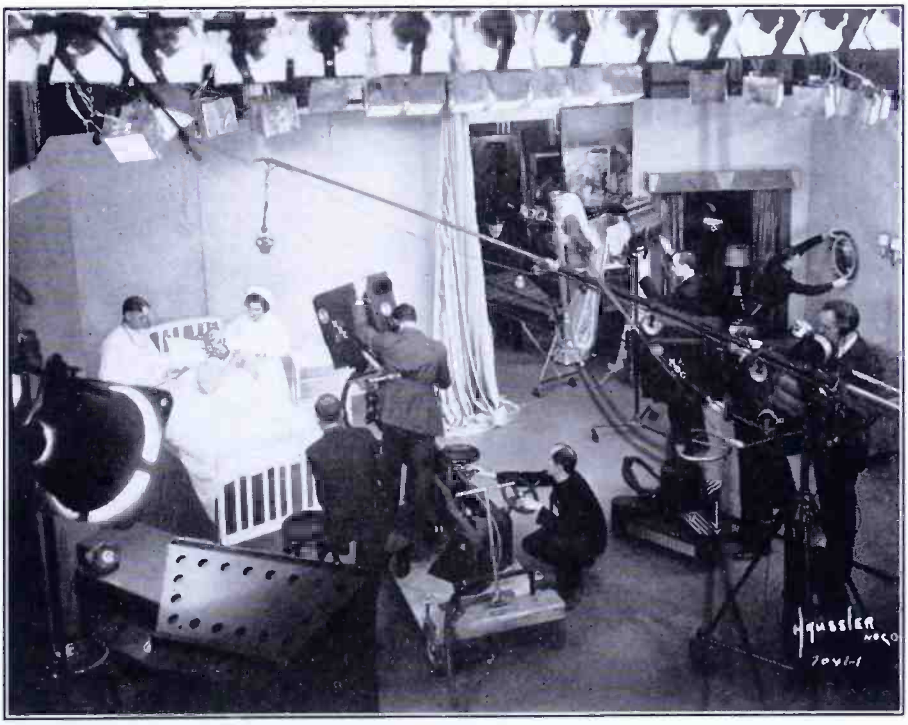


Fig. 6—*(Left)* Scene on the air. *(Right)* setting up for next scene.

desired shots in the proper sequence. The production director instructs the staff and personnel from the control room, speaking over a public-address system. Each shot is worked out and its camera location marked on the floor. At times, the actors may unconsciously depart slightly from the rehearsed routine during an actual show; the camera operator must be prepared and alert to make the best of the situation regardless of all previous floor markings. Continuity is so planned that while one camera is taking the action, another camera is moving to a new location and composing a new shot to be switched on at the proper time. This frees the first camera, which can now move to a third location, and so on. Sometimes during a twenty-minute performance each camera may take twenty different shots. Of course, besides different floor locations, the height and angle of the cameras must be varied to comply with good composition. During rehearsals, timing must frequently be revised to allow for the actual camera movements.

Finally, a dress rehearsal is scheduled. The complete program is televised, including any film subjects or slides that may be needed to complete the program. Frequently the program will begin with a short film leader, followed immediately by a newsreel or a short subject, the film portion of the program coming from the film-televising studio. While the film is running, the live-talent studio is continuously warned as to the time remaining before it must take over the program. Once the studio program goes on the air the production director is no longer able to use the public address system to communicate with the personnel in the studio. Instead, he uses a telephone circuit to his assistant in the studio, and, through the video engineer, communicates by phone with the camera operators.

Another standby warning is usually given when there is one minute to go. Then, as the cue to begin comes, the green light on the title camera is lighted. From this point, continuity must be rigidly preserved. As titles move from one to another, appropriate music is cued in and actors are sent to their opening positions.

With the completion of titles, the image is faded out electrically and cameras are switched to the opening shot. Performers begin their action on a silent cue from the assistant director, who is instructed from the control room. During this first scene, the camera previously picking up titles moves quickly into position to shoot a second view of the action. Again cameras are switched, permitting the first to move to a new position; and so the action proceeds. If the play has several scenes, the concluding shot of the first scene is taken by one camera while others line up on the new scene and wait for the switch. Frequently, there are outdoor scenes. These are filmed during the first stages of rehearsal for transmission from the film studio at the proper time during the performance. The switch to film is handled exactly as another camera switch, except that the switch is to the film studio instead of to one of the studio cameras. The projectionist must be warned in advance to have his projector up to speed and "on the air" at the proper instant to preserve the production continuity. This requires very critical timing, as you can well appreciate. When the film is completed the studio cameras again take over the next interior scene.

Upon completion of the studio portion of the program, one camera lines up on the final studio title, which usually returns the program to the film studio for a concluding film subject.

Since the first program on July 7, 1936, many television programs have been produced. Each has been a serious attempt at something new. Although much has been accomplished, there remain a vast number of unknowns to be answered before it can be said that television's potentialities have been even partially realized. Today, as this paper has indicated, television bears many points of similarity to motion pictures. As a matter of fact, it is likely that television would be somewhat handicapped if it were unable to borrow heavily from a motion picture production technic that has been built up by capable minds and at great expense over a period of many years. Infant television is indeed fortunate to have such a wealth of information at its disposal. Possibly continued experimentation will lead us toward a new technic distinctive of television. During its early years, however, television must borrow from all in creating for itself a book of rules. The first chapter of that book is scarcely written.

TELEVISION LIGHTING*

By

William C. Eddy

National Broadcasting Company

Summary—*Lighting a television production presents many problems peculiar to this new field of public entertainment. These problems have necessitated the redesign of lighting equipment and the establishment of a simplified technic for handling the equipment that differs radically from moving picture practice.*

To cope properly with the lighting requirements of the continuous action sequences, characterizing television productions, a system employing inside silvered incandescent lamps in a standardized unit was developed by NBC engineers. Based on multiple standardized group of 1½ kw each, these units are used in both the foundation light and modeling equipment of the television studios in Radio City, thus insuring quantitative as well as qualitative control of lighting by the personnel.

With cameras generally in motion and an average duration of pick-up from one camera a matter of seconds, the problem of modeling in the sets becomes acute. This appears to be satisfactorily solved by the technic now in use wherein the major interest is centered around the close-up camera. Even this solution, however, required new and ingenious equipment to maintain light in the sets and still give floor precedence to the cameras and sound equipment.

While NBC at the present time has appeared to have standardized on the inside silvered lamp, exhaustive tests were carried out in an attempt to utilize more orthodox equipment. Actual tests under production conditions proved, however, that certain requirements of space, weight, and flexibility could not be had without a serious sacrifice of foot-candles on the set, resulting in the present set-up of equipment and personnel that are handling the television lighting assignment in the East.

ALTHOUGH the practical application of lighting to the presentation of television studio programs will admittedly be subject to further improvement, the imminence of a public television service warrants a description of the lighting equipment and operating technic which the National Broadcasting Company has worked out as a result of several years of experimentation in this field.

This description covers primarily the lighting developments since 1935 when the Radio Corporation of America launched an extensive experimental field-test of television. Of considerably greater scope

* Reprinted from the *Journal of the Society of Motion Picture Engineers.* July, 1939.

than previous tests, it was designed to permit a pre-commercial analysis of the art through a combined appraisal of the laboratory-reared electrical system and a comprehensive survey of the problems introduced by regular production of programs.

Starting with studio lighting equipment similar to that used in moving pictures, we have gradually evolved a reasonably satisfactory solution of our illumination problem that has resulted in a new and interesting layout of equipment applicable to the demands imposed by television studio operation. This was achieved largely through simplification of the equipment and the technic involved in handling it.

To permit both engineer and director to discuss the lighting set-up with a common terminology, and thus facilitate presentations, we also simplified the existing abstract definitions of light into two separate and distinct classifications: namely, foundation and modeling light.

Foundation light, according to our standards, is the non-characteristic flat illumination of a set, irrespective of its origin or amount. It is primarily the light energy necessary to create an electrical picture in the cameras and provide a foundation to which we can add the characteristic or dynamic quality of modeling light.

Modeling light is any illumination that adds to the contrast or delineation of the picture. It may be from overhead, from the floor, or from the back, but according to our definition, it must create some characteristic highlight or shadow, as opposed to the flat illumination function of the foundation lights.

It was, then, the creation of a satisfactory lighting installation for television rather than the adaptation of equipment and technics geared to an older art that paced our developmental work. It may help to follow the reasoning behind our transition from motion picture lighting into the present installation of incandescent sources, if we consider chronologically the television studio work at Radio City during the formative period from 1935 to the fall of 1938.

A rough analysis of the requirements for a satisfactory system seemed to indicate that flexibility and efficiency were the paramount factors to be considered, although glare and radiant heat from the units had to be taken into account. Of necessity, the light produced had to be a high-level diffused illumination in quantities encountered only in the color-film studios. In addition, television required that the operation, upkeep, and maneuvering of this light be of such simplicity that one or two men could satisfactorily handle routine productions. We naturally turned to the standardized fixtures of the moving picture lots for our first tests. In the Radio City studio we installed routine spots and broads. Due to the limitation of a nineteen-foot ceiling, a practical light bridge was out of the question. As a substitute, the major portion of our lighting equipment was installed on portable stands. Figures 1 and 2 show the arrangement of the apparatus for our first television program from Radio City in 1936.

Fig. 1—A stage set-up in the television studio.

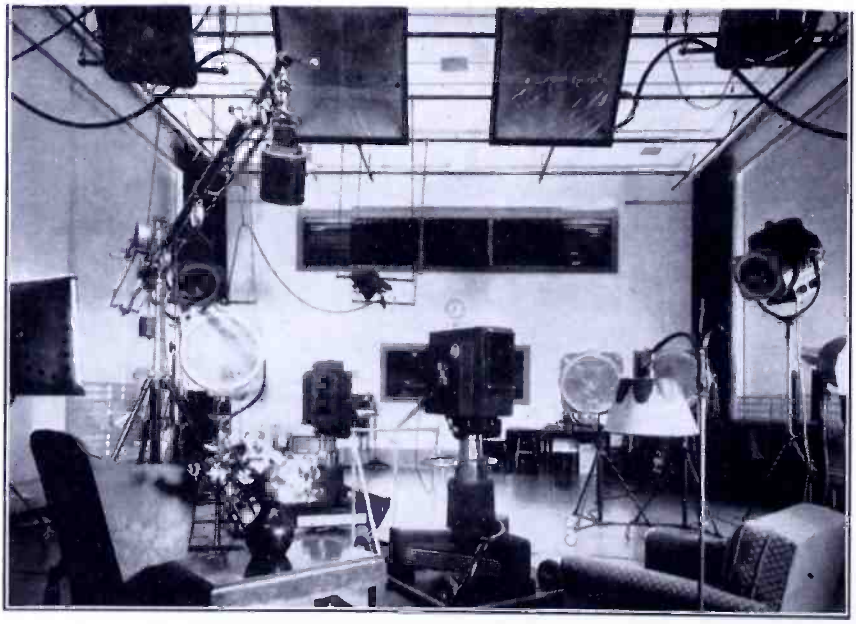


Fig. 2—View of studio showing equipment.

From a quantitative standpoint, we had little to criticize in this installation, but it was immediately apparent that the excessive glare and operational requirements of such a battery of lights precluded their general use in television. An attempt was made to redesign and redistribute these units, but with little or no success, indicating conclusively that equipment of such power and concentration could not be left unattended throughout a television sequence and that the proper manipulation of this type of illumination required a lighting personnel of considerable magnitude.

Our next step was a gradual conversion from the concentrated type of unit to the more diffused and uniform light produced by scoop reflectors and floor broads. Focusing spots and suns were still maintained in the studio, but their function was limited to modeling rather than producing the foundation illumination. Lack of space for opera-

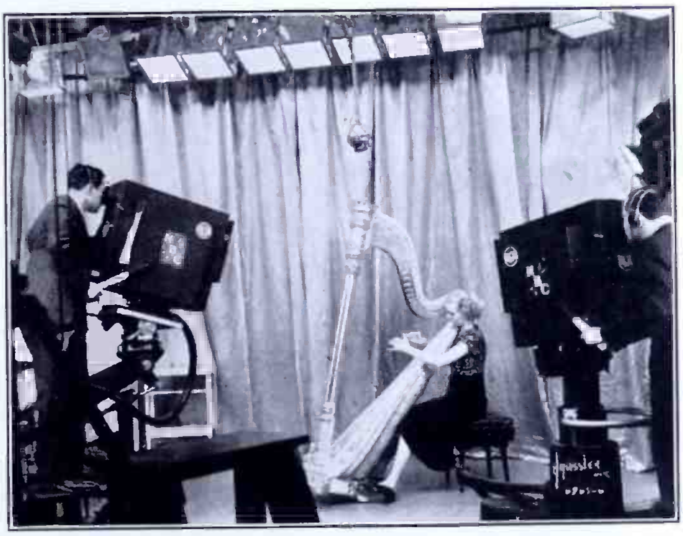


Fig. 3—Illumination by battery of 500-watt units.

tion, weight, and their general inefficiency coupled with unbearable glare on the set soon proved their impracticability even though the unattended light produced by high-efficiency lamps met the requirements of the production staff. During this period, little attempt was made to do more than spill into the sets a predetermined quantity of shadowless light lacking the characteristic modeling that might prove embarrassing in certain sequences. Such a technic reduced the personnel to a minimum, to be sure, but it also produced a television picture in the field that was flat, non-dimensional, and on the whole, highly unsatisfactory from the program standpoint.

Our next experimental step toward a television lighting system came with the installation of a battery of 500-watt units (Figure 3), each equipped with separate reflector and lens systems. These lights were positioned on a gridiron over a single set in such a manner that they would produce a cube of uniform, nondirectional illumination that, it was hoped, would approximate the character and modeling obtained under high-intensity diffused light. Needless to say, the resultant picture showed the effect of flat front lighting. Again the spots and suns were brought out from the storeroom and put into operation as modeling units in an attempt to create above this pedestal of 1500 foot-candles the highlights and shades that had been destroyed by the basic arrangement of the foundation-light installation. Because this system of multi-unit lighting was the first radical departure from orthodox lighting practice and the forerunner of our present studio equipment, it might be well to go into more detail concerning its advantages and shortcomings.

Fig. 4—The single-six mounting.

Coupled with the failure of this installation to produce the required quality of light were several equally important deficiencies: namely, lack of flexibility, excess weight, and great heat radiation. By reason of the bulk of the single unit alone it was necessary to select a certain area to be illuminated, a limitation that required the program group to parade their subjects within the confines of a limited stage. This placed a definite limitation on the efforts of this program group. The weight of the installation closely approached the safe load limit of our acoustical ceiling, making impossible the addition of further equipment above the set to reinforce the existing light or to create special light effects. The unit inefficiency of each lamp, lens, and exterior reflector created an ambient heat problem that severely taxed the air-conditioning service to this particular studio. These deficiencies made the adoption of this system inadvisable but did indicate the direction of our next step.

Photometric tests, conducted in the studio, have already indicated that the new inside silvered spotlight would deliver into an area more light per watt than the lens, lamp, and reflector assembly or the standard incandescent bulb and exterior scoop. This new bulb was light in weight and of relatively small envelope size in the wattage required. It remained to design a fixture that would permit simple adjustment in elevation and direction to satisfy the requirements of the multi-set productions proposed by the program staff. Figure 4 shows such a mounting, known as the "single six." It incorporates

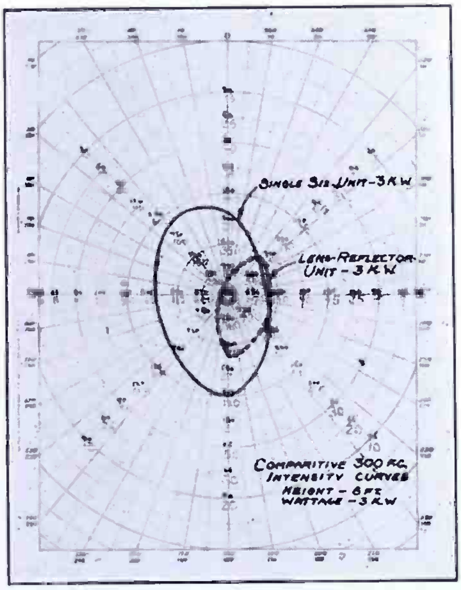


Fig. 5—Light distribution curves of single-six unit and lens reflector unit.

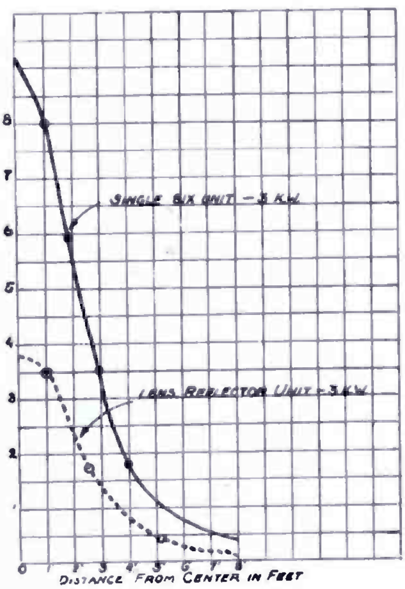


Fig. 6—Photometric distribution of the beam about the center-line.

six 500-watt spotlights on a framework of thin-walled steel tubing, so arranged that the center-to-center distance between lights is ten inches. This insures that the light-beams interlock at a distance of eight feet from the fixture and that the light arriving on the set is relatively free from spots and secondary shadows. The total weight of the fixture, equipped with spots, is slightly less than 19 pounds and lamped for three kilowatts produces an index of 18,000 units, compared with an index of 7650 units registered by an equivalent grouping of lens, lamp, and reflector units. Roughly, this amounts to an increase in usable light per watt consumed of approximately 240 per cent. The distribution of these two test fixtures is best demonstrated by referring to polar coördinate curves projected on an area of approximately 200 square-feet from a height of eight feet. In Figure 5 the 300-foot-candle intensity curve for the "single six" is indicated by the solid line; that of the competitive fixture is shown dotted. Areas within these limits serve to indicate relative efficiencies, as the wattage, arrangement, and length of throw were held constant in obtaining the data. Figure 6, with the solid line again indicating the "single six," gives a general idea of the photometric distribution of the beam about the center line.

The mechanical arrangement for flexibility consists of a universal clamp for attaching the supporting arm to a gridiron, with rotational freedom possible at the fixture itself. A single adjusting screw allows

Fig. 7—The double-three unit.

the operator to set the bank for any desired angle or direction of throw with the framework arranged either horizontally or in a vertical position relative to the studio floor.

The first of the standardized installations consisted of eighteen of these "single-six" units mounted on the gridiron in such a manner that they could quickly and easily be brought into play on any acting area selected by the production group. As a space-conserving measure a few of these long units were reassembled in two rows of three (Figure 7), designated as "double threes." In certain sets where the light-concentration was high and space at a minimum, this arrangement was found to be more satisfactory from an operational standpoint. This type of construction was later mounted on portable stands for use as floor broads.

The "single three" (Figure 8), one-half of the "single six," was next brought into use for reinforcing light, background flooding, and as a general-purpose strip-light of minimum dimensions.

By standardizing the construction of our unit assembly we were assured of uniform spectral characteristics and distribution from each fixture rather than a spotty heterogeneous mixture of several types of light requiring careful blending on the set. A common standard of light-producing unit also allowed us to familiarize ourselves with

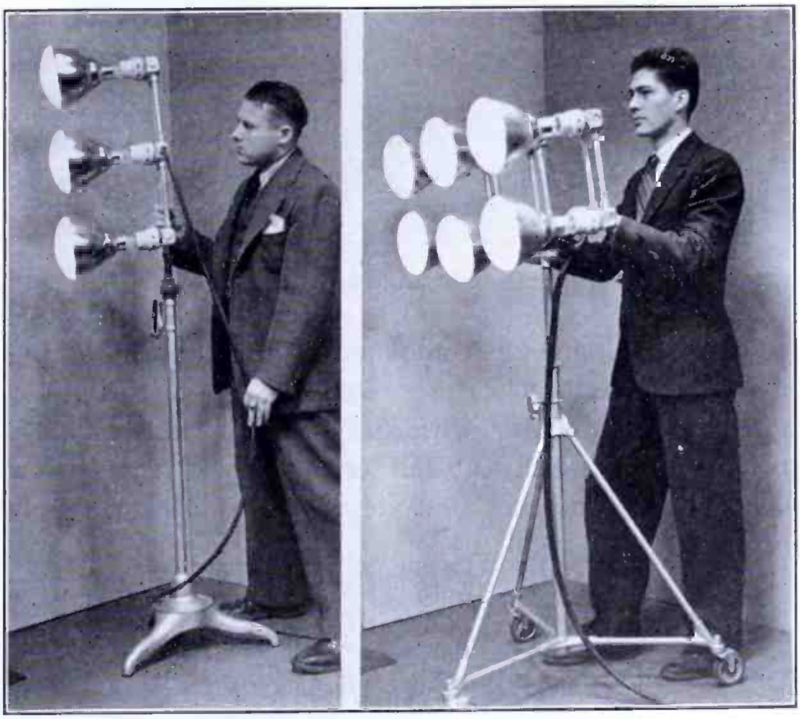


Fig. 8 *(Left)*—The single-three unit.
Fig. 9 *(Right)*—The floor broad.

the operation of the fixture and, by simple addition or subtraction, to meet the studio's quantitative light problems.

Shortly after completing the foundation-light installation we turned to the more complex problem of supplying the characteristic, or modeling, light from the floor. Here again, several problems confronted us, resulting in a partial redesign of the standardized mounting.

The floor broad (Figure 9) is identical with the overhead array except that it is mounted on a portable floor stand. Two of these units are used normally as reinforcing lights from stage right and left to create a rough modeling angle or to temper the shadows on the backdrops. In all cases, however, it was required that the operation of these lights should give floor precedence to camera movement. They are, therefore, brought into play and taken out frequently during the course of a single sequence. The diffused characteristic of this light permits such an unorthodox procedure to be satisfactorily carried out without leaving an apparent hole in the set illumination.

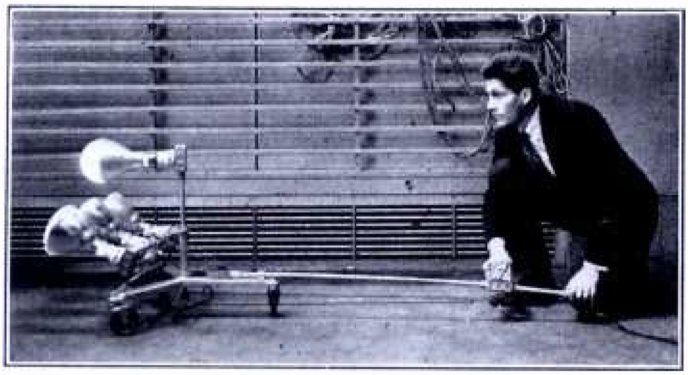


Fig. 10—Portable foot light.

Our modeling equipment is completed by the addition of two other units, the portable foot light (Figure 10) and the hand light (Figure 11). This floor light, working with and ahead of the close-up camera, is maneuvered to highlight the subject properly from this camera angle. Such a technic decrees that the intimate close-ups which pro-

Fig. 11—The hand light.

duce the best delineation of halftone value shall benefit by the best lighting. It is impossible, of course, to light each shot of each camera from the optimum angle in a studio where we find the duration of pick-up from a single camera sometimes a matter of seconds. We have, therefore, made it a practice to work toward the camera that best displays our wares, after making sure that the foundation lighting over the set is so arranged as to supply satisfactory illumination for the other cameras.

The hand light (Figure 11) is used to reinforce floor light in such sequences where a single camera shot can be safely modeled to the contrast limit. It is normally used on the wide-angle close-up camera and can be fitted with either a spotlight for contrast highlights or a diffusing lamp for the more subtle modeling.

We do not attempt to approach the contrasts common on the stage and in motion pictures. In television we are confronted with a highly compressed contrast range that permits modeling, to be sure, but also holds as a penalty for exaggeration a wash-out or a complete black. It is therefore necessary that we work well within these limits, since the review and criticism of our lighting technic is by the audience in the field rather than by a cutting-room jury. This, however, has not restricted the use of modeling light; the trend, on the other hand, being toward the greater contrast that the electrical system will accept, in preference to the flat non-dimensional pictures of past years. Experience gained by operation and observation appears to be the only rule in the use of these modeling fixtures even though we have endeavored to take guesswork out of the equipment.

Our failure to mention back-lighting does not mean that we have overlooked the possibilities of this type of illumination. In the studio sets we have yet to arrive at a reasonable system of back-lighting that will answer all the requirements of flexibility, weight, and operation. It is true that we now are using, in our main studios, an advanced type of remotely controlled ceiling light that appears to solve the problem, but since our findings to date are not conclusive, we felt that discussion of this system should be held for the future.

We make use of one other type of light that merits consideration. This equipment is known as the "portrait table," used as the name implies: in cases where the picture is primarily a portrait. Four lights are arranged at the outer rim of the announcing desk on flexible goosenecks adjustable as to height, angle, and throw. By substitution of various types of bulbs and variations of the wattage, detailed modeling of the face can be effected with a minimum of difficulty. This equipment also has portable back-lighting, which again is controllable, making the work shot of this table the television equivalent of a studio portrait.

This enumeration completes the catalog of our lighting technic and equipment in the National Broadcasting Company television studios. We have tested all reasonable systems of light production and are still carrying on these investigations. Lately we have been interested in vapor-lamps as a possible adjunct to the system, but the complications inherent in a three-phase power-supply and a watercooling system would appear to make further consideration of present models impracticable.

There have been many statements and many more conjectures as to the light used in television studios. We quote pertinent figures based on our last six-month period of operation. Our average set illumination was in the neighborhood of 1200 foot-candles of incident light. Our average modeling ratio was 2 to 1, while the average light load was slightly more than 50 kw of 110-volt d-c. Our lowest foundation lighting level was 800 foot-candles, a play in which the contrast throughout the set was carried to the upper limit of the Iconoscope. The highest foot-candle reading recorded was slightly less than 2500 foot-candles, a continuity where, obviously, little modeling was attempted.

In our work of the past three years, we feel that we have established a substantial foundation in television studio lighting on which we hope to base an even simpler system. If we appear to have standardized certain assemblies and particular light-sources, this does not mean that our development work has ceased. It continues with renewed vigor as we see our experiments bearing fruit.

SELECTIVE SIDE-BAND TRANSMISSION IN TELEVISION

By

R. D. Kell and G. L. Fredendall

RCA Manufacturing Company, Inc., Camden, N. J.

Summary—Reproduction of television detail in a selective side band system is treated as a function of the modulation factor and the ratio of vestigial side band to transmitted side band. A comparison with double side-band transmission is included.

THE fundamental limitation placed upon the amount of detail which may be obtained ultimately in a television picture is the width of the radio-frequency spectrum allotted to a television channel. Six megacycles has been adopted as a standard width by the Radio Manufacturers' Association (R.M.A.). After provision has been made for the sound channel and guard bands and account taken of practical circuit considerations in receiver design there remains a spectrum about 5.25 megacycles wide for the transmission of picture signals.

A problem of first importance is a determination of the position of the television carrier in a spectrum of fixed width and the amplitude and phase characteristics over the spectrum that lead to the transmission of the greatest amount of detail. The following discussion is a mathematical analysis of the problem based upon certain reasonable simplifications.

The amount of picture detail refers to the fidelity of reproduction at the receiver of abrupt changes in intensity of half-tones in the picture at the transmitter. Figure 1 illustrates the typical abrupt changes which may occur in the direction of scanning (horizontal detail). The transmission of similar changes which occur at right angles to the direction of scanning (vertical detail) does not involve the transmission characteristics of the system and thus need not be considered. In (a) and (b) the single abrupt change in intensity is assumed to be isolated to the extent that the corresponding signal is not perceptibly influenced by preceding or following detail. Such detail occurs at the junctions between relatively large areas having different half-tone values.

The pulses in (c) and (d) have a width of the order of a scanning line and correspond to an isolated narrow line perpendicular to the direction of scanning.

Two pulses not necessarily of the same height, but separated by a distance comparable to the width of the pulse are shown in (e) and (f). These correspond to two closely spaced vertical lines in the picture.

Since all types in Figure 1 are fundamental in the building of detail in a television picture, no type can be safely excluded from a study of television transmission.

Previous Studies of Selective Side-band Transmission

Almost from the beginning of the transmission and reception of television images it was found that a better picture could be obtained with the receiver tuned so that the carrier was located on one side of the selectivity curve.

Poch and Epstein[1] have demonstrated by laboratory measurement the improvement in the reception of detail (e) resulting from moving

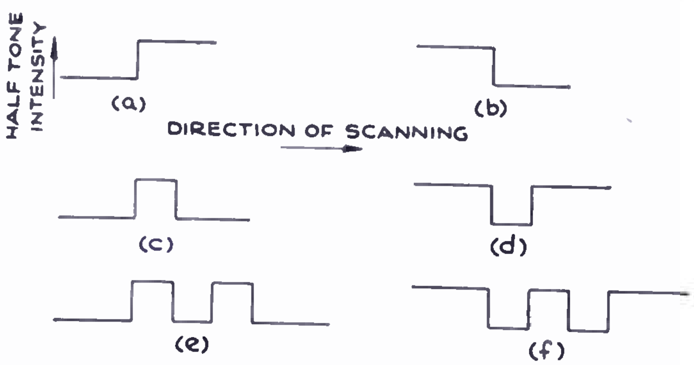


Fig. 1—Television detail.

the carrier to one side of the selectivity curve of a receiver. They gave a mathematical confirmation in the nature of a steady-state analysis of the phase and amplitude characteristics of the video signal corresponding to a low percentage of modulation of the carrier by a single video frequency. In addition to these steady-state conditions, it is important to know the response to the wave forms as shown in Figure 1.*

Goldman[2] has presented a mathematical analysis of the transmission of the details in Figure 1 by a selective side-band system. In his analysis the carrier was varied over a channel of one specific transmission characteristic.

Preliminary Considerations

If the details shown in Figure 1 could be scanned by a pick-up device having an aperture of infinitesimal dimension in the direction of scanning, the video signal generated would have the same wave-

* A paper, "Effect of the Quadrature Component in Single Side-Band Transmission," by H. Nyquist and K. W. Pfleger has been published in *B.S.T.J.*, January, 1940, since this manuscript was accepted by the publisher.

form as the transitions in half-tone. The effect of a finite symmetrical aperture of a practical scanning device may be obtained by imagining that the signal generated by the infinitesimal aperture is passed through an electrical filter having no phase distortion, but an amplitude distortion characteristic of the particular finite aperture. It has been found possible in practice to compensate electrically for the amplitude distortion at least up to the highest video frequency which conceivably could be accommodated in the proposed television channels. Hence, it shall be assumed in the following treatment that suitable correction has been made and that the transmitting aperture is not a controlling factor in shaping the transmitted signal.

The video amplifiers at the transmitter are regarded as distortionless up to an abrupt cut-off frequency f_0 (less than the highest frequency of aperture compensation) and, therefore, become the controlling factor in modifying the waveform of a picture signal before modulation of the radio-frequency carrier. A linear detector responding to the envelope of the intermediate-frequency signal is commonly employed in television receivers and is assumed here.

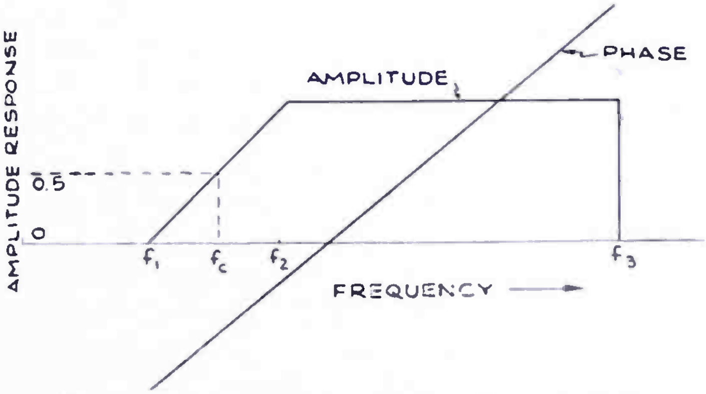


Fig. 2—Idealized transmission characteristic $(f_0 - f_c) < (f_1 + f_c)$.

If a receiving aperture of infinitesimal width in the direction of scanning were possible, the variation in light intensity along a scanning line would have the same wave shape as the intermediate-frequency envelope. In practice the receiving aperture is finite, but the amplitude distortion thus introduced may be compensated electrically to a degree that justifies the assumption of an infinitesimal aperture.

With the above suppositions, the envelope of the signal at the input to the detector becomes a direct criterion of the fidelity of transmission.

Position of the Carrier on the Transmission Characteristic

The overall transmission characteristic of the system properly includes the characteristics of the radio-frequency circuits at the transmitter and the radio- and intermediate-frequency circuits of the receiver. Figure 2 shows idealized overall characteristics which (although not physically compatible) presumably could be approximated in an actual system. The amplitude characteristic shows partial suppression of one side band; the phase shift is a linear function of the frequency. The latter assumption is desirable because thereby the best transmission associated with a given amplitude characteristic will be obtained.

If the steady-state amplitude and phase characteristics are derived by determining the envelope of the system when modulated with various video frequencies, one at a time, it is found that the position of the carrier frequency exerts a large effect on the amplitude characteristic. If the carrier is near f_1 in Figure 2, the high-frequency portion is accentuated; if near f_2, the low-frequency portion is accentuated; and if f_c is half way between f_1 and f_2, the frequency response is flat. Harmonics of the modulation frequency are always generated when one side-band is partially suppressed; hence, the conventional frequency response of the system may be misleading unless properly qualified. If the percentage of modulation is sufficiently low, the magnitudes of harmonics are negligible.

In the analysis of selective side-band transmission the carrier f_c shall be fixed at the point of 50 per cent response which gives a flat frequency response. Frequent comparisons will be made between double and selective side-band transmissions.

Reproduction of a Unit Function Detail

Figure 3 shows a carrier wave modulated by a signal which is the response of the video frequency amplifiers at the transmitter to a unit function detail. As a consequence of the finite cut-off frequency of the video amplifier, the response (Figure 3b) is not a unit function, but has the same form as that of a low-pass filter to a unit function. The solution for the envelope of the modulated carrier at the output (at the receiver) of the idealized selective side-band system of Figure 2 is derived in Appendix 1 and summarized in Equations (8) and (9).

A quantity $(f_3 - f_c)$ $(t - \tau)$ is the independent variable in which

t is the time,

τ is the time delay of the envelope equal to the slope of the linear phase shift curve,

$(f_3 - f_c)$ is the video band width.

The envelope is found to be a function of two parameters, the ratio $(f_c - f_1)/(f_3 - f_c)$ and a modulation factor m. The modulation factor is determined by the relative amplitudes of the carrier before and after the scanning of an abrupt edge. This is illustrated in Figure 3c. Thus m is equal to unity when one of the carrier levels is zero.

The square root of the sum of the squares of the in-phase component P plus a constant and the quadrature component Q define the shape of the envelope (Equation 9). As the modulation factor approaches zero, the magnitude of the in-phase component becomes large compared with the quadrature component and the envelope approaches the case of double side-band transmission. In double side-band transmission the quadrature component is zero and the in-phase component has the same form as in Equation (8). Thus, the effect of the partial suppression of one side band is to introduce distortion in the form of the quadrature component.

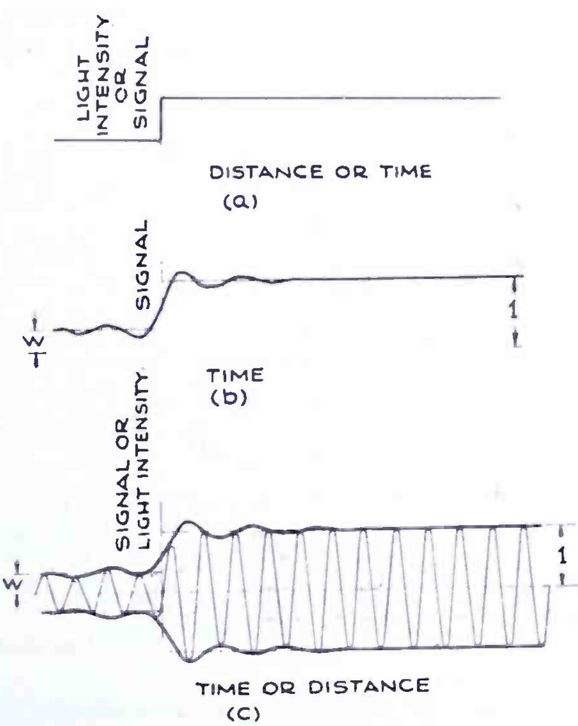


Fig. 3—(a) Unit function detail.
(b) Response of idealized amplifier to (a).
(c) Response of selective side-band system to (b).

$$\text{Modulation factor } m = \frac{1 - W}{1 + W}.$$

In Figure 4 a family of envelopes have been plotted according to Equation (9) in which the partially suppressed side band $(f_c - f_1)$ is the parameter and m is equal to unity. A fixed band width of 5.25 megacycles is used corresponding to the standards of the R.M.A. A more explicit independent variable $(t - \tau)$ is used in place of the generalized form $(f_3 - f_c)\ (t - \tau)$.

Under the conditions laid down initially, the variation of intensity along the scanning line has the same wave shape as the envelope of the electrical response. Hence, the axis of ordinates may be regarded as the intensity and the axis of abscissas as the distance along the scanning line.

A distance equal to one scanning line pitch corresponds to 0.12 microsecond (R.M.A. Standards).

Several observations may be made when $m = 1$.

(1) The steepness of rise for different values of $(f_c - f_1)$ do not differ significantly in the interval -0.15 to 0 microseconds, the range of greatest variation in the response.

(2) The amount of "transient" overswing of each envelope increases as $(f_c - f_1)$ decreases. Under the R.M.A. standard that white correspond to zero carrier, the overswing and damped oscillation would appear as striations of alternate light and dark bands superimposed

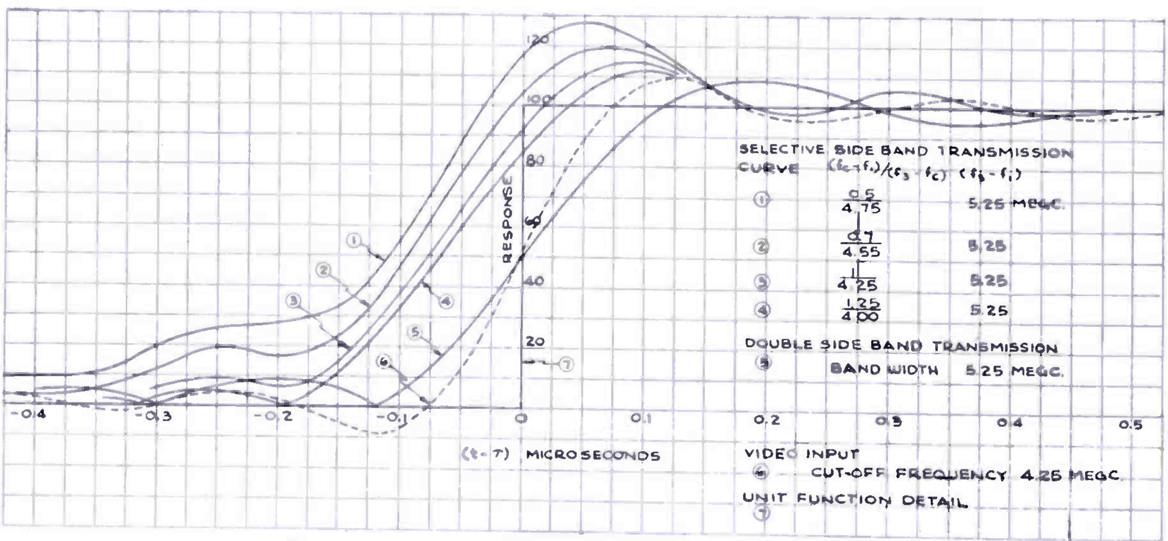


Fig. 4—Transmission of unit function detail.

in the gray region of the white to gray transition. The first striation is wider than a picture element and would, therefore, be visible at the correct viewing distance from the screen. Succeeding striations are of the order of a picture element and, therefore, may not be distinguishable.

(3) The principle rise that largely identifies the location of the transition in the received picture is preceded by an anticipatory step which is more pronounced as $(f_c - f_1)$ is decreased. This step gives the visual impression of a blurred transition. In this respect systems represented by curves (3) or (4) are definitely superior.

(4) The fidelity with which a unit function is transmitted through a system characterized by a particular value of $(f_c - f_1)$ may be judged by comparing the envelope of the response with the corresponding video signal which supplies the modulation. It is recalled that the video signal at the transmitter is the output of an idealized

amplifier having an abrupt cut-off frequency equal to $f_o = (f_3 - f_c)$. Thus, envelope (3) drawn for $(f_c - f_1) = 1$ megacycle must be compared with curve (6), the video response to a unit function detail of an idealized amplifier for which $(f_3 - f_c) = 4.25$ megacycles.

The ratio of the average slope (in the region of principal rise) of curve (3) to that of curve (6) is about 1.6. Thus, an abrupt transition between half-tones appears less abrupt when received in the selective side-band system (curve 3) than when applied as a modulating signal at the transmitter (curve 6).

(5) Curve (5), the envelope of the response of a double side-band system 5.25 megacycles wide, has an average steepness comparable with those of the envelopes (3) and (4) for selective side-band trans-

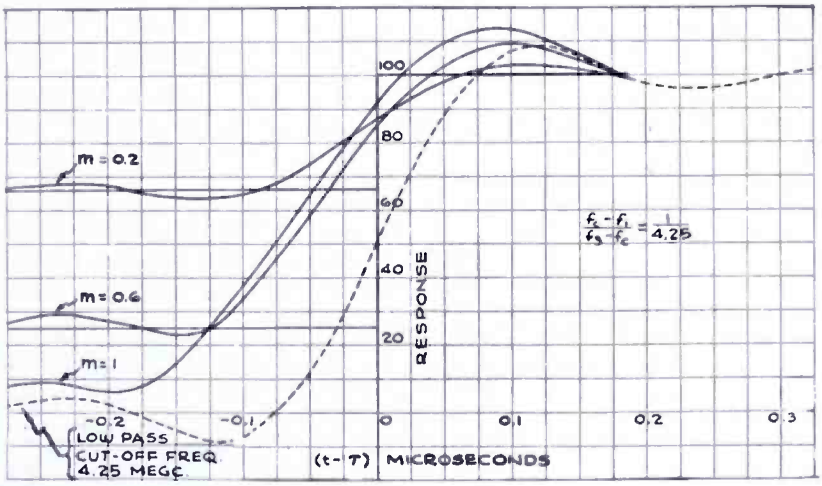


Fig. 5—Transmission of unit function detail as function of modulation factor m.

mission. This means that for a fixed band width there is no superiority of a selective side-band system over a double side-band system for the transmission of a unit function detail at percentages of modulation near 100 (m near unity).

In a video signal as in an audio signal the average percentage modulation is low. Many of the abrupt transitions in a television subject take place between two half-tones neither of which is white, that is, the value of the modulation factor m is not unity. Figure 5 is a family of envelopes drawn for $m = 0.2$, 0.6, and 1. $(f_c - f_1)$ is taken equal to 1 megacycle and the band width is 5.25 megacycles as in Figure 4. It is observed that the envelopes properly scaled (Figure 6) approach the video signal as m approaches zero. That is, the received signal resembles more and more the modulating video signal (cut-off frequency $= (f_3 - f_c)$ and in this sense becomes distortionless in the limit. This signifies that the fidelity of the idealized selective side-band system approaches that of a double side-band system having a band width of 8.5 megacycles for small percentages of modulation and that of a double side-band system 5.25 megacycles wide for percentages of modulation near 100.

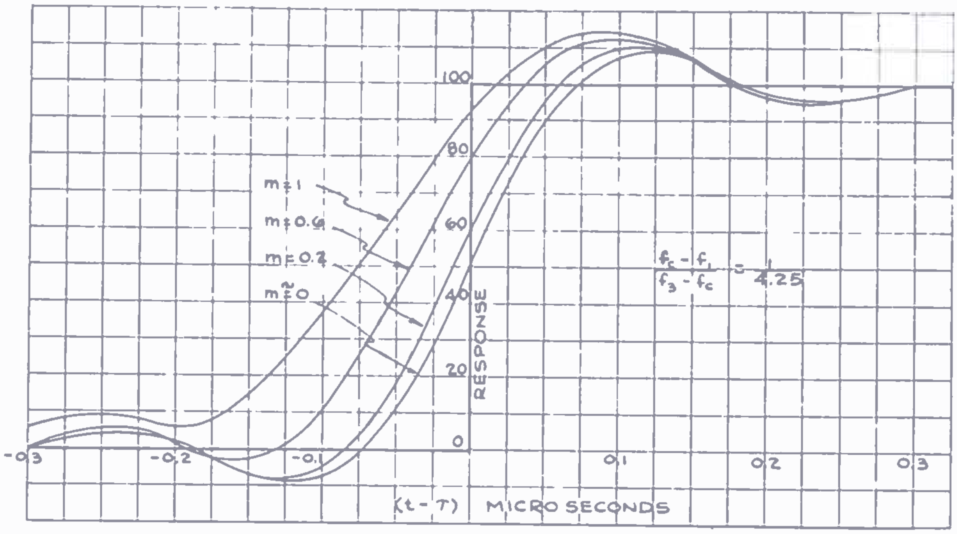


Fig. 6—Transmission of unit function detail as function of modulation factor m.

Reproduction of Narrow Line Detail

A narrow line perpendicular to the direction of scanning is ideally represented by the square pulse in Figure 7(1). When the signals generated by the scanning device are limited in the amplifiers to a band width of 4.25 megacycles, the video waveforms for a pulse 0.15 microsecond long are shown in Figure 7, curve (2). This pulse corresponds to a line having a width approximately equal to a picture element in the present television system. Except for the negative loops which should be reflected in the time axis curve (2) is also the envelope of a carrier modulated by the pulse with a modulation factor equal to one and transmitted double side band with a band width of 8.5 mega-

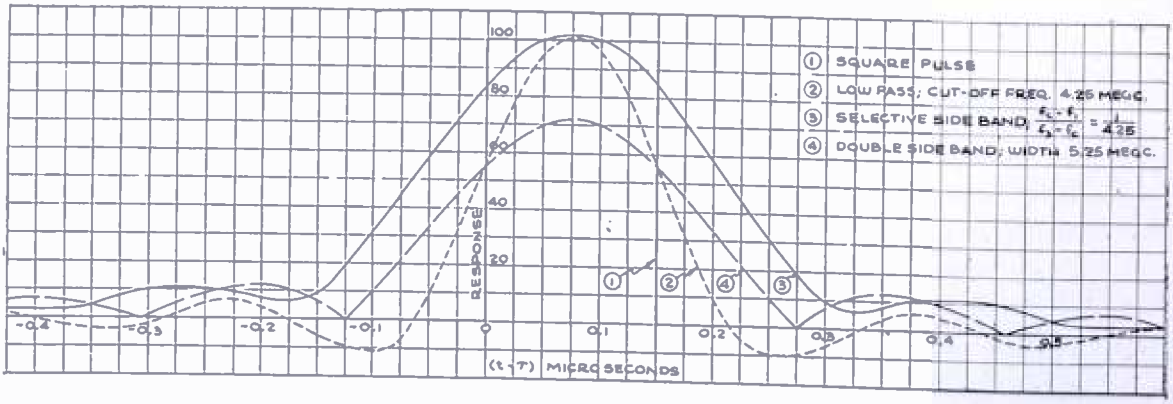


Fig. 7—Transmission of narrow line detail.

cycles. Curve (3) indicates that the maximum amplitude of the response of a selective side-band system 5.25 megacycles wide is the same as that of a double side-band system 8.5 megacycles wide. Curve (4) is the response when a band width of 5.25 megacycles is used for double side-band transmission. Comparing the two modes of transmission on the same band width it is observed that a narrow line is reproduced at only 70 per cent of its proper intensity in the double side-band case.

The apparent width is also a significant characteristic of lines of the order of a picture element wide. There is an apparent elongation of the pulse after transmission through a selective side-band system.

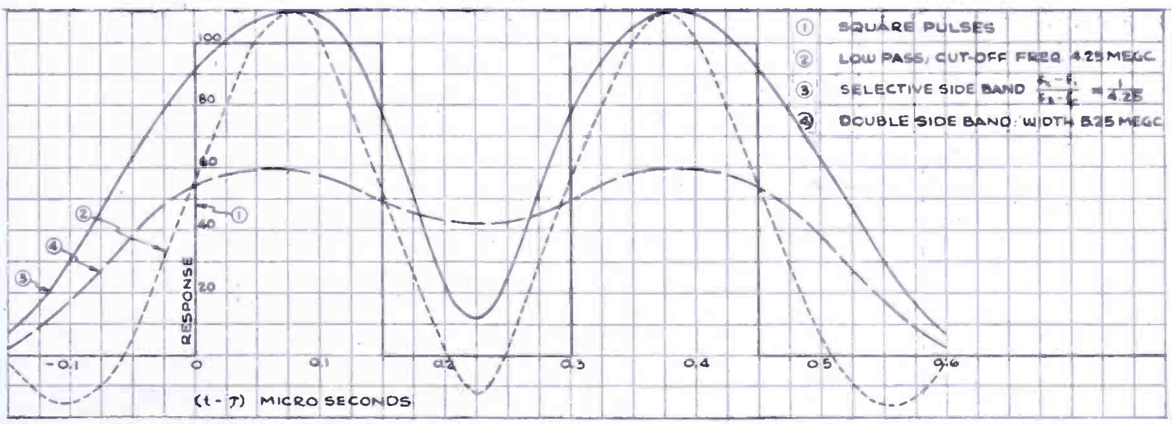


Fig. 8—Transmission of two closely spaced narrow lines.

As the modulation factor is reduced there is less and less distortion, and in the limit the transmission is the same as through a double side-band system 8.5 megacycles wide.

Reproduction of Closely Spaced Narrow Lines

The term "resolution" is applied most frequently to the property of a system to distinguish between closely spaced narrow lines as represented in Figure 1e. Mathematical expressions have been developed in Appendix 3 that afford a revealing comparison of the fidelity of selective and double side-band systems for the reproduction of such fine line detail. A series of envelopes corresponding to two square pulses 0.15 microsecond long and separated by an interval of the same length are shown in Figure 8. It is observed that the video signal, curve (2), provides resolution of the two lines by sinking to a sustained low value between lines. Curve (3) represents the response of a selective side-band system and does contain an interval of low signal value, but curve (4) for double side-band transmission over a band of equal width does not indicate appreciable resolution. This result is certainly not predicted in Curve (3), Figure 7, which illustrated the case of a single narrow line. The sloping edge of the reproduced single pulse itself exceeded the separation of pulses in Figure 8. An examination of the mathematical expression for the envelope (Equation 11) affords the explanation of the apparent contradiction. The quadrature or distorting components due to the first and second pulses partially cancel near the center of the separating interval and exactly cancel at the center.

Economy in Band Width

The economy in band width obtained by means of partial suppression of one side band depends therefore upon the range of the modulation factor. Thus, curves may be drawn for each type of television detail showing the relation between the modulation factor and the band width required for equal fidelity of reproduction of horizontal and vertical detail. This would involve a method of measuring the sharpness of transition between halftones in the unit function detail, the width of a narrow line, and the resolution of two narrow lines. The subject of the measures of vertical and horizontal resolutions in a television picture is variously treated by writers and it is not the purpose of this discussion to enter into the merits of different methods. The permissible range of the modulation factor in a television picture will represent a compromise between fidelity, band width, and intensity of the transmitted signal.

Concluding Remarks

The intent of this analysis has been to demonstrate the characteristic differences of double and selective side-band transmissions of television detail and in particular to find a favorable transmission characteristic for the latter. The analysis was developed on the hypothesis of a linear phase characteristic throughout the pass band in order to obtain the optimum envelope associated with a given amplitude characteristic. Thus, the envelopes derived here cannot be duplicated point for point in a physically possible system in which some phase distortion always resides, but the broad aspects of the idealized treatment may be realized in practice when careful phase compensation has been provided. The preceding work may be summarized as follows:

(1) If the modulation factor is near 1, a unit function detail is transmitted most faithfully in a selective side-band system when the ratio of the partially suppressed side band to the completely transmitted side band $(f_c - f_1)/(f_3 - f_c)$ has a value lying between 1/4.25 and 1.25/4. Comparable fidelity is obtained in a double side-band system of equal band width that requires a video band width appreciably less than in the selective side-band example. As the modulation factor

becomes less than 1, the fidelity of the selective side-band system for the transmission of the unit function increases, whereas that for double side-band transmission (equal band width) does not increase. In the limit as the modulation factor approaches zero, the sharpness of reproduction in the selective side-band system is about 1.6 times greater.

(2) When the modulation factor is equal to 1 the width of the input video signal corresponding to a single narrow line is increased about equally after transmission through either system operated over equal band widths, but there is a reduction in amplitude of the envelope in a double side-band system. The extension in width approaches zero as the factor is made progressively smaller in the selective side-band system, but there is no change in the other system.

(3) Two narrow lines are resolved more completely by selective side-band transmission for any value of the modulation factor than by double side-band transmission over an equal band width. As the factor becomes less than 1, the remarks above also apply for the resolution of narrow lines.

Appendix

I. Response of a Selective Side-band System to a Unit Function Detail

Figure 3b may be regarded as the limit of a square wave $E(t)$ as the fundamental frequency approaches zero and the upper limit of the frequency spectrum is held constant.

$$E(t) = \frac{1}{1+m}\left[1 + \frac{4m}{\pi}\sum_{1}^{N}\frac{\sin(2n-1)}{(2n-1)}\omega t\right] \tag{1}$$

$$(2N-1)\omega = \omega_o$$

A sine-wave carrier modulated by $E(t)$ has the form

$$e_1(t) = \frac{1}{1+m}\left\{\sin\omega_c t + \frac{2m}{\pi}\sum_{1}^{N}\cos\left[\omega_c - (2n-1)\omega\right]t\right.$$

$$\left.\frac{-2m}{\pi}\sum_{1}^{N}\cos\left[\omega_c + (2n-1)\omega\right]t\right\}.$$

If $e_1(t)$ is impressed on a linear system that alters the amplitude and phase, there results

$$e(t) = \frac{1}{1+m}\left\{A_c\sin(\omega_c t + \theta_c) + \frac{2m}{\pi}\sum_{1}^{N}\frac{A_{(2n-1)}}{2n-1}\cos\left\{\left[\omega_c - (2n-1)\omega\right]t + \theta(2n-1)\right\}\right. \tag{2}$$

$$\frac{-2m}{\pi}\sum_{1}^{N}\frac{B_{(2n-1)}}{2n-1}\cos\left\{[\omega_c+(2n-1)\omega]t+\beta(2n-1)\right\}.$$

If the phase shift is linear then

$$\theta_{(2n-1)} = \tau[\omega_c - (2n-1)\omega] + b$$

$$\beta_{(2n-1)} = -\tau[\omega_c + (2n-1)\omega] + b$$

$$T = (t-\tau).$$

(2) becomes

$$\begin{aligned} e(t) = {} & \frac{1}{1+m} A_c \sin(\omega_c T + b) \\ & + \frac{2m}{\pi}\cos(\omega_c T + b)\sum_{1}^{N}\frac{A_{(2n-1)}}{2n-1}\cos(2n-1)\omega T \\ & + \frac{2m}{\pi}\sin(\omega_c T + b)\sum_{1}^{N}\frac{A_{(2n-1)}}{2n-1}\sin(2n-1)\omega T \\ & - \frac{2m}{\pi}\cos(\omega_c T + b)\sum_{1}^{N}\frac{B_{(2n-1)}}{2n-1}\cos(2n-1)\omega T \\ & + \frac{2m}{\pi}\sin(\omega_c T + b)\sum_{1}^{N}\frac{B_{(2n-1)}}{2n-1}\sin(2n-1)\omega T]. \end{aligned} \quad (3)$$

(3) has the form

$$\left(P + \frac{A_c}{1+m}\right)\sin(\omega_c T + b) + Q\cos(\omega_c T + b) = \sqrt{\left(P + \frac{A_c}{1+m}\right)^2 + Q^2}.$$

$$\cos[\omega_c T + b + \varepsilon] \quad (4)$$

where

$$P = \frac{1}{1+m}\ \frac{2m}{\pi}\sum_{1}^{N}\left\{\frac{A_{(2n-1)}}{2n-1} + \frac{B_{(2n-1)}}{2n-1}\right\}\sin(2n-1)\omega T \quad (5)$$

$$Q = \frac{1}{1+m}\,\frac{2m}{\pi}\sum_{1}^{N}\left\{\frac{A_{(2n-1)}}{2n-1} - \frac{B_{(2n-1)}}{2n-1}\right\}\cos(2n-1)\omega T.$$

The envelope of the modulated carrier is the coefficient

$$\sqrt{\left(P + \frac{A_c}{1+m}\right)^2 + Q^2}.$$

$A_{(2n-1)}$ and $B_{(2n-1)}$ may be assigned values in accordance with the amplitude characteristic of Figure 2.

$$\frac{A_{(2n-1)}}{2n-1} + \frac{B_{(2n-1)}}{2n-1} = \frac{1}{2n-1} \text{ over the characteristic.}$$

P becomes

$$\frac{1}{1+m}\,\frac{2m}{\pi}\sum_{1}^{N}\frac{\sin(2n-1)\omega T}{2n-1}.$$

$$\frac{A_{(2n-1)}}{2n-1} - \frac{B_{(2n-1)}}{2n-1} = -\frac{\omega}{\omega_c - \omega_1}$$ on the sloping part of the characteristic and

$$\frac{A_{(2n-1)}}{2n-1} - \frac{B_{(2n-1)}}{2n-1} = -\frac{1}{2n-1}$$ on the straight part of the characteristic.

Q becomes

$$\frac{1}{1+m}\left[-\frac{2m}{\pi}\,\frac{\omega}{(\omega_c-\omega_1)}\sum_{1}^{p}\cos(2n-1)\omega T - \frac{2m}{\pi}\sum_{p+1}^{N}\frac{1}{2n-1}\cos(2n-1)\omega T\right] \tag{6}$$

where $(2p-1)\omega = (\omega_c - \omega_1)$; $(2N-1)\omega = \omega_o = (\omega_3 - \omega_c)$.

The first sum in (6) may be simplified by using the following proposition[4]

$$\sum_{1}^{\frac{K+2}{2}}\cos(2n-1)\theta = \frac{\frac{1}{2}\sin(K+2)\theta}{\sin\theta}.$$

There results

$$\sum_{1}^{p}\cos(2n-1)\omega T = \frac{\frac{1}{2}\sin 2p\omega T}{\sin\omega T} = \frac{\frac{1}{2}\sin(\omega_c-\omega_1+\omega)T}{\sin\omega T}.$$

If $\omega \to o$ there results

$$P = \frac{m}{1+m} \frac{1}{\pi} \int_o^{(\omega_3 - \omega_c)T} \frac{\sin x}{x} dx$$

$$Q = \frac{m}{1+m} \frac{1}{\pi} \left[- \frac{\sin (\omega_c - \omega_1) T}{(\omega_c - \omega_1) T} - \int_{(\omega_c - \omega_1) T}^{(\omega_3 - \omega_c) T} \frac{\cos x}{x} dx \right. \tag{7}$$

$$= \frac{m}{1+m} \frac{1}{\pi} \left[- \frac{\sin (\omega_c - \omega_1) T}{\omega_c - \omega_1} - \int_{(\omega_c - \omega_1) T}^{\infty} \frac{\cos x}{x} dx + \int_{(\omega_3 - \omega_c) T}^{\infty} \frac{\cos x}{x} dx. \right.$$

If a change of independent variable is made in (7)

$$(\omega_3 - \omega_c)\ T = \eta$$

and if

$$\frac{\omega_c - \omega_1}{\omega_3 - \omega_c} = \delta$$

more general forms for P and Q are (8)

$$P = \frac{m}{1+m} \frac{1}{\pi} \int_o^{\eta} \frac{\sin x}{x} dx$$

$$Q = \frac{m}{1+m} \frac{1}{\pi} \left[- \frac{\sin \eta\delta}{\eta\delta} - \int_{\delta\eta}^{\infty} \frac{\cos x}{x} dx + \int_{\eta}^{\infty} \frac{\cos x}{x} dx \right].$$

δ and m are parameters.

These integrals have been tabulated extensively[5].

According to (4) the envelope is the coefficient

$$\sqrt{\left\{ P + \frac{1}{2} (1+m) \right\}^2 + Q^2}. \tag{9}$$

Part 2. Response of a Selective Side-band System to a Square Pulse

The equation of a square pulse T_1 seconds long is obtained by adding a unit function having an amplitude $\left(- \frac{2m}{1+m} \right)$ and delayed T_1 seconds to $E(t)$, Equation (1). The solution for the corresponding envelope follows in a manner similar to the development in Part 1. The result is

$$\text{envelope} = \sqrt{\rho^2 + \partial^2}$$

$$\text{where } \rho = \frac{1}{2(1+m)} - \frac{m}{2(1+m)} + P(T) - P(T - T_1) \qquad (10)$$

$$\partial = Q(T) - Q(T - T_1).$$

The P and Q functions are defined by (7).

Part 3. Response of a Selective Side-band System to Two Square Pulses

Two pulses illustrated in Figure 1c are formed by adding unit functions of the following descriptions to $E(t)$:

$$\text{amplitude } \frac{-2m}{1+m}\text{; delayed } T_1 \text{ seconds}$$

$$\text{amplitude } \frac{2m}{1+m}\text{; delayed } T_2 \text{ seconds}$$

$$\text{amplitude } \frac{-2m}{1+m}\text{; delayed } T_3 \text{ seconds.}$$

The envelope of the response is

$$\sqrt{\rho^2 + \partial^2}$$

where

$$\rho = \frac{1}{2(1+m)} - \frac{m}{2(1+m)} + P(T) - P(T - T_1) + P(T - T_2) - P(T - T_3) \qquad (11)$$

$$\partial = Q(T) - Q(T - T_1) + Q(T - T_2) - Q(T - T_3).$$

Block-shaped signals of any description may be expressed by suitably combining functions of the unit function type. The envelopes will be given by P and Q functions defined in (7).

Part 4. Response of Low-Pass Systems to Television Detail

a. Unit Function (Figure 1a)

$$e(t) = \frac{1}{2} + \frac{1}{\pi} \int_0^{2\pi f_0 T} \frac{\sin x}{x}\, dx. \qquad (12)$$

b. Square Pulse (Figure 1c)

$$e(t) = \frac{1}{\pi}\left[\int_{o}^{2\pi f_o T} \frac{\sin x}{x}\,dx - \int_{o}^{2\pi f_o (T-T_1)} \frac{\sin x}{x}\,dx\right]. \quad (13)$$

c. Two Square Pulses (Figure 1e)

$$e(t) = \frac{1}{\pi}\left[\int_{o}^{2\pi f_o T} \frac{\sin x}{x}\,dx - \int_{o}^{2\pi f_o (T-T_1)} \frac{\sin x}{x}\,dx + \right. \quad (14)$$

$$\left. \int_{o}^{2\pi f_o (T-T_2)} \frac{\sin x}{x}\,dx - \int_{o}^{2\pi f_o (T-T_3)} \frac{\sin x}{x}\,dx\right]$$

V. Response of Double Side-band Systems to Television Detail

Same as in Part 4 if $f_o = \dfrac{\text{band width}}{2}$.

List of References

[1] Poch and Epstein, "Partial Suppression of One Side Band in Television Reception," RCA Review, Vol. I, p. 19, 1937.

[2] Goldman, "Television Detail and Selective Side-band Transmission," I. R. E. Fall Convention (1938), Rochester, New York.

[3] Poch and Epstein, *loc. cit.*

[4] Chrystal's Algebra II, p. 273.

[5] Jahnke-Emde, "Tables of Functions," second revised edition, p. 78. B. G. Teubner, Leipsig and Berlin, (1933).

FLUCTUATIONS IN SPACE-CHARGE-LIMITED CURRENTS AT MODERATELY HIGH FREQUENCIES

BY

B. J. THOMPSON, D. O. NORTH AND W. A. HARRIS
RCA Manufacturing Company, Inc., Harrison, N. J.

PART II—DIODES AND NEGATIVE-GRID TRIODES

BY DWIGHT O. NORTH

Summary—*A quantitative theory of shot effect in the parallel-plane diode is formulated for any degree of space charge, and for frequencies such that transit-time effects are of no concern. Beginning with the steady-state description of a diode showing a Maxwell-Boltzmann distribution of emission velocities, the theory is founded upon a determination of the new steady state which results from the injection of a small additional emission comprised of electrons of a specified velocity.*

The fluctuations in a diode current I are expressed by

$$\overline{i^2} = \Gamma^2 \cdot 2eI\Delta f.$$

For a temperature-limited diode it has long been known that $\Gamma^2 = 1$. *For a diode with anode potential sufficiently negative (retarding field),* $\Gamma^2 = 1$ *also. For the usual space-charged-limited case, i.e., any instance in which there is a virtual cathode,* Γ^2 *is less than unity, but in this paper* Γ^2 *is computed by numerical integration for only those instances in which the anode current is a small fraction of the emission.*

In this last case the shot-effect formula is also written, phenomenologically, to correspond with Nyquist's well-known expression for thermal agitation in a passive network of conductance g, thus:

$$\overline{i^2} = \theta \cdot 4kTg\Delta f.$$

Here T is absolute cathode temperature and g is diode conductance. The dimensionless factor θ *is found to be virtually a constant* $\left(\theta \approx \frac{2}{3}\right)$ *for the whole range of normal anode potentials with asymptotic value,*

$$\theta \sim 3\left(1 - \frac{\pi}{4}\right)$$

in the limit of high anode potential. That is, the mean-square noise generated by emission fluctuations is roughly numerically equal to two-thirds of the noise of thermal agitation generated by a resistance of magnitude equal to the a-c resistance of the diode and possessing a temperature equal to the cathode temperature.

The theory is extended to cover shot effect in the anode circuit of a triode with negative grid. Unless the amplification factor μ *is very low, the same formulas apply, except that g is now interpreted as the con-*

ductance of the "equivalent diode". Following recent practice, the anode-circuit shot effect is exressed in terms of an ohmic resistance R_{eq} at room temperature T_0 in the grid circuit, the thermal agitation of which produces an equal fluctuation in the anode circuit, thus:

$$R_{eq} = \frac{\theta}{\sigma} \cdot \frac{T}{T_0} \cdot \frac{1}{g_m}.$$

where g_m is the transconductance and $\sigma = \frac{g_m}{g}$. For modern tubes, σ lies between 0.5 and unity.

The diode in thermal equilibrium (diode of any geometry, all at one temperature, no source of energy) is analyzed to show that the fluctuations can be quantitatively reckoned as either shot effect or thermal agitation with the same numerical results. It is argued that here only can these distinct concepts be merged into one.

Diode measurements show noise consistently higher than the theoretical predictions, the factor of disagreement amounting to an order of magnitude at high currents. This is qualitatively accounted for in a satisfactory manner in terms of a small amount of elastic reflection of electrons at the anode, an hyopthesis supported by published experimental work.

Negative-grid triodes are usually not subject to such complications; this is evidenced by good agreement between the triode formulas and measurements of a variety of modern tubes, including not only the usual quasi-cylindrical structures, but even one flat structure with a V-type filament. The theory is, therefore, believed to be extensively valuable as a guide to low-noise amplifier design.

THEORY

INTRODUCTION AND REVIEW OF STEADY STATE

IN PART I it was shown that an adequate analysis of space-charge-reduced shot effect must recognize that the reduction associated with fluctuations in emission of electrons having a specified emission velocity is, *inter alia*, a function of that velocity. This investigation will, therefore, be confined initially to ascertaining the net fluctuations in current which occur as a consequence of true shot fluctuations in the emission of electrons of a *specified velocity*. The total reduced shot effect, measurable in a suitably connected circuit, will then be established by summing over the whole gamut of initial velocities.

It can be seen immediately that the character of the results will be determined largely by the Maxwell-Boltzmann (M-B) laws of distribution. At the same time it becomes apparent that the theory will have to be founded upon a *steady-state* description which embodies these laws. As often happens, the processes of analysis are impeded least by confining ourselves to a parallel-plane model, for the desired description of the steady state is well known.[1,2,3,4] Whether or not the results obtained can be applied to other types of structure is a question relegated to subsequent experimental studies. The treatment and notation of Fry and of Langmuir will be employed here inasmuch as these references are likely to be most accessible. To conserve space and to avoid unreasonable repetition of these published works, it will be assumed henceforth that the reader is acquainted with the general nature of these papers, particularly the work of Fry.

The steady-state potential within a space-charge-limited parallel-plane diode, the cathode of which produces an unlimited supply of electrons with *zero* emission velocity, is represented by the well-known expression,

$$E = \left(9\pi \sqrt{\frac{m}{2e}} I \right)^{2/3} x^{4/3}, \tag{1}$$

in which E is the potential at a distance x from the cathode, I is the cathode current per unit area, m is the mass of an electron, and e its charge (throughout, this symbol implies a positive quantity). The potential of such a hypothetical case is illustrated by curve A of Figure 1.

In actuality there is a finite emission current I_s per unit area possessing a M-B distribution described by the usual exponential law. If we let v_s be the normal component of an electron's initial velocity, and if we define two useful equivalent potentials, thus:

$$\frac{1}{2} m v_s^2 = e V_s, \quad \frac{kT}{e} = V_e, \tag{2}$$

one aspect of the M-B distribution can be written

$$dI_s = \frac{I_s}{V_e} \varepsilon^{-\frac{V_s}{V_e}} dV_s. \tag{3}$$

In (2), k is Boltzmann's constant ($= 1.37 \times 10^{-16}$ ergs/°K), and T is the absolute cathode temperature. Equation (3) defines the portion

[1] P. S. Epstein, "Theory of Space-Charge Effects," *Verh. d. Deut. Phys. Gesell*, Vol. 21, p. 85, (1919).

[2] T. C. Fry, "The Thermionic Current Between Parallel Plane Electrodes: Velocities of Emission Distributed According to Maxwell's Law," *Phys. Rev.*, Vol. 17, p. 441, (1921).

[3] I. Langmuir, "The Effect of Space Charge and Initial Velocities on the Potential Distribution and Thermionic Current Between Parallel Plane Electrodes," *Phys. Rev.* Vol. 21, p. 419, (1923).

[4] I. Langmuir and K. T. Compton, "Electrical Discharges in Gases," *Rev. Mod. Phys.*, Vol. 3, p. 191, (1931).

dI_s, of the total emission current I_s, composed of electrons whose initial normal velocities, expressed in terms of the equivalent potential V_s, lie within the narrow limits V_s and $V_s + dV_s$. The actual potential function, therefore, has the appearance of curve B, the potential minimum E_m adjacent to the cathode serving as a gate which permits electrons of higher velocity to proceed to the plate. These constitute the cathode (or anode) current I. The rest of I_s, consisting of electrons whose emission velocities are such that $\frac{1}{2} m v_s^2 < |e E_m|$, is returned

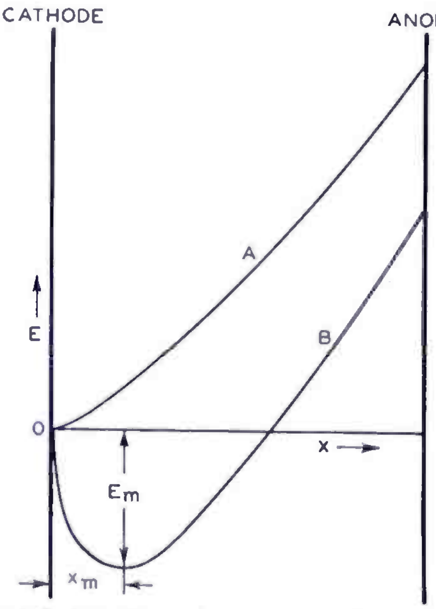


Fig. 1—Potential functions in a space-charged-limited diode.

to the cathode without having crossed the potential minimum or "virtual" cathode.

The potential function in this case is best described by replacing E and x by quantities η and ξ, respectively, having the following definitions,

$$\eta = \frac{E - E_m}{V_e} \tag{4a}$$

$$\xi = 4 \left(\frac{\pi}{2kT} \right)^{3/4} m^{1/4} e^{1/2} I^{1/2} (x - x_m). \tag{4b}$$

It should be noted that the origin of coordinates is thereby shifted to the potential minimum. In the α-space, to the left of the virtual cathode, ξ is negative; in the β-space, to the right of the virtual cathode, ξ is positive; while η is positive in both spaces. The steady-state potential distribution then appears, in Fry's notation, as the solution of the cryptic differential equation,

$$\left(\frac{d\eta}{d\xi}\right)^2 = \phi\ (\eta), \tag{5}$$

$\phi\ (\eta)$ being shorthand for either of two expressions,

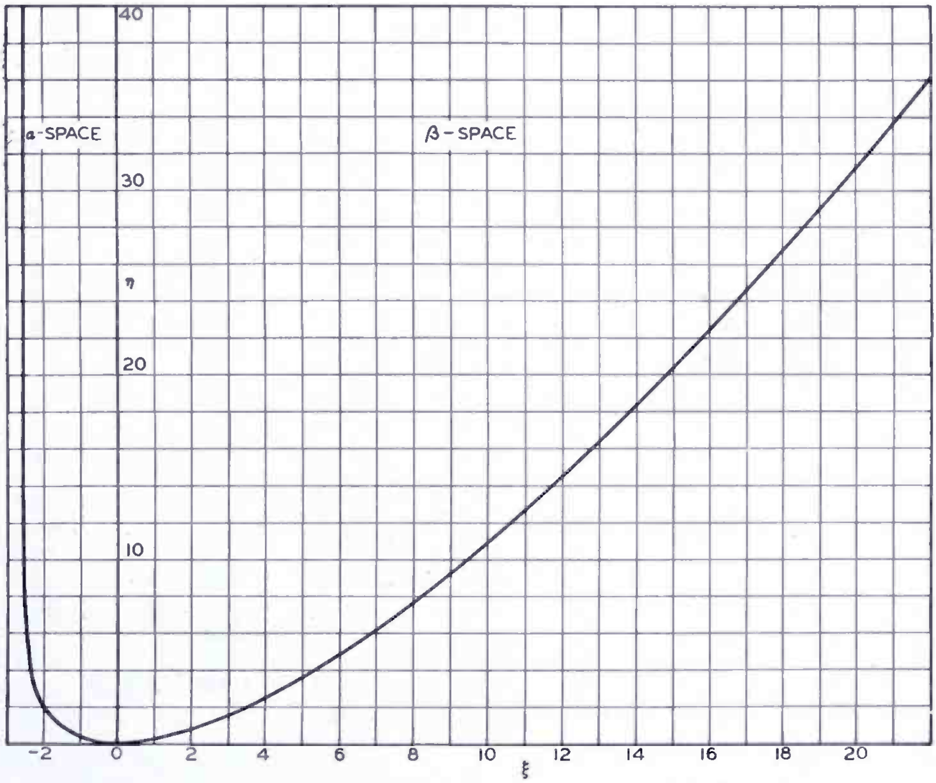


Fig. 2—Solutions of the differential equation: $\left(\frac{d\eta}{d\xi}\right)^2 = \phi(\eta)$.

$$\phi\ (\eta) = \varepsilon^{\eta} - 1 \mp \left(\frac{2}{\sqrt{\pi}}\,\eta^{1/2} - \varepsilon^{\eta}\,\mathrm{erf}\,\eta^{1/2}\right). \tag{6}$$

The upper sign is used for the α-space, the lower sign for the β-space, and the error function is defined:

$$\mathrm{erf}\ y = \frac{2}{\sqrt{\pi}} \int_o^y e^{-x^2}\,dx \tag{7}$$

The solution of this equation has already been effected by mechanical means, with the assistance of series approximations. It is plotted in Figure 2 which, in terms of η and ξ, embodies all plots of E against x, such as curve B, Figure 1. The solution is tabulated in Table I, taken intact from Langmuir's paper.[1] In Table II there are listed the values of ϕ_α for values of $\eta < 10$, and of ϕ_β for values of $\eta < 100$. These were calculated directly from (6) with the help of tables of the error function, except for values of ϕ_β for $\eta > 1$ which were computed from the asymptotic series. The series expansions are very valuable in checking the tables, and, in addition, give much information as to the behavior of the functions.

TABLE I

η	$-\xi_\alpha$	ξ_β	η	$-\xi_\alpha$	ξ_β	η	$-\xi_\alpha$	ξ_β
0.00	0.0000	0.0000	3.0	2.2338	4.4750	20	2.5538	14.8260
0.05	0.4281	0.4657	3.2	2.2650	4.6524	25	2.5539	17.1931
0.10	0.5941	0.6693	3.4	2.2930	4.8261	30		19.4253
0.15	0.7167	0.8296	3.6	2.3183	4.9963	35		21.5522
0.20	0.8170	0.9674	3.8	2.3410	5.1634	40		23.5939
0.25	0.9028	1.0909	4.0	2.3615	5.3274	45		25.5643
0.30	0.9785	1.2042	4.5	2.4044	5.7259	50		27.4740
0.35	1.0464	1.3098	5.0	2.4376	6.1098	60		31.141
0.40	1.1081	1.4092	5.5	2.4634	6.4811	70		34.642
0.45	1.1648	1.5035	6.0	2.4834	6.8416	80		38.007
0.50	1.2173	1.5936	6.5	2.4990	7.1924	90		41.258
0.6	1.3120	1.7636	7.0	2.5112	7.5345	100		44.412
0.7	1.3956	1.9224	7.5	2.5206	7.8690	150		59.086
0.8	1.4704	2.0725	8.0	2.5280	8.1963	200		72.479
0.9	1.5380	2.2154	9.0	2.5382	8.8323	300		96.877
1.0	1.5996	2.3522	10	2.5444	9.4465	400		119.185
1.1	1.6561	2.4839	11	2.5481	10.0417	500		140.068
1.2	1.7081	2.6110	12	2.5504	10.6204	600		159.885
1.4	1.8009	2.8539	13	2.5518	11.1845	700		178.861
1.6	1.8813	3.0842	14	2.5526	11.7355	800		197.146
1.8	1.9515	3.3040	15	2.5531	12.2747	900		214.850
2.0	2.0134	3.5151	16	2.5534	12.8032	1000	2.5539	232.054
2.2	2.0681	3.7187	18	2.5537	13.8313			
2.4	2.1168	3.9158						
2.6	2.1602	4.1071						
2.8	2.1990	4.2934						

When $\eta << 1$,

$$\phi_\beta{}^\alpha(\eta) = \left[\eta + \frac{1}{2}\eta^2 + \frac{1}{6}\eta^3 + \cdots\right] \pm \frac{2}{\sqrt{\pi}}\eta^{1/2}\left[\frac{2}{3}\eta + \frac{4}{15}\eta^2 + \cdots\right]$$

When $\eta >> 1$,

$$\phi_\alpha(\eta) \sim 2\varepsilon^\eta - \frac{2}{\sqrt{\pi}}\eta^{1/2} - 1 - \frac{1}{\sqrt{\pi}\,\eta^{1/2}}\left[1 - \frac{1}{2\eta} + \frac{3}{4\eta^2} - \frac{15}{8\eta^3} + \cdots\right]$$

$$\phi_\beta(\eta) \sim \frac{2}{\sqrt{\pi}}\eta^{1/2}\left[1 + \frac{1}{2\eta} - \frac{1}{4\eta^2} + \frac{3}{8\eta^3} - \frac{15}{16\eta^4} + \cdots\right] - 1 \qquad (8)$$

[1] Langmuir, loc. cit.

TABLE II

η	$\phi_\alpha(\eta)$	$\phi_\beta(\eta)$	η	$\phi_\beta(\eta)$
0.00	0.0000	0.0000	10	2.75
0.01	0.0109	0.0093	15	3.51
0.02	0.0224	0.0180	20	4.17
0.03	0.0345	0.0265	25	4.75
0.04	0.0469	0.0347	30	5.29
0.05	0.0599	0.0427	35	5.77
0.06	0.0731	0.0505	40	6.23
0.07	0.0867	0.0583	45	6.65
0.08	0.1009	0.0657	50	7.06
0.09	0.1153	0.0731	60	7.81
0.1	0.1300	0.0804	70	8.51
0.2	0.2944	0.1484	80	9.15
0.3	0.4897	0.2101	90	9.77
0.4	0.7164	0.2672	100	10.32
0.5	0.9764	0.3210		
0.6	1.2722	0.3720		
0.7	1.6068	0.4208		
0.8	1.9836	0.4674		
0.9	2.4067	0.5125		
1.	2.8806	0.5560		
2.	11.846	0.932		
3.	36.94	1.23		
4.	105.68	1.50		
5.	293.1	1.77		
6.	802.8	1.99		
7.	2,190	2.20		
8.	5,960	2.39		
9.	16,200	2.57		

It will be convenient to use Subscripts 1 and 2 to denote values of quantities such as ξ or η at the cathode and anode, respectively. This brief description of the steady state is then completed by observing that

$$I/I_s = \varepsilon^{-\eta_1}. \tag{9}$$

When I, I_s and T are specified, the operating voltage $(E_2 - E_1)$ can be found[1] by first evaluating η_1 from (9). The value of ξ_1 is then located in Table I. Knowledge of the distance $(x_2 - x_1)$ between electrodes permits ξ_2 to be derived from (4b). The value of η_2 is then located in Table I, and the desired $(E_2 - E_1)$ from (4a). The process of finding I, when the operating potential is specified, follows the same tactics; various values of I are assumed and the corresponding potentials determined; the I which corresponds to the specified $(E_2 - E_1)$ is then located by interpolation or read from a current-voltage plot.

[1] Langmuir and Compton, loc. cit.

ANALYSIS OF SHOT FLUCTUATIONS

Upon this steady-state foundation the succeeding analysis of fluctuations will be constructed. Inasmuch as all of the work entails numerical integration, one very important simplifying assumption will be made. It will be supposed that

$$I/I_s \rightarrow 0. \tag{10}$$

In effect, this means that ξ_1 will always be assigned the asymptotic value, -2.554, which is suggested by Figure 2 and tabulated in Table I. Although this may at first appear to be a most severe limitation, it is not, in fact. For example, reference to Table I shows that ξ_1 is only 16% below its asymptotic value when $\eta_1 = 3.0$, which is to say, in view of (9), when $I/I_s = \varepsilon^{-3} \approx 1/20$. One may expect to run across disagreement between theory and measurement on this account in case thoriated or pure-tungsten emitters are used. But, ordinarily, oxide-coated cathodes produce enough emission to place I/I_s far below the figure cited, so that, in such applications, the assignment of discrepancies to this limitation of theory will not be just. Nevertheless one should not lose sight of the assumption, for it is tantamount to what might be termed "complete space-charge limitation", and the final expressions to be evolved are valid only in this domain. If, for instance, the noise is measured while the cathode temperature (emission) is slowly dropped, one should expect the measured shot effect to depart eventually from the theoretical value given in this paper, and to move to higher levels through an uncharted region, finally arriving at a magnitude prescribed by Schottky's original formula for true temperature-limited shot effect simultaneously with the disappearance of the last vestige of a virtual cathode. The transition region is left unanalyzed in this paper, not because it is not understood, but simply because it requires much crank turning, yet promises little of practical interest.

The only further limitation of note is the confinement of discussion to frequencies low compared with the reciprocal of the electron transit time. Although there is every need for an understanding of shot fluctuations at ultra-high frequencies, it comes not on the wings of the morning, but only through a comprehension of the phenomena at frequencies unmolested by transit angles. This particular limitation simplifies the treatment tremendously. For, as pointed out in Part I, we shall consequently be concerned with fluctuations of a duration great enough that the analysis need only search for the description of the new *equilibrium state* into which the diode settles when, in addition to its original emission I_s, one admits a *small* steady increment $i_s(v_s)$ possessing a specified normal emission velocity v_s. Despite the fact that fluctuations in emission really do occur at random over the surface of the cathode, so that, to be exact, every emerging electron should be regarded as a "noise" electron, we shall nevertheless suppose i_s to be spread uniformly over the cathode area, even as I_s itself. That no precision is sacrificed through use of this artifice was proven in Part I, where the question was discussed in detail.

Hereafter, for brevity, we shall suppose the cathode potential to be zero ($E_1 = 0$). As before, we shall replace v_s by an equivalent potential V_s defined in the manner of (2),

$$\frac{1}{2} m v_s^2 = e\,V_s. \tag{11}$$

Let us further define a dimensionless quantity,

$$\lambda = \frac{V_s + E_m}{Ve}, \tag{12}$$

it being understood that E_m is negative. This quantity can be used in place of v_s to designate the emission velocity of electrons comprising i_s. If v_s be chosen so that i_s crosses the potential minimum to strike the anode ($V_s > -E_m$), λ is positive. For smaller choice of v_s such that the electrons proceed only part way to the potential minimum, stop, and return to the cathode, λ is negative; in fact, $-\eta_1 \leqslant \lambda \leqslant 0$, the lower limit corresponding to $v_s = 0$. In the η, ξ diagram, Figure 2, the turning point for these electrons will, of course, be a function of λ and will be that point for which $\eta = -\lambda$.

We are now prepared to construct a—for the moment—purely formal expression for the space-charge-reduced shot current fluctuations in the anode circuit. For every $i_s(\lambda)$ we shall find the new equilibrium current $\hat{I}$ (identified with a fixed emission I_s, and therefore *not* including $i_s(\lambda)$ itself) which flows in the presence of $i_s(\lambda)$. This permits a determination of the net increase in anode current over the steady-state I which flows in the absence of $i_s(\lambda)$. Both $\hat{I}$ and I are to be evaluated under the assumption that the electrode spacing and operating voltage are held fixed.* The ratio of this net increase in anode current to the quantity $i_s(\lambda)$ will, for small i_s, be a function of λ and not i_s. It will be denoted by $\gamma(\lambda)$ and is of value

* In Part I it was shown that the basic quantity to be determined is the fluctuation current. The voltage it produces across a connected circuit is only a simple problem in algebra involving both the external circuit and the internal impedance of the tube. Fluctuations of the latter, although certainly present, being initiated by fluctuations in space charge, represent a negligible contribution to the fluctuation voltage and will not be further discussed.

in that it represents the linear reduction factor by which a change in emission is converted into a change in plate current. Now confining our attention to the emission current dI_s composed of electrons whose normal emission velocities lie between λ and $\lambda + d\lambda$, we find from (3), (9), and (12) that its magnitude is

$$dI_s = I\,\varepsilon^{-\lambda}\,d\lambda. \tag{13}$$

This differential-sized emission current exhibits true shot fluctuations, expressible by the following well-known formula for the mean-square fluctuation current $\overline{i^2}$ in a band width Δf:

$$d(\overline{i^2})_s = 2e(dI_s)\triangle f, \tag{14}$$

or, in terms of (13):

$$d(\overline{i^2})_s = 2eI\triangle f(\varepsilon^{-\lambda}\,d\lambda). \tag{15}$$

But every variation in emission is, as we have said, *linearly* reduced at the anode by the factor $\gamma(\lambda)$. The resulting fluctuations in anode current to be associated with electrons of this velocity class are, therefore,

$$d(\overline{i^2}) = 2eI\triangle f\,[\gamma^2(\lambda)\,\varepsilon^{-\lambda}\,d\lambda]. \tag{16}$$

Since the fluctuations in emission of one velocity class are presumably independent of fluctuations in all other classes, the total fluctuation current in the anode circuit is the integral of (16) over λ:

$$\overline{i^2} = 2eI\triangle f\int_{-\eta_1}^{\infty}\gamma^2(\lambda)\,\varepsilon^{-\lambda}\,d\lambda. \tag{17}$$

The quantity $2eI\triangle f$ is recognized as the true shot effect to be expected from a temperature-limited current I. The mean-square fluctuations in the present space-charge-limited current I are, therefore, reduced below the true shot effect by the factor,

$$\Gamma^2 = \int_{-\eta_1}^{\infty}\gamma^2(\lambda)\,\varepsilon^{-\lambda}\,d\lambda. \tag{18}$$

This is the quantity of principal interest. The evaluation of the integral will be treated in two sections, thus:

$$\left.\begin{aligned}&\Gamma^2 = \Gamma_\alpha^2 + \Gamma_\beta^2\\ &\Gamma_\alpha^2 = \int_{-\eta_1}^{o}\gamma^2(\lambda)\,\varepsilon^{-\lambda}\,d\lambda, \quad \Gamma_\beta^2 = \int_{o}^{\infty}\gamma^2(\lambda)\,\varepsilon^{-\lambda}\,d\lambda.\end{aligned}\right\} \tag{19}$$

The subscripts α and β are appropriately used to signify the reduction factors associated with the two main groups of electrons, those whose emission velocities are insufficient to take them across the virtual cathode and out of the α-space, and those whose velocities permit them to enter the β-space and thus proceed to the anode. It should be observed that there is an intrinsic distinction to be made here, namely, fluctuations $i_s(\lambda > 0)$ contribute *directly* to the anode current, whereas fluctuations $i_s(\lambda < 0)$ do not.

More precisely,

$$\left.\begin{aligned} \gamma(\lambda) &= \frac{\hat{I} - I}{i_s(\lambda)}, -\eta_1 \leqslant \lambda \leqslant 0, \alpha\text{-group}, \\ \gamma(\lambda) &= 1 + \frac{\hat{I} - I}{i_s(\lambda)}, 0 \leqslant \lambda \leqslant \infty, \beta\text{-group}. \end{aligned}\right\} \tag{20}$$

Inasmuch as the great majority of α-group electrons are returned before coming very near to the potential minimum, we should expect to find $\Gamma_\alpha^2 << \Gamma_\beta^2$. This will be borne out in the succeeding analysis. Furthermore, for either group, one should expect $\lim\limits_{\lambda = 0} \gamma(\lambda) = -\infty$, since an electron which comes permanently to rest* in the virtual cathode produces a permanent decrease in plate current. And yet, for the β-group, $\lim\limits_{\lambda \to \infty} \gamma(\lambda) = 1$, since the influence of an electron in transit must be a monotonic decreasing function of its transit time. It follows that there must be a certain $\lambda > 0$ such that $\gamma(\lambda) = 0$. One should reasonably expect this λ to correspond to a velocity at the virtual cathode *roughly* equal to the mean velocity of the M-B current I at the same point. In other words, one should predict $\gamma(\lambda) = 0$, for $\lambda \sim \pi/4$. This also will be verified.

Evaluation of Γ_β^2

When, in addition to the M-B emission I_s, a current $i_s(\lambda > 0)$ is produced at the cathode, the anode current $\hat{I}$ which is derived from I_s can be represented as the solution of

$$\left(\frac{d\eta}{d\xi}\right)^2 = \frac{i_s}{\hat{I}} \cdot \frac{2}{\sqrt{\pi}} \left[\sqrt{\eta + \lambda} - \sqrt{\lambda} \right] + \phi(\eta) \tag{21}$$

* The nature of the potential function in the vicinity of the virtual cathode, (8), permits this view, for it shows that an electron just able to reach the potential minimum requires an infinite transit time.

or

$$\left(\frac{d\eta}{d\xi}\right)^2 = \frac{i_s}{\hat{I}} F(\eta, \lambda) + \phi(\eta) \tag{22}$$

where

$$F(\eta, \lambda) = \frac{2}{\sqrt{\pi}} \left[\sqrt{\eta + \lambda} - \sqrt{\lambda} \right]. \tag{23}$$

This expression is valid in both the α- and β-spaces, with the understanding that the appropriate form of $\phi(\eta)$ be employed in each case, cf. (6). The construction of (21) will not be entered into; the reader will find it simple enough if he avails himself of the methods employed by Fry in developing the basic equation (5). That (21) is correct can be *inferred* from the following useful identity:

$$\int_0^\infty F(\eta,\lambda)\, \varepsilon^{-\lambda}\, d\lambda = \phi_\beta(\eta). \tag{24}$$

This means, naturally, that if, instead of confining i_s to a specified velocity λ, we distribute its velocities so that at the virtual cathode i_s has the same velocity distribution as $\hat{I}$, (21) reduces to the same form as (5) in the β-space, which not only is absolutely necessary, but should be sufficiently convincing that (21) is properly constructed.

Now, since we always assume $i_s << \hat{I}$, and I, we shall continually neglect all but the lowest power of i_s/I or $i_s/\hat{I}$. The differential equation can, therefore, be written

$$d\xi = \pm \left[1 - \frac{1}{2} \frac{i_s}{I} \frac{F(\eta,\lambda)}{\phi(\eta)} \right] \frac{d\eta}{\phi(\eta)^{1/2}}, \tag{25}$$

where the proper sign is to be used for each space.

For the β-space:—

Integrate (25), $0 < \xi < \hat{\xi}_2$, $0 < \eta < \hat{\eta}_2$.

$$\hat{\xi}_2 = \int_0^{\hat{\eta}_2} \frac{d\eta}{\phi_\beta(\eta)^{1/2}} - \frac{1}{2} \frac{i_s}{I} \int_0^{\hat{\eta}_2} \frac{F(\eta,\lambda)}{\phi_\beta(\eta)^{3/2}}\, d\eta. \tag{26}$$

Solution of the unperturbed problem, (5), gives

$$\xi_2 = \int_o^{\eta_2} \frac{d\eta}{\phi_\beta(\eta)^{1/2}}. \tag{26a}$$

The differences $(\hat{\eta} - \eta)$ and $(\hat{\xi} - \xi)$ evaluated at end points are also to be looked upon as first-order infinitesimals. The difference between (26) and (26a) is, therefore,

$$\hat{\xi}_2 - \xi_2 = \frac{\hat{\eta}_2 - \eta_2}{\phi_\beta(\eta_2)^{1/2}} - \frac{1}{2}\frac{i_s}{I}\int_o^{\eta_2} \frac{F(\eta,\lambda)}{\phi_\beta(\eta)^{3/2}}\,d\eta. \tag{27}$$

For the α-space:—

A similar treatment shows

$$\hat{\xi}_1 - \xi_1 = -\frac{\hat{\eta}_1 - \eta_1}{\phi_\alpha(\eta_1)^{1/2}} + \frac{1}{2}\frac{i_s}{I}\int_o^{\eta_1} \frac{F(\eta,\lambda)}{\phi_\alpha(\eta)^{1/2}}\,d\eta. \tag{28}$$

We shall set up the difference between (27) and (28). First, however, we define

$$D \equiv \frac{1}{2}(\xi_2 - \xi_1) + \phi_\beta(\eta_2)^{-1/2} + \phi_\alpha(\eta_1)^{-1/2}. \tag{29}$$

Then observe that, for small variations, and with the aforementioned stipulation that the operating potential is invariant, there follows from the definitions, (4a) and (4b):

$$\left.\begin{aligned} \hat{\eta}_2 - \eta_2 = \hat{\eta}_1 - \eta_1 = \frac{E_m - \hat{E}_m}{V_e} = \log\frac{I}{\hat{I}} = -\frac{\hat{I} - I}{I} \\ (\hat{\xi}_2 - \hat{\xi}_1) - (\xi_2 - \xi_1) = (\xi_2 - \xi_1)\left(\sqrt{\frac{\hat{I}}{I}} - 1\right) = \frac{1}{2}(\xi_2 - \xi_1)\frac{\hat{I} - I}{I}. \end{aligned}\right\} \tag{30}$$

When (29) and (30) are used in taking the difference of (27) and (28), rearrangement of terms shows

$$\frac{\hat{I} - I}{i_s} = -\frac{1}{2D}\left[\int_o^{\eta_1} \frac{F(\eta,\lambda)}{\phi_\alpha(\eta)^{3/2}}\,d\eta + \int_o^{\eta_2} \frac{F(\eta,\lambda)}{\phi_\beta(\eta)^{3/2}}\,d\eta\right].$$

The negative sign confirms the view that every transit of an electron of this group effects a momentary drop in the *M-B* space current. With regard to (20), we have finally:

$$\gamma_{\lambda \geqslant 0}(\lambda) = 1 - \frac{1}{2D}\left[\int_o^{\eta_1} \frac{F(\eta,\lambda)}{\phi_\alpha(\eta)^{3/2}}\,d\eta + \int_o^{\eta_2} \frac{F(\eta,\lambda)}{\phi_\beta(\eta)^{3/2}}\,d\eta\right]. \quad (31)$$

From this point evaluation of $\gamma(\lambda)$ proceeds, perforce, numerically. And there is now introduced, for simplicity, the previously discussed assumption that the emission I_s is copious enough to warrant postulating

$$I/I_s = 0,$$

which reduces in (31) to the equivalent statements,

$$\eta_1 = \infty,\ \xi_1 = -2.554,\ \phi_\alpha(\eta_1) = \infty\,,$$

so that

$$D = \frac{1}{2}(\xi_2 + 2.554) + \phi_\beta(\eta_2)^{-1/2}\,, \quad (32)$$

a function of η_2 alone. Likewise $\gamma(\lambda)$ itself reduces to a function of η_2 and λ alone.

The numerical work was carried out with the help of a computing machine and Tables I and II, together with auxiliary tables constructed from these, e.g., $\phi_\alpha(\eta)^{3/2}$ and $\phi_\beta(\eta)^{3/2}$. The details are for the most part uninteresting and will be described only in brief. After a value of λ was chosen, $\gamma(\lambda)$ was found from (31) for a series of values of η_2, $5 \leqslant \eta_2 \leqslant 100$, by trapezoidal integration. This computation was repeated for a large number of selected λ's, $0.05 \leqslant \lambda \leqslant 5$. Both of the integrands in (31) become infinite at the origin; hence, in the range $0 \leqslant \eta \leqslant 0.01$, these integrals were evaluated directly, using the series expansions (8). The outcome is a numerical representation of γ as a function of two parameters, η_2 and λ, in the domain:

$$5 \leqslant \eta_2 \leqslant 100$$

$$0.05 \leqslant \lambda \leqslant 5.$$

The representation cannot be carried to $\lambda = 0$, for γ exhibits a logarithmic discontinuity at this point. It could have been carried to zero for η_2, but there is little point in it for when $\eta_2 = 0$, the virtual cathode has moved over to the anode and there is no "gate" action—hence no reduction in shot effect—so that we know *a priori* that $\Gamma^2 = \Gamma_\beta{}^2 = 1$. More will be said of this situation later. The range of η_2 was limited to values thought to be of practical interest. For example, V_e is about

0.1 volt for oxide-type cathodes, whereas the operating potential is typified by 3 volts, in which event η_2 is approximately 30.

This stage of the calculations, being intermediate, merits only momentary attention to the nature of the function γ. In Figure 3, curve B shows $\gamma(\lambda)$ when $\eta_2 = 30$. The unbroken line contains the calculated values; the dashed line shows the logical extrapolation in both directions. As predicted, the point for which $\gamma = 0$ lies near $\pi/4$. Plots of γ for other choices of η_2 show precisely the same qualitative character and need not be exhibited. It is now patent from Figure 3

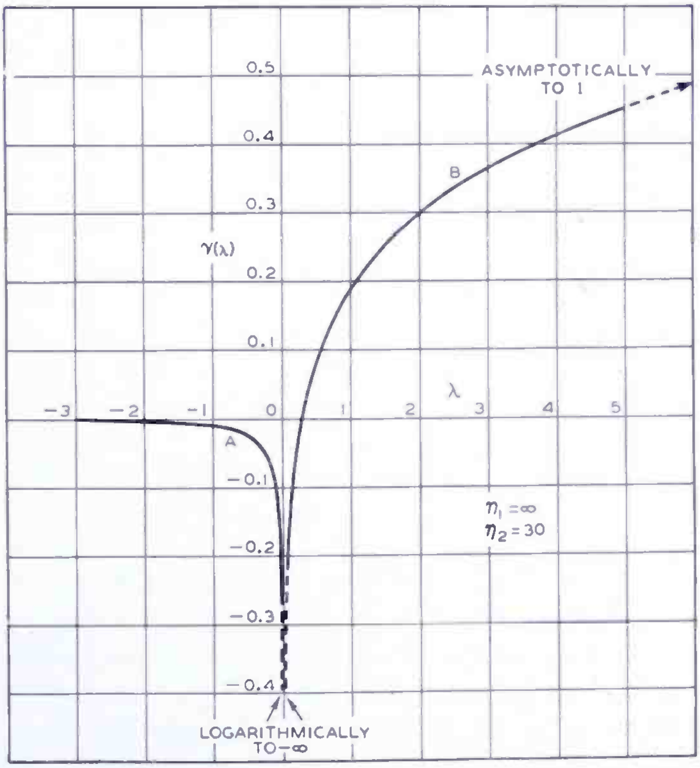


Fig. 3—Shot-effect reduction factor γ as a function of λ (velocity) for A—the α-group of electrons which fail to reach the virtual cathode. B—the β-group of electrons which traverse the virtual cathode.

that the reduction factor varies so widely with the emission velocity of the electrons that it is absolutely necessary to investigate each velocity class separately. The point is emphasized by the offensive behavior of fluctuations associated with electrons barely able to cross the virtual cathode (λ slightly > 0), these fluctuations proving to be so over-compensated ($\gamma < -1$, $\gamma^2 > 1$) that their *net fluctuations are in excess of the true or temperature-limited shot effect.*

This portion of the analysis is completed with a numerical integration of (19), giving Γ_β^2 as a function of η_2. For example, the value of Γ_β^2 when $\eta_2 = 30$ was found by squaring the ordinates of curve B, Figure 3, multiplying by $\epsilon^{-\lambda}$ and making a trapezoidal integration over the range $0.05 \leqslant \lambda \leqslant 5$, for which numerical values of $\gamma(\lambda)$ had been computed. The additional contributions, $0 \leqslant \lambda \leqslant 0.05$ and $5 \leqslant \lambda \leqslant \infty$, were calculated by direct integration from series approximations. The former can be handled in no other way because of the logarithmic discontinuity at the origin; the latter is a relatively insignificant contribution since $\varepsilon^{-\lambda}$ vanishes rapidly. The result is tabulated in Table III and plotted in Figure 4. The heavy line contains the calculated values, while the light line shows the logical extrapolation. The dashed

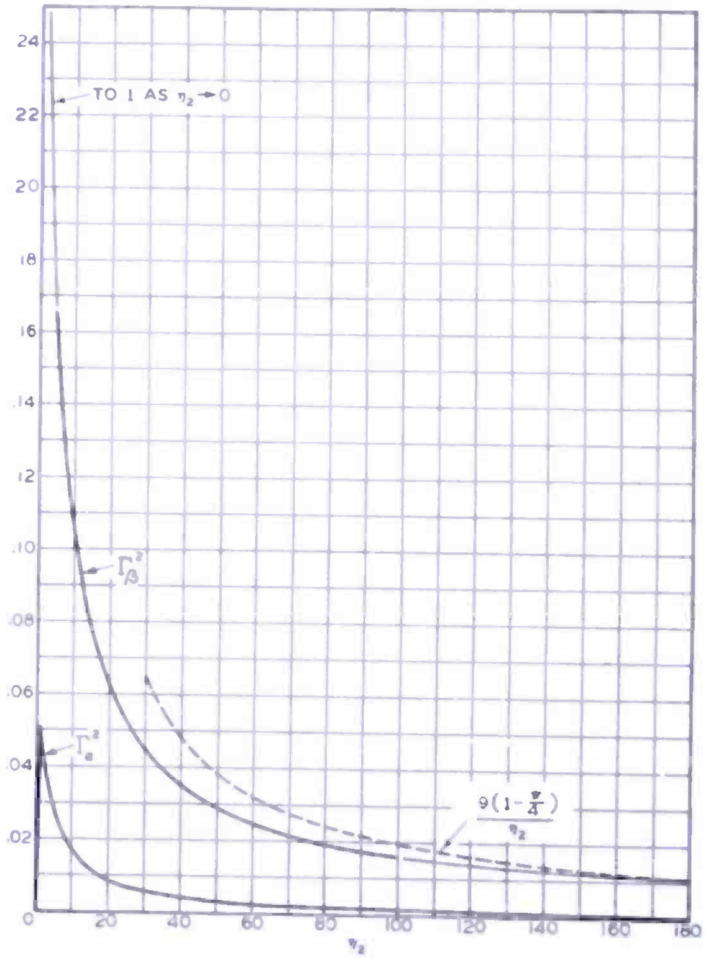


Fig. 4—Shot-effect reduction factors for the case, $I/I_s << 1$.

line shows the asymptotic form approached for large η_2. This is found from (31) through consideration of the largest η_2-terms only, and use of the asymptotic form of $\phi_\beta(\eta)$ as given in (8). In brief, for large η_2, the first indicated integral in (31) vanishes in comparison with the second, and $2D$ can be replaced by ξ_2. The second integral has already been shown to possess a logarithmic discontinuity for $\lambda = 0$. This can be overlooked, however, for the integral is otherwise an increasing function of η_2 so that, for any finite λ, choice of a sufficiently large η_2 permits that portion of the integral which depends on η_2 to overshadow the portion which depends on $\log\lambda$. Assuming finally that $\lambda << \eta_2$, one finds

$$\gamma \underset{\substack{\lambda > 0 \\ \eta_2 \to \infty}}{\sim} \frac{3}{\sqrt{\eta_2}} \left(\sqrt{\lambda} - \frac{1}{2} \sqrt{\pi} \right). \tag{33}$$

It is quite apparent that this form bears only a very slight resemblance to the curve of Figure 3 and approaches incorrect limits for both large and small λ. The difference in appearance means only that $\eta_2 = 30$ is too small to permit the use of (33). Of the incorrect lower limit enough has just been said to justify the omission from (33) of the additive term which contains the proper logarithmic discontinuity. That the upper limit is incorrect is only natural in view of the assumption $\lambda << \eta_2$. In spite of this, relation (33) may be used to find the asymptotic form of Γ_β^2 because of the rapid attenuation furnished by $\varepsilon^{-\lambda}$. The result of putting (33) into (19) is

$$\Gamma_\beta^2 \underset{\eta_2 \to \infty}{\sim} \frac{9}{\eta_2} \left(1 - \frac{\pi}{4} \right) \tag{34}$$

Evaluation of Γ_α^2

This study parallels the other so closely that its description can be made brief. Since the current i_s consists, in this case, of electrons which have normal emission velocities insufficient to carry them over the virtual cathode $(-\eta_1 < \lambda < 0)$, and since those electrons of a specified velocity λ turn about at the point $\eta = -\lambda$, it will be understandable that, in contrast with (21), there are now three distinct differential equations, the simultaneous solution of which determines the *M-B* anode current $\hat{\imath}$ derived from the *M-B* I_s when an additional $i_s(\lambda < 0)$ is produced at the cathode.

$$\left.\begin{aligned}
&\text{For the } \beta\text{-space:} && \left(\frac{d\eta}{d\xi}\right)^2 = \phi_\beta(\eta),\ 0 \leqslant \eta \leqslant \hat{\eta}_2 \\
&\text{For the } \alpha\text{-space:} && \left(\frac{d\eta}{d\xi}\right)^2 = \phi_\alpha(\eta),\ 0 \leqslant \eta \leqslant -\lambda \\
& && \left(\frac{d\eta}{d\xi}\right)^2 = \frac{i_s}{\hat{\imath}} G(\eta,\lambda) + \phi_\alpha(\eta),\ -\lambda \leqslant \eta \leqslant \hat{\eta}_1
\end{aligned}\right\} \tag{35}$$

where
$$G(\eta,\lambda)=\frac{4}{\sqrt{\pi}}\sqrt{\eta+\lambda}\,. \tag{36}$$

As before, the construction of these equations will not be described: it is simple, but lengthy. The stamp of credibility is affixed by noting

$$\int_{-\eta}^{o} G(\eta,\lambda)\,\varepsilon^{-\lambda}d\lambda=-2\left[\frac{2}{\sqrt{\pi}}\eta^{1/2}-\varepsilon^{\eta}\,erf\eta^{1/2}\right]=\phi_{\alpha}(\eta)-\phi_{\beta}(\eta), \tag{37}$$

which the reader can interpret, like (24), to mean that under the proper conditions (M-B distribution of i_s) the set (35) reverts to the form (5).

Repeating the process of taking the infinitesimal difference between the solution of (35) for $\bar{I}$ and the unperturbed solution of (5) for I, we arrive by familiar stages at the statement,

$$\underset{\eta\leqslant 0}{\gamma(\lambda)}=-\frac{1}{2D}\int_{-\lambda}^{\eta_1}\frac{G(\eta,\lambda)}{\phi_{\alpha}(\eta)^{3/2}}d\eta\,. \tag{38}$$

Again it is convenient to suppose the emission I_s very copious, so that D takes the form (32) and the upper limit of the integral above is made infinite. The integral then becomes a function of λ alone, and one numerical integration, therefore, suffices to permit representation of γ for all values of η_2, since η_2 influences the function only through the definition of D.

These integrations were performed for chosen values of λ in the range $0.01\leqslant-\lambda\leqslant 3$, observing the same precautions at discontinuities and methods similar to those outlined earlier. A sample plot for $\eta_2=30$ is shown in Figure 3, curve A.

Finally Γ_{α}^2 is determined from a numerical integration of (19) with the result:

$$\Gamma_{\alpha}^2=\frac{0.729}{D^2}\,. \tag{39}$$

This function has a marked maximum at $\eta_2\approx 0.8$, when $\Gamma_{\alpha}^2\approx 0.0513$, but vanishes rapidly, of course, as η_2 approaches zero. The values of Γ_{α}^2 versus η_2 are also listed in Table III and plotted in Figure 4. The expectation that $\Gamma_{\alpha}^2<<\Gamma_{\beta}^2$ is confirmed. For a practical instance, say $\eta_2=30$, $\Gamma_{\alpha}^2/\Gamma^2=0.11$; for larger η_2, the ratio diminishes slowly.

Composition and Interpretation

The measuring device cannot discriminate between Γ_α^2 and Γ_β^2. The important quantity, their sum Γ^2, is shown in Table III. And in Figure 5, Γ is plotted against η_2. This curve, in effect, describes the ratio of r-m-s current fluctuations appearing in the anode current of a space-charge-limited parallel-plane diode to the fluctuations in an equal temperature-limited current. Once more, the heavy line contains the calculated values, while the light line shows a reasonable extrapolation. The asymptotic curve for large η_2 is again included for comparison. The trend with both temperature and applied potential should be self-explanatory.

Rough agreement with experience is already noticeable, for most reports of noise measurements on modern amplifying tubes indicate an r-m-s reduction of 1/3 to 1/5. But Figure 5 is an inconvenient basis for comparison because the parameter η_2 is difficult to ascertain. It requires knowledge of the depth of the potential minimum, which, in turn, requires a determination of the ratio I/I_s. A plot of Γ against $(\eta_2 - \eta_1)$ is hardly an improvement because of serious errors that may arise when the applied voltage is not corrected for contact potential. A much more useful universal parameter is found as follows. Starting with

$$\eta_2 - \eta_1 = \frac{E_2}{V_e}$$

$$\xi_2 - \xi_1 \propto I^{1/2}$$

$$\eta_1 = \log \frac{I_s}{I},$$

and defining the diode conductance by $g \equiv \dfrac{\partial I}{\partial E_2}$ (i.e., the reciprocal of the a-c diode resistance), one can demonstrate that, under the assumption of invariance of I_s and T when I is altered by varying E_2,

$$\frac{I}{gV_e} = D\phi_\beta(\eta_2)^{1/2}. \tag{40}$$

The quantity on the left will become our new parameter, replacing η_2 and being defined by (40) which is principally a function of η_2 only, and becomes *entirely* independent of η_1 if it is again supposed that $I/I_s \to 0$. This assumption was made in plotting $I/g\,V_e$ against η_2 in Figure 5.

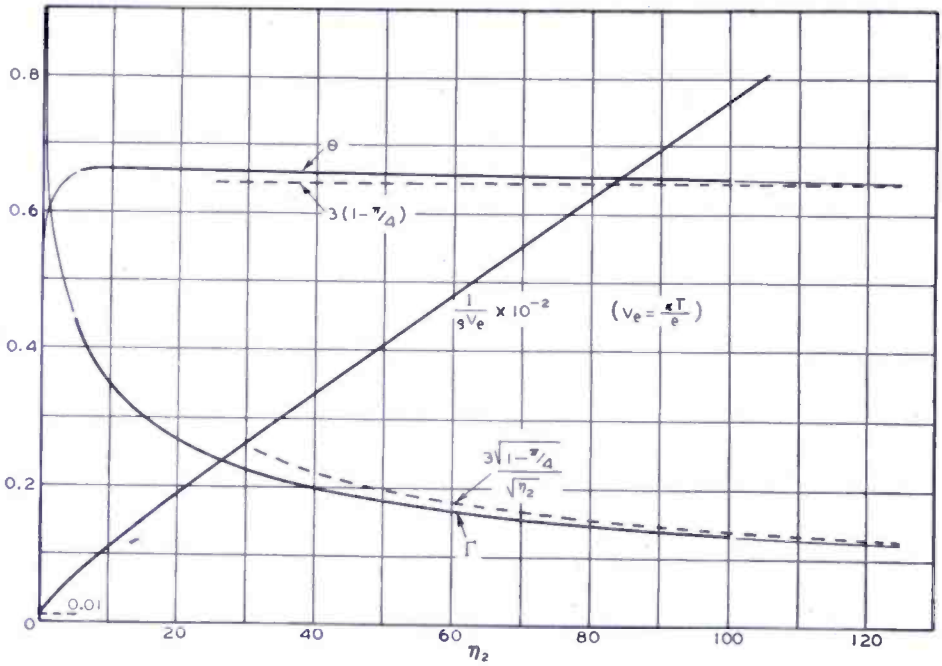


Fig. 5—Shot-effect reduction factors for the case, $I/I_s << 1$.

From the two curves of Figure 5, the plot of Γ in Figure 6 was made. This curve is entirely suitable for experimental checks of the analysis since current, conductance and cathode temperature are accessible parameters. Too, the reader has no doubt foreseen that representation of Γ in this way permits a test of the theory against non-planar structures.

The analysis might logically terminate at this point were it not for widespread curiosity over the connection (if any) between this type of shot effect and the current fluctuations of thermal origin (Brownian movement) known to exist in any short-circuited ohmic resistance. According to Nyquist's formula,[1] the thermal current fluctuations in an ohmic resistance whose conductance is g, and whose temperature is T are represented by

$$\overline{i^2} = 4kTg\triangle f.$$

(k = Boltzmann's constant, f = band width)

Let us then suppose that our space-charge-limited diode exhibits pseudo-thermal fluctuations (that they cannot be considered truly thermal has been emphasized in Part I). We, therefore, describe the space-charge-reduced current fluctuations by the empirical formula,

$$\overline{i^2} = \theta \cdot 4kTg\triangle f\,, \tag{41}$$

[1] H. Nyquist, "Thermal Agitation of Electric Charge in Conductors," *Phys. Rev.*, Vol. 32, p. 110, July, (1928).

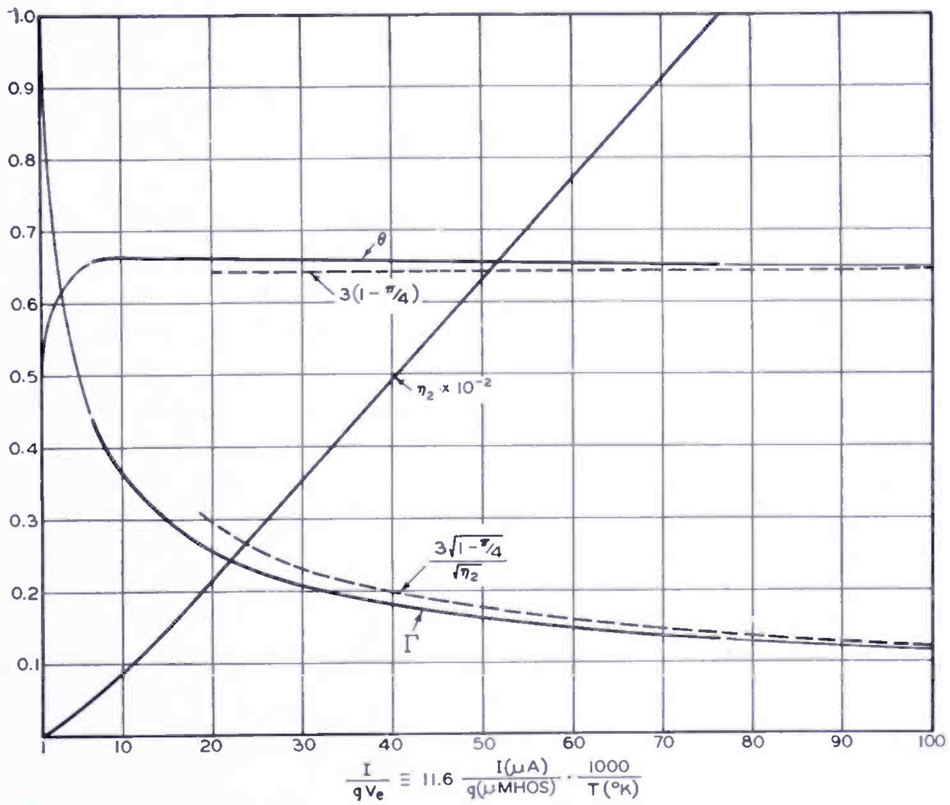


Fig. 6—Shot-effect reduction factors for the case, $I/I_s << 1$.

TABLE III

η_2	ξ_2	$\phi_\beta(\eta_2)^{1/2}$	D	Γ_α^2	Γ_β^2	Γ^2	Γ	$\frac{I}{gVe}$	θ
0	0	0	∞	0	1	1	1		
5	6.110	1.33	5.08	0.0282	0.1663	0.1945	0.441	6.76	0.657
10	9.447	1.66	6.60	0.0167	0.1043	0.1210	0.348	10.96	0.663
15	12.275	1.87	7.95	0.0115	0.0775	0.0890	0.298	14.9	0.663
20	14.826	2.04	9.18	0.0086	0.0621	0.0707	0.266	18.7	0.661
25	17.193	2.18	10.33	0.0068	0.0519	0.0587	0.242	22.5	0.660
30	19.425	2.30	11.43	0.0056	0.0448	0.0504	0.224	26.3	0.663
35	21.55	2.40	12.47	0.0047	0.0394	0.0441	0.210	29.9	0.659
40	23.59	2.50	13.47	0.0040	0.0352	0.0392	0.198	33.7	0.661
45	25.56	2.58	14.45	0.0035	0.0319	0.0354	0.188	37.3	0.660
50	27.47	2.66	15.39	0.0031	0.0291	0.0322	0.179	40.9	0.658
60	31.14	2.79	17.21	0.0025	0.0248	0.0273	0.165	48.0	0.655
70	34.64	2.92	18.94	0.0020	0.0217	0.0237	0.154	55.3	0.655
80	38.01	3.02	20.61	0.0017	0.0192	0.0209	0.145	62.2	0.650
90	41.26	3.13	22.23	0.0015	0.0173	0.0188	0.137	69.6	0.654
100	44.41	3.21	23.80	0.0013	0.0157	0.0170	0.130	76.4	0.649

$I/I_s \rightarrow 0$ so that $\xi_1 = -2.55$, $\eta_1 \rightarrow \infty$

and our problem is to discover the relationship between θ and η_2 or $I/g\,V_e$. Since we have formerly written

$$\overline{i^2} = \Gamma^2 \cdot 2eI\triangle f\,, \tag{42}$$

identification of (41) with (42) shows that

$$\theta = \frac{1}{2}\Gamma^2 \frac{I}{g\,V_e} = \frac{1}{2}\Gamma^2 D\phi_\beta(\eta_2)^{1/2}\,. \tag{43}$$

The right-hand side is a combination of known functions of η_2 and serves to define θ. Values of θ, calculated in this manner, are listed in Table III and plotted in Figures 5 and 6 against η_2 and $I/g\,V_e$, respectively. The asymptotic value is also shown (dashed line); its magnitude is found from (34) together with

$$\frac{I}{g\,V_e} = D\phi_\beta(\eta_2)^{1/2} \underset{(\eta_2\to\infty)}{\sim} \frac{2}{3}\eta_2\,, \tag{40a}$$

so that

$$\underset{(\eta_2\to\infty)}{\theta \sim} 3\left(1-\frac{\pi}{4}\right) = 0.6438\,. \tag{44}$$

In this place it should be recorded that the asymptotic solution was first developed by W. A. Harris of these laboratories. Working along similar lines and independently, he arrived at the above expression in March, 1936, prior to the completion of the author's analysis. Indeed, the simplicity of his result, together with its substantial accord with contemporary experimental work,[1] added incentive to the more detailed study just described.

The value of θ for vanishing η_2 (incipient retarding field condition) is also found easily from the following limit values:

$$\lim_{\eta_2\to 0}\left\{\begin{array}{l} \dfrac{I}{g\,V_e} \equiv D\phi_\beta(\eta_2)^{1/2} = 1 \\[2ex] \Gamma\alpha^2 = 0,\ \Gamma^2 = \Gamma_\beta{}^2 = 1, \text{ so that} \\[2ex] \theta = \dfrac{1}{2}\,. \end{array}\right. \tag{45}$$

[1] G. L. Pearson, "Shot Effect and Thermal Agitation in an Electron Current Limited by Space Charge," *Physics*, Vol. 6, p. 6, January, (1935).

It is interesting to note how rapidly θ rises and how closely it corresponds to the asymptotic value (44) throughout the whole practical working range. (Slight fluctuations in the tabulated values are to be attributed wholly to small errors arising in the numerical work.) Because of this fortunate property, equation (41) recommends itself as the formula for space-charge-reduced shot noise most useful in engineering application; and for most purposes sufficient accuracy will be obtained by assuming $\theta(\eta_2)$ to be simply a constant equal to the asymptotic value, namely, 0.644. Although, whenever one is principally interested in the noise *reduction* attributable to space charge, it will be convenient to revert to (43):

$$\Gamma^2 = 2\,\theta \frac{g\,V_e}{I} \sim 1.29 \frac{g\,V_e}{I}\,. \tag{43a}$$

Inasmuch as I increases more rapidly than g as (say) voltage is raised, it is apparent that although Γ^2 becomes smaller it decreases more slowly than the true shot effect increases so that the net result is a steadily increasing noise proportional to the diode conductance. It may be said then that *the mean-square noise generated by emission fluctuations in a space-charge-limited diode is roughly numerically equal to two-thirds of the noise of thermal agitation generated by a resistance of magnitude equal to the a-c resistance of the diode and possessing a temperature equal to the cathode temperature*[1]. Yet the two phenomena must not be confused in concept. For, thermal agitation is known to be a form of Brownian movement, and finds its origin in the equipartition of energy among the various mechanical and electrical degrees of freedom of a substance in thermal (i.e., kinetic or statistical) equilibrium. The diode, on the other hand, while clearly in a stationary state, cannot be regarded as a system in a condition of thermal equilibrium so long as there is a battery providing plate voltage and energy. The mechanics of the two phenomena are, therefore, distinct; the formulas alone exhibit a resemblance. The only instance in which complete identification can be made is that in which the plate voltage is zero and the whole diode (anode as well as cathode) is brought to a common temperature. This interesting situation will be discussed briefly in a later section.

It is of interest to perfect the picture of shot noise by considering the case of a retarding field. The mean-square noise at the threshold

[1] In our oral report of this week (Rochester Convention, November, 1936), we quoted "six-tenths" instead of the present "two-thirds," (cf. *Electronics*, Vol. 9, No. 11, p. 31, November, 1936). The alteration follows discovery of a numerical error which required that all values of Γ_α^2 be raised by the factor 16/3.

of a retarding field condition has already been shown, (45), to be precisely half the thermal-agitations value. One can easily demonstrate that equations (45) are all valid throughout the retarding-field region. There is no virtual cathode, hence no space-charge reduction; and since $I/g\,V_e = 1$, it is numerically immaterial whether one regards the noise as true shot effect expressed by

$$\overline{i^2} = 2eI\triangle f,$$

or half-thermal fluctuations expressed by

$$\overline{i^2} = \frac{1}{2} \cdot 4kTg\triangle f.$$

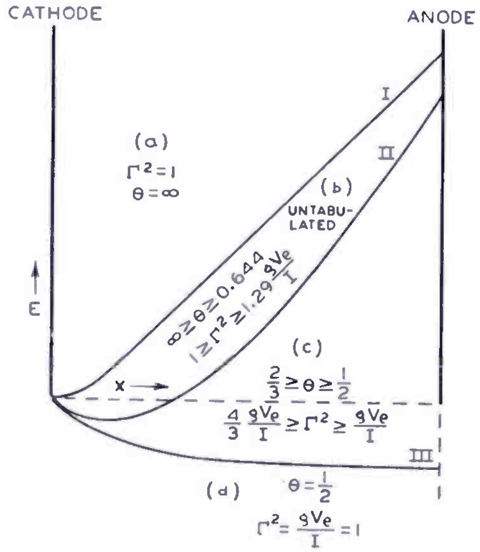


Fig. 7—Schematic survey of shot effect in a parallel-plane diode, all modes of operation.

This identification was first remarked in print by F. C. Williams,[1] who also showed good experimental agreement with the theory, and pointed out the nature of the modifications necessary when the formulas are to be adapted to cylindrical structures. These modifications are of no great consequence unless the ratio of anode to cathode diameter is much larger than customary in conventional equipotential-cathode structures.[2]

We are now in a position to diagram the shot effect for virtually all modes of operation of a diode which approximates a parallel-plane structure. The applicable formulas will be either or both of (41) and (42). In Figure 7 four operating regions, (a)—(d), are schematically demarcated by three curves of space potential. Curve I represents

---

[1] "Fluctuation Noise in Vacuum Tubes Which Are Not Temperature-Limited," *Journ. I.E.E.*, Vol. 78, No. 471, p. 326, March (1936).

[2] Cf. footnote, p. 468.

that potential distribution for which the virtual cathode coincides with the cathode $\left(\frac{\partial E}{\partial x} = 0 \text{ at cathode}\right)$; curve *I* shows qualitatively a mode of operation such that $I/I_s = 1/5$ (say); curve *III* gives the potential distribution for which the virtual cathode coincides with the anode $\left(\frac{\partial E}{\partial x} = 0 \text{ at anode}\right)$, which is the condition of incipient retarding field. Region (a): there is no virtual cathode, and the diode is, therefore, temperature-limited so that the classical shot-effect formula obtains, i.e., $\Gamma = 1$. No significance can be attached to (41) for, except for field currents or other anomalous behavior, $g = 0$. Region (b): no tabulation has been made for this region where I/I_s is not sufficiently small that the present analysis can be used with negligible error. Close study is so lengthy and provides so little additional engineering information that mention will be made here of only a few rough estimates of error incurred if the tabulated values of Γ^2 and θ are employed. The value $I/I_s = 1/5$ is of course an arbitrary choice for the boundary curve *II*. If $\eta_2 = 10$ (anode potential approximately 1 volt, for a cathode temperature of $1000°K$), the tabulated value of Γ^2 is low by approximately 1 per cent and θ is low by approximately 1½ per cent, when $I/I_s = 1/5$. Larger values of η_2 (higher anode potentials or lower cathode temperatures) reduce the error. For $I/I_s = 1/3$ and the same η_2, the errors in Γ^2 and θ are approximately 4 and 6 per cent, respectively. And for $I/I_s = 1/2$ the errors are only about 10 and 15 per cent, respectively. It is clear, then, that Γ^2 departs appreciably from the tabulated values only when the tube is nearly temperature-limited and then climbs rapidly to unity as I approaches I_s.[1] Region (c): this is the region of practical operation of virtually all biased diodes giving high emission. The virtual cathode effects a reduction in emission fluctuations, and the tabulated values of Γ^2 and θ should show negligible error. Region (d): with the appearance of a monotonic retarding field, the virtual cathode and, therefore, the reduction of shot effect simultaneously vanish. Throughout this region $\Gamma^2 = 1$, $\theta = 1/2$, and when used in their respective formulas, they produce identical numerical results.

[1]Spenke has carried out the complete numerical integration for two examples and shows for them the sharp rise of Γ to unity described here; E. Spenke, "The Space-Charge-Reduction of Shot Effect," *Wiss. Veröff. aus den Siemens-Werken*, Vol. 16, No. 2, p. 19, July, (1937). In a more recent study, Rack has shown graphically the influence of the ratio I/I_s for the complete range, $0.05 \leqslant I/I_s \leqslant 1$; A. J. Rack, "Effect of Space Charge and Transit Time on the Shot Noise in Diodes," *B.S.T.J.*, Vol. 17, No. 4, p. 592, October, (1938).

The Diode in Thermal Equilibrium

It was mentioned above that one could expect the shot effect in a diode to be completely identified with the thermal agitations of its resistance only when the diode is unenergized and the whole of it put at one temperature. Let us demonstrate this identity by showing that the formula for shot effect proves to be, in this instance, identical with Nyquist's expression for the thermal agitation. Consider an electrode (1) surrounded by an electrode (2), the material of which each is constructed being unimportant. The electrodes are short-circuited and the diode is maintained at temperature T. Both surfaces emit, but at rates which are not necessarily equal, since, for example, the work functions are not prescribed. Yet the current I_1, that portion of the emission from the first electrode which is absorbed by the second, must be equal to I_2 (similarly defined) *on the average*. This statement will be recognized as a direct outcome of the second law of thermodynamics. The average current through the short-circuiting link is therefore zero, but fluctuations are present nevertheless. If these fluctuations are traced back to variations in emission, they are properly labeled "shot effect"; yet, if they are simply accepted as a manifestation of Brownian motion, they are just as properly termed "thermal agitation".

Adhering for the moment to the latter point of view, we find that Nyquist's formula provides a valid measure of the fluctuations. For, we need only observe that we have here a two-terminal, unenergized circuit element at a uniform temperature. Its short-circuit fluctuations are, therefore,

$$\overline{i^2} = 4kTg\Delta f, \tag{46}$$

where g is the real part of the diode admittance at the frequency in question. Confining ourselves, as usual, to frequencies at which transit-time effects are negligible, we may say that g is simply the low-frequency conductance. It is not difficult to obtain an expression for g. Suppose that a voltage E is included in the circuit, making the outer electrode more positive. In case the situation is not complicated by individual collisions,[1] it can readily be shown by kinetic theory that

$$I_1/I_2 = \varepsilon \frac{E}{V_e}, \tag{47}$$

[1] This assumption is reasonable, except for exceedingly high space-charge densities, not likely to be encountered in practice. Cf. I. Langmuir and K. T. Compton, "Electrical Discharges in Gases," *Rev. Mod. Physics*, Vol. 3, No. 2, p. 220, (1931).

regardless of the geometrical configuration, coefficients of reflection, contact potential, etc. Consider now variations of I_1 and I_2 against E with T fixed:

$$\frac{\partial I_1}{I_1} - \frac{\partial I_2}{I_2} = \frac{\partial E}{V_e}.$$

The conductance is, therefore,

$$g = \left[\frac{\partial (I_1 - I_2)}{\partial E} \right]_{\substack{E=0 \\ I_1 = I_2 = I}} = \frac{I}{V_e}, \tag{48}$$

where I is the equilibrium current emitted by either electrode and absorbed by the other. And this expression is valid whether or not there exists a virtual cathode in the equilibrium state. Consequently, the thermal fluctuations can be written

$$\overline{i^2} = 4Ie\Delta f. \tag{49}$$

Turning now to the kinematic viewpoint, we note that the emission current from either electrode exhibits true shot effect, so that the mean-square current fluctuations in the short-circuiting link are simply the sum of the separate shot effects associated with the currents I_1 and I_2. To calculate this sum, we must first determine the space-charge-reduction factor Γ for each current. Suppose that the equilibrium state is perturbed by the injection of an additional small steady emission i_s from the first electrode. The emission velocities of the electrons comprising i_s will be permitted any values whatever. The question is: In what way is the original equilibrium affected? Now except for individual encounters, which we suppose highly improbable, the electrons in i_s disturb the equilibrium only in so far as they introduce a potential field superimposed upon that which existed initially. The diode, therefore, settles into a new equilibrium state such that the space-charge density is everywhere slightly altered, yet, except for the injected electrons i_s, at every point in space the M-B distribution still obtains. The originally equal equilibrium currents I_1 and I_2 are now somewhat reduced, but still equal. It follows that whatever fraction of i_s is absorbed on the second electrode is accompanied by no net increment in the equilibrium current, and this is true whether or not there exists a virtual cathode in the equilibrium state. There is consequently no space-charge reduction of noise, $\Gamma = 1$, and

$$\overline{i^2} = 2e(I_1 + I_2)\Delta f = 4eI\Delta f. \tag{50}$$

In the equilibrium case, then, the two viewpoints are only alternative aspects of the same phenomenon. The identification is logical and complete. Where space-charge densities are low enough to make individual collisions unimportant, all of the dynamical processes which maintain the electrons in thermodynamical equilibrium occur within the electrode material. The space between electrodes is only a passageway through which electrons travel without incident.

With this in mind, we can now understand thermodynamically that the shot effect of a normal diode operating in the retarding-field condition should have turned out to be precisely half-thermal. Consider a diode in thermal equilibrium, and, if there is a virtual cathode, let the emission from the electrode of higher work function become smaller by raising the work function until the virtual cathode has disappeared. If the diode is parallel-plane, this emission is now temperature-limited, i.e., it contributes nothing[1] to the conductance g given by (48). On the other hand, it is *still* responsible for half the shot effect as represented by (50), because $I_1 = I_2$ is still valid, nothing having been done yet to destroy thermal equilibrium. But now we remove the temperature-limited current altogether, say, by lowering the emitter to room temperature. The noise is halved, the conductance is unaltered, and the half-thermal fluctuations of a parallel-plane diode operating in retarding-field condition are provided with thermodynamical significance.

NEGATIVE-GRID TRIODES

The foregoing study is, of course, but a preliminary to the problem of chief interest, namely, noise in amplifying tubes. To see how the theory of diodes can be adapted to a triode, imagine the conversion of a parallel-plane diode by adding grid wires one at a time. When only two or three grid wires are in place, and biased negatively, the space potential is badly disrupted. Moreover, electrons on the plate side of the grid wires still contribute materially to the potential in the vicinity of the virtual cathode. Noise analysis at this stage would be painfully difficult. Hence, there is little to be said about very low-μ tubes, except,

[1] It should be noted that this statement is partly hypothesis. If the coefficient of reflection is a function of velocity, the conductance of any unilateral current flow will be affected, whether the field be accelerating or retarding. The nature, even the existence of such a function, is still moot (cf. Compton and Langmuir, *Rev. Mod. Phys.*, Vol. 2, No. 2, p. 171, 1930; also W. B. Nottingham, *Phys. Rev.*, Vol. 49, p. 83, 1936). Overlooking this difficulty as well as the possibility of field emission, the reader should observe that the statement is still not applicable to all geometrical configurations. As long as there are any electrons which, from an energy standpoint alone, would fall on the collector, but which fail to do so because of momentum considerations, the conductance is finite for a temperature-limited current, and is not easily formulated for any mode of operation. It is on this account that the present analysis is restricted to parallel-plane structures.

perhaps, to set limits for the noise. When, with the addition of more grid wires, the control grid stands complete, with a μ of say, 5 or more, the space potential can generally be considered very regular again, except in the immediate neighborhood of the grid wires. A first approximation to the shot current fluctuations, exact in the limit of an infinite μ, would then amount to neglecting the grid-plate region altogether, and applying the diode analysis directly to the grid-cathode space. The anode potential of the diode, E_2, would have to be interpreted now as the "effective" potential of the grid plane, E_a, i.e., that potential which, when applied to a solid sheet in the grid-plane, would draw the same cathode current. In addition, g would now be viewed as the conductance of this equivalent diode. This procedure recognizes the contribution which electrons in the grid-plate space make to the steady-state space potential on the other side of the grid, but does not take account of their contribution to Γ^2. For this reason, theoretical values of shot effect so determined should be a little too high. The error, however, should be practically unimportant where high-μ tubes are concerned. For, in the first place, an electron's contribution to noise reduction must vary inversely as its velocity and its distance from the virtual cathode. Second, the grid mesh is a good electric shield.

$$\left.\begin{aligned}
E_a &= \sigma\left(E_{c1} + \frac{1}{\mu_o}E_b\right)\\
\sigma &= \left\{1 + \frac{1}{\mu_o}\left[1 + \frac{4}{3}y(1+h) - \frac{1}{3}h^2(6+4h+h^2)\right]\right\}^{-1}\\
y &= \frac{x_p}{x_c} = \frac{\text{grid-anode spacing}}{\text{cathode-grid spacing}}\\
h &= \frac{\tau_p}{\tau_c} = \frac{\text{grid-anode transit time}}{\text{cathode-grid transit time}}\\
\mu_o &= \mu\left(1 - \frac{h^3}{y}\right)\\
\mu &= \text{"cold", i.e., electrostatic amplification factor}
\end{aligned}\right\} \quad (51)$$

To put this procedure into effect, one must relate E_a and g to the grid and anode voltages. It will be assumed throughout that the grid has negative bias so that the plate alone collects electrons. The best formulation of E_a to date is (51), derived by Llewellyn,[1] which takes into account the influence of space charge upon space potential, but assumes zero emission velocities.

The grid-plate transconductance (for negligible transit angles) is, therefore,

$$g_m = \sigma g, \tag{52}$$

and since it is easily measured, whereas g cannot be measured at all, g_m will be used in place of g in the expressions for shot effect. Thus

(41) becomes
$$\overline{i^2} = \frac{\theta}{\sigma} \cdot 4kTg_m \Delta f, \tag{41a}$$

and (43a) becomes
$$\Gamma^2 = 2 \frac{\theta}{\sigma} \frac{g_m V_e}{I} \sim \frac{1.29}{\sigma} \frac{g_m V_e}{I}. \tag{43b}$$

Under ordinary circumstances σ will decrease somewhat with an increase in current, because h increases. This behavior may be marked for low-μ tubes, but becomes less pronounced as μ is increased, since

$$\lim_{\mu \to \infty} \sigma = 1 \quad .$$

It will not in general be correct, therefore, to assume that σ is strictly a constant; for conventional tubes, it will usually lie between 0.5 and 1. Inasmuch as h is usually less than 0.5, rough estimates of σ for parallel-plane structures may be made by supposing $h = 0$, whence[2]

$$\sigma \sim \left\{ 1 + \frac{1}{\mu} \left[1 + \frac{4}{3} y \right] \right\}^{-1}. \tag{51a}$$

No rigorous formula analogous to (51) has yet been developed for cylindrical structures. The analogue of (51a), however, has long been known:[2]

$$\sigma_{\text{cyl.}} \sim \left\{ 1 + \frac{1}{\mu} \left[1 + \frac{2}{3} \log \frac{r_p}{r_g} \right] \right\}. \tag{51b}$$

This expression is limited to structures in which the ratio of grid to cathode radius is larger than about 10. For smaller ratios the formula for plane structures will suffice.

The current fluctuations in the plate circuit (without load) are now seen to be principally a function of conditions between cathode and grid. And in contrast with the diode fluctuations, it is no longer pos-

[1] F. B. Llewellyn, "Operation of Ultra-High-Frequency Vacuum Tubes," *B.S.T.J.*, Vol. 14, p. 659, October, (1935).

[2] B. D. H. Tellegen, "The Value of Cathode Current in a Triode," *Physica*, Vol. 3, p. 301, (1925).

sible to find a simple empirical relation between triode shot effect and thermal agitation in an ohmic resistance of magnitude equal to the anode resistance. Following (41a), attention is more properly focused upon the transconductance.

The shot voltage $\overline{e_o^2}$ appearing across a load Z in the output circuit of a negative-grid triode whose plate resistance is r_p, is from (41a),

$$\overline{e_o^2} = \frac{\theta}{\sigma} \,.\, 4kTg_m\triangle f. \left| \frac{r_pZ}{r_p + Z} \right|^2 \qquad (53)$$

Inasmuch as the principal sources of local noise in amplifiers are thermal agitation in the input circuit and shot effect in the first tube, it has often been remarked that, instead of (53), a more useful formula for shot effect is found by referring the plate-current fluctuations back to the input circuit. This effective input fluctuation voltage is then that voltage which, applied between cathode and grid, produces in the output fluctuations equal in magnitude to those actually generated by shot effect. The effective input shot effect can then be compared directly with the thermal agitation of the input circuit. In fact, instead of expressing the shot effect as a voltage, we may symbolize it by a resistance R_{eff}, i.e., that resistance which, at ambient temperature, exhibits a thermal agitation voltage equal to the effective input shot voltage of the tube. This we do as follows. Since the transconductance, g_m, represents the ratio of output current (without load) to input voltage, we have from (41a),

$$\overline{ei^2} \equiv \frac{\overline{i^2}}{g_m{}^2} = \frac{\theta}{\sigma} \,.\, \frac{4kT}{g_m} \,.\, \triangle f. \qquad (54)$$

Nyquist's expression for thermal agitation voltage across a resistance R_{eff} at room temperature T_o is

$$\overline{e_i^2} = 4kT_o\, R_{eff}\triangle f. \qquad (55)$$

Equating the two voltages, we obtain

$$R_{eff} = \frac{\theta}{\sigma} \,.\, \frac{T}{T_o} \,.\, \frac{1}{g_m} \qquad (56)$$

A brief numerical estimate will serve to indicate the order of magnitude of R_{eff}. Using figures typical of coated-cathode receiving tubes, let us choose $T = 1000°K$, $T_o = 300°K$, $\theta = 2/3$, $\sigma = 3/4$. Then,

$$R_{eff} \approx 3/g_m.$$

With transconductances of 1000, 1500, and 10,000 micromhos, respectively, such a tube would exhibit shot effect equivalent to the fluctuations arising from thermal agitation in resistances connected between grid and cathode of magnitude 3000, 2000, and 300 ohms, respectively.

(The close approximation of these estimates to measured values of R_{eff} already published at once attests the validity of the theory and implies its successful adaptation to cylindrical structures.) So long as the input impedance actually employed presents a real component R $(Z = R + j X)$ two or three times greater than R_{eff}, the limit of useful amplification is set by thermal agitation; this is generally the state of affairs in the case of sharply tuned circuits and high-gain tubes, e.g., in broadcast receivers properly designed. When the reverse is true, and $R_{eff} \geq R$, the shot-effect problem becomes acute and anything which can be done to restore the original inequality, $R_{eff} << R$, without slighting performance requirements will increase the intelligibility of small signals. A case in point is the television receiver; in brief, the requirement of a pass-band width of several megacycles necessitates a low tuned-input impedance comparable in magnitude to the R_{eff} of conventional tubes, so that shot effect may be a serious concern.

A more detailed study of this and other noise problems is postponed (Part V) until we have completed our analysis with a description of shot-effect phenomena peculiar to multi-collector tubes, i.e., those in which the cathode-current stream is collected at two or more electrodes (Part III). For the present it will be sufficient simply to note the obvious recommendations of (56). As stated previously, θ is practically a constant and little can be done to reduce it. Similarly σ is practically a constant and generally offers, at most, a maximum of 50 per cent improvement. Furthermore, T is at present confined to rather limited regions in the vicinity of $1000°K$ for coated cathodes and $1900°K$ for thoriated tungsten. Other things being equal, coated cathodes are, therefore, to be preferred. It is, of course, possible to lower the temperature below the recommended value and still maintain performance, but in general a drop of 200° or so brings the tube so near the temperature-limited condition that the formulas given are vitiated, and the signal-to-noise ratio decreases sharply (see Figure 11). Fortunately, there always seem to be ways to improve transconductance; although the mean-square output noise is proportional to g_m, the equivalent input noise varies as $1/g_m$, so that it is profitable to use as high a transconductance as possible. This notion is hardly novel, for it has long been recognized that increase in gain, *per se*, would improve the input signal-to-noise ratio. The derived proportionality of mean-square output noise to g_m serves to render this expected improvement not inexistent, but simply smaller than conceived in dreams. Other things being equal, then, a change in structure which doubles g_m should cut the mean-square input noise of a triode in half —not a magnificent improvement, but the only thoroughly practical method available.

(To be continued)

A METHOD OF MEASURING FREQUENCY DEVIATION

By

MURRAY G. CROSBY

Engineering Department, R.C.A. Communications, Inc.

***Summary**—A method of calibrating the frequency deviation of a frequency modulator is described in which the carrier is heterodyned to a beat note and filtered so that the carrier zero points may be observed as modulation is applied. From the known conditions existing at the carrier zero points the frequency deviation may be determined.*

THE method of measuring frequency deviation about to be described here has been used extensively by the writer in measuring frequency and phase deviations and is believed by him to be the simplest and most reliable of any system used. The method is based upon the amplitude characteristic of the carrier in a frequency-modulated wave when a single sinusoidal modulating wave is applied. It is well known that the amplitudes of the carrier and side frequencies of a frequency-modulated wave vary according to Bessel Functions as shown in Figure 1. The carrier is proportional to the zero-order Bessel Function, $J_o(F_d/F_m)$, where F_d is the peak frequency deviation, F_m is the modulating frequency, and the ratio F_d/F_m is called the modulation index. The frequency deviation concerned in this case is the amount the frequency varies away from the unmodulated carrier frequency on one side, and would be one-half the total excursion of the wave on both sides of the carrier frequency. The total excursion is sometimes called the "swing".

The first side frequency is proportional to the first-order Bessel Function, $J_1(F_d/F_m)$, the second side frequency is proportional to the second-order function, $J_2(F_d/F_m)$, and so on. As an example, if the frequency deviation is 10,000 cycles and the applied modulating frequency has a frequency of 5,000 cycles, the modulation index would be equal to 2. For this index, the carrier has an amplitude of 0.224 times the amplitude of the unmodulated carrier (see Figure 1). The first side frequencies have an amplitude of 0.577 times, and so on.

Figure 2 shows how the amplitude of the carrier varies as the modulation index is varied when the phase reversals are disregarded. It will be noted that the carrier has unit amplitude when the index is zero ($F_d = 0$). As the index is increased, either by increasing F_d or decreasing F_m, the carrier amplitude decreases and becomes zero when the index has the value 2.405, 5.52, 8.654, etc. It can be seen that if a method of detecting these zero points is available, each point furnishes an ideal calibration point if the frequency of the modulating tone is known. For instance, if a 1,000-cycle modulating tone is applied and modulation increased until the carrier is at the first zero point, it will be known that the frequency deviation is 2,405 cycles. This follows since it is known that for this condition the modulation

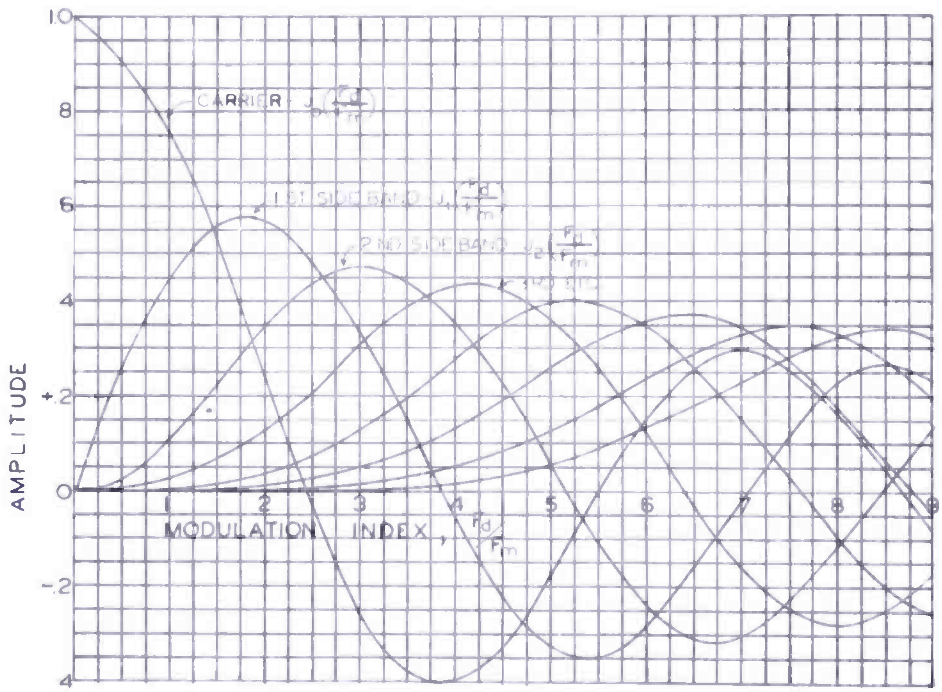


Fig. 1—Variation of the amplitudes of the carrier and side frequencies of a frequency modulated wave as the modulation index is varied. It will be noted that these Bessel Functions, which give the carrier and side-band amplitudes, are quite similar to the ordinary sine and cosine functions. Thus instead of having sin X and determining the value of sin X from a table of sine and cosine functions, we have Jn(X) and use a table of Bessel Functions.

index, F_d/F_m, is equal to 2.405. Then $F_d = 2.405$ times 1,000 or 2,405 cycles. Likewise, at the second minimum the deviation will be 5,520 cycles for the case of a 1,000-cycle modulating tone. If the modulating tone were increased to 2,000 cycles, the first zero point would indicate a frequency deviation of 4,810 cycles, the second 11,040 cycles, and so on. Thus, the modulation frequency may be varied to obtain the frequency deviation of a modulation input level which would not normally produce a carrier zero.

Figure 3 shows a diagram of the apparatus required for the detection of the carrier zeros. The unmodulated carrier is tuned in on a receiver which has a beating oscillator capable of heterodyning the carrier to an audio beat note. This beat note is passed through an audio filter which removes the side bands of the wave from the frequency-modulation generator being calibrated. For instance, if the modulating tone were 1,000 cycles, the filter would be something less than 2,000 cycles wide so as to remove the side bands spaced from either side of the carrier by 1,000 cycles. The indicator at the output of the filter may be headphones or a loud-speaker.

The measurement procedure consists of recording the modulation input levels which correspond to the carrier minimums. The unmodulated carrier is first tuned through the audio filter to give

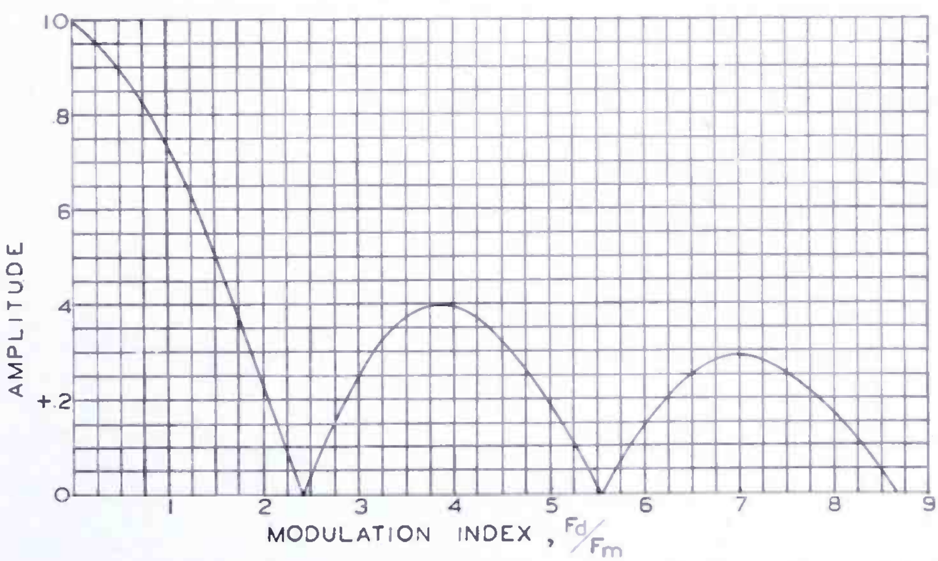


Fig. 2—Variation of the amplitude of the carrier alone as the modulation index is varied. The phase reversals indicated in Figure 1 are disregarded in order to more clearly picture the amplitude characteristic of the carrier as the modulation index is varied.

maximum output in the phones or loud-speaker. Modulation is then applied until the first carrier zero is obtained. At this point the frequency deviation is equal to 2.405 times the frequency of the modulating tone. Increasing the modulation still further produces the second minimum at which point the frequency deviation is equal to 5.52 times the frequency of the modulating tone. The following is a list of the modulation indexes obtained at the carrier zeros as determined from a Bessel Function table[1]:

[1] The most complete set of Bessel Function Tables known to the writer is contained in the British Association Report, 1915, p. 29-32. Photostatic copies of these tables may be obtained from the Engineering Societies Library at 29 West 39th Street, New York City at a cost of seventy-eight cents. The following books also contain somewhat less extensive tables: Jahnke-Emde, "Tables of Functions". Gray and Mathews, "Treatise on Bessel Functions".

Zero	*Index*	*Zero*	*Index*
1	2.405	6	18.071
2	5.520	7	21.212
3	8.654	8	24.353
4	11.792	9	27.494
5	14.931	10	30.635

It will be noted that the first zero differs from the second by an amount equal to 3.115 which is almost equal to 3.1416 or π. This difference approaches π for the higher zeros and for practical usage in measuring deviation may be taken as equal to π (with an error of less than 0.5 per cent).

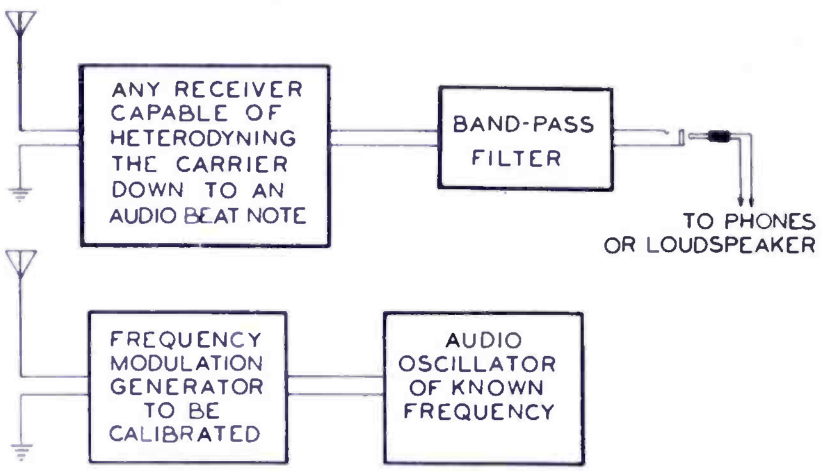


Fig. 3—Apparatus used to measure frequency deviation by the carrier zero method.

The band-pass filter does not have very rigid requirements. A selectivity about equal to that obtainable with a single audio tuned circuit has been found to be sufficient and the author has successfully observed the zeros with the aid of the natural resonance of a poor pair of headphones. If a high modulation frequency is used, the receiver itself can be depended upon to remove the side bands sufficiently. It is usually the lower modulation frequencies which require the highest selectivity.

It may be found that the carrier shifts out of the filter as the modulation is applied. When this occurs the modulation must be raised slowly and the receiver carefully tuned to follow the shift. Apparently a small amount of shift may occur with a modulator which gives low distortion otherwise.

It is apparent that one of the side bands may be tuned in and the zero points of it observed. The zero points of the side bands will, of course, be different than those of the carrier and will have to be obtained from the Bessel Function Tables. However, use of the side bands permits the possibility of getting the wrong side band. Also, there is the possibility of asymmetrical modulation, due to concomitant amplitude modulation, affecting the reading. Likewise, the carrier maxima or side-band maxima may be used, but of course the setting for the minima can be performed with a higher degree of accuracy.

THE LIMITS OF INHERENT FREQUENCY STABILITY

By

Walter Van B. Roberts

Patent Department, Radio Corporation of America

GREAT improvements in oscillator stability have been made in recent years by careful attention to the mechanical design and layout of component parts and also by the development of a number of compensation arrangements. An example of the latter is the compensation for a change in plate voltage by a suitable change in screen voltage. By the proper use of such refinements and by other precautions, frequency variations due to variations of input and output impedances of the tube may be much reduced. However, it seems obvious that if the circuit could be made more stable in the first place, the addition of these schemes would bring about still better final results.

The many causes of frequency variations may be divided into three groups: First, changes in the constants of the frequency determining circuit itself; second, changes introduced by the loading on the circuit; and third, changes in the effective input and output impedances of the oscillator tube which are reflected into the circuit by the necessary coupling of the tube to the circuit. In what follows, only the third group will be considered, and the term "inherent stability" is used to refer to the extent to which the frequency is independent of small changes in the effective tube impedances. The object of this investigation is therefore to determine just how far one can go in reducing the effect of given capacitance changes in a tube on the frequency of any ordinary oscillator circuit.

Figure 1 shows a representative simple feed-back oscillator in which the small capacitances C_g and C_p represent the maximum *variations* that may be expected in the input and output circuits of the tube. It is of course possible that these variations may sometimes take place in opposite senses so as to tend to compensate for each other. However, in order to deal with the worst case possible they will be assumed to take place in the same sense and at the same time. In this case the resulting frequency shift, measured in cycles per second, may readily be shown to be given approximately by the expression

$$\frac{f(\omega M_g)^2 C_g}{2L} + \frac{f(\omega M_P)^2 C_P}{2L} \tag{1}$$

in which f is the oscillator frequency in cycles per second, ω is $2\pi f$, the inductance L and the mutual inductances M_g and M_P are measured in henrys and the capacitances are in farads.

If now the mutuals are reduced to the point where the system just barely oscillates and if, furthermore, their ratio is adjusted to give the least possible frequency shift when C_g and C_P disappear or reappear, then the expression

$$f\left(\frac{r}{L}\right)\left(\frac{\sqrt{C_g C_P}}{g}\right) \tag{2}$$

gives the smallest frequency shift that can be obtained in the presence of the capacitance variations C_g and C_P. In this expression r is in ohms and g is the transconductance of the tube in mhos. The derivation of the expression will be given in the appendix.

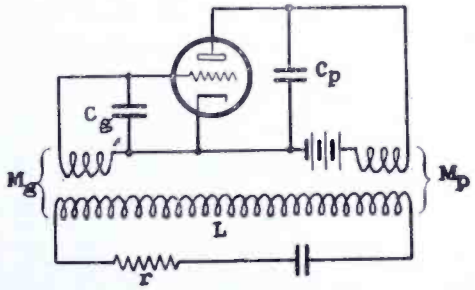


Fig. 1

From expression 2 it can be seen that the stability is limited by three independent factors. First of all, the minimum possible frequency shift in cycles is seen to be proportional to the frequency of operation, which is natural enough and is merely a way of saying that the percentage frequency shift is independent of frequency. Secondly, this shift is proportional to the ratio r/L which means that a good coil is desirable, a conclusion that again is less than startling. Finally, the shift is proportional to the ratio of the geometric mean of the tube capacitance variations to the transconductance. This result is a little less obvious and might lead to the choice of a tube not ordinarily considered particularly well suited to oscillator use. For example, a certain tube may have a rather large variation of input capacitance, say 1.0 mmf. Nevertheless, if its output capacitance is constant to within 0.01 mmf the geometric mean variation is only 0.1 mmf and the tube is preferable, other things being equal, to one which has only say 0.2 mmf variation at most, but has this much variation of both its input and output capacitances.

In the foregoing it was assumed that the circuit was just barely oscillating, and the looser the couplings can be made the greater the stability up to the limit given by expression 2. In practice, of course, the couplings would be made somewhat closer to allow a factor of safety in starting the oscillation, and also, to obtain a sufficiently strong oscillation to be of some use. However, for any given factor of safety, expression 2 will be proportional to the actual frequency variation so that conclusions drawn from it will still be valid.

Harmonic Operation

When an oscillator is used to obtain excitation in several frequency bands, it is common practice to run it at the frequency of the lowest band, or even a submultiple thereof, and to obtain excitation for the other bands by frequency multiplication. Let us see what conclusions can be drawn from expression 2 regarding this mode of operation. To be specific, suppose the fundamental frequency is in the range from about 830 to 1020 kilocycles, say 900 kc. (This range is very easily calibrated by beating with broadcast stations and also by beating the sixth harmonic of 833.333 kc and the fifth harmonic of 1000 kc with the 5-megacycle transmissions of WWV). The number of cycles shift is given by expression 2, and the number of cycle shift of the 28.8 megacycle harmonic is thirty-two times as great since the latter frequency is the thirty-second harmonic of the oscillator. Thus, the formula for the shift in the ten-meter band would be 28,800,000 $\left(\frac{r}{L}\right)\left(\frac{\sqrt{C_g C_P}}{g}\right)$. But this is the same formula that would be used if the oscillator were running at 28.8 megacycles as its fundamental except that the ratio r/L and the ratio $\frac{\sqrt{C_g C_P}}{g}$ were evaluated at 900 kilocycles in the one case and at 28.8 megacycles in the other. Hence, it is seen that expression 2 may be generalized to take care of harmonic operation as follows:

$$f_{\text{radiated}} \left(\frac{r}{L} \frac{\sqrt{C_g C_P}}{g}\right)_{\text{fundamental}} \qquad (3)$$

where the subscripts indicate that the frequency is taken as the radiated frequency while the rest of the expression is evaluated at the fundamental oscillation frequency.

The interesting thing about expression 3 is that it indicates that for any given radiated frequency the actual number of cycles shift caused by tube variations can be reduced in theory by using a low enough fundamental oscillation frequency. This is partly because the ratio of tube capacitance changes to transconductance is somewhat lower at the lower frequencies, but mostly because a low-frequency coil can be made to have a very much lower r/L ratio than a high-frequency coil of the same physical size, as is evident from the fact that the selectivity in cycles of a tuned circuit is proportional to r/L, and the fact that low-frequency circuits are much more selective than high-frequency circuits in terms of actual cycles. In practice of course, it would be too complicated to multiply all the way up from audio frequency for example, but a great improvement may be obtained by multiplying from reasonably low frequencies such as the range from 850 to 1000 kilocycles. To illustrate, let us substitute some reasonable values in expression 3. If the fundamental frequency is between 850 and 1000 kc, and a "Q" of 200 is assumed, the ratio r/L is about 30,000. Taking g as 3000×10^{-6} and the mean capacitance variation as 1/10 micromicrofarad, the number of cycles shift given by expression 3 is, approximately, numerically equal to the output frequency measured in megacycles. Thus, at 14 Mc the frequency variation would be 14 cycles. This is 14 times better than if the oscillator had been run at a 14 Mc fundamental, using a coil of the same "Q", and assuming the same capacitance variation and transconductance.

Since the amount of tuning capacitance has not appeared in the expressions for minimum shift, it may be concluded that the stiffness of the circuit is of no importance unless it affects the ratio r/L. Data already published indicate that this ratio will be very little different in coils of the same size and shape, but wound for different inductances, using the optimum wire size in each case. Thus, there would seem to be a good deal of latitude in the amount of capacitance that may be used. If the variable condenser is very large, however, it is likely to have large and relatively flexible plates and small clearances, all of which may introduce vibration troubles and changes of the calibration curve with aging. Hence, it does not appear desirable to approach maximum stability by using a "high C" circuit with the tube electrodes connected across the whole circuit, as the amount of capacitance required for maximum stability in this type of circuit may be many thousands of micromicrofarads.

Practical Circuits

The circuit of Figure 1 was chosen for simplicity of analysis. In practice it is likely to give parasitic oscillations. The same is true of many circuits where grid, cathode, and plate are tapped to the coil at points close together in order to loosen the couplings, as was assumed at the start of the derivation of expression 2 from expression 1. Parasitics may, of course, be suppressed by inserting resistances at suitable points, but this is likely to increase the effective resistance at the desired frequency and hence increase the r/L ratio. It is preferable to use a circuit which does not develop parasitics. Such a circuit is included in Figure 2 which shows the essentials of an exciter that has been in use for some time. The 100 mmf condenser is adjusted until the desired frequency band, 850 to 1000 kc, is just a little more than covered by variation of the 50 mmf condenser after which the

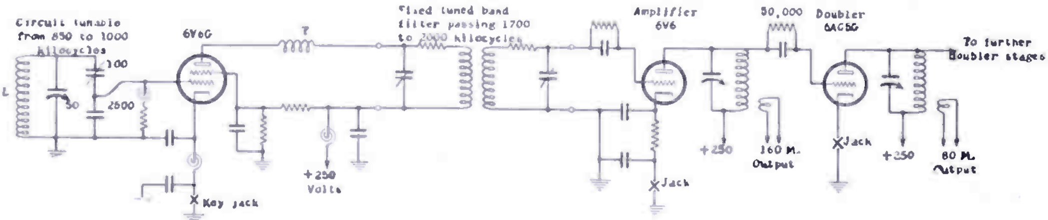


Fig. 2—L is a coil of #24 wire wound 32 turns per inch on a bakelite tube two and a quarter inches in diameter, the length of winding being two and three-quarters inches. T is the feed-back or tickler coil and consists of about six turns wound over the grounded end of L. An aluminum box contains the entire tuned circuit, grid choke and leak, coil T, and the band-pass filter.

former is left severely alone so as to keep the calibration of the oscillator unchanged. These two condensers are physically a single unit made for band-spread use, and only the 50-mmf section has a shaft on it. The other data shown in connection with Figure 2 are what are actually in use, but have not been worked out by cut and try to their best values. The whole arrangement is merely illustrative and doubtless could be considerably polished up. In particular the band-pass filter designed to pass with fair uniformity all frequencies between 1700 and 2000 kc could be much improved by experimenting with different damping resistors and varying the coupling between coils. This filter is fixed-tuned so as to avoid any tuning reaction on the oscillator. Incidentally the grid leaks on the doublers really are connected as shown since the amplification constant of these tubes is so high that with the leaks connected from grid to ground the plate currents fall to nothing when the oscillator key is up. By connecting as shown, the plate currents stay up and keep the load on the power supply nearly constant during keying, and also, no r-f chokes are needed in series with the leaks. As is evident from the diagram, excitation for any band can be obtained by connecting a transmission line to any of the tank links, the only other change when changing bands being that it is well to pull out the tube following the link selected so as to get all the power available from the tank.

While Figure 2 represents the arrangement in use at present, a slightly different scheme for getting the power out of the oscillator, as shown in Figure 3, is believed to be better and avoids the band filter. In Figure 3 the oscillator tube has enough cathode bias to bring the operating point on the steepest part of the grid voltage-plate current characteristic curve in the absence of oscillations. The oscillator tube should be one requiring a large bias for cut-off while the following tube should be one requiring less bias for cut-off so that with no

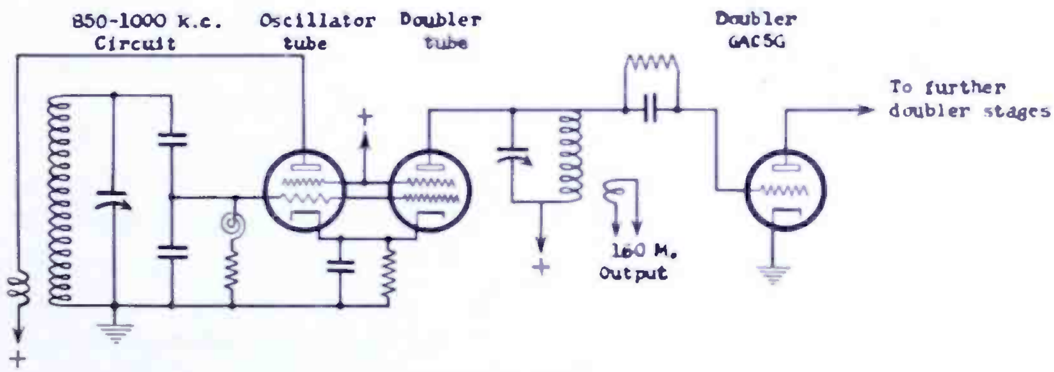


Fig. 3

oscillations the plate current of the second tube would be just cut off by the normal bias of the oscillator. Then with even very feeble oscillations the following tube would act as an efficient doubler while with stronger oscillations, that would develop more bias on both tubes by way of the grid leak, the harmonic output of the following tube would be still further increased. The following tube should of course be well screened to prevent reaction of its plate circuit upon the oscillator.

CONCLUSIONS

To recapitulate, in order to obtain the greatest inherent stability:

1. Make the fundamental frequency as low as possible.

2. Make the "Q" of the coil as large as possible at the fundamental frequency. This means that the coil should be as large physically as there is room for within the shield can, subject to clearance of at least half a diameter, as well as that the coil design should be good in other respects.

3. Use the loosest couplings between the tuned circuit and the tube that will give the required output, and use a low enough bias resistor so that the effective transconductance in the oscillating condition is not seriously reduced.

4. Choose for the oscillator tube one which has a high ratio of transconductance to capacitance fluctuations when operating at the required level.

5. Keeping the oscillation strength constant, vary the ratio between the grid and plate couplings. The best ratio depends on the ratio between the capacitance variations of the grid and plate.

After having obtained a good inherent stability, any or all the tricks known to the trade may be added. Some of these are: Temperature compensation in the tuned circuit, or at least arranging this circuit where it will not be heated by the tube or other parts of the transmitter, supporting the tuned circuit on a single rigid member to avoid bending and vibration of its parts, reducing the power taken from the oscillator as much as possible and preferably taking output at a harmonic frequency, supplying screen voltage from a voltage divider whose two portions have resistances chosen to form the combination that best compensates for variations in supply voltage, and stabilizing the supply voltage.

By starting with an oscillator of high inherent stability and then adding the refinements to it, sufficient stability may be obtained for many purposes, even including oscillator keying for output at the highest frequencies used for long-distance communication.

Appendix

Let $x = \dfrac{C_p}{C_g}$ and $y = \dfrac{M_p}{M_g}$. Then expression 1 may be written in the form $\dfrac{f}{2} \dfrac{\omega^2 M_g M_p}{L} \sqrt{C_g C_p}\ (1/xy + xy)$. For any given value of x the minimum value of $(1/xy + xy)$ that may be obtained by varying y is 2, and this minimum occurs when $y = 1/x$, that is, $\dfrac{M_p}{M_g} = \dfrac{C_g}{C_p}$.

Furthermore, in order for oscillations to occur, a given grid voltage must cause at least an equal voltage to be fed back, whence the condition for oscillation is approximately $\omega^2 M_g M_p \geq r/g$. Substituting this value of $\omega^2 M_g M_p$ and the minimum value of $(1/xy + xy)$ already obtained into the expression above, expression 2 results.

DESIGN OF SUPERHETERODYNE INTERMEDIATE-FREQUENCY CIRCUITS

By

F. E. Spaulding, Jr.

Engineering Department, Radiomarine Corporation of America

WIDELY differing selectivity requirements of the many present-day receiver applications necessitate careful study of the intermediate-frequency circuits.

The methods by which these requirements may be satisfied must be analyzed with a view to their effect on other important receiver characteristics.

The following discussion of intermediate frequency circuit design is based to a considerable extent on the requirements established by marine telegraph and telephone service.

Consideration of the conditions under which a receiver is to be used will usually indicate the approximate amount of selectivity to be desired. The ten-kilocycle separation of assigned frequencies in the American broadcast band, for example, immediately points to the desirability of attenuating the adjacent carrier frequencies. Marine radiotelephone receivers to operate with coastal and harbor stations in the 2500-2600 kc range must be free from adjacent carrier interference in spite of 8-kc carrier separation. Marine direction-finding receivers require excellent selectivity in the 285-315 kc beacon band, as the transmitting frequencies are only 2 kc apart. Receivers having less critical fidelity requirements, as for example, commercial receivers for telegraph reception, can make use of the extremely sharp characteristic afforded by crystal i-f filters and thereby extend their useful range to weak signals which would otherwise be obscured by noise and interference.

Broadcast receivers incorporating push-button tuning should have broad-nosed curves, even if some sacrifice need be made in adjacent carrier attenuation in order that reasonable amounts of oscillator-frequency drift may occur without seriously impairing fidelity.

When good fidelity is of prime importance, use may be made of triple-tuned i-f transformers to obtain broad-nosed and yet steep-sided characteristics which will pass more of the high audio tones and yet provide good attenuation to adjacent carriers, or possibly some system of variable coupling may be employed. It is also possible to design the audio system to compensate for highs lost in the i-f system.

After obtaining a rough idea of the overall selectivity to be desired, thought should be given to its distribution between r-f and i-f stages. As the discrimination of tuned circuits for a specified number of kilocycles off resonance decreases with increasing frequency, it is apparent that the usual lower-frequency i-f system will contribute the greater portion of the overall selectivity. At r-f circuit frequencies of about two megacycles and above, the i-f selectivity practically constitutes the overall selectivity.

Factors Governing Choice of Intermediate Frequency

The intermediate frequency to be used should be chosen only after having given careful consideration to the following characteristics:

1. The degree of overall selectivity desired.
2. Image rejection.
3. I-f rejection.
4. Tuning range.
5. Harmonic interference.
6. Number of tuned circuits required.
7. Overall sensitivity.

Maximum possible sharpness results when crystal filters are incorporated in the i-f system. These require careful adjustment by operating personnel for satisfactory results. Band widths in the order of a few hundred cycles may be obtained even when the i-f system frequency is fairly high, as 600-800 kc. Such devices are usually restricted to telegraph service because of the narrow pass band. If sharpness approaching this amount is to be obtained without crystal filters, a very low intermediate frequency becomes imperative to avoid a prohibitive number of tuned circuits. Frequencies in the order of 50 to 200 kc are used in such instances. The recent development of composite band pass filters for i-f systems has shown it possible to very nearly reach the ideal characteristic of a flat-topped curve with nearly vertical sides.

The intermediate frequency chosen has a direct bearing on the image-rejection ratio. Interference from the image frequency, being the outstanding undesired response in superheterodynes, must be suppressed so as to become negligible. Other spurious responses are usually of much smaller magnitude and do not usually give trouble. Suitable image rejection may be obtained in several ways. One is to provide good r-f selectivity, using loosely coupled transformers with sacrifice in r-f stage gain, followed by relatively low frequency i-f transformers. This means that while the image frequency is relatively close to the fundamental, the r-f circuit selectivity is made sufficiently great to suppress images. A better way is to use a high intermediate frequency, thus removing the image farther from the fundamental so that even poor r-f selectivity will give adequate rejection. Some suitable compromise between these two extremes can usually be found. The latter method permits choice of r-f gain through coupling adjustment without too seriously impairing image ratio. This is desirable from the standpoint of signal-to-noise ratio, as high r-f gain assists in maintaining an optimum ratio. If a gain of five times or more is maintained in the r-f circuit, the total receiver internal noise is comprised mainly of the first grid-circuit noise which is materially less than the noise generated in the mixer tube.

At radio frequencies above 5 Mc, the r-f selectivity is usually so small as to necessitate intermediate frequencies in the order of 1500-2000 kc for adequate image suppression in commercial receivers.

Special r-f circuits have been devised to provide very great attenuation to image frequencies. These may be satisfactorily applied in certain instances.

It has occasionally been found necessary to use two intermediate frequencies in the same receiver, a high intermediate frequency (1500-5000 kc) in the first i.f. to provide good image rejection, followed by a fixed frequency oscillator converting this to a low frequency (50-200 kc) second i-f section to provide a sharp selectivity characteristic.

When an incoming carrier happens to be at the intermediate frequency, it will cause interference if it is able to reach the mixer grid without being sufficiently attenuated. By attention to r-f circuit design, shielding, and use of i-f trap circuits, it is possible to keep this interference at a low level. It is obvious, however, that a judicious choice of the intermediate frequency will avoid use of any wavelengths in frequent use by powerful shore or ship transmitters. Marine telegraph frequencies in frequent use are 285-315 kc, 375 kc, 425 kc, 500 kc, and some others.

It is not usually feasible to attempt to include the intermediate frequency within the tuning range of a receiver due to the production of spurious responses in this region.

Interference may result from harmonics of both the first and second detector beating with incoming signals to produce undesired responses. In order to minimize the former, it is advisable to restrict the r-f stage gain to that amount, giving adequate reduction of internal noise, as previously mentioned. Attention to oscillator-circuit design is also important.

Interference due to harmonics of the i-f heterodyning with the desired signals may be minimized by careful shielding of the second detector from the input circuits and attention to diode load filtering, but may sometimes be avoided by choice of intermediate frequency. If, for example, a 450-kc i.f. is used, a whistle may be heard in some instances while tuning through the vicinity of 900 kc, the second harmonic, and occasionally 1350 kc, the third harmonic. The choice of 175 kc, standard practice several years ago, minimized this type of interference in broadcast receivers, as the second and third harmonics fall below the broadcast band while those of higher order are correspondingly weaker.

Receivers for service above 20 Mc should have higher intermediate frequencies than those for broadcast use, not only for improved image suppression, but also for ease of tuning and alignment. Attempting to tune in signals of this order with a 455-kc intermediate frequency becomes extremely difficult, even with high-ratio vernier dials and very stable oscillators. Under such conditions, the percentage difference between signal and oscillator frequencies becomes very small and even normally stable oscillators show a tendency to "lock in" and to be otherwise influenced by the signal circuit during the tuning process. Even small amounts of coupling between oscillator and signal circuits, whether through space-charge coupling within the tube or from other causes, will sometimes produce appreciable amounts of current at oscillator frequency to flow in the latter and introduce harmful affects. Use of 1600 kc or higher will prevent any of these difficulties.

From the standpoint of gain and sensitivity, it is not generally necessary to use more than a single i-f stage for frequencies below 500 kc. The maximum stable tube gain, with the tubes commercially available, may be in the order of 400 for tubes with low grid-plate capacitance.

To improve selectivity, however, it becomes necessary to increase the number of tuned circuits. This, in turn, makes additional tubes necessary to maintain the desired overall sensitivity. When intermediate frequencies above 500 kc are used, the additional stages become necessary as both the selectivity and the maximum stable gain per stage decrease with increasing frequency. When several stages are used in cascade, the overall voltage gain is equal to the product of the gains of individual stages, and similarly the overall selectivity may be approximated by multiplying the attenuations of the various stages at corresponding frequencies off resonance. This is assuming, of course, that regeneration is kept at a minimum by sufficient by-passing and shielding.

Diode Transformer Considerations

It must be remembered that the transformer feeding the diode is not capable of providing much more than fifty per cent of the selectivity it could if it were not loaded by the diode circuit. In some instances, with low diode load impedances, this may not even exceed thirty per cent. This is particularly true when a-v-c power must be supplied from this transformer. It has been found best to design this stage to meet a-v-c requirements, and thus avoid overload difficulties, rather than to strive for maximum selectivity. This means fairly close coupling in the diode transformer and operation of the last i-f tube with little or no a.v.c. Since 30 to 40 volts of automatic volume control is required to cut off tubes with remote characteristics, there will be ample for audio requirements. It may even be necessary to tap down on the diode load to feed the following audio stage.

The amount of direct current developed across the diode load will vary from sixty to eighty-five per cent of the peak carrier voltage impressed upon the diode, depending on the circuit constants.

The peak audio voltage of a fifty-per-cent modulated carrier will be fifty per cent of the d-c voltage obtained. One other point in connection with a discussion of the diode circuit is that good modulation capability, the ability to detect signals faithfully which are deeply modulated, demands that the ratio of a-c to d-c diode load impedance be as near unity as possible. This generally necessitates as high a value of a-v-c filter resistor as the time constant permits, and prevents the use of diode load resistance values of much over 500,000 ohms. For the purpose of measurements on a diode transformer, the loading may be closely approximated by placing a resistance equal to one-half of the normal diode load resistor across the secondary.

Full selectivity of the last transformer may be obtained in the case of infinite impedance detectors and other biased types which do not place power requirements upon the input circuits. Such detectors do not generally provide ready means of obtaining automatic volume control, however, and some have other disadvantages such as low efficiency, poor modulation capability, together with overloading and distortion which limit their usefulness.

General Coupled Tuned-Circuit Theory

A study of the theory of coupled tuned circuits provides expressions which permit computation of their performance in conventional circuits. The fundamental expression for the voltage gain of an r-f pentode working into a finite load is:

Voltage Gain $= g_m Z_1$.

where $g_m =$ Grid-plate transconductance (in mhos) of the tube.

and $Z_1 =$ Load impedance in ohms.

When the plate load is a single tuned circuit, at resonance, the load impedance becomes the resonant impedance ωLQ of the circuit

$$\text{Voltage Gain} = g_m \omega LQ$$

where $\omega = 2\pi f$,

$f =$ Megacycles

$L =$ Microhenries

$g_m =$ Mhos

If the pentode is feeding into a single tuned transformer with secondary tuning, the overall gain at resonance becomes approximately as follows:

$$\text{Voltage Gain} = g_m \omega M Q_s$$

where $g_m =$ Grid-plate transconductance (in mhos) of the tube.

For the case of a double tuned transformer at resonance, the following expression is a close approximation.

$$\text{Voltage Gain} = g_m k \frac{\omega\sqrt{L_s L_p}}{k^2 + \dfrac{1}{Q_p Q_s}}$$

where $g_m =$ Grid-plate transconductance (in mhos) of tube.

L_s and L_p are secondary and primary inductance in microhenries.

Q_s and Q_p are secondary and primary Figures of Merit, equal to $\omega L/R$

$k =$ Coefficient of coupling

k reaches a maximum of one at critical coupling and is equal to $\dfrac{M}{\sqrt{L_p L_s}}$ for less than critical coupling.

An approximate formula for computing selectivity has been shown to be as follows:

$$\Delta f = \frac{f_r}{Q}\sqrt{\left(\frac{E_r}{E}\right)^2 - 1}$$

The band width Δf is that corresponding to the ratio of resonant to off-resonant output voltage, and f_r is the resonant frequency.

When shunted by the plate resistance for a pentode, it is modified slightly as follows:

$$\Delta f = f_r\left(\frac{1}{Q}+\frac{\omega L}{r_p}\right)\sqrt{\left(\frac{E_r}{E}\right)^2-1}$$

For two loosely coupled tuned circuits, it becomes

$$\Delta f = \frac{f_r}{\sqrt{Q_1Q_2}}\sqrt{\frac{E_r}{E}-1}$$

The preceding expressions point out two things. First, that the stage gain is a function of tuned-circuit impedance $Q\omega L$ and that for a given Q the highest L/C ratio is desired for maximum gain, and secondly, that selectivity at a given frequency is a function of Q alone.

Attempts to increase Q and L/C ultimately reach practical limits. Use of Litz wire and iron cores permit improvement in Q for frequencies in the order of 200 to 500 kc, over solid wire and air-core coils. Large-diameter forms, space-wound coils, attention to wire size and choice of optimum length to diameter ratios help improve Q in the higher frequency ranges.

The inductance cannot be increased beyond the point at which it resonates to the desired frequency with the minimum circuit capacitance including the tube, coil distributed capacity, trimmer, and lead capacities, etc. In most instances, it cannot be increased to this value because of instability brought about by variations in the preceding items.

Effect of Vacuum-Tube Admittance Variations

Of these several items, the most serious and usually the limiting factor is the variation in the input capacitance offered by the tube, since this quantity may vary appreciably with changes in the electrode potentials.

Much investigation in recent years has disclosed the fact that relatively large changes in tube input admittances may occur under certain conditions. These may be both capacity and resistance variations, and may result in harmful effects upon the associated circuits.

The tube input capacitance is the sum of three components. The first is the cold input capacitance. This is the direct capacitance from the signal grid to all other grounded electrodes, such as cathode, screen, suppressor, etc., and is the capacitance which appears when the tube is not operating. The second component appears when the tube is in normal operation and results from the presence of a space charge between the grid and cathode. A capacitive component of grid current is induced by the presence of a stream of moving electrons. (In converter tubes, this electron stream may be modulated by the oscillator elements and hence produce grid current at oscillator frequency in the signal grid circuit.) The third component of the input capacitance is the grid-to-plate capacitance. The current flowing through this latter capacitance is affected by the magnitude of $C_{g\text{-}p}$ and $C_{p\text{-}k}$ and also by the grid-plate transconductance (g_m) and by both magnitude and phase of the plate load. For the case of a resistive plate load, as for example, a tuned circuit at resonance, the input admittance is a pure capacitance. If a reactive plate load is present, such as a tuned circuit above or below resonance, a resistive component is introduced which will be negative for inductive loads and positive for capacitive loads. This checks with general oscillator theory, which shows that an inductive plate load furnishes the negative resistance in the grid circuit necessary to sustain oscillations. (This means a parallel resonant circuit operated above its resonant frequency).

The total input capacitance is independent of frequency. It may, however, be of such magnitude as to allow a prohibitive amount of current at high frequencies.

The resistive component of the input admittance varies inversely with frequency, and hence becomes a severe load at high frequencies. It is one of the major factors determining the maximum usable frequency of tubes. This component appears as an appreciable conductance in parallel with the existing grid-circuit impedance.

Recent investigation at the higher frequencies indicates that input loading in these ranges is influenced appreciably by electron transit-time and cathode-lead inductance effects.

Terman writes the following expressions for input capacitance and resistance:

$$\text{Input Capacitance} = C_g = C_{g\text{-}f} + C_{g\text{-}p}(1 + A \cos\theta)$$

where $C_{g\text{-}f}$ = Grid-to-cathode capacitance.

$C_{g\text{-}p}$ = Grid-to-plate capacitance.

$$A = \frac{E \text{ Load}}{E \text{ Input}} = \text{Tube gain.}$$

and θ = Angle of load impedance

For purely inductive load, $\theta = +90°$

For purely resistive load, $\theta = 0$

For purely capacitive load, $\theta = -90°$

$$\text{Input resistance} = R_g = \frac{-\dfrac{1}{\omega C_{g\text{-}p}}}{A \operatorname{Sin} \theta}$$

where $\omega = 2\pi f$

A study of these two expressions shows that when the plate load consists of a tuned circuit and the frequency of the applied voltage is varied through resonance, there is an increase of input capacitance at resonance. At frequencies above resonance, a negative resistance is reflected into the grid circuit, while below resonance a positive resistance is indicated.

It is possible then to produce two undesirable effects. First, the grid circuit tunes to a different frequency at resonance than off resonance. Second, the symmetry of the resonance curve is destroyed, as, with negative resistance above resonance the effective grid circuit Q is increased, while below resonance the positive resistance lowers the circuit Q and broadens the curve.

The above effects may become appreciable under certain conditions, as may be seen from an actual case which developed in practice using commercially available tubes and components.

A 6K7 was used with 455-kc air-core i-f transformers whose inductance was 1.8 microhenries and $Q = 110$. It was found that a change of minus 8.5 volts in the grid bias resulted in a shift of nearly 5 kc in the resonant frequency of the grid circuit. At the same time, it was observed that when the trimmer capacitance of the plate transformer was increased, in other words the load was made inductive, oscillation started. When this circuit was returned to resonance at the original frequency, the oscillation ceased. Reducing the magnitude of the plate load impedance by lowering L to 1 microhenry and increasing C correspondingly corrected both of the undesirable characteristics. Materially reducing the transconductance of the tube accomplished the same result.

Freeman's investigation of the capacitance changes indicates a similar degree of magnitude. He also shows how choice of a proper value of unbypassed-cathode resistor provides compensation for these capacitance variations.

Insertion of approximately 20 ohms of cathode resistance, unbypassed, and returning the grid trimmer to cathode instead of ground were found, in the example previously cited, to result in negligible shift of grid-circuit resonance.

A similar means is used to compensate for grid-circuit loading at higher frequencies.

Another device which accomplishes this result is that of tapping down the grid on its resonant circuit. This practice has several advantages, and is finding increased use where more than one i-f stage is needed. Any capacitance or resistance reflected from the plate circuit is no longer impressed directly across the entire grid-tuned circuit and, by suitable choice of the tap, may be made substantially negligible. In addition to this, it becomes possible to lower the overall i-f system gain and still maintain any degree of coupling in the transformers. This is helpful since it frequently happens that a system with several

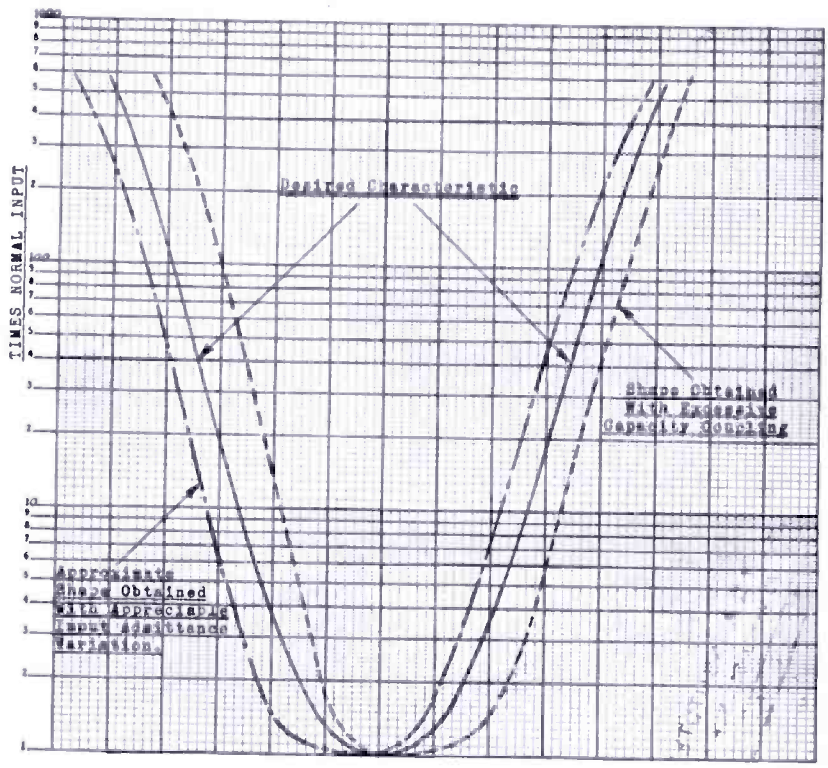


Fig. 1

stages, necessary to provide sufficient selectivity, will furnish more than the desired amount of gain. Other considerations, such as automatic volume control and overloading, make it desirable to operate the tubes near their maximum transconductance, in other words near rated plate current.

In tapping down on a tuned circuit as mentioned, the shape of its resonance curve is not altered, and remains a function of its Q.

Another possible cause of asymmetry in i-f curve shape may be due to an excessive amount of capacitance coupling between plate and grid windings of the same transformer.

It has been general practice to arrange the position of coils and condensers so as to make use of the inductive coupling only while holding the capacitive component to a minimum. While it is true that a small amount may be tolerated, a greater amount may become troublesome. An instance was found wherein a distorted curve was obtained, as shown in the accompanying diagram. Additional shielding and filtering or bypassing did not help, and reduction of tube gain was of no avail. The asymmetry was still present even though only one tube and its output transformer were used, and it was further found that use of a tapped transformer did not improve the shape. It was observed that the side of the curve which was broadened was for frequencies above resonance. This indicated that input admittance variations could not be the cause, since their effect is to sharpen the curve above resonance frequency.

The clue which pointed to capacitive coupling was provided when it was discovered that moving the grid and plate leads closer together aggravated the condition. It was then observed that the spacing between the two condensers was small enough to result in a direct capacitance of 5 micro-microfarads between them. Changing the position of one condenser and attention to grid and plate lead separation was found to restore the correct symmetry to the selectivity characteristic. Curves are shown which illustrate both this effect and the effect of input-admittance variation.

The foregoing discussion has been an attempt to bring together many of the factors requiring some degree of consideration by the designer so as to present a broad view of the general subject.

No attempt has been made to treat the economic aspects of the problem as they are many and varied and are beyond the scope of this article.

Bibliography

1. I-F Selectivity in Receivers for Commercial Radio Services, J. B. Moore and H. A. Moore, *RCA Review*, January, 1940.

2. Spurious Responses in Superheterodynes, H. K. Morgan, *Proc. I. R. E.*, October, 1935.

3. Image Suppression in Superheterodyne Receivers, H. A. Wheeler, *Proc. I. R. E.*, June, 1935.

4. Recent Developments in Diversity Receiving Equipment, J. B. Moore, *RCA Review*, July, 1937.

5. Diode Detector Design, H. A. Wheeler, *Proc. I. R. E.*, June, 1938.

6. Radio Engineering Principles, F. E. Terman, *McGraw-Hill Publishing Co.*

7. Intermediate Frequency Transformer Design, F. H. Sheer, *Proc. I. R. E.*, December, 1935.

8. Input Loading of Receiving Tubes at R. F., *Application Note No. 101*, RCA Radiotron Company, Harrison, N. J., January, 1939.

9. Input Resistance of R. F. Receiving Tubes, G. Grammer, *Q. S. T.*, May, 1939.

10. Use of Feedback to Compensate for Vacuum-Tube Input Capacitance Variations With Bias., R. L. Freeman, *Proc. I. R. E.*, November, 1938.

DEVELOPMENT AND PRODUCTION OF THE NEW MINIATURE BATTERY TUBES

By

Newell R. Smith and Allen H. Schooley

Research and Engineering Department, RCA Manufacturing Company, Inc., Harrison, N. J.

Summary—A new line of miniature battery tubes, including a converter, a radio frequency amplifier, a diode-pentode and a power output pentode has been made available. These new tubes are designed to operate efficiently on a 45-volt "B" supply with the filament operating directly from a single dry cell. The tubes are about two inches long and less than three quarters of an inch in diameter.

A feature of this new line of tubes is a decrease in size without an increase in cost. This is accomplished by using a simplified envelope design which in the majority of cases permits standard size electrodes to be assembled using conventional manufacturing procedure. A new button stem in which the external leads serve as base connections contributes materially to the reduction in tube dimensions.

The small size and efficient operating characteristics of the new miniature tubes make them especially applicable to compact communication equipment as well as portable broadcast receivers. Also they may find application in special fields such as hearing aids, meteorological service, or other places where size and weight are a consideration.

THE trend in radio receiver design, during the past two years has been toward small, low-priced models. This, together with the increasing popularity of portable, self-contained, battery receivers, has created a demand for smaller and less expensive radio tubes. The new miniature tubes, with their simplified construction and great reduction in size, provide a logical answer to this demand.

Although several types of small tubes are now available, these, in general, use very small parts with close electrode spacing requiring slow, careful, assembly by highly skilled operators and special fabricating processes. Such tubes have application wherever the requirement for special characteristics warrant their increased cost. However, the inherently higher cost of producing tubes of this character is sufficient to prohibit their general use in popular-priced broadcast receivers. Consequently, the problem of obtaining a small, low-cost tube depends for its solution upon the selection of a satisfactory design of simplified construction which will not only offer the desired reduction in size, but also lend itself to usual manufacturing operations.

In the design of this new line of tubes every effort has been directed toward producing the smallest glass tubes which can feasibly be manufactured by present high-speed methods of fabrication. The variations required in manufacturing technique have, therefore, been carefully restricted to procedures which have been proven by prior experience.

In so far as possible, all parts and materials which serve no functional purpose in the completed tube have been eliminated.

The new miniature line consists of four tubes, the RCA-1R5, converter; RCA-1T4, radio-frequency pentode; RCA-1S5, diode-pentode; and RCA-1S4, power-output pentode; and provides a complete complement for receiver design. All of the tubes operate efficiently from a small, 45-volt "B" battery and their filaments are designed for operation at 50 milliamperes (except the RCA-1S4 which requires 100 milliamperes) supplied directly from a single dry-cell.

The reduction in tube dimensions accomplished by the new design is illustrated by the photograph, Figure 1, in which a miniature tube is

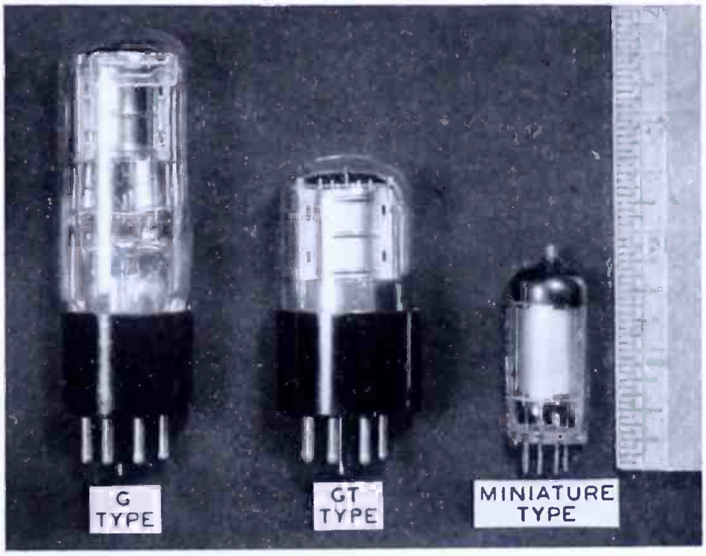


Fig. 1—New Miniature Tube compared with smallest present equivalent types.

compared with a G-type and a GT-type having similar characteristics. It is interesting to note that the miniature tube is about two inches long and less than ¾ inch in diameter and only displaces about 20 per cent of the receiver space required by the GT-type equivalent. This large reduction in volume should recommend these new tubes for those applications where compactness is essential. They should be especially desirable in the design of portable broadcast receivers, pocket receivers and police equipment. They may prove useful in meteorological work, hearing-aid and other special applications where size, weight and cost must be considered.

The manner in which the reduction in size has been effected together with the various special features of design and manufacturing technique employed in the production of the new miniature tubes are discussed in the ensuing paragraphs.

Stem Design

In Figure 2 (top) the various parts and an assembled unit of a conventional-type stem are shown. In this stem, the electrode leads are located in a single plane through the center of the preformed flare and the vacuum seal is made by pinch-pressing the molten glass onto the short sections of special seal wire. The particular stem shown in Figure 2 (top) is representative of one of the shortest stems in general use. In this case the length of the glass flare from the sealing line to the top of the press is approximately 14 millimeters (0.55 inch). To

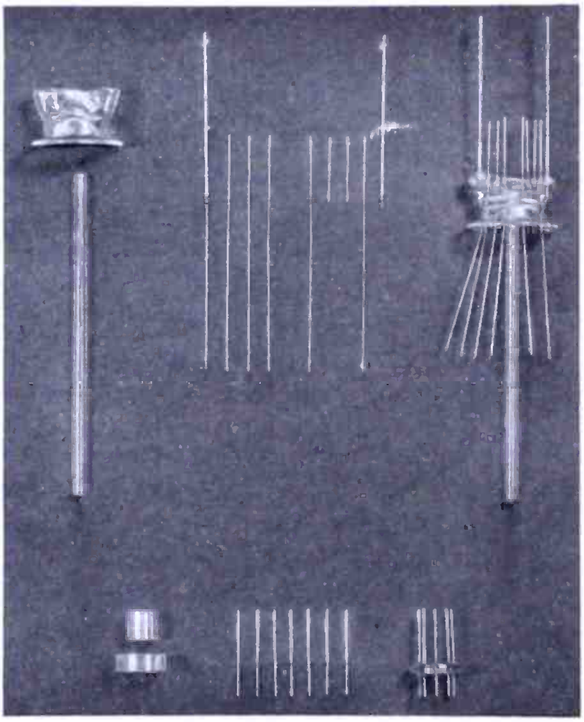


Fig. 2—(Top)—Parts and an assembled unit of a conventional stem. (Bottom)—Miniature button stem and component parts.

this value must be added sufficient length to provide for trimming and forming the leads and making the welds of the leads to the tube electrodes. This additional length of approximately 10 millimeters (0.39 inch) makes the total distance from the point of sealing to the lower edge of the mount approximately 24 millimeters (0.94 inch). It will be evident that this long length is non-essential to the electrical characteristics, but results from the mechanical limitations of this type of stem design. The stem length has been greatly reduced in the design of the miniature button stem by making the sealing position to the enclosure coincident with the vacuum seal line as shown in Figure 2 (bottom). The distance from the sealing line to the lower edge of the mount has by this construction been reduced by 65 per cent and results in shortening the overall length of the finished tube by 14 to 15 millimeters (.55 to .59 inch).

The use of the "button-type" stem also contributes to a reduction in tube diameter. The conventional-type stem shown in Figure 2 (top) requires a flange diameter which exceeds the width of the stem press. In the GT-type tube of Figure 3, the flange diameter is approximately 1 inch and a bulb diameter of 1⅛ inches is, therefore, required for successful assembly. In comparison, the miniature type has seven leads located on a ⅜ inch circle in a glass button only 6/10 inch in diameter. This stem can be sealed into a bulb having an outside dimension of approximately 11/16 inch.

Although the diameter of the tube is reduced, the lead spacing has

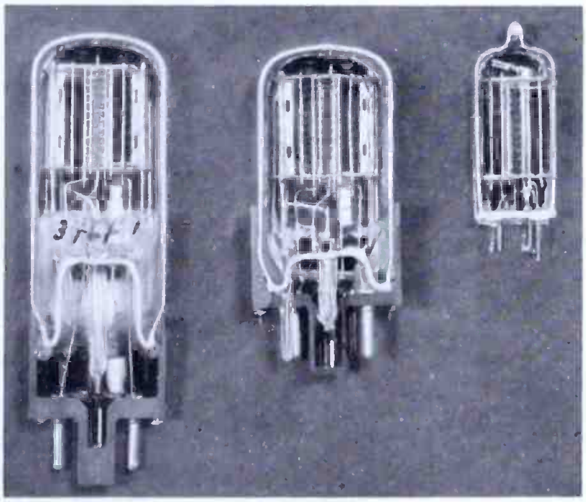


Fig. 3—Cut-away view of G-Type (left), GT-Type (center), and Miniature Type (right).

been improved by use of the button-type stem. For example: seven leads sealed into a conventional-type stem cannot have an average spacing much in excess of 0.110 inch between leads. In the miniature-tube design seven leads are spaced 45° apart on a circle 0.375 inch in diameter. Thus, the leads are spaced 0.147 inch apart, an increase of 33 per cent more than can be obtained by use of the larger conventional-type stem. From this explanation, it is evident that in those instances where three or more leads are required, a much smaller diameter stem of the button type can be designed to have a greater lead spacing than would be possible with the conventional flat-press stem. The wider spacing reduces the possibility of electrical leakage and minimizes capacitance coupling between leads.

It is evident that in addition to the reduction in size realized through the use of the miniature button stem that a considerable saving in materials has also been achieved by departing from the conventional pinch-press stem. For example: only about 1/5 of the amount of glass used by the small conventional type stems is required for the miniature stem. The glass does not need to be preformed, but is cut directly from tubing of the correct diameters to the short lengths required. All leads used in the miniature stem are identical regardless of tube type. Only three types of parts are required to produce the miniature stems as compared to six types of parts for the conventional stem. This standardization of parts is a decided advantage in manufacture.

Since the miniature stem was designed for fabrication on the same

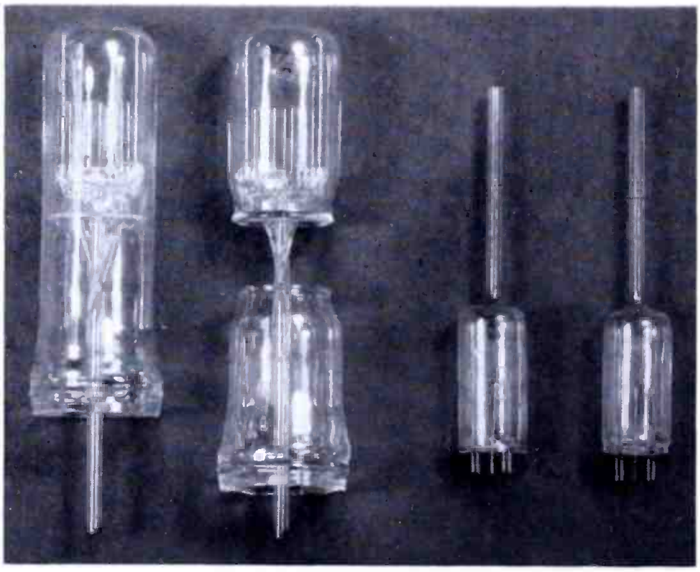


Fig. 4—(Left Half)—Conventional type tube before and after sealing, with discarded collet.
(Right Half)—Miniature-type tube before and after sealing, with no collet.

equipment and under the same technique employed for the production of metal-tube button stems, only slight modifications of existing equipment were required.

Bulb

The miniature bulb (shown as part of Figure 4) is molded from glass tubing and has no collet. This method of fabrication is economical in that no glass is wasted and close tolerances in the bulb and the stem combine to make sealing a relatively simple operation.

The miniature stem has no tubulation and it is, therefore, necessary to provide an exhaust tubing in the bulb. This is done by reviving the bulb seal-off which was used by the incandescent lamp and radio tube industries for many years.

MOUNT

In the past, the designers of small tubes have concentrated upon obtaining a reduction in mount size by close electrode spacings and very small parts, the primary reason being to reduce inter-electrode capacity and lead inductance in order to improve high frequency operation. Tubes of this design are difficult and slow to assemble in the factory and must necessarily be expensive to produce. Examination of

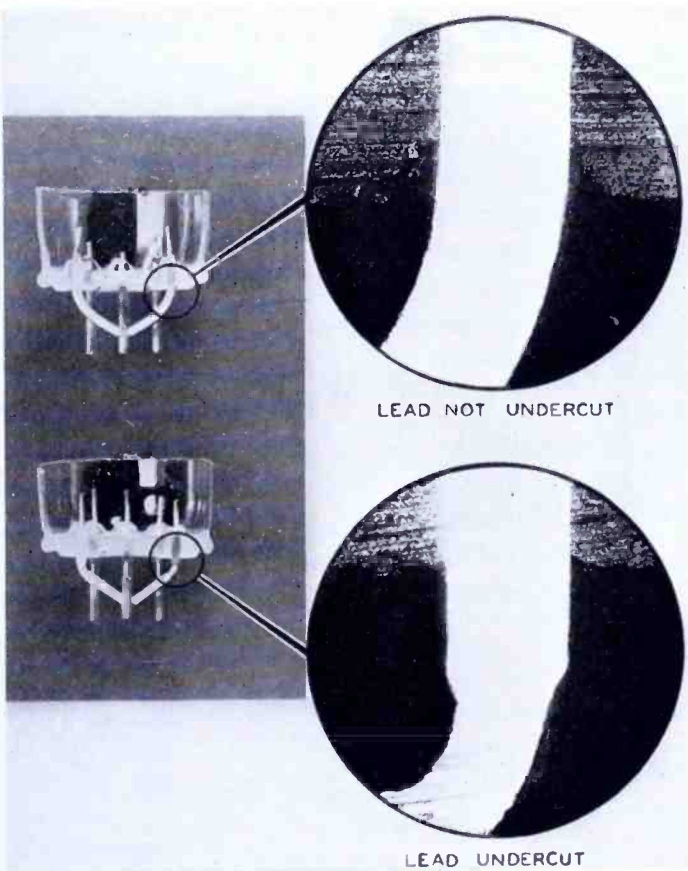


Fig. 5—Photographs showing how undercutting leads of miniature stems protects the glass seal under extreme conditions of bending.

the three tubes shown in Figure 3 will disclose that the mounts in all three types are identical for mount parts and electrode spacings except, in the case of the miniature, that the anode has been changed to a simple cylindrical sleeve so as to permit insertion into the bulb. In the miniature tube, the getter has been mounted above the top mica to eliminate any possibility of leakage across the stem due to getter. Maintenance of mount dimensions identical to those used on other types of receiving tubes, provides standardization of parts and materials and facilitates factory operations by permitting assembly on common fixtures and at usual mounting speeds.

Baseless Construction

From a comparison of the tubes shown in Figure 3, it will be apparent that the reduction in size of the miniature tubes has been obtained almost entirely from the stem and enclosure, either by redesign or elimination of parts. Some of these changes may be considered as elimination of safeguards previously considered essential to tube construction. For instance, the tubulated bulb used in early incandescent lamp and radio tube practice has been incorporated in the miniature design. The bakelite base with is rigid pins for socket contacts separated by intermediate flexible connectors from the vacuum seal has been replaced by short pins sealed directly into the glass of the stem. Although laboratory tests have indicated that the glass seal is sufficiently strong to withstand the normal pressures which will be exerted on the base pins in various applications, it has been deemed advisable until experience can be obtained from the field to incorporate an additional safety factor of undercut pins as illustrated in Figure 5. This device consists of reducing the diameter of a short lengh of the pin at some point below the glass so that in case the lead is accidentally deflected from its normal position the bending will be localized in the weakened part of the pin and the vacuum seal will not be affected. If a solid lead is bent, the fulcrum point of the bend is located at the edge of the glass seal as is shown in Figure 5 (top). This causes chipping of the glass and may result in a ruptured seal. An undercut pin which is similarly bent (Figure 5 bottom) has absorbed the distortion entirely in the weakened section and has not disturbed the vacuum seal. The advantage of this artifice is that the strength of the contact pin can be reduced to well within the safe limits permitted by the strength of the glass seal. The amount of this lead under-cutting can be varied or eliminated entirely without altering the contact diameter of the pins.

This development provides a new line of small tubes having very efficient operating characteristics at low battery voltages. These tubes are designed especially for use in compact, lightweight, portable equipment.

The co-operative efforts of the many engineers involved in this development are recognized and appreciated, although it is impractical to acknowledge the contributions individually.

RCA MEN HONORED

FORTY-SEVEN engineers associated with RCA laboratories were among the 572 American industrial engineers and scientists recently given awards as "Modern Pioneers on American Frontiers of Industry" by the National Association of Manufacturers in connection with a nation-wide observance of the 150th Anniversary of the Founding of the American Patent System. Special national awards were given to nineteen of those receiving honors. Dr. V. K. Zworykin of the RCA Laboratories at Camden, N. J., was the recipient of one of these.

The Association's Committee on Awards was comprised of Karl T. Compton, President, Massachusetts Institute of Technology; Forest R. Moulton, Permanent Secretary, American Association for the Advancement of Science, George B. Pegram, Dean, Graduate Faculties of Columbia University, and John T. Tate, Dean, College of Science, Literature and the Arts of the University of Minnesota.

The names of the RCA engineers who received awards are:

Randall Clarence Ballard
Max Carter Batsel
Alda Vernon Bedford
George Lisle Beers
Harold H. Beverage
Rene Albert Braden
George Harold Brown
Irving F. Byrnes
Wendell LaVerne Carlson
Philip S. Carter
Lewis Mason Clement
Murray G. Crosby
Glenn Leslie Dimmick
James L. Finch
Dudley E. Foster
Clarence Weston Hansell
O. B. Hanson
Ralph Shera Holmes
Harley A. Iams
Ray David Kell
Edward Washburn Kellogg
Winfield Rudolph Koch
Fred H. Kroger
E. Anthony Lederer
Humboldt W. Leverenz
Nils Erik Lindenblad
Loris E. Mitchell
Gerrard Mountjoy
Harry Ferdinand Olson
Richard T. Orth
Harold O. Peterson
Walter Van B. Roberts
George M. Rose, Jr.
Bernard Salzberg
Otto H. Schade
Stuart W. Seeley
Terry M. Shrader
Browder J. Thompson
Harry C. Thompson
William Arthur Tolson
George L. Usselman
Arthur William Vance
Arthur F. Van Dyck
Julius Weinberger
Irving Wolff
Charles Jacob Young
Vladimir Kosma Zworykin

ERRATA

The Service Range of Frequency Modulation

BY

MURRAY G. CROSBY

RCA Review Vol. IV, No. 3, January 1940

Sixth line from bottom of Page 350, change the constant "2.21" to "3550".

Delete the last sentence of the third paragraph reading: "However, for this type of noise, etc.," on Page 369.

Replace Figures 7, 8, 9, and 10 with the following corrected diagrams.

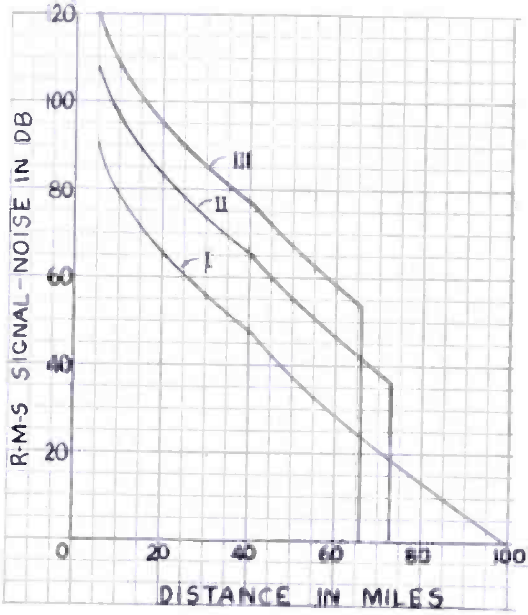


Fig. 7

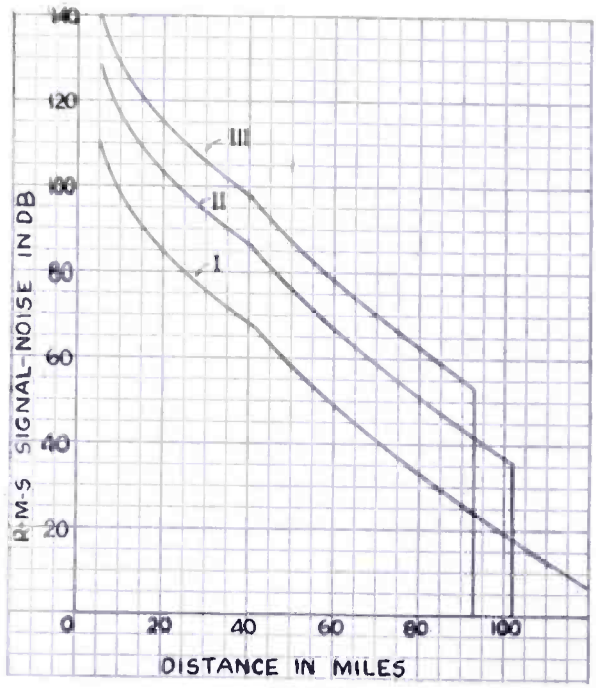


Fig. 8

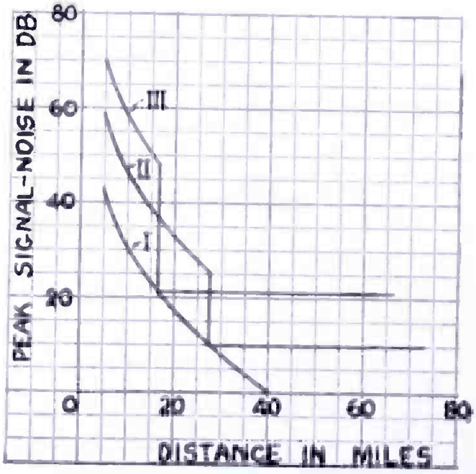


Fig. 9

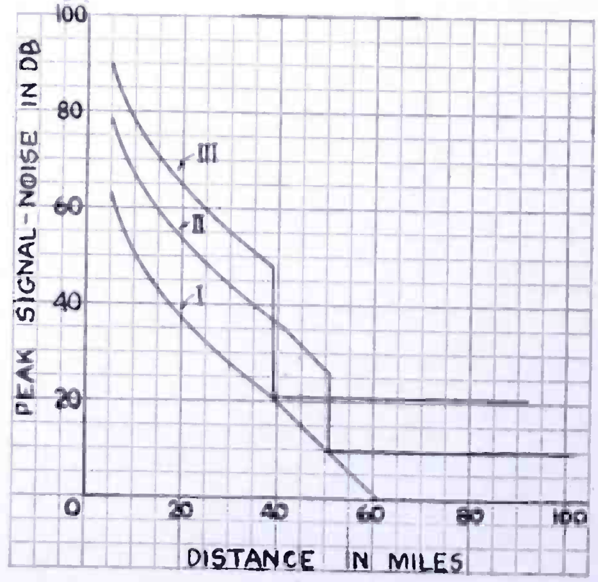
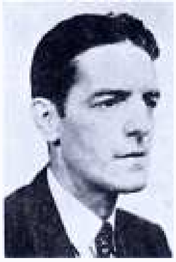


Fig. 10

OUR CONTRIBUTORS

MURRAY G. CROSBY joined the branch of the Radio Corporation of America which is now R.C.A. Communications, Inc., in 1925. He was engaged in the operating and design departments of that branch until 1926, when he took a leave of absence and returned to the University of Wisconsin for one semester and received his degree of B.S. in Electrical Engineering. Since that time he has remained in the research and development division of R.C.A. Communications, Inc. Mr. Crosby is a member of the Institute of Radio Engineers and a fellow of the Radio Club of America.

WILLIAM C. EDDY is a native of Saratoga Springs, N. Y. He was graduated from the United States Naval Academy in 1926, and served as Communications Officer, Submarine Division, from 1929 to 1934. While in that service he specialized in sonic and supersonic detector systems for submarines and developed a binaural integrator for sound detection. He retired from the United States Navy in 1934 and went with the Farnsworth Company as research engineer, later doing transmitter work and studio engineering. In 1937 Mr. Eddy joined the National Broadcasting Company where he since has been engaged as lighting engineer and video effects engineer.



GORDON L. FREDENDALL received the B.S. degree in electrical engineering from the University of Wisconsin in 1931 and the Ph.D. degree from the same institution in 1936. From 1932 to 1936 he held teaching positions in the Department of Electrical Engineering and Mathematics at the University of Wisconsin. Since 1936 Dr. Fredendall has been a member of the Research Division of the RCA Manufacturing Company.

RAY D. KELL received his B.S. degree in electrical engineering from the University of Illinois in 1926. From 1926 to 1930 he was engaged in television research in the radio consulting laboratory of the General Electric Company. From 1930 to the present time he has been a member of the Research Division of the RCA Manufacturing Company, where he has continued his work on various television problems. Mr. Kell is a member of Sigma Xi and an associate member of the Institute of Radio Engineers.

DWIGHT O. NORTH received his B.S. degree from Wesleyan University in 1930 and his Ph.D. degree from the California Institute of Technology in 1933. Since 1934, Dr. North has been with the Research and Engineering Department of the RCA Manufacturing Company at Harrison, N. J., engaged principally in research studies of tube and circuit noise. He is a member of The Institute of Radio Engineers and a member of the American Physical Society.

ALBERT W. PROTZMAN has been connected with radio since 1922. He became Field Supervisor for station WEAF in 1924, and later Field Supervisor for NBC. Mr. Protzman served as Assistant Sound Director for the Fox Film Corporation in Hollywood from 1929 to 1936, when he joined the Television staff of the National Broadcasting Company, where he since has been engaged as a Television Technical Director.

WALTER VAN B. ROBERTS is a graduate of Princeton University. Before the World War he was connected with the Western Electric Company. He became Head of the Department of Radio and Signalling in the School of Military Aviation in June, 1917, and from March, 1918 to the close of the war he was technical officer of Sound Ranging Section Number 1, on the American front. He later taught in Princeton University from 1919 to 1924. In 1924 he joined the RCA Technical and Test Department and in 1927 transferred to the Patent Department of RCA. Dr. Roberts is a fellow of the Institute of Radio Engineers.

ALLEN H. SCHOOLEY was graduated with the degree of B.S. in Electrical Engineering from Iowa State College in 1931. In 1932 he received his M.S. degree from Purdue Unviersity. Between 1932 and 1936 he did radio servicing, was a computor for the United States Coast and Geodetic Survey, and spent a year at the State University of Iowa doing graduate work in engineering and physics. Mr. Schooley joined the RCA Radiotron Division in 1936, and is now an engineer in the Research and Engineering Department of the RCA Manufacturing Company at Harrison, N. J. He is an associate member of I.R.E., and a member of Sigma Xi.

NEWELL R. SMITH received his degree of B.S. in Electrical Engineering from Ohio University in 1926. Following graduation he served a few months with the Cleveland Illuminating Company prior to joining the Vacuum Tube Standardizing Department of the General Electric Company at Nela Park in Cleveland, Ohio. In 1930 this activity was transferred to the Radiotron Division of the RCA at Harrison, N. J. In 1933 he was made assistant in charge of Standardization in which capacity he remained until 1937 when he became associated with the receiving tube design activity.

FRANK E. SPAULDING, JR., attended Yale Engineering School and received his B.S. degree in electrical engineering from the Worcester Polytechnic Institute in 1933. In 1930 he was employed by the Brooklyn Edison Company, and from 1933 to 1935 was with the Underwood Elliot Fisher Company. From 1935 to 1938 he was in the radio receiver engineering department of the General Electric Company at Bridgeport, Conn. Since 1938, Mr. Spaulding has been associated with the Radiomarine Corporation of America, where he is engaged in the development of marine radio equipment. He is an associate member of I.R.E.

GILBERT S. WICKIZER, a native of Pennsylvania, received the B.S. degree in electrical engineering from the Pennsylvania State College in 1926. The same year he joined the Radio Corporation of America, Operating Division, and in 1927 transferred to the Receiver Research and Advanced Development Section of R.C.A. Communications, Inc. Mr. Wickizer is a member of Eta Kappa Nu and an associate member of the Institute of Radio Engineers.

TECHNICAL ARTICLES BY RCA ENGINEERS

Published First Quarter, 1940

BURRILL, CHARLES M.—New Equipment to Measure Intensity of Radio Noise—*Broadcast News*, March.

CARLSON, W. L.—see HARVEY and CARLSON.

CROSBY, MURRAY G.—The Service Range of Frequency Modulation—*RCA Review*, January.

DREW, R. O., and E. W. KELLOGG—Starting Characteristics of Speech Sounds—*Journal Society of Motion Picture Engineers*, January.

FOSTER, D. E.—(Coauthor with C. V. Aggers, Westinghouse Elec. & Mfg. Co., and C. S. Young, Pennsylvania Power and Light Co.) Instruments and Methods of Measuring Noise —*Electrical Engineering*, March.

GOLDSMITH, DR. ALFRED N.—Future Development in the Field of the Projectionist—*Journal Society of Motion Picture Engineers*, February.

———Television Economics, Part XII—*Communications*, January.

———Television Economics, Part XIII—*Communications*, February.

HAEFF, A. V., and L. S. NERGAARD—A Wide-Band Inductive-Output Amplifier—*Proceedings Institute of Radio Engineers*, March.

HARVEY, R. L. and W. L. CARLSON—Polystyrene Applied to Radio Apparatus—*Radio Technical Digest*, January-February.

HEROLD, E. W.—Superheterodyne Converter System Considerations in Television Receivers—*RCA Review*, January.

HOLLY, G. F.—Portable Emergency Sound System—*International Projectionist*, January.

HOLMES, R. S.—Television Reception in an Airplane—*RCA Review*, January.

IAMS, HARLEY—Television Today—*Stanford Illustrated Review*, February.

KELLOGG, E. W.—see DREW and KELLOGG.

KIMBALL, C. N.—see SEELEY and KIMBALL.

KOWALSKI, R. J.—Instrument for Reduction of Flutter—*International Projectionist*, February.

LEDERER, E. A.—Recent Advances in Barium Getter Technique—*RCA Review*, January.

MOORE, H. A.—see J. B. MOORE and H. A. MOORE.

MOORE, J. B., and H. A. MOORE—I-F Selectivity in Receivers for Commercial Radio Services—*RCA Review*, January.

MORRICAL, K. C.—Evaluation of Sound Insulation in Buildings—*Architectural Record*, January.

MORRIS, R. M.—The New Standard Volume Indicator and Reference Level—*Broadcast News*, March.

———(Coauthor with H. A. Chinn, Columbia Broadcasting System, and D. K. Gannett, Bell Telephone Laboratories) A New Standard Volume Indicator and Reference Level—*Proceedings Institute of Radio Engineers*, January.

MORTON, ALFRED H.—NBC Television Activities—*Radio and Television*, March.

MORTON, G. A.—see PIORE and MORTON.

MOUNTJOY, GARRARD—Simplified Television I-F Systems—*RCA Review*, January.

NERGAARD, L. S.—see HAEFF and NERGAARD.

PHELPS, W. D.—Acoustic Line Loudspeaker—*Electronics*, March.

PIORE, E. E., and G. A. MORTON—The Behavior of Willemite Under Electron Bombardment—*Journal Applied Physics*, February.

PREISMAN, ALBERT—Some Unusual Features of Our Television System—*Communications*, January.

ROSE, ALBERT—Electron Optics of Cylindrical Electric and Magnetic Fields—*Proceedings Institute of Radio Engineers*, January.

SARNOFF, DAVID—Vision of Future Developments—*RCA Review*, January.

SEELEY, S. W., and C. N. KIMBALL—A New Method for Determining Sweep Linearity—*RCA Review*, January.

SHELBY, R. E.—A Cathode-Ray Frequency Modulation Generator—*Electronics*, February.

SMITH, T. A.—RCA Television Field Pickup Equipment—*RCA Review*, January.

THOMPSON, B. J.—Fluctuations in Space-Charge-Limited Currents at Moderately High Frequencies; Part I, General Survey—*RCA Review*, January.

WALTER, J. C.—Calculation of T.I.F. for Transmitter Loads—*Broadcast News*, March.

WHITEMAN, R. A.—Application of Abelian Finite Group Theory to Electromagnetic Refraction —*RCA Review*, January.

RCA REVIEW

A Quarterly Journal of Radio Progress

INDEX

Papers and Authors

Volume IV

1939 - 1940

RCA INSTITUTES TECHNICAL PRESS
75 VARICK STREET NEW YORK

INDEX TO VOLUME IV

AUTHORS OF VOLUME IV

RCA

INSTITUTES

INC

TECHNICAL PRESS